Pamela S. Soltis
Douglas E. Soltis *Editors*

Polyploidy and Genome Evolution

Springer

Polyploidy and Genome Evolution

Pamela S. Soltis · Douglas E. Soltis
Editors

Polyploidy and Genome Evolution

Springer

Editors
Pamela S. Soltis
Florida Museum of Natural History
University of Florida
Gainesville, FL
USA

Douglas E. Soltis
Florida Museum of Natural History and
Department of Biology
University of Florida
Gainesville, FL
USA

ISBN 978-3-642-31441-4 ISBN 978-3-642-31442-1 (eBook)
DOI 10.1007/978-3-642-31442-1
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012945474

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Polyploidy (whole-genome duplication; WGD) is common in plants and has long been considered as both an important speciation mechanism and a crucial component of plant genome structure. Analyses of chromosome numbers and hypothesized breaks between diploid and polyploid base numbers have suggested anywhere from 30 to 80 % of all angiosperms are polyploid. While recent polyploids may be easily detected through comparison of chromosome numbers, various processes of diploidization or fractionation may substantially alter chromosome numbers and structure, ultimately masking the evolutionary history of duplication events. In contrast, other footprints of ancient WGD may remain in the genome, even when chromosome numbers no longer carry the signature of past WGDs. Genome sequences and other sources of genomic data tell us that, in fact, all angiosperms, as well as all seed plants, have undergone one or more rounds of polyploidy. Furthermore, ancient WGD characterizes all vertebrates, with subsequent, more recent polyploidization in fishes and amphibians. Ancient WGD is also evident in the genomes of yeast and other fungi. While more common in plants than other major lineages of life, polyploidy is now recognized as a fundamental process in all crown eukaryotes. Polyploidy plays a major role in shaping genome structure and organization and in establishing patterns and mechanisms of gene regulation. In fact, it is now impossible to construct models of genome evolution that do not account for genomic content and genetic interactions contributed by WGD.

It has been over 30 years since the publication of a comprehensive treatment of polyploidy [*Polyploidy: Biological Relevance*, W. H. Lewis (ed.), 1980]. The intervening years have witnessed a technological revolution with a transition from the early days of recombinant DNA to nearly routine genome sequencing of non-model organisms and from limited biological computing to high-performance computing networks for the biological sciences. These transformations in methodology and computation permit fresh perspectives on polyploidy and the ability to ask old questions with new tools.

Over the past decade, it has been a dream of ours to publish a book that synthesizes the rapid progress in understanding the role of polyploidy in genome

evolution, and this book is now a reality. In the current volume, we have compiled the expertise of scientists studying polyploid genome evolution from multiple perspectives in phylogenetically diverse organisms. Topics range from the conceptual and theoretical underpinnings of polyploidy (chapters by McGrath and Lynch, Birchler) to processes at work in polyploid genomes (Zielinski and Mittelsten Scheid, Finigan et al., Evans et al.), to patterns of ancient polyploidy and its detection (Burleigh, Paterson et al.), to a series of case studies that both document attributes of genome evolution in focal species and address general properties of polyploid genomes, from ancient polyploids [maize (Schnable and Freeling), legumes (Doyle), vertebrates (Cañestro), fishes (Braasch and Postlethwait), yeast (Hudson and Conant)] to classic model polyploids [cotton (Wendel et al.), tobacco (Kovarik et al.), wheat (Feldman et al.)] to very recent ones [*Spartina* (Ainouche et al.), *Senecio* (Hegarty et al.), and *Tragopogon* (Soltis et al.)]. The emerging paradigm from these studies is that polyploidy—through alterations in genome structure and gene regulation, some of which occur shortly after polyploid formation—generates genetic and phenotypic novelty that manifests itself at the chromosomal, physiological, and organismal levels, with long-term ecological and evolutionary consequences.

We thank our many colleagues, students, and postdocs for lively and challenging discussions on polyploidy and its many evolutionary consequences. We further acknowledge the support of the U.S. National Science Foundation (Grants 9624643, 0346437, 0614421, 0919254, and 0922003) and thank the National Evolutionary Synthesis Center for its hospitality during the preparation of this book.

Gainesville, April 2012

Pamela S. Soltis
Douglas E. Soltis

Contents

Chapter 1
Evolutionary Significance of Whole-Genome Duplication

C. L. McGrath and M. Lynch

Abstract Whole-genome duplication (WGD) appears to be a widespread phenomenon, occurring in diverse taxa including many of the model organisms used in molecular, cellular, and developmental biology. It is therefore essential to understand the potential evolutionary consequences for individual duplicated genes, as well as for the lineage as a whole. For example, duplicate genes may undergo pseudogenization or may be maintained due to neofunctionalization, subfunctionalization, or selection for increased dosage or dosage balance. Duplicates created via WGD are maintained at higher rates than single-gene duplicates, perhaps due to dosage-balance constraints. Duplicate-gene maintenance may lead to heterodimerization of an existing homodimer or to the divergence of an entire duplicated network or pathway. Allopolyploids and autopolyploids are likely to undergo different evolutionary pressures due to increased divergence between allopolyploid paralogs and an increased prevalence of multivalent formation at meiosis in autopolyploids. Perhaps most importantly, duplicate-gene loss following a WGD may significantly increase the rate of reproductive isolation between geographically isolated subpopulations and may therefore temporarily increase the speciation rate within polyploid lineages.

1.1 Introduction

One of the major findings of the new field of evolutionary genomics is that duplication events involving individual genes or multigene segments arise at rates comparable to the rate of mutation at single-nucleotide sites (Lynch and Conery

C. L. McGrath · M. Lynch (✉)
Department of Biology, Indiana University, Bloomington, IN, USA
e-mail: milynch@indiana.edu

P. S. Soltis and D. E. Soltis (eds.), *Polyploidy and Genome Evolution*,
DOI: 10.1007/978-3-642-31442-1_1,

2000, 2003a, b), or possibly at even higher rates (Lipinski et al. 2011). Such observations lend credibility to Ohno's (1970) early speculation that gene duplication is a major resource for the origin of evolutionary novelties. Moreover, it is now clear that whole-genome duplication (WGD) events have occurred in a wide diversity of phylogenetic lineages, including most of the model systems relied upon in molecular, cellular, and developmental biology. For example, budding yeast is a descendant of an ancient genome duplication (Wolfe and Shields 1997; see Chap. 15, this volume), as is the frog *Xenopus laevis* (Morin et al. 2006; see Chap. 18, this volume) and the zebrafish (Postlethwait et al. 2000; see Chap. 17, this volume). Many ray-finned fish lineages have experienced additional rounds of WGD (Meyer and Van de Peer 2005; see Chap. 17, this volume), and Ohno's (1970) suggestion that two WGD events preceded the radiation of the vertebrate lineage has become increasingly credible (Panopoulou and Poustka 2005; Hughes and Liberles 2008; Putnam et al. 2008; see Chap. 16, this volume). Finally, three WGD events are recorded within the genome of *Arabidopsis thaliana* (Simillion et al. 2002), and nearly all other land-plant genomes appear to harbor a legacy of at least one polyploidization event (Doyle et al. 2008), with a proposed WGD in the ancestor of all seed plants and another in the ancestor of all angiosperms (Jiao et al. 2011). Thus, it is clear that understanding the mechanisms of origin and preservation of duplicate genes promises to reveal not only the ways in which genes acquire new functions and organisms respond to natural selection, but also the roots of organismal diversity across the tree of life.

Because genome duplication adds thousands of duplicate genes to the genome, understanding the evolutionary forces that act on individual duplicate genes is critical to our understanding of polyploidization. Processes such as neofunctionalization and subfunctionalization have the potential to influence all gene duplicates, whether created through polyploidization or smaller scale duplication events. It has become increasingly clear, however, that duplicates that arise via polyploidization are subject to unique evolutionary forces, such as increased retention due to dosage-balance constraints. Further, there may be processes that are exclusive to gene duplicates that arise via specific types of polyploidization, such as changes in duplicate-gene expression due to the genomic merger that occurs with allopolyploidization. The relative contributions of these evolutionary forces that give rise to the maintenance and evolution of duplicate genes that arise via WGD, or to the evolution of the genome or species as a whole, are currently unknown. However, discriminating between these forces and their effects is likely to be the subject of much research over the next several years.

1.2 Fates of Duplicate Genes

The fate of the vast majority of duplicate genes arising by segmental duplication is nonfunctionalization of one member of the pair (Lynch and Conery 2000, 2003a, b), and this is expected to occur within a few million years in the absence of any

intrinsic advantage of a duplicate copy (Watterson 1983; Lynch et al. 2001). Despite this, most genomes that have been studied contain a large number of duplicate genes, some of which are clearly quite ancient (Lynch and Conery 2000). Based on this observation, several mechanisms have been proposed for the permanent preservation of duplicate genes (Hughes 1994; Force et al. 1999; Lynch et al. 2001; Taylor and Raes 2004; Lynch 2007; Innan and Kondrashov 2010): (1) neofunctionalization, whereby one copy acquires a novel, beneficial function at the expense of an essential ancestral function; (2) subfunctionalization, whereby complementary mutations lead to a partitioning of independently mutable subfunctions in the ancestral gene; (3) selection for increased gene product; and (4) the masking of nonfunctional alleles.

When a duplicate is maintained by selection for increased gene product, it experiences purifying selection (and may also undergo repeated gene conversion) in order to maintain its ancestral function; this process is likely responsible for the multiple copies of ribosomal RNA genes present in many genomes (e.g., Pinhal et al. 2011). Neofunctionalization, on the other hand, is thought to involve positive selection for the mutation(s) responsible for the new function, generally arising at the expense of an essential original function, thereby preserving both copies. There are many examples of neofunctionalization giving rise to novel gene functions in a variety of organisms, including *Arabidopsis* (Erdmann et al. 2010), fish (Ngai et al. 1993), vertebrates (Layeghifard et al. 2009), and yeast (Byrne and Wolfe 2007; Tirosh and Barkai 2007). Because one duplicate is undergoing positive selection for a new function while the other is under purifying selection to maintain the ancestral function, asymmetric evolutionary rates between duplicates are often thought to be a hallmark of neofunctionalization (Johnson and Thomas 2007; Han et al. 2009), though purely stochastic mechanisms can also give rise to apparent rate asymmetry (Lynch and Katju 2004).

Subfunctionalization may involve positive selection acting on both duplicates if the partitioning of the ancestral functions leads to relaxation of pleiotropic constraints, enabling each ancestral function to be fine-tuned and improved through mutation independently in each copy (Piatigorsky and Wistow 1991; Hughes 1994; Des Marais and Rausher 2008). Alternatively, subfunctionalization may be a completely neutral process if each duplicate copy simply acquires a degenerative mutation that renders it unable to perform one of the ancestral functions (Force et al. 1999). At this point, both copies are needed in order to provide the organism with all of the functionality of the original, single-copy gene, and so both will be maintained in the genome by selection. Although identifying definitive cases of subfunctionalization requires determining that the ancestral gene carried multiple functions that have been partitioned in the daughter duplicates, there are nonetheless several compelling examples (e.g., Force et al. 1999; Altschmied et al. 2002; Yu et al. 2003; Adams and Liu 2007; MacNeil et al. 2008; Semon and Wolfe 2008; Buggs et al. 2010; Deng et al. 2010; Hickman and Rusche 2010; Colon et al. 2011; Froyd and Rusche 2011).

In addition to these cases of qualitative subfunctionalization, where duplicates eventually come to be expressed in different tissues or at different times or carry out

different functions relative to each other, quantitative subfunctionalization, in which reduction-of-expression (Force et al. 1999) or activity-reducing mutations (Stoltzfus 1999; Scannell and Wolfe 2008) affect both duplicates, is also possible. In quantitative subfunctionalization, both duplicates acquire partial loss-of-function mutations that affect the same function, again rendering both copies essential for the proper dosage or activity of the gene products. In this case, both copies are preserved while retaining the ancestral gene function. Although few studies demonstrating quantitative subfunctionalization exist, Qian et al. (2010) estimated that this process has been responsible for the maintenance of a large proportion of duplicates in yeast and mammals, whereas Woolfe and Elgar (2007) postulated that sequence evolution in *cis*-regulatory elements may have caused quantitative subfunctionalization among *Fugu* duplicates.

A related consequence of gene duplication is that it can allow for the differentiation of multimeric subunits, such as the evolution of heterodimers from homodimers. Consider a gene whose protein product forms a homodimer. After duplication of this gene, protein subunits produced by the two duplicates (denoted *A* and *B*) may randomly associate to make mixtures of dimers in the ratio 1 *AA*: 2 *AB*: 1 *BB*. If the duplicate genes are identical initially, the *AA*, *AB*, and *BB* dimers will be identical as well. However, subsequent differentiation of the duplicate genes causes the three types of dimer to become distinct. This differentiation could be neutral, or it could be selective. If, for example, there were pleiotropic constraints on the form or function of the pre-duplication homodimer, duplication could allow for escape from these constraints in the *AB* heterodimer, as each subunit (*A* and *B*) can now evolve independently. This can be viewed as a special type of subfunctionalization of duplicates. Winter et al. (2002) showed that a class-B floral protein heterodimer had evolved from an ancestral homodimer via this mechanism during the gymnosperm/angiosperm transition. In gymnosperms, GGM2-like genes form homodimers, while the duplicated homologs in eudicots, DEF-like genes and GLO-like genes, form heterodimers. Monocots also have duplicated DEF-like genes and GLO-like genes, but, interestingly, it appears the GLO-like proteins of monocots can both homodimerize and heterodimerize with DEF-like proteins, perhaps representing the transition between the homo- and heterodimerized states (Winter et al. 2002; Kanno et al. 2003; Soltis et al. 2006).

A similar process appears to have occurred several times in the evolution of the DUF606 family of transmembrane proteins in bacteria (Lolkema et al. 2008). In bacteria with a single DUF606 gene, the DUF606 proteins are able to insert into the membrane in both orientations, and functional homodimers are formed by two subunits in opposite (antiparallel) orientations. Other species of bacteria, however, have duplicated DUF606 genes located tandemly in an operon. In all of these latter cases, the two protein subunits each have a fixed but opposite orientation in the membrane, and they heterodimerize to form the necessary antiparallel two-domain complex. A phylogenetic analysis of the DUF606 gene family reveals that this process of duplication followed by heterodimerization likely occurred five different times in the history of this gene family lineage (Lolkema et al. 2008). Other proposed examples of this mechanism include SMC proteins (Surcel et al. 2008),

adenylyl cyclases (Sinha et al. 2005), and mitochondrial peptidases (Brown et al. 2007), all gene families that contain duplicates that form heterodimers in eukaryotes (or eukaryotic mitochondria) with single-copy homologs that form homodimers in prokaryotes.

1.3 Fates of Duplicate Genes Arising via WGD

In addition to the general preservational processes just mentioned, paralogs resulting from WGD events are subject to unique mechanisms of duplicate-gene maintenance and evolution (Force et al. 1999; Lynch and Conery 2000; Yang et al. 2003; Davis and Petrov 2005; Veitia et al. 2008). Well-studied polyploid species commonly exhibit 25–75 % retention of paralogous gene pairs from the most recent WGD event (reviewed in Lynch 2007; Otto 2007), budding yeast being an exception with only ~8 % duplicate-gene preservation (Wolfe and Shields 1997). These are surprisingly high preservational levels, when, as discussed above, the fate of the vast majority of duplicate genes arising by segmental duplication is nonfunctionalization of one duplicate (Lynch and Conery 2000, 2003a, b). Although it is possible that many polyploid species have not yet reached equilibrium and are still in an ongoing phase of duplicate-gene loss, it has become increasingly clear that there are likely to be additional forces acting to preserve duplicate genes arising via WGD.

A simple explanation for the large number of preserved duplicates within polyploids is that, unlike single-gene duplicates, WGD duplicates exhibit complete conservation of surrounding regulatory sequences, chromosomal environments, etc. Although this likely contributes somewhat to the pattern of higher duplicate retention in polyploids, it does not explain the observation that different types of genes seem to be preserved following WGD compared to smaller scale duplications. This fact can be better explained by selection for dosage balance among proteins. Due to stoichiometric relationships with other interacting genes (e.g., multi-subunit complexes and numerous pathways involved in metabolism and transcriptional regulation), the functions of a subset of protein-coding loci can be highly influenced by dosage imbalances (Veitia 2002; Papp et al. 2003; Birchler et al. 2005; Veitia et al. 2008). In such cases, duplication of just a single member of a gene interaction may be detrimental and actively selected against. In contrast, following a WGD event, most stoichiometric relationships are initially intact, and therefore subsequent losses of interacting paralogs will be inhibited by selection for proper dosage relationships. Thus, for dosage-dependent genes, the dosage-balance hypothesis predicts an under-representation among duplicates created by single-gene duplications, but an over-representation among those created by WGD (Yang et al. 2003; Davis and Petrov 2005; Veitia et al. 2008). For example, Davis and Petrov (2005) showed that the pool of preserved duplicates from the WGD event in *S. cerevisiae* is enriched for ribosomal genes (which form a large complex) and regulatory genes encoding transferases, kinases, and transcription factors, while

those involved in ion transport are under-represented. Likewise, the *Paramecium tetraurelia* genome exhibits elevated retention of duplicate genes involved in known complexes (Aury et al. 2006) and in metabolic pathways (Gout et al. 2009). As in yeast, ribosomal genes, transferases, and kinases are over-represented among surviving paralogs, while ion-transport genes are underrepresented. In *Paramecium*, there also appears to be an additional effect whereby highly expressed genes are over-retained in duplicate following the most recent polyploidization event (Gout et al. 2010). That certain types of genes are maintained preferentially following a WGD has achieved fairly convincing empirical support from other studies as well (Papp et al. 2003; Yang et al. 2003; Barker et al. 2008; Liang et al. 2008; Qian and Zhang 2008; Edger and Pires 2009), including studies in *Arabidopsis* (Blanc and Wolfe 2004; Maere et al. 2005; Thomas et al. 2006), vertebrates (Makino and McLysaght 2010), and across divergent species (Paterson et al. 2006). Selection to maintain dosage balance following WGD has also been hypothesized to be the driving force behind the original selective advantage of the WGD in the *Saccharomyces cerevisiae* lineage (Conant and Wolfe 2007). In this scenario, the maintenance of glycolytic genes and the loss of non-glycolytic genes following WGD might have increased the relative dosage of glycolytic genes, thereby increasing flux through the glycolysis pathway and providing polyploid yeast with a growth advantage over non-polyploids due to increased glucose fermentation ability.

Duplicate genes that arise via WGD are further unique in that entire (or partial) duplicated pathways or networks of interacting proteins can diverge in concert. For example, Evlampiev and Isambert (2007) modeled the evolution of protein–protein interaction networks following WGD and concluded that such networks grow under exponential, rather than time-linear, dynamics following WGD. Interestingly, they also found that these exponential dynamics relied on asymmetric divergence between duplicates.

Another intriguing possibility is that following WGD, a whole ancestral network may become neofunctionalized or subfunctionalized following polyploidization, with one set of paralogs carrying out one task or reaction and a parallel set of paralogs carrying out a related, but largely independent, task. Obviously, such innovations require the establishment of multiple mutations and the avoidance of pathway crosstalk. Although the essential population genetic theory remains to be worked out, several examples of such paralog coevolution appear to have followed the WGD in yeast: parallel paralogous networks have been identified where the expression of each gene is highly correlated with the other genes within its network but poorly correlated with its paralog (Blanc and Wolfe 2004; Conant and Wolfe 2006). In this way, polyploidy provides a unique mechanism for the evolution of gene networks with new (or subdivided) functions.

A final consideration in duplicate-gene evolution is whether the forces that act to preserve duplicates change over evolutionary time. For example, it seems possible that following WGD, a large proportion of genes could be initially maintained due to dosage-balance constraints. Subsequently, however, over longer periods of evolutionary time, some duplicates might accumulate mutations that could lead to neofunctionalization or subfunctionalization. Because these genes

are dosage sensitive (hence their initial preservation due to selection for dosage balance), it is likely that such neo- or subfunctionalizing mutations would need to be preceded or rapidly followed by mutations affecting the dosage of one or both copies. For a more detailed example, imagine proteins *A* and *B* that must interact in a 1:1 ratio for proper functioning. Both genes become duplicated during a WGD, giving rise to duplicates *A1* and *A2* and *B1* and *B2*. Initially, all four genes are preserved by selection for dosage balance, as loss of any one gene interrupts the 1:1 interaction ratio. Over evolutionary time, however, slightly deleterious mutations in the *A1* promoter that decrease its expression level become fixed due to drift. To compensate, mutations in the *A2* promoter that increase its expression level are fixed, which helps to restore the 1:1 *A*/*B* ratio. At this point, *A1* is contributing fewer products to the overall *A* protein pool. A subsequent mutation that changes the function of *A1*, allowing it to take on a new role completely, is now more easily accommodated, as *A2* is better able to compensate and take on the full load of the ancestral *A* activity. Note that, instead of *A1* and *A2* dosage evolving in concert, as above, *A1* and *B1* dosage could also evolve in concert to maintain the proper 1:1 *A*/*B* ratio, allowing both *A1* and *B1* to take on new functions.

While still just a verbal theory, this scenario has two advantages in terms of allowing for neofunctionalization (or subfunctionalization) of WGD duplicates. First, there is a longer time frame in which neo- or subfunctionalizing mutations can arise, as duplicates are maintained for longer time-scales without becoming nonfunctionalized. This is important because neofunctionalization requires the accumulation of beneficial mutation(s), which are thought to be rare. Second, this process would allow for neo- or subfunctionalization of dosage-sensitive duplicates, both of which might otherwise be constrained to maintain their ancestral function indefinitely following WGD.

1.4 Autopolyploidy Versus Allopolyploidy

Whether polyploidization occurs by autopolyploidy or allopolyploidy can have a significant impact on the expression and evolution of duplicate genes. Autopolyploids arise when there is an increase in ploidy within a single species (often within a single individual), while allopolyploids are created by hybridization between two different species, each of which contributes a full complement of chromosomes to the hybrid, thus doubling the genome (reviewed in Coyne and Orr 2004). Many plant and frog polyploids are the result of allopolyploidization (Adams 2007; Evans 2008), while the yeast WGD appears to have been an autopolyploidization event (Scannell et al. 2007), although in practice it is difficult to ascertain the ancestral state once paralog divergence has become high.

It has long been assumed that autopolyploids would initially form multivalents at meiosis, with all four homologous chromosomes pairing randomly, while allopolyploids would be more likely to form bivalents, with homologous

chromosomes from each diploid ancestor pairing independently. This would mean that duplicate copies in autopolyploids would not represent true paralogs as the term is usually understood, but would instead represent a doubling of the number of homologs (i.e., four homologs instead of two). The presence of multivalents is significant biologically, as multivalent pairing can lead to intergenomic recombination via segregation, crossing-over, and double reduction. Certain duplicates from one diploid parent could be lost completely via this process, leaving only duplicates from the other diploid parent. This would not represent gene silencing as it is typically understood then, but would rather be a byproduct of multivalent formation and segregation. Evidence from plants indicates that multivalent pairing is indeed more prevalent among autopolyploids, though the difference between the two forms of polyploidy is perhaps less than originally expected: a survey of plant polyploids indicated that the mean percent occurrence of multivalents is 28.8 % in autopolyploids and 8.0 % in allopolyploids (Ramsey and Schemske 2002). Although multivalent formation occurs at a lower rate in allopolyploids, it may be more biologically significant than multivalent formation in autopolyploids, as intergenomic recombination is likely to have a greater effect when genomes are more divergent. Over time, divergence between duplicated chromosomes would lead to increased bivalent formation.

Because allopolyploids are the result of a genomic merger between two species, duplicate genes in allopolyploids are already differentiated to some extent immediately after polyploidization, while duplicates in autopolyploids are likely to be more similar in sequence and may even be identical. Allopolyploids often exhibit immediate changes in gene expression due to the genetic differentiation present between homeologs. This can lead to changes in methylation (Salmon et al. 2005; Gaeta et al. 2007), changes in heterochromatin formation and transposable element suppression (Josefsson et al. 2006), biased expression of homeologs (Adams et al. 2003; Bottley et al. 2006; Tate et al. 2006; Udall et al. 2006; Rapp et al. 2009), and non-additive expression effects between homeologs (Hegarty et al. 2006; Wang et al. 2006; Rapp et al. 2009). These initial expression differences between homeologs can, in turn, impact the long-term evolution of duplicates, as selection pressures may be expected to act differently on genes that are differentially expressed. For example, Anderson and Evans (2009) showed that in octoploid and dodecaploid *Xenopus* species, paralogs of RAG1β were more likely to become pseudogenized than paralogs of RAG1α (the homeolog of RAG1β from an earlier allopolyploidy event), and they inferred that this was due to differences in ancestral expression between RAG1α and RAG1β.

Many of these effects seen in allopolyploids are believed to be due to the hybridization between two divergent genomes, rather than genome doubling *per se*. Flagel et al. (2008) estimated that of the genes with biased expression between homeologs in the allopolyploid *Gossypium hirsutum*, 24 % exhibit a bias due to the genomic merger (i.e., the bias existed immediately when the allopolyploidization occurred, at time zero), while the bias in the remaining 76 % is due to long-term evolutionary forces such as neofunctionalization and subfunctionalization. The relationship between the magnitude of these alterations in gene expression and the

genetic divergence between the two parental genomes is not well understood, however, as demonstrated by *Brassica* allopolyploids (Pires and Gaeta 2011). While the parental species that gave rise to the allopolyploid *Brassica napus* are more similar to each other than those that gave rise to *B. juncea*, resynthesized *B. napus* polyploids exhibit more genomic rearrangements, changes in gene expression, and epigenetic alterations than do resynthesized *B. juncea* polyploids.

1.5 Polyploidization and Speciation

Perhaps the most pivotal role that polyploidization plays in evolution is in the creation of new species. The polyploidization event itself can lead to instantaneous reproductive isolation and speciation, as the cross between a new tetraploid ($4n$) and its diploid progenitor ($2n$) yields triploid ($3n$) offspring, which are often sterile due to problems with chromosome pairing/segregation during meiosis and the production of aneuploid gametes (reviewed in Coyne and Orr 2004). It is for this reason that models predict that species capable of self-fertilization are more likely to give rise to a successful polyploid lineage (Rodriguez 1996; Baack 2005; Rausch and Morgan 2005).

Perhaps more importantly, however, once a polyploid lineage is established, subsequent silencing of duplicate genes can lead to further reproductive isolation among subpopulations of the polyploids themselves and, therefore, give rise to additional daughter species (Oka 1988; Werth and Windham 1991; Lynch and Conery 2000; Lynch and Force 2000). In this model, we assume a pair of fully functional and redundant duplicate genes, *A* and *B*, in an ancestral population, such that each member of the initial population has the genotype *AABB* (Fig. 1.1). If two subpopulations become geographically isolated and one duplicate becomes nonfunctionalized in each subpopulation, there is a 50 % probability that a different duplicate copy will be lost in each of the two groups. This reciprocal gene loss (or divergent resolution) would result in the genotypes *aaBB* and *AAbb* for the two subpopulations, where *a* and *b* denote null alleles. Hybridization between the two groups would then lead to offspring with the genotype *AaBb*. Gametes produced by these F_1 individuals would have a 1/4 probability of carrying an *ab* genotype and would therefore be inviable if a functional copy of the *A/B* gene were essential for gamete survival or function. Even if this were not the case, 1/16 of the F_2 individuals would have the *aabb* genotype and, if a functional copy were essential for zygote viability or sterility, would be inviable or sterile, whereas another 1/4 would have three null alleles and might experience reduced viability or sterility. Up to 50–65 % of the genes encoding transcription factors, membrane receptors, and members of macromolecular protein complexes are estimated to be haploinsufficient (Jimenez-Sanchez et al. 2001; Veitia 2002), suggesting that only one functional allele of such genes is indeed likely to be deleterious.

An appealing aspect of the divergent resolution model is that it is a natural consequence of degenerative mutations, requiring no adaptive evolution at the

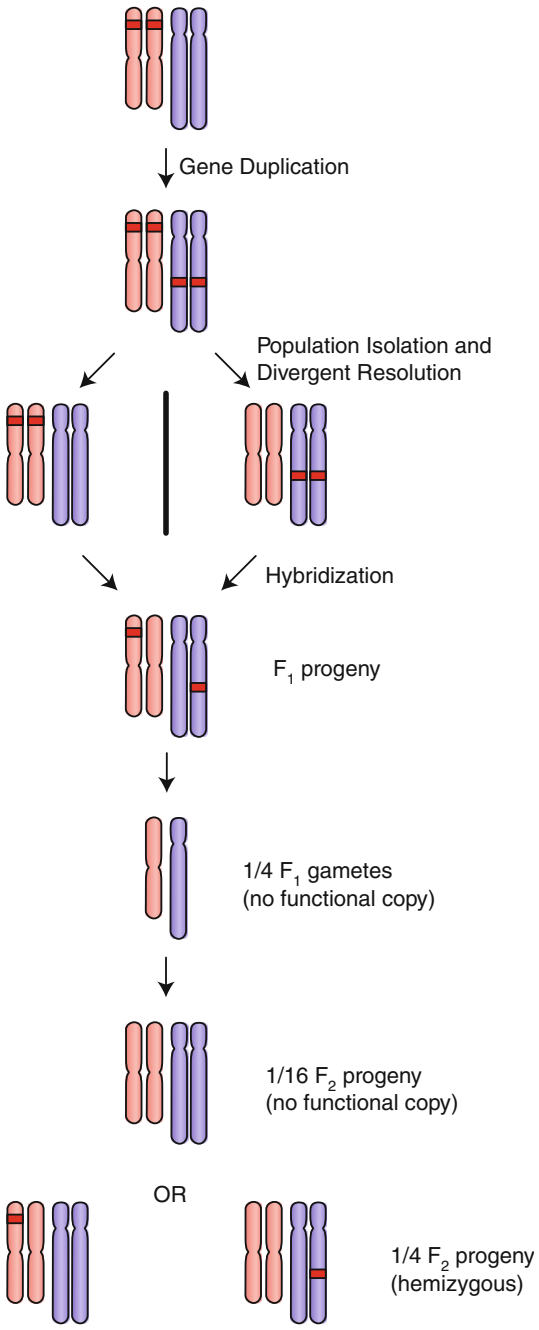

Fig. 1.1 Divergent resolution of duplicate genes can lead to hybrid sterility/inviability and reproductive isolation. *Red bar* represents a gene that becomes duplicated. See text for details

molecular level for speciation to occur. Moreover, in addition to genes that are reciprocally silenced, duplicate pairs that undergo neofunctionalization or subfunctionalization may also contribute to hybrid sterility/inviability in a similar fashion—for example if a different duplicate becomes neofunctionalized and loses its ancestral function in each subpopulation, or if the two duplicates become subfunctionalized in complementary ways in the two subpopulations (Lynch and Force 2000).

The process of reciprocal gene loss has been shown to be responsible for male sterility between hybrids of *Drosophila melanogaster* and *D. simulans* (Masly et al. 2006). A gene essential for male fertility, *JYAlpha*, is located on the fourth chromosome in *D. melanogaster* and on the third chromosome in *D. simulans*. This translocation presumably occurred via duplication of the *JYAlpha* gene and subsequent silencing of one copy. The difference in chromosomal location of the gene in the two species causes a proportion of hybrids to completely lack *JYAlpha*, leading to their sterility.

Two similar cases were recently identified in rice. The first involves reproductive isolation between two subspecies of *Oryza sativa*. The ancestral *O. sativa* genome appears to have had a pair of duplicates termed *DOPPELGANGER1* (*DPL1*) and *DOPPELGANGER2* (*DPL2*) (Mizuta et al. 2010). The subspecies *japonica* and *indica* have experienced independent losses of one copy each: *DPL1* has become a pseudogene in *indica*, while *DPL2* has been nonfunctionalized in *japonica*. Hybrid pollen lacking a functional copy of either *DPL1* or *DPL2* is nonfunctional and does not germinate, contributing to the partial reproductive isolation present between the subspecies. This validates an earlier hypothesis by Oka (1988) that F_1 sterility between *japonica* and *indica* was caused by "duplicate gametophytic sterility genes", *japonica* being homozygous for one nonfunctional copy and *indica* being homozygous for another nonfunctional copy. In the second rice example, reciprocal loss of one of the duplicated nuclear genes encoding mitochondrial ribosomal protein L27 in *O. sativa* and *O. glumaepatula* again causes a proportion of the pollen produced by F_1 hybrids to be sterile (Yamagata et al. 2010).

The final example of reproductive isolation through reciprocal gene loss comes from *A. thaliana*, where the histidinol-phosphate amino-transferase gene appears in different chromosomal locations (as in the *Drosophila* example, presumably via duplication and subsequent silencing of one copy) in the Columbia and Cape Verde Island accessions (Bikard et al. 2009). F_2 offspring homozygous for both null alleles completely lack the gene's product, HPA, which results in arrested embryo development and seed abortion. In addition, in at least one intermediate heterozygote, a quantitative phenotype termed "weak root" was observed, suggesting that the presence of three null alleles is somewhat deleterious in this cross. As these four examples constitute $\sim 1/3$ of the dozen or so successful searches for the genes underlying the speciation process (most in *Drosophila* species; Presgraves 2010), there now seems little question that the passive nonfunctionalization of duplicate genes is a major mechanism of speciation.

These examples demonstrate that the divergent resolution of even one duplicated gene can lead to detectable reproductive isolation. However, genetic incompatibility between two populations can be magnified substantially when reciprocal gene loss occurs at hundreds or thousands of duplicated loci simultaneously, as is the case in polyploid lineages. The probability that an F_2 offspring obtained by outcrossing will be double null for at least one of n pairs of divergently resolved loci is $1-(15/16)^n$, which takes on values of 0.063, 0.276, 0.476, and 0.998 for $n = 1$, 5, 10, and 100, respectively. Moreover, in species that undergo autogamy or selfing, such as *Paramecium,* this probability can be as high as $1-(3/4)^n$, giving probabilities of 0.250, 0.763, 0.944, and ≈ 1 for $n = 1$, 5, 10, and 100. Speciation events will continue to occur as long as duplicates are still being resolved between subpopulations, leading to nested rounds of speciation, and, because a large number of duplicates are thought to be silenced quickly following WGD (Scannell et al. 2006), a cluster of speciation events might occur within a brief period of time. The net result is the expected generation of a species radiation following a WGD event.

It has been suggested that this nested speciation process might be responsible for the radiations of the polyploid yeast species (Scannell et al. 2006), teleost fishes (Semon and Wolfe 2007; see Chap. 15, this volume), angiosperms (Soltis et al. 2009; though see Mayrose et al. 2011), and the *Paramecium aurelia* species complex (Aury et al. 2006). This mechanism may also be responsible for reproductive isolation between mutagenized lines of an experimentally derived allotetraploid created by hybridizing two species of *Saccharomyces* (Maclean and Greig 2010).

1.6 Unsolved Problems

The maintenance of duplicate genes via selection for increased gene product, neofunctionalization, and subfunctionalization has been hypothesized for nearly 40 years (Ohno 1970). Recent genetic and genomic data have now identified compelling examples of these processes and have further contributed to our understanding of the prevalence of whole-genome duplications and the dosage-balance theory of duplicate maintenance. However, a number of unresolved questions related to WGDs and duplicate maintenance merit further scrutiny.

The first avenue for future study involves a more comprehensive understanding of the relative importance of the forces behind duplicate-gene maintenance, including maintenance for increased dosage, dosage-balance constraints, neofunctionalization, and subfunctionalization. All of these mechanisms have been demonstrated to be responsible for duplicate maintenance in certain cases, but it remains unclear which, if any, is responsible for maintaining the majority of duplicate genes or how such contributions vary among phylogenetic lineages. Most likely, there will be no single driving force for duplicate maintenance but the relative strength of these forces will differ among taxonomic groups or among functional classes of genes. For example, subfunctionalization of duplicates may be more

likely within species that have evolved a modular (and therefore independently mutable) regulatory structure. Such modular systems are predicted to arise more easily within species with smaller population sizes (Force et al. 2005), demonstrating how species-level features of an organism may influence the evolutionary forces acting upon duplicate genes. More comprehensive studies of large numbers of duplicates from a variety of organisms are required to address what other features might influence the relative strengths of mechanisms of duplicate maintenance. Such studies must not only detail what genes remain duplicated vs. single-copy, but must also detail whether existing duplicates have the same function as each other (to assess rates of neofunctionalization) or share functions with the pre-duplicated ancestor (to assess rates of subfunctionalization) (Fig. 1.2).

There are few data on the rate of duplicate-gene loss over time following a WGD, though data from polyploid yeast species suggest that the rate changes over time (Scannell et al. 2006). Data from additional taxa would aid in determining whether this is a general pattern among all WGDs (Fig. 1.3). A related unresolved question is whether the evolutionary forces controlling duplicate maintenance change over time following a WGD, e.g., whether a dosage-sensitive gene may initially be preserved due to selection for dosage balance, but then evolve a new function concurrent with its release from such dosage constraints. An analysis of this question could be made by comparing the fates of duplicate genes in multiple lineages descended from a single WGD event. Such an analysis might identify duplicates that had been maintained due to dosage constraints in the majority of daughter lineages but that had become neofunctionalized in one lineage, perhaps suggesting a secondary mechanism of retention.

The unsolved question that promises to be the hardest to answer is why certain lineages or taxonomic groups appear to contain more WGD events than others. This is not the same as asking why certain groups contain more polyploid species, as this may be a simple reflection of the fact that WGD may promote subsequent reproductive isolation and speciation. The pattern remains, however, that some phylogenetic groups seem to contain more independent WGD events than others in their evolutionary pasts. For example, a recent analysis estimated that among ferns, 31 % of speciation events involve polyploidization, while the value for angiosperms is only 15 % (Wood et al. 2009). Similarly, in the history of the *Xenopus* lineage, there are many more instances of WGD events than compared to, say, mammals. Several factors could contribute to such patterns, such as the ability to hybridize and form allopolyploids, or the ability to self-fertilize (at least transiently) or undergo asexual reproduction, which helps a polyploid lineage become abundant in a surrounding world of diploids. It is not even understood whether mechanistic reasons (at meiosis, say) or differences in developmental programs would facilitate or hinder creation of a viable polyploid in certain lineages, or whether discrepancies in ecological persistence of polyploid species alone are able to explain the patterns that we see. Perhaps the best way to approach such a question is to study closely related lineages where one exhibits several WGD events and the other does not, though teasing apart the mechanistic and ecological differences is certain to remain a challenge for decades to come.

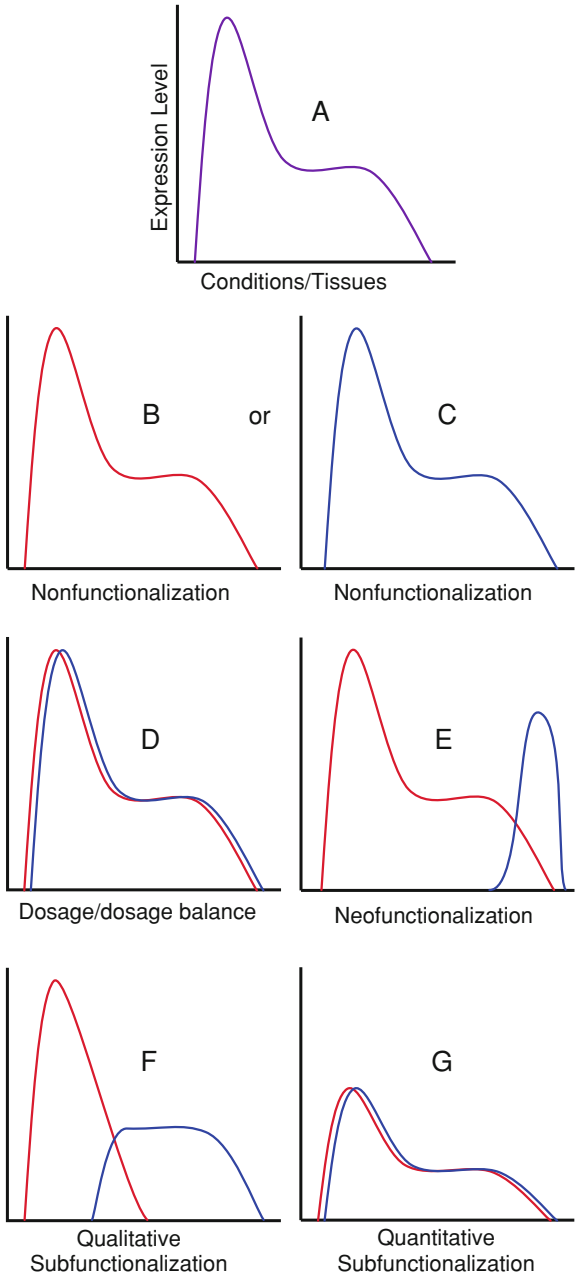

Fig. 1.2 Distinguishing between the evolutionary forces that maintain duplicate genes. Panel A shows the expression level of a gene (*purple*) across different conditions or tissues before duplication. The *bottom* six panels (B–G) show patterns that might be seen for the two copies (*red* and *blue*) once the gene has been duplicated and what evolutionary processes these patterns would indicate. Note that panel D might indicate maintenance for increased dosage in the case of a single-gene duplicate or maintenance for dosage balance in the case of a duplicate arising via WGD

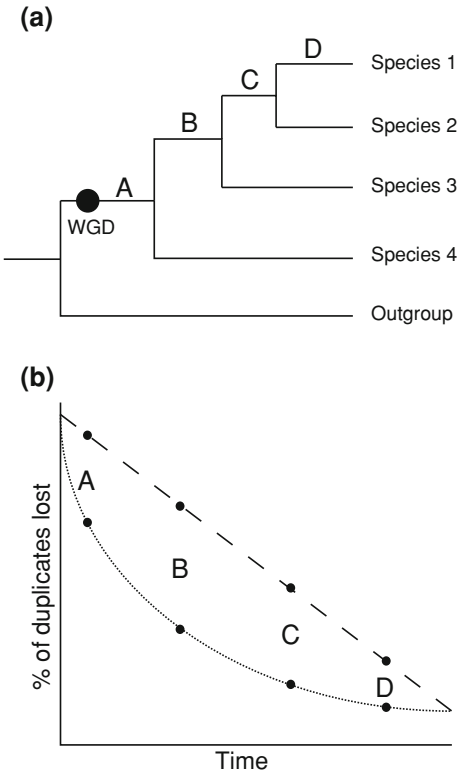

Fig. 1.3 Determining the rate of duplicate-gene loss over time. **a** An example of a tree for four species that share a whole-genome duplication (*WGD*). Duplicate-gene presence/absence information for each of the four species could be used to infer the number of duplicate genes lost on each of the branches labeled *A*, *B*, *C*, and *D*. **b** The data on gene retention/loss gathered in (**a**) could be used to plot the percentage of duplicate genes lost per unit time (or divergence). Depending on where points *A*, *B*, *C*, or *D* land on the graph, the data may indicate that the rate of gene loss remains constant (*top dashed line*) or changes (*bottom dotted line*) over time

1.7 Conclusions

Whole-genome duplications are widespread across the tree of life and appear in the evolutionary history of a large number of model organisms. Processes such as neo- and subfunctionalization affect retention of individual gene duplicates, and dosage-balance constraints promote the retention of large sets of genes following polyploidization. Allopolyploidization, through hybridization and subsequent changes or biases in homeolog expression, has the ability to instantaneously create a population of individuals that are ecologically and epigenetically unique from either parent lineage, providing a new lineage upon which natural selection can act. Both allo- and autopolyploidization provide a unique opportunity for the differentiation of new gene networks and pathways through concerted evolution of duplicated, interacting proteins. Most importantly, WGD can lead to reproductive isolation through divergent resolution of duplicated genes, thus creating new species and species groups. Further understanding of the relative importance and the temporal properties of the forces acting on polyploid species and the duplicate genes within their genomes promises to enhance our knowledge of the origins of species as well as genetic, protein network, and organismal complexity.

References

Adams KL (2007) Evolution of duplicate gene expression in polyploid and hybrid plants. J Hered 98(2):136–141

Adams KL, Cronn R et al (2003) Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proc Nat Acad Sci U S A 100:4649–4654

Adams KL, Liu Z (2007) Expression partitioning between genes duplicated by polyploidy under abiotic stress and during organ development. Curr Biol 17(19):1669–1674

Altschmied J, Delfgaauw J et al (2002) Subfunctionalization of duplicate mitf genes associated with differential degeneration of alternative exons in fish. Genetics 161(1):259–267

Anderson DW, Evans BJ (2009) Regulatory evolution of a duplicated heterodimer across species and tissues of allopolyploid clawed frogs (*Xenopus*). J Mol Evol 68:236–247

Aury J-M, Jaillon O et al (2006) Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. Nature 444:171–178

Baack EJ (2005) To succeed globally, disperse locally: effects of local pollen and seed dispersal on tetraploid establishment. Heredity 94:538–546

Barker MS, Kane NC et al (2008) Multiple paleopolyploidizations during the evolution of the compositae reveal parallel patterns of duplicate gene retention after millions of years. Mol Biol Evol 25(11):2445–2455

Bikard D, Patel D et al (2009) Divergent evolution of duplicate genes leads to genetic incompatibilities within *A. thaliana*. Science 323:623–626

Birchler JA, Riddle NC et al (2005) Dosage balance in gene regulation: biological implications. Trends Genet 21(4):219–226

Blanc G, Wolfe KH (2004) Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. Plant Cell 16:1679–1691

Bottley A, Xia GM et al (2006) Homoeologous gene silencing in hexaploid wheat. Plant J 47:897–906

Brown MT, Goldstone HMH et al (2007) A functionally divergent hydrogenosomal peptidase with protomitochondrial ancestry. Mol Microbiol 64(5):1154–1163

Buggs RJA, Elliott NM et al (2010) Tissue-specific silencing of homoeologs in natural populations of the recent allopolyploid *Tragopogon mirus*. New Phytol 186:175–183

Byrne KP, Wolfe KH (2007) Consistent patterns of rate asymmetry and gene loss indicate widespread neofunctionalization of yeast genes after whole-genome duplication. Genetics 175(3):1341–1350

Colon M, Hernandez F et al (2011) *Saccharomyces cerevisiae* Bat1 and Bat2 aminotransferases have functionally diverged from the ancestral-like *Kluyveromyces lactis* orthologous enzyme. PLoS ONE 6(1):e16099

Conant GC, Wolfe KH (2006) Functional partitioning of yeast co-expression networks after genome duplication. PLoS Biol 4(4):0545–0554

Conant GC, Wolfe KH (2007) Increased glycolytic flux as an outcome of whole-genome duplication in yeast. Mol Syst Biol 3:129

Coyne JA, Orr HA (2004) Speciation. Sinauer Associates, Inc, Sunderland, MA

Davis JC, Petrov DA (2005) Do disparate mechanisms of duplication add similar genes to the genome? Trends Genet 21(10):548–551

Deng C, Cheng CHC et al (2010) Evolution of an antifreeze protein by neofunctionalization under escape from adaptive conflict. Proc Nat Acad Sci U S A 107(50):21593–21598

Des Marais DL, Rausher MD (2008) Escape from adaptive conflict after duplication in an anthocyanin pathway gene. Nature 454(7205):762–765

Doyle JJ, Flagel LE et al (2008) Evolutionary genetics of genome merger and doubling in plants. Annu Rev Genet 42:443–461

Edger PP, Pires JC (2009) Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. Chromosome Res 17:699–717

Erdmann R, Gramzow L et al (2010) GORDITA (AGL63) is a young paralog of the Arabidopsis thaliana B (sister) MADS box gene ABS (TT16) that has undergone neofunctionalization. Plant J 63(6):914–924

Evans BJ (2008) Genome evolution and speciation genetics of clawed frogs (*Xenopus* and *Silurana*). Front Biosci 13:4687–4706

Evlampiev K, Isambert H (2007) Modeling protein network evolution under genome duplication and domain shuffling. BMC Syst Biol 1:49

Flagel LE, Udall JA et al (2008) Duplicate gene expression in allopolyploid gossypium reveals two temporally distinct phases of expression evolution. BMC Biol 6:16

Force A, Cresko WA et al (2005) The origin of subfunctions and modular gene regulation. Genetics 170(1):433–446

Force A, Lynch M et al (1999) The preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531–1545

Froyd CA, Rusche LN (2011) The duplicated deacetylases sir2 and hst1 subfunctionalized by acquiring complementary inactivating mutations. Mol Cell Biol 31(16):3351–3365

Gaeta RT, Pires JC et al (2007) Genomic changes in resynthesized brassica napus and their effect on gene expression and phenotype. Plant Cell 19(11):3403–3417

Gout J-F, Duret L, et al. (2009) Differential retention of metabolic genes following whole-genome duplication. Mol Biol Evol 26(5):1067–1072

Gout J-F, Kahn D, et al. (2010) The relationship among gene expression, the evolution of gene dosage, and the rate of protein evolution. PLoS Genetics 6(5):e1000944

Han MV, Demuth JP et al (2009) Adaptive evolution of young gene duplicates in mammals. Genome Res 19(5):859–867

Hegarty MJ, Barker GL et al (2006) Transcriptome shock after interspecific hybridization in *Senecio* is ameliorated by genome duplication. Curr Biol 16:1652–1659

Hickman MA, Rusche LN (2010) Transcriptional silencing functions of the yeast protein Orc1/Sir3 subfunctionalized after gene duplication. Proc Nat Acad Sci U S A 107(45):19384–19389

Hughes AL (1994) The evolution of functionally novel proteins after gene duplication. Proc R Soc Lond B Biol Sci 256:119–124

Hughes T, Liberles DA (2008) Whole-genome duplications in the ancestral vertebrate are detectable in the distribution of gene family sizes of tetrapod species. J Mol Evol 67(4):343–357

Innan H, Kondrashov F (2010) The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet 11:97–108

Jiao YN, Wickett NJ et al (2011) Ancestral polyploidy in seed plants and angiosperms. Nature 473(7345):97–100

Jimenez-Sanchez G, Childs B et al (2001) Human disease genes. Nature 409:853–855

Johnson DA, Thomas MA (2007) The monosaccharide transporter gene family in Arabidopsis and rice: a history of duplications, adaptive evolution, and functional divergence. Mol Biol Evol 24(11):2412–2423

Josefsson C, Dilkes B et al (2006) Parent-dependent loss of gene silencing during interspecies hybridization. Curr Biol 16:1322–1328

Kanno A, Saeki H et al (2003) Heterotopic expression of class B floral homeotic genes supports a modified ABC model for tulip (*Tulipa gesneriana*). Plant Mol Biol 52:831–841

Layeghifard M, Pirhaji L et al (2009) Adaptive evolution in the Per gene family of vertebrates: neofunctionalization by positive Darwinian selection after two major gene duplications. Biol Rhythm Res 40(6):433–444

Liang H, Plazonic KR et al (2008) Protein under-wrapping causes dosage sensitivity and decreases gene duplicability. PLoS Genet 4(1):0072–0077

Lipinski KJ, Farslow JC et al (2011) High spontaneous rate of gene duplication in *Caenorhabditis elegans*. Curr Biol 21:306–310

Lolkema JS, Dobrowolski A et al (2008) Evolution of antiparallel two-domain membrane proteins: tracing multiple gene duplication events in the DUF606 family. J Mol Biol 378:596–606

Lynch M (2007) The origins of genome architecture. Sinauer Associates, Sunderland, MA

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290(5494):1151–1155
Lynch, M, Conery JS (2003a). The evolutionary demography of duplicate genes. In: Meyer A, Van de Peer Y (eds) Genome evolution. Kluwer Academic Publishers, Dordrecht, 35–44
Lynch M, Conery JS (2003) The origins of genome complexity. Science 302:1401–1404
Lynch M, Force AG (2000) The origin of interspecific genomic incompatibility via gene duplication. Am Naturaliste 156(6):590–605
Lynch M, Katju V (2004) The altered evolutionary trajectories of gene duplicates. Trends Genet 20:544–549
Lynch M, O'Hely M et al (2001) The probability of preservation of a newly arisen gene duplicate. Genetics 159:1789–1804
Maclean CJ, Greig D (2010) Reciprocal gene loss following experimental whole-genome duplication causes reproductive isolation in yeast. Evolution 65(4):932–945
MacNeil AJ, McEachern LA et al (2008) Gene duplication in early vertebrates results in tissue-specific subfunctionalized adaptor proteins: CASP and GRASP. J Mol Evol 67(2):168–178
Maere S, De Bodt S et al (2005) Modeling gene and genome duplications in eukaryotes. Proc Nat Acad Sci U S A 102(15):5454–5459
Makino T, McLysaght A (2010) Ohnologs in the human genome are dosage balanced and frequently associated with disease. Proc Nat Acad Sci U S A 107:9270–9274
Masly JP, Jones CD et al (2006) Gene transposition as a cause of hybrid sterility in *Drosophila*. Science 313:1448–1450
Mayrose I, Zhan SH et al (2011) Recently formed polyploid plants diversify at lower rates. Science 333(6047):1257
Meyer A, Van de Peer Y (2005) From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). BioEssays 27:937–945
Mizuta Y, Harushima Y et al (2010) Rice pollen hybrid incompatibility caused by reciprocal gene loss of duplicated genes. Proc Nat Acad Sci U S A 107(47):20417–20422
Morin RD, Chang E et al (2006) Sequencing and analysis of 10,967 full-length cDNA clones from Xenopus laevis and Xenopus tropicalis reveals post-tetraploidization transcriptome remodeling. Genome Res 16:796–803
Ngai J, Dowling MM et al (1993) The family of genes encoding odorant receptors in the channel catfish. Cell 72:657–666
Ohno S (1970) Evolution by gene duplication. Springer, Berlin
Oka HI (1988). Functions and genetic bases of reproductive barriers. Origin of cultivated rice. Japan Scientific Societies Press/Elsevier, HI Oka, Tokyo, pp 181–209
Otto SP (2007) The evolutionary consequences of polyploidy. Cell 131:452–462
Panopoulou G, Poustka AJ (2005) Timing and mechanism of ancient vertebrate genome duplications—the adventure of a hypothesis. Trends Genet 21:559–567
Papp B, Pal C et al (2003) Dosage sensitivity and the evolution of gene families in yeast. Nature 424:194–197
Paterson AH, Chapman BA et al (2006) Many gene and domain families have convergent fates following independent whole-genome duplication events in *Arabidopsis*, *Oryza*, *Saccharomyces* and *Tetraodon*. Trends Genet 22(11):597–602
Piatigorsky J, Wistow G (1991) The recruitment of crystallins—new functions precede gene duplication. Science 252(5009):1078–1079
Pinhal D, Yoshimura TS et al (2011) The 5S rDNA family evolves through concerted and birth-and-death evolution in fish genomes: an example from freshwater stingrays. BMC Evol Biol 11:151
Pires JC, Gaeta RT (2011) Structural and functional evolution of resynthesized polyploids. In: Schmidt R, Bandcroft I (eds) Genetics and genomics of the brassicaceae, Springer, New York, 9:195–214
Postlethwait JH, Woods IG et al (2000) Zebrafish comparative genomics and the origins of vertebrate chromosomes. Genome Res 10:1890–1902

Presgraves DC (2010) The molecular evolutionary basis of species formation. Nat Rev Genet 11:175–180
Putnam NH, Butts T et al (2008) The amphioxus genome and the evolution of the chordate karyotype. Nature 453:1064–1071
Qian W, Zhang J (2008) Gene dosage and gene duplicability. Genetics 179:2319–2324
Qian WF, Liao BY et al (2010) Maintenance of duplicate genes and their functional redundancy by reduced expression. Trends Genet 26(10):425–430
Ramsey J, Schemske DW (2002) Neopolyploidy in flowering plants. Annu Rev Ecol Syst 33:589–639
Rapp RA, Udall JA et al (2009) Genomic expression dominance in allopolyploids. BMC Biol 7:18
Rausch JH, Morgan MT (2005) The effect of self-fertilization, inbreeding depression, and population size on autopolyploid establishment. Evolution 59(9):1867–1875
Rodriguez DJ (1996) A model for the establishment of polyploidy in plants. Am Naturalist 147(1):33–46
Salmon A, Ainouche ML et al (2005) Genetic and epigenetic consequences of recent hybridization and polyploidy in *Spartina* (Poaceae). Mol Ecol 14:1163–1175
Scannell DR, Byrne KP et al (2006) Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. Nature 440:341–345
Scannell DR, Frank AC et al (2007) Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication. Proc Nat Acad Sci U S A 104(20):8397–8402
Scannell DR, Wolfe KH (2008) A burst of protein sequence evolution and a prolonged period of asymmetric evolution follow gene duplication in yeast. Genome Res 18(1):137–147
Semon M, Wolfe KH (2007) Reciprocal gene loss between Tetraodon and zebrafish after whole genome duplication in their ancestor. Trends Genet 23(3):108–112
Semon M, Wolfe KH (2008) Preferential subfunctionalization of slow-evolving genes after allopolyploidization in *Xenopus laevis*. Proc Nat Acad Sci U S A 105(24):8333–8338
Simillion C, Vandepoele K et al (2002) The hidden duplication past of *Arabidopsis thaliana*. Proc Nat Acad Sci U S A 99:13627–13632
Sinha SC, Wetterer M et al (2005) Origin of asymmetry in adenylyl cyclases: structures of *Mycobacterium tuberculosis* Rv1900c. EMBO J 24:663–673
Soltis DE, Albert VA et al (2009) Polyploidy and angiosperm diversification. Am J Bot 96(1):336–348
Soltis PS, Soltis DE et al (2006) Expression of floral regulators in basal angiosperms and the origin and evolution of ABC-function. Adv Bot Res 44:483–506
Stoltzfus A (1999) On the possibility of constructive neutral evolution. J Mol Evol 49:169–181
Surcel A, Zhou X et al (2008) Long-term maintenance of stable copy number in the eukaryotic SMC family: origin of a vertebrate meiotic SMC1 and fate of recent segmental duplicates. J Syst Evol 46(3):405–423
Tate JA, Ni Z et al (2006) Evolution and expression of homeologous loci in *Tragopogon miscellus* (Asteraceae), a recent and reciprocally formed allopolyploid. Genetics 173:1599–1611
Taylor JS, Raes J (2004) Duplication and divergence: The evolution of new genes and old ideas. Annu Rev Genet 38:615–643
Thomas BC, Pedersen B et al (2006) Following tetraploidy in an Arabidopsis ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. Genome Res 16:934–946
Tirosh I, Barkai N (2007) Comparative analysis indicates regulatory neofunctionalization of yeast duplicates. Genome Biol 8(4):R50
Udall JA, Swanson JM et al (2006) A novel approach for characterizing expression levels of genes duplicated by polyploidy. Genetics 173:1823–1827
Veitia RA (2002) Exploring the etiology of haploinsufficiency. BioEssays 24(2):175–184

Veitia RA, Bottani S et al (2008) Cellular reactions to gene dosage imbalance: genomic, transcriptomic, and proteomic effects. Trends Genet 24(8):390–397

Wang J, Tian L et al (2006) Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. Genetics 172:507–517

Watterson GA (1983) On the time for gene silencing at duplicate loci. Genetics 105(3):745–766

Werth CR, Windham MD (1991) A model for divergent, allopatric speciation of polyploid Pteridophytes resulting from silencing of duplicate-gene expression. Am Naturalist 137(4):515–526

Winter KU, Weiser C et al (2002) Evolution of class B floral homeotic proteins: obligate heterodimerization orginated from homodimerization. Mol Biol Evol 19(5):587–596

Wolfe KH, Shields DC (1997) Molecular evidence for an ancient duplication of the entire yeast genome. Nature 387:708–713

Wood TE, Takebayashi N et al (2009) The frequency of polyploid speciation in vascular plants. Proc Nat Acad Sci US A 106(33):13875–13879

Woolfe A, Elgar G (2007) Comparative genomics using *Fugu* reveals insights into regulatory subfunctionalization. Genome Biol 8(4):R53

Yamagata Y, Yamamoto E et al (2010) Mitochondrial gene in the nuclear genome induces reproductive barrier in rice. Proc Nat Acad Sci U S A 107(4):1494–1499

Yang J, Lusk R et al (2003) Organismal complexity, protein complexity, and gene duplicability. Proc Nat Acad Sci U S A 100(26):15661–15665

Yu W-P, Brenner S et al (2003) Duplication, degeneration and subfunctionalization of the nested *synapsin-Timp* genes in *Fugu*. Trends Genet 19:180–183

Chapter 2
Genetic Consequences of Polyploidy in Plants

James A. Birchler

Abstract Most eukaryotes have a history of whole-genome multiplication events followed by a progressive return to a more diploid state. The initial state of polyploidization, in which more than two copies of the genome are present, is considered here and the various types of genetic consequences that occur depending on the nature of the polyploid formed. The degree of association of chromosomes in meiosis is determined by the relative homology and will affect the segregation of the chromosome which determines the genetic properties. If all the chromosomes are quite similar and form associations of like type, this situation is referred to as autopolyploidy. If the different sets of multiple chromosomes are sufficiently dissimilar to each other, then the homologs will pair in meiosis with themselves and segregate independently of the different but related chromosome pair. This situation is referred to as allopolyploidy. Gene expression in ploidal series typically follows a per cell level correlated more or less with the number of sets of chromosomes present. Variation of individual chromosomes, or aneuploidy, produces a greater number of modulations of gene expression in parallel to classical studies noting that aneuploids have greater impact on the phenotype than changes in the copy number of the whole genome. The genetic properties of odd-number ploidies, such as triploids, are also described as well as higher ploidal levels such as hexaploidy and octoploidy.

J. A. Birchler (✉)
Division of Biological Sciences, University of Missouri, 311 Tucker Hall,
65211 Missouri, Colombia
e-mail: birchlerj@missouri.edu

P. S. Soltis and D. E. Soltis (eds.), *Polyploidy and Genome Evolution*,
DOI: 10.1007/978-3-642-31442-1_2,

2.1 Introduction

Most eukaryotes have a history of polyploidization followed by fractionation back to a near diploid level (Wolfe and Shields 1997; Simillion et al. 2002; Bowers et al. 2003; Blanc and Wolfe 2004; Chapman et al. 2006; Maere et al. 2005; Blomme et al. 2006; Freeling and Thomas 2006; Barker et al. 2008). Thus, at the least, polyploidy in essence is a matter of degree, and it has played an important role in the composition of the gene repertoire of many species. Typically, it is defined as the presence of more copies of the whole genome than the normal two that constitute a diploid (Stebbins 1947). However, from the standpoint of gene content, the determination of whether a species is a polyploid is somewhat arbitrary and dependent on the time before the present when the copy number of the genome was increased. Nevertheless, for evolutionarily "recent" events, certain principles can apply which will be summarized in this chapter.

In the polyploidy literature, the basic chromosome number is designated by x and consists of the complete set of chromosomes, or a genome. The number of chromosomes in the gametophyte generation and hence the gametes is referred to as the gametic chromosome number or n. In diploids, $x = n$, but at higher levels of polyploidy, this is not the case.

Polyploidy is typically divided into at least two categories that are determined by the type of chromosome pairing in meiosis I and the distribution of chromosomes during this process. Indeed, the type of chromosome pairing that occurs in meiosis affects the genetic properties of the species so such classifications have value. If the increase in genome copy number results from the combination of chromosome sets from divergent species, the different types of chromosomes will usually not pair with each other in prophase of meiosis I. In the case of tetraploids, if both divergent genomes are doubled by whatever means, those sets of chromosomes that are similar or identical will preferentially pair with each other to the exclusion of the other genome. This type of pairing is referred to as "disomic" in analogy with the situation in a diploid. A species with this type of scenario is referred to as an allopolyploid because the contributing genomes are different from each other.

2.2 Allopolyploids

Genetic ratios in an allotetraploid depend on the constitution of each genome (Clausen and Goodspeed 1925; Clausen 1941; Clausen and Cameron 1944). The different sets of related chromosomes are referred to as homoeologues. If both the homoeologues possess the homoeologous gene copies that are expressed similarly, then both would need to be mutant in order to express a recessive phenotype. Under these circumstances, duplicate gene ratios would typically be observed. In other words, recessive phenotypes would be found in 1/16 (1/4 × 1/4) of the F_2

from a self of an F_1 between parental types that are dominant and recessive. However, if one of the gene copies is missing or expressed in other tissues from one of the homoeologous chromosomes, then genetic ratios typical of a diploid will be found because only one genome will have different alleles in an F_1, and they will segregate to produce a 3:1 ratio because the single genome will behave as a diploid.

2.3 Autopolyploids

If on the other hand the increase in genome copy number in a polyploid results from the same species such that the chromosomes are all quite similar, the pairing in prophase of meiosis I forms conglomerates that switch pairing partners along the length of the chromosome (Fig. 2.1). This type of pairing is referred to as "quadrivalent" pairing because all four chromosomes present can be involved with each other. However, 3:1 and 2:2 associations are also observed. The segregation in this case will depend on the position of the locus in question in the chromosome and relative to the respective centromere (Blakeslee et al. 1923; Haldane 1930; Bartlett and Haldane 1934; Mather 1935, 1936; Randolph 1935; Little 1945, 1958; Doyle 1973). Those genes near the centromere will be distributed to the diploid gametes based on the usual case that pairs of centromeres will separate from each other at meiosis I and that the two sets from each chromosome of the complement will do so at random. A homozygous dominant autotetraploid (*AAAA*) is referred to as a quadruplex and the homozygous recessive (*aaaa*) as a nulliplex. There are three types of heterozygotes: *AAAa* (triplex), *AAaa* (duplex), and *Aaaa* (simplex). If one designates a hybrid autotetraploid as $AA'aa'$, then there are six types of possible gametes that will be formed: AA', Aa, Aa', aA, $a'A'$, aa'. The frequency of diploid homozygous gametes under these circumstances is 1/6 (0.167). A self-pollination will produce 2.77 % of the progeny that are homozygous for the recessive allele (Fig. 2.2). However, as the distance of a gene from the centromere increases, recombination between the locus and the centromere will randomize the distribution of the different alleles into the diploid gametes to the point that the frequency of homozygous diploid gametes will be $(4/8 \times 3/7 = 0.21)$ as a maximum. In this case, a self-pollination will produce 4.41 % of the progeny that are homozygous.

Recombination between the monitored locus and the centromere can also produce homozygous spores from a triplex heterozygote (*AAAa*) to produce *aa* gametes (Catcheside 1956). This process is called double reduction. Again, this result is affected by the position of the locus under consideration from the centromere with greater double reduction increasing with distance. Another factor affecting segregation in autotetraploids is aneuploidy, i.e., altered copy number of individual chromosomes. This circumstance would change the pairing and segregation properties of individual chromosomes. Autotetraploids can also generate spontaneous diploid progeny via parthenogenesis (Randolph and Fischer 1939).

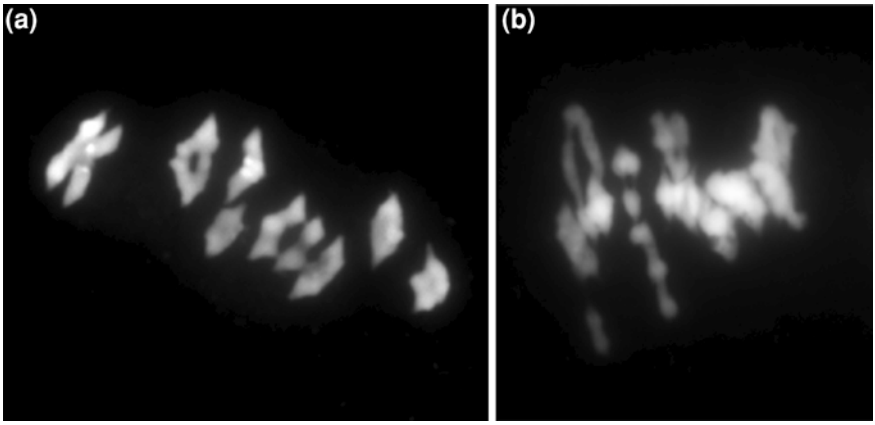

Fig. 2.1 Comparison of meiotic anaphase I in matched diploid and tetraploid maize plants. **a** Gray value image of anaphase I of diploid inbred line B73. Note that membrers of each pair of homologs separate from each other. Modified from (Birchler 2011). **b** Gray value image of anaphase I of a tetraploid derivative of inbred line B73. Note the multivalent associations of chromosomes. Photos by Zhi Gao and Fangpu Han

	AA	Aa	Aa	Aa	Aa	aa
AA	AA AA	AA Aa	AA Aa	AA Aa	AA Aa	AA aa
Aa	AA Aa	Aa Aa	Aa Aa	Aa Aa	Aa Aa	Aa aa
Aa	AA Aa	Aa Aa	Aa Aa	Aa Aa	Aa Aa	Aa aa
Aa	AA Aa	Aa Aa	Aa Aa	Aa Aa	Aa Aa	Aa aa
Aa	AA Aa	Aa Aa	Aa Aa	Aa Aa	Aa Aa	Aa aa
aa	AA aa	aa Aa	aa Aa	aa Aa	aa Aa	aa aa

Fig. 2.2 Genotypes in a self-pollination of an autotetraploid heterozygote AAaa. In an autotetraploid, the gametes are diploid. With two chromosomes carrying the dominant **A** allele and two carrying the recessive **a** allele, the distribution of gametes from the heterozygote is shown across the *top* and along the side. The combinations of these gametes to produce the tetraploid progeny are shown in the grid. Only one out of 36 are homozygous for **A**, and one out of 36 are homozygous for the recessive **a**. Other combinations of **A** and **a** are shown. These conditions hold for genes closely linked to centromeres as described in the text

2.4 Segmental Allopolyploids

A third classification based upon empirical chromosome associations is segmental allopolyploid (Stebbins 1947). In this case, some chromosomes exhibit bivalent pairing and others show quadrivalent pairing. The basis of such behavior was not clear until recently. Xiong and colleagues (Xiong et al. 2011) found that in resynthesized *Brassica napus* derived from the diploid progenitors, *B. oleracea* and *B. rapa*, different lineages could form compensating nullisomic-tetrasomic configurations for different chromosomes. In this case, the tetraploid will have some chromosomes that are basically identical and other members of the set that will be divergent. Similar results were reported for naturally occurring tetraploid *Tragopogon miscellus* (Chester et al. 2012). Such a species will be a composite of allo- and autotetraploid chromosomes for different members of the karyotype and would be expected to exhibit the pairing characterized by a segmental allopolyploid.

2.5 Heterosis and Ploidy

Because of the chromosome pairing considerations noted above, allopolyploids will have a diversity of gene products "fixed" in their genomic structure. This circumstance will basically maintain the essence of hybrid vigor even though technically an otherwise high degree of homozygosity might be present. Thus, allopolyploids typically exhibit robust biomass and excellent fertility compared to the diploid progenitor species. Nevertheless, crosses between different isolates of allopolyploids can show even greater heterotic effects when each genome is heterozygous as well (Gustafson 1946). Autopolyploids, which can have up to four different alleles at one locus, also exhibit hybrid vigor and with increasing diversity of alleles present, a phenomenon known as progressive heterosis, there is increasing biomass and fertility (Busbice and Wilsie 1966; Levings et al. 1967; Mok and Peloquin 1975; Groose et al. 1989). However, autotetraploids can be subject to inbreeding depression in which the potential exists for all copies of a chromosome to become homozygous (Busbice and Wilsie 1966; Sockness and Dudley 1989a, b). Polyploids that are entirely homozygous exhibit extreme depression and reduction of stature and fertility (Busbice and Wilsie 1966; Riddle et al. 2006; Abel and Becker 2007; Stupar et al. 2007; Redei 1964; d'Erfurth et al. 2009; Yao et al. 2011). This situation is unlikely under natural circumstances.

A discussion of the vigor of polyploids needs to consider the intersection with heterosis or hybrid vigor. In recent years, it has been possible to produce ploidy series for completely or highly homozygous materials (Riddle et al. 2006; Abel and Becker 2007; Stupar et al. 2007; d'Erfurth et al. 2009; Yao et al. 2011). The general rule that emerges from these studies is that with increasing ploidy and the maintenance of homozygosity, there is usually a decline in stature and fertility. The cell and pollen size increases with ploidy and the plants typically take on a

"stocky" appearance but with extreme ploidies, the plants are depauperate (Blakeslee 1941; Randolph 1942; Rhoades and Dempsey 1966; d'Erfurth et al. 2009; Yao et al. 2011). In contrast, hybrids with increasing ploidy tend to exhibit greater biomass and in the species in which it has been examined closely, there is an increase in heterosis with increasing diversity of alleles, i.e., progressive heterosis (Busbice and Wilsie 1966; Mok and Peloquin 1975; Levings et al. 1967; Groose et al. 1989; Bingham et al. 1994; Riddle and Birchler 2008). The common view that polyploids exhibit more robust stature derives from experience with allopolyploids or with heterotic autopolyploids, which are the situations most commonly encountered.

2.6 Aneuploidy Relative to Ploidy

In contrast to a ploidy series, changes in dosage of individual chromosomes (or substantial parts of chromosomes) have a more dramatic effect on the phenotype (Blakeslee et al. 1920; Blakeslee 1934). Typically, the removal of a chromosome or chromosomal segment has the strongest effects and is lethal in some cases (Kush and Rick 1968; Vizir and Mulligan 1999). All of the monosomics for each of the ten chromosomes in maize have been recovered and studied (Weber 1983), but this is not the case in other species in which this issue has been examined such as tomato (Kush and Rick 1968) and *Arabidopsis* (Vizir and Mulligan 1999). The addition of a chromosome to produce a trisomic usually also has a detrimental effect on plant vigor but the usual circumstance is that the impact is much less than monosomics (Lee et al. 1996). Indeed, full sets of trisomics have been produced for many plant species (Singh 1993). Tetrasomics for whole chromosome arms, otherwise called secondary trisomics, have been produced in *Datura* by recovery of extra chromosomes that are duplicated for one or the other arm of the progenitor chromosome (Blakeslee 1934). These secondary trisomics usually have more intensified phenotypic effects and are more intensified when present in haploids (Satina et al. 1937a, b).

Extra or missing chromosomes in higher ploidies have less severe phenotypic effects. A comprehensive set of aneuploids was generated in hexaploid wheat (Sears 1944; Sears 1953, 1954). Monosomics and trisomics are regularly produced, and because of the high ploidy state, nullisomics, which are missing both copies of a chromosome, can be produced (Sears 1953, 1954). Nullisomics have a more severe effect than the corresponding monosomic. Tetrasomics can be produced and have a more severe effect than the respective trisomic. Compensating nullisomics for one homoeologue and tetrasomics for another return to a more normal phenotype than exhibited by the nullisomic or tetrasomic alone (Sears 1953, 1954). Newly synthesized *B. napus* (Xiong et al. 2011) and natural neopolyploids of *T. miscellus* (Chester et al. 2012) will exhibit aneuploidy that resolves into compensating 4:0 or 3:1 contributions from different progenitor genomes illustrating that the compensating balanced condition is favored in

laboratory or natural selection. Together, these results further illustrate that the greater the deviation from the standard set of chromosomes, the more severe the impact on the phenotype.

2.7 Gene Expression Studies

Studies on gene expression in ploidy and aneuploid series parallel the phenotypic results. When individual genes are sampled in a ploidy series, the expression level is more or less proportional to the ploidal level, although there are examples of genes whose expression deviates from this trend both positively and negatively (Birchler and Newton 1981; Guo et al. 1996). Genome-wide studies of gene expression in ploidy series demonstrate a similar pattern (Wang et al. 2004; Albertin et al. 2005; Stupar et al. 2007; Riddle et al. 2010; Yu et al. 2010). In contrast, sampling of individual genes or protein patterns in aneuploids reveals a greater set of changes from the diploid level of expression (Birchler 1979; Birchler and Newton 1981; Guo and Birchler 1994). A dosage series for a particular chromosomal region would alter the amount of expression of a portion of the total gene products encoded across the genome. The effects could be positive or negative correlations with the change in dosage. The more common effect especially with trisomics was a negative correlation between the dosage and the target gene expression (Birchler 1979; Birchler and Newton 1981; Guo and Birchler 1994). Thus, the gene expression patterns show changes in a ploidy series but aneuploid series exhibit greater effects in parallel with the phenotypic relationships.

2.8 Genomic Balance

This gene expression relationship led to the suggestion that the stoichiometry of regulatory genes affected the outcome of gene expression (Birchler and Newton 1981) and ultimately the phenotype (Guo and Birchler 1994). Studies to identify single genes that would mimic the aneuploid effects using a partial loss of function mutation in the *white* eye color gene in *Drosophila* produced single-gene mutations that would modulate the target's expression either positively or negatively (Rabinow et al. 1991; Birchler et al. 2001). The molecular identification of many of these genes revealed them to be transcription factors, chromatin modifiers, and components of signal transduction (Birchler et al. 2001).

Interestingly, these same classes of genes are typical of those that exhibit preferential retention following a polyploidization event (Blanc and Wolfe, 2004; Freeling and Thomas, 2006) and underrepresentation in segmental duplications (Maere et al. 2005; Freeling et al. 2008). Thus, this evidence suggests that if these classes of genes are out of register with each other, there is a negative fitness

consequence. Thus, the phenotypic, gene expression and evolutionary studies form a coherent picture that these types of genes form a balance. When individual components exhibit a dosage effect, this will ultimately produce a fitness consequence due to the impact of the altered gene expression on the phenotype (Birchler et al. 2001; Veitia 2002; Veitia 2004; Birchler et al. 2005, 2007; Veitia et al. 2008; Birchler and Veitia 2007, 2010).

2.9 Triploids

Triploids are a polyploid level between diploid and tetraploid. They arise from crosses between diploid and tetraploids of the same or related species or from unreduced gametes from one diploid parent. In meiosis, the chromosomes associate in trivalents, which consists of pairing of any two chromosomes at any one point (McClintock 1929; Punyasingh 1947; Upcott 1935). The distribution of chromosomes is nearly random, resulting in spores that range from $1x$ to $2x$. As a consequence, the gametophytes are mostly highly aneuploid and in some cases abort (Satina and Blakeslee 1937a, b). Fertilization involving gametes of different chromosome numbers in the endosperm will often cause endosperm abortion (Satina et al. 1938; Punyasingh 1947; Brink and Cooper 1947; Cooper 1951). The gametes that are successful tend to be those at or near the $1x$ or $2x$ level. Because of the variability of the chromosome numbers in gametes from triploid individuals, this ploidal level is not stable.

2.10 Higher Ploidal Levels

Ploidal levels above the tetraploid level most commonly involve hexaploids and octoploids although much higher levels have been documented. Chromosome pairing in allohexaploids has been studied in detail, for example in wheat, which is ordinarily disomic in nature (Kihara 1919; Lilienfeld 1951; Dvorak et al. 1988). The Ph system insures pairing of homologs and against pairing of homoeologues but homoeologues can pair in mutant plants (Yousafzai et al. 2010; see Chap. 7, this volume). The wheat genome is composed of three different slightly diverged genomes tracing back through the joining of an allotetraploid composed of two genomes with the third. The Ph system maintains disomic pairing and hence excellent fertility. Octoploids, using sugar cane as an example, have variable chromosome numbers due to the minimal detrimental effects of aneuploidy at this level (Piperidis et al. 2010). In contrast, triticale, which is an octoploid consisting of hexaploid wheat with the addition of a rye genome, exhibits faithful chromosomes numbers.

2.11 Concluding Remarks

The genetics of polyploids depends essentially on the pairing properties of the multiple chromosomes in meiosis. If the multiple copies of a genome are sufficiently dissimilar from each other, they tend to pair among themselves and maintain the genetic variation within each genome. If the multiple copies of a genome are similar to each other, then all copies are free to pair and recombine among themselves. In this circumstance, the genetic behavior of a particular gene is dependent on its position on the chromosome and the fidelity of the pairing of homologs. Aneuploidy, i.e., the variation of a single chromosome or chromosomal segment, can have more severe consequences than varying the whole genome. However, as the background ploidy increases, the effect of the same chromosome change of aneuploidy becomes less. The phenotypic effects, gene expression patterns and the evolutionary results of differential gene retention following whole-genome duplications versus segmental duplication suggest the importance of genomic balance.

Acknowledgments Research supported by National Science Foundation grant DBI 0733857.

References

Abel S, Becker HC (2007) The effect of autopolyploidy on biomass production in homozygous lines of *Brassica rapa* and *Brassica oleracea*. Plant Breed 126:642–643

Albertin W, Brabant P, Catrice O, Eber F, Jenczewski E, Chèvre AM, Thiellement H (2005) Autopolyploidy in cabbage (*Brassica oleracea L.*) does not alter significantly the proteomes of green tissues. Proteomics 5:2131–2139

Barker MS, Kane NC, Matvienko M, Kozik A, Michelmore RW, Knapp SJ, Rieseberg LH (2008) Multiple paleopolyploidizations during the evolution of the compositae reveal parallel patterns of duplicate gene retention after millions of years. Mol Biol Evol 25:2445–2455

Bartlett MS, Haldane JBS (1934) The theory of inbreeding in Autotetraploids. J Genet 29:175–180

Bingham ET, Groose RW, Woodfield DR, Kidwell KK (1994) Complementary gene interactions in alfalfa are greater in autotetraploids than diploids. Crop Sci 34:823–829

Birchler JA (1979) A study of enzyme activities in a dosage series of the long arm of chromosome one in maize. Genetics 92:1211–1229

Birchler JA (2011) Epigenetic aspects of centromere function in plants. Curr Opin Plant Biol 14:217–222

Birchler JA, Newton KJ (1981) Modulation of protein levels in chromosomal dosage series of maize: the biochemical basis of aneuploid syndromes. Genetics 99:247–266

Birchler JA, Bhadra U, Pal Bhadra M, Auger DL (2001) Dosage dependent gene regulation in multicellular eukaryotes: implications for dosage compensation, aneuploid syndromes and quantitative traits. Dev Biol 234:275–288

Birchler JA, Riddle NC, Auger DL, Veitia RA (2005) Dosage balance in gene regulation: biological implications. Trends Genet 21:219–226

Birchler JA, Veitia RA (2007) The gene balance hypothesis: from classical genetics to modern genomics. Plant Cell 19:395–402

Birchler JA, Veitia RA (2010) The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. New Phytol 186:54–62

Birchler JA, Yao H, Chudalayandi S (2007) Biological consequences of dosage dependent gene regulatory systems. Biochem Biophys Acta-Gene Struct Expr 1769:422–428
Blakeslee AF (1934) New Jimson weeds from old chromosomes. J Hered 24:80–108
Blakeslee AF (1941) Effect of induced polyploidy in plants. Am Nat 75:117–135
Blakeslee AF, Belling J, Farnham ME (1920) Chromosomal duplication and Mendelian phenomena in Datura mutants. Science 52:388–390
Blakeslee AF, Belling J, Farnham ME (1923) Inheritance in tetraploid Daturas. Bot Gaz 76:329–373
Blanc G, Wolfe KH (2004) Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. Plant Cell 16:1679–1691
Blomme T, Vandepoele K, De Bodt S, Simillion C, Maere S, Van de Peer Y (2006) The gain and loss of genes during 600 million years of vertebrate evolution. Genome Biol 7:43
Bowers JE, Chapman BA, Rong J, Paterson AH (2003) Unraveling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 422:433–438
Busbice TH, Wilsie CP (1966) Inbreeding depression and heterosis in autotetraploids with application to *Medicago sativa* L. Euphytica 15:52–67
Brink RA, Cooper DC (1947) The endosperm in seed development. Bot Rev 13:423–541
Catcheside DG (1956) Double reduction and numerical non-disjunction in tetraploid maize. Heredity 10:205–218
Chapman BA, Bowers JE, Feltus FA, Paterson AH (2006) Buffering of crucial functions by paleologous duplicated genes may contribute cyclicality to angiosperm genome duplication. Proc Natl Acad Sci USA 103:2730–2735
Chester M, Gallagher JP, Symonds VV, Cruz da Silva AV, Mavrodiev EV, Leitch AR, Soltis PS, Soltis DE (2012) Extensive chromosomal variation in a recently formed natural allopolyploid, *Tragopogon miscellus* (*Asterasceae*). Proc Natl Acad Sci USA (in press)
Clausen RE (1941) Polyploidy in *Nicotiana*. Am Nat 75:291–306
Clausen RE, Cameron DR (1944) Inheritance in *Nicotiana tabacum*. XVIII. Monosomic analysis. Genetics 29:447–477
Clausen RE, Goodspeed TH (1925) Interspecific hybridization in *Nicotiana* II. A tetraploid *glutinosa-tabacum* hybrid. An experimental verification of Winge's hypothesis. Genetics 10:278–284
Cooper DC (1951) Caryopsis development following matings between diploid and tetraploid strains of *Zea mays*. Am J Bot 38:702–708
Doyle GG (1973) Autotetraploid segregation. Theor Appl Genet 43:139–146
d'Erfurth I, Jolivet S, Froger N, Catrice O, Novatchkova M, Mercier R (2009) Turning meiosis into mitosis. PLoS Biol 7:1000124
Dvorak J, McGuire PE, Cassidy B (1988) Apparent sources of the a genomes of wheats inferred from the polymorphism in abundance and restriction fragment length of repeated nucleotide sequences. Genome 30:680–689
Freeling M, Lyons E, Pedersen B, Alam M, Ming R, Lisch D (2008) Many or most genes in *Arabidopsis* transposed after the origin of the order Brassicales. Genome Res 18:1924–1937
Freeling M, Thomas BC (2006) Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. Genome Res 16:805–814
Guo M, Birchler JA (1994) Trans-acting dosage effects on the expression of model gene systems in maize aneuploids. Science 266:1999–2002
Guo M, Davis D, Birchler JA (1996) Dosage effects on gene expression in a maize ploidy series. Genetics 142:1349–1355
Groose RW, Talbert LE, Kojis WP, Bingham ET (1989) Progressive heterosis in autotetraploid alfalfa: studies using two types of inbreds. Crop Sci 29:1173–1177
Gustafson A (1946) The effect of heterozygosity on viability and vigor. Hereditas 32:263–286
Haldane JBS (1930) The theoretical genetics of autopolyploids. J Genet 22:359–372
Kihara H (1919) Uber zytologische Studien bei einigen Getreidearten. Bot Mag Tokyo 33
Kush GS, Rick CM (1968) Cytogenetic analysis of the tomato genome by means of induced deficiencies. Chromosoma 23:452–484

Lee EA, Darrah LL, Coe EH (1996) Dosage effects on morphological and quantitative traits in maize aneuploids. Genome 39:898–908
Levings CS, Dudley JW, Alexander DE (1967) Inbreeding and crossing in autotetraploid maize. Crop Sci 7:72–73
Lilienfeld FA (1951) H. Kihara: Genome analysis in *Triticum* and *Aegilops*. Concluding review. Cytologia 16:101–123
Little TM (1945) Gene segregation in autotetraploids. Bot Rev 11:60–85
Little TM (1958) Gene segregation in autotetraploids. II. Bot Rev 24:319–339
Maere S, DeBodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y (2005) Modeling gene and genome duplications in eukaryotes. Proc Natl Acad Sci USA 102:5454–5459
Mather K (1935) Reductional and equational separation of the chromosomes in bivalents and tetravalents. J Genet 30:53–78
Mather K (1936) Segregation and linkage in autotetraploids. J Genet 32:287–314
McClintock B (1929) A cytological and genetical study of triploid maize. Genetics 14:180–227
Mok DWS, Peloquin SJ (1975) Breeding value of 2n pollen (diploandroids) in tetraploid x diploid crosses in potato. Theor Appl Genet 46:307–314
Piperidis G, Piperidis N, D'Hont A (2010) Molecular cytogenetic investigation of chromosome composition and transmission in sugarcane. Mol Genet Genomics 284:65–73
Punyasingh K (1947) Chromosome numbers in crosses of diploid, triploid and tetraploid maize. Genetics 32:541–554
Rabinow L, Nguyen-Huynh AT, Birchler JA (1991) A trans-acting regulatory gene that inversely affects the expression of the *white*, *brown* and *scarlet* loci in *Drosophila melanogaster*. Genetics 129:463–480
Randolph L (1935) Cytogenetics of tetraploid maize. J Agri Res 50:591–605
Randolph LF (1942) The influence of heterozygosis on fertility and vigor in autotetraploid maize. Genetics 27:163
Randolph LF, Fischer HE (1939) The occurrence of parthenogenetic diploids in tetraploid maize. Proc Natl Acad Sci USA 25:161–164
Redei G (1964) Crossing experiences with polyploids. Arabidopsis Inf Serv 1:13
Rhoades MM, Dempsey E (1966) Induction of chromosome doubling at meiosis by the *elongate* gene in maize. Genetics 54:505–522
Riddle NC, Birchler JA (2008) Comparative analysis of inbred and hybrid maize at the diploid and tetraploid levels. Theor Appl Genet 116:563–576
Riddle NC, Kato A, Birchler JA (2006) Genetic variation for the response to ploidy change in *Zea mays* L. Theor Appl Genet 114:101–111
Riddle NC, Jiang H, An L, Doerge RW, Birchler JA (2010) Gene expression analysis at the intersection of ploidy and hybridity in maize. Theor Appl Genet 120:341–353
Satina S, Blakeslee AF (1937a) Chromosome behavior in triploid *Datura stramonium* I. The male gametophyte. Am J Bot 24:518–527
Satina S, Blakeslee AF (1937b) Chromosome behavior in triploid *Datura stramonium* II. The female gametophyte. Am J Bot 24:621–627
Satina S, Blakeslee AF, Avery AG (1937) Balanced and unbalanced haploids in Datura. J Hered 28:192–202
Satina S, Blakeslee AF, Avery AG (1938) Chromosome behavior in triploid Datura. III. The seed. Am J Bot 25:595–602
Sears ER (1944) Cytogenetic studies with polyploid species of wheat. II. Additional chromosome aberrations in *Triticum vulgare*. Genetics 29:232–246
Sears ER (1953) Nullisomic analysis in common wheat. Am Nat 87:245–252
Sears ER (1954) The aneuploids of common wheat. Univ Mo Res Bull 572:1–58
Simillion C, Vandepoele K, Montagu MC, Zabeau M, Van de Peer Y (2002) The hidden duplication past of *Arabidopsis thaliana*. Proc Natl Acad Sci USA 99:13627–13632
Singh RJ (1993) Plant cytogenetics. CRC Press Inc, Boca Raton
Sockness BA, Dudley JW (1989a) Performance of single and double cross autotetraploid maize hybrids with different levels of inbreeding. Crop Sci 29:875–879

Sockness BA, Dudley JW (1989b) Morphology and yield of isogenic diploid and tetraploid maize inbreds and hybrids. Crop Sci 29:1029–1032

Stebbins GL Jr (1947) Types of polyploids: their classification and significance. Adv Genet 1:403–429

Stupar RM, Bhaskar PB, Yandell BS, Rensink WA, Hart AL, Ouyang S, Veilleux RE, Busse JS, Erhardt RJ, Cr Buell, Jiang J (2007) Phenotypic and transcriptomic changes associated with potato autopolyploidazation. Genetics 176:2055–2067

Upcott M (1935) The cytology of triploid and tetraploid *Lycopersicum esculentum*. J Genet 27:105–132

Veitia RA (2002) Exploring the etiology of haploinsufficiency. BioEssays 24:175–184

Veitia RA (2004) Gene dosage balance in cellular pathways: implications for dominance and gene duplicability. Genetics 168:569–574

Veitia RA, Bottani S, Birchler JA (2008) Cellular reactions to gene dosage imbalance: genomic, transcriptomic and proteomic effects. Trends Genet 24:390–397

Vizir IY, Mulligan BJ (1999) Genetics of gamma-irradiation-induced mutations in *Arabidopsis thaliana*: large chromosomal deletions can be rescued through fertilization of diploid eggs. J Hered 90:412–417

Wang J, Tina L, Madlung A, Lee HS, Chen M, Lee JJ, Watson B, Kagochi T, Comai L, Chen ZJ (2004) Stochastic and epigenetic changes of gene expression in *Arabidopsis* polyploids. Genetics 167:1961–1973

Weber DF (1983) Monosomic analysis in diploid crop plants. In: Swaminathan MS, Gupta PK, Sinha U (eds) Cytogenetics of crop plants. Macmillan India Limited, New Delhi, pp 351–378

Wolfe KH, Shields DC (1997) Molecular evidence for an ancient duplication of the entire yeast genome. Nature 387:708–713

Xiong Z, Gaeta RT, Pires JC (2011) Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. Proc Natl Acad Sci USA 108:7908–7913

Yao H, Kato A, Mooney B, Birchler JA (2011) Phenotypic and gene expression analysis of a ploidy series of maize inbred Oh43. Plant Mol Biol 75:237–251

Yousafzai FK, Al-Kaff N, Moore G (2010) The molecular features of chromosome pairing at meiosis: the polyploidy challenge using wheat as a reference. Funct Integr Genomics 10:147–156

Yu Z, Haberer G, Matthes M, Rattei T, Mayer KF, Gierl A, Torres-Ruiz RA (2010) Impact of natural genetic variation on the transcriptome of autotetraploid *Arabidopsis thaliana*. Proc Natl Acad Sci USA 107:17809–17814

Chapter 3
Meiosis in Polyploid Plants

Marie-Luise Zielinski and Ortrun Mittelsten Scheid

Abstract Meiosis is an obligate process during sexual reproduction, which involves the combination of parental genomes and the coordinated segregation of the recombined chromosomes to the gametes. Polyploidy has direct and fundamental consequences on meiosis, which are gradually and individually different between the extreme cases of auto- and allopolyploids. Multiple chromosome complements have a major impact, especially on chromosome pairing during pachytene and on the segregation of genotypes and phenotypes in progeny. At the same time, irregularities during meiosis are a major source of naturally occurring polyploidization events by the formation of unreduced gametes. Although individuals originating from nonhaploid gametes may suffer from reduced vigor and fecundity, their gametogenesis can produce many more chromosomal combinations than regular diploids, and thereby expose more diversity to natural selection. A more relaxed control of pairing and segregation in polyploids, possibly also with increased recombination rates, might be an important contribution to evolution and adaptation potential, especially under drastic or frequent changes in environmental conditions.

3.1 Introduction

Polyploidization can affect the size of the nucleus and the cell, stimulate genetic rearrangements, or modify gene expression patterns, but these changes very likely operate by the same mechanistic principles of nuclear organization, recombination, or gene regulation, as in diploid cells. In contrast, reducing the number of

M.-L. Zielinski · O. Mittelsten Scheid (✉)
Gregor Mendel Institute of Molecular Plant Biology, Vienna, Austria
e-mail: ortrun.mittelsten_scheid@gmi.oeaw.ac.at

P. S. Soltis and D. E. Soltis (eds.), *Polyploidy and Genome Evolution*,
DOI: 10.1007/978-3-642-31442-1_3,

chromosomes during the formation of gametes requires a more elementary modification in polyploids. At the same time, irregular gamete formation is a major trigger for polyploidization. Both processes converge in meiosis. This is a sequence of usually two special, subsequent cell divisions in all sexually reproducing eukaryotic organisms in which the number of chromosomes is reduced to half, before two gametes of different parental origin fuse to form a zygote to produce the next generation.

Like mitosis, the cell division of somatic cells, meiosis is also preceded by the replication of the nuclear DNA into duplicate sister chromatids which are attached at the centromeres. While in mitosis the replicated chromosomes are arranged individually on the metaphase plate, and chromatids become separated in the subsequent anaphase, each replicated chromosome in meiotic cells is first aligned with its respective homologous partner. This pairing is followed by programed induction of DNA double-strand breaks (DSB), from which at least one per chromosome becomes converted into a recombination event between non-sister chromatids (crossover, CO). Later, these COs become cytologically visible as chiasmata. Recombination within chromosome pairs results in new and unique combinations of paternal and maternal genetic information per chromosome, while it is also prerequisite for subsequent organized chromosome segregation. Paired chromosomes become physically connected by the synaptonemal complex (SC). The process of pairing, synapsis, and recombination takes most of the time during prophase I, which is subdivided into leptotene, zygotene, pachytene, diplotene, and diakinesis according to chromosome condensation and configuration. In the subsequent metaphase I, the recombined meiotic chromosome pairs are arranged on the equatorial plane, sister kinetochores of each pair attach to opposite spindles and are pulled apart in anaphase I. Correct and synchronous pairing is a prerequisite for equal distribution of the chromosomes to the two daughter cells. After telophase I and prophase II, sister chromatids become separated during meta-, ana-, and telo-phase II, resulting in a classical meiosis with four nuclei. Faultless meiosis is an essential factor for fertility and thereby decisive for evolutionary success. Therefore, it is not surprising that it is a tightly controlled process and intensely studied in many organisms.

The majority of meiosis research is performed with diploid organisms, and there are excellent reviews available that describe morphological, genetic, and mechanistic aspects of meiosis in many different systems (Dawe 1998; Bhatt et al. 2001; Armstrong et al. 2003; Mezard et al. 2007; Mercier and Grelon 2008; Harrison et al. 2010; Pawlowski 2010). However, characteristic differences in all stages of meiosis in many different polyploids have been observed repeatedly (reviewed in Ramsey and Schemske 2002). Therefore, we will focus on those aspects that are known or expected to be different between meiosis in diploids and polyploids. Although mechanistically intertwined, we will separate three aspects, (1) chromosome pairing, (2) recombination and crossover, and (3) segregation. Whenever relevant, we will distinguish between autopolyploids, with multiples of similar chromosomes, and allopolyploids, with chromosomes of different origin, and therefore higher divergence.

Since polyploidy in animal germ lines is restricted to relatively few groups, and the research literature about meiotic mechanisms is scarce, most of the evidence summarized here stems from investigations in fungi and plants. There, polyploidy is by far most common among angiosperms, but genome research provides growing evidence for ancient polyploidization events in other groups (Sundstrom et al. 2008; Jiao et al. 2011). Independent of how and when polyploids originate, mastering meiosis is an important checkpoint for survival and evolutionary success.

3.2 Recognition, Pairing, and Synapsis

Early in meiosis, homologous chromosomes have to find and recognize each other. Interphase chromosomes are thought to occupy nonrandom territories (Schubert and Shaw 2011), and nonrandom spatial organization of homologous chromosomes has been recognized (reviewed in Avivi and Feldman 1980). In many eukaryotes, telomeres and centromeres can cluster at opposite poles of the nucleus (Rabl 1885; Cowan et al. 2001). This Rabl-configuration (Fig. 3.1a) is found in wheat, rye, oats, and barley, but not in maize or *Arabidopsis* (reviewed in Schubert and Shaw 2011). It is assumed to be related to large genome size and/or chromosome length (Dong and Jiang 1998). Rabl-configuration may support gene expression control and the onset of meiosis (Cowan et al. 2001).

After the premeiotic DNA replication, the two sister chromatids are held together by sister chromatid cohesion and gradually become condensed. During leptotene, their chromatin is folded into loops attached to a protein fiber core, the axial element (AE) (Harper et al. 2004). During transition to zygotene, the so-called telomere bouquet is formed, in which the chromosome ends attach to the inner nuclear envelope and cluster (Fig. 3.1b). This association is likely to be actively regulated (Scherthan 2007). Recognition between the homologous chromosomes appears to occur between subtelomeric regions (Corredor et al. 2007). The telomere bouquet, found in almost all studied organisms except *Drosophila* and *C. elegans* (Harper et al. 2004), is not absolutely required for the following steps but is proposed to make progression of meiosis more efficient. Both Rabl-configuration and telomere bouquet bundle chromosomes and thereby help reduce the spatial distance between them. After telomere bouquet formation in yeast, the centromeric regions are distributed throughout the nucleus and oscillate to enhance finding homologous sequences (Bass et al. 1997; Scherthan 2007). Chromosome movements are extremely dynamic at this stage (Sheehan and Pawlowski 2009), which increases physical encounters and makes meeting and recognition of the homologs more likely (Pawlowski and Cande 2005).

3.2.1 Homology Recognition in Polyploids

It seems plausible to expect that higher chromosome numbers, and larger nuclei in general, but especially more potential homologous partner chromosomes in

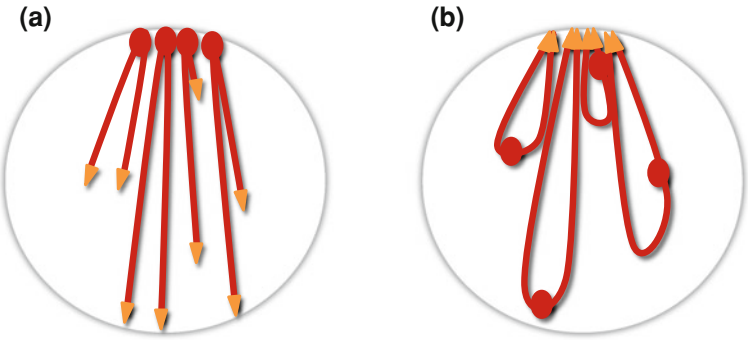

Fig. 3.1 Chromosome arrangement (*red circles* centromeres; *orange triangles* telomeres). **a** Rabl-configuration in interphase, with centromeres clustered at one pole of the nucleus and telomeres oriented toward the other. **b** Bouquet formation during meiosis at the onset of zygotene, with telomeres clustered at the inner nuclear envelope

autopolyploids, could delay progression through meiosis, since the early stages of recognition and pairing can be more complex. Indeed, meiosis in autotetraploid *Saccharomyces cerevisiae* strains is delayed in comparison to its diploid counterpart (Trelles-Sticken et al. 2003). While higher C-values (genomic DNA content) in different plant species also increase the duration of meiosis when compared at the same ploidal level, the comparison of plants with different ploidy within the same species revealed a surprising, negative correlation: higher polyploidy seems to shorten the time needed for completing meiosis (Bennett 1977). These discrepancies might be created by defining the duration only by cytologically visible events. Analysis with higher resolution techniques and consideration of molecular events (Carlton et al. 2006) at the onset of meiosis might provide more accurate information on meiosis kinetics in polyploids. In spite of individual differences, it is remarkable that a wide range of genome and nuclear size variation does not modify the length of the recognition process (Moore and Shaw 2009). It appears that homology recognition is not the limiting step, due to supportive elements like Rabl-configuration and telomere bouquet , specialized chromosome pairing sites as found in *D. melanogaster* and *C. elegans* (McKee 1996; McKim 2007), and intensive chromosome movements that are effective in diploids and polyploids.

3.2.2 Pairing in Autopolyploids

There is one important difference that distinguishes the pachytene stage in autopolyploids: the possibility of pairing between more than two chromosomes, resulting in the formation of multivalents. This requires a certain degree of homology between pairing partners, and this is reflected in different frequency of multivalent occurrence depending on the type of polyploidy, species, individual chromosomes, and chromosome segments (Sybenga 1996). While allopolyploids

with genomes composed of genetically divergent parents often form only bivalents aligned over the whole length of the chromosome, autopolyploids may have multivalents. These are generated by simultaneous alignment of different partners and at different chromosome ends. The probability depends on the degree of ploidy and homology (random-end pairing model by John and Henderson 1962) and is controlled by genetic factors (see below), but it is independent of chromosome length (Morrison and Rajhathy 1960). Although multivalent formation is more often associated with autopolyploidy, many exceptions delimit the universality of this correlation: newly formed autopolyploids often have a lower rate of multivalents than expected, and allopolyploids may form multivalents to some extent (Ramsey and Schemske 2002). A few of these cases might be due to tri- or tetrasomy compensated by lack of other chromosomes, resulting in apparent euploidy (Mestiri et al. 2010).

Progressive pairing starting simultaneously from opposite ends can result in multivalents with pairing partner switches (PPS), so that one chromosome can be aligned with two or more others in different segments (Fig. 3.2a). The distribution of PPS is variable but probably not totally random. More than one switch per chromosome (Fig. 3.2b) indicates the existence of additional, autonomous pairing sites (APS) along the chromosomes, each with a uniform and low probability of generating a PPS (Jones and Vincent 1994). The number of switches indicates the minimal number of pairing sites (Loidl 1995). These are likely not only determined genetically, since a higher number of PPS in autotriploid *Crepis capillaris*, compared to autotetraploid (Jones 1994), indicates other factors than just APS distribution, but probably some interference between pairing initiation sites. Pairing with one partner at one pairing site preferentially promotes continuation of pairing in a zipper-like manner, due to steric constraints. However, this preference is valid only once synapsis has been initiated.

According to the random-end pairing model (John and Henderson 1962), alignment of two chromosome ends enhances the two remaining to pair, although the opposite ends of the chromosome still can pair randomly, resulting in 2/3 multivalents and 1/3 bivalents in tetraploids (only chromosome ends are considered under this model). Exceptions with fewer than expected multivalents (Weiss and Maluszynska 2000; Santos et al. 2003; Carvalho et al. 2010) or bivalent formation due to selective pairing (Simioni and do Valle 2011, and references within) reflect genetic and/or epigenetic heterozygosity between parental chromosome sets, due to ongoing diploidization. This can be different between individual plants or individual chromosomes (Santos et al. 2003) and extremely divergent in complex polyploids, as deduced from genetic maps of sugarcane (Jannoo et al. 2004). Higher than expected ratios of multivalents can also occur, as in newly generated autotetraploids of *Arabidopsis thaliana*, indicating multiple active pairing initiation sites even in small chromosomes (Santos et al. 2003).

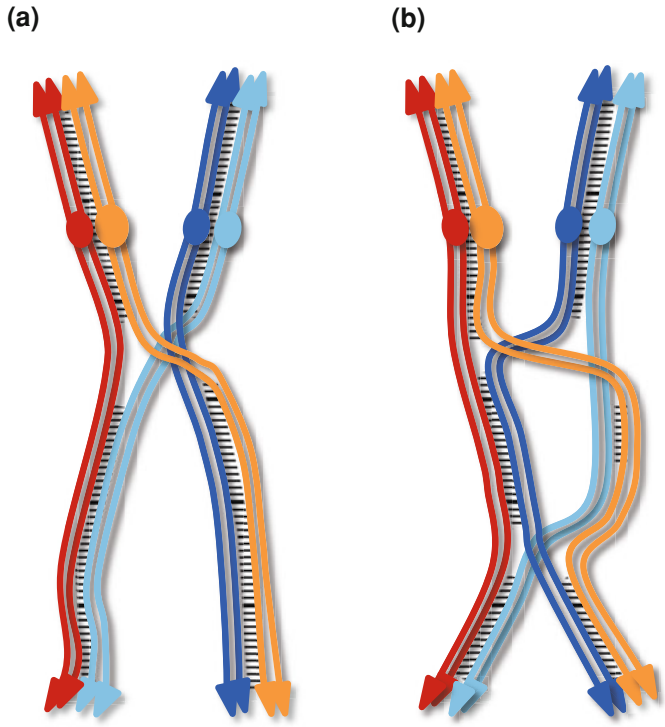

Fig. 3.2 Pairing partner switches during multivalent formation in polyploids. Alignment and partial synapsis between four chromosomes with one (**a**) or two (**b**) pairing partner switches (*PPS*) on the long arm

3.2.3 Pairing in Allopolyploids

Although newly formed hybrids combining different genomes and different chromosome numbers and shapes can exhibit erroneous pairing or lack of alignment (Ozkan and Feldman 2009), established allopolyploids show usually diploid-like pairing and formation of bivalents. Nevertheless, multivalents or segmental multivalent pairing can be found in allopolyploids, depending on external or genetic factors. In many polyploids, a genetic control system has evolved that regulates the pairing behavior during meiosis (Watanabe 1981; Gupta and Fedak 1985; Martinez-Perez et al. 2001; Comai et al. 2003; Martinez-Perez et al. 2003; Jenczewski and Alix 2004). Probably the best studied example was discovered in the hexaploid bread wheat (*Triticum aestivum*, composed of genomes A, B, and D). In the presence of the *Ph*1 locus, pairing and recombination occur preferentially between homologous chromosomes (Riley and Chapman 1958); lack of *Ph*1 leads to a substantial increase of homoeologous pairing. The *Ph*1 locus from wheat can also affect pairing after introgression into rye (Lukaszewski and Kopecky 2010). The absence of *Ph*1 in diploids suggests its emergence through polyploidization (Griffiths et al. 2006). The

*Ph*1 locus is a complex rearrangement between subtelomeric heterochromatin translocated within a cluster of *cdk2*-like genes localized on one of the chromosomes of the B genome (Griffiths et al. 2006).

*Ph*1 controls meiosis at several levels. While telomere pairing is not affected, *Ph*1 prevents pairing between centromeres of nonhomologous chromosomes, keeps centromeres at the nuclear periphery, synchronizes chromatin condensation, and controls the expression of synapsis and recombination components (Naranjo and Corredor 2004; Moore and Shaw 2009; Knight et al. 2010; Yousafzai et al. 2010).

3.2.4 Synapsis

As soon as meiotic chromosomes align, they become joined by a stable proteinaceous structure, the SC. Its formation starts mainly at telomeres (Stack and Anderson 2002) and is intimately connected with DSBs and recombination (see next paragraph). The SC consists of three components: the two parallel AEs, the connecting central element, and periodically occurring recombination nodules (RNs) (Lohmiller et al. 2008). The RNs are multiprotein complexes thought to be involved in synapsis and recombination and can be distinguished by their structure and appearance as early or late RNs (Stack and Anderson 2002; Anderson and Stack 2005).

Synapsis is expected to follow once pairing is established (Loidl 1995). However, unequal chromosome numbers can make a difference in polyploids: in autotriploid, but not in autotetraploid yeast, trivalent SC formation was observed (Loidl 1995). Such fixed multivalents are prone to nondisjunction at metaphase I. Multivalents with equal numbers seem to resolve more often into regular bivalents, although SC connections between several AEs can occur and become fixed by crossovers. Without crossover at the region of the PPS, the central region is twisted and bivalents are formed (von Wettstein et al. 1984). When, and how, are interesting questions, especially for meiosis in autopolyploids.

In summary, recognition, pairing, and synapsis of homologous chromosomes in diploids and polyploids are different in several aspects and determined by various parameters. The high rate of multivalent formation in autopolyploids depends on the type and history of the polyploidization, the number of pairing initiation sites, and odd or even chromosome multiplication. Allopolyploids rather generally form bivalents, but segment-, chromosome-, or genome-specific multivalent formation, and genetic and environmental influence, can also result in meiotic progression different from that in diploids.

3.3 Recombination and Crossover

Meiotic recombination in plants is initiated by multiple DSBs. Some of the breaks are repaired by homologous recombination using the non-sister chromatid as a template, resulting in reciprocal strand exchange (CO events). As the number of

initial DSBs exceeds the number of COs, the majority of DSBs are processed otherwise, either by separating double Holliday junction into non-crossover (NCO) products or by the synthesis-dependent strand annealing (SDSA) pathway. There are excellent recent reviews that describe details of DSB processing and components of meiotic recombination (Hamant et al. 2006; Mercier and Grelon 2008; Sanchez-Moran et al. 2008; De Muyt et al. 2009; Edlinger and Schlogelhofer 2011; Osman et al. 2011), and there is no reason to assume that the principal mechanisms are different in polyploids. However, there is growing evidence that the frequency with which DSBs give rise to COs can be regionally or generally modified by polyploidy. Therefore, we will describe some factors that influence crossover frequencies.

Due to the role of recombination in the regulation of chromosome pairing and segregation, there appears to be a minimum of one CO per chromosome in many species (Youds and Boulton 2011). Beyond this, and in spite of variation between different species, there seems to be no correlation between the length of chromosomes and the number of COs (Brubaker et al. 1999; Mezard 2006). The genomes are rather composed of 'hot spots' and 'cold spots', with high and low probabilities for meiotic recombination (Drouaud et al. 2006; Mezard 2006; Kim et al. 2007). Even small local sequence divergence, like a transgene insertion, can modify meiotic recombination locally (Sun et al. 2008). Variation in recombination or CO frequency among *Arabidopsis* accessions (Barth et al. 2001; Sanchez-Moran et al. 2002) also indicates genetic components that regulate the meiotic recombination frequency in *trans*. Further, temperature and age of flowers within the plant can modify CO events (Francis et al. 2007). Another aspect to consider is CO interference, the inhibition of additional recombination events by COs in their proximity or even along the whole chromosome (Holliday 1977; van Veen and Hawley 2003; Baudat and de Massy 2007; Youds and Boulton 2011).

Differences in the number of CO events in polyploids have been described in several systems. The A and D genome components of allotetraploid *Gossypium* (cotton) are very different in size but marker pairs have, nevertheless, comparable genetic distances if compared between diploid or allopolyploid mapping populations, respectively. Surprisingly, the comparison for the same markers between diploids and allotetraploids indicated a higher recombination rate in the latter (Brubaker et al. 1999; Desai et al. 2006). F_1 hybrids between the diploid parents *Brassica oleracea* and *Brassica rapa*, compared with F_1 plants derived from the same hybrid with doubled chromosome numbers after colchicine treatment, produced more progeny with intergenome recombination (Szadkowski et al. 2010, 2011). A comparison of the CO number among diploid, allotriploid, and allotetraploid hybrids from crosses between *Brassica oleracea* and *Brassica rapa* revealed the highest number for allotriploids, intermediate values for allotetraploids, and the lowest numbers in diploids (Nicolas et al. 2008; Leflon et al. 2010). This indicates stimulation of recombination by hybridity and/or polyploidy, however, in a nonlinear correlation with the number of homologous chromosomes. This might be coupled with the occurrence of univalents, in addition to bivalents, in the triploids, whereas the tetraploids exclusively form

bivalents (Leflon et al. 2010), indicating a control mechanism to prevent multivalents. A good candidate is the *PrBn* locus (Jenczewski et al. 2003; Nicolas et al. 2009; Cifuentes et al. 2010). Exploring an assay system for meiotic recombination based on fluorescent proteins expressed in seeds (Melamed-Bessudo et al. 2005), a comparison among diploids, isogenic autotetraploids, and allotetraploids generated by interspecies hybridization revealed an unexpected increase of recombination between the markers for both types of polyploids (Pecinka et al. 2011). The presence of mainly multivalent formation in one, and bivalent formation in the other, argue against a correlation with the pairing behavior.

In summary, the frequency of recombination between any two markers is determined by interplay of physical distance, local *cis*- and *trans*-acting genetic control elements, the degree of overall genetic divergence, and sex-specific differences (Vizir and Korol 1990; De Vicente and Tanksley 1991; Drouaud et al. 2007; Nelson et al. 2005; Pecinka et al. 2011), regardless of ploidy. The few briefly described examples of increased recombination frequency in polyploid plants suggest that pairing behavior, potential modification of crossover interference by PPS, distinctive segregation patterns (see below), or other still unknown factors add to the complexity. However, it is tempting to speculate that high(er) recombination rates, at least in part, contributed to the prevalence of polyploidy among angiosperms and crop plants bred within the last 12,000 years.

3.4 Chromosome and Allele Segregation

3.4.1 Chromosome Segregation

Following pairing and recombination in prophase I, chromosomes are arranged at the equatorial plane during metaphase I. The following anaphase I is distinct from that in mitosis since both centromeres of sister chromatids attach to spindles of the same pole and retain their cohesion when the chromosome pairs get dragged apart, moving to opposite poles. This 'reductional' division, therefore, halves the chromosome number. It is followed by metaphase II with the formation of two new equatorial planes, and anaphase II, in which the sister chromatids now become separated and distributed to the resulting four postmeiotic nuclei. A tightly controlled order of maintenance and stepwise release of cohesion specifies all stages of meiosis (Sakuno and Watanabe 2009), but there is no evidence that an increased chromosome number in polyploids affects this control.

In contrast, due to the effects of polyploidy on pairing and/or recombination, there is a much higher probability of unequal segregation of chromosomes into the postmeiotic nuclei. Unequal distribution of the chromosomes can occur in many different ways (Pagliarini 2000). It can decrease the regular nuclear DNA content, if unresolved multivalents, unpaired or laggard chromosomes, or parts of chromosomes are not included or translocated to other chromosomes (Madlung et al. 2005;

Charles et al. 2010; Gaeta and Pires 2010; Wang et al. 2010). Extreme cases of systematic chromosome elimination were observed in the allopolyploid *Paspalum subciliatum* (Adamowski et al. 1998) and in pentaploid *Brachiaria decumbens* (Ricci et al. 2010), due to kinetic asynchrony of the two different genomes after diakinesis. Alternatively, the nuclear DNA may be increased if restitution of nuclei occurs prior to division or around incompletely separated chromosomes (Ramsey 2007; Brownfield and Kohler 2011). This can occur either after the first or the second meiotic division, differing in the degree of heterozygosity transmitted to the progeny (Peloquin et al. 2008).

While loss of chromosomal material is often deleterious for the resulting cell and leads to cell death or reduced viability and fertilization, a gain, and optimally a balanced multiplication of all chromosomes, is much less detrimental. Accordingly, unreduced gametes occur in numerous species and are a major source of polyploidization events (Leitch and Leitch 2008; Koehler et al. 2010). The frequency of unreduced gametes varies significantly between hybrids (e.g., Ortiz 1997; Lim et al. 2004), between different ploidal levels (e.g., Burton and Husband 2001; Ramsey 2007), and also depends on genetic factors (reviewed in Brownfield and Kohler 2011), external conditions, and stress factors (reviewed in Ramsey and Schemske 1998). The frequent occurrence of unreduced gametes has been explained by the lack of, or less effective, 'pachytene checkpoint' in plants (Li et al. 2009), but may be an inherent element of circumventing meiotic problems in interspecies hybrids by polyploidization.

The occurrence of diploid or other unreduced post-meiotic cells also has consequences for the subsequent phase of the life cycle. Haploid gametophytes from diploid parents depend on functionality of each single genomic copy that encodes factors for proper gametophyte viability and fertilization potential. Gametophytes from polyploid species contain more than one copy, providing backups for defective alleles and the potential for heterozygosity even during the two and three postmeiotic cell divisions forming the male and female gametophytes, respectively. Since at least pollen selection is efficient at several levels (Ottaviano et al. 1990), avoiding haploidy might provide reduced selection pressure and increased variability of transmitted genetic information.

Not directly related to meiosis but with consequences for gamete formation and fertility is the phenomenon of cytomixis, the fragmentation of chromatin and transfer of fragments among cells through cytoplasmic channels. Observed in many species, and especially in pollen mother cells (e.g., Mursalimov and Deineko 2011), one published comparison between diploid and tetraploid varieties of *Withania somnifera* indicated that the resulting reduction of fertility is much less pronounced in the tetraploids, along with a reduced extent of intercellular connections (Singhal and Kumar 2008). However, a correlation with polyploidy in other species awaits investigation.

3.4.2 Allelic Segregation

The genetic segregation of traits in polyploids can be quite different from that in diploids. Heterozygosity and multiple alleles per gamete increase the combinatorial possibilities in the progeny population substantially and make the segregation patterns deviate from the Mendelian ratios for diploids. This makes QTL analysis or any mapping process complex, as in the extreme case of highly polyploid and partially aneuploid sugarcane (Andru et al. 2011). Naturally, analysis of inheritance requires consideration of pairing behavior and the resulting chromosome segregation. Variation among hybrids, individuals, chromosomes, and chromosome segments does not allow establishing general rules. Therefore, we will depict the extremes: autotetraploids with quadrivalent formation, unbiased chromosome segregation, and tetrasomic inheritance on one hand, and allopolyploids with strict bivalent formation and disomic inheritance on the other. A Punnett square for segregation of one trait in the progeny of heterozygous autotetraploid parents with two copies of two different alleles has 36 fields (Fig. 3.3b), rather than four in the case of a heterozygous diploid (Fig. 3.3a). Deleterious recessive alleles become apparent as affected phenotypes in only 1:35, rather than 1:3 in diploids. Each individual tetraploid can have one, two, three, or four identical alleles (termed simplex, duplex, triplex, or quadruplex) at each locus, and the expected segregation of the different genotypes upon crossing with a partner of the same genotype is 1:8:18:8:1 (Fig. 3.3b). In the case of allopolyploids with strict bivalent formation between the chromosomes of common origin, each gamete carries one of each chromosome type so that the progeny are uniform (Fig. 3.3c). If the chromosomes carry different alleles, autotetraploids can produce 19 different genotypes at one locus (Fig. 3.4a), while disomic inheritance in allotetraploids is limited to 9 combinations (Fig. 3.4b). While exclusive tetrasomic or disomic inheritance in polyploids does occur, intermediate forms and changes in both directions over time are frequently observed in nature and need to be considered in genetic and population studies (Stift et al. 2008).

A genetic peculiarity of polyploid meiosis that does not occur in diploid organisms is the chance of double reduction (Darlington 1929a; Butruille and Boiteux 2000). This term describes the possibility that regions from two sister chromatids become combined in the same gamete. Double reduction results from recombination events between the locus under observation and the centromere. If the two chromosomes that have recombined move to the same pole in anaphase I, there is a high probability that the distal end of the recombined chromatid will be included in the same nucleus as the distal end of the non-recombined chromatid, after anaphase II (locus B in 2 gametes in Fig. 3.5). The frequency of double reduction for a genetic locus depends on its distance from the centromere and the frequency of multivalent formation.

Since segregation of traits in polyploids is so complex, but of high importance for population geneticists and plant breeders, there were many attempts to approach the problem via mathematical modeling. Ground-breaking theoretical work (Mather 1936; Fisher 1947) has more recently stimulated several

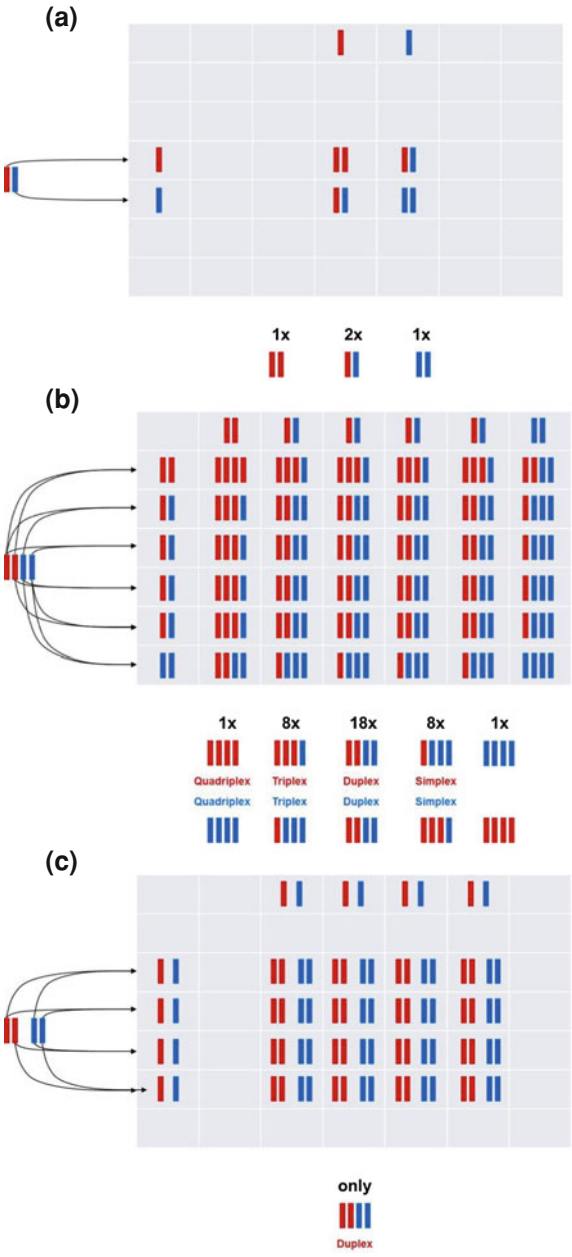

Fig. 3.3 Comparison of segregation for an individual locus with two alleles upon bivalent or quadrivalent formation. Two different alleles (*red*, *blue*) in diploid parents are transmitted with equal probability, resulting in three genotypes, forming twice as many heterozygotes as each homozygote (**a**). Two different alleles with two copies each in tetraploids with random pairing between each of the four chromosomes result in five genotypes, with only one homozygote among 36 individuals (**b**). The same situation in tetraploids with strict bivalent formation between the more similar chromosome pairs renders only one combination (**c**)

computational methods based on maximum-likelihood and other statistical approaches to estimate double reduction frequency and recombination and to support mapping of quantitative trait loci in polyploids with multivalent formation (Doerge and Craig 2000; Ridout et al. 2001; Wu et al. 2001a, b; Luo et al. 2004,

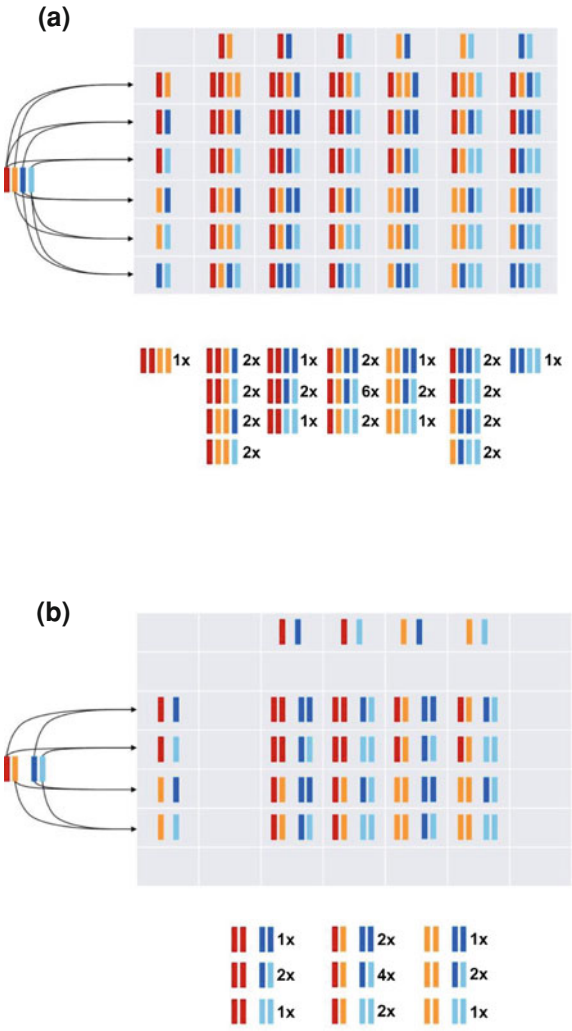

Fig. 3.4 Comparison of segregation for an individual locus with four alleles upon bivalent or quadrivalent formation. Four different alleles (*red*, *orange*, *dark blue*, *light blue*) in tetraploids with random pairing between each of the four chromosomes result in nineteen genotypes (**a**). The same situation in tetraploids with strict bivalent formation between the more similar chromosome pairs renders only nine combinations (**b**)

2006; Li et al. 2010). However, different modeling approaches do not always support the same conclusions (Ma et al. 2002; Cao et al. 2004), and it is likely that we need to extend and refine experimental data collection as well as modeling approaches to provide satisfying tools. Already difficult for strict autopolyploids with mostly polysomic inheritance, the situation is even more complex for polyploids with intermediate types of inheritance, as mentioned above. These can be caused either by segment- or chromosome-specific homology differences, or by stochastic pairing differences. Not even the situation in tetraploid yeast strains, where tetrad analysis provides excellent resolution of segregation analysis, is easy to interpret (Albertin et al. 2009; Stift et al. 2010). Allelic segregation in

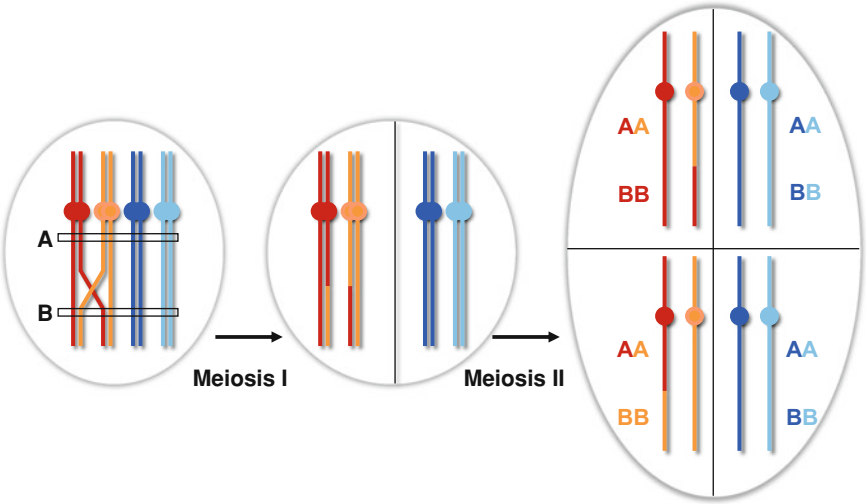

Fig. 3.5 Double reduction. Limited to polyploids, there is a certain probability that distal regions (around locus B) from two sister chromatids are included in the same gamete. This exceptional configuration occurs if two recombined chromosomes move to the same pole in anaphase I, which is excluded in diploids

allopolyploid plants, as described, depends very much on the ratio of multivalent formation at the genetic locus under investigation.

3.5 External Influence

As for other developmental and physiological processes, progress and efficiency of meiosis can be modified by environmental influences. Temperature has been most extensively studied, with a range of different responses. Higher temperatures reduced regular meiosis in diploid *Allium ursinum* (Loidl 1989) and *Rosa* (Pecrix et al. 2011) and increased the number of unreduced pollen (reviewed in Ramsey and Schemske 1998). This could enhance the formation of polyploids under heat stress and thereby support the described adaptive role of polyploidization under adverse conditions. However, data for polyploids are rare, and partially divergent. In allopolyploid *Brassica* hybrids, more unreduced pollen was observed at lower temperatures (Mason et al. 2011), and meiosis in wheat was not affected within a physiological range of temperatures (Bayliss and Riley 1972).

A more uniform effect across species (yeast, worms, and plants) appears to be a stimulation of meiotic recombination rates at higher temperatures (Rose and Baillie 1979; Borner et al. 2004; Francis et al. 2007). It is likely that adaptation to habitats with more or less drastic environmental challenges has selected a matching range of meiotic responses. Further, the more recently established

polyploid hybrids generated by breeders might respond quite differently than polyploids originating from spontaneous events long ago.

3.6 Bypassing Meiosis

Ancient polyploidization events and the prevalence of polyploids among living plants leave no doubt that the ecological and metabolic advantages of multiple chromosome complements preponderate over possible disadvantages (Soltis and Soltis 1999; Ramsey and Schemske 2002; Comai 2005; Otto 2007; Parisod et al. 2010; Jiao et al. 2011). As described, successful meiosis and gamete formation are mastered in different ways. Nevertheless, the survival and success of each new hybridization or polyploidization event depends on multiple factors during meiosis. Besides the requirements to adapt chromosome pairing and segregation, asynchrony, incompatibility of protein complex subunits, dosage effects, or differences in gene regulation can present meiotic barriers.

One solution in plants to avoid such a decrease in reproduction is to avoid meiosis by apomixis. Apomixis, the asexual production of seeds, leads to offspring that is genetically identical to the mother plant and originate via different modifications of the double fertilization pathway (Gustaffson 1946; Asker and Jerling 1992). Gametophytic apomixis (Nogler 1984) is divided into *apospory*, where the unreduced embryo sac is formed by a cell of the nucellus, and *diplospory*, where the unreduced embryo sac is formed by bypassing meiosis I during megasporogenesis. Both result in an embryo sac that contains an unreduced egg cell, which develops into an embryo independent of fertilization. Although apomicts are found in diploids and polyploids, they are more common among polyploids (Asker and Jerling 1992). Apomixis is thought to have evolved several times from sexual ancestors in over 400 plant species (Nogler 1984). It is clearly associated with changes in gene expression, kinetics, or epigenetic regulation (Carman 1997; Sharbel et al. 2010). Whether polyploidization is a driving force toward apomixis, or a consequence of apomictic propagation, is a matter of debate. Apomixis harbors the potential to avoid sexual sterility caused by multivalent formation in polyploids (Horandl et al. 2011). On the other hand, and as described, polyploidy promotes the formation of unreduced gametes (Ramsey and Schemske 1998) and might thereby foster the establishment of higher ploidal levels. Polyploidization of apomicts could also protect them against extinction, by buffering against the irreversible accumulation of deleterious mutations in the absence of recombination as depicted by the metaphor of "Muller's ratchet" (Darlington 1929b). Nevertheless, it is likely that these conditions are relevant in varying and individual combinations, with additional components such as highly efficient DNA repair system in asexually reproducing species (Schoen and Martens 1998) or occasional or conditional sexual reproduction (D'Souza et al. 2004).

3.7 Manipulating Meiosis

While natural selection in sexually reproducing species rapidly eradicates failures in meiosis, these are of interest for plant breeders. If not the seed, but the surrounding fruit is the wanted product, for example in *Citrus* species or melons, customers prefer products with small, reduced, or no seeds. Besides exploiting developmental mutations, one effective and popular way to achieve such fruits is the generation of triploids (Sanchez-Moran et al. 2002) by crossing tetraploids with diploids. Triploid embryos are prone to abort subsequently (Kamiri et al. 2011). Since fruit development is often regulated by phytohormones originating from the developing seeds (Dorcey et al. 2009), this strategy depends on their seed-independent substitution (Pandolfini 2009). However, once generated, triploid plants are often seedless, due to a high number of univalents in meiosis, a resulting low number of balanced gametes, and poor pollination and fertilization rates. This principle is applied in banana (Heslop-Harrison and Schwarzacher 2007) and watermelon (Beaulieu and Lea 2006), or is combined with selection for developmental mutants as for squash (Menezes et al. 2005).

Once a beneficial combination of genetic traits by crossing is achieved, breeders want to fix this for further progeny (Wijnker and de Jong 2008; Chan 2010). Programed switches between crossing and apomixis would, therefore, be desirable (Spillane et al. 2004), especially for polyploids with their complex segregation patterns. Recent advances in provoking apomixis have been made by strategies to disturb meiosis. A mutation in the *Arabidopsis* gene *SWI/DYAD* (Mercier et al. 2001; Agashe et al. 2002) leads to the formation of unreduced egg cells, albeit at low frequencies. This is due to failure in female meiosis and maintains complete maternal heterozygosity in the triploid progeny (Ravi et al. 2008). Combining three mutations affecting different steps of meiosis (*osd*1/*Atspo*11-1/*Atrec*8) turned meiosis into mitosis and resulted in unreduced male and female gametes (d'Erfurth et al. 2009). Consequently, ploidy in subsequent generations was doubled, accompanied by reduced fertility. Nevertheless, the achievement of fertilization-independent seed development is a step toward introducing apomixis into crop plants, and of special interest for polyploid plants.

The work in the OMS lab was supported by grants from the Austrian Science Fund FWF P18986 and I489.

References

Adamowski ED, Pagliarini MS, Batista LAR (1998) Chromosome elimination in paspalum subciliatum (Notata group). Sex Plant Reprod 11(5):272–276

Agashe B, Prasad CK, Siddiqi I (2002) Identification and analysis of DYAD: a gene required for meiotic chromosome organisation and female meiotic progression in Arabidopsis. Development 129(16):3935–3943

Albertin W, Marullo P, Aigle M, Bourgais A, Bely M, Dillmann C, De Vienne D, Sicard D (2009) Evidence for autotetraploidy associated with reproductive isolation in *Saccharomyces cerevisiae*: towards a new domesticated species. J Evol Biol 22(11):2157–2170

Anderson LK, Stack SM (2005) Recombination nodules in plants. Cytogenet Genome Res 109 (1–3):198–204

Andru S, Pan YB, Thongthawee S, Burner DM, Kimbeng CA (2011) Genetic analysis of the sugarcane (Saccharum spp.) cultivar LCP 85-384'. I. Linkage mapping using AFLP, SSR, and TRAP markers. Theor Appl Genet 123(1):77–93

Armstrong SJ, Franklin FCH, Jones GH (2003) A meiotic time-course for *Arabidopsis thaliana*. Sex Plant Reprod 16(3):141–149

Asker S, Jerling L (1992) Apomixis in plants. CRC Press, Boca Raton

Avivi L, Feldman M (1980) Arrangement of chromosomes in the interphase nucleus of plants. Hum Genet 55(3):281–295

Barth S, Melchinger AE, Devezi-Savula B, Lubberstedt T (2001) Influence of genetic background and heterozygosity on meiotic recombination in *Arabidopsis thaliana*. Genome 44(6): 971–978

Bass HW, Marshall WF, Sedat JW, Agard DA, Cande WZ (1997) Telomeres cluster de novo before the initiation of synapsis: a three-dimensional spatial analysis of telomere positions before and during meiotic prophase. J Cell Biol 137(1):5–18

Baudat F, de Massy B (2007) Regulating double-stranded DNA break repair towards crossover or non-crossover during mammalian meiosis. Chromosome Res 15(5):565–577

Bayliss MW, Riley R (1972) Analysis of temperature-dependent asynapsis in *Triticum aestivum*. Genet Res 20(2):193

Beaulieu JC, Lea JM (2006) Characterization and semiquantitative analysis of volatiles in seedless watermelon varieties using solid-phase microextraction. J Agric Food Chem 54(20):7789–7793

Bennett MD (1977) The time and duration of meiosis. Philos Trans R Soc Lond B Biol Sci 277(955):201–226

Bhatt AM, Canales C, Dickinson HG (2001) Plant meiosis: the means to 1 N. Trends Plant Sci 6(3):114–121

Borner GV, Kleckner N, Hunter N (2004) Crossover/noncrossover differentiation, synaptonemal complex formation, and regulatory surveillance at the leptotene/zygotene transition of meiosis. Cell 117(1):29–45

Brownfield L, Kohler C (2011) Unreduced gamete formation in plants: mechanisms and prospects. J Exp Bot 62(5):1659–1668

Brubaker CL, Paterson AH, Wendel JF (1999) Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. Genome 42(2):184–203

Burton TL, Husband BC (2001) Fecundity and offspring ploidy in matings among diploid, triploid and tetraploid *Chamerion angustifolium* (Onagraceae): consequences for tetraploid establishment. Heredity 87:573–582

Butruille DV, Boiteux LS (2000) Selection-mutation balance in polysomic tetraploids: impact of double reduction and gametophytic selection on the frequency and subchromosomal localization of deleterious mutations. Proc Nat Acad Sci USA 97(12):6608–6613

Cao DC, Osborn TC, Doerge RW (2004) Correct estimation of preferential chromosome pairing in autotetraploids. Genome Res 14(3):459–462

Carlton PM, Farruggio AP, Dernburg AF (2006) A link between meiotic prophase progression and crossover control. PLoS Genet 2:119–128

Carman JG (1997) Asynchronous expression of duplicate genes in angiosperms may cause apomixis, bispory, tetraspory, and polyembryony. Biol J Linn Soc 61(1):51–94

Carvalho A, Delgado M, Baroa A, Frescatada M, Ribeiro E, Pikaard CS, Viegas W, Neves N (2010) Chromosome and DNA methylation dynamics during meiosis in the autotetraploid *Arabidopsis arenosa*. Sex Plant Reprod 23(1):29–37

Chan SWL (2010) Chromosome engineering: power tools for plant genetics. Trends Biotechnol 28(12):605–610

St Charles J, Hamilton ML, Petes TD (2010) Meiotic chromosome segregation in triploid strains of *Saccharomyces cerevisiae*. Genetics 186(2):537–550

Cifuentes M, Grandont L, Moore G, Chevre AM, Jenczewski E (2010) Genetic regulation of meiosis in polyploid species: new insights into an old question. New Phytol 186(1):29–36

Comai L (2005) The advantages and disadvantages of being polyploid. Nat Rev Genet 6(11):836–846

Comai L, Tyagi AP, Lysak MA (2003) FISH analysis of meiosis in Arabidopsis allopolyploids. Chromosome Res 11(3):217–226

Corredor E, Lukaszewski AJ, Pachon P, Allen DC, Naranjo T (2007) Terminal regions of wheat chromosomes select their pairing partners in meiosis. Genetics 177(2):699–706

Cowan CR, Carlton PM, Cande WZ (2001) 'The polar arrangement of telomeres in interphase and meiosis. Rabl organization and the bouquet. Plant Physiol 125(2):532–538

d'Erfurth I, Jolivet S, Froger N, Catrice O, Novatchkova M, Mercier R (2009) Turning meiosis into mitosis. PLoS Biol 7(6):e1000124

D'Souza TG, Storhas M, Schulenburg H, Beukeboom LW, Michiels NK (2004) Occasional sex in an 'asexual' polyploid hermaphrodite. Proc Biol Sci 271(1543):1001–1007

Darlington CD (1929a) Chromosome behaviour and structural hybridity in the tradescantiae. J Genet 21(2):207–286

Darlington CD (1929b) Polyploids and polyploidy. Nature 124:62–64

Dawe RK (1998) Meiotic chromosome organization and segregation in plants. Annu Rev Plant Physiol Plant Mol Biol 49:371–395

De Muyt A, Mercier R, Mezard C, Grelon M (2009) Meiotic recombination and crossovers in plants. Genome Dyn 5:14–25

Desai A, Chee PW, Rong J, May OL, Paterson AH (2006) Chromosome structural changes in diploid and tetraploid A genomes of Gossypium. Genome 49(4):336–345

De Vicente MC, Tanksley SD (1991) Genome-wide reduction in recombination of backcross progeny derived from male versus female gametes in an interspecific cross of tomato. Theor Appl Genet 83(2):173–178

Doerge RW, Craig BA (2000) Model selection for quantitative trait locus analysis in polyploids. Proc Nat Acad Sci USA 97(14):7951–7956

Dong F, Jiang J (1998) Non-Rabl patterns of centromere and telomere distribution in the interphase nuclei of plant cells. Chromosome Res 6(7):551–558

Dorcey E, Urbez C, Blazquez MA, Carbonell J, Perez-Amador MA (2009) Fertilization-dependent auxin response in ovules triggers fruit development through the modulation of gibberellin metabolism in Arabidopsis. Plant J 58(2):318–332

Drouaud J, Camilleri C, Bourguignon PY, Canaguier A, Berard A, Vezon D, Giancola S, Brunel D, Colot V, Prum B et al (2006) Variation in crossing-over rates across chromosome 4 of *Arabidopsis thaliana* reveals the presence of meiotic recombination "hot spots". Genome Res 16(1):106–114

Drouaud J, Mercier R, Chelyshev L, Bérard A, Falque M, Martin O, Zanni V, Brunel D, Mézard C (2007) Sex-specific crossover distributions and variations in interference level along *Arabidopsis thaliana* chromosome 4. PLoS Genet 3(6):e106

Edlinger B, Schlogelhofer P (2011) Have a break: determinants of meiotic DNA double strand break (DSB) formation and processing in plants. J Exp Bot 62(5):1545–1563

Fisher RA (1947) The theory of linkage in polysomic inheritance. Philos Trans R Soc Lond B Biol Sci 233(594):55–87

Francis KE, Lam SY, Harrison BD, Bey AL, Berchowitz LE, Copenhaver GP (2007) Pollen tetrad-based visual assay for meiotic recombination in Arabidopsis. Proc Nat Acad Sci USA 104(10):3913–3918

Gaeta RT, Pires JC (2010) Homologous recombination in allopolyploids: the polyploid ratchet. New Phytol 186(1):18–28

Griffiths S, Sharp R, Foote TN, Bertin I, Wanous M, Reader S, Colas I, Moore G (2006) Molecular characterization of Ph1 as a major chromosome pairing locus in polyploid wheat. Nature 439(7077):749–752

Gupta PK, Fedak G (1985) Genetic control of meiotic chromosome pairing in polyploids in the genus Hordeum. Can J Genet Cytol 27(5):515–530
Gustaffson A (1946) Apomixis in higher plants. Lunds Universitets Arsskrift 42(2):1–370
Hamant O, Ma H, Cande WZ (2006) Genetics of meiotic prophase I in plants. Annu Rev Plant Biol 57:267–302
Harper L, Golubovskaya I, Cande WZ (2004) A bouquet of chromosomes. J Cell Sci 117(Pt 18): 4025–4032
Harrison CJ, Alvey E, Henderson IR (2010) Meiosis in flowering plants and other green organisms. J Exp Bot 61(11):2863–2875
Heslop-Harrison JS, Schwarzacher T (2007) Domestication, genomics and the future for banana. Ann Bot 100(5):1073–1084
Holliday R (1977) Recombination and meiosis. Philos Trans R Soc Lond B Biol Sci 277(955): 359–370
Horandl E, Cosendai AC, Rodewald J (2011) Origin and distribution of autopolyploids via apomixis in the alpine species *Ranunculus kuepferi* (Ranunculaceae). Taxon 60(2):355–364
Jannoo N, Grivet L, David J, D'Hont A, Glaszmann JC (2004) Differential chromosome pairing affinities at meiosis in polyploid sugarcane revealed by molecular markers. Heredity 93(5):460–467
Jenczewski E, Alix K (2004) From diploids to allopolyploids: the emergence of efficient pairing control genes in plants. Crit Rev Plant Sci 23(1):21–45
Jenczewski E, Eber F, Grimaud A, Huet S, Lucas MO, Monod H, Chevre AM (2003) PrBn, a major gene controlling homologous pairing in oilseed rape (*Brassica napus*) haploids. Genetics 164(2):645–653
Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Yi, Liang H, Soltis PS et al (2011) Ancestral polyploidy in seed plants and angiosperms. Nature 473(7345):97–100
John B, Henderson SA (1962) Asynapsis and polyploidy in *Schistocerca paranensis*. Chromosoma 13(2):111–147
Jones GH (1994) Meiosis in autopolyploid crepis-capillaris 3. Comparison of triploids and tetraploids: evidence for nonindependence of autonomous pairing sites. Heredity 73:215–219
Jones GH, Vincent JE (1994) Meiosis in autopolyploid crepis capillaris 2. Autotetraploids. Genome 37(3):497–505
Kamiri M, Stift M, Srairi I, Costantino G, Moussadik AE, Hmyene A, Bakry F, Ollitrault P, Froelicher Y (2011) Evidence for non-disomic inheritance in a citrus interspecific tetraploid somatic hybrid between *C. reticulata* and *C. limon* using SSR markers and cytogenetic analysis. Plant Cell Rep 30(8):1415–1425
Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S, Ecker JR, Weigel D, Nordborg M (2007) Recombination and linkage disequilibrium in *Arabidopsis thaliana*. Nat Genet 39(9):1151–1155
Knight E, Greer E, Draeger T, Thole V, Reader S, Shaw P, Moore G (2010) Inducing chromosome pairing through premature condensation: analysis of wheat interspecific hybrids. Funct Integr Genomics 10(4):603–608
Koehler C, Mittelsten OS, Erilova A (2010) The impact of the triploid block on the origin and evolution of polyploid plants. Trends Genet 26(3):142–148
Leflon M, Grandont L, Eber F, Huteau V, Coriton O, Chelysheva L, Jenczewski E, Chevre AM (2010) Crossovers get a boost in brassica allotriploid and allotetraploid hybrids. Plant Cell 22(7):2253–2264
Leitch AR, Leitch IJ (2008) Genomic plasticity and the diversity of polyploid plants. Science 320(5875):481–483
Li J, Das K, Fu G, Tong C, Li Y, Tobias C, Wu R (2010) Em algorithm for mapping quantitative trait Loci in multivalent tetraploids. Int J Plant Genomics 2010:216547
Li XC, Barringer BC, Barbash DA (2009) The pachytene checkpoint and its relationship to evolutionary patterns of polyploidization and hybrid sterility. Heredity 102(1):24–30

Lim KB, Shen TM, Barba-Gonzalez R, Ramanna MS, Van Tuyl JM (2004) Occurrence of SDR 2 N-gametes in Lilium hybrids. Breed Sci 54(1):13–18

Lohmiller LD, De Muyt A, Howard B, Offenberg HH, Heyting C, Grelon M, Anderson LK (2008) Cytological analysis of MRE11 protein during early meiotic prophase I in Arabidopsis and tomato. Chromosoma 117(3):277–288

Loidl J (1989) Effects of elevated temperature on meiotic chromosome synapsis in *Allium ursinum*. Chromosoma 97(6):449–458

Loidl J (1995) Meiotic chromosome pairing in triploid and tetraploid *Saccharomyces cerevisiae*. Genetics 139(4):1511–1520

Lukaszewski AJ, Kopecky D (2010) The Ph1 locus from wheat controls meiotic chromosome pairing in autotetraploid rye (*Secale cereale L.*). Cytogenet Genome Res 129(1–3):117–123

Luo ZW, Zhang RM, Kearsey MJ (2004) Theoretical basis for genetic linkage analysis in autotetraploid species. Proc Nat Acad Sci USA 101(18):7040–7045

Luo ZW, Zhang Z, Zhang RM, Pandey M, Gailing O, Hattemer HH, Finkeldey R (2006) Modeling population genetic data in autotetraploid species. Genetics 172(1):639–646

Ma CX, Casella G, Shen ZJ, Osborn TC, Wu RL (2002) A unified framework for mapping quantitative trait loci in bivalent tetraploids using single-dose restriction fragments: a case study from alfalfa. Genome Res 12(12):1974–1981

Madlung A, Tyagi AP, Watson B, Jiang HM, Kagochi T, Doerge RW, Martienssen R, Comai L (2005) Genomic changes in synthetic Arabidopsis polyploids. Plant J 41(2):221–230

Martinez-Perez E, Shaw P, Aragon-Alcaide L, Moore G (2003) Chromosomes form into seven groups in hexaploid and tetraploid wheat as a prelude to meiosis. Plant J 36(1):21–29

Martinez-Perez E, Shaw P, Moore G (2001) The Ph1 locus is needed to ensure specific somatic and meiotic centromere association. Nature 411(6834):204–207

Mason AS, Nelson MN, Yan G, Cowling WA (2011) Production of viable male unreduced gametes in Brassica interspecific hybrids is genotype specific and stimulated by cold temperatures. BMC Plant Biol 11:103

Mather K (1936) Segregation and linkage in autotetraploids. J Genetics 32(2):287–314

McKee BD (1996) The license to pair: identification of meiotic pairing sites in Drosophila. Chromosoma 105(3):135–141

McKim KS (2007) Meiotic pairing: a place to hook up. Curr Biol 17(5):R165–R168

Melamed-Bessudo C, Yehuda E, Stuitje AR, Levy AA (2005) A new seed-based assay for meiotic recombination in *Arabidopsis thaliana*. Plant J 43(3):458–466

Menezes CB, Maluf WR, Azevedo SM, Faria MV, Nascimento IR, Nogueira DW, Gomes LA, Bearzoti E (2005) Inheritance of parthenocarpy in summer squash (*Cucurbita pepo L.*). Genet Mol Res 4(1):39–46

Mercier R, Grelon M (2008) Meiosis in plants: ten years of gene discovery. Cytogenet Genome Res 120(3–4):281–290

Mercier R, Vezon D, Bullier E, Motamayor JC, Sellier A, Lefevre F, Pelletier G, Horlow C (2001) SWITCH1 (SWI1): a novel protein required for the establishment of sister chromatid cohesion and for bivalent formation at meiosis. Genes Dev 15(14):1859–1871

Mestiri I, Chague V, Tanguy AM, Huneau C, Huteau V, Belcram H, Coriton O, Chalhoub B, Jahier J (2010) Newly synthesized wheat allohexaploids display progenitor-dependent meiotic stability and aneuploidy but structural genomic additivity. New Phytol 186(1):86–101

Mezard C (2006) Meiotic recombination hotspots in plants. Biochem Soc Trans 34:531–534

Mezard C, Vignard J, Drouaud J, Mercier R (2007) The road to crossovers: plants have their say. Trends Genet 23(2):91–99

Moore G, Shaw P (2009) Improving the chances of finding the right partner. Curr Opin Genet Dev 19(2):99–104

Morrison JW, Rajhathy T (1960) Frequency of quadrivalents in autotetraploid plants. Nature 187(4736):528–530

Mursalimov SR, Deineko EV (2011) An ultrastructural study of cytomixis in tobacco pollen mother cells. Protoplasma 248(4):717–724

Naranjo T, Corredor E (2004) Clustering of centromeres precedes bivalent chromosome pairing of polyploid wheats. Trends Plant Sci 9(5):214–217
Nelson MN, Nixon J, Lydiate DJ (2005) Genome-wide analysis of the frequency and distribution of crossovers at male and female meiosis in *Sinapis alba* L. (white mustard). Theor Appl Genet 111(1):31–43
Nicolas SD, Leflon M, Liu Z, Eber F, Chelysheva L, Coriton O, Chevre AM, Jenczewski E (2008) Chromosome 'speed dating' during meiosis of polyploid Brassica hybrids and haploids. Cytogenet Genome Res 120(3–4):331–338
Nicolas SD, Leflon M, Monod H, Eber F, Coriton O, Huteau V, Chevre AM, Jenczewski E (2009) Genetic regulation of meiotic cross-overs between related genomes in *Brassica napus* haploids and hybrids. Plant Cell 21(2):373–385
Nogler GA (1984) Gametophytic apomixis. In: Johri BM (ed) Embryology of angiosperms. Springer, Berlin
Ortiz R (1997) Occurrence and inheritance of 2n pollen in Musa. Ann Bot 79(4):449–453
Osman K, Higgins JD, Sanchez-Moran E, Armstrong SJ, Franklin F, Chris H (2011) Pathways to meiotic recombination in *Arabidopsis thaliana*. New Phytol 190(3):523–544
Ottaviano E, Sari Gorla M, Mulcahy DL (1990) Pollen selection: efficiency and monitoring. In: Ogita ZI, Markert CL (eds) Isozymes: structure, function, and use in biology and medicine. Wiley-Liss, New York, pp 575–588
Otto SP (2007) The evolutionary consequences of polyploidy. Cell 131(3):452–462
Ozkan H, Feldman M (2009) Rapid cytological diploidization in newly formed allopolyploids of the wheat (*Aegilops-Triticum*) group. Genome 52(11):926–934
Pagliarini MS (2000) Meiotic behavior of economically important plant species: the relationship between fertility and male sterility. Genet Mol Biol 23(4):997–1002
Pandolfini T (2009) Seedless fruit production by hormonal regulation of fruit set. Nutrients 1(2):168–177
Parisod C, Holderegger R, Brochmann C (2010) Evolutionary consequences of autopolyploidy. New Phytol 186(1):5–17
Pawlowski WP, Cande WZ (2005) Coordinating the events of the meiotic prophase. Trends Cell Biol 15(12):674–681
Pawlowski WP (2010) Chromosome organization and dynamics in plants. Curr Opin Plant Biol 13(6):640–645
Pecinka A, Fang W, Rehmsmeier M, Levy Avraham A, Mittelsten Scheid O (2011) Polyploidization increases meiotic recombination frequency in Arabidopsis. BMC Biol 9:24
Pecrix Y, Rallo G, Folzer H, Cigna M, Gudin S, Le Bris M (2011) Polyploidization mechanisms: temperature environment can induce diploid gamete formation in Rosa sp. J Exp Bot 62(10): 3587–3597
Peloquin SJ, Boiteux LS, Simon PW, Jansky SH (2008) A chromosome-specific estimate of transmission of heterozygosity by 2n gametes in potato. J Hered 99(2):177–181
Ramsey J (2007) Unreduced gametes and neopolyploids in natural populations of *Achillea borealis* (Asteraceae). Heredity 98(3):143–150
Ramsey J, Schemske DW (1998) Pathways, mechanisms, and rates of polyploid formation in flowering plants. Annu Rev Ecol Syst 29:467–501
Ramsey J, Schemske DW (2002) Neopolyploidy in flowering plants. Annu Rev Ecol Syst 33:589–639
Ravi M, Marimuthu MP, Siddiqi I (2008) Gamete formation without meiosis in Arabidopsis. Nature 451(7182):1121–1124
Ricci GCL, Pagliarini MS, Valle CB (2010) Genome elimination during microsporogenesis in two pentaploid accessions of *Brachiaria decumbens* (Poaceae). Genet Mol Res 9(4): 2364–2371
Ridout MS, Bell JA, Simpson DW (2001) Analysis of segregation data from selfed progeny of allopolyploids. Heredity 87:537–543
Riley R, Chapman V (1958) Genetic control of the cytologically diploid behaviour of hexaploid wheat. Nature 182(4637):713–715

Rose AM, Baillie DL (1979) Effect of temperature and parental age on recombination and nondisjunction in *Caenorhanditis elegans*. Genetics 92(2):409–418
Sakuno T, Watanabe Y (2009) Studies of meiosis disclose distinct roles of cohesion in the core centromere and pericentromeric regions. Chromosome Res 17(2):239–249
Sanchez-Moran E, Armstrong SJ, Santos JL, Franklin FCH, Jones GH (2002) Variation in chiasma frequency among eight accessions of *Arabidopsis thaliana*. Genetics 162(3): 1415–1422
Sanchez-Moran E, Osman K, Higgins JD, Pradillo M, Cunado N, Jones GH, Franklin FCH (2008) ASY1 coordinates early events in the plant meiotic recombination pathway. Cytogenet Genome Res 120(3–4):302–312
Santos JL, Alfaro D, Sanchez-Moran E, Armstrong SJ, Franklin FC, Jones GH (2003) Partial diploidization of meiosis in autotetraploid *Arabidopsis thaliana*. Genetics 165(3):1533–1540
Scherthan H (2007) Telomere attachment and clustering during meiosis. Cell Mol Life Sci 64(2): 117–124
Schoen I, Martens K (1998) DNA repair in ancient asexuals: a new solution to an old problem? J Nat Hist 32:943–948
Schubert I, Shaw P (2011) Organization and dynamics of plant interphase chromosomes. Trends Plant Sci 16(5):273–281
Sharbel TF, Voigt ML, Corral JM, Galla G, Kumlehn J, Klukas C, Schreiber F, Vogel H, Rotter B (2010) Apomictic and sexual ovules of Boechera display heterochronic global gene expression patterns. Plant Cell 22(3):655–671
Sheehan MJ, Pawlowski WP (2009) Live imaging of rapid chromosome movements in meiotic prophase I in maize. Proc Natl Acad Sci USA 106(49):20989–20994
Simioni C, do Valle CB (2011) Meiotic analysis in induced tetraploids of *Brachiaria decumbens* Stapf. Crop Breed Appl Biotech 11(1):43–49
Singhal VK, Kumar P (2008) Cytomixis during microsporogenesis in the diploid and tetraploid cytotypes of *Withania somnifera* (*L.*) *Dunal*, 1852 (Solanaceae). Comp Cytogenet 2(1):85–92
Soltis DE, Soltis PS (1999) Polyploidy: recurrent formation and genome evolution. Trends Ecol Evol 14(9):348–352
Spillane C, Curtis MD, Grossniklaus U (2004) Apomixis technology development-virgin births in farmers fields? Nat Biotechnol 22(6):687–691
Stack SM, Anderson LK (2002) Crossing over as assessed by late recombination nodules is related to the pattern of synapsis and the distribution of early recombination nodules in maize. Chromosome Res 10(4):329–345
Stift M, Reeve R, van Tienderen PH (2010) Inheritance in tetraploid yeast revisited: segregation patterns and statistical power under different inheritance models. J Evol Biol 23(7):1570–1578
Stift Marc, Berenos C, Kuperus P, van Tienderen PH (2008) Segregation models for disomic, tetrasomic and intermediate inheritance in tetraploids: a general procedure applied to Rorippa (Yellow cress) microsatellite data. Genetics 179(4):2113–2123
Sun X, Zhang Y, Yang S, Chen JQ, Hohn B, Tian D (2008) Insertion DNA promotes ectopic recombination during meiosis in Arabidopsis. Mol Biol Evol 25(10):2079–2083
Sundstrom G, Larsson TA, Larhammar D (2008) Phylogenetic and chromosomal analyses of multiple gene families syntenic with vertebrate Hox clusters. BMC Evol Biol 8:254
Sybenga J (1996) Chromosome pairing affinity and quadrivalent formation in polyploids: do segmental allopolyploids exist? Genome 39(6):1176–1184
Szadkowski E, Eber F, Huteau V, Lode M, Coriton O, Jenczewski E, Chevre AM (2011) Polyploid formation pathways have an impact on genetic rearrangements in resynthesized *Brassica napus*. New Phytol 191(3):884–894
Szadkowski E, Eber F, Huteau V, Lode M, Huneau C, Belcram H, Coriton O, Manzanares-Dauleux MJ, Delourme R, King GJ et al (2010) The first meiosis of resynthesized *Brassica napus*, a genome blender. New Phytol 186(1):102–112
Trelles-Sticken E, Loidl J, Scherthan H (2003) Increased ploidy and KAR3 and SIR3 disruption alter the dynamics of meiotic chromosomes and telomeres. J Cell Sci 116(12):2431–2442

van Veen JE, Hawley RS (2003) Meiosis: when even two is a crowd. Curr Biol 13(21): R831–R833

Vizir IY, Korol AB (1990) Sex difference in recombination frequency in Arabidopsis. Heredity 65(3):379–383

von Wettstein D, Rasmussen SW, Holm PB (1984) The synaptonemal complex in genetic segregation. Annu Rev Genet 18:331–413

Wang Y, Jha AK, Chen R, Doonan JH, Yang M (2010) Polyploidy-associated genomic instability in *Arabidopsis thaliana*. Genesis 48(4):254–263

Watanabe K (1981) Studies on the control of diploid-like meiosis in polyploid taxa of *Chrysanthemim japonese*. Cytologia 46(3):459–498

Weiss H, Maluszynska J (2000) Chromosomal rearrangement in autotetraploid plants of *Arabidopsis thaliana*. Hereditas 133(3):255–261

Wijnker E, de Jong H (2008) Managing meiotic recombination in plant breeding. Trends Plant Sci 13(12):640–646

Wu RL, Gallo-Meagher M, Littell RC, Zeng ZB (2001a) A general polyploid model for analyzing gene segregation in outcrossing tetraploid species. Genetics 159(2):869–882

Wu SS, Wu RL, Ma CX, Zeng ZB, Yang MC, Casella G (2001b) A multivalent pairing model of linkage analysis in autotetraploids. Genetics 159(3):1339–1350

Youds JL, Boulton SJ (2011) The choice in meiosis: defining the factors that influence crossover or non-crossover formation. J Cell Sci 124(Pt 4):501–513

Yousafzai FK, Al-Kaff N, Moore G (2010) The molecular features of chromosome pairing at meiosis: the polyploid challenge using wheat as a reference. Funct Integr Genomics 10(2):147–156

Chapter 4
Origins of Novel Phenotypic Variation in Polyploids

Patrick Finigan, Milos Tanurdzic and Robert A. Martienssen

Abstract Polyploid species represent a special type of organism in nature, one that can survive and compete with three or more full sets of homologous chromosomes. While less common in the animal and fungal kingdoms, polyploid species are highly prevalent in the plant kingdom. Indeed, most agricultural crops are polyploids, typically because polyploidy confers greater robustness and therefore higher yields. Among many examples of novel phenotypic variation exhibited by polyploids are the production of larger fruits, reduced tillering, delays in the reproductive transition, and even the creation of visually stunning flower pigmentation patterns coveted by gardeners. The source of this novel variation in polyploids is still largely unclear. However, multiple cellular mechanisms have been proposed, with some supporting evidence, to explain novel variation. We review some of these mechanisms here.

4.1 Prevalence and Significance of Polyploids

Polyploid organisms have three or more complete sets of homologous chromosomes (Winge 1917; Ramsey and Schemske 2002). Polyploidy, or whole-genome duplication (WGD), can arise through multiple ways—most often from unreduced gametes following meiosis (Ahloowalia and Garber 1961; Harlan and deWet 1975; Jørgensen 1928; Newton and Pellew 1929; Skalinska 1946; Ramsey and Schemske 1998; Bretagnolle and Thompson 1995). In mammals, WGD events are considered to be rare and typically lethal, presumably due to deleterious effects associated

P. Finigan · M. Tanurdzic · R. A. Martienssen (✉)
Cold Spring Harbor Laboratory, One Bungtown Road, Cold Spring Harbor, NY 11724, USA
e-mail: martiens@cshl.edu

P. S. Soltis and D. E. Soltis (eds.), *Polyploidy and Genome Evolution*,
DOI: 10.1007/978-3-642-31442-1_4,

with dosage (Bertrand et al. 2010; Van de Peer et al. 2009; see also Chap. 18, this volume). However, genomic sequencing has revealed evidence for a few widespread ancient WGD events in the animal kingdom (Maere et al. 2005; Dehal and Boore 2005; Kellis et al. 2004; Ohno 1970; Meyer and Van de Peer 2005; Meyer and Schartl 1999) (see also Chaps. 16 and 17, this volume). Extant vertebrate genomes in particular are believed to be the result of two or three separate WGD events during evolution (Maere et al. 2005; Dehal and Boore 2005; Kellis et al. 2004; Ohno 1970; Meyer and Van de Peer 2005; Meyer and Schartl 1999) (Chap. 16, this volume). In plants, polyploidy is generally tolerated, and most plant species are recent or ancient polyploids (Lexer et al. 2003; Soltis et al. 2009; Jiao et al. 2011). Polyploidy is thought to play a major role in speciation (Clausen et al. 1945; Grant 1981; Lumaret 1988), as polyploids are often reproductively isolated from their progenitors—crosses between diploids and tetraploids generally yield triploid progeny with unbalanced meiosis, leading to polyploid infertility (Darlington 1963; Grant 1981).

On a simplified level, polyploid organisms can be either autopolyploids or allopolyploids (Kihara and Ono 1926). Following one definition, an autopolyploid is a polyploid organism in which all of the chromosome sets are derived from the same species, whereas an allopolyploid organism has chromosome sets derived from different species (Kihara and Ono 1926). Allopolyploids are therefore permanent hybrids, but with a complete chromosome set from each parental species (Kihara and Ono 1926). The complete parental chromosome sets in allopolyploids allow for proper pairing of homologous chromosomes during meiosis (disomic segregation), rather than missegregation of trivalents and quadrivalents as may be found in autopolyploids (multisomic segregation) (Stebbins 1971; Ramsey and Schemske 2002). Thus, allopolyploids have often been considered more stable than autopolyploids, and "fix" hybrid genotypes in successive generations, along with any beneficial or detrimental phenotypes that result (Winge 1932). However, it has been recognized that autopolyploids may have functional (either disomic or multisomic) chromosome pairing and be highly fertile (Soltis et al. 2007; Soltis and Rieseberg 1986).

Many important agricultural crops are polyploids (Eigsti 1957). Polyploid crops typically demonstrate increased growth, including flower and fruit size, and novel variation compared to their diploid counterparts that can make them better suited as agricultural products (Stebbins 1971; Ramsey and Schemske 2002; Grant 1981). The modern-day bread wheat, *Triticum aestivum*, is an allohexaploid that is the combination of three different diploid species (Dubcovsky and Dvorak 2007) (see also Chap. 7, this volume). *T. aestivum* has largely replaced its diploid progenitors and now accounts for about 95 % of the entire wheat crop produced around the world (Dubcovsky and Dvorak 2007). In addition to wheat, many staple crops (including corn, cotton, coffee, oat, canola, rye, apple, banana, watermelon, potato, sugar cane, and soybean) are also recent or ancient polyploids (Gaut and Doebley 1997; Stebbins 1971; Ohno 1970; Shoemaker et al. 1996; Lagercrantz and Lydiate 1996; Eigsti 1957). The reduced seed set of certain polyploids (odd ploidal levels) can make them more desirable for consumption, like seedless watermelons

and bananas. Moreover, allopolyploids can overcome hybrid incompatibility between different species, allowing production of new varieties and introgression of favorable traits into important crops (Eigsti 1957). Allohexaploid *Nicotiana* plants were used as a bridge to transfer the gene responsible for tobacco mosaic virus resistance from wild to commercial varieties of *Nicotiana* species (Eigsti 1957). This was only possible through a polyploid intermediate, as the interspecies *Nicotiana* hybrids were sterile (Eigsti 1957) (see also Chap. 11, this volume).

4.2 Origin of Novel Variation in Neopolyploids

The prevalence of polyploids in the plant kingdom despite the rarity of their formation (Ramsey and Schemske 1998) indicates that polyploids may have a fitness advantage compared to their progenitors (Stebbins 1950). The explanation for this fitness advantage stems from the observation that many polyploids display novel phenotypic variation compared to their progenitors (Randolph 1941; Levin 1983, 2002; Lumaret 1988; Ramsey and Schemske 2002; Müntzing 1936). This novel phenotypic variation is believed to enable polyploids to exploit different environmental niches better than their progenitors; this variation bestows a fitness advantage on the polyploids (Clausen et al. 1945; Stebbins 1950). One caveat of this hypothesis is that studies documenting the geographic distribution of polyploids and their diploid progenitors used extant diploid relatives instead of the exact diploid progenitors (usually unknown) (Stebbins 1971; Ramsey and Schemske 2002). Thus, it is possible the real progenitors had the same geographic range distribution as the current polyploids. To address this problem, many current studies of polyploids use synthetic polyploids that can be compared to their exact diploid progenitors (Stebbins 1971; Ramsey and Schemske 2002). Future studies should help to clarify if range expansion relative to their progenitors is a general characteristic of polyploids.

The emergence of novel phenotypes in polyploids, especially neoallopolyploids (defined as early-generation polyploids), compared to their progenitors has been of great interest because of its application in plant breeding and relevance to biodiversity and evolution (Eigsti 1957; Stebbins 1971; Ramsey and Schemske 2002; Grant 1981). Genetic mechanisms encompassing gene dosage/allelic combinations, novel gene interactions, genomic alterations, and epigenomic reorganization have all been proposed to explain the origin of novel phenotypes of polyploids (Fig. 4.1). The relevance of these different mechanisms for causing novel variation is still largely unclear, with important distinctions between autopolyploids and allopolyploids.

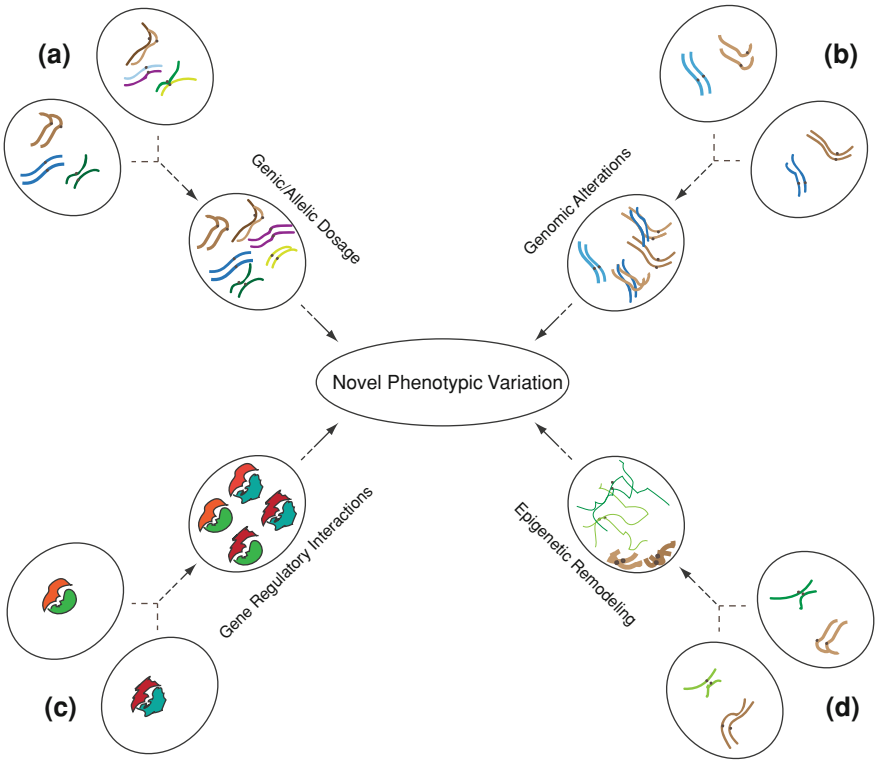

Fig. 4.1 Molecular mechanisms that may contribute to *novel phenotypic variation*. Four distinct molecular mechanisms have been suggested to contribute to changes in gene expression and function that may result in novel phenotypic variation in neoallopolyploids compared to their progenitors—adapted from (Osborn et al. 2003). The figure diagrams each mechanism by showing the two progenitor states in comparison with their resultant polyploid offspring. **a** Changes in *genic* or *allelic dosage* are diagramed by the inheritance of novel chromosome sets. **b** *Genomic alterations* are shown by the inheritance of recombined chromosomes. **c** Novel *gene regulatory interactions* are diagrammed by unique protein–protein interactions. **d** Epigenomic reorganization is shown as the inheritance of large-scale changes in chromatin compaction. The progenitors and their polyploid offspring are diagrammed by ellipses, with the progenitors represented by the ellipses furthest from the central "novel phenotypic variation" ellipse. The polyploids are represented by the ellipses closest to the central ellipse.

4.2.1 Gene Dosage and Allelic Combinations May Result in Novel Variation in Neopolyploids

Genome-wide increases in DNA content have been proposed to contribute to novel phenotypes in neopolyploids (Randolph 1941). The "gigas effect" is a well-documented Phenotype associated with polyploid formation, where certain organs are larger in the polyploids than their progenitors—but not necessarily the entire plant (Randolph 1941; Stebbins 1971). The gigas effect is believed to be a consequence

of increasing the genome size, leading to the increase in cell size and reduction of cell divisions that are commonly associated with neopolyploids compared to their progenitors (Noggle 1946; Stebbins 1971). Additionally, some evidence suggests that neopolyploids have a protracted reproductive growth phase, with a delay in the onset of flowering and longer reproductive growth (Stebbins 1971). While these changes in morphology and growth may be the result of simply increasing the amount of DNA in neopolyploids, they may also stem from increasing the relative level of gene and protein expression.

Gene and protein expression may be sensitive to gene dosage effects, whereby increases in ploidy result in changes in expression that are not proportional to the increase in gene dosage. Gene and protein expression studies in *Zea*, *Solanum*, *Helianthus*, and *Arabidopsis* autoploid series have suggested that the vast majority of changes in expression are proportional to gene dosage (Yu et al. 2010; Pignatta et al. 2010; Wang et al. 2006a, b; Stupar et al. 2007; Guo et al. 1996; Yao et al. 2011; Riddle et al. 2010; Church and Spaulding 2009). Pignatta et al. (2010) created independent neo-autotetraploid lines of *Arabidopsis thaliana* (Columbia) from the same homogeneous diploid progenitor. The researchers performed genome-wide gene expression studies to identify reproducible changes in gene expression in the autotetraploids compared to the diploid progenitor. Only a few genes, out of 26,107, were found that potentially displayed changes in gene expression that were not proportional to the ploidal levels. Of note, resolution or sampling limitations may have missed some dosage-sensitive genes (Pignatta et al. 2010). A recent study with Columbia and Landsberg accessions of *Arabidopsis* ploidal series found similar results to those above, with slightly higher numbers of genes whose expression was disproportionately affected by changes in ploidy (Yu et al. 2010). A greater role for gene dosage was identified in potato and maize studies, as increasing the ploidal levels resulted in roughly 10 % of genes with disproportionate effects on expression (Stupar et al. 2007; Guo et al. 1996; Yao et al. 2011; Riddle et al. 2010; Church and Spaulding 2009). Taken together, these studies suggest that most gene expression is proportional to ploidal level. None of these studies identified dosage-sensitive genes whose increase in expression correlated with any ploidy-dependent Phenotype. One unanswered question is: what effect does increasing the overall level of gene expression, without changing the relative abundance of gene products, have on biological pathways?

In theory, increased gene dosage and mixtures in polyploids could result in unique allelic combinations that would expand phenotypic diversity (Bingham 1979; Grant 1981; Stebbins 1971). Instead of possessing only two alleles for a given gene as in a diploid, a tetraploid could possess four alleles that may allow additional variability for a phenotypic trait. This mechanism of allelic interaction is a natural extension of phenotypic variation in diploids attributed to combinations of different alleles that affect quantitative traits (Guo and Birchler 1994; Birchler et al. 2001; Osborn et al. 2003). However, it remains unclear what allelic combinations contribute to novel variation in polyploids (Osborn et al. 2003), and indeed allelic variation in neopolyploids may be limited due to genetic "bottlenecking" in their formation.

4.2.2 Hybrid Regulatory Interactions Can Cause Novel Phenotypic Variation

Hybrid gene interactions may contribute to phenotypic variation in allopolyploids (Osborn et al. 2003; Ramsey and Schemske 2002). Similar to interspecies hybrids, allopolyploids have divergent genetic contributions from their progenitors. This heterogeneity could result in perturbations in regulatory networks that could dramatically alter their outcomes (Ramsey and Schemske 2002; Stebbins 1947, 1950; Osborn et al. 2003). In synthetic *Arabidopsis* allopolyploids, altered regulatory interactions between the two progenitor-derived chromosome sets are thought to be responsible for delaying flowering time (Wang et al. 2006a, b). The *A. arenosa FRIGIDA* (*FRI*) gene is a transcription factor that *trans*-activates the *A. thaliana* (Landsberg) *Flowering Locus C* (*FLC*) gene (Wang et al. 2006a, b). *FLC* is a major negative regulator of the reproductive transition in *Arabidopsis* species, and its over-expression is predicted to delay flowering in the *Arabidopsis* neoallopolyploids (Shindo et al. 2005; Simpson and Dean 2002). While *FRI* and *FLC* genes are present in both the progenitors, *A. thaliana* (Landsberg) *FRI* and *A. arenosa FLC* alleles have reduced expression or function (Wang et al. 2006a, b). Therefore, it is possible that the flowering delay might be explained by the novel genic interaction between *A. arenosa FRI* and *A. thaliana FLC* in the neoallopolyploids (Wang et al. 2006a, b).

4.2.3 Chromosome Missegregation Can Lead to Novel Variation in Neopolyploids

Perturbations in chromosome segregation have been proposed to be a major source of novel phenotypic variation in polyploids (Stebbins 1971; Müntzing 1937; Ramsey and Schemske 2002; Song et al. 1993, 1995; Soltis and Soltis 1999). During meiotic prophase, the increased chromosome sets in polyploids will often lead to multivalent chromosome pairing that can result in slower cell division, unequal chromosome partitioning to daughter cells, and even homeologous recombination (Storchova et al. 2006; Stebbins 1971; Ramsey and Schemske 2002). For example, autotetraploids have four homologous chromosomes instead of two, like their diploid counterparts. These four homologous chromosomes may form univalents, trivalents, or quadrivalents during prophase that will cause lagging or unequal chromosome partitioning during anaphase and result in aneuploid gametes (Ramsey and Schemske 2002). The prevalence of multivalent formation in polyploids is widespread; a survey of the available literature by Ramsey and Schemske (2002) found that the mean multivalent frequency for auto- and allopolyploids was estimated to be 28.8 and 8.0 %, respectively. The lower percentage of allopolyploids with multivalent pairing versus autopolyploids, is most

likely the result of sequence divergence between homeologous chromosomes in allopolyploids that favors disomic pairing.

Multivalent pairing is expected to result in a much higher degree of aneuploidy than disomic pairing and ultimately reduce the fertility of polyploids (Ramsey and Schemske 2002; Stebbins 1971). Ramsey and Schemske (2002) compared the occurrence of aneuploidy in gametic and sporophytic cells in different polyploid species. They found that the mean frequency of aneuploid pollen was approximately 40 %, while the mean frequency of aneuploid progeny was an estimated 29 %. Importantly, there were no studies on the frequency of aneuploid ovules, but if one were to assume a similar ratio of aneuploidy as found in pollen, it would imply that 64 % of all zygotes would be aneuploids. These differences in the expected and observed percentages of aneuploid progeny suggest that some aneuploid gametes or progeny do not survive and that chromosome segregation defects correlate with the reduced fertility of polyploid. In contrast to the very different multivalent frequencies observed at meiosis, the frequency of aneuploid gametes and progeny was similar in auto- and allopolyploids. If aneuploidy requires multivalent pairing, it is unclear how allopolyploids could have a similar level of aneuploidy as Autopolyploids. Either there is a much higher level of multivalent pairing in allopolyploids than reported, or other factors contribute to the unexpected level of aneuploidy in allopolyploids (Ramsey and Schemske 2002; Chester et al. 2012).

In addition to aneuploidy, multivalent pairing of homeologous chromosomes can also result in genetic recombination that can severely disrupt genome organization and contribute to phenotypic variation (Stebbins 1971; Ramsey and Schemske 2002; Song et al. 1993, 1995). In allopolyploids, unequal recombination between chromosomes can result from synteny between homeologous chromosomes. Changes in gene dosage, gene expression, the epigenetic landscape, and even gene conversion are all possible outcomes of homeologous recombination in allopolyploids. Homeologous recombination has been observed in many different resynthesized allopolyploid species (Wendel 2000; Osborn et al. 2003; Doyle et al. 2008; Hegarty and Hiscock 2008; Leitch and Leitch 2008; Soltis and Soltis 2009; Stebbins 1971; Grant 1981; Ramsey and Schemske 2002; Chester et al. 2012; Pires et al. 2004; Buggs et al. 2009), suggesting it may be a major mechanism of genome alteration and evolution. In synthetic *Arabidopsis* allopolyploids derived from *A. thaliana* and *A. arenosa* progenitors, the nucleolus organizer region (NOR) and 5S rDNA genetic regions have been shown to be lost or recombined between the chromosomes derived from *A. thaliana* and *A. arenosa* by the F3 generation (Pontes et al. 2004). Rearrangements of the NOR and 5S rDNA regions were also demonstrated in the naturally occurring *Arabidopsis* allopolyploid, *A. suecica*. Genetic rearrangements in resynthesized *Brassica napus* allopolyploids have been linked to phenotypic variation (Gaeta et al. 2007; Pires et al. 2004; Xiong et al. 2011; Song et al. 1995). In one striking example, variation for flowering time between two different *B. napus* allopolyploid lines was found to correlate with a genomic rearrangement disrupting the expression of *FLC* (Pires et al. 2004). These allopolyploid lines also displayed phenotypic variation in flowering time that

exceeded the range of their progenitors (Gaeta et al. 2007; Pires et al. 2004). Together, these results provide powerful evidence for the role of genomic rearrangements in neopolyploids driving the creation of novel phenotypic variation that is not present in their diploid progenitors.

4.2.4 *Epigenetic Remodeling Can Result in Novel Variation in Neopolyploids*

Epigenetic remodeling of polyploid genomes has also been suggested to play a role in the origin of novel Phenotypes (Osborn et al. 2003). Epigenetic changes involving DNA methylation, histone post-translational modifications (PTMs), histone replacement, and sRNA-mediated silencing have all been demonstrated to affect gene expression levels in diploids (Wolffe and Matzke 1999; Calarco and Martienssen 2011). Moreover, the dynamic nature of these epigenetic modifications in combination with the instability of neopolyploid genomes would suggest that there is the potential for rampant epigenetic remodeling in polyploids that could affect gene expression (Osborn et al. 2003; Wolffe and Matzke 1999; Ramsey and Schemske 2002; Matzke et al. 1999).

Concurrently, there have been many studies documenting novel changes in DNA methylation (Li et al. 2010; Chen et al. 2008; Lukens et al. 2006; Xu et al. 2009; Wang et al. 2004, 2009; Liu and Wendel 2003; Madlung et al. 2002; Kenan-Eichler et al. 2011; Parisod et al. 2009; Yaakov et al. 2011) as well as changes in sRNA profiles (Kenan-Eichler et al. 2011; Ha et al. 2009; Preuss et al. 2008) within allopolyploid genomes compared to their progenitors, but there has been limited data demonstrating links of these epigenetic changes to phenotypic consequences. One classic example of epigenetic reprogramming in allopolyploids is the occurrence of nucleolar dominance among interspecies hybrids, including allopolyploids (McStay 2006; Preuss and Pikaard 2007). Nucleolar dominance occurs when rRNA genes from one parent, or progenitor species, are preferentially silenced in a hybrid; this silencing is neither the result of random inactivation, nor correlated with imprinting and sexual parent of origin (McStay 2006; Preuss and Pikaard 2007). In *Arabidopsis* allopolyploids, small interfering RNAs (sRNA), DNA methylation, and histone deacetylation result in the silencing of *A. thaliana*-derived rRNA genes in *A. suecica* (Preuss and Pikaard 2007, 2008; Earley et al. 2006). While no clear phenotype is associated with nucleolar dominance, it is believed to be important for genome stability and prevention of premature aging, which has been linked to perturbations in rDNA genic regions (Finigan and Martienssen 2008).

For histone PTMs, a few studies in *Arabidopsis* neoallopolyploids demonstrate changes in histone modification patterns that correlate with changes in gene expression (Wang et al. 2006a, b; Ni et al. 2009). Interestingly, these changes in gene expression patterns were linked to phenotypic changes in flowering time

(Wang et al. 2006a, b) and chlorophyll and sugar content (Ni et al. 2009). However, these perturbations are most likely the downstream consequences of novel regulatory interactions (Ni et al. 2009; Wang et al. 2006a, b). Additionally, a tentative link between seed death and perturbations in histone H3 lysine 27 methylation (H3K27me) has been identified in *Arabidopsis* neoallopolyploids (Josefsson et al. 2006; Walia et al. 2009). Josefsson et al. (2006) reported that the expression of *PHERES1*, *MEDEA*, and other imprinted genes were deregulated in developing *Arabidopsis* neoallopolyploid seeds. In addition to the imprinted genes, Athila retrotransposons from the pollen parent were also deregulated in *Arabidopsis* allotriploid seed, which are inviable. This deregulation of imprinted genes and Athila elements was suggested to result from a lack of silencing of their pollen parent copies by the *POLYCOMB REPRESSIVE COMPLEX* (*PRC*), which is responsible for H3K27me in *Arabidopsis* plants (Josefsson et al. 2006). Subsequent work suggested that the PRC complex was not functioning in these neoallopolyploids because of a 2–10-fold down-regulation in the expression of *FERTILIZATION-INDEPENDENT SEED 2* (*FIS2*), one of the subunits of the PRC (Walia et al. 2009). Why the PRC complex, and FIS2, are misregulated in the developing neoallopolyploids is still an unanswered question, but major disruptions of at least H3K27 methylation patterns in these neoallopolyploids would be predicted by this model (Josefsson et al. 2006; Walia et al. 2009). Future studies to analyze H3K27 methylation patterns in the neoallopolyploids will be important to validate this hypothesis. This misregulation of the PRC complex is only present in the neoallopolyploids but absent in its autopolyploid progenitors and depends on both the ploidy level and genetic diversity of the progenitors (Josefsson et al. 2006; Walia et al. 2009).

So far, the limited evidence for epigenetic remodeling has been primarily identified in allo-, and not autopolyploids. While allopolyploids inherit epigenetic and genomic divergence from their progenitors, one might still expect to see epigenomic remodeling in autopolyploids due to their dynamic nature. One role for epigenetic mechanisms in autopolyploids is in the origin of unreduced gametes from which they arise. In the ovules of most sexual flowering plants, female gametogenesis is initiated from a single surviving gametic cell, the functional megaspore, formed after meiosis of the somatically derived megaspore mother cell (MMC). The *Arabidopsis* small RNA binding protein ARGONAUTE 9 (AGO9) controls female gamete formation by restricting the specification of gametophyte precursors to the MMC (Olmedo-Monfil et al. 2010; Durán-Figueroa et al. 2010). Mutations in AGO9 lead to the differentiation of diploid (unreduced) gametic cells from the surrounding ovule that are able to initiate gametogenesis. Mutations in the maize AGO9 homolog also lead to unreduced gametes, through a related but distinct mechanism (Singh et al. 2011). AGO9 preferentially interacts with 24-nucleotide sRNAs derived from transposable elements (TEs), and its activity is necessary to silence TEs in female gametes and their accessory cells (Olmedo-Monfil et al. 2010; Durán-Figueroa et al. 2010). That AGO9-dependent sRNA silencing is also crucial to specify ploidy in the gametes indicates that epigenetic reprogramming may link transposon silencing to germ cell fate (Slotkin et al. 2009).

This may reflect the coevolution of transposon activity and sexual reproduction, as well as the increased tolerance of polyploids to mutations caused by TE insertions (Martienssen 2010).

4.2.5 Nonadditive Gene Expression and Novel Phenotypic Variation

Mechanisms encompassing gene dosage/allelic combination, novel gene interactions, genomic alterations, and epigenomic reorganization must all converge on gene expression and function to induce novel phenotypic variation. While little is known about perturbations in gene function in neopolyploids, changes in gene expression (Wang et al. 2006a, b; Akhunova et al. 2010; Pumphrey et al. 2009; Chaudhary et al. 2009; Chagué et al. 2010; Flagel et al. 2008; Flagel and Wendel 2010; Hegarty et al. 2006) and protein levels (Albertin et al. 2006; Ng et al. 2011) have been demonstrated in allopolyploids compared to their progenitors.

Transcript profiling experiments in synthetic *Arabidopsis* allopolyploids were the first demonstrations of genome-wide nonadditive gene expression in allopolyploids (Wang et al. 2006a, b). Nonadditive gene expression occurs when the expression level of genes in hybrids differs from the average expression level of the orthologous genes in the parental species (i.e. midparent level). In *Arabidopsis* neoallopolyploids, an estimated 5–35 % of all genes were nonadditively expressed—depending on the statistical method employed (Wang et al. 2006a, b). Subsequent research in other neoallopolyploid species has suggested that nonadditive gene expression is a general trend found associated with recent allopolyploidization; however, the amount of nonadditive gene expression in specific allopolyploids is highly variable (Hegarty et al. 2006; Chaudhary et al. 2009; Flagel et al. 2008; Flagel and Wendel 2010; Akhunova et al. 2010; Pumphrey et al. 2009; Chagué et al. 2010). For example, in bread wheat allohexaploids nonadditive gene expression has been reported as anywhere from 7 to 40 % (Pumphrey et al. 2009; Chagué et al. 2010; Kashkush et al. 2002; Akhunova et al. 2010).

The discovery of nonadditive gene expression in neoallopolyploids implies that the progenitor-derived chromosome sets must be interacting to result in novel gene expression patterns. How these chromosome sets interact is just beginning to be unraveled; but, by definition, these novel interactions must stem directly from changes in *cis*- and *trans*-regulatory divergence between the progenitor-derived genomes. *Cis*- and/or *trans*-regulatory differences could result in changes in the expression of homeologous genes (homeoalleles) from each progenitor-derived chromosome set. *Cis*-regulatory divergence can act directly on single genes or on localized chromatin domains, such as promoters or enhancers, resulting in asymmetric accumulation of homeologous transcripts in allopolyploids. *Trans*-regulatory divergence between the parental genomes will affect homeologous genes equally and result in equal accumulation of homeologous transcripts.

Thus, both *cis*- and *trans*-regulatory differences could result in nonadditive homeologous gene expression, but only *cis*-regulatory differences could lead to biased homeoallele-specific gene expression. The relative contribution of *cis*- and *trans*-regulatory divergence to nonadditive gene expression is still unclear, but some evidence has arisen recently implicating a major role of *cis*-regulatory differences for this phenomenon.

Transcriptome profiling studies in *Gossypium* (Flagel and Wendel 2010; Doyle et al. 2008), *Arabidopsis* (Wang et al. 2006a, b), and maize (Schnable et al. 2011), and small-scale analyses in *Tragopogon* (Buggs et al. 2011; Tate et al. 2006) polyploids have suggested that there is a common theme of "genome dominance" in homeologous gene expression in allopolyploids (Rapp et al. 2009). In the genome dominance model, one of the progenitor transcription profiles outcompetes the other to shift the overall expression in the allopolyploids toward a progenitor-specific expression profile. In *Gossypium* allopolyploids, five different allopolyploids have been suggested to have a 54–60 % transcriptional bias in favor of the D genome versus the A genome (Flagel and Wendel 2010). A similar phenomenon has been reported in *Arabidopsis* allopolyploids, where a 55 % bias toward *A. arenosa* over *A. thaliana* homeoallele expression has been reported for natural allopolyploids (Chang et al. 2010) and suggested for synthetic allopolyploids (Wang et al. 2006a, b). In contrast, the evidence for genome dominance has been inconclusive in *Triticum* (Akhunova et al. 2010; Pumphrey et al. 2009; Chagué et al. 2010). However, this phenomenon is still un-resolved, because the majority of the evidence for genome dominance are based on comparing the expression profiles between the progenitors and their allopolyploid offspring, without distinguishing between homeoalleles in the allopolyploids (Flagel et al. 2008; Flagel and Wendel 2010; Wang et al. 2006a, b). Further, in cases where genome-wide profiles of homeologous gene expression have suggested biased expression of one genome over the other (Akhunova et al. 2010; Chang et al. 2010; Schnable et al. 2011), the exact progenitors of these established allopolyploids were extinct and true comparisons could not be made. Future studies should address these discrepancies and help to unravel the phenotypic consequences of nonadditive gene expression.

4.3 Technical Considerations When Comparing Gene Expression Studies in Polyploids

Many of the controversies surrounding genome dominance and the variability of nonadditive gene expression in neoallopolyploids may stem from the differences in the methodologies employed and the organisms studied. One important distinction is whether polyploidization preceded hybridization or vice versa (Hegarty and Hiscock 2008). If polyploidization precedes hybridization then the resulting neoallopolyploid will have duplicate chromosomes that can accurately pair at meiosis

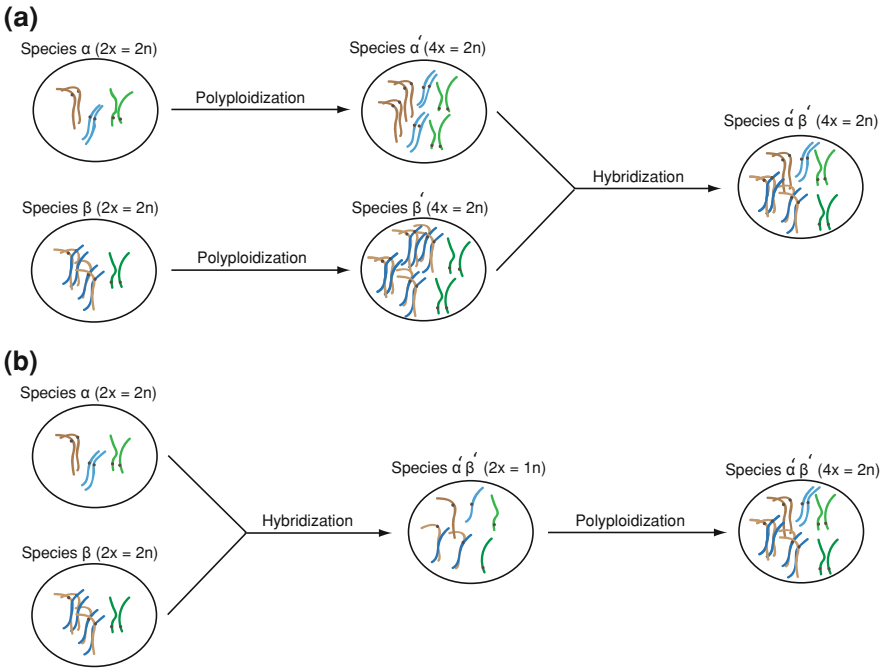

Fig. 4.2 The origin of allopolyploids. Two types of allopolyploids are distinguished by the relative order of polyploidization and hybridization. **a** Type I allopolyploids arise when polyploidization precedes hybridization, typically via unreduced gametes. Type I neoallopolyploids will have duplicate chromosomes that properly pair in meiosis. **b** Type II allopolyploids are created when hybridization precedes polyploidization. Type II allopolyploids transit through a hybrid state, which is usually sterile, as chromosomes do not pair during meiosis. A subsequent polyploidization event restores 2*n* chromosome number in allopolyploids

(Fig. 4.2). Synthetic *Arabidopsis* allopolyploids are an example, and natural *Arabidopsis* allopolyploids are believed to have arisen this way (Hegarty and Hiscock 2008; Josefsson et al. 2006). On the other hand, if polyploidization follows hybridization, then the resulting hybrid will first transit through an unbalanced chromosome state, most likely severely infertile, before restoration with a polyploidization event to create a neoallopolyploid (Fig. 4.2) (Hegarty et al. 2006; Hegarty and Hiscock 2008). *Gossypium*, *Triticum*, *Tragopogon*, *Spartina*, and *Senecio* allopolyploids are examples of the latter type (Hegarty et al. 2006). Transcriptome profiling studies in *Senecio* (Hegarty et al. 2006) and *Triticum* (Feldman and Levy 2005) hybrids before and after polyploidization suggested that the homologous gene expression levels were more perturbed (vs. the midparent) in the sterile hybrids before chromosome doubling. These results may suggest that gene expression profiles in the resulting allopolyploid may be influenced by the creation of a hybrid intermediate.

Another important distinction is the nature of the progenitor species used to compare with allopolyploid expression profiles. Many allopolyploids do not have

extant progenitors, and relatives must be used instead (Chang et al. 2010; Rapp et al. 2009; Flagel and Wendel 2010). These relatives may not be exact genetic matches (Chang et al. 2010; Rapp et al. 2009; Flagel and Wendel 2010) and may contain epigenetic differences as well, as is the case for different accessions of *A. thaliana* (Vaughn et al. 2007; Kliebenstein et al. 2006). A recent study illustrates this point. An estimated 6,790 homologous genes were differentially expressed between *A. suecica* and synthetic *Arabidopsis* allopolyploids (Chang et al. 2010), and comparative genome sequencing revealed that 938 homeologous genes were missing from *A. suecica* compared to the synthetic allopolyploid (Chang et al. 2010). These genomic differences are comparable to those between different *Arabidopsis thaliana* accessions (Vaughn et al. 2007; Kliebenstein et al. 2006) and may represent differences between the *A. suecica* progenitors and their extant relatives.

Additionally, some studies employ progenitor species that have been severely genetically manipulated (Akhunova et al. 2010; Chang et al. 2010; Kerber 1964; Wang et al. 2006a, b). Previous analysis of fifth-generation *Arabidopsis* allopolyploids have relied on a tetraploidized *A. thaliana* (Landsberg accession) line produced spontaneously from root explant regenerants that underwent a callus phase and cell culture treatment (Chang et al. 2010; Valvekens et al. 1988; Wang et al. 2006a, b; Comai et al. 2000). Plant lines regenerated from callus/cell cultures typically display heritable changes in genetic and epigenetic regulation—namely, somaclonal variation (Phillips et al. 1994; Meins and Thomas 2003; Mohan Jain 2001; Tanurdzic et al. 2008). These changes could have a profound effect on gene expression and epigenetic regulation in neoallopolyploids derived from such progenitors. Classical approaches based on colchicine treatment have also been applied to the generation of tetraploid *A. thaliana* lines (Henry et al. 2005; Santos et al. 2003). Studies in *Gossypium* (Rapp et al. 2009; Chaudhary et al. 2009), *Triticum* (Akhunova et al. 2010; Pumphrey et al. 2009; Chagué et al. 2010), and *Tragopogon* (Tate et al. 2009) have also relied on colchicine treatments to polyploidize sterile triploid hybrids after fertilization. Importantly, there was no difference in genetic or DNA methylation changes between spontaneous or colchicine produced *Brassica* polyploids (Gaeta et al. 2007).

4.4 Conclusions

The prevalence of polyploidy in nature suggests it has been positively selected for during evolution. Indeed, the higher yields of polyploids among agricultural crops and their usefulness to bridge species barriers have led to their widespread use and creation for the agricultural industry (Eigsti 1957). Understanding the emergent properties of neopolyploids, such as novel phenotypic variation and reproductive barriers, is vitally important to further advance the usefulness of polyploids for crop breeding. While we are beginning to understand the molecular mechanisms that contribute to novel variation in polyploids, there is still a lack of specific links

between genes, proteins, and phenotypes. We also do not understand the interplay between these different mechanisms and the important drivers of phenotypic diversity. What role do genomic alterations or allelic diversity play in the origin of this novel phenotypic variation? Many homeologous genes are nonadditively expressed in neoallopolyploids, but what is the source of this differential expression? If nonadditive gene expression is a result of genome dominance, than how is such a mechanism established? The answers to these questions are likely to emerge in the coming years.

References

Ahloowalia B, Garber F (1961) The genus *Collinsia*. XIII. Cytogenetic studies of interspecific hybrids involving species with pediceled flowers. Bot Gaz 122:219

Akhunova A, Matniyazov R, Liang H, Akhunov E (2010) Homoeolog-specific transcriptional bias in allopolyploid wheat. BMC Genomics 11(1):505

Albertin W, Balliau T, Brabant P, Chevre A-M, Eber F, Malosse C, Thiellement H (2006) Numerous and rapid Nonstochastic modifications of Gene products in newly synthesized *Brassica napus Allotetraploids*. Genetics 173(2):1101–1113. doi: 10.1534/genetics.106.057554

Bertrand D, Gagnon Y, Blanchette M, El-Mabrouk N (2010) Reconstruction of ancestral genome subject to whole genome duplication, speciation, rearrangement and loss. Paper presented at the Proceedings of the 10th international conference on algorithms in bioinformatics, Liverpool, UK

Bingham E (1979) Maximizing heterozygosity in autopolyploids. Basic Life Sci 13:471–489

Birchler JA, Bhadra U, Bhadra MP, Auger DL (2001) Dosage-dependent Gene regulation in multicellular Eukaryotes: implications for dosage compensation, Aneuploid syndromes, and quantitative traits. Dev Biology 234(2):275–288. doi: 10.1006/dbio.2001.0262

Bretagnolle F, Thompson J (1995) Tansley review no. 78. Gametes with the somatic chromosome number: mechanisms of their formation and role in the evolution of Autopolyploid plants. New Phytol 129:1

Buggs RJA, Doust AN, Tate JA, Koh J, Soltis K, Feltus FA, Paterson AH, Soltis PS, Soltis DE (2009) Gene loss and silencing in *Tragopogon miscellus* (*Asteraceae*): comparison of natural and synthetic *Allotetraploids*. Heredity 103 (1):73–81. doi: http://www.nature.com/hdy/journal/v103/n1/suppinfo/hdy200924s1.html

Buggs Richard JA, Zhang L, Miles N, Tate Jennifer A, Gao L, Wei W, Schnable Patrick S, Barbazuk WB, Soltis Pamela S, Soltis Douglas E (2011) Transcriptomic shock generates evolutionary novelty in a newly formed. Natural allopolyploid plant. Curr Biol 21(7):551–556. doi: 10.1016/j.cub.2011.02.016

Calarco JP, Martienssen RA (2011) Genome reprogramming and small interfering RNA in the *Arabidopsis germline*. Curr Opin Genet Dev 21(2):134–139. doi: 10.1016/j.gde.2011.01.014

Chagué V, Just J, Mestiri I, Balzergue S, Tanguy A-M, Huneau C, Huteau V, Belcram H, Coriton O, Jahier J, Chalhoub B (2010) Genome-wide gene expression changes in genetically stable synthetic and natural wheat allohexaploids. New Phytol 187(4):1181–1194. doi: 10.1111/j.1469-8137.2010.03339.x

Chang P, Dilkes B, McMahon M, Comai L, Nuzhdin S (2010) Homoeolog-specific retention and use in *allotetraploid Arabidopsis suecica* depends on parent of origin and network partners. Genome Biol 11(12):R125

Chaudhary B, Flagel L, Stupar R, Udall J, Verma N, Springer N, Wendel J (2009) Reciprocal silencing, transcriptional bias and functional divergence of homeologs in polyploid cotton (*Gossypium*). Genetics 182:503–517

Chen M, Ha M, Lackey E, Wang J, Chen Z (2008) RNAi of met1 reduces DNA methylation and induces genome-specific changes in gene expression and centromeric small RNA accumulation in *Arabidopsis* Allopolyploids. Genetics 178:1845–1858

Chester M, Gallagher JP, Symonds VV, Cruz da Silva AV, Mavrodiev EV, Leitch AR, Soltis PS, Soltis DE (2012) Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon* miscellus (*Asteraceae*). Proc Nat Acad Sci 109(4):1176–1181. doi: 10.1073/pnas.1112041109

Church SA, Spaulding EJ (2009) Gene expression in a wild Autopolyploid sunflower series. J Hered 100(4):491–495. doi: 10.1093/jhered/esp008

Clausen J, Keck D, Hiesey W (1945) Experimental studies on the nature of species. II. Plant evolution through *Amphiploidy* and autoploidy, with examples from the *Madiinae*. Carnegie Inst Wash Publ 564

Comai L, Tyagi AP, Winter K, Holmes-Davis R, Reynolds SH, Stevens Y, Byers B (2000) Phenotypic instability and rapid Gene silencing in newly formed *Arabidopsis Allotetraploids*. Plant Cell 12(9):1551–1568. doi: 10.1105/tpc.12.9.1551

Darlington CD (1963) Chromosome botany, and the origins of cultivated plants. Allen and Unwin, London

Dehal P, Boore JL (2005) Two rounds of whole genome duplication in the *Ancestral Vertebrate*. PLoS Biol 3(10):e314

Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF (2008) Evolutionary genetics of genome merger and doubling in plants. Annu Rev Genet 42(1): 443–461. doi: 10.1146/annurev.genet.42.110807.091524

Dubcovsky J, Dvorak J (2007) Genome plasticity a key factor in the success of polyploid wheat under domestication. Science 316(5833):1862–1866. doi: 10.1126/science.1143986

Durán-Figueroa N, Vielle-Calzada J-P (2010) *ARGONAUTE9*-dependent silencing of transposable elements in pericentromeric regions of *Arabidopsis*. Plant Signal Behav 5(11):1476–1479

Earley K, Lawrence RJ, Pontes O, Reuther R, Enciso AJ, Silva M, Neves N, Gross M, Viegas W, Pikaard CS (2006) Erasure of histone acetylation by *Arabidopsis* HDA6 mediates large-scale gene silencing in nucleolar dominance. Genes Dev 20(10):1283–1293. doi: 10.1101/gad.1417706

Eigsti OJ (1957) Induced Polyploidy. Am J Bot 44(3):272–279

Feldman M, Levy AA (2005) Allopolyploidy—a shaping force in the evolution of wheat genomes. Cytogenet Genome Res 109(1–3):250–258

Finigan P, Martienssen RA (2008) Nucleolar dominance and DNA methylation directed by small interfering RNA. Mol Cell 32(6):753–754

Flagel L, Udall J, Nettleton D, Wendel J (2008) Duplicate gene expression in allopolyploid *Gossypium* reveals two temporally distinct phases of expression evolution. BMC Biol 6:16

Flagel LE, Wendel JF (2010) Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during *allotetraploid* cotton speciation. New Phytol 186(1):184–193. doi: 10.1111/j.1469-8137.2009.03107.x

Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC (2007) Genomic changes in Resynthesized *Brassica napus* and their effect on Gene expression and Phenotype. Plant Cell Online 19(11):3403–3417. doi: 10.1105/tpc.107.054346

Gaut BS, Doebley JF (1997) DNA sequence evidence for the segmental *allotetraploid* origin of maize. Proc Nat Acad Sci U S A 94(13):6809–6814

Grant V (1981) Plant speciation, 2nd edn. Columbia University Press, New York

Guo M, Birchler JA (1994) Trans-acting dosage effects on the expression of model Gene systems in Maize Aneuploids. Science 266(5193):1999–2002. doi: 10.1126/science.266.5193.1999

Guo M, Davis D, Birchler JA (1996) Dosage effects on Gene expression in a Maize Ploidy series. Genetics 142(4):1349–1355

Ha M, Lu J, Tian L, Ramachandran V, Kasschau KD, Chapman EJ, Carrington JC, Chen X, Wang X-J, Chen ZJ (2009) Small RNAs serve as a genetic buffer against genomic shock in *Arabidopsis* interspecific hybrids and allopolyploids. Proc Nat Acad Sci 106(42): 17835–17840. doi: 10.1073/pnas.0907003106

Harlan J, deWet J (1975) On Ö. Winge and a prayer: the origins of polyploidy. Bot Rev 41 (4):361–390. doi: 10.1007/bf02860830

Hegarty MJ, Barker GL, Wilson ID, Abbott RJ, Edwards KJ, Hiscock SJ (2006) Transcriptome shock after interspecific hybridization in senecio is ameliorated by genome duplication. Curr Biol 16(16):1652–1659. doi: 10.1016/j.cub.2006.06.071

Hegarty MJ, Hiscock SJ (2008) Genomic clues to the evolutionary success of polyploid plants. Curr Biol: CB 18(10):R435–R444

Henry IM, Dilkes BP, Young K, Watson B, Wu H, Comai L (2005) Aneuploidy and genetic variation in the *Arabidopsis thaliana* Triploid response. Genetics 170(4):1979–1988. doi: 10.1534/genetics.104.037788

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, Soltis DE, Clifton SW, Schlarbaum SE, Schuster SC, Ma H, Leebens-Mack J, dePamphilis CW (2011) Ancestral polyploidy in seed plants and angiosperms. Nature 473 (7345):97–100. doi: http://www.nature.com/nature/journal/v473/n7345/abs/10.1038-nature09916-unlocked.html#supplementary-information

Jørgensen C (1928) The experimental formation of heteroploid plants in the genus *Solanum*. J Genet 11:133

Josefsson C, Dilkes B, Comai L (2006) Parent-dependent loss of Gene silencing during interspecies hybridization. Curr Biol 16(13):1322–1328. doi: 10.1016/j.cub.2006.05.045

Kashkush K, Feldman M, Levy A (2002) Gene loss, silencing and activation in a newly synthesized wheat *Allotetraploid*. Genetics 160:1651–1659

Kellis M, Birren BW, Lander ES (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces Cerevisiae*. Nature 428 (6983):617–624. doi: http://www.nature.com/nature/journal/v428/n6983/suppinfo/nature02424_S1.html

Kenan-Eichler M, Leshkowitz D, Tal L, Noor E, Melamed-Bessudo C, Feldman M, Levy AA (2011) Wheat hybridization and Polyploidization results in deregulation of small RNAs. Genetics 188(2):263–272. doi: 10.1534/genetics.111.128348

Kerber E (1964) Wheat: reconstitution of the tetraploid component (AABB) of hexaploids. Science 143:53–255

Kihara H, Ono T (1926) Chromosomenzahlen und systematische Gruppierung der Rumex-Arten. Z Zellforsch Mikr Anat 4:475

Kliebenstein DJ, West MAL, van Leeuwen H, Kim K, Doerge RW, Michelmore RW, St. Clair DA (2006) Genomic survey of Gene expression diversity in *Arabidopsis thaliana*. Genetics 172(2):1179–1189. doi: 101534/genetics.105.049353

Lagercrantz U, Lydiate DJ (1996) Comparative genome mapping in *Brassica*. Genetics 144(4):1903–1910

Leitch AR, Leitch IJ (2008) Genomic plasticity and the diversity of polyploid plants. Science 320(5875):481–483. doi: 10.1126/science.1153585

Levin DA (1983) Polyploidy and novelty in flowering plants. Am Nat 122(1):1–25

Levin DA (2002) The role of chromosomal change in plant evolution. Oxford University Press, Oxford

Lexer C, Welch ME, Raymond O, Rieseberg LH (2003) The origin of ecological divergence in *Helianthus paradoxus* (*Asteraceae*): selection on transgressive characters in a novel hybrid habitat. Evolution 57(9):1989–2000. doi: 10.1111/j.0014-3820.2003.tb00379.x

Li X, Guo W, Wang B, Li X, Chen H, Wei L, Wang Y, Wu J, Long H (2010) Instability of chromosome number and DNA methylation variation induced by hybridization and amphidiploid formation between Raphanus sativus L. and *Brassica Alboglabra Bailey*. BMC Plant Biol 10(1):207

Liu B, Wendel JF (2003) Epigenetic phenomena and the evolution of plant allopolyploids. Mol Phylogenet Evol 29(3):365–379. doi: 10.1016/s1055-7903(03)00213-6

Lukens LN, Pires JC, Leon E, Vogelzang R, Oslach L, Osborn T (2006) Patterns of sequence loss and cytosine methylation within a population of newly Resynthesized *Brassica napus* allopolyploids. Plant Physiol 140(1):336–348. doi: 10.1104/pp.105.066308

Lumaret R (1988) Adaptive strategies and ploidy levels. Acta Oecol Oecol Plant 9:83

Madlung A, Masuelli RW, Watson B, Reynolds SH, Davison J, Comai L (2002) Remodeling of DNA Methylation and Phenotypic and transcriptional changes in synthetic *Arabidopsis Allotetraploids*. Plant Physiol 129(2):733–746. doi: 10.1104/pp.003095

Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y (2005) Modeling gene and genome duplications in Eukaryotes. Proc Nat Acad Sci U S A 102(15):5454–5459. doi: 10.1073/pnas.0501102102

Martienssen RA (2010) Heterochromatin, small RNA and post-fertilization dysgenesis in allopolyploid and interploid hybrids of *Arabidopsis*. New Phytol 186(1):46–53. doi: 10.1111/j.1469-8137.2010.03193.x

Matzke MA, Scheid OM, Matzke AJM (1999) Rapid structural and epigenetic changes in polyploid and aneuploid genomes. BioEssays 21(9):761–767. doi: 10.1002/(sici)1521-1878(199909)21:9<761:aid-bies7>3.0.co;2-c

McStay B (2006) Nucleolar dominance: a model for rRNA gene silencing. Genes Dev 20(10):1207–1214. doi: 10.1101/gad.1436906

Meins F, Thomas M (2003) Meiotic transmission of epigenetic changes in the cell-division factor requirement of plant cells. Development 130(25):6201–6208. doi: 10.1242/dev.00856

Meyer A, Schartl M (1999) Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. Curr Opin Cell Biol 11(6):699–704. doi: 10.1016/s0955-0674(99)00039-3

Meyer A, Van de Peer Y (2005) From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). BioEssays 27(9):937–945. doi: 10.1002/bies.20293

Mohan Jain S (2001) Tissue culture-derived variation in crop improvement. Euphytica 118(2):153–166. doi: 10.1023/a:1004124519479

Müntzing A (1936) The evolutionary significance of Autopolyploidy. Hereditas 21(2–3):363–378. doi: 10.1111/j.1601-5223.1936.tb03204.x

Müntzing A (1937) The effects of chromosomal variation in *Dactylis*. Hereditas 23:113

Newton W, Pellew C (1929) Primula kewensis and its derivatives. J Genet 20:405

Ng DWK, Zhang C, Miller M, Shen Z, Briggs SP, Chen ZJ (2011) Proteomic divergence in *Arabidopsis* autopolyploids and allopolyploids and their progenitors. Heredity. doi: http://www.nature.com/hdy/journal/vaop/ncurrent/suppinfo/hdy201192s1.html

Ni Z, Kim E-D, Ha M, Lackey E, Liu J, Zhang Y, Sun Q, Chen ZJ (2009) Altered circadian rhythms regulate growth vigour in hybrids and Allopolyploids. Nature 457 (7227):327–331. doi: http://www.nature.com/nature/journal/v457/n7227/suppinfo/nature07523_S1.html

Noggle G (1946) The physiology of polyploidy in plants. Lloydia 9:153

Ohno S (1970) Evolutoin by gene duplication. Springer, Berlin

Olmedo-Monfil V, Duran-Figueroa N, Arteaga-Vazquez M, Demesa-Arevalo E, Autran D, Grimanelli D, Slotkin RK, Martienssen RA, Vielle-Calzada J-P (2010) Control of female gamete formation by a small RNA pathway in *Arabidopsis*. Nature 464 (7288):628–632. doi: http://www.nature.com/nature/journal/v464/n7288/suppinfo/nature08828_S1.html

Osborn T, Pires J, Birchler J, Auger D, Chen Z, Lee H, Comai L, Madlung A, Doerge R, Colot V, Martienssen R (2003) Understanding mechanisms of novel gene expression in polyploids. Trends Genet 19:141–147

Parisod C, Salmon A, Zerjal T, Tenaillon M, Grandbastien M-A, Ainouche M (2009) Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in *Spartina*. New Phytol 184(4):1003–1015. doi: 10.1111/j.1469-8137.2009.03029.x

Phillips RL, Kaeppler SM, Olhoft P (1994) Genetic instability of plant tissue cultures: breakdown of normal controls. Proc Nat Acad Sci U S A 91(12):5222–5226

Pignatta D, Dilkes BP, Yoo S-Y, Henry IM, Madlung A, Doerge RW, Jeffrey Chen Z, Comai L (2010) Differential sensitivity of the Arabidopsis *thaliana* transcriptome and enhancers to the effects of genome doubling. New Phytol 186(1):194–206. doi: 10.1111/j.1469-8137.2010.03198.x

Pires JC, Zhao J, Schranz ME, Leon EJ, Quijada PA, Lukens LN, Osborn TC (2004) Flowering time divergence and genomic rearrangements in resynthesized *Brassica* polyploids (*Brassicaceae*). Biol J Linn Soc 82(4):675–688. doi: 10.1111/j.1095-8312.2004.00350.x

Pontes O, Neves N, Silva M, Lewis MS, Madlung A, Comai L, Viegas W, Pikaard CS (2004) Chromosomal locus rearrangements are a rapid response to formation of the *allotetraploid Arabidopsis suecica* genome. Proc Nat Acad Sci U S A 101(52):18240–18245. doi: 10.1073/pnas.0407258102

Preuss S, Pikaard CS (2007) rRNA gene silencing and nucleolar dominance: insights into a chromosome-scale epigenetic on/off switch. Biochimica et Biophysica Acta (BBA)—Gene structure and expression 1769 (5–6):383–392. doi: 10.1016/j.bbaexp.2007.02.005

Preuss SB, Costa-Nunes P, Tucker S, Pontes O, Lawrence RJ, Mosher R, Kasschau KD, Carrington JC, Baulcombe DC, Viegas W, Pikaard CS (2008) Multimegabase silencing in nucleolar dominance involves siRNA-directed DNA methylation and specific methylcytosine-binding proteins. Mol Cell 32(5):673–684. doi: 10.1016/j.molcel.2008.11.009

Pumphrey M, Bai J, Laudencia-Chingcuanco D, Anderson O, Gill B (2009) Nonadditive expression of homoeologous genes is established upon polyploidization in hexaploid wheat. Genetics 181:1147–1157

Ramsey J, Schemske DW (1998) Pathways, mechanisms, and rates of polyploid formation in flowering plants. Annu Rev Ecol Syst 29(1):467–501. doi: 10.1146/annurev.ecolsys.29.1.467

Ramsey J, Schemske DW (2002) Neopolyploidy in flowering plants. Annu Rev Ecol Syst 33(1):589–639. doi: 10.1146/annurev.ecolsys.33.010802.150437

Randolph L (1941) An evaluation of induced polyploidy as a method of breeding crop plants. Am Nat 75:347

Rapp R, Udall J, Wendel J (2009) Genomic expression dominance in allopolyploids. BMC Biol 7:18

Riddle N, Jiang H, An L, Doerge R, Birchler J (2010) Gene expression analysis at the intersection of ploidy and hybridity in maize. Theor Appl Genet 120(2):341–353. doi: 10.1007/s00122-009-1113-3

Santos JL, Alfaro D, Sanchez-Moran E, Armstrong SJ, Franklin FCH, Jones GH (2003) Partial diploidization of meiosis in Autotetraploid *Arabidopsis thaliana*. Genetics 165(3):1533–1540

Schnable JC, Springer NM, Freeling M (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proc Nat Acad Sci 108(10):4069–4074. doi: 10.1073/pnas.1101368108

Shindo C, Aranzana MJ, Lister C, Baxter C, Nicholls C, Nordborg M, Dean C (2005) Role of FRIGIDA and flowering locus c in determining variation in flowering time of *Arabidopsis*. Plant Physiol 138(2):1163–1173. doi: 10.1104/pp.105.061309

Shoemaker RC, Polzin K, Labate J, Specht J, Brummer EC, Olson T, Young N, Concibido V, Wilcox J, Tamulonis JP, Kochert G, Boerma HR (1996) Genome duplication in soybean (*Glycine* subgenus soja). Genetics 144(1):329–338

Simpson GG, Dean C (2002) *Arabidopsis*, the rosetta stone of flowering time? Science 296(5566):285–289. doi: 10.1126/science.296.5566.285

Singh M, Goel S, Meeley RB, Dantec C, Parrinello H, Michaud C, Leblanc O, Grimanelli D (2011) Production of viable gametes without Meiosis in Maize deficient for an ARGONAUTE protein. Plant Cell Online 23(2):443–458. doi: 10.1105/tpc.110.079020

Skalinska M (1946) Polyploidy in valeriana officinalis Linn. In relation to its ecology and distribution. J Linn Soc London, Bot 53:159

Slotkin RK, Vaughn M, Borges F, Tanurdzic M, Becker JD, Feijo JA, Martienssen RA (2009) Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell 136 (3):461–472. doi: S0092-8674(08)01644-9 [pii] 10.1016/j.cell.2008.12.038

Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, dePamphilis CW, Wall PK, Soltis PS (2009) Polyploidy and angiosperm diversification. Am J Bot 96(1):336–348. doi: 10.3732/ajb.0800079

Soltis DE, Rieseberg LH (1986) Autopolyploidy in *Tolmiea menziesii* (*Saxifragaceae*): genetic insights from enzyme electrophoresis. Am J Bot 73(2):310–318

Soltis DE, Soltis PS (1999) Polyploidy: recurrent formation and genome evolution. Trends Ecol Evol 14(9):348–352. doi: 10.1016/s0169-5347(99)01638-9

Soltis DE, Soltis PS, Schemske DW, Hancock JF, Thompson JN, Husband BC, Judd WS (2007) Autopolyploidy in angiosperms: have we grossly underestimated the number of species? Taxon 56(1):13–30

Soltis PS, Soltis DE (2009) The role of hybridization in plant speciation. Annu Rev Plant Biol 60(1):561–588. doi: 10.1146/annurev.arplant.043008.092039

Song K, Lu P, Tang K, Osborn TC (1995) Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. Proc Nat Acad Sci 92(17):7719–7723

Song K, Tang K, Osborn T (1993) Development of synthetic *Brassica* amphidiploids by reciprocal hybridization and comparison to natural amphidiploids. Theor Appl Genet 86:811

Stebbins G (1947) Types of polyploids: their classification and significance. Adv Genet 1:403

Stebbins GL (1950) Variation and evolution in plants. Columbia University Press, New York

Stebbins GL (1971) Chromosomal evolution in higher plants [by] G. Ledyard Stebbins. Contemporary biology, vol Accessed from http://nla.gov.au/nla.cat-vn1859678. Edward Arnold, London

Storchova Z, Breneman A, Cande J, Dunn J, Burbank K, O'Toole E, Pellman D (2006) Genome-wide genetic analysis of polyploidy in yeast. Nature 443 (7111):541–547. doi: http://www.nature.com/nature/journal/v443/n7111/suppinfo/nature05178_S1.html

Stupar RM, Bhaskar PB, Yandell BS, Rensink WA, Hart AL, Ouyang S, Veilleux RE, Busse JS, Erhardt RJ, Buell CR, Jiang J (2007) Phenotypic and Transcriptomic changes associated with Potato Autopolyploidization. Genetics 176(4):2055–2067. doi: 10.1534/genetics.107.074286

Tanurdzic M, Vaughn MW, Jiang H, Lee T-J, Slotkin RK, Sosinski B, Thompson WF, Doerge RW, Martienssen RA (2008) Epigenomic consequences of immortalized plant cell suspension culture. PLoS Biol 6(12):e302. doi: 10.1371/journal.pbio.0060302

Tate JA, Symonds VV, Doust AN, Buggs RJA, Mavrodiev E, Koh J, Soltis PS, Soltis DE (2009) Synthetic polyploids of *Tragopogon miscellus* and *T. mirus* (*Asteraceae*): 50 + years after Ownbey's discovery. Am J Bot 96:979–988

Tate JA, Ni Z, Scheen A-C, Koh J, Gilbert CA, Lefkowitz D, Chen ZJ, Soltis PS, Soltis DE (2006) Evolution and expression of Homeologous Loci in Tragopogon miscellus (*Asteraceae*), a recent and reciprocally formed allopolyploid. Genetics 173(3):1599–1611. doi: 10.1534/genetics.106.057646

Valvekens D, Montagu MV, Lijsebettens MV (1988) *Agrobacterium tumefaciens*-mediated transformation of *Arabidopsis thaliana* root explants by using kanamycin selection. Proc Nat Acad Sci U S A 85(15):5536–5540

Van de Peer Y, Maere S, Meyer A (2009) The evolutionary significance of ancient genome duplications. Nat Rev Genet 10(10):725–732

Vaughn MW, Tanurdzic M, Lippman Z, Jiang H, Carrasquillo R, Rabinowicz PD, Dedhia N, McCombie WR, Agier N, Bulski A, Colot V, Doerge RW, Martienssen RA (2007) Epigenetic natural variation in Arabidopsis thaliana. PLoS Biol 5 (7):e174. doi: 06-PLBI-RA-2115 [pii] 10.1371/journal.pbio.0050174

Walia H, Josefsson C, Dilkes B, Kirkbride R, Harada J, Comai L (2009) Dosage-dependent deregulation of an AGAMOUS-LIKE gene cluster contributes to interspecific incompatibility. Curr Biol: CB 19(13):1128–1132

Wang H, Chai Y, Chu X, Zhao Y, Wu Y, Zhao J, Ngezahayo F, Xu C, Liu B (2009) Molecular characterization of a rice mutator-phenotype derived from an incompatible cross-pollination reveals transgenerational mobilization of multiple transposable elements and extensive epigenetic instability. BMC Plant Biol 9(1):63

Wang J, Tian L, Lee H-S, Chen ZJ (2006a) Nonadditive regulation of FRI and FLC Loci mediates flowering-time variation in *Arabidopsis Allopolyploids*. Genetics 173(2):965–974. doi: 10.1534/genetics.106.056580

Wang J, Tian L, Lee H-S, Wei NE, Jiang H, Watson B, Madlung A, Osborn TC, Doerge RW, Comai L, Chen ZJ (2006b) Genomewide nonadditive Gene regulation in Arabidopsis *Allotetraploids*. Genetics 172(1):507–517. doi: 10.1534/genetics.105.047894

Wang J, Tian L, Madlung A, Lee H-S, Chen M, Lee JJ, Watson B, Kagochi T, Comai L, Chen ZJ (2004) Stochastic and Epigenetic changes of Gene expression in *Arabidopsis* Polyploids. Genetics 167(4):1961–1973. doi: 10.1534/genetics.104.027896

Wendel JF (2000) Genome evolution in polyploids. Plant Mol Biol 42(1):225–249. doi: 10.1023/a:1006392424384

Winge O (1917) The chromosomes: their number and general importance. C.R. Trav Lab, Carlsberg

Winge Ö (1932) On the origin of constant species-hybrids. Sven Bot Tidskr 26:107

Wolffe AP, Matzke MA (1999) Epigenetics: regulation through repression. Science 286(5439):481–486. doi: 10.1126/science.286.5439.481

Xiong Z, Gaeta RT, Pires JC (2011) Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. Proc Nat Acad Sci 108(19):7908–7913. doi: 10.1073/pnas.1014138108

Xu Y, Zhong L, Wu X, Fang X, Wang J (2009) Rapid alterations of gene expression and cytosine methylation in newly synthesized *Brassica napus* allopolyploids. Planta 229(3):471–483. doi: 10.1007/s00425-008-0844-8

Yaakov B, Kashkush K (2011) Massive alterations of the methylation patterns around DNA transposons in the first four generations of a newly formed wheat allohexaploid. Genome 54(1):42–49. doi: 10.1139/g10-091

Yao H, Kato A, Mooney B, Birchler J (2011) Phenotypic and gene expression analyses of a ploidy series of maize inbred Oh43. Plant Mol Biol 75(3):237–251. doi: 10.1007/s11103-010-9722-4

Yu Z, Haberer G, Matthes M, Rattei T, Mayer KFX, Gierl A, Torres-Ruiz RA (2010) Impact of natural genetic variation on the transcriptome of autotetraploid *Arabidopsis thaliana*. Proc Nat Acad Sci 107(41):17809–17814. doi: 10.1073/pnas.1000852107

Chapter 5
Identifying the Phylogenetic Context of Whole-Genome Duplications in Plants

J. Gordon Burleigh

Abstract Although evolutionary biologists have long recognized the transformative evolutionary potential of whole-genome duplications (WGDs) in plants, identifying the precise phylogenetic location of WGDs presents many challenges. This chapter reviews some new approaches to map WGDs on a phylogeny, the first step for understanding the large-scale evolutionary and ecological consequences of WGDs in plants. Specifically, it examines approaches for using chromosome and gene copy number data, gene trees, and other genomic insights to identify the evolutionary location of WGDs. The abundance of genomic sequence data and advances in phylogenetic methods present unprecedented opportunities to place WGDs within the plant tree of life. Still, there exist few direct tests to identify and place WGDs, and analyses of complex data are often susceptible to error.

A central challenge in evolutionary biology is to determine the genetic mechanisms that generate species diversity as well as new traits, functions, and adaptations. Plant evolutionary biologists have long recognized the transformative evolutionary potential of polyploidy or whole-genome duplication (WGD) (e.g., Stebbins 1950; Grant 1963; Levin 2002; Soltis and Soltis 2000; Soltis et al. 2009). However, to study the evolutionary consequences of whole-genome duplications (WGDs) and link WGDs to phenotypic changes or diversification, the WGDs must first be placed in a phylogenetic context. Unfortunately, the rapid gene loss and genome rearrangements that frequently follow WGDs erase evidence of WGD (e.g., Wolfe 2001; Doyle et al. 2008). This process of diploidization brings about a paradox of the study of polyploidy; despite its apparent pervasiveness throughout the evolutionary history of land plants and its evolutionary importance, clear, unambiguous evidence of ancient WGDs can be remarkably difficult to locate within a phylogeny.

J. G. Burleigh (✉)
Department of Biology, University of Florida, Gainesville, FL 32611, USA
e-mail: gburleigh@ufl.edu

P. S. Soltis and D. E. Soltis (eds.), *Polyploidy and Genome Evolution*,
DOI: 10.1007/978-3-642-31442-1_5,

WGD in plants has been and studied for more than 100 years (e.g., Digby 1912; Winge 1917). Although sometimes cryptic, evidence of WGDs may be gleaned from such disparate sources as chromosome counts (e.g., Stebbins 1971), guard cell size (Masterson 1994), age distributions of gene copies (e.g., Blanc and Wolfe 2004), or large segmental duplications within genomes (e.g., Vision et al. 2000). Still, it is difficult to infer the phylogenetic location of the WGDs with any precision from these observations alone. To place WGDs in an evolutionary context requires a robust, and ideally well–sampled, phylogenetic hypothesis, large-scale comparative data indicating WGDs, and models and methods to map these data onto a phylogeny. Only recently have advances in phylogenetics, comparative methods, and genome sequencing made this possible on a large scale. These advances have enabled the first studies addressing some of the basic evolutionary questions about polyploidy, such as the frequency of polyploid speciation (Wood et al. 2009) or the effect of WGDs on diversification (Vamosi and Dickinson 2006; Soltis et al. 2009; Mayrose et al. 2011), in a rigorous phylogenetic framework. This chapter reviews some new approaches to map WGDs on a phylogeny, the first step for understanding the large-scale evolutionary and ecological consequences of WGDs in plants.

5.1 Chromosome Evolution

Many early surveys of polyploidy in plants used chromosome counts to estimate the percentage of polyploid species (e.g., Grant 1963, 1982; Stebbins 1938, 1950, 1971; Goldblatt 1980). This work laid the foundation for understanding the role of WGDs in plant evolution. However, lacking a phylogenetic context, it is difficult to estimate the frequency of WGD events, let alone the evolutionary placement of WGDs simply from the percentage of polyploids (although see Meyers and Levin 2006; Otto and Whitton 2002). For example, if a plant family has 100 species, 50 of which are recent polyploids, this could be the result of as few as one WGD or as many as 50 WGDs (assuming at most one WGD per lineage). Understanding the relationships among species is necessary to infer the history of WGD. Furthermore, chromosome number alone is not necessarily indicative of polyploidy. *Zea mays*, with a haploid chromosome number of 10, has had multiple WGDs in the last $\sim$20 million years (Gaut and Doebley 1997; Gaut 2001), and *Arabidopsis thaliana*, with a haploid chromosome number of five, has experienced 3–5 WGDs since the origin of seed plants (Vision et al. 2000; Blanc et al. 2003; Bowers et al. 2003; Jiao et al. 2011). Yet few chromosome number surveys would have considered either polyploid.

Mapping chromosome numbers on a phylogenetic hypothesis can help reveal the frequency and evolutionary placement of WGD events. Informal phylogenetic observations were first used to deduce ancestral chromosome numbers. For example, numerous studies surmised a low base chromosome number for angiosperms based on the chromosome counts of the "basal" angiosperm lineages (e.g.,

Ehrendorfer et al. 1968; Stebbins 1971; Walker 1972; Raven 1975). These observations implied a history of WGD near the root of angiosperms, although they could not map this precisely; indeed, the relationships among these lineages were unknown. With the growth of phylogenetic methods and data, more formal maximum parsimony approaches were used to reconstruct ancestral chromosome number on the inner nodes of phylogenetic trees (e.g., Stace et al. 1997; Schultheis 2001; Mishima et al. 2002; Guggisberg et al. 2006; Hipp et al. 2007). To do this, chromosome number can be treated as a discrete variable, and the ancestral states can be reconstructed using linear or squared change parsimony. If the ancestral states are far higher than the base chromosome number, then that ancestral node may represent a polyploid. It is possible to construct elaborate chromosome number transition matrices for parsimony analyses, for example, allowing chromosome doubling as well as single chromosome changes, or to weight different changes, like down-weighting chromosome losses, but these analyses are rarely performed. In any case, parsimony reconstructions often have difficulty accounting for multiple transitions on a single branch or quantifying uncertainty in ancestral state reconstructions.

More recently, probabilistic models of chromosome number evolution have been developed (Meyers and Levin 2006; Mayrose et al. 2010). In a simple formulation, the chromosome models allow transitions that add a chromosome, remove a chromosome, or double the number of chromosomes (Mayrose et al. 2010). Thus, although the transition matrix among chromosome states (chromosome numbers) may be extremely large, the evolutionary process can be modeled with only a few parameters. The performance of these models has not been characterized in detail; however, they appear to infer more ancient WGDs than parsimony methods (Wood et al. 2009) and may also provide quite different ancestral state reconstructions of chromosome number (Cusimano et al. 2012). In the future, these models may link chromosome evolution to diversification rates or phenotypes related to the frequency of WGDs to obtain even more accurate estimates of WGDs in a phylogeny.

Studies of plant chromosome numbers have provided a wealth of insight into polyploidy in plants and have contributed substantially to canonical views of plant speciation and evolution. Yet chromosome number is a sort of summary statistic for WGD, a simple observation that is meant to represent a complex, large-scale genomic change, and chromosome number alone may not be sufficient to detect WGDs. A small chromosome number, as in *Arabidopsis thaliana*, does not necessarily imply the absence of historical WGDs, and high chromosome numbers are not necessarily evidence of WGDs. Without additional cytological or genetic data, it is impossible to distinguish between a WGD and increasing dysploidy, a change in the chromosome number that is not associated with a change in the amount of genetic material, based solely on chromosome counts.

Also, as with any phenotype, there are limitations and biases associated with ancestral state reconstruction (e.g., Schluter et al. 1997; Ané 2008). Often reconstruction is most difficult for characters with high rates of evolution or high degrees of homoplasy. Chromosome numbers may be unstable following a WGD

(see Lim et al. 2008; Chester et al. 2012) and can decrease quickly following a WGD. Thus, modeling approaches likely will have difficulty for identifying ancient WGDs. In fact, it appears that the frequency of chromosome loss and diploidization was not always appreciated in studies that only examined chromosome numbers, and even with large-scale genomic data, the mechanisms for rapid chromosome loss are not clear (Doyle et al. 2008). This lack of appreciation for the lability of chromosome numbers may have contributed to the failure to detect, or even surmise, the extent of ancient WGDs, and also may have encouraged the idea that WGDs were evolutionary dead ends (e.g., Stebbins 1950).

Despite the limitations of chromosome numbers alone, data are available for many thousands of plant species, for example on the online Index to Plant Chromosome Numbers (IPCN) database (http://www.tropicos.org/Project/IPCN). Thus, until large-scale genomic sequence data sets become available for thousands of phylogenetically diverse taxa, chromosome number may provide the best opportunity to identify putative WGD events, especially recent events, with phylogenetic precision and to examine the macroevolutionary consequences of WGD throughout the history of all plants.

5.2 Gene Copy Models

With the availability of large-scale genomic data from an increasing number of plant species, estimates of copy numbers for gene families are increasingly available for many plant species. Since WGD should change not only the chromosome numbers but also the copy numbers for all gene families, gene family copy number should provide more data to estimate a WGD than simply a single chromosome number. Like chromosome number evolution, ancestral gene copy numbers can be reconstructed using parsimony methods (Snel et al. 2002; Kunin and Ouzounis 2003; Mirkin et al. 2003; Csürös 2010; Ames et al. 2012; Librado et al. 2012). The parsimony models can be implemented in numerous ways, including weighting gains and losses differently. It is not necessarily easy to find evidence of WGDs based on the number of gene gains or losses on a branch in the species tree, but we might expect WGDs will result in far more gains, and subsequently losses, per unit time than are found on other branches. Hahn et al. (2005) developed a stochastic birth and death model that assumes a homogeneous process of duplication and loss throughout the species tree. The ML implementation of this model in CAFÉ
(De Bie et al. 2006), as well as a similar Bayesian approach (Liu et al. 2011), estimates gene family gain and loss rates across the tree and can identify anomalous gene families and branches on the tree. These branches may reflect the effects of WGDs. More complex models that account for heterogeneity in the rates of duplications and losses across lineages, and in some cases also allow gains of genes or gene families by lateral transfer, also have been proposed (Iwasaki and Takagi 2007; Csurös 2010; Ames et al. 2012; Librado et al. 2012).

In spite of much recent work on developing models of gene family copy number, all of the gene copy models assume that gene duplications or gene gains are independent. Thus, a WGD in a plant might be viewed as 20,000 gene duplications rather than a single duplication event. Consequently, while an increased duplication rate on a branch or increased loss rates on subsequent branches may suggest a WGD, there is no definitive test of WGD, and it may be difficult to distinguish WGD from simply an elevated duplication rate or a large-scale duplication. One approach may be to create a model that could estimate a rate of doubling for all gene family numbers. This transition matrix could be applied to different branches to test for either a rate of WGD greater than zero or different rates between clades or branches.

Although gene copy number provides a more detailed assessment of genomic content than chromosome numbers, inferring the histories of gene copy number and chromosome number have similar limitations. For example, gene copies appear to be rapidly silenced and lost immediately following a WGD (e.g., Tate et al. 2006; Buggs et al. 2009, 2012; see Chap. 14, this volume), which may quickly obscure the evidence for WGDs. However, simply obtaining accurate estimates of gene copy number for extant taxa may be a challenge. Without complete genome sequencing, it can be difficult to distinguish a gene loss from a failure to sample a gene. In fact, the lack of complete sequencing across a broad range of plant species may explain the lack of studies of gene copy number evolution in plants. Even with complete genome sequences, estimates of gene copy number depend on the vagaries of the extremely complex genome annotation and gene family circumscription problems. Furthermore, there are high levels of intraspecific variation in gene copy number in some plants (e.g., Springer et al. 2009; Debolt 2010; Zheng et al. 2011). It is possible to account for uncertainty in the gene copy numbers or incomplete sampling in a likelihood model, although such approaches have not been implemented.

Still, gene copy number does not always provide direct evidence for the location of gene or WGDs. For example, take the case in Fig. 5.1, in which an outgroup has a single gene copy, and two sister taxa each have two gene copies. The parsimonious explanation for these data would be that a gene duplication preceded the most recent common ancestor of the sister taxa, although it is possible that there were independent duplications in each sister lineage. In this case, the gene topologies can provide much more insight into the history of duplication than simply looking at copy number and can easily distinguish between the two duplication scenarios in Fig. 5.1. The additional information from evolutionary history of the genes can further help identify the placement of historical duplications and WGDs.

5.3 Gene Tree Reconciliation

The general problem of gene tree reconciliation is based on the observation that population-level processes, such as coalescence (lineage sorting), as well as evolutionary events, such as gene duplications and loss, recombination,

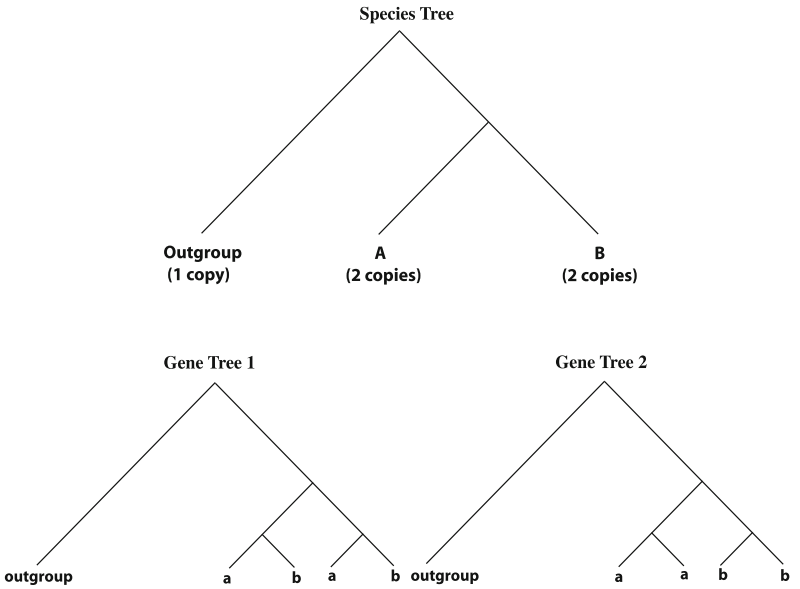

Fig. 5.1 Gene copy number data and corresponding gene trees. The gene copy number data mapped on a species tree implies at least one gene duplication, but do not specify the location of the duplication(s). Gene tree 1 implies a single duplication preceding the most recent common ancestor of A and B. Gene tree 2 implies independent duplications in the lineages leading to A and B

hybridization, or lateral gene transfer can result in gene topologies that differ from the phylogeny of the species in which the genes evolve (e.g., Maddison 1997). The general challenge of gene tree reconciliation is to find the evolutionary scenario that best explains the gene tree topologies. For the case of WGD, we might ask if a collection of gene trees is consistent with a WGD event, or at least a large number of duplications, at a particular point in the species phylogeny. While this approach seeks to directly map the location of gene duplications, and thus provide more direct evidence of WGD than simply examining changes in chromosome or gene copy numbers, in practice it often is complicated by the number of different scenarios that can cause gene tree incongruence and the inherent difficulty of accurately inferring a gene tree from only a gene sequence alignment (for more detailed, general reviews see Eulenstein et al. 2010; Doyon et al. 2011b).

5.3.1 Parsimony Approaches

Much of the initial work in gene tree reconciliation was based on optimizing a parsimony criterion; that is, finding a mapping of a gene tree topology onto the species tree that implies the fewest evolutionary events. The gene duplication model was first

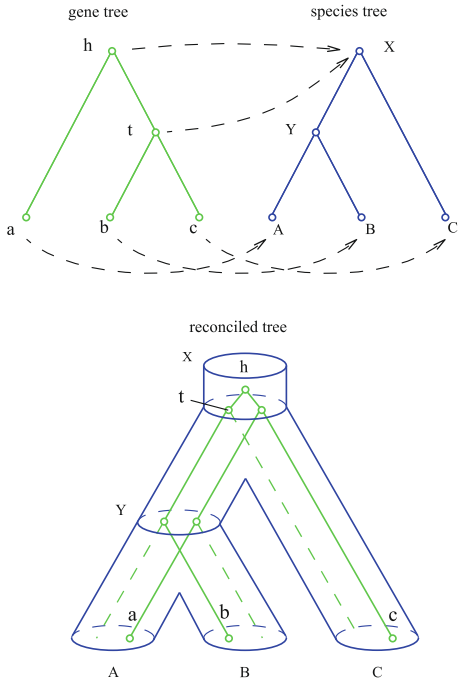

Fig. 5.2 An example of lca-mapping of gene duplications. The parent and child nodes in the gene tree (h and t) map to the same node (X) in the species tree. This implies a single duplication occurring prior to or at node X in the species tree

introduced by Goodman et al. (1979; also see Page 1994 and Guigó et al. 1996) to find the minimum number of gene duplications needed to explain the incongruence between a gene tree and species tree. To do this, a gene tree can be embedded into a species tree through least common ancestor mapping (lca-mapping), which maps every node in the gene tree (tips and internal nodes) to the most recent node in the species tree that could have contained the gene node (Fig. 5.2). A duplication occurred if a parent and child node in the gene tree share the same lca-mapping in the species tree. The lca-mapping identifies the most recent possible location of the gene duplication in the species tree, but this often is not the only possible location of the gene duplication. In many cases, the duplication could have preceded the lca-mapping, although the earlier placement of the duplication may imply additional gene losses. However, as in the gene copy number analyses, gene losses are often difficult to distinguish from incomplete sampling.

This simple gene reconciliation can provide a direct estimate of the phylogenetic location of gene duplication events with a precision that is impossible with chromosome or gene copy number data. Also, it does not even require the presence of duplicated, paralogs genes; incongruence between a single-copy gene tree topology and the species phylogeny may be evidence of a hidden history of gene duplication and loss. On the other hand, gene duplications and losses are not the only explanation of gene tree incongruence. For example, incomplete lineage sorting or reticulation also can confound gene tree topologies. In this case, the gene duplication model may mistakenly imply a large number of duplications

preceding rapid cladogenesis in the species tree, where we might expect high levels of incomplete lineage sorting.

Perhaps the most difficult problem underlying the gene reconciliation approaches is that, in many cases, the incongruence between a gene tree and the species phylogeny may simply be due to error. The gene tree model will interpret any topological error as evidence of duplications. Consequently, this approach often implies far more duplications rather than biologically plausible (e.g., Rasmussen and Kellis 2011). The errors in the gene tree tend to place erroneously large numbers of duplications near the root of the species tree, which may falsely suggest large-scale duplication events at the origins of major clades (Hahn 2007; Burleigh et al. 2010). In the parsimony context, several strategies have been proposed to ameliorate the problems of gene tree error. First, poorly supported clades in the gene tree may be collapsed. Several algorithmic approaches have extended the gene duplication model to deal with reconciling nonbinary trees (Berglund-Sonnhammer et al. 2006; Chang and Eulenstein 2006; Durand et al. 2006). Also, several approaches have been developed to allow minor modifications of the gene tree topology if they reduce the number of implied duplications (e.g., Chen et al. 2000; Chaudhury et al. 2011, 2012; Gorecki and Eulenstein 2012). For example, Chaudhury et al. (2012) introduced an algorithm that, given a gene tree and a species tree, finds a gene topology in a subtree pruning and regrafting (SPR) neighborhood of the original gene tree that minimizes the number of implied duplications. These local rearrangements can massively reduce the number of estimated gene duplication events.

In spite of the many issues related to gene tree reconciliation, simple and informal gene tree reconciliations have been effective at helping to identify the phylogenetic location of WGDs in plants. These approaches are usually limited to small gene trees with paralogs, that is, gene trees in which at least one duplication must have occurred. In a simple three-taxon approach, a gene tree is constructed with a pair of paralogs genes from a test taxon, and the best homologs from a sister taxon and from an outgroup taxon (e.g., Bowers et al. 2003). If paralogs from the test taxon form a clade, they diverged after the common ancestor with the sister taxon; if they do not, they diverged before the most recent common ancestor. This three-taxon phylogenetic approach provides only a limited phylogenetic context for the duplications, but it has been used to determine the timing of WGDs in *Arabidopsis* relative to its divergence from pines, rice, and other eudicots (Bowers et al. 2003) and rice relative to its divergence from pines, *Arabidopsis*, and other monocots (Vandepoele et al. 2003; Chapman et al. 2004). More recently, Jiao et al. (2011) counted the gene trees that were consistent with different scenarios of WGD to infer WGDs at the root of angiosperms and seed plants.

5.3.2 Parsimony Methods to Identify WGDs

The gene duplication problem described above treats each duplication independently. Although it may identify places in the species tree with high numbers of

gene duplications, it does not attempt to find large-scale duplication events. Several proposed approaches attempt to identify the minimum number of gene duplication events, where an event may include duplications of many or all genes, rather than simply the number of duplications. One indirect approach is to examine all the possible locations of each gene duplication and find a mapping that minimizes the number of locations (nodes) on the species tree where gene duplications occur (Guigó et al. 1996; Page and Cotton 2002; Burleigh et al. 2009; Luo et al. 2009). This approach does not directly infer WGDs; but ideally, it can identify places in the species tree that are possible locations of clusters of many duplications. With a limited number of gene trees, this approach appears to be effective at identifying some locations of ancient WGDs in plants (Burleigh et al. 2009). Unfortunately, with a large number of gene trees, it is likely that all possible duplication mappings will require duplication events at every node in the species tree. In this case, every possible mapping of gene duplications will be equally optimal, and this approach will be uninformative.

Another approach seeks to find a gene duplication mapping that implies the fewest gene duplication episodes (Guigó et al. 1996; Page and Cotton 2002; Bansal and Eulenstein 2008; Luo et al. 2009). Given a single gene tree and species tree, any set of gene duplications, from the same or different gene trees, that occur on the same node in a species tree can be explained by a single gene duplication episode (or event) as long as none of the gene duplications in the set have an ancestor–descendant relationship with each other (Fig. 5.3). This approach appears to help identify WGDs, which should be very large episodes, but the largest episode is simply the largest episode found on any single gene tree (Page and Cotton 2002; Burleigh et al. 2010). In practice, the mapping that minimizes the number of episodes is determined by only the largest gene trees. Furthermore, randomizing the leaf labels (taxon names) on the gene trees can result in gene tree mappings that imply fewer episodes (Burleigh et al. 2010). Thus, although the notion of finding gene tree mappings that are consistent with large-scale duplications is desirable, it is not clear that this problem has been properly formulated.

5.3.3 Likelihood-Based Approaches

If the gene trees are accurate, the parsimony criterion for mapping gene duplications appears to perform well when the rates of duplication and loss are low (Åkerborg et al. 2009; Doyon et al. 2009). However, these approaches do not consider branch lengths in the gene or species trees, and they have a limited ability to allow multiple duplications and losses on a single branch. Perhaps more important, it is difficult to incorporate the parsimony criterion into a rigorous statistical framework to examine evolutionary hypotheses associated with gene duplication. Numerous likelihood-based models of gene duplication and loss for reconciling gene trees and species trees have been proposed (e.g., Arvestad et al. 2003, 2004, 2009; Åkerborg et al. 2009; Doyon et al. 2009, 2011a; Rasmussen and

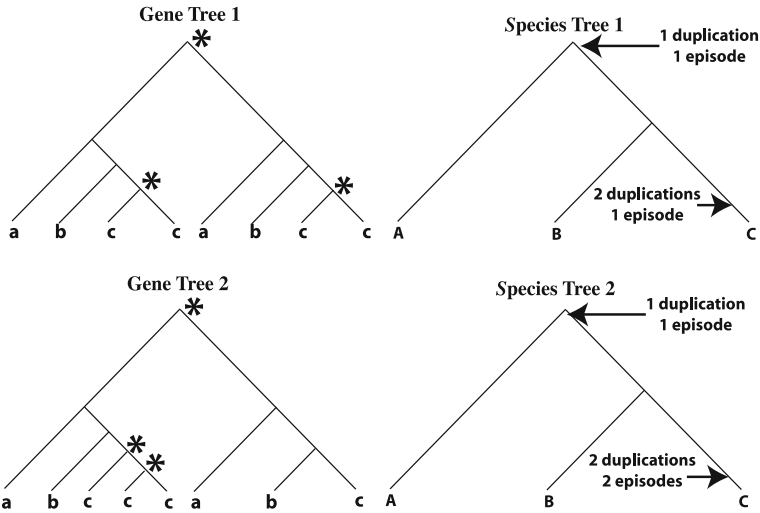

Fig. 5.3 Examples of gene duplication episodes. Duplications in the gene trees are noted with a * followed by their location in the species tree. Both gene trees have three duplications, one at the root node and two at species node C. In gene tree 1, the two duplications at species node C can be explained by a single duplication episode (since there is not a parent–child relationship between the duplication nodes). However, in gene tree 2, the two duplication episodes must have occurred at species node C since one duplication preceded the other

Kellis 2011; Gorecki et al. 2011). Most modeling approaches are based on using birth–death processes to model duplications and losses of genes as they evolve within a species tree (e.g., Arvestad et al. 2003, 2004, 2009; Åkerborg et al. 2009; Rasmussen and Kellis 2011). Also, Doyon et al. (2009, 2011a) introduced a reconciliation algorithm that calculates the likelihood of possible reconciliations based on a constant duplication and loss rate model, and this appears to produce similar reconciliations as a parsimony approach. Gorecki et al. (2011) used a Poisson model to identify the most likely distribution of reconciliations across branches in the tree based on the species branch lengths and the gene tree topologies.

Most of the likelihood model-based approaches require as input a species tree with branch lengths. The method of Gorecki et al. (2011) also requires gene tree topologies. Thus, it may be susceptible to the same problems with gene tree error as parsimony approaches. Several approaches, however, take a gene sequence alignment as input and use a Markov chain Monte Carlo (MCMC) approach to simultaneously obtain the posterior distributions of gene tree topologies and gene duplication and loss mappings (e.g., Arvestad et al. 2004, 2009; Åkerborg et al. 2009; Rasmussen and Kellis 2011). This approach provides an elegant, although computationally difficult, way to incorporate gene tree uncertainty into the gene tree reconciliation. Rasmussen and Kellis (2011) demonstrated that this approach can produce more accurate gene trees and consequently greatly reduce the number of implied duplication events compared to a parsimony approach.

The computational complexity of these likelihood-based approaches raises concerns that they may have difficulty exploiting the magnitude of new genomic data. Yet, in many ways, they still greatly simplify the complexity of genome evolution. Rasmussen and Kellis (2012) have described the first models of both duplication and coalescence, and other modeling approaches estimate the effects of hybridization and coalescence, but not duplication and loss (e.g., Meng and Kubatko 2009; Gerard et al. 2011). These represent important steps in simultaneously accounting for the many processes that affect gene tree topologies. Still, all of the modeling approaches for duplications and losses assume that the genes and all gene duplications are independent. Thus, they may detect branches in the species tree with high rates of duplication or loss, but they do not directly assess the likelihood of WGDs. The development of such models that allow for simultaneous duplications across many genes can allow for rigorous statistical tests of the placement of WGD events.

5.4 Other Genomic Data

From availability of large-scale genomic data, evidence of cryptic ancient WGDs often comes from either identifying large, syntenic (duplicated) blocks within a single genome, or by looking at the age distribution of duplicated genes within a chromosome (see Van de Peer 2004). Since these approaches use only data from a single species and are focused more on identifying ancient WGDs than placing the WGDs in an evolutionary context, I will not cover them in detail. The presence of duplicated chromosomal segments may provide direct, unambiguous evidence of WGDs that may be difficult to obtain from simply gene copy numbers or gene trees (Vision et al. 2000). However, in practice, rapid gene losses and rearrangements after polyploidy can make it extremely difficult to detect such duplications, and different methods of detecting duplicated blocks and using different criteria for defining a syntenic block can greatly affect interpretations of the history of large-scale duplications (see Durand and Hoberman 2006). Although simply examining the genome of a single species cannot reveal the phylogenetic context of a WGD, the dates of the ancient divergences can be estimated based on the molecular divergence of paralogs. It may be possible to map the evolution of large duplicated segments on a tree, but in plants, this may require extending the taxonomic sampling of species with adequate genomic mapping data. Perhaps the greater contribution of these duplicated regions is that they can define sets of paralogs that originated from WGDs, and this information can be used to validate mappings of duplications from gene copy number or gene reconciliation analyses.

The rapidly increasing abundance of large-scale transcriptome data sets for plants provides an opportunity to define WGDs based on the age distribution of duplicated genes (see Cui et al. 2006). The methods first can define pairs of most recent gene duplicates using methods such as an all-by-all BLAST. If gene duplication and loss occur at a constant rate, the frequency of duplicated genes in a genome will decrease exponentially with time. In contrast, a large-scale duplication

event, like a WGD, should result in an overrepresentation of duplicated gene pairs at the time corresponding to the large-scale duplication event. In practice, in a plot of the age distribution, usually represented by synonymous substitution distance, of duplicated genes, peaks in the age distribution curves or evidence of multiple distributions, may indicate WGDs. Again, it is difficult to precisely place a WGD just from the pairwise divergence of duplicated sequences, but with data available from many taxa, comparisons of these age plots from related species can be informative. In some cases, analyses of the age distribution plots have failed to detect known WGDs (e.g., Blanc and Wolfe 2004; Paterson et al. 2004). However, unlike gene tree reconciliation methods, they will not be misled by incomplete lineage sorting or gene tree error.

5.5 Conclusions

New genomic sequence data and advances in phylogenetic methods presents unprecedented opportunities to place WGDs within the plant tree of life. Still, there is much work to do. Although numerous data sources and methods may provide evidence of WGDs, there exist few statistical tests of WGD. A rigorous statistical framework still must be developed to examine hypotheses about the locations of WGDs. Also, examinations of the phylogenetic placement of WGDs often are based on available data; data sets are rarely generated solely for the purpose of placing the location of WGDs. Thus, there has been little discussion about the optimal methods or optimal data sets for mapping WGDs. Indeed, this is a complex issue. For example, gene trees may allow direct observations of the patterns of gene duplication and loss, but they also are susceptible to many errors and biases that may not be problems with simpler data, such as gene copy number. Ideally, learning more about the evolutionary context and implications of WGDs in plants (e.g., their effect on diversification rates and their relationship to phenotypes such as life history or mating system) will also help to identify and place WGDs in plants.

Acknowledgments This chapter was developed and written in parts with support from the Gene Tree Reconciliation Working Group at NIMBioS through NSF award EF-0832858, with additional support from the University of Tennessee. Discussions with members of the working group including Cecile Ané, Oliver Eulenstein, Pawel Gorecki, and Brian O'Meara were helpful for this manuscript.

References

Åkerborg Ö, Sennlad B, Arvestad L, Lagergren J (2009) Simultaneous bayesian gene tree reconstruction and reconciliation analysis. Proc Natl Acad Sci USA 106:5714–5719

Ames RM, Money D, Ghatge VP, Whelan S, Lovell SC (2012) Determining the evolutionary history of gene families. Bioinformatics (In press)

Ané C (2008) Analysis of comparative data with hierarchical autocorrelation. Ann Appl Stat 2:107–1102

Arvestad L, Berglund A-C, Lagergren J, Sennblad B (2003) Bayesian gene/species tree reconciliation and orthology analysis using MCMC. Bioinformatics 19:i7–i15

Arvestad L, Berglund A-C, Lagergren J, Sennblad B (2004) Gene tree reconstruction and orthology analysis based on an integrated model for duplications and sequence evolution. RECOMB 2004:326–335

Arvestad L, Lagergren J, Sennblad B (2009) The gene evolution model and computing its associated probabilities. J ACM 56:7

Bansal MS, Eulenstein O (2008) The multiple gene duplication problem revisited. Bioinformatics 24:i132–i138

Berglund-Sonnhammer A-C, Steffansson P, Betts MJ, Liberles DA (2006) Optimal gene-trees from sequences and species trees using a soft interpretation of parsimony. J Mol Evol 63:240–250

Blanc G, Wolfe KH (2004) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell 16:1093–1101

Blanc G, Hokamp K, Wolfe KH (2003) A recent polyploidy superimposed on older large-scale duplications in the A*rabidopsis* genome. Genome Res 13:137–144

Bowers JE, Chapman BA, Rong J, Paterson AH (2003) Unravelling angiosperm genome eolution by phylogenetic analysis of chromosomal duplication events. Nature 422:433–438

Buggs RJA, Doust AN, Tate JA, Koh J, Soltis K, Feltus FA, Paterson AH, Soltis PS, Soltis DE (2009) Gene loss and silencing in *Tragopogon miscellus* (Asteraceae): comparison of natural and synthetic allotetraploids. Heredity 103:73–81

Buggs RJA, Chamala S, Wu W, Tate JA, Schnable PS, Soltis DE, Soltis PS, Barbazuk WB (2012) Rapid, repeated, and clustered loss of duplicated genes in allopolyploid plant populations of independent origin. Curr Biol 22:1–5

Burleigh JG, Bansal MS, Wehe A, Eulenstein O (2009) Locating large-scale gene duplication events through reconciled trees: implications for identifying ancient polyploidy in plants. J Comput Biol 16:1071–1083

Burleigh JG, Bansal M, Eulenstein O, Vision TJ (2010) Inferring species trees from gene duplication episodes. Proc BCB 2010:198–203

Chang W-C, Eulenstein O (2006) Reconciling gene trees with apparent polytomies. COCOON 2006. LNCS 4112:235–244

Chapman BA, Bowers JE, Schulze SR, Paterson AH (2004) A comparative phylogenetic approach for dating whole genome duplication events. Bioinformatics 20:180–185

Chaudhary R, Burleigh JG, Eulenstein O (2011) Algorithms for rapid error correction for the gene duplication problem (ISBRA) 2011. LNCS 6674:184−196

Chaudhary R, Burleigh JG, Eulenstein O (2012) Efficient error correction algorithms for gene tree reconciliation based on duplication, duplication and loss, and deep coalescence. BMC Bioinformatics 13:s11

Chen K, Durand D, Farach-Colton M (2000) Notung: a program for dating gene duplications and optimizing gene family trees. J Comput Biol 7:429–447

Chester M, Gallagher JP, Symonds VV, Cruz da Silva AV, Mavrodiev EV, Leitch AR, Soltis PS, Soltis DE (2012) Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragapogon miscellus* (Asteraceae). Proc Nat Acad Sci USA 109:1176–1181

Csurös M (2010) Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. Bioinformatics 26:1910–1912

Cui L, Wall PK, Leebens-Mack JH, Lindsay BG, Soltis DE, Doyle JJ, Soltis PS, Carlson JE, Arumuganathan K, Barakat A, Albert VA, Ma H, de Pamphilis CW (2006) Widespread genome duplications throughout the history of flowering plants. Genome Res 16:738–749

Cusimano N, Sousa A, Renner SS (2012) Maximum likelihood inference implies a high, not a low, ancestral haploid chromosome number in araceae, with a critique of the bias introduced by 'x'. Ann Bot 109:681−692

De Bie T, Cristianini N, Demuth JD, Hahn MW (2006) CAFÉ: a computational tool for the study of gene family evolution. Bioinformatics 22:1269–1271

DeBolt S (2010) Copy number variation shapes genome diversity in A*rabidopsis* over immediate family generational scales. Genome Biol Evol 2:441–453

Digby L (1912) The cytology of *Primula kewensis* and of other related *Primula* hybrids. Ann Bot 26:357–388

Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF (2008) Evolutionary genetics of genome merger and doubling in plants. Annu Rev Genet 42:443–461

Doyon J-P, Chauve C, Hamel S (2009) Space of gene/species tree reconciliations and parsimonious models. J Comput Biol 16:1399–1418

Doyon J-P, Hamel S, Chauve C (2011a) An efficient method for exploring the space of gene tree/species tree reconciliations in a probabilistic framework. IEEE/ACM Trans. Comput Biol Bioinform 99: (In press)

Doyon J-P, Ranwez V, Daubin V, Berry V (2011b) Models, algorithms and programs for phylogeny reconciliation. Briefings Bioinform 12:392–400

Durand D, Halldórsson B, Vernot B (2006) A hybrid micro-macroevolutionary approach to gene tree reconstruction. J Comput Biol 13:320–335

Durand D, Hoberman R (2006) Diagnosing duplications—can it be done? Trends Genet 22:156–164

Ehrendorfer F, Krendl F, Habeler E, Sauer W (1968) Chromosome numbers and evolution in primitive angiosperms. Taxon 17:337–468

Eulenstein O, Huzurbazar S, Liberles DA (2010) Reconciling phylogenetic trees. In: Dittmar K, Liberles D (eds) Evolution after gene duplication. Wiley, Hoboken, pp 185–206

Gaut BS (2001) Patterns of chromosomal duplication in maize and their implications for comparative maps of the grasses. Genome Res 11:55–66

Gaut BS, Doebley JF (1997) DNA sequence evidence for the segmental allotetraploid origin of maize. Proc Natl Acad Sci USA 94:6809–6814

Gerard D, Gibbs HL, Kubatko L (2011) Estimating hybridization in the presence of coalescence using phylogenetic intraspecific sampling. BMC Evol Biol 11:291

Goldblatt P (1980) Polyploidy in angiosperms: monocotyledons. In: Lewis WH (ed) Polyploidy: biological relevance. Plenum Press, New York, pp 219–239

Goodman M, Czelusniak J, Moore GW, Romero-Herrera AE, Matsuda G (1979) Fitting the gene lineage into its species lineage, a parsimony strategy illustrated by cladograms constructed by globin sequences. Syst Zool 28:132–163

Gorecki P, Eulenstein O (2012) Simultaneous error correction and rooting for gene tree reconciliation and the gene duplication problem. BMC Bioinformatics (In press)

Gorecki P, Eulenstein O, Burleigh JG (2011) Maximum likelihood models and algorithms for gene tree evolution with duplications and losses. BMC Bioinform 12:S15

Grant V (1963) The origin of adaptations. Columbia University Press, New York

Grant V (1982) Periodicities in the chromosome numbers of the angiosperms. Bot Gaz 143:379–389

Guggisberg A, Mansion G, Kelso S, Conti E (2006) Evolution of biogeographic patterns, ploidy levels, and breeding systems in a diploid-polyploid species complex in primula. New Phytol 171:617–632

Guigó R, Muchnik I, Smith TF (1996) Reconstruction of ancient molecular phylogeny. Mol Phylogenet Evol 6:189–213

Hahn MW (2007) Bias in phylogenetic tree reconciliation methods: implications for vertebrate genome evolution. Genome Biol 8:R141

Hahn MW, De Bie T, Stajich JE, Nguyen C, Cristianni N (2005) Estimating the tempo and mode of gene family evolution from comparative data. Genome Res 15:1153–1160

Hipp AL, Rothrock PE, Reznicek AA, Berry PE (2007) Chromosome number changes associated with speciation in sedges: a phylogenetic study in *Carex* section *Ovales* (Cyperaceae) using AFLP data. Aliso 23:193–203

Iwasaki W, Takagi T (2007) Reconstruction of highly heterogeneous gene-content evolution across the three domains of life. Bioinformatics 23:i230–i239

Jiao Y, Wickett NJ, Ayampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, Soltis DE, Clifton SW, Schlarbaum SE, Schuster SC, Ma H, Leebens-Mack J, de Pamphilis CW (2011) Ancestral polyploidy in seed plants and angiosperms. Nature 473:97–102

Kunin V, Ouzounis CA (2003) GeneTRACE-reconstruction of gene content of ancestral species. Bioinformatics 19:1412–1416

Levin DA (2002) The role of chromosomal change in plant evolution. Oxford University Press, New York

Librado P, Vieira FG, Rozas J (2012) BadiRate: estimating family turnover rates by likelihood-based methods. Bioinformatics 28:279–281

Lim KY, Soltis DE, Soltis PS, Tate J, Matyasek R, Srubarova H, Kovarik A, Pires JC, Xiong Z, Leitch AR (2008) Rapid chromosome evolution in recently formed polyploids in *Tragapogon* (Asteraceae). PLoS ONE 3:e3353

Liu L, Yu L, Kalavacharla V, Liu Z (2011) A bayesian model for gene family evolution. BMC Bioinform 12:426

Luo CW, Chen MC, Chen YC, Yang RWL, Liu HF, Chao KM (2009) Linear-time algorithms for the multiple gene duplication problems. IEEE/ACM Trans Comput Biol Bioinform 99:5555

Maddison WP (1997) Gene trees in species trees. Syst Biol 46:523–536

Masterson J (1994) Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. Science 264:421–424

Mayrose I, Barker MS, Otto SP (2010) Probabilistic models of chromosome evolution and the inference of polyploidy. Syst Biol 59:132–144

Mayrose I, Zhan SH, Rothfels CJ, Magnus-Ford K, Barker MS, Rieseberg LH, Otto SP (2011) Recently formed polyploidy plants diversify at lower rates. Science 333:1257

Meng C, Kubatko LS (2009) Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: a model. Theor Popul Biol 75:35–45

Meyers LA, Levin DA (2006) On the abundance of polyploids in flowering plants. Evolution 60:1198–1206

Mirkin BG, Fenner TI, Galperin MY, Koonin EV (2003) Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes. BMC Evol Biol 3:2

Mishima M, Ohmido N, Fukui K, Yahara T (2002) Trends in site-number change of rDNA loci during polyploidy evolution in *Sanguisorba* (Rosaceae). Chromosoma 110:550–558

Page RDM (1994) Maps between trees and cladistic analysis of historical associations among genes, organisms, and areas. Syst Biol 43:58–77

Page RDM, Cotton JA (2002) Vertebrate phylogenomics: reconciled trees and gene duplication. Pac Symp Biocomput, 536–547

Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Natl Acad Sci 101:9903–9908

Rasmussen MD, Kellis M (2011) A Bayesian approach for fast and accurate gene tree reconstruction. Mol Biol Evol 28:273–290

Rasmussen MD, Kellis M (2012) Unified modeling of gene duplication, loss and coalescence using a locus tree. Genome Res 22:755−765

Raven PH (1975) The bases of angiosperm phylogeny: cytology. Ann Mo Bot Gard 62:724–764

Schluter D, Price T, Mooers AØ, Ludwig D (1997) Likelihood of ancestor states in adaptive radiation. Evolution 41:1239–1251

Schultheis LM (2001) Systematics of *Downingia* (Campanulaceae) based on molecular sequence data: implications for floral and chromosome evolution. Syst Bot 26:603–621

Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Patterson AH, Zheng C, Sankoff D, de Pamphilis CW, Wall PK, Soltis PS (2009) Polyploidy and angiosperm diversification. Am J Bot 96:336–348

Soltis PS, Soltis DE (2000) The role of genetic and genomic attributes in the success of polyploids. Proc Natl Acad Sci 97:7051–7057

Snel B, Bork P, Huynen MA (2002) Genomes in flux: the evolution of archael and proteobacterial gene content. Genome Res 12:17–25

Springer NM, Ying K, Fu Y, Ji T, Yeh C-T, Jia Y, Wu W, Richmond T, Kitzman J, Rosenbaum H, Iniguez AL, Barbazuk WB, Jeddeloh JA, Nettleton D, Schnable PS (2009) Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence (PAV) in genome content. PLoS Genet 5:e1000734

Stace HM, Chapman AR, Lemson KL, Powell JM (1997) Cytoevolution, phylogeny, and taxonomy in Epacridaceae. Ann Bot 79:283–290

Stebbins GL (1938) Cytological characteristics associated with the different growth habits in the dicotyledons. Am J Bot 25:189–198

Stebbins GL (1950) Variation and evolution in plants. Columbia University Press, New York

Stebbins GL (1971) Chromosomal evolution in higher plants. Addison-Wesley, London

Tate JA, Ni Z, Scheen A-C, Koh J, Gilbert CA, Lefkowitz D, Chen ZJ, Soltis PS, Soltis DE (2006) Evolution and expression of homeologous loci in *Tragopogon miscellus* (Asteraceae), a recent and reciprocally formed allopolyploid. Genetics 173:1599–1611

Vamosi JC, Dickinson TA (2006) Polyploidy and diversification: a phylogenetic investigation in Rosaceae. Int J Plant Sci 167:349–358

Vandepoele K, Simillion C, Vande Peer Y (2003) Evidence that rice and other cereals are ancient aneuploids. Plant Cell 15:2192–2202

Van de Peer Y (2004) Computational approaches to unveiling ancient genome duplications. Nat Rev Genet 5:752–763

Vision TJ, Brown DG, Tanksley SD (2000) The origins of genomic duplication in arabidopsis. Science 290:2114–2117

Walker JW (1972) Chromosome numbers, phylogeny, phytogeography of the Annonaceae and their bearing on the (original) basic chromosome number of angiosperms. Taxon 21:57–65

Winge Ö (1917) The chromosomes. Their numbers and general importance. Comptes Rendus des Travaux Laboratoire Carlsberg 13:131–275

Wolfe KH (2001) Yesterday's polyploids and the mystery of diploidization. Nat Rev Genet 2:333–341

Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH (2009) The frequency of polyploidy speciation in vascular plants. Proc Natl Acad Sci USA 106:13875–13879

Zheng L-Y, Guo X-S, He B, Sun L-J, Peng Y, Dong S-S, Liu T-F, Jiang S, Ramachandran S, Liu C-M, Jing H-C (2011) Genome-wide patterns of genetic variation in sweet and grain sorghum *(Sorghum bicolor*). Genome Biol 12:R114

Chapter 6
Ancient and Recent Polyploidy in Monocots

Andrew H. Paterson, Xiyin Wang, Jingping Li and Haibao Tang

Abstract At least two whole-genome duplications (WGD) have profoundly influenced the evolution of most, if not all, grass (Poaceae) genomes, with the most recent of these predating the divergence of these lineages by 20 million or more years. Taxa within each major lineage of Poaceae (e.g., Panicoideae, Ehrhartoideae, Pooideae) have independently experienced additional polyploidizations that have been of central importance to the evolution and productivity of some of our most important crop plants [for example, sugarcane (*Saccharum* spp.), and durum and bread wheat (*Triticum* spp.)]. Following polyploidy, adaptation to the duplicated state is evident at the levels of transmission genetics, chromosome structure, and gene repertoire. While most duplicated chromosomal regions re-establish largely independent evolution within a few million years, 70-million-year-old duplicated chromosome segments in one unusual region of the rice genome and its orthologs in other grasses have continued to exhibit concerted evolution more recently than the divergence of rice subspecies *japonica* and *indica* an estimated 400,000 years ago.

A. H. Paterson (✉) · X. Wang · J. Li · H. Tang
Plant Genome Mapping Laboratory, University of Georgia, Athens, GA, USA
e-mail: paterson@plantbio.uga.edu

X. Wang
e-mail: wang.xiyin@gmail.com

J. Li
e-mail: jingpingli@gmail.com

H. Tang
e-mail: tanghaibao@gmail.com

X. Wang
Center for Genomics and Computational Biology, Hebei United University, Tangshan, People's Republic of China

H. Tang
J. Craig Venter Institute, Rockville, MD, USA

P. S. Soltis and D. E. Soltis (eds.), *Polyploidy and Genome Evolution*,
DOI: 10.1007/978-3-642-31442-1_6,

6.1 Monocot Comparative Genomics

Monocotyledons, also known as monocots, are one of the major clades of angiosperms. Most recent phylogenetic analyses (based largely on plastid sequence data) reveal a successive series of basal angiosperm lineages, with monocots often sister to *Ceratophyllum* plus eudicots (e.g., Soltis et al. 2011). According to the IUCN there are 59,300 species of monocots (http://cmsdata.iucn.org/). The most species-rich family in this clade (and indeed, one of the largest of all angiosperms) are the orchids (family Orchidaceae), with more than 20,000 species (Raven et al. 2005). Among the best-studied monocots are the grasses, family Poaceae (Gramineae), which provide much of the world's food and plant biomass and include economically important grains, such as rice (*Oryza*), wheat (*Triticum*), maize (*Zea*), barley (*Hordeum*), and sorghum (*Sorghum*), turf and forage/pasture grasses, sugar cane, and the bamboos. Other economically important monocot families are the palms (Arecaceae), bananas (Musaceae), gingers (Zingiberaceae), and the onion family Alliaceae.

As of this writing, the vast majority of monocot genetic and genomic information, including all whole-genome-scale DNA sequences of sufficient contiguity for synteny analysis, are for members of Poaceae, necessarily constraining the focus of this chapter. However, this constraint is expected to be relieved in the very near future, with genome sequences in progress for several non-grass monocots (e.g. *Phoenix dactylifera*, *Musa acuminata*, *Ananas comosus*, *Asparagus officinalis*, *Phalaenopsis equestris, Zostera marina*), adding important new dimensions to understanding of monocot gene and genome evolution. A draft of the date palm (*P. dactylifera*) genome is publicly available, but is presently of too low contiguity for robust synteny analysis.

High-quality sequences for representatives of all three major grass clades have been published, including rice (International Rice Genome Sequencing Project 2005; Yu et al. 2005) (Ehrhartoideae), sorghum (Paterson et al. 2009b) and maize (Schnable et al. 2009) (Panicoideae), and *Brachypodium* (The International Brachypodium Initiative 2010) (Pooideae) (Fig. 6.1). Draft genome sequence of barley (*Hordeum*) and wheat (*Triticum*) group 1 chromosomes were recently made available (Mayer et al. 2011; Wicker et al. 2011).

Initial analyses of the available genome sequences available for Poaceae have shown at least two whole-genome duplications (WGD) influencing most, if not all, grass genomes, with the most recent of these predating the divergence of these lineages by an estimated 20 million or more years. Taxa within each lineage have independently experienced additional polyploidizations and readaptation to the duplicated state (for example, sugarcane and durum and bread wheat), with the model genomes constituting a good starting point for accelerating progress in the study and improvement of many additional taxa. Although chloridoid and arundinoid grasses are explored only at the EST level to date, these data show that at least the Chloridoideae experienced an evolutionary history similar to those of the other major grass clades, including both of the two WDGs shared by other

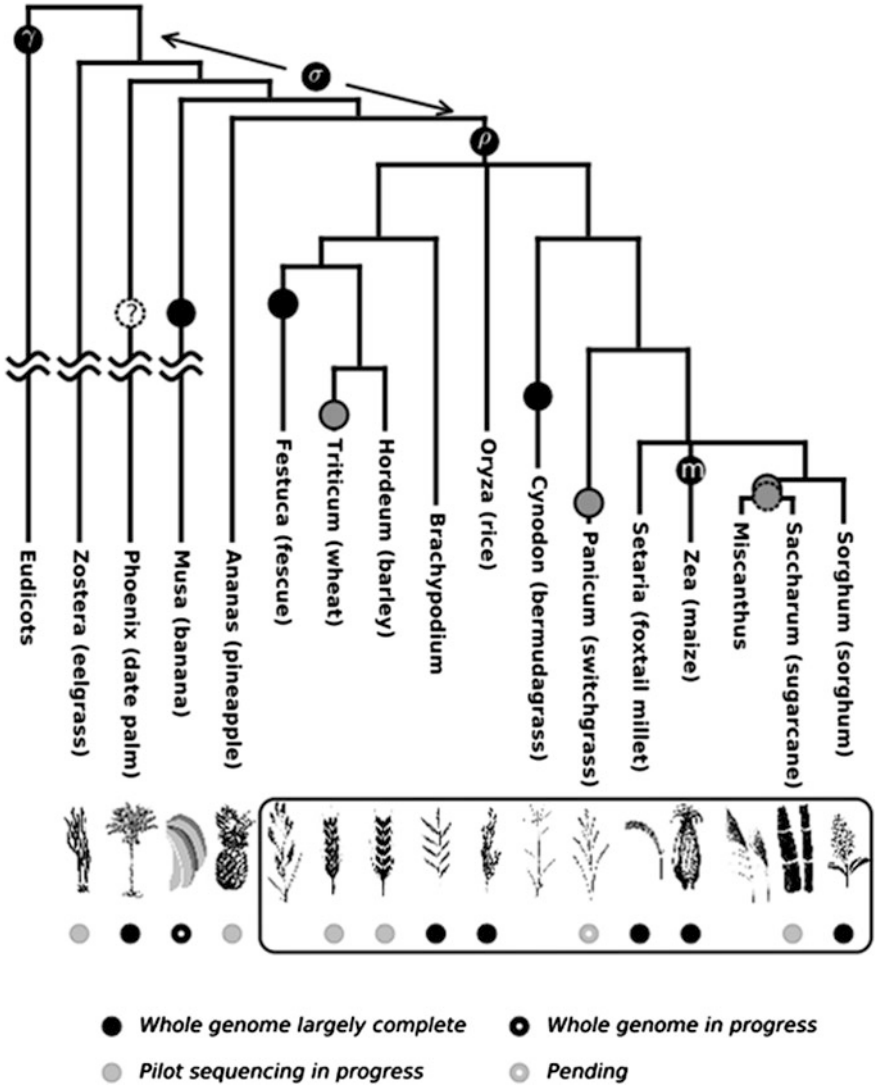

Fig. 6.1 Phylogenetic relationships among major lineages of monocots and estimated positions of early paleopolyploidy events (modified from Paterson et al. 2009a). Notice that some lack sufficient contiguity (*Phoenix*) or data (*Ananas, Zostera*) to infer genome duplication, and accordingly the position of sigma (σ) (Tang et al. 2010) cannot be precisely determined at this time; this uncertainty is indicated by *arrows*

Poaceae. Chloridoids also have an additional lineage-specific genome duplication (Kim et al. 2009).

By developing and using comparative genomics tools, research starting from the sequenced genomes may transitively shed light on monocots that are still

lacking de novo genomic data (Van de Peer 2004; Wang et al. 2006; Lohithaswa et al. 2007; Tang et al. 2008b). The sequences of additional grasses and non-grass monocots such as *Elaeis* (Arecaceae), *Musa* (Musaceae), and *Zostera* (Zosteraceae) will clarify the functional innovation of their gene sets, further elucidating the structural and functional evolution of this ecologically and economically important plant family.

6.2 Rho (ρ), the Most Recent Pan-Poaceae Polyploidy

The notion that even relatively small and structurally simple grass genomes may be paleopolyploid is of long-standing interest. Indeed, secondary associations of grass chromosomes have been known for 80 years (Lawrence 1931), and genetic mapping nearly 2 decades ago suggested duplication of scattered chromosome segments in both rice (Kishimoto et al. 1994; Nagamura et al. 1995) and sorghum (Chittenden et al. 1994).

Paleopolyploidy in the grasses was demonstrated beyond a reasonable doubt with the first availability of the whole-genome sequence of rice, *Oryza sativa* ssp. *japonica* (Goff et al. 2002). A brief controversy about whether the scope of duplication was confined to specific chromosomes (Vandepoele et al. 2003) or the whole genome (Paterson et al. 2003, 2004) was soon reconciled, by analyzing the genome sequence of another rice cultivar, *O. sativa* ssp. *indica* (Yu et al. 2005), with the controversy attributed to differences in the approaches used to infer the duplicated blocks (Wang et al. 2005). This polyploidy event (ρ) was dated to $\sim$70 million years ago (mya) based on putatively neutral DNA substitution rates between duplicated genes, and suggested to be shared by all main lineages of grasses.

The genome sequences of sorghum (*Sorghum bicolor*) and *Brachypodium* (The International Brachypodium Initiative 2010) confirmed that these taxa share the paleoduplication first discerned in rice, and that neither has been affected by additional polyploidization after the evolutionary split of their respective lineages. Moreover, *Brachypodium* and rice have preserved a very high level of gene collinearity (Paterson et al. 2009b; The International Brachypodium Initiative 2010), making it possible to take them as a single genetic system to perform transitive genetics research across different grasses (Freeling 2001). Only a small fraction of genes shows differential gene losses after the split of rice (1.8 %) and sorghum (3.1 %). These findings suggest that, after the 70-mya polyploidization event, the genome of the last common ancestor of grasses had already experienced most gene loss and reached a relatively stable state prior to the divergence of the extant major grass lineages about 50 mya (Paterson 2008; The International Brachypodium Initiative 2010). Some prior reports of deviations from collinearity in the grasses may be accounted for by these low rates of differential gene loss on homoeologous chromosomes.

6.3 Sigma (σ), and Clarifying the Genome Composition of the Last Universal Common Ancestor of Monocots

Rho accounted for much, but not all, of the paleo-duplication that could be discerned in modern cereal genomes (Paterson et al. 2004). Shortly after the discovery of rho, an early report (Zhang et al. 2005), later supported by two independent studies (Jaillon et al. 2007; Salse et al. 2008) hinted at the presence of additional, earlier monocot duplications.

Detailed elucidation of sigma (σ), an additional duplication event preceding rho in the grass lineage, utilized a 'bottom–up' approach, first collapsing 15,640 rice genes and 15,636 sorghum genes into 13,308 rho-nodes that computationally reverse post-rho gene loss, increasing the sensitivity of subsequent analysis.

These nodes (genes) were compared among themselves, revealing collinear patterns of correspondence that involve all nine major synteny blocks resulting from the rho duplication. Some collinear patterns between pairs of rho-blocks are one-to-one, while others are higher order, suggesting that multiple events may have been identified. The eight largest synteny blocks that retained collinearity following sigma contain a total of 4,168 sigma-nodes, covering 5,747 rice genes and 5,738 sorghum genes ($\sim$20 % of the rice and sorghum transcriptomes).

Further study of these sigma-duplicated regions highlights an important constraint in monocot genomics, albeit one that soon will be relieved. There is little remaining intragenomic correspondence between sigma-derived rice segments, although relationships between some of these duplicated segments can still be identified through transitive comparisons of cereal genomes to outgroups. At present, however, the only available outgroups are eudicots such as grape (*Vitis*). Similarities between monocot and eudicot genomes resulting from common ancestry have been obscured by many paleo-polyploidy events and numerous genome rearrangements (Liu et al. 2001; Jaillon et al. 2007). The availability of non-cereal monocot outgroup genomes would significantly increase power to study these segments, particularly outgroups with genomes that have not experienced lineage-specific duplication and associated fractionation of ancestral gene orders.

The lack of an ideal (non-cereal non-reduplicated monocot) outgroup notwithstanding, better understanding of monocot paleopolyploidy has improved our ability to compare monocot and eudicot genomes, a long-standing goal (Paterson et al. 1996) that has been complicated by paleopolyploidy events. Toward this goal, we applied a hierarchical clustering approach, first dividing the chromosomes into small bins and then comparing all pairs of rice and grape bins. Duplicated segments retained in grape following the eudicot 'gamma' hexaploidy event (3), and homologous segments retained in rice following at least two rounds of duplication (rho and sigma), contain 38 "putative ancestral regions" (PAR) that collectively explain 19.1 % of all observed homolog pairs and 31.0 % of reciprocal best hits between grape and rice genes, $\sim$10-fold more than would be explicable by chance. The PARs interleave multiple grape and rice genomic regions collectively

covering ~70 % of each genome. By consolidating much of the redundancy in each genome, the PARs create syntenic blocks with much less ambiguity and in most cases show association between one gamma block and one sigma block. We did not find any PAR that simultaneously mapped to two different gamma or sigma blocks (Tang et al. 2010). Some "ghost duplications" (Van de Peer 2004) in rice that we failed to identify through intragenomic comparisons (due to reciprocal gene losses in largely complementary fashion) are much clearer in cross-species comparisons (PARs).

Compared to the WGD events in grape where 94.5 % of the genome appears duplicated (Jaillon et al. 2007), the rice WGDs are more complicated and degenerate. The 38 grape–rice PARs are a qualitative advance toward a global view of monocot-eudicot synteny. Collinearity appears to be disrupted around the peri-centromeric regions in 10 of the 12 rice chromosomes, suggesting dynamic reorganization of heterochromatic portions of the rice genome that may result from massive transpositions or gene losses (Bowers et al. 2005).

Our unique approach to synteny analysis provides new insight into the number(s) of WGD events experienced by modern cereal genomes. In many lineages, the existence and the numbers of WGD events have been contentious. In 22 of the 38 PARs, grape–rice collinearity is clear, with 12 PARs being threefold redundant in grape, consistent with hexaploidy (Jaillon et al. 2007). The level of redundancy in rice is less clear, ranging from as little as twofold (1 PAR) to sevenfold (3 PARs) and eightfold (5 PARs). In line with the intragenomic evidence from our bottom–up analysis, these high redundancies suggest that the rice lineage experienced more than two, perhaps three, rounds of ancient WGD that have collectively been lumped into the single event that we are presently referring to as 'sigma' (σ). Once again, the availability of non-cereal and non-reduplicated monocot outgroup genomes would significantly improve our understanding of monocot evolutionary history.

6.4 Lineage-Specific Monocot Polyploidies and Their Consequences

Polyploidy in monocots is, of course, not limited to ancient events, and indeed more recent events have been integral to the evolution and productivity of many major crops. A classical textbook example of a more recent polyploidization is the evolution of wheat (see also Chap. 7, this volume), with formation of a tetraploid and subsequently a hexaploid leading to independent crops, durum wheat used for pasta (*Triticum turgidum*) and bread wheat (*T. aestivum*), respectively (Feldman and Levy 2005). Indeed, the synthesis by humans of octoploid triticale (× *Triticosecale* a hybrid between *Triticum* and *Secale*) in the past century illustrates that there may be further gains to be made by mimicking the natural tendency of some plant lineages to form polyploids.

Polyploidy has been particularly closely associated with productivity in autopolyploids such as many forage and biomass grasses. One of the most extreme

cases of a successful autoploid is sugar cane (*Saccharum* spp.), in which cultivated forms are typically octoploid—dodecaploid (i.e., with 8–12 chromosome sets, and sometimes variations in between!). These are interspecific hybrids (and therefore not strict autopolyploids, but a combination of auto- and allopolyploidy) composed of about 85–90 % chromatin from *Saccharum officinarum* and 10–15 % from *S. spontaneum*, a wild relative. Studies using DNA markers to quantify 'dosages' of homologous genomic regions in sugar cane show clearly nonlinear consequences of allele dosage, with one or two copies of favorable alleles at homologous loci usually having favorable effects, but additional copies yielding diminishing or even reduced returns (Ming et al. 2001, 2002a, b). Comparative genomic studies, in particular with sorghum, show that the *Saccharum* lineage experienced at least one WGD since its divergence from the *Sorghum* lineage something less than 10 mya (Ming et al. 1998; Jannoo et al. 2007).

Ploidy in *Saccharum* is particularly interesting, in that it has been further complicated by a high frequency of chromosome non-reduction, yielding $2n + n$ progeny in *S. officinarum* (female) × *S. spontaneum* (male) crosses (Bremer 1923). In efforts to broaden genetic variability in 'noble canes' that were the foundation of sugar cane production until the early part of the twentieth century, F_1 progeny of interspecific crosses between *S. officinarum* and *S. spontaneum* were found to be distinctively more robust than either parent. When *S. officinarum* clones were used as the female parent, progeny tended to be larger stalked, higher in sucrose levels, and generally more vigorous than when *S. spontaneum* clones were used as the female parent. The pollination of noble cane with *S. spontaneum* followed by repeated backcrosses to the noble canes has come to be called 'nobilization', with selected hybrid progenies referred to as 'nobilized' canes (Bremer 1961). A key event in the evolution of modern sugarcane cultivars was the production of the nobilized cultivar, 'POJ2878', of Proefstation Oost, Java, in 1921 (Jeswiet 1929).

Further study of the clade that includes *Saccharum* may prove intriguing. This clade also includes *Miscanthus*, which has biomass yields that are similarly high as *Saccharum*, but is better adapted to temperate climates and therefore has stimulated much interest for bioenergy production in the USA, Europe, and China (Heaton et al. 2008). *Miscanthus* species have a basic set of $n = x = 19$ chromosomes, versus the $x = 10$ that is characteristic of many Saccharinae including *Saccharum*. One attractive hypothesis to explain the transition from 10 to 19 chromosomes is that *Miscanthus*, like *Saccharum* (Ming et al. 1998), may have experienced a polyploidization event in the 8–9 my since its divergence from sorghum—but unlike *Saccharum,* which is largely autopolyploid, *Miscanthus* homologs have diverged sufficiently that they no longer normally pair with one another—that is, there is now preferential pairing of chromosomes. Genome evolution may have included a chromosomal fusion to get from $n = 20$ to 19. If *Miscanthus* and *Saccharum* shared a genome doubling event, a possibility that we are continuing to investigate, it would be an intriguing and perhaps unprecedented case in which one lineage (*Miscanthus*) adapted to the duplicated state by re-establishing disomy, while a sister lineage continued to have the option of polysomy.

6.5 Adapting to the Polyploid State

6.5.1 Centromeric Divergence and Restoration of Disomy

One important challenge facing a newly formed autopolyploid may be that the presence of sets of four homologous chromosomes may tend to hinder purging of deleterious alleles. Transition to diploid inheritance would, in principle, allow more rapid allele frequency changes and reduced genetic load.

Centromeric divergence may have been a mechanism by which paleopolyploid grasses restored diploid inheritance (Bowers et al. 2005). A high concentration of rice genes duplicated by ancient polyploidy falls near K_s 0.85, while rice gene pairs with K_s 0.2–0.6 tend to be located in peri/centromeric regions. This suggests that shortly after polyploidization, a substantial restructuring of centromeric regions began that lasted until about 16 mya [based on the synonymous substitution rate used (Lynch and Conery 2000)]. About 18 % of the rice genome shows highly significant concentrations of matches ($p < 1 \times 10^{-5}$) in the K_s range of 0.2–0.6. The restructuring of non-syntenic regions largely involves migration of DNA between pericentromeric regions of different chromosomes. The concentrations of relatively recent single–locus duplications near the centromeres may, but does not necessarily, reflect a higher duplication rate in these regions. Alternatively, recent duplications may be preserved more frequently in pericentromeric regions, i.e., there may be more rapid loss of single-gene duplications in euchromatic regions (Bowers et al. 2005).

6.5.2 Karyotype Evolution

A polyploidy event may result in genomic instability, consequently incurring a process of diploidization characterized by widespread DNA rearrangements often accompanied by large-scale gene losses (Paterson et al. 2004; Van de Peer 2004; Wang et al. 2006; Xiong et al. 2011). These DNA rearrangements may result in chromosome number variations. Grasses range from $n = 2$ to 18 in their basic chromosome sets (Soderstrom et al. 1987; Hilu 2004). In the sequenced genomes, rice, sorghum, and *Brachypodium* have $n = 12$, 10, and 5 chromosomes, respectively. Although maize (*Zea mays*) experienced a WGD since divergence from sorghum, modern-day maize retains the same chromosome number ($n = 10$) as sorghum. Some lineages of *Sorghum* have experienced chromosome condensations even in the absence of polyploidization (Spangler et al. 1999; Spangler 2003). Comparison of grass genomes has shed light on chromosome number evolution and ancestral grass karyotypes (Salse et al. 2009; Murat et al. 2010). An ancestral karyotype of $n = 5$ chromosomes was inferred (Salse et al. 2009; Murat et al. 2010) before the pan-grass polyploidization, with $n = 2x = 10$ chromosomes after the duplication, then two fissions to result in $n = 2x = 12$ chromosomes in the common

ancestor of major cereals. However, the authors noted that an ancestral karyotype of $n = 6–7$ was also possible. They inferred that chromosome number variation/reduction from the common ancestor may be attributable to non-random centric double-strand break repair events. It was suggested that centromeric/telomeric illegitimate recombination between non-homologous chromosomes led to nested chromosome fusions and synteny break points, and concluded that these break points were meiotic recombination hotspots that corresponded to high sequence turnover loci through repeat invasion. These rules seem to explain most chromosome number changes in the grass genomes sequenced so far, especially the previously observed nested chromosome fusions in *Brachypodium* (The International Brachypodium Initiative 2010). However, many details related to dynamics of centromeres and telomeres during the rearrangements remain unclear, and the wide range in possible ancestral karyotypes (from 12 to 24) suggests that further revision of thinking on this subject is likely.

6.5.3 Gene Retention and Loss

Among the genes duplicated in the polyploidy events shared by the grass lineage, the vast majority of pairs have lost at least one duplicated copy. The finding that only a small fraction of genes show differential gene losses after the split of rice (1.8 %) and sorghum (3.1 %) demonstrates empirically that, after the 70-mya polyploidization, the genome of the last universal common ancestor of grasses had already experienced most gene loss and reached a relatively stable state prior to the divergence of the major grass lineages about 50 mya (Paterson 2008; The International Brachypodium Initiative 2010).

Gene losses have often occurred in a complementary and segmental manner, that is, with non-random patterns of retention/loss on corresponding duplicated DNA segments, in a process known as fractionation (Thomas et al. 2006). Genes may be removed by a short-DNA deletion mechanism (Woodhouse et al. 2010), and in a pair of duplicates, gene loss may be universally biased to preserve the gene that is responsible for the majority of expression (Schnable et al. 2011). For more than 90 % of the preserved duplicated genes, the two copies have the same transcriptional orientations (Wang et al. 2005), and the exceptions may be a result of local DNA inversions or differential gains/losses of new tandemly duplicated genes in the paleo-duplicated regions.

In grasses, and indeed across angiosperms, we find three ‘fates’ of individual gene pairs following duplication (see also Chap. 1, this volume). Most gene functional groups show post-duplication gene preservation/loss rates that are indistinguishable from the genome-wide average. Such ‘neutral’ loss of duplicated genes presumably involves inactivating mutations opposed by very weak selection (Haldane 1933), as the fate of the vast majority of duplicated genes. Population genetic models suggest that loss of duplicated genes may happen over a few million years (Lynch and Conery 2000). Genes in some specific functional categories duplicate and reduplicate

(Blanc and Wolfe 2004; Seoighe and Gehring 2004; Maere et al. 2005; Chapman et al. 2006; Paterson et al. 2006; Tang et al. 2008a), and in many instances can be related to the "gene balance" hypothesis, that stoichiometry among members of pathways and networks is important to biological function (Birchler et al. 2005; Birchler and Veitia 2007; Veitia et al. 2008). Other specific genes and gene functional groups show more extensive loss of duplicate copies than the genome-wide average, and this loss has often been convergent following independent duplications separated by hundreds of millions of years during the evolution of grasses, *Arabidopsis*, yeast, and *Tetraodon* (pufferfish) (Paterson et al. 2006). Much greater knowledge of gene functions, particularly regarding those less-explored genes that are recurrently restored to the singleton state, may provide new insights into the 'adaptation' of a newly formed polyploid to the duplicated state.

6.6 The Unique Case of Rice Chromosomes 11 and 12

Rice chromosomes 11 and 12 (R11 and R12) are a striking exception among chromosomes affected by the 70-mya polyploidization (Wang et al. 2011). R11 and R12 share a ~3-Mb duplicated DNA segment at the termini of their short arms, the formation of which had been dated based on synonymous substitutions to ~5–7 mya (The Rice Chromosomes 11 and 12 Sequencing Consortia 2005; Wang et al. 2005; Yu et al. 2005). Remarkably, in the second grass genome sequence, sorghum, the corresponding region(s) of its orthologous chromosomes (S5 and S8, respectively) also contained such an apparently recent duplication despite having diverged from rice about 50 mya (Paterson et al. 2009b). Physical and genetic maps also suggest shared terminal segments of the corresponding chromosomes in wheat (4, 5), foxtail millet (VII, VIII), and pearl millet (linkage groups 1, 4) (Devos et al. 2000; Singh et al. 2007). It would be exceedingly unlikely for segmental duplications to happen independently at such closely corresponding locations in reproductively isolated lineages. A much more parsimonious hypothesis is that the R11/12 and S5/8 regions each resulted from the pan-grass duplication 70 mya but have an unusual evolutionary history (Paterson et al. 2009b).

Detailed analysis of R11 and R12 suggested that illegitimate recombination has continued for millions of years after the 70-mya divergence of these homoeologs, indeed remaining ongoing in the past 400,000 years since divergence of subspecies *japonica* and *indica* (Wang et al. 2007). Gradual and step-by-step restrictions on recombination, starting around the time of the 70-mya polyploidization, have resulted in 'strata' along the chromosome pair that differ in the degree of DNA sequence similarity between homoeologous genes (Wang et al. 2011). Sequence similarity between homoeologs in the strata reflects the time(s) of recombination suppression rather than the times of their origination. Indeed, the most terminal stratum in rice (RSA) appears <0.5 my old, while two more internal strata (RSB, RSC) appear to have been restricted in their ability to recombine 9.4 and 39.1 mya,

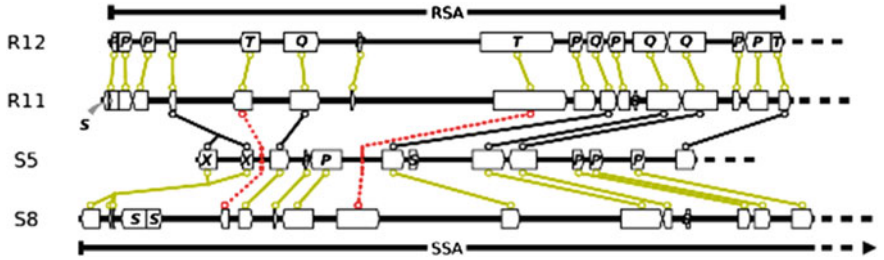

Fig. 6.2 Gene repertoire and organization in a terminal segment of homoeologous rice and sorghum chromosomes experiencing concerted evolution. This region includes 4 rice-sorghum quartets (Q), 3 triplets [T: losses on S5 (2), S8], 13 pairs of taxon-specific genes [P: 5 sorghum, 8 rice]; 8 taxon-specific singletons [S: 6 sorghum, 2 rice], and one lineage-specific duplication (X: two terminal genes on S5 and S8 share common ancestor with single rice genes). Classification letters for a gene family appear in only one member (members connected by lines). Reprinted from Wang et al. (2011) with permission. Copyright is owned by the American Society of Plant Biology

respectively. In sorghum, even the most terminal (of only two) strata dates to 13.4 mya, reflecting the parallel but independent evolution of this unusual chromosome segment in divergent lineages (Wang et al. 2011). The corresponding regions in maize and *Brachypodium* also show prominent homoeologous recombination. However, widespread chromosomal rearrangement, especially in maize after its lineage-specific polyploidization, makes the stratification patterns more difficult to compare than in rice and sorghum.

Both intriguing and perplexing is that the distal chromosomal region with the greatest DNA similarity between surviving duplicated genes also has the highest concentration of lineage-specific gene pairs found anywhere in these genomes, and a significantly elevated gene evolutionary rate (Wang et al. 2011). Of 33 and 23 rice- and sorghum-specific gene pairs on these chromosomes, a respective 100 and 90 % of them are in the young strata. Both members of a remarkable 50 % of the 16 duplicated RSA gene pairs are absent from sorghum, and 15 (38 %) of 39 SSA pairs are absent from rice (Fig. 6.2). Gene losses on either one of a pair of homoeologs experiencing concerted evolution may be commuted to the other, perhaps explaining the more than tenfold higher rate of gene loss in the RSA and SSA regions than the genome-wide averages of 1.8 % in rice and 3.1 % in sorghum since their divergence about 50 mya.

The singular evolutionary history of this pair of grass chromosomes needs further exploration. Elevated gene loss rates and elevated evolutionary rates of the preserved genes in young strata may facilitate speciation in that the loss of alternative copies of duplicated genes leads to reproductive isolation (Werth and Windham 1991; Lynch and Force 2000; Scannell et al. 2006; see also Chap. 1, this volume). The recently proposed inter-relationship between reproductive isolation and autoimmune responses (Bomblies et al. 2007; Yin et al. 2008) draws attention to the finding that orthologs R11 and S5 each contain ~25 % of the NBS-LRR resistance genes (Zhou et al. 2004; Paterson et al. 2009b). Second, a high level of

concerted evolution, associated stratification of chromosomal segments, and extensive homoeologous gene loss are each characteristics of sex chromosomes in organisms from divergent branches of the tree of life, including humans (Lahn and Page 1999), chickens (Lawson Handley et al. 2006), fungi (Charlesworth 2002), and plants (Ming and Moore 2007). Moreover, unexpectedly close proximity between, and co-expansion of, NBS-LRR and several sex-determining gene analogs is found, particularly on S5 (Wang et al. 2011). A hypothesis for further study is whether genes on the various orthologs and paralogs of these chromosomes (or regions therein) could have some 'functional coherence' resembling that of the human Y chromosome (Lahn and Page 1999).

References

Birchler JA, Veitia RA (2007) The gene balance hypothesis: from classical genetics to modern genomics. Plant Cell 19:395–402

Birchler JA, Riddle NC, Auger DL, Veitia RA (2005) Dosage balance in gene regulation: biological implications. Trends Genet 21:219–226

Blanc G, Wolfe KH (2004) Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. Plant Cell 16:1679–1691

Bomblies K, Lempe J, Epple P, Warthmann N, Lanz C, Dangl JL, Weigel D (2007) Autoimmune response as a mechanism for a dobzhansky-muller-typeincompatibility syndrome in Plants. PLOS Biology 5:1962-1972

Bowers JE, Arias MA, Asher R, Avise JA, Ball RT, Brewer GA, Buss RW, Chen AH, Edwards TM, Estill JC, Exum HE, Goff VH, Herrick KL, Steele CLJ, Karunakaran S, Lafayette GK, Lemke C, Marler BS, Masters SL, McMillan JM, Nelson LK, Newsome GA, Nwakanma CC, Odeh RN, Phelps CA, Rarick EA, Rogers CJ, Ryan SP, Slaughter KA, Soderlund CA, Tang HB, Wing RA, Paterson AH (2005) Comparative physical mapping links conservation of microsynteny to chromosome structure and recombination in grasses. Proc Nat Acad Sci USA 102:13206–13211

Bremer G (1923) A cytological investigation of some species and species-hybrids of the genus *Saccharum*. Genetica 5:273–326

Bremer G (1961) Problems in breeding and cytology of sugar cane. 4. Origin of increase of chromosome number in species hybrids of *Saccharum*. Euphytica 10:325–342

Chapman BA, Bowers JE, Feltus FA, Paterson AH (2006) Buffering crucial functions by paleologous duplicated genes may impart cyclicality to angiosperm genome duplication. Proc Nat Acad Sci USA 103:2730–2735

Charlesworth B (2002) The evolution of chromosomal sex determination. Novartis Found Symp 244:207–219 (discussion 220–204, 253–207)

Chittenden LM, Schertz KF, Lin YR, Wing RA, Paterson AH (1994) A detailed Rflp map of *Sorghum-bicolor* X S-propinquum, suitable for high-density mapping, suggests ancestral duplication of sorghum chromosomes or chromosomal segments. Theor Appl Genet 87:925–933

Devos KM, Pittaway TS, Reynolds A, Gale MD (2000) Comparative mapping reveals a complex relationship between the pearl millet genome and those of foxtail millet and rice. Theor Appl Genetics 100:190–198

Feldman M, Levy AA (2005) Allopolyploidy—a shaping force in the evolution of wheat genomes. Cytogenet Genome Res 109:250–258

Freeling M (2001) Grasses as a single genetic system: reassessment 2001. Plant Physiol 125:1191–1197

Goff SA, Ricke D, Lan TH, Presting G, Wang RL, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchinson D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong JP, Miguel T, Paszkowski U, Zhang SP, Colbert M, Sun WL, Chen LL, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu YS, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T, Oliphant A, Briggs S (2002) A draft sequence of the rice genome (Oryza sativa L. ssp japonica). Science 296:92–100

Haldane JBS (1933) The part played by recurrent mutation in evolution. Am Nat 67:5–19

Heaton EA, Dohleman FG, Long SP (2008) Meeting US biofuel goals with less land: the potential of *Miscanthus*. Glob Change Biol 14:2000–2014

Hilu KW (2004) Phylogenetics and chromosomal evolution in the Poaceae (grasses). Aust J Bot 52:10

International Rice Genome Sequencing P (2005) The map-based sequence of the rice genome. Nature 436:793–800

Jaillon O, Aury JM et al (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463–467

Jannoo N, Grivet L, Chantret N, Garsmeur O, Glaszmann JC, Arruda P, D'Hont A (2007) Orthologous comparison in a gene-rich region among grasses reveals stability in the sugarcane polyploid genome. Plant J 50:574–585

Jeswiet J (1929) The development of selection and breeding of the sugarcane in Java. Int Soc Sugar Cane Technol 3:44–57

Kim C, Tang H, Paterson AH (2009) Duplication and divergence of grass genomes: integrating the chloridoids. Trop Plant Biol 2:51–62

Kishimoto N, Higo H, Abe K, Arai S, Saito A, Higo K (1994) Identification of the duplicated segments in rice chromosomes 1 and 5 by linkage analysis of cDNA markers of known functions. Theor Appl Genet 88:722–726

Lahn BT, Page DC (1999) Four evolutionary strata on the human X chromosome. Science (New York) 286:964–967

Lawrence WJC (1931) The secondary association of chromosomes. Cytologia 2:352–384

Lawson Handley LJ, Hammond RL, Emaresi G, Reber A, Perrin N (2006) Low Y chromosome variation in Saudi-Arabian hamadryas baboons (*Papio hamadryas hamadryas*). Heredity 96:298–303

Liu H, Sachidanandam R, Stein L (2001) Comparative genomics between rice and *Arabidopsis* shows scant collinearity in gene order. Genome Res 11:2020–2026

Lohithaswa HC, Feltus FA, Singh HP, Bacon CD, Bailey CD, Paterson AH (2007) Leveraging the rice genome sequence for comparative genomics in monocots. Theor Appl Genetics 115:237–243

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290:1151–1155

Lynch M, Force AG (2000) The origin of interspecific genomic incompatibility via gene duplication. Am Nat 156:590–605

Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y (2005) Modeling gene and genome duplications in eukaryotes. Proc Nat Acad Sci USA 102:5454–5459

Mayer KFX, Martis M, Hedley PE, Simkova H, Liu H, Morris JA, Steuernagel B, Taudien S, Roessner S, Gundlach H, Kubalakova M, Suchankova P, Murat F, Felder M, Nussbaumer T, Graner A, Salse J, Endo T, Sakai H, Tanaka T, Itoh T, Sato K, Platzer M, Matsumoto T, Scholz U, Dolezel J, Waugh R, Stein N (2011) Unlocking the barley genome by chromosomal and comparative genomics. The Plant Cell 23:1249–1263

Ming R, Moore PH (2007) Genomics of sex chromosomes. Curr Opin Plant Biol 10:123–130

Ming R, Liu SC, Lin YR, da Silva J, Wilson W, Braga D, van Deynze A, Wenslaff TF, Wu KK, Moore PH, Burnquist W, Sorrells ME, Irvine JE, Paterson AH (1998) Detailed alignment of

Saccharum and sorghum chromosomes: comparative organization of closely related diploid and polyploid genomes. Genetics 150:1663–1682

Ming R, Liu SC, Moore PH, Irvine JE, Paterson AH (2001) QTL analysis in a complex autopolyploid: genetic control of sugar content in sugarcane. Genome Res 11:2075–2084

Ming R, Del Monte TA, Hernandez E, Moore PH, Irvine JE, Paterson AH (2002a) Comparative analysis of QTLs affecting plant height and flowering among closely-related diploid and polyploid genomes. Genome 45:794–803

Ming R, Wang YW, Draye X, Moore PH, Irvine JE, Paterson AH (2002b) Molecular dissection of complex traits in autopolyploids: mapping QTLs affecting sugar yield and related traits in sugarcane. Theor Appl Genet 105:332–345

Murat F, Xu JH, Tannier E, Abrouk M, Guilhot N, Pont C, Messing J, Salse J (2010) Ancestral grass karyotype reconstruction unravels new mechanisms of genome shuffling as a source of plant evolution. Genome Res 20:1545–1557

Nagamura Y, Inoue T, Antonio B, Shimano T, Kajiya H, Shomura A, Lin S, Kuboki Y, Harushima Y, Kurata N, Minobe Y, Yano M, Sasaki T (1995) Conservation of duplicated segments between rice chromosomes 11 and 12. Breed Sci 45:373–376

Paterson AH (2008) Paleopolyploidy and its impact on the structure and function of modern plant genomes. Genome Dyn 4:1–12

Paterson AH, Lan TH, Reischmann KP, Chang C, Lin YR, Liu SC, Burow MD, Kowalski SP, Katsar CS, DelMonte TA, Feldmann KA, Schertz KF, Wendel JF (1996) Toward a unified genetic map of higher plants, transcending the monocot-dicot divergence. Nat Genet 14:380–382

Paterson A, Bowers J, Peterson D, Estill J, Chapman B (2003) Structure and evolution of cereal genomes. Curr Opin Genet Dev 13:644–650

Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Nat Acad Sci USA 101:9903–9908

Paterson AH, Chapman BA, Kissinger J, Bowers JE, Feltus FA, Estill J, Marler BS (2006) Convergent retention or loss of gene/domain families following independent whole-genome duplication events in *Arabidopsis*, *Oryza*, *Saccharomyces*, and *Tetraodon*. Trends Genet 22:597–602

Paterson AH, Bowers JE, Feltus FA, Tang H, Lin L, Wang X (2009a) Comparative genomics of grasses promises a bountiful harvest. Plant Physiol 149:125–131

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob ur R, Ware D, Westhoff P, Mayer KF, Messing J, Rokhsar DS (2009b) The Sorghum bicolor genome and the diversification of grasses. Nature 457:551–556

Raven PH, Evert RF, Eichhorn SE (2005) Biology of plants, 7th edn. W. H. Freeman, New York

Salse J, Abrouk M, Bolot S, Guilhot N, Courcelle E, Faraut T, Waugh R, Close TJ, Messing J, Feuillet C (2009) Reconstruction of monocotelydoneous protochromosomes reveals faster evolution in plants than in animals. Proceedings of the National academy of sciences of the United States of America 106:14908–14913

Salse J, Bolot S, Throude M, Jouffe V, Piegu B, Quraishi UM, Calcagno T, Cooke R, Delseny M, Feuillet C (2008) Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. Plant Cell 20:11–24

Scannell DR, Byrne KP, Gordon JL, Wong S, Wolfe KH (2006) Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. Nature 440:341–345

Schnable PS, Ware D, Fulton RS, Stein JC, Wei FS, Pasternak S, Liang CZ, Zhang JW, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du FY, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen WZ, Yan L,

Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He RF, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin JK, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren LY, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Van Buren P, Vaughn MW, Ying K, Yeh CT, Emrich SJ, Jia Y, Kalyanaraman A, Hsia AP, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia JM, Deragon JM, Estill JC, Fu Y, Jeddeloh JA, Han YJ, Lee H, Li PH, Lisch DR, Liu SZ, Liu ZJ, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang LX, Yu Y, Zhang LF, Zhou SG, Zhu Q, Bennetzen JL, Dawe RK, Jiang JM, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA, Wilson RK (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115

Schnable JC, Springer NM, Freeling M (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proc Nat Acad Sci USA 108:4069–4074

Seoighe C, Gehring C (2004) Genome duplication led to highly selective expansion of the *Arabidopsis* thaliana proteome. Trends Genet 20:461–464

Singh NK, Dalal V, Batra K, Singh BK, Chitra G, Singh A, Ghazi IA, Yadav M, Pandit A, Dixit R, Singh PK, Singh H, Koundal KR, Gaikwad K, Mohapatra T, Sharma TR (2007) Single-copy genes define a conserved order between rice and wheat for understanding differences caused by duplication, deletion, and transposition of genes. Funct Integr Genomics 7:17–35

Soderstrom TR, Hilu KW, Campbell CS, Barkworth MA (1987) Grass systematics and evolution. Smithsonian Institution Press, Washington

Soltis DE, Smith S, Cellinese N, Refulio-Rodriquez NF, Olmstead R, Crawley S, Black C, Diouf D, Hilu KW, Latvis M, Wurdack K, Xi Z, Davis C, Donoghue M, Soltis PS (2011) Inferring angiosperm phylogeny: a 17-gene analysis. Am J Bot 98:704–730

Spangler R (2003) Taxonomy of Sarga, *Sorghum*, and Vacoparis (Poaceae: Andropogoneae). Aust Syst Bot 16:279–299

Spangler R, Zaitchik B, Russo E, Kellogg E (1999) Andropogoneae evolution and generic limits in *Sorghum* (Poaceae) using ndhF sequences. Syst Bot 24:267–281

Tang H, Wang X, Bowers JE, Ming R, Alam M, Paterson AH (2008) Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. Genome Res 18:1944–1954

Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH (2008) Synteny and colinearity in plant genomes. Science 320:486–488

Tang HB, Bowers JE, Wang XY, Paterson AH (2010) Angiosperm genome comparisons reveal early polyploidy in the monocot lineage. Proc Nat Acad Sci USA 107:472–477

The International Brachypodium Initiative (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463:763–768

Thomas BC, Pedersen B, Freeling M (2006) Following tetraploidy in an *Arabidopsis* ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. Genome Res 16:934–946

Van de Peer Y (2004) Computational approaches to unveiling ancient genome duplications. Nat Rev Genet 5:752–763

Vandepoele K, Simillion C, Van de Peer Y (2003) Evidence that rice and other cereals are ancient aneuploids. Plant Cell 15:2192–2202

Veitia RA, Bottani S, Birchler JA (2008) Cellular reactions to gene dosage imbalance: genomic, transcriptomic and proteomic effects. Trends Genet 24:390–397

Wang X, Shi X, Hao B, Ge S, Luo J (2005) Duplication and DNA segmental loss in the rice genome: implications for diploidization. New Phytol 165:937–946

Wang XY, Shi XL, Li Z, Zhu QH, Kong L, Tang W, Ge S, Luo JC (2006) Statistical inference of chromosomal homology based on gene colinearity and applications to arabidopsis and rice. BMC Bioinformatics 7:447

Wang X, Tang H, Bowers JE, Feltus FA, Paterson AH (2007) Extensive concerted evolution of rice paralogs and the road to regaining independence. Genetics 177:1753–1763

Wang X, Tang H, Paterson AH (2011) Seventy million years of concerted evolution of a homoeologous chromosome pair, in parallel, in major Poaceae lineages. Plant Cell 23:27–37

Werth CR, Windham MD (1991) A model for divergent, allopatric speciation of polyploid pteridophytes resulting from silencing of duplicate-gene expression. Am Nat 137:515–526

Wicker T, Mayer KFX, Gundlach H, Martis M, Steuernagel B, Scholz U, Simkova H, Kubalakova M, Choulet F, Taudien S, Platzer M, Feuillet C, Fahima T, Budak H, Dolezel J, Keller B, Stein N (2011) Frequent gene movement and pseudogene evolution is common to the large and complex genomes of wheat, barley, and their relatives. The Plant Cell 23:1706–1718

Woodhouse MR, Schnable JC, Pedersen BS, Lyons E, Lisch D, Subramaniam S, Freeling M (2010) Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homologs. PLoS Biol 8:e1000409

Xiong Z, Gaeta RT, Pires JC (2011) Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. Proc Nat Acad Sci USA 108:7908–7913

Yin T, Difazio SP, Gunter LE, Zhang X, Sewell MM, Woolbright SA, Allan GJ, Kelleher CT, Douglas CJ, Wang M, Tuskan GA (2008) Genome structure and emerging evidence of an incipient sex chromosome in Populus. Genome Res 18:422–430

Yu J, Wang J, Lin W, Li SG, Li H, Zhou J, Ni PX, Dong W, Hu SN, Zeng CQ, Zhang JG, Zhang Y, Li RQ, Xu ZY, Li ST, Li XR, Zheng HK, Cong LJ, Lin L, Yin JN, Geng JN, Li GY, Shi JP, Liu J, Lv H, Li J, Deng YJ, Ran LH, Shi XL, Wang XY, Wu QF, Li CF, Ren XY, Wang JQ, Wang XL, Li DW, Liu DY, Zhang XW, Ji ZD, Zhao WM, Sun YQ, Zhang ZP, Bao JY, Han YJ, Dong LL, Ji J, Chen P, Wu SM, Liu JS, Xiao Y, Bu DB, Tan JL, Yang L, Ye C, Zhang JF, Xu JY, Zhou Y, Yu YP, Zhang B, Zhuang SL, Wei HB, Liu B, Lei M, Yu H, Li YZ, Xu H, Wei SL, He XM, Fang LJ, Zhang ZJ, Zhang YZ, Huang XG, Su ZX, Tong W, Li JH, Tong ZZ, Li SL, Ye J, Wang LS, Fang L, Lei TT, Chen C, Chen H, Xu Z, Li HH, Huang HY, Zhang F, Xu HY, Li N, Zhao CF, Dong LJ, Huang YQ, Li L, Xi Y, Qi QH, Li WJ, Hu W, Zhang YL, Tian XJ, Jiao YZ, Liang XH, Jin JA, Gao L, Zheng WM, Hao BL, Liu SQ, Wang W, Yuan LP, Cao ML, McDermott J, Samudrala R, Wong GKS, Yang HM (2005) The genomes of Oryza sativa: a history of duplications. Plos Biology 3:266–281

Zhou JH, Wang JL, Xu JC, Lei CL, Ling ZZ (2004) Identification and mapping of a rice blast resistance gene Pi-g(t) in the cultivar Guangchangzhan. Plant Pathol. 53:191–196

Zhang Y, Xu GH, Guo XY, Fan LJ (2005) Two ancient rounds of polyploidy in rice genome. J Zhejiang Univ Sci B 6:87–90

Chapter 7
Genomic Plasticity in Polyploid Wheat

Moshe Feldman, Avraham Levy, Boulos Chalhoub and Khalil Kashkush

Abstract

The importance of hybridization and polyploidization in wheat speciation has been recognized for close to a century (Sakamura 1918; Kihara 1919, 1924, 1954; Percival 1921; Sax 1927). Following these pioneering works, it quickly became apparent that polyploid wheats are not the sum of their constituent genomes. This is not unexpected because the nascent hybrids/polyploids are equipped with a complex set of regulatory elements and of copy number variation that originate from two or more divergent genomes and that generate novel types of interactions and dosage effects. Moreover, they have to adjust at the cytological level, at the level of gene expression, and at the protein level. They also have to maintain genome stability through the regulation of meiotic pairing and recombination, the orchestration of cell division, and the silencing of transposons. The recent studies described here provide an impressive account with regard to the extent and the rapid time course at which a new genetic variant was established upon hybridization and polyploidization. We describe here the current knowledge on the changes that occurred in the wheat genome upon allopolyploidization, starting from the early evolutionary and cytological studies to the recent genomic analyses.

M. Feldman · A. Levy
Plant Sciences Department, The Weizmann Institute of Science,
76100, Rehovot, Israel

B. Chalhoub
UMR INRA 1165—CNRS 8114—UEVE,
Organization and Evolution of Plant Genomes (OEPG),
Unité de Recherche En Génomique Végétale (URGV),
2 Rue Gaston Crémieux, 91057, Evry Cedex, France

K. Kashkush (✉)
Department of Life Sciences, Ben-Gurion University, 84105,
Beer-Sheva, Israel
e-mail: kashkush@bgu.ac.il

P. S. Soltis and D. E. Soltis (eds.), *Polyploidy and Genome Evolution*,
DOI: 10.1007/978-3-642-31442-1_7,

Table 7.1 Superimposition of Sakamura's (1918) finding of the right chromosome number on Schulz's (1913) natural classification of the wheats

Ploidy level	Series	Wild progenitors	Domesticated wheats	
			Spelt (hulled) wheats	Naked (free threshing) wheats
Diploids ($2n = 2x = 14$)	Einkorn (one-grained wheat)	*T. aegilopoides* — >	*T. monococcum*	None
Tetraploids ($2n = 4x = 28$)	Emmer (two-grained wheat)	*T. dicoccoides* — >	*T. dicoccum* — >	*T. durum* *T. turgidum* *T. polonicum*
Hexaploids ($2n = 6x = 42$)	Dinkel	None	*T. spelta* — >	*T. compactum* *T. vulgare*

7.1 The Wheat Group: Natural and Synthetic Polyploids

The discovery of wild emmer wheat, the progenitor of most domesticated wheats (Aaronsohn and Schweinfurth 1906; Aaronsohn 1910), made it possible for Schulz (1913) to assemble the first natural classification of the wheats. He divided the genus *Triticum* into three major groups: einkorn, emmer, and dinkel. This classification was supported by the pioneering cytological study of Sakamura (1918), who was the first to determine the correct chromosome number of the wheats. Sakamura discovered that Schulz's three groups of wheats differ in their chromosome number: the einkorns are diploids ($2n = 14$), the emmers are tetraploids ($2n = 28$), and the dinkels are hexaploids ($2n = 42$) (Table 7.1). It then became obvious that the species of *Triticum* represent a polyploid series with diploid, tetraploid, and hexaploid species.

Since then, the species of the wheat group (*Triticum* and its closely related genus *Aegilops*) have been subjected to extensive taxonomic, cytogenetic, genetic, biochemical, molecular, and evolutionary study by numerous scientists (see reviews of Kihara 1954; Mac Key 1966; Morris and Sears 1967; Kimber and Sears 1987; Feldman et al. 1995; Feldman 2001; Gupta et al. 2005; Dvorak 2009). Due to these extensive studies the allopolyploid species of the wheat group became a classic example of evolution through allopolyploidy. The cytogenetic studies of Kihara (1919, 1924, 1954), Percival (1921), Sax (1927), and others on chromosome pairing in hybrids among species of different ploidy levels showed that all the polyploid species of the group form an allopolyploid series based on $x = 7$. Each allopolyploid species was identified as a product of hybridization followed by chromosome doubling (Fig. 7.1). Von Tschermak and Bleier (1926) were the first to identify a spontaneous chromosome doubling in the cross of wild emmer (*T. turgidum* ssp. *dicoccoides*) with *Aegilops geniculata,* thus demonstrating the possibility of species formation via allopolyploidy in the wheat group. Subsequent studies showed that the frequency of unreduced gametes in intergeneric hybrids of wheat could be in some hybrids as high as 50 % (Kihara and Lilienfeld 1949). Therefore, one might assume that there is a high potential for the frequent and

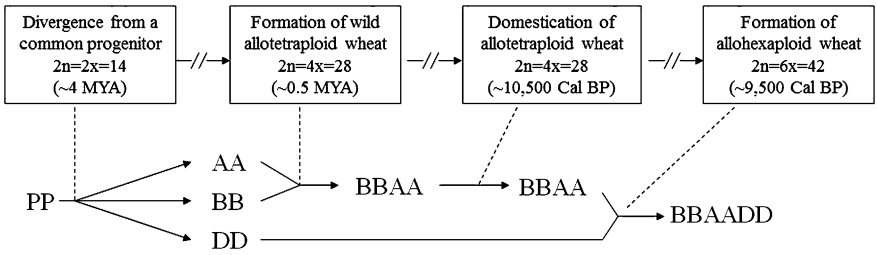

Fig. 7.1 Evolutionary history of allotetraploid and allohexaploid wheat. Diploid wheats ($2n = 2x = 14$) from the *Triticum-Aegilops* group diverged ~4 million years ago (MYA) from a common progenitor (Huang et al. 2002). Interspecific hybridization between the diploid *T. urartu* (genome AA) as male and the donor of the B genome (an unknown species similar to *Ae. speltoides*) as female, followed by chromosome doubling, gave rise, ~0.5 MYA, to wild allotetraploid wheat, *Triticum turgidum* ssp. *dicoccoides* ($2n = 4x = 28$, genome BBAA), an allotetraploid considered as the direct progenitor of durum wheat. Domestication of allotetraploid wheat took place ~10,500 YA and was rapidly followed (~9,500 YA) by a second round of intergeneric hybridization and chromosome doubling between domesticated allotetraploid wheat and the donor of the D genome, *Ae. tauschii* ($2n = 2x = 14$, genome DD), giving rise to bread wheat, an allohexaploid with $2n = 6x = 42$ chromosomes (genome BBAADD). Adapted from Levy and Feldman (2004)

recurrent formation of interspecific or intergeneric hybrids and allopolyploids in the wheat group.

The discovery by Blakeslee (1937) that colchicine can induce chromosome doubling opened new possibilities for the study of wheat evolution through allopolyploidization. It also provided an easy method to synthesize different wheat allopolyploids, some of which have similar genomes to natural allopolyploids and others having new genomic combinations (see examples in Ozkan et al. 2001). Synthetic allopolyploids, either induced or occurring spontaneously, offer excellent tools to mimic the evolutionary speciation events that occurred in nature and to test in a controlled manner; the new features of the hybrid/polyploid genome compared to those of its parents.

Since the discovery that the polyploid species of wheat comprise an allopolyploid series, attempts have been made to identify the diploid donors of the different genomes to the allopolyploids of the wheat group. Most of these attempts used the cytogenetic approach of genome analysis, developed by Kihara (1919, 1924) and based on the concept of genome stability, assuming that the genomes of the allopolyploid species remain similar to those of their diploid parents. However, the accumulating cytogenetic and molecular evidence has indicated that this is not the case; while one genome remains relatively unchanged, the second genome(s) of the allopolyploids have changed considerably from those of their parental diploids. These genomes were termed modified genomes by Kihara (1954) and other wheat cytogeneticists. Every allopolyploid species of *Aegilops* and *Triticum* contains an unchanged genome side-by-side with a modified one whose diploid origin has been intricate and difficult to trace (Zohary and Feldman 1962).

Nevertheless, genome analysis studies revealed that allotetraploid wheat (genome BBAA) originated from hybridization events involving two diploid progenitors classified in the genera *Aegilops* and *Triticum*. Genome B, which is a modified genome, was derived from *Ae. speltoides* and underwent changes on the polyploid level or, more likely, from a closely related species to *Ae. speltoides* which is extinct or extant (Feldman et al. 1995). Genome A, which has been modified relatively little, was derived from *T. urartu* (Chapman et al. 1976; Dvorak 1976). Allohexaploid wheat (genome BBAADD) originated from hybridization between allotetraploid wheat and *Ae. tauschii*, the donor of the D genome (Kihara 1944; McFadden and Sears 1944, 1946). The evolution of the wheats is presented in Fig. 7.1 (for details, see Feldman et al. 1995; Feldman 2001). Modern classification for the *Triticum* group (Van Slageren 1994; Table 7.2) recognizes two diploid species, *T. monococcum* L. and *T. urartu* Tum. ex Gand., two tetraploid species, *T. turgidum* L. and *T. timopheevii* (Zhuk.) Zhuk., and two hexaploid species, *T. aestivum* L. and *T. zhukovskyi* Men. & Er. The economically important wheats are *T. aestivum* (bread wheat, comprising 95 % of the global wheat production) and *T. turgidum* (macaroni wheat).

7.2 Cytological Diploidization

Because of the close relationship of the progenitors and the similarity of the two different genomes of allotetraploid wheat and of the three different genomes of allohexaploid wheat (Morris and Sears 1967), the successful establishment of the polyploid species in nature required the acquisition during the allopolyploidization process of molecular and genetic systems that would prevent intergenomic pairing and recombination. By restricting pairing to fully homologous chromosomes (intragenomic pairing), the cytological-diploidizing systems ensure exclusive bivalent pairing at meiosis and, consequently, regular segregation of genetic material, complete fertility, and genetic stability. Cytological diploidization has been brought about in allopolyploid wheat by two independent systems that complement each other: One system is based on genetic control of pairing, and the second is based on physical divergence of chromosomes.

Historically, the first mechanism of cytological diploidization that received much attention was a genetic system involved in sustaining the exclusive bivalent pairing in allopolyploids. It consisted of the activity of the genetic loci *Ph1* on chromosome arm 5BL and *Ph2* on chromosome arm 3DS (Sears 1976). These loci suppress pairing of homoeologous chromosomes while allowing homologs to pair regularly. The mechanism controlling the *Ph1* mode of action is still unclear. Mapping data suggests the involvement of a complex locus that affects cell cycle progression through the regulation of Cyclin-dependent kinases (Griffiths et al. 2006); however, direct evidence through complementation studies is still missing. It was suggested by several authors that *Ph*-like genes exist in diploid species of the wheat group (Okamoto and Inomata 1974; Avivi 1976; Waines 1976; Maan

Table 7.2 The nomenclature of the commercially cultivated wheats and their immediate wild relatives (after van Slageren, 1994)[a]

Ploidy level	Genome	Species and subspecies	Common name
Diploids ($2n = 2x = 14$)	A^mA^m	*Triticum monococcum*	
		ssp. *aegilopoides*	Wild einkorn
		ssp. *monococcum*	Domesticated einkorn or small spelt
	AA	*Triticum urartu*	None (wild form)
Tetraploids ($2n = 4x = 28$)	GGAA	*Triticum timopheevii*	
		ssp. *armeniacum*	Wild timopheevii
		ssp. *timopheevii*	Domesticated timopheevii
	BBAA	*Triticum turgidum*	
		ssp. *dicoccoides*	Wild emmer
		ssp. *dicoccon*	Domesticated emmer
		ssp *paleocochicum*	Georgian wheat
		ssp. *parvicoccum*[b]	None
		ssp. *durum*	Macaroni or hard wheat
		ssp. *turgidum*	Rivet, cone or pollard wheat
		ssp. *polonicum*	Polish wheat
		ssp. *turanicum*	Khorassan wheat
		ssp. *carthlicum*	Persian wheat
Hexaploids ($2n = 6x = 42$)	GGAA A^mA^m	*Triticum zhukovskyi*	None
	BBAADD	*Triticum aestivum*	
		ssp. *spelta*	Dinkel or large spelt
		ssp. *macha*	None
		ssp. *aestivum*	Common or bread wheat
		ssp. *compactum*	Club wheat
		ssp. *sphaerococcum*	Indian dwarf or short wheat

[a] Taxa derived from a single spontaneous or induced mutation and are not commercially cultivated, such as diploid *T. sinskajae* Filat and Kurk., tetraploid *T. militinae* Zhuk. and Migush., and hexaploid *T. vavilovii* (Tum.) Jakubz., are not included

[b] Extinct, described by Kislev 1980

1977) and they became more effective at the polyploid level as a result of duplication. This dosage-dependent effect might have been selected to improve the fertility of the allopolyploid. This genetic system superimposes itself on, takes advantage of, and thereby reinforces the system of physical homoeologous differentiation already in existence and described below. The genetic system is very effective in suppressing homoeologous pairing in interspecific and intergeneric F1 hybrids. However, its suppressive effect on homoeologous pairing in allopolyploid wheat might not be essential since in plants deficient for *Ph1* there is relatively very little such pairing (Sears 1976). Interestingly, and in accord with the above, gene(s) like *Ph* were not found in all the allopolyploid species of the closely related genus *Aegilops*. In spite of this, these species, relying solely on the structural homoeologous differentiation, exhibit exclusive bivalent pairing of fully homologous chromosomes, i.e., exclusive intragenomic pairing. It might be that

Table 7.3 Types and characteristics of genomic changes in allopolyploid wheat

Revolutionary changes (triggered by allopolyploidization)	Evolutionary changes (Facilitated by allopolyploidy)
Occur during or immediately after allopolyploidization	Occur during the life of the allopolyploid species
Genetic and epigenetic changes	Mostly genetic changes
Species-specific	Population- or biotype-specific
Lead to cytological diploidization	Promote genetic diversity, flexibility, and adaptability
Improve harmonic functioning of the divergent genomes	
Stabilize the nascent allopolyploid and facilitate its establishment as a new species in nature	

stringent selection for fertility under domestication has favored the development of two systems to ensure suppression of multivalent formation and to promote bivalent pairing. Moreover, the cytological diploidization of the allopolyploid wheats that leads to disomic inheritance prevents independent segregation of genes from the different genomes. This mode of inheritance leads to permanent maintenance of favorable intergenomic genetic interactions. It enables fixation of heterotic interaction between genomes and sustained division of tasks (genome asymmetry) between genomes. A series of DNA rearrangements in the allopolyploid further contributes to the physical divergence between the homoeologous chromosomes and to the strengthening of the disomic genetic system. These changes are described below.

7.3 Genomic Structural Changes

Studies with synthetic polyploids as well as genome sequencing data indicate that a broad range of DNA rearrangements occurred during, or soon after, hybridization and polyploidization. What triggers these changes is a fascinating and still open question, but what is clear is that these changes are extensive, including DNA loss, transposon activation, gene duplication, and pseudogenization, and are relatively rapid. Feldman and Levy (2005) have distinguished between the revolutionary changes that occur rapidly and evolutionary changes that take place throughout the evolution of the allopolyploid (Tables 7.3 and 7.4). Note that evolutionary changes might also occur in an accelerated manner, thanks to the buffering of mutations in the polyploid background (Mac Key 1954, 1958; Sears 1972; Thompson et al. 2006), leading to rapid neo- or subfunctionalization of genes and to a process of diploidization and of divergence from the diploid progenitor genomes.

Table 7.4 Revolutionary changes induced by allopolyploidization in wheat

Level	Genetic	Epigenetic
Structural	• Elimination of low-copy DNA sequences	• Chromatin remodeling • Chromatin modifications
	• Elimination, reduction, or amplification of high-copy sequences	• Heterochromatinization • DNA methylation
	• Inter-genomic invasion of DNA sequences • Elimination of rRNA and 5S RNA genes	• Small RNAs activation or repression
Functional	• Gene loss	• Methylation (leading to silencing)
	• Rewiring of gene expression through novel inter-genomic interactions	• Demethylation (leading to gene activation)
	• New dosage response (positive, negative, dosage compensation)	• Release transposons from silencing
	• Gene suppression or activation	• Silencing transposons
	• Transcriptional activation of transposons (that may affect nearby genes in *cis*)	
	• New transpositions of transposons	

7.3.1 Revolutionary changes

Allopolyploidization causes immediate nonrandom elimination of specific non-coding, lowcopy, and high-copy DNA sequences. These sequences are present in all the diploid species of *Aegilops* and *Triticum* but occur in only one pair of chromosomes (chromosome-specific sequences) or in several chromosome pairs of one genome (genome-specific sequences) at the polyploid level (Fig. 7.2) (Feldman et al. 1997; Liu et al. 1998a; Liu et al. 1998b; Ozkan et al. 2001; Shaked et al. 2001; Han et al. 2003; Salina et al. 2004; Han et al. 2005). The extent of DNA elimination was estimated by the determination of the nuclear DNA amount in natural allopolyploids and in their diploid progenitors as well as in newly synthesized allopolyploids and in their parental plants (Ozkan et al. 2003; Eilam et al. 2008; Eilam et al. 2010). Natural wheat allopolyploids contain 2–10 % less DNA than the sum of their diploid parents, and synthetic allopolyploids exhibit a similar loss, indicating that DNA elimination occurs soon after allopolyploidization (Eilam et al. 2008; Eilam et al. 2010). Also, from the very little variation in DNA amount that exists at the intraspecific level, it was concluded that the reduction of DNA content occurred immediately after the formation of the polyploids, and that after this there was almost no change in DNA amount during the life of the allopolyploid species. In triticale (an allopolyploid between wheat and rye, *Secale cereale*), Boyko et al. (1984, 1988) found that there was a great reduction in DNA content in the course of triticale formation with about 9 % for octoploid and 28–30 % for hexaploid triticale. The different genomes were not affected equally in triticale; wheat genomic sequences were relatively conserved, whereas rye genomic sequences were predominantly involved in a very high level of variation and elimination (Ma et al. 2004; Ma and Gustafson 2005, 2006).

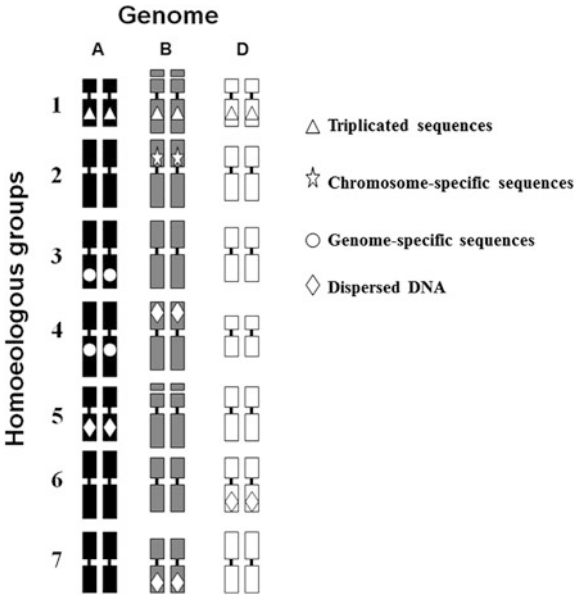

Fig. 7.2 Schematic representation of the wheat karyotype. The wheat karyotype is arranged into genomes A, B, and D and into seven homoeologous groups. Examples of the different types of sequences are drawn on top of the chromosomes, namely: triplicated sequences (group-specific sequences), chromosome-specific sequences (*CSSs*) that are present in only one chromosome pair, genome-specific sequences (*GSSs*) that can be on more than one chromosome pair but only in one of the genomes, and dispersed repeats that are present on both homoeologous and nonhomoeologous chromosomes. Adapted from Levy and Feldman (2004)

Also in hexaploid wheat the genomes were not affected equally; genome D underwent considerable reduction in DNA amount whereas the wheat A and B genomes did not shrink in size (Eilam et al. 2008; Eilam et al. 2010). DNA elimination seems to be nonrandom at the intrachromosomal level as well: Liu et al. (1997) found in allohexaploid wheat that the chromosome-specific sequences of chromosome arm 5BL are not distributed at random along this chromosome arm but cluster in terminal (subtelomeric), subterminal, and interstitial regions of this arm, making these regions extremely chromosome-(homologous-)specific. Hence, it was tempting to suggest that these homologous-specific regions play a critical role in homology search and initiation of meiotic pairing (the classical pairing initiation sites) (Feldman et al. 1997). In some studies, DNA elimination was not observed in the first generations of synthetic allopolyploids (Mestiri et al. 2010; Zhao et al., 2011). It is possible that the timing, as well as the extent of elimination depends on the type of genomic combinations involved or on the type of sequences analyzed.

Table 7.5 Evolutionary changes facilitated by allopolyploidy in wheat

Level	Type
Structural	• Chromosomal repatterning (intra- and inter-genomic translocations)
	• Introgression of chromosomal segments from alien genomes and production of recombinant genomes
Functional	• Nonfunctionalization (deletion or pseudogenization)
	• Subfunctionalization
	• Neofunctionalization
	• Copy number changes
	• New allelic variation

7.3.2 Evolutionary Changes

Several structural changes occur in the allopolyploid wheat genomes, which generate new variants that could not take place in the diploid parental genomes and that occur almost exclusively in an allopolyploid background (Table 7.5). This includes events, such as intergenomic horizontal transfer of chromosomal segments, repetitive sequences, transposons, or genes among the constituent genomes. These events may occur sporadically throughout the history of the allopolyploid species. Intergenomic translocations that are population- or biotype-specific are widespread in allohexaploid wheat (Maestra and Naranjo 1999). Invasion of the A genome by sequences from the B genome—most probably transposons—was detected in wild allotetraploid wheat using GISH (Belyayev et al. 2000). The possibility of intergenomic transfer adds to the allopolyploid genomes' plasticity and enables the creation of new genetic combinations that are beyond the addition of two genomes.

Moreover, in contrast to diploids, which are genetically isolated from each other and have undergone divergent evolution, allopolyploids in the wheat group exhibit convergent evolution, because they contain genetic material from two or more different diploid genomes and can exchange genes with each other via hybridization and introgression, resulting in the production of new genomic combinations. Examples of such introgression between allotetraploid *Aegilops* species that share one genome and differ in the other genome(s) were provided by Zohary and Feldman (1962) and Feldman (1965a, b, c). Such hybridizations are eased by the shared genome, which acts as a buffer and ensures some fertility in the resulting hybrids. In such hybrids, chromosomes of two dissimilar genomes, brought together from different parents, may pair and exchange genetic material and recombine (Feldman 1965a). Additional evidence for the existence of introgressed genomes in allopolyploid *Aegilops* was obtained from C-banding analysis (Badaeva et al. 2004). Introgression of a DNA sequence from allopolyploid wheat to the allotetraploid *Aegilops* species, *Ae. peregrina,* was described by Weissmann et al. (2005).

In addition to evolutionary changes that are almost unique to an allopolyploid background, it might be that other types of mutations that can cause structural or functional changes on an evolutionary scale (e.g., point mutations, satellite

instability, transposition, etc.) can occur in an accelerated manner in the polyploid background. The presence of duplication or triplication of the genetic material in wheat allopolyploids might have relaxed constraints on gene structure and function. Thus, the accumulation of genetic variation through mutations or hybridization might be tolerated more readily in allopolyploid than in diploid species. While there is no direct evidence for this assertion, there is indirect support from experimental data showing a higher resistance of allohexaploid wheat to irradiation compared to the diploid progenitors (Mac Key 1954, 1958; Sears 1972). Such increase in resistance to mutation with increased ploidy level was shown to be correlated with increased evolvability and fitness in yeast (Thompson et al. 2006).

7.4 Functional Changes

Allopolyploidy affects gene or protein function through a variety of mechanisms (Table 7.4). It has been widely suggested that following polyploidization, individual genes follow one of many possible evolutionary fates including *nonfunctionalization* (deletion or pseudogenization), *neofunctionalization* (evolution of novel functions among alleles or homoeoalleles), generating new phenotypes, or *subfunctionalization* (evolution of partitioned ancestral functions among alleles or homoeoalleles) (Lynch and Force 2000; Prince and Pickett 2002; Chaudhary et al. 2009). Functional changes are often regulated by genetic and epigenetic interactions among homoeoalleles and might provide the plasticity that is required to improve the fitness and adaptation of the newly formed allopolyploid and to increase its competitive efficiency with its parental species as well as other plant species, leading to its successful establishment in nature. Moreover, as discussed above, polyploidization seems to facilitate gene evolvability.

7.4.1 Functional Diploidization

Increased gene dosage may lead to redundancy or in some cases may have a deleterious effect, for example, due to the formation of an unbalanced system (Veitia et al. 2008). Regulating gene action in a duplicated genome might be achieved through dosage compensation (nonlinear gene dosage response). A regulatory process that can bring redundant or unbalanced gene systems in polyploids toward a diploid-like mode of expression is functional diploidization (Ohno 1970). **Functional diploidization** is the process whereby existing genes in multiple doses can be eliminated, become inactive via mutations (base substitution, insertion, or deletion leading to pseudogenization) or diverted to new functions.

Examples of functional diploidization in polyploid wheat involve mainly genes that code for structural or storage proteins, e.g., histones, subunits of tubulins,

subunits of glutenins and gliadins, and ribosomal RNA (and possibly also tRNA). In such genes, expression of all homoeoalleles might be redundant and even deleterious, due to overproduction and inefficiency. Also, activity of all homoeoalleles may produce intermediate phenotypes in several traits that decrease the viability of the plants. In this case, traits controlled by genes from only one genome may have a higher adaptive value. It is, therefore, expected that such gene loci would have been targets for genetic diploidization. In hexaploid wheat, intergenomic suppression, as seen by the disappearance of a storage protein subunit, was observed immediately upon formation of a wheat allohexaploid (genome BBAADD; see Galili and Feldman 1984; Galili et al. 1986). This is a common way to instantaneously reduce the negative effect of overproduction and inefficiency of genes that exist in super-optimal dose. Fascinatingly, suppression was reversible: the storage protein reappeared upon extraction of the tetraploid BBAA genomes and disappeared when the D genome was added. Similarly, attempts to transfer a leaf-rust resistance gene from tetraploid to hexaploid wheat failed because of a suppressor gene that was mapped to the D genome (Kerber and Green 1980). Intergenomic suppression of disease resistance genes is a common phenomenon as was noticed in natural and in several newly formed allopolyploids (Anikster, Y., Manisterski, J., and Feldman, M., unpublished data). Comparable results were obtained by Dhaliwal and co-workers (Aghaee-Sarbarzeh et al. 2001) who found that in *Triticum durum-Aegilops* amphiploids dominant leaf rust and stripe rust resistance genes from the *Aegilops* parents were suppressed by genes on the AB genomes of the wheat parent. Another well-studied example of intergenomic suppression is the silencing of rye ribosomal RNA genes in the presence of the wheat genome. Cytosine methylation is involved in this silencing, as suggested by reactivation of the rye ribosomal RNA genes upon treatment with 5-aza-cytidine and analyzed by the use of methylation sensitive/insensitive isoschizomers (Houchins et al. 1997).

Recent analysis of the sequences of wheat group 1 chromosomes shows significant deviations from synteny with many of the nonsyntenic genes representing pseudogenes (Wicker et al. 2011). Part of this pseudogenization might have occurred after polyploidization as suggested from the analysis of the Q-factor. A recent study shows that a combination of mutations in Q genes contributed to the domestic spike phenotype, namely nonfragile, soft glumes, and free threshing (Zhang et al. 2011). The mutation with the most significant phenotypic effect is an amino acid substitution in the protein coded by the 5A locus, while other mutations, such as pseudogenization of the locus on 5B or subfunctionalization of the locus on 5D, also contributed to the domestication phenotype, but to a lesser extent (Zhang et al. 2011). Remarkably, these mutations occurred after polyploidization. The *Hardness* (*Ha*) gene constitutes another example of genetic diploidization, through gene deletion, in polyploid wheat (Chantret et al. 2005).

Genetic diploidization might also be achieved through epigenetic control. Epigenetic silencing can be brought about via cytosine methylation of DNA sequences (Kashkush et al. 2002) or chromatin modifications or remodeling as well as the activity of small RNA molecules (Kenan-Eichler et al. 2011). Gene

silencing can also be achieved via novel regulatory interactions such as intergenomic suppression (Galili and Feldman 1984). Shaked et al. (2001) and Kashkush et al. (2002) reported the occurrence of alterations in cytosine methylation in 13 % of the loci affecting both repetitive DNA sequences and low-copy DNA in approximately equal proportions. Changes in microRNAs, such as miR168 which targets the *Argonaute1* gene, were shown to occur in newly synthesized wheat (Kenan-Eichler et al. 2011).

7.4.2 Subfunctionalization Through Partitioning and Compensation of Duplicated Gene Expression

Functional diversification of duplicated genes (subfunctionalization), i.e., differential or partitioning of expression of homoeoalleles in different tissues and/or in different developmental stages, is also a form of genetic diploidization (Adams et al. 2003). Subfunctionalization, an important aspect of allopolyploidy, has been studied relatively little in allopolyploid wheat. Koebner and co-workers (Bottley et al. 2006) found that differential expression of homoeoalleles in different plant tissues is common in hexaploid wheat. The activity of silenced genes could be restored in aneuploid lines, suggesting that no mutation was involved but rather new *cis–trans* interactions or reversible epigenetic alterations were responsible. The data of Bottley et al. (2006) suggest that for leaf transcripts, there is a modest bias toward silencing of the D genome copies, but this pattern does not extend to root transcripts. Mochida et al. (2006) also presented evidence for differential expression of homoeoalleles in wheat and suggested that inactivation of homoeoalleles is a nonrandom effect. The molecular basis for these cases of instant subfunctionalization remains to be determined. Similarly, subfunctionalization of all three homoeologs of the *Q*-locus in hexaploid wheat was recently described (Zhang et al. 2011).

7.4.3 New Interactions Between Maintained Genes/Proteins

In allohexaploid wheat many gene loci exist in triplicate dose, and the extra gene dosage per se may produce improved or even novel traits. Homoeoalleles may differ from one another by allelic variation, and in this case, activity of all the duplicated genes may produce desirable intergenomic interactions and heterotic effects. These are mainly genes that code for functional proteins (enzymes). This occurrence of enzyme diversity (isozymes) increases the biochemical potential of the allopolyploids. In fact, in wheat allopolyploids the expression of most homoeoalleles coding for functional enzymes is retained (Mitra and Bhatia 1971; Hart 1983a, b, 1987). Also, intergenomic gene interactions may be, in some cases,

expressed in novel traits that do not exist in their parental diploids. Some of these traits may have great adaptive value. Intergenomic gene interactions have direct relevance also to wheat cultivation. For example, the baking quality of allohexaploid wheat (bread wheat) is due to the unique properties of its gluten—a product derived from the combined contribution of the three genomes of hexaploid wheat, and thus exists only at the hexaploid level. In addition, the combination of a large number of spikelets per spike derived from *T. urartu* (the donor of the A genome), with several fertile florets per spikelet originating from the donors of the B genome, facilitated the high fertility of durum and bread wheat. Likewise, merging of *Ae. tauschii* (the donor of the D genome) displaying the cold hardiness phenotype with the prolific nature of allotetraploid wheat (the donor of the A and B genomes) enabled the expansion of wheat cultivation into colder regions. Enzyme multiplicity, derived from the activity of all homoeoalleles, increases the ability of the allopolyploid to adapt to a wider range of environments. This might account for the very wide distribution of wheat under cultivation—much wider than that of any other cultivated plant.

Genome-wide reprograming in gene expression may occur in a polyploid as a result of new interactions among regulatory factors of the parents. For example, novel interactions between the *trans* factor from one species and the *cis* or *trans* factors of the other parental species, as was shown in yeast in a cross between wide hybrids (Tirosh et al. 2009), may account for the observed cases of gene repression or for cases of activation via overexpression.

The consequences of allopolyploidy on gene expression have been widely studied at the genome-wide level in several natural and synthetic allopolyploids of *Triticum*. The majority of studies relied on the comparison of the expression level in the allopolyploid to those of its parents and/or to the average of its parents, expressed as the mid-parental value (MPV). In hexaploid wheat, Pumphrey et al. (2009) found that approximately 16 % of the 825 analyzed genes displayed nonadditive expression in the first generation of synthetic hexaploid wheat. Chagué et al. (2010) analyzed 55,052 transcripts in two lines of synthetic allohexaploid wheat and found that 7 % of the genes had nonadditive expression, while Akhunova et al. (2010) found in synthetic allohexaploid wheat that about 19 % of the studied genes showed nonadditive expression. Similar studies (He et al. 2003) showed that the expression of a significant fraction of genes (7.7 %) was altered in the synthetic allohexaploid *T. turgidum-Ae. tauschii*, and that *Ae. tauschii* genes were affected much more frequently than those of *T. turgidum*. Strikingly, these different studies show that deviation of gene expression in allopolyploids from their parents or average of parents appears to be a common feature, although evaluated in various allopolyploids and using different techniques and approaches. Interestingly, silencing of the same genes was also found in natural hexaploid wheat, i.e., in the variety Chinese Spring (He et al. 2003). Chagué et al. (2010) suggested, based on similar gene expression patterns observed between natural and synthetic wheat allohexaploids, that regulation of gene expression is established immediately after allohexaploidization and maintained over generations. It is of interest to note that several genes that are silent in

the parental species became active in the newly formed allohexaploid (He et al. 2003). Similarly, cDNA-AFLP gels also revealed several cDNAs that were expressed only in the allopolyploids and not in the diploid progenitors (Shaked et al. 2001; Kashkush et al. 2002).

The proportion of genes for which expression in the polyploids is different from the average of the parents is underestimated in these studies, as most of the technologies used do not allow separate analysis of the expression level of each of the gene copies (homoeoalleles) and their respective contribution to overall gene expression (Pumphrey et al. 2009; Chagué et al. 2010; Akhunova et al. 2010). It is not possible to detect, for example, a situation where repression of one homoeoallele is compensated by the activation of the other using microarray technologies. The progress towards sequencing of wheat genomes and the promise of next-generation technologies should allow in the future the resolution of expression at the level of individual homoeologues.

7.5 Genome asymmetry

The rapid processes of cytological and genetic diploidization allow for the development and occurrence of two contrasting and highly important genetic phenomena in allopolyploid wheat that contribute to the evolutionary success of these polyploids: (1) build up and maintenance of enduring intergenomic favorable genetic combinations, and (2) genome asymmetry in the control of a variety of morphological, physiological, and molecular traits, i.e., complete or principal control of certain traits by only one of the constituent genomes. However, while the first phenomenon was taken for granted by plant geneticists, genomic asymmetry in interspecific hybrids and allopolyploids was mainly known in ribosomal RNA genes (reviewed in Pikaard 2000) and only recently has also been documented in other traits (Peng et al. 2003a, b; Fahima et al. 2006; Feldman and Levy 2009; Flagel et al. 2009; Rapp et al. 2009; Flagel and Wendel 2010). The phenomena of diploidization and of dominance in gene expression lead to genome asymmetry, which is manifested in a clear-cut division of tasks among the constituent genomes of allopolyploid wheat (Levy and Feldman 2004; Feldman and Levy 2009; Feldman et al., in press). Genome A controls morphological traits while genome B in allotetraploid wheat and genomes B and D in allohexaploid wheat control the reaction to biotic and abiotic factors (Tables 7.6 and 7.7). Intergenomic pairing would lead to both disruption of the linkage of the homoeoalleles that contribute to positive intergenomic interactions and segregation of genes that participate in the control of certain traits by a single genome. Intergenomic recombination may, therefore, result in many intermediate phenotypes that may affect, in a negative manner, the functionality, adaptability, and stability of the allopolyploids.

Table 7.6 Genome asymmetry in the control of various traits in the wild allotetraploid wheat, *T. turgidum* subsp. *dicoccoides* (genome BBAA)

Traits under control of Genome A	Traits under control of Genome B
Inflorescence morphology	Regulation of ecological adaptation
Free caryopsis	Double the number of disease resistance genes
Glumes with keels	Contains more stress–related genes?
The shape of the edge of the glumes (beaked glumes)	Higher polymorphism of molecular markers
Hairs at the base of every spikelet	Higher polymorphism of HMW glutenin genes
Plant habitus	Larger amount of repetitive sequences
Growth habit	Activity on nucleolar organizers
Autogamous behavior	Larger number of rRNA genes
Many domestication genes	

7.6 Response of Transposable Elements to Allopolyploidization

In the preceding sections, we have discussed structural changes and changes in gene expression resulting from genetic or epigenetic phenomena. Transposable elements (TEs) are discussed here separately due to their specific mode of action, their abundance (up to 90 % of the wheat genomes; Sabot et al. 2005), and their impact on both structure and expression of the genome.

7.6.1 Transcriptional Activation of Retrotransposons in Synthetic Wheat Polyploids

It is now clear that some eukaryote retrotransposon promoters retain activity under normal conditions and initiate either read-in transcripts of the transposon itself or read-out transcripts into flanking host sequences (Vicient et al. 2001; Kashkush et al. 2002; Nigumann et al. 2002; Kashkush et al. 2003; Kashkush and Khasdan 2007). Following allopolyploidization events in wheat, the steady-state level of expression of LTR retrotransposons was massively elevated (Kashkush et al. 2002, 2003, unpublished data), similarly to what was observed in synthetic *Arabidopsis* allopolyploid hybrids (Madlung et al. 2005). In addition, the transcriptional activity of a LTR element termed *Wis2-1A* (Lucas et al. 1992) leads to the production of read-out transcripts toward flanking host DNA sequences, a process that occurred in a genome-wide manner (Kashkush et al. 2003). In many cases, these read-out transcripts were associated with the expression of adjacent genes, depending on their orientation: knocking down or knocking out the gene product if the read-out transcript was in the antisense orientation relative to the orientation of

Table 7.7 Genome asymmetry in the control of agronomic traits in domesticated durum (genome BBAA) and bread wheat (genome BBAADD)[a]

Traits	Traits under control of		
	Genome A	Genome B	Genome D
Inflorescence morphology			
Elongated glumes	*Eg P*1 on 7AL	*Eg P*2 on 7BL (?)	
Branched spikes	*Bh* on 2AS		
Nonbrittle rachis	*br A*1 on 3AS *br* A2 on 2A	*br B*1 on 3BS	
Nontenacious glume (lax glume)		tg2 on 2BS	tg1 on 2DS
Reduce plant height	*Rht*7 on 2A; *Rht*12 on 5AL;	*Rht B*1 *on 4BS; Rh*4 on 2Bl; *Rht*5 on 3BS; *Rht*9 on 7BS; *Rht*13 on 7BS	*Rht D*1 on 4DS; *Rht*8 on 2DL;
Grain protein content		*Gpc B*1 on 6BS	*Pro*1 on 5DL; *Pro*2 on 5Ds
Grain hardness			*Ha* on 5DS
Puroindolines and grain softness protein			*Pin D*1 on 5DS
Gibberellic acid response		*Ga*1, *Ga*3 on 4BS	Ga2 on 4DS
Waxiness		*W*1 on 2BS	*W*2 on 2DS (?)
Epistatic inhibitors of waxiness		$W1^I$ on 2BS; $W3^I$ on 1BL	$W2^I$ on 2DS
Male sterility	*Ms*3 on 5AS; *ms*5 on 3A	*ms*1 on 4BS	*Ms*2 on 4DS; *Ms*4 on 4DS
Pairing homoeologous		*Ph*1 on 5BL	*Ph*2 on 3DS
Hybrid necrosis		*Ne*1 on 5BL; *Ne*2 on 2BS	
Hybrid chlorosis	*Ch*1 on 2A		*Ch*2 on 3DL
Aluminum tolerance			*Alt*2 on 4DL
Boron tolerance		*Bo*1 on 7BL	
Low cadmium uptake		*Cdu*1 on 5BL	
Iron deficiency		*Fe*2 on 7BS	*Fe*1 on 7DL
Herbicide response			
Difenzoquat insensitivity		*Dfg* 1 on 2BL	
Chlortoluron insensitivity	*Imi*3 on 6AL	*Su*1 on 6BS	*Imi*1 on 6DL
Imidazolinone resistance		*Imi*2 on 6BL	
Response to photoperiod		*Ppd-B*1 on 2BS	*Ppd-D*1 on 2DS
Response to vernalization	*Vrn-A*1 on 5AL	*Vrn-B*1 on 5BL; Vrn-B3 on 7BS	*Vrn-D*1 on 5DL; Vern-D4 on 5DL;Vern-D5 on 5DL
Response to salinity			*Kna*1 on 4DL
Frost resistance	*Fr*1 on 5AL		*Fr*2 on 5DL
Number of resistance genes to diseases and pests	45	88	51

[a] Data from the 2008 Wheat Gene Catalogue (http://wheat.pw.usda.gov/GG2/index.shtml)

the gene transcript (such as the *iojap-like* gene) or overexpressing the gene if the read-out transcript was in the sense orientation (such as the *puroindoline-b* gene). The mechanisms by which transcriptional activation of TEs influences the expression of neighboring genes are poorly understood. In some cases, the correlation between the reduction of the sense expression of the gene and the production of the antisense strand that initiated from the adjacent transposon promoter (Kashkush et al. 2003; Puig et al. 2004) might indicate that post-transcriptional gene silencing is a major mechanism for inactivating adjacent genes. Recent studies on tracking methylation changes around a LTR retrotransposon in the first four generations of a newly formed wheat allopolyploid (Kraitshtein et al. 2010) may indicate that this read-out activity is restricted to the first generations of the nascent polyploid species.

7.6.2 Massive Methylation of TE-Adjacent DNA Sequences Following Allopolyploidization

Alterations in the genomic methylation patterns following allopolyploidization have been examined in several polyploid systems, including *Arabidopsis* (Madlung et al. 2002; Belzile et al. 2009), *Spartina* (Salmon et al. 2005; Parisod et al. 2009), *Brassica* (Lukens et al. 2006; Wang et al. 2009), and wheat (Shaked et al. 2001). The methylation alterations are either hyper- or hypomethylation, depending on the sequence analyzed, and are reproducible. Recent studies in wheat have investigated in detail the methylation of CCGG sites flanking several TE families (Kraitshtein et al. 2010; Yaakov and Kashkush 2011a, b; Zhao et al. 2011). In one study (Kraitshtein et al. 2010), transposon methylation display (TMD) analysis was applied (see Kashkush and Khasdan 2007) to analyze a terminal-repeat retrotransposon in miniature (TRIM), termed *Veju*, in *Triticum turgidum* ssp. *durum* (genome AABB) and *Aegilops tauschii* (genome DD), and the first four generations of the derived allohexaploid. It was estimated that over 50 % of the CCGG sites flanking *Veju* elements showed altered TMD patterns in the first four generations of the newly formed allohexaploid. Hypomethylation of *Veju*-flanking CCGG sites was predominant in the first generation of the newly formed allohexaploid, while hypermethylation was predominant in subsequent generations. This might indicate reduced *Veju* transcriptional activity after the third generation of the synthetic allohexaploid. A similar pattern of hypomethylation of *Veju* elements was also observed in the first three generations of a synthetic allotetraploid that was derived from a cross between *Ae. sharonensis* (genome S^lS^l) and *T. monococcum* ssp. *aegilopoides* (genome A^mA^m) (Yaakov and Kashkush 2011b). However, unlike in the synthetic allohexaploid, *Veju* elements remained hypomethylated up to the fourth generation of the synthetic allotetraploid. In support of these studies, Zhao et al. (2011) observed massive methylation changes around *Veju* in three different combinations of newly formed wheat allohexaploids.

The methylation patterns of three TEs [*Balduin* (belonging to the *CACTA* superfamily), *Apollo* (belonging to the *MuDR/Foldback* superfamily), and *Thalos* (a *stowaway*-like *MITE* belonging to the Tc1/*mariner* superfamily)] have also been analyzed in allopolyploid wheat. The CCGG sites flanking the three elements underwent massive hypermethylation in the first four generations of the synthetic allohexaploid (Yaakov and Kashkush 2011a, b), while they underwent massive hypomethylation in the first four generations of the synthetic allotetraploid (Yaakov and Kashkush 2011b). The massive hypermethylation of these elements in the synthetic allohexaploid might be connected to the lack of transpositional activity (Yaakov and Kashkush 2011a). It is important to mention that transcriptional activation of LTR retrotransposons does not correlate with transpositional activity (Kashkush et al. 2003).

7.6.3 Changes in TE Composition Following Allopolyploidization

The prevalence of TEs and their inherent sequence similarity make them a prime target for illegitimate and nonhomologous recombination. TEs have been shown to undergo rearrangements following allopolyploidization in *Spartina* (Parisod et al. 2009), tobacco (Petit et al. 2010), and *Triticale* (Bento et al. 2008). Recent data for synthetic allohexaploid wheat indicate that rearrangements of retrotransposon-containing sequences occur rapidly and reproducibly (repeated in independently newly formed allopolyploid lines) in the first generations following polyploidization. In addition, a change in the methylation status (usually hypomethylation) in the first generation was followed by deletion of retrotransposon-containing sequences in subsequent generations (Kraitshtein et al. 2010). These data suggest a correlation between methylation and post-allopolyploidization rearrangements that occur via a mechanism that has yet to be identified. One possible explanation is that hypomethylation confers an open chromatin structure to the TE sequences, which exposes these demethylated elements to be targeted for deletion by the host. There is evidence that small RNAs corresponding to *Veju* elements might play a pivotal role in *Veju* methylation in the newly formed wheat allohexaploid (Kenan-Eichler et al. 2011).

Despite the altered methylation status and transcriptional activation of TEs following allopolyploidization, there are very few reports on the transpositional activity of transposons. Madlung et al. (2005) showed both methylation alterations and limited transpositional activation of a *Sunfish* transposon in polyploid *Arabidopsis*. Petit et al. (2010) reported an increase in the copy number of a *Tnt1* retrotransposon in allotetraploid tobacco. No transposition bursts were reported in *Spartina* (Parisod et al. 2009) or in wheat (Kashkush et al. 2003; Kraitshtein et al. 2010; Yaakov and Kashkush 2011a). These reports suggest that the transpositional activity of TEs following allopolyploidization might be restricted to specific TE families (Parisod et al. 2010). Recently, it was shown that the immense loss of *Veju* sequences in the first generation of the synthetic allohexaploid is probably

followed by retrotransposition in subsequent generations, a process that causes new insertions to accumulate in allohexaploids (Kraitshtein et al. 2010). The same study also suggests that these new insertions are targeted for methylation. Methylation of the new *Veju* elements protects the genome from deleterious transposon insertions. Investigating the scale of eliminated DNA sequences, including TE sequences, by identifying the deletion breakpoints will allow a better understanding of the mechanism(s) involved and of the nature of the connection between methylation and rearrangements.

In summary, different classes of wheat TEs appear to respond differently to the allopolyploidization event. Table 7.8 summarizes the type of response of 12 different class I and class II elements. It can be seen clearly that epigenetic response, mainly methylation changes, is a common factor for all studied TEs, while genetic response that includes rearrangements and/or transpositions is restricted to specific TE families. Transcriptional data for most studied TEs are still lacking (Table 7.8). However, there is a good basis for suggesting a connection among methylation changes in TEs with alteration in their expression patterns.

7.7 Concluding Remarks

The studies reviewed above indicate that, in the wheat group, hybridization and chromosome doubling induce a burst of genomic alterations, some of which could not occur at the diploid level. Some of these changes might improve the ability of the newly formed allopolyploids to survive in nature and to compete with their parental species, corresponding thus to phenotypic and adaptive novelty. Other changes are probably deleterious. TEs seem to play an important role in the various responses to hybridization and polyploidization due to their abundance and also due to their tendency to be dysregulated as a result of genomic shocks. The balance between the beneficial and deleterious changes associated with allopolyploidization is probably what determines the fate of the nascent species.

The formation of an allopolyploid species is accomplished rapidly via the combined processes of hybridization and genome doubling, but its establishment in nature as a successful species probably depends on a high degree of plasticity that enables it to overcome potential incompatibilities and to gain new traits. The studies reported here suggest that wheat can achieve genomic plasticity through the induction of a series of cardinal nonadditive genomic changes. Some changes, genetic and epigenetic, are rapid and nonMendelian, occurring during or immediately after the formation of the F1 hybrid or the allopolyploid (revolutionary changes). Other changes occur sporadically over a long time period during the life of the allopolyploid species (evolutionary changes). From a population point of view, the chance of a new individual, such as a nascent hybrid/allopolyploid, to establish itself as a new species is almost nil, unless it has some increased fitness over its parents. This fitness advantage must be manifested within a few generations of formation or the new species will rapidly be extinct. The revolutionary

Table 7.8 Summary of the genetic and epigenetic responses of several TE families to allopolyploidization in wheat

Family	TE	Genetic and epigenetic alterations in synthetic allopolyploids (compared to parental lines)[a]				Reference
		Methylation	Transcription	Rearrangement[b]	Transposition[c]	
TRIM	*Veju*	✔	✔	✔	?	Kraitshtein et al. (2010) Kenan-Eichler et al. (2011) Zhao et al. (2011)
Copia	*Angela*	✔	NA	NA	×	Unpublished
	BARE-1	✔	NA	×	×	Zhao et al. (2011)
	Wis2-1A	NA	✔	×	×	Kashkush et al. (2003)
Gypsy	*Sabrina*	✔	NA	×	×	Unpublished
CACTA	*Balduin*	✔	NA	×	×	Yaakov and Kashkush (2011a)
MuDR/ Foldback	*Apollo*	✔	NA	×	×	Yaakov and Kashkush (2011a)
Stowaway-like MITE	*Thalos*	✔	NA	×	×	Yaakov and Kashkush (2011a)
	Eos	✔	NA	✔	?	Unpublished
	Minos	✔	NA	✔	✔	Unpublished
	Oleus		NA	✔	?	Unpublished
	Fortuna	✔	NA	✔	?	Unpublished

[a] ✔ altered, × no change, ? not validated, *NA* data not available
[b] Include deletion and/or insertion
[c] Typical TE transposition

changes described here may contribute to the establishment of the new species. Instantaneous elimination of sequences from one genome in the newly formed allopolyploids increases the divergence of the homoeologous chromosomes, and thus leads to exclusive intragenomic pairing that improves fertility. Mechanisms such as loss of deleterious genes (e.g. genetic incompatibilities) or positive dosage effects or new intergenomic heterotic interactions may all rapidly increase the fitness of the nascent species. The evolutionary changes, on the other hand, contribute to the build-up of genetic variability and thus increase adaptability, fitness, competitiveness, and colonizing ability. It is clear that most hybridization events in

nature do not lead to the formation of a new species, but remarkably, the wheat group is equipped with a battery of molecular mechanisms that provide the potential for phenotypic novelty and for successful speciation to occur. Future work should enable better understanding of the role of specific genes and DNA sequences in speciation, the mechanisms that confer robustness of the genome to the shock of allopolyploidy and to the activation of TEs, and the mechanisms that enable the orchestration of chromosome division and the control of bivalent pairing during meiosis.

Altogether, the reported revolutionary and evolutionary genomic changes emphasize the dynamic plasticity of the wheat allopolyploid genome with regard to both structure and function. Presumably, these changes have improved the adaptability of the newly formed allopolyploids and facilitated their rapid colonization of new ecological niches. No wonder, therefore, that cultivated allopolyploid wheats exhibit a wider range of genetic flexibility than diploid wheats and could adapt themselves to a great variety of environments. In contrast to Stephens (1951), who had the insight that allopolyploidy might lead to new evolutionary opportunities, Stebbins (1971, 1980) stated that while polyploidy has been of great importance for the origin of species it has contributed little to progressive evolution. He assumed that polyploids evolve more slowly than their diploid relatives. Stebbins (1971, 1980) did not take into consideration that allopolyploidization triggers a burst of genomic alterations that are not feasible at the diploid level and that lead to new evolutionary opportunities. Allopolyploidy has proved to be a powerful evolutionary factor that has played a decisive role in the evolution of the wheat group.

Acknowledgments We would like to thank Beery Yaakov for his critical reading of the manuscript and Hakan Ozkan for providing part of the seed material for the project. This work was supported by grants from the Israel Science Foundation (grant # 142/08 to K.K, and grant # 616/09 to A.A.L).

References

Aaronsohn A (1910) Agricultural and botanical explorations in Palestine. Bull Plant Ind 180:1–63

Aaronsohn A, Schweinfurth G (1906) Die auffindung des wilden emmers (*Triticum dicoccum*) in Nordpalästina. Altneuland Monatsschrift für die irtschaft. Erschliessung Palästinas 7(8):213–220

Adams KL, Cronn R, Percifield R, Wendel JF (2003) Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proc Natl Acad Sci U S A 100(8):4649–4654

Aghaee-Sarbarzeh M, Dhaliwal HS, Harjit-Singh (2001) Suppression of rust resistance genes from distantly related species in *Triticum durum-Aegilops* amphiploids. In: Johnson R, Yahyaoui A, Wellings C, Saidi A, Ketata H (eds) Meeting the challenge of yellow rust in cereal crops. Proceedings of the First Regional Conference on Yellow Rust in the Central and West Asia and North Africa Region, Karaj, Iran. pp 8–14

Akhunova AR, Matniyazov RT, Liang H, Akhunov ED (2010) Homoeolog-specific transcriptional bias in allopolyploid wheat. BMC genomics 11:505

Avivi L (1976) The effect of genes controlling different degrees of homoeologous pairing on quadrivalent frequency in induced autotetraploid lines of *Triticum* longissimum. Can J Genet Cytol 18:357–364

Badaeva ED, Amosova AV, Samatadze TE, Zoshchuk SA, Shostak NG, Chikida NN, Zelenin AV, Raupp WJ, Friebe BR, Gill BS (2004) Genome differentiation in *Aegilops*. 4. Evolution of the U-genome cluster. Plant Syst Evol 246:45–76

Belyayev A, Raskina O, Korol A, Nevo E (2000) Coevolution of A and B genomes in allotetraploid *Triticum dicoccoides*. Genome 43(6):1021–1026

Belzile F, Beaulieu J, Jean M (2009) The allotetraploid Arabidopsis thaliana-Arabidopsis lyrata subsp petraea as an alternative model system for the study of polyploidy in plants. Mol Genet Genomics 281(4):421–435

Bento M, Pereira HS, Rocheta M, Gustafson P, Viegas W, Silva M (2008) Polyploidization as a retraction force in plant genome evolution: sequence rearrangements in *Triticale*. PLoS ONE 3:1402–1413

Blakeslee AF (1937) Redoublement du nombre de chromosomes chez les plantes par traitement chimique. Compt Rend Acad Sci Paris 205:476–479

Bottley A, Xia GM, Koebner RMD (2006) Homoeologous gene silencing in hexaploid wheat. Plant J 47(6):897–906

Boyko EV, Badaev NS, Maximov NG, Zelenin AV (1984) Does DNA content change in the course of triticale breeding. Cereal Res Commun 12(1–2):99–100

Boyko EV, Badaev NS, Maximov NG, Zelenin AV (1988) Regularities of genome formation and organization in cereals. I. DNA quantitative changes in the process of allopolyploidization. Genetika 24:89–97

Chague V, Just J, Mestiri I, Balzergue S, Tanguy AM, Huneau C, Huteau V, Belcram H, Coriton O, Jahier J, Chalhoub B (2010) Genome-wide gene expression changes in genetically stable synthetic and natural wheat allohexaploids. New phytol 187(4):1181–1194

Chaudhary B et al (2009) Reciprocal silencing, transcriptional bias and functional divergence of homeologs in polyploid cotton (gossypium). Genetics 182:503–517

Chantret N, Salse J, Sabot F, Rahman S, Bellec A, Laubin B, Dubois I, Dossat C, Sourdille P, Joudrier P, Gautier MF, Cattolico L, Beckert M, Aubourg S, Weissenbach J, Caboche M, Bernard M, Leroy P, Chalhoub B (2005) Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (*Triticum* and *Aegilops*). Plant Cell 17(4):1033–1045

Chapman V, Miller TE, Riley R (1976) Equivalence of the A genome of bread wheat and that of *Triticum urattu*. Genet Res 27:69–76

Dvorak J (1976) The relationship between the genome of *Triticum urattu* and the A and B genomes of *Triticum aestivum*. Can J Genet Cytol 18:371–377

Dvorak J (2009) Triticeae genome structure and evolution. In: Feuiller C, Muehlbauer GJ (eds) Genetics and genomics of the *Triticeae*, plant genetics and genomics: crops and models 7. Springer, Berlin. pp 685–711

Eilam T, Anikster Y, Millet E, Manisterski J, Feldman M (2008) Nuclear DNA amount and genome downsizing in natural and synthetic allopolyploids of the genera *Aegilops* and *Triticum*. Genome 51(8):616–627

Eilam T, Anikster Y, Millet E, Manisterski J, Sagi-Assif O, Feldman M (2010) Genome size in diploids, allopolyploids, and autopolyploids of mediterranean triticeae. doi:10.1155/2010/341380

Fahima T, Cheng JP, Peng JH, Nevo E, Korol A (2006) Asymmetry distribution of disease resistance genes and domestication synrome QTLs in tetraploid wheat genome. 8th International Congress of Plant Molecular Biology, Adelaide, Australia

Feldman M (1965a) Chromosome pairing between differential genomes in hybrids of tetraploid *Aegilops* species. Evolution 19:563–568

Feldman M (1965b) Fertility of interspecific F1 hybrids and hybrid derivatives involving tetraploid species of *Aegilops* Section *Pleionathera*. Evolution 19:556–562

Feldman M (1965c) Further evidence for natural hybridization between tetraploid tetraploid species of *Aegilops* Section Pleionathera. Evolution 19:162–174
Feldman M (2001) The origin of cultivated wheat. In: Bonjean A, Angus W (eds) The wheat book. Lavoisier Tech and Doc, Paris, pp 1–56
Feldman M, Levy AA (2005) Allopolyploidy—a shaping force in the evolution of wheat genomes. Cytogenet Genome Res 109(1–3):250–258
Feldman M, Levy AA (2009) Genome evolution in allopolyploid wheat—a revolutionary reprogramming followed by gradual changes. J Genet Genomics 36(9):511–518
Feldman M, Levy AA, Fahima T, Korol A (2012) Genomic asymmetry in allopolyploid plants - wheat as a model. J. Exp. Bot. (in press)
Feldman M, Liu B, Segal G, Abbo S, Levy AA, Vega JM (1997) Rapid elimination of low-copy DNA sequences in polyploid wheat: a possible mechanism for differentiation of homoeologous chromosomes. Genetics 147(3):1381–1387
Feldman M, Lupton FGH, Miller TE (1995) Wheats. In: Smartt J, Simmonds NW (eds) Evolution of crop plants, 2nd edn. Longman Scientific, London, pp 184–192
Flagel LE, Chen LP, Chaudhary B, Wendel JF (2009) Coordinated and fine-scale control of homoeologous gene expression in allotetraploid cotton. J Hered 100(4):487–490
Flagel LE, Wendel JF (2010) Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. New Phytol 186(1):184–193
Galili G, Feldman M (1984) Inter-genomic suppression of endosperm- protein genes in common wheat. Can J Genet Cytol 26:651–656
Galili G, Levy AA, Feldman M (1986) Gene-dosage compensation of endosperm proteins in hexaploid wheat *Triticum aestivum*. Proc Natl Acad Sci U S A 83:6524–6528
Griffiths S, Sharp R, Foote TN, Bertin I, Wanous M, Reader S, Colas I, Moore G (2006) Molecular characterization of Ph1 as a major chromosome pairing locus in polyploid wheat. Nature 439(7077):749–752
Gupta PK, Kulwal PL, Rustgi S (2005) Wheat cytogenetics in the genomics era and its relevance to breeding. Cytogenet Genome Res 109(1–3):315–327
Han FP, Fedak G, Guo WL, Liu B (2005) Rapid and repeatable elimination of a parental genome-specific DNA repeat (pGcIR-1a) in newly synthesized wheat allopolyploids. Genetics 170(3):1239–1245
Han FP, Fedak G, Ouellet T, Liu B (2003) Rapid genomic changes in interspecific and intergeneric hybrids and allopolyploids of *Triticeae*. Genome 46(4):716–723
Hart GH (1983a) Genetic and evolution of mulilocus isozymes in hexaploid wheat. In: Ratazzi MC, Scandalios JG, Whitt GS (eds) Isozymes: current topics in biological and medical research, vol 10., Genetics and Evolution Alan R. Liss., Inc., New York, pp 365–380
Hart GH (1983b) Hexaploid wheat (*Triticum aestivum L. em Thell.*). In: Tanksley SD, Orton TJ (eds) Isozymes in plant genetics and breeding, Part. B, Elsvier Science Publishers B.V., Amsterdam, pp 35–56
Hart GH (1987) Genetic and biochemical studies of enzymes. In: Heyne EG (ed) Wheat and wheat improvement, Second Ed., Amer. Soc. Agronomy, Madison, Wisconsin, USA
He P, Friebe BR, Gill BS, Zhou JM (2003) Allopolyploidy alters gene expression in the highly stable hexaploid wheat. Plant Mol Biol 52(2):401–414
Houchins K, ODell M, Flavell RB, Gustafson JP (1997) Cytosine methylation and nucleolar dominance in cereal hybrids. Mol Gen Genet 255(3):294–301
Huang S, Sirikhachornkit A, Su X, Faris J, Gill B, Haselkorn R, Gornicki P (2002) Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the Triticum/Aegilops complex and the evolutionary history of polyploid wheat. Proc Natl Acad Sci U S A 99(12):8133–8138
Kashkush K, Feldman M, Levy AA (2002) Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. Genetics 160(4):1651–1659
Kashkush K, Feldman M, Levy AA (2003) Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. Nature Genet 33(1):102–106

Kashkush K, Khasdan V (2007) Large-scale survey of cytosine methylation of retrotransposons, and the impact of readout transcription from LTRs on expression of adjacent rice genes. Genetics 177:1975–1985

Kenan-Eichler M, Leshkowitz D, Tal L, Noor E, Melamed-Bessudo C, Feldman M, Levy AA (2011) Wheat hand polyploidization results in deregulation of small RNAs. Genetics 188:263–272

Kerber ER, Green GJ (1980) Suppression of stem rust resistance in hexaploid wheat cv Canthach by chromosome 7DL. Can J Bot 58:1347–1350

Kihara H (1919) Über cytologische studien bei einigen getreidearten. I. Species-bastarde des weizens und weizenroggen-bastarde. Bot Mag 33:17–38

Kihara H (1924) Cytologische und genetische studien bei wichtigen getreidearten mit besonderer rücksicht ouf das verhalten der chromosomen und die sterilitat in den bastarden. Mem Cell Sci, Kyoto Imp University, B1: 1–200

Kihara H (1944) Discovery of the DD-analyser, one of the ancestors of *Triticum vulgare*. Agric Hortic 19:13–14

Kihara H (1954) Considerations on the evolution and distribution of *Aegilops* species based on the analyzer-method. Cytologia 19:336–357

Kihara H, Lilienfeld F (1949) A new synthesized 6x-wheat. In: Larsson GBaR (ed) Proceedings of Eighth International Congress of Genetics, Stockholm, Sweden, 1949. Hereditas (Suppl), pp 307–319

Kimber G, Sears ER (1987) Evolution in the genus *Triticum* and the origin of cultivated wheat. In: Heyne EG (ed) Wheat and wheat improvement. American Society of Agronomy, Madison, pp 154–164

Kislev ME (1980) *Triticum parvicoccum* sp. *nov*., the oldest naked wheat. Isr J Bot 28:95–107

Kraitshtein Z, Yaakov B, Khasdan V, Kashkush K (2010) Genetic and epigenetic dynamics of a retrotransposon after allopolyploidization of wheat. Genetics 186(3):U801–U889

Levy AA, Feldman M (2004) Genetic and epigenetic reprogramming of the wheat genome upon allopolyploidization. Biol J Linn Soc 82(4):607–613

Liu B, Segal G, Vega JM, Feldman M, Abbo S (1997) Isolation and characterization of chromosome-specific DNA sequences from a chromosome arm genomic library of common wheat. Plant J 11(5):959–965

Liu B, Vega JM, Feldman M (1998a) Rapid genomic changes in newly synthesized amphiploids of *Triticum* and *Aegilops*. II. Changes in low-copy coding DNA sequences. Genome 41(4):535–542

Liu B, Vega JM, Segal G, Abbo S, Rodova H, Feldman M (1998b) Rapid genomic changes in newly synthesized amphiploids of *Triticum* and *Aegilops*. I. Changes in low-copy noncoding DNA sequences. Genome 41(2):272–277

Lucas H, Moore G, Murphy G, Flavell RB (1992) Inverted repeats in the long-terminal repeats of the wheat retrotransposon wis 2–1A. Mol Bio Evol 9(4):716–728

Lukens LN, Pires JC, Leon E, Vogelzang R, Oslach L, Osborn T (2006) Patterns of sequence loss and cytosine methylation within a population of newly resynthesized *Brassica napus* allopolyploids. Plant Physiol 140(1):336–348

Lynch M, Force A (2000) The probability of duplicate gene preservation by subfunctionalization. Genetics 154:459–473

Ma XF, Fang P, Gustafson JP (2004) Polyploidization-induced genome variation in *triticale*. Genome 47(5):839–848

Ma XF, Gustafson JP (2005) Genome evolution of allopolyploids: a process of cytological and genetic diploidization. Cytogenet Genome Res 109(1–3):236–249

Ma XF, Gustafson JP (2006) Timing and rate of genome variation in triticale following allopolyploidization. Genome 49(8):950–958

Maan SS (1977) Fertility of amphiploids in *Triticinae*. J Heredity 68:87–94

Mac Key J (1954) Mutation breeding in polyploid cereals. Acta Agriculturae Scandinavica 4:549–557

Mac Key J (1958) Mutagenic response in *Triticum* at different lrvels of ploidy. In: Jenkins CB (ed) Proceedings 1st

Mac Key J (1966) Species relationship in *Triticum*. Proceedings 2nd International Wheat Genetics Symposium, Lund 1963, Hereditas Suppl. 2, pp 237–276

Madlung A, Masuelli RW, Watson B, Reynolds SH, Davison J, Comai L (2002) Remodeling of DNA methylation and phenotypic and transcriptional changes in synthetic *Arabidopsis* allotetraploids. Plant Physiol 129(2):733–746

Madlung A, Tyagi AP, Watson B, Jiang HM, Kagochi T, Doerge RW, Martienssen R, Comai L (2005) Genomic changes in synthetic *Arabidopsis* polyploids. Plant J 41(2):221–230

Maestra B, Naranjo T (1999) Structural chromosome differentiation between *Triticum timopheevii* and *T-turgidum* and *T-aestivum*. Theor Appl Genet 98(5):744–750

McFadden ES, Sears ER (1944) The artificial synthesis of *Triticum spelta*. Records Genet Soc Amer 13:26–27

McFadden ES, Sears ER (1946) The origin of *Triticum spelta* and its free-threshing hexaploid relatives. J Heredity 37(81–89):107–116

Mestiri I, Chague V, Tanguy AM, Huneau C, Huteau V, Belcram H, Coriton O, Chalhoub B, Jahier J (2010) Newly synthesized wheat allohexaploids display progenitor-dependent meiotic stability and aneuploidy but structural genomic additivity. New phytol 186(1):86–101

Mitra R, Bhatia C (1971) Isoenzymes and polyploidy. 1. Qualitative and quantitative isoenzyme studies in the *Triticinae*. Genet Res Camb 18:57–69

Mochida K, Kawaura K, Shimosaka E, Kawakami N, Shin-I T, Kohara Y, Yamazaki Y, Ogihara Y (2006) Tissue expression map of a large number of expressed sequence tags and its application to in silico screening of stress response genes in common wheat. Mol Genet Genomics 276(3):304–312

Morris R, Sears ER (1967) The cytogenetics of wheat and its relatives. In: Quisenberry KS, Reitz LP (eds) Wheat and wheat improvement. Madison, U.S.A., pp 19–87

Nigumann P, Redik K, Matlik K, Speek M (2002) Many human genes are transcribed from the antisense promoter of L1 retrotransposon. Genomics 79(5):628–634

Ohno S (1970) Evolution by gene duplication. Springer, Berlin

Okamoto M, Inomata N (1974) Possibility of 5B-like effect in diploid species. Wheat Inform Serv 38:15–16

Ozkan H, Levy AA, Feldman M (2001) Allopolyploidy-Induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. Plant Cell 13:1735–1747

Ozkan H, Tuna M, Arumuganathan K (2003) Nonadditive changes in genome size during allopolyploidization in the wheat (*Aegilops-Triticum*) group. J Hered 94(3):260–264

Parisod C, Alix K, Just J, Petit M, Sarilar V, Mhiri C, Ainouche M, Chalhoub B, Grandbastien MA (2010) Impact of transposable elements on the organization and function of allopolyploid genomes. New Phytol 186(1):37–45

Parisod C, Salmon A, Zerjal T, Tenaillon M, Grandbastien MA, Ainouche M (2009) Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in *Spartina*. New Phytol 184(4):1003–1015

Peng I, Ronin Y, Fahima T, Röder MS, Li Y, Nevo E, Korol A (2003a) Genomic distribution of domestication QTLs in wild emmer wheat, *Triticum dicoccoides*. In Proceedings 10th International Wheat Genetics Symposium, Paestum, Italy, pp 34–37

Peng JH, Ronin Y, Fahima T, Roder MS, Li YC, Nevo E, Korol A (2003b) Domestication quantitative trait loci in *Triticum dicoccoides*, the progenitor of wheat. Proc Natl Acad Sci U S A 100(5):2489–2494

Percival J (1921) The wheat plant. E.P. Dutton and Company, New York, pp 1–463

Petit M, Guidat C, Daniel J, Denis E, Montoriol E, Bui QT, Lim KY, Kovarik A, Leitch AR, Grandbastien MA, Mhiri C (2010) Mobilization of retrotransposons in synthetic allotetraploid *tobacco*. New Phytol 186(1):135–147

Pikaard CS (2000) The epigenetics of nucleolar dominance. Trends Genet 16(11):495–500

Prince VE, Pickett FB (2002) Splitting pairs: the diverging fates of duplicated genes. Nat Rev Genet 3:827–837

Puig M, Caceres M, Ruiz A (2004) Silencing of a gene adjacent to the breakpoint of a widespread Drosophila inversion by a transposon-induced antisense RNA. P Natl Acad Sci U S A 101(24):9013–9018

Pumphrey M, Bai J, Laudencia-Chingcuanco D, Anderson O, Gill BS (2009) Nonadditive expression of homoeologous genes is established upon polyploidization in hexaploid wheat. Genetics 181(3):1147–1157

Rapp RA, Udall JA, Wendel JF (2009) Genomic expression dominance in allopolyploids. Bmc Biol 7

Sabot F, Guyot R, Wicker T, Chantret N, Laubin B, Chalhoub B, Leroy P, Sourdille P, Bernard M (2005) Updating of transposable element annotations from large wheat genomic sequences reveals diverse activities and gene associations. Mol Genet Gen 274(2):119–130

Sakamura T (1918) Kurze mitteilung über die chromosomenzahalen und die verwandtschafts-verhältnisse der *Triticum* Arten. Bot Mag 32(1918):151–154

Salina EA, Numerova OM, Ozkan H, Feldman M (2004) Alterations in subtelomeric tandem repeats during early stages of allopolyploidy in wheat. Genome 47(5):860–867

Salmon A, Ainouche ML, Wendel JF (2005) Genetic and epigenetic consequences of recent hybridization and polyploidy in Spartina (Poaceae). Mol Ecol 14(4):1163–1175

Sax K (1927) Chromosome behavior in *Triticum* hybrids, Verhandlungen des V Int. Kongresses für Vererbungswissenchaft, Berlin, 2:1267–1284

Schulz A (1913) Die geschichte der kultivierten getreide. Nebert, Halle

Sears ER (1972) The nature of mutation in hexaploid wheat. Symp Biol Hung 12:73–82

Sears ER (1976) Genetic control of chromosome pairing in wheat. Annu Rev Genet 10:31–51

Shaked H, Kashkush K, Ozkan H, Feldman M, Levy AA (2001) Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. Plant Cell 13:1749–1759

Stebbins GLJ (1980) Polyploidy in plants: unsolved problems and prospect, in polyploidy—biological relevance. In: Lewis WH (ed) Plenum Press, New York

Stebbins GLJ (1971) Chromosomal evolution in higher plants. Addison-Wesley, New York

Stephens SG (1951) Possible significance of duplication in evolution. Adv Genet 4:247–265

Thompson DA, Desai MM, Murray AW (2006) Ploidy controls the success of mutators and nature of mutations during budding yeast evolution. Curr Biol: CB 16(16):1581–1590

Tirosh I, Reikhav S, Levy AA, Barkai N (2009) A yeast hybrid provides insight into the evolution of gene expression regulation. Science 324(5927):659–662

Van Slageren MW (1994) Wild wheats: a monograph of *Aegilops L.* and Amblyopyrum (Jaub. and Spach) Eig (*Poaceae*). Agricultural University, Wageningen, The Netherlands

Veitia RA, Bottani S, Birchler JA (2008) Cellular reactions to gene dosage imbalance: genomic, transcriptomic and proteomic effects. Trends Genet 24(8):390–397

Vicient CM, Jaaskelainen MJ, Kalendar R, Schulman AH (2001) Active retrotransposons are a common feature of grass genomes. Plant Physiol 125(3):1283–1292

Von Tschermak E, Bleier H (1926) Über fruchtbare *Aegilops*-weizenbastarde, der deutsch. Bot Ges 44:110–132

Waines JG (1976) A model for the origin of diploidizing mechanisms in polyploid species. Amer Natur 110:415–430

Wang JB, Xu YH, Zhong L, Wu XM, Fang XP (2009) Rapid alterations of gene expression and cytosine methylation in newly synthesized *Brassica napus* allopolyploids. Planta 229(3):471–483

Weissmann S, Feldman M, Gressel J (2005) Sequence evidence for sporadic intergeneric DNA introgression from wheat into a wild *Aegilops* species. Mol Biol Evol 22:2055–2062

Wicker T, Mayer KFX, Gundlach H, Martis M, Steuernagel B, Scholz U, Simkova H, Kubalakova M, Choulet F, Taudien S, Platzer M, Feuillet C, Fahima T, Budak H, Dolezel J, Keller B, Stein N (2011) Frequent gene movement and pseudogene evolution is common to the large and complex genomes of wheat, barley, and their relatives. Plant Cell 23(5):1706–1718

Yaakov B, Kashkush K (2011a) Massive alterations of the methylation patterns around DNA transposons in the first four generations of a newly formed wheat allohexaploid. Genome 54(1):42–49

Yaakov B, Kashkush K (2011b) Methylation, transcription, and rearrangements of transposable elements in synthetic allopolyploids. Int J Plant Genomics. doi:10.1155/2011/569826

Zhang ZC, Belcram H, Gornicki P, Charles M, Just J, Huneau C, Magdelenat G, Couloux A, Samain S, Gill BS, Rasmussen JB, Barbe V, Faris JD, Chalhoub B (2011) Duplication and partitioning in evolution and function of homoeologous Q loci governing domestication characters in polyploid wheat. P Natl Acad Sci U S A 108(46):18737–18742

Zhao N, Zhu B, Li M, Wang L, Xu L, Zhang H, Zheng S, Qi B, Han F, Liu B (2011) Extensive and heritable epigenetic remodeling and genetic stability accompany allohexaploidization of wheat. Genetics. doi:10.1534/genetics.111.127688

Zohary D, Feldman M (1962) Hybridization between amphiploids and the evolution of polyploids in the wheat (*Aegilops-Triticum*) group. Evolution 16:44–61

Chapter 8
Maize (*Zea Mays*) as a Model for Studying the Impact of Gene and Regulatory Sequence Loss Following Whole-Genome Duplication

James C. Schnable and Michael Freeling

Abstract Modern maize ($2n = 20$) is functionally diploid, and its chromosomes pair normally, forming 10 bivalents during meiosis. Sufficient genomic rearrangement has occurred that no two maize chromosomes are homologous across their entire lengths. Yet comparisons of genetic maps, duplicate gene sequences, and later genome assemblies revealed maize is descended from a polyploid ancestor which lived 5–12 million years ago. In the time since that polyploid ancestor lived 8,000–9,000 genes conserved at syntenic positions in other grass species have been reduced to single copy in maize while 4,000–5,000 genes are still retained as homologous gene pairs. The consequences of this polyploidy are continuing to resolve in modern maize accessions. With a wide range of data sets generated by an active research community, maize is an unparalleled model for the in silico study of the changes in genome structure, gene content, and gene regulation that a successful polyploidy brings about in a plant lineage.

8.1 Background on the Maize Polyploidy

Suspicion of a polyploid origin for maize first came from the prevalence of duplicate mutants identified at unlinked locations throughout the genome; these have been observed since the early days of maize genetics (Rhoades 1951). Over time, the evidence that the maize lineage was descended from an ancient polyploid grew to include the arrangement of duplicate genes and markers in similar orders on multiple maize chromosomes reported in a number of genetic maps developed

J. C. Schnable · M. Freeling (✉)
Department of Plant and Microbial Biology, University of California, Berkeley, USA
e-mail: freeling@uclink.berkeley.edu

P. S. Soltis and D. E. Soltis (eds.), *Polyploidy and Genome Evolution*,
DOI: 10.1007/978-3-642-31442-1_8,

during the 1980s (Goodman et al. 1980; Wendel et al. 1986; Helentjaris et al. 1988). With the availability of abundant sequence data for cloned duplicate genes, it became possible to definitively classify maize as an ancient polyploid (Gaut and Doebley 1997).

In the absence of diploid species showing higher genetic similarity to one maize subgenome or the other, it will be impossible to conclusively prove maize is not an ancient autopolyploid. The most closely related genus to *Zea*, *Tripsacum*, is descended from the same polyploid ancestor (Bomblies and Doebley 2005). The most closely related species with a sequenced genome is *Sorghum bicolor*, which shows an equal divergence from both subgenomes of maize (Swigoňová et al. 2004; Woodhouse et al. 2010). Almost no molecular data are available for the most closely related, and apparently diploid, genera to maize, *Elionurus* (13 sequences in GenBank) and *Coelorachis* (14 sequences in GenBank). The question of allotetraploid versus allo-polyploidy may never be conclusively answered for maize.

8.2 Timing of the Maize Polyploidy

Unlike other important crops species, such as bread wheat and potato, the maize polyploidy was not directly associated with domestication, but occurred millions of years earlier. Analysis of divergence between duplicate genes on opposite maize subgenomes—known as homeologs—has placed the split of the two progenitor genomes found within modern maize at ~12 million years before present (Swigoňová et al. 2004). The date of divergence between the two subgenomes of maize does not necessarily reflect how long ago the actual event of polyploidization occurred, as fertile allopolyploids can form between related species that have been evolving independently for millions of years. However, based on a single case of gene conversion it can be concluded that the two subgenomes of maize have shared a single nucleus for at least 5 million years (Swigoňová et al. 2004).

Recent evidence has allowed the phylogenetic placement of the maize tetraploidy to be narrowed to a discrete interval within the diversification of the Andropogoneae—the tribe of grasses within which maize is placed (Mathews et al. 2002). Phylogenetic trees of the homologous genes *zfl1* and *zfl2* show that these whole genome duplicates had already diverged in the common ancestor of the genus *Zea* (maize and teosinte) and the sister genus *Tripsacum* (Bomblies and Doebley 2005). Comparison of genetic maps in maize and sorghum prior to the publication of complete genome sequences for these species demonstrated that sorghum—and by extension other relatives within the "core" Andropogoneae—did not share the maize polyploidy (Wei et al. 2007). Following the publication of complete genome assembles for both maize (Schnable et al. 2009) and sorghum (Paterson et al. 2009), a study of thousands of maize homologous gene pairs found that both maize subgenomes appear equally diverged from sorghum (Woodhouse et al. 2010). Given these constraints, the phylogenetic placement of the maize tetraploidy can be inferred (Fig. 8.1).

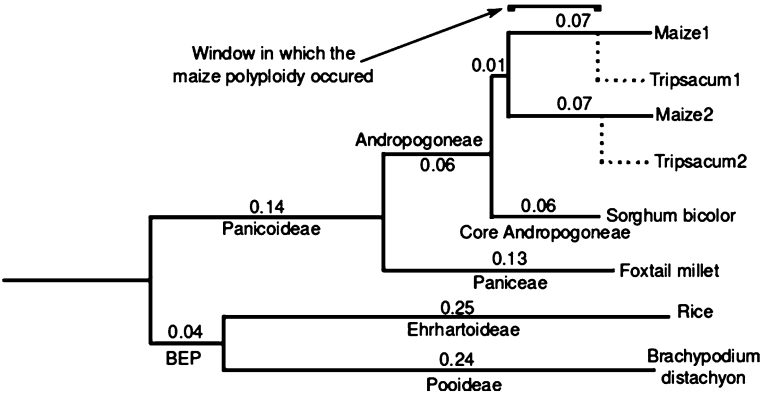

Fig. 8.1 A highly pruned tree of grass species. Branches are scaled by modal synonymous substitutions per site among all syntenic orthologs as measured by SynMap and QuotaAlign (Lyons et al. 2008b; Tang et al. 2011). The position of the branch between *Zea* and *Tripsacum* is only approximated, as the genome of *Tripsacum* has not yet been sequenced

8.3 Changes in Genome Arrangement Following Polyploidy in Maize

While all plant genomes sequenced to date have experienced at least one ancient polyploidization (Paterson et al. 2010), in the vast majority of cases, duplicate regions have been too heavily rearranged to reconstruct which regions can be grouped together as originating from the same parental genome. Such was not the case with maize. The polyploid ancestor of maize possessed 20 chromosomes, 2 equivalent to each of the 10 chromosomes of sorghum. Researchers could assign segments of the ten chromosomes of modern maize back to each of these ancestral chromosomes even before the completion of either the sorghum or maize genomes (Wei et al. 2007). With the publication of the completed maize and sorghum genomes, the same reconstruction can now be carried out using web-based tools by anyone with a fondness for puzzles (Fig. 8.2).

A cursory examination of Fig. 8.2 will reveal for any position in the sorghum genome there are two syntenic orthologous regions in maize, while for any position in the maize genome there is only one syntenic orthologous region in sorghum. In no case are the two regions orthologous to the same region in sorghum are present on the same maize chromosome. In 15 of the 20 inferred ancestral chromosomes, all segments are contained within a single chromosome of modern maize. In the remaining five cases, orthologous regions belonging to a single inferred ancestral maize chromosome are split between two—or in one case three—modern maize chromosomes. Fusion of chromosomes has often resulted from the insertion of one chromosome in between the arms of another—often linked with inversions of whole chromosome arms—as previously observed in the reduction of the *Brachypodium distachyon* genome to five chromosomes from the

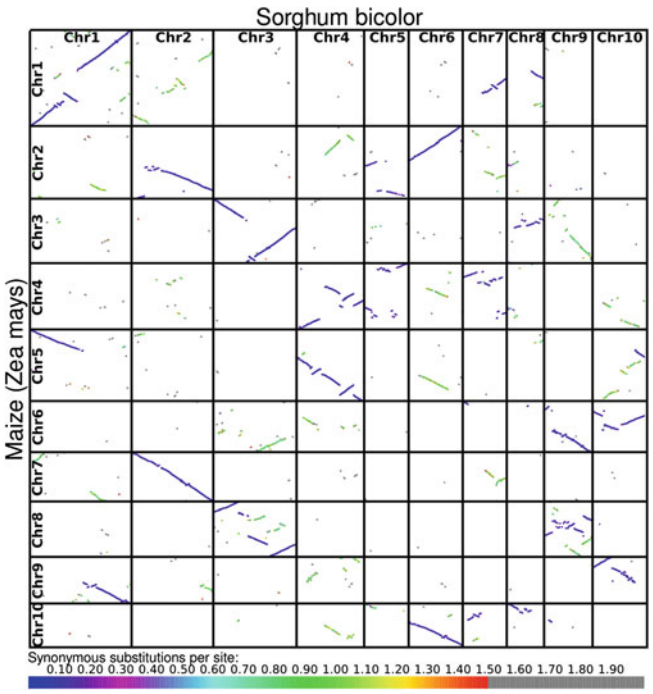

Fig. 8.2 A dotplot generated using SynMap (Lyons et al. 2008b) to compare the arrangement of orthologous regions of the sorghum (x-axis) and maize (y-axis) genome. Each *dot* represents a pair of homologous genes in maize and sorghum within a syntenic region, and each dot is color coded by synonymous substitutions per site between those two genes. In general, *purple lines* represent syntenic orthologous regions between the maize and sorghum genomes while*green diagonals* represent homologous syntenic regions from a more ancient whole-genome duplication shared by all grass species. An interactive version of this graphic can be regenerated using the following link: http://genomevolution.org/r/3vpl

inferred ancestral number of 12 (The International Brachypodium Initiative 2010). Researchers can search for differences across whole pairs of reconstructed maize ancestral chromosomes, and, by using differential gene loss as a mark to distinguish chromosome copies, even compare entire subgenomes (Schnable et al. 2011).

8.4 Ancient Gene Loss in Maize

Being able to compare pairs of complete ancestral chromosomes enabled a thorough investigation of an odd observation reported for *Arabidopsis* paleopolyploidy: the loss of genes was biased between homologous regions of the genome

("biased fractionation"), with one copy of each region retaining more total genes than the other (Thomas et al. 2006). Depending on the definition of a functional gene, between 4,000 and 5,000 homologous gene pairs are retained within the maize genome (Schnable et al. 2009; Woodhouse et al. 2010). Another 8,000–9,000 genes conserved at syntenic location in both rice and sorghum have fractionated back to a single copy in maize (Schnable et al. 2011). The same fractionation bias was observed across each of the ten pairs of ancestral maize chromosomes, with one chromosome copy retaining 70–90 % of the ancestral grass gene content and the other chromosome copy retaining only 40–60 %. By grouping high gene loss ancestral chromosomes together into one parental subgenome and low gene loss ancestral chromosomes into the other parental subgenome and using the wealth of maize RNA-seq data being produced by other research projects, it was shown that genes on the subgenome with low gene loss, referred to as maize1, tend to be expressed at higher levels than their duplicates on the subgenome with high gene loss, or maize2 (Schnable et al. 2011). This inequality of expression between parental subgenomes has previously been reported in allopolyploid cotton—among other species—where it was termed "genomic dominance" (Flagel and Wendel 2010).

The correlation between bias in gene expression levels and bias in gene loss rates suggests a simple explanation for the bias in gene loss rates observed in all ancient polyploid species studied to date (Sankoff et al. 2010). The deletion of the gene copy that contributed the majority of total gene pair expression should be more likely to have a functional impact than the deletion of the gene copy which contributes the minority of total gene pair expression. Therefore, even if the base rate of gene deletion is equal on both subgenomes, null alleles of more highly expressed gene copies would be more likely to be purged from the population by purifying selection. Over time, deletion of redundant gene copies would tend to cluster on the less expressed subgenome. In support of this hypothesis, it was noted that known maize morphological mutants are significantly more likely to result from the disruption of genes on the high-expression subgenome than expected given the overall distribution of expressed genes between the subgenomes (Schnable and Freeling 2011).

8.5 Ongoing Gene Loss in Maize

In addition to ancient fixed patterns of gene loss, fractionation continues within maize today, with genes present in some maize inbreds lost in other accessions. Concurrent with the publication of the maize genome, the first report emerged describing a high incidence of presence–absence variation (PAV) between different maize inbreds. A comparison of two of the most-studied maize lines, B73 and Mo17 (the former being the line used to generate the reference genome), determined that thousands of sequences were present in the former but entirely absent from the latter, including at least 180 single-copy genes (Springer et al. 2009).

A follow-up study which examined 33 inbreds and accessions of wild teosinte identified 3,410 high-confidence genes in the genome of B73 which had been lost in one or more of these lines—a total that represents more than 10 % of all high-confidence genes present in the maize genome (Swanson-Wagner et al. 2010). While most PAV in maize was initially observed using comparative genomic hybridization to microarrays, the same variation in gene content was observed in a recent study that resequenced six inbreds using Illumina short read technology (Lai et al. 2010).

While many genes which exhibit PAVs between B73 and other maize inbreds possess no syntenic orthologs in other grass species and may have recently inserted into their current locations, 4–6 % of maize genes with syntenic orthologs in both sorghum and rice—indicating these genes have been functionally conserved for at least 50 million years—are also involved in PAV between diverse maize lines (Schnable et al. 2011). Among genes where both homologous are still retained in B73, genes on the 'high-gene-loss, low-expression' subgenome (maize2) are more likely to be involved in PAV (Swanson-Wagner et al. 2010; Schnable et al. 2011).

8.6 Sequence Deletion in Maize

The loss of genes is the result of the deletion of sequence from the genome. Given the extensive observation of chromosomal mispairing, rearrangements, and partial or complete aneuploidy in synthetic and recent natural polyploids, it might be expected that much duplicate gene loss is the result of large deletions which remove whole chromosomal segments from one subgenome. The fact that only a portion of the genomes of ancient polyploid species such as rice—65.7 % of the genes covered by duplicated blocks (Yu et al. 2005)—and *Arabidopsis*—89 % of genes covered (Bowers et al. 2003)—would seem consistent with this expectation.

However, the pattern of gene loss in maize is not consistent with large deletions following whole-genome duplication. A search for regions of the sorghum genome represented by only one syntenic orthologous region within maize revealed only one putative deletion of $\geq$30 genes (Schnable et al. 2012). A comparison of the patterns of gene loss observed in duplicate regions of the maize genome to simulations assuming different lengths of sequence deletion found that the pattern observed is consistent with >85 % of deletions removing only a single gene, or a portion there of, with the remainder of deletions removing two, or occasionally three, adjacent genes (Woodhouse et al. 2010). The present genome of maize is littered with the partially deleted fragments of homologous genes (Fig. 8.3a).

The prevalence of short direct repeats flanking deletions within these fragmentary genes indicates the loss of genes in polyploids is a result of intrastrand non-homologous recombination: short direct repeats found by chance throughout the genome pair with each other, splicing out any intervening sequence, and leaving a single copy of the repeat sequence (Woodhouse et al. 2010).

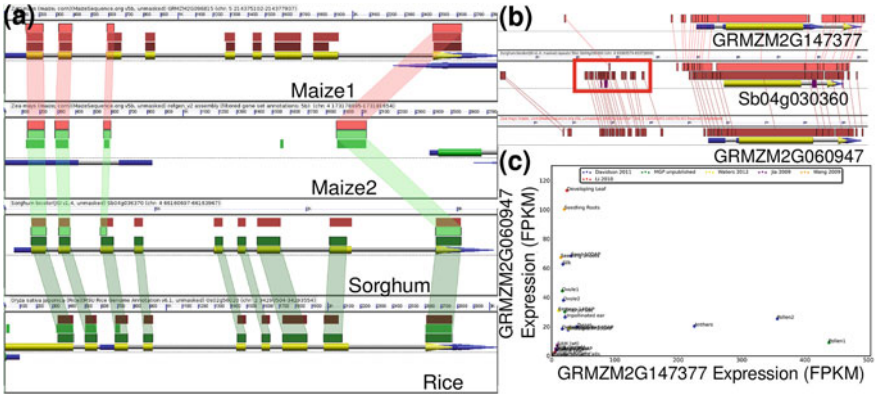

Fig. 8.3 **a** GEvo panel (Lyons et al. 2008a) illustrating an example of a gene conserved in rice, sorghum, and one of two maize subgenomes (maize1), while on the second maize subgenome (maize 2) only the 5' and 3' ends of the gene remain conserved. The colored boxes mark regions of sequence similarity, as determined by blastn, between the orthologous genomic regions in this figure. An interactive version of the same panel can be regenerated at the following link: http://genomevolution.org/r/3vbu. **b** A second GEvo panel comparing a single sorghum gene to its two coorthologs in maize. *Purple rectangles* represent functionally constrained noncoding sequences identified by the comparison of this gene to its syntenic ortholog in rice. The *red box* highlights a cluster of conserved noncoding sequences conserved upstream of one maize gene but lost from the promoter of the other. **c** A comparison of the relative expression of these two maize genes in a wide range of tissues using RNA-seq data produced by a number of maize research groups (Wang et al. 2009; Jia et al. 2009; Li et al. 2010; Davidson et al. 2011; Waters et al. 2011)

Fractionation and sequence deletion are not confined to only the protein-coding regions of genes. Sequence deletion, presumably by the same mechanism of intra-strand non-homologous recombination, can also remove conserved noncoding regulatory sequences from one of two duplicate copies of a maize gene (Fig. 8.3b). These genes often show different patterns of tissue-specific expression in existing maize RNA-seq data sets, allowing researchers to develop testable hypotheses about the function of specific regulatory sequences (Fig. 8.3c) (Freeling et al. 2012).

8.7 Future Prospects

The mechanism responsible for the unequal expression of duplicate genes from different parents in polyploid species remains unknown. With an extensive collection of well-characterized mutants, including knockouts of a variety of epigenetic mechanisms, maize is an excellent system for the investigation of this inexplicable behavior. If the whole-genome duplication in maize was indeed the result of allo-polyploidy, the identification of a wild species more closely related to one maize subgenome than the other remains an exciting possibility. The wealth of RNA-seq data being generated in maize for unrelated purposes creates an opportunity to

investigate how the expression patterns of homologous gene pairs in maize have diverged in silico. Soon, it may also be possible to study how the functions of duplicate genes have diverged in the 5–12 million years since the maize polyploidy in silico, as high-resolution genome-wide associate studies begin to identify the loci responsible for variation in a wide range of maize phenotypes. In addition to a higher likelihood of being lost entirely, has the reduced importance of less expressed genes on the non-dominant subgenome also given these genes greater freedom for innovation, even if their new role comes at the express of their ancestral function? We predict that higher levels of neofunctionalization will be observed on the non-dominant subgenome, but this hypothesis remains untested.

Acknowledgments We thank Vincent Li, a high school intern in the Freeling lab from Project SEED, for identifying the gene fragment shown in Fig. 8.3 and Addie M. Thompson for critical reading of an early version of this text. Funding provided by NSF Plant Genome Research Program grant 0701871 to MF and a Chang-Lin Tien Graduate Fellowship to JCS.

References

Bomblies K, Doebley JF (2005) Molecular evolution of FLORICAULA/LEAFY orthologs in the *Andropogoneae (Poaceae)*. Mol Biol Evol 22:1082–1094

Bowers JE et al (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 422:433–438

Davidson RM et al (2011) Utility of RNA sequencing for analysis of maize reproductive transcriptomes. Plant Genome 4:191–203

Flagel LE, Wendel JF (2010) Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. New Phytol 186:184–193

Freeling M et al (2012) Fractionation mutagenesis and similar consequences of mechanisms removing dispensable or less-expressed DNA in plants. Curr Opin Plant Biol, Advance Online Publication

Gaut BS, Doebley JF (1997) DNA sequence evidence for the segmental allotetraploid origin of maize. Proc Natl Acad Sci USA 94:6809–6814

Goodman MM et al (1980) Linkage relationships of 19 enzyme Loci in maize. Genetics 96:697–710

Helentjaris T et al (1988) Identification of the genomic locations of duplicate nucleotide sequences in maize by analysis of restriction fragment length polymorphisms. Genetics 118:353–363

Jia Y et al (2009) Loss of RNA–dependent RNA polymerase 2 (RDR2) function causes widespread and unexpected changes in the expression of transposons, genes, and 24-nt small RNAs. PLoS Genet 5:e1000737

Lai J et al (2010) Genome-wide patterns of genetic variation among elite maize inbred lines. Nat Genet 42:1027–1030

Li P et al (2010) The developmental dynamics of the maize leaf transcriptome. Nat Genet 42:1060–1067

Lyons E et al (2008a) Finding and comparing syntenic regions among *Arabidopsis* and the outgroups papaya, poplar, and grape: CoGe with *Rosids*. Plant Physiol 148:1772–1781

Lyons E et al (2008b) The value of non-model genomes and an example using SynMap within CoGe to dissect the hexaploidy that predates the Rosids. Trop Plant Biol 1:181–190

Mathews S et al (2002) Phylogeny of *Andropogoneae* inferred from phytochrome B, GBSSI, and ndhF. Int J Plant Sci 163:441–450

Paterson AH et al (2010) Insights from the comparison of plant genome sequences. Annu Rev Plant Biol 61:349–372

Paterson AH et al (2009) The sorghum bicolor genome and the diversification of grasses. Nature 457:551–556
Rhoades MM (1951) Duplicate genes in maize. Am Nat 85:105–110
Sankoff D et al (2010) The collapse of gene complement following whole genome duplication. BMC Genomics 11:313
Schnable JC et al (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proc Natl Acad Sci USA 108:4069–4074
Schnable JC et al (2012) Genome-wide analysis of syntenic gene deletion in the grasses. Genome Biol Evol 4:265–277
Schnable JC, Freeling M (2011) Genes identified by visible mutant phenotypes show increased bias toward one of two subgenomes of maize. PLoS ONE 6:e17855
Schnable PS et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115
Springer NM et al (2009) Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. PLoS Genet 5:e1000734
Swanson-Wagner RA et al (2010) Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. Genome Res 20:1689–1699
Swigoňová Z et al (2004) Close split of sorghum and maize genome progenitors. Genome Res 14:1916–1923
Tang H et al (2011) Screening synteny blocks in pairwise genome comparisons through integer programming. BMC Bioinf 12:102
The International Brachypodium Initiative (2010) Genome sequencing and analysis of the model grass Brachypodium distachyon. Nature 463:763–768
Thomas BC et al (2006) Following tetraploidy in an Arabidopsis ancestor, genes were removed preferentially from one homologue leaving clusters enriched in dose-sensitive genes. Genome Res 16:934–946
Wang X et al (2009) Genome-wide and organ-specific landscapes of epigenetic modifications and their relationships to mRNA and small RNA transcriptomes in maize. Plant Cell 21:1053–1069
Waters AJ et al (2011) Parent-of-origin effects on gene expression and DNA methylation in the maize endosperm. Plant Cell 23:4221–4233
Wei F et al (2007) Physical and genetic structure of the maize genome reflects its complex evolutionary history. PLoS Genet 3:e123
Wendel JF et al (1986) Duplicated chromosome segments in maize (Zea mays L.): further evidence from hexokinase isozymes. Theoret Appl Genet 72:178–185
Woodhouse MR et al (2010) Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homeologs. PLoS Biol 8:e1000409
Yu J et al (2005) The Genomes of *Oryza sativa*: a history of duplications. PLoS Biol 3:e38

Chapter 9
Polyploidy in Legumes

Jeff J. Doyle

Abstract Legumes are the third largest family of flowering plants, with over 700 genera and more than 19,000 species. Genomic evidence has shown that a whole-genome duplication (WGD) occurred shortly after the origin of the family, in an ancestor that gave rise to the papilionoids, the clade that comprises 65 % of the genera and 71 % of the species, including nearly all of the economically important crop legumes. This polyploidy event may have been associated with the origin of nitrogen-fixing symbiosis (nodulation) in the papilionoids. Nodulation most likely evolved independently in other legumes outside the papilionoids, hence there appears to be no requirement for polyploidy in the evolution of this important symbiosis. More recent polyploidy, as inferred from chromosome counts, occurs in approximately a quarter of all legume genera for which data are available. In most cases, polyploidy is confined to individual genera, species within genera, or cytotypes within species. An exception is the core clade of the genistoid legumes, a major papilionoid group that includes lupines (*Lupinus*). This group is probably fundamentally polyploid and also has a propensity for further polyploidy and aneuploidy in many of its genera. The frequency of polyploidy varies considerably among clades of the family, being most common (outside the genistoids) in the largely temperate, herbaceous Hologalegina (including pea and clover), and low in woody tropical groups such as the caesalpinioids.

J. J. Doyle (✉)
Department of Plant Biology, Cornell University, Ithaca, NY 14850, USA
e-mail: jjd5@cornell.edu

P. S. Soltis and D. E. Soltis (eds.), *Polyploidy and Genome Evolution*,
DOI: 10.1007/978-3-642-31442-1_9,

9.1 Introduction

The legumes (Leguminosae or, less preferably, Fabaceae, according to Lewis et al. 2005) are the third largest family of flowering plants, and are tremendously diverse ecologically, morphologically, chemically, and cytologically (Doyle and Luckow 2003; Lewis et al. 2005). Not surprisingly, the family is also cytologically diverse. As in other families, polyploidy is implicated as a major force at all levels of legume evolution, from the early stages of radiation in the family to the origin and recent diversification of modern genera, such as *Glycine* (soybean and allies) and species within genera, such as the *Medicago sativa* complex (alfalfa and allies).

After summarizing progress in understanding the phylogeny of the family, this review will discuss the role of paleopolyploidy during the early stages of the radiation of the entire family and the possible connections with nodulation. The occurrence of polyploidy in each of the major clades of the family will then be reviewed.

9.2 A Brief Overview of Legume Phylogeny

Along with Polygalaceae, Surianaceae, and Quillajaceae, Leguminosae form the order Fabales, one of eight-orders in the Fabidae clade of rosid eudicots (Wang et al. 2009). Bello et al. (2009) suggested that the Fabales are the product of a rapid radiation, with legumes probably sister to Surianaceae plus Quillajaceae.

The Leguminosae has been the focus of considerable phylogenetic study, culminating in solid, chloroplast based, working hypotheses of generic relationships (Fig. 9.1), notably those of Wojciechowski et al. (2004) and Bruneau et al. (2008). The older classification of the family into three subfamilies, Caesalpinioideae, Mimosoideae, and Papilionoideae (sometimes treated as separate families), is not supported by molecular phylogenetic studies, in that although the Mimosoideae (mimosas, acacias) and Papilionoideae (pea, bean, soybean, etc.) are monophyletic, the former is embedded in one clade of a paraphyletic caesalpinioid grade. Relationships at the base of the family are uncertain and differ among the studies of Wojciechowski et al. (2004), which focused most heavily on Papilionoideae, and Bruneau et al. (2008), which emphasized caesalpinioids and included few papilionoids. However, both studies identified caesalpinioids as the earliest-diverging lineages, including such taxa as the tribe Cercideae, which includes *Cercis* (the redbud or Judas tree) and the large genus, *Bauhinia* (orchid tree). Relationships at the bases of the two monophyletic subfamilies are also uncertain.

Fossil evidence places the origin of the family in the Paleocene, around 60 million years ago (MYA; see Lavin et al. 2005 for discussion; see also Bell et al. 2010). Divergence times of all major groups within the family have been estimated from fossil-calibrated molecular data (Lavin et al. 2005; Bruneau et al. 2008) and suggest rapid diversification of many clades, such that within

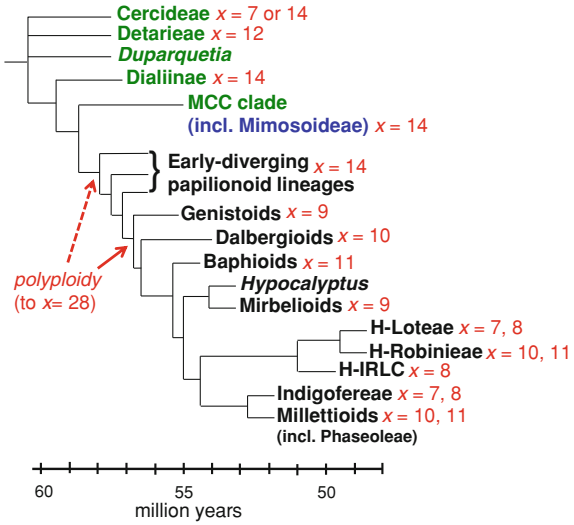

Fig. 9.1 Phylogeny of legumes. Caesalpinioids are shown in *green*, papilionoids in *black*. Relationships among caesalpinoid lineages are shown as unresolved due to conflict among published studies. Papilionoid taxa marked "H" are members of the Hologalegina. Divergence dates for the origins of major clades are from Lavin et al. (2005). Base chromosome numbers are given for groups with published counts. Minimum (*solid arrow*) and maximum (*dashed arrow*) dates for the papilionoid polyploidy event are shown

10 million years of divergence from a common ancestor, all of the major lineages in the family had evolved, including the two monophyletic subfamilies and all of the major clades within the Papilionoideae.

9.3 Polyploidy and the Early Diversification of Legumes

To what degree has polyploidy shaped the radiation of legumes? Given the growing understanding that polyploidy can drive phenotypic diversification (e.g., Freeling et al. 2006) and has played a role in the preservation of lineages during periods of extinction (Fawcett et al. 2009), it might be expected that polyploidy would be an important feature of evolution in a family that is "successful" as judged by its sheer size and ecological dominance in some tropical biomes (e.g., rain forests, woody savannas, and dry forests), and in which a significant adaptive novelty—the symbiotic association with nitrogen-fixing soil bacteria, termed nodulation—has arisen (e.g., Doyle 2011).

9.3.1 Chromosome Number Evidence for Polypoidy in Legumes

Until the advent of genomic data, chromosome number was the prime source of information available for inferring the existence of polyploidy. Goldblatt's (1981) review of the distribution of chromosome numbers in Leguminosae, published in *Advances in Legume Systematics, Part* 2, remains the most comprehensive treatment of chromosomal variation in the family, and includes hypotheses concerning the base numbers and ploidy levels of its constituent subfamilies and tribes. The information from which his summary was drawn was included in the descriptions of genera in *Advances in Legume Systematics, Part* 1 (Polhill and Raven 1981), known to researchers in the family as the legume "bible." The taxonomic treatments provided by this key resource were recently updated in *Legumes of the World* (Lewis et al. 2005), taking into account the rapid progress in legume phylogenetics. The phylogenetic studies that have revolutionized our understanding of relationships within the family also provide a new phylogenetic context for understanding chromosome number evolution that was not available previously, but unfortunately the otherwise excellent *Legumes of the World* does not include any cytological information.

The key contribution of objective phylogenetic data to our understanding of cytological evolution in the family is the confirmation of caesalpinioid legumes as a grade rather than as a natural subfamily. Chromosome numbers for the major clades that comprise the caesalpinioid grade are relatively constant, principally $2n = 24$–28 (Fig. 9.1). Standing out from these higher chromosome numbers are *Chamaecrista* and *Cercis*. The large, mostly, pantropical genus, *Chamaecrista*, is cytologically complex, with $2n = 14$, 16, and 28. Goldblatt (1981) considered its lower numbers to be the products of aneuploid reduction, and this hypothesis has been recently supported (Torres et al. 2011, see below). Phylogenetic studies (Wojciechowski et al. 2004; Bruneau et al. 2008) now nest *Chamaecrista* and other Cassieae s.s. within the Mimosoideae-Caesalpinioideae-Cassieae (MCC) clade ($2n = 28$), supporting this hypothesis.

Cercis is a small genus (10 species) with disjunct worldwide distribution and $2n = 14$. It is a member of the Cercideae, all other members of which are $2n = 28$, including the large pantropical genus, *Bauhinia* s.l. (ca. 250 species). Phylogenetic studies show that *Cercis* is sister to the remaining genera (Bruneau et al. 2008), which may be consistent with Goldblatt's (1981) conclusion that it is diploid and the remainder of the tribe is fundamentally polyploid. This is of some importance given the relatively early divergence of Cercideae in some phylogenies. In the *rbcL* phylogeny of Kajita et al. (2001) the tribe was sister to the remainder of the family, though with relatively weak support, and this topology also appears in the phylogenetic summary of Lewis et al. (2005). The chloroplast *matK* tree of Wojciechowski et al. (2004), which emphasized Papilionoideae, placed Cercideae, and Detarieae (mainly $2n = 24$) together as the first-diverging legume lineage. In these topologies, it is possible that, as Goldblatt (1981) suggested, the legumes are fundamentally $x = n = 7$, with subsequent independent

chromosomal increase both within Cercideae and in the ancestor of all remaining legumes.

In contrast, the concatenated chloroplast *matK/trnK* + *trnL-F* tree of Bruneau et al. (2008) placed Detarieae as the first branch in the legume phylogeny, sister to a trichotomy composed of Cercideae, *Duparquetia*, and the remainder of the family. In this phylogeny, then, the base number for the family would be $x = n = 12$, with $2n = 14$ in *Cercis* representing a reduction. Interestingly, the genome size of *Cercis canadensis* is comparable to measurements from the several species of *Bauhinia* in the Kew C-value database (http://data.kew.org/cvalues/; Leitch and Doyle, unpublished data), supporting this reduction hypothesis.

Even a high base number for early diverging lineages, as suggested by the Bruneau et al. (2008) topology, would not definitively suggest polyploidy at the base of the family, given what is known of chromosome numbers from other Fabales. No information is available for *Quillaja* in the Index of Plant Chromosome Numbers (IPCN; http://www.tropicos.org/Project/IPCN), but Surianaceae is represented by a single species of *Stylobasium,* with a number of $2n = 30$, suggesting that the common ancestor of legumes and Surianaceae could have had a high chromosome number.

Patterns of chromosomal evolution among major groups of legumes are complex even outside of the earliest branching. The bulk of the family belongs to two sister clades: the MCC clade and the Papilionoideae. The two tribes that comprise the MCC clade along with Mimosoideae (Caesalpinieae and Cassieae s.s.) are both diploid based on $x = 14$. Given the presence of taxa with $2n = 28$ in the grade at the base of Papilionoideae, it is likely that the common ancestor of that group and the MCC was diploid based on $x = 14$ as well. The majority of papilionoids, however, have lower base chromosome numbers, ranging from $x = 7$–11, depending on the tribe. These presumably represent reductions in chromosome number, as discussed below; they certainly give no evidence for polyploidy.

9.3.2 Genetic and Genomic Evidence for Polyploidy in the Early Evolution of Legumes

Genomic studies, starting with linkage maps and continuing through studies of expressed sequence tags (ESTs) and genome sequencing, have revolutionized understanding of polyploidy in seed plants. Although it was long known that diploidization can erase chromosomal evidence of polyploidy over time, it is now clear that plant genomes comprise nested sets of WGD. The common ancestor of all seed plants underwent a polyploid duplication, with a later WGD in the ancestor of all angiosperms (Jiao et al. 2011) and numerous lineage-specific duplications in various groups of flowering plants (Soltis et al. 2009).

It has been known for some time that cryptic polyploidy occurs in legume genomes. For example, Shoemaker et al. (1996) used linkage map information to

hypothesize that the soybean genome shows evidence of a more ancient duplication than the one that is responsible for its high chromosome number relative to allied phaseoloid genera (millettioid clade, Fig. 9.1). In 2004, two different groups mined the extensive EST collections of soybean and the diploid model legume, *Medicago truncatula* ($2n = 14$; a member of the Hologalegina IRLC clade), to search for the genomic signature of ancient polyploidy events (Blanc and Wolfe 2004; Schleueter et al. 2004). This signature is produced when all genes in the genome are duplicated by autopolyploidy or when homoeologous loci are brought together by allopolyploidy. It is observed by plotting the frequency distribution of pairwise K_s (synonymous substitutions per synonymous site—a stand-in for time) values for hundreds to thousands of paralogous gene pairs. Simple gene duplication is an ongoing phenomenon in all eukaryotes, but most duplicates are purged from the genome rapidly, producing a characteristic distribution with many recent duplicates with low K_s values and relatively few older pairs with high K_s (Lynch and Conery 2003). Polyploid duplications appear as additional components ("peaks") against this background; the mode of such a K_s peak is taken as an estimate for the age of the polyploid, though it is generally an overestimate of that age (Doyle and Egan 2010).

Both Blanc and Wolfe (2004) and Schleueter et al. (2004) identified two K_s peaks in soybean, as expected (Fig. 9.2); both reported similar K_s modes for these peaks but because they used different substitution rates for plant nuclear genes, this led to different estimates of the age of polyploidy (or homoeologue divergence). The Schleueter et al. (2004) estimates are more in keeping with divergence dates for papilionoid legume taxa (Lavin et al. 2005) and are preferred for that reason (Shoemaker et al. 2006); in addition, the rate used by Schleueter et al. (2004) is much closer to the rate recently estimated for *Arabidopsis* (Ossowski et al. 2010).

Of great interest was the finding, by both groups, of two K_s peaks in the *M. truncatula* EST collection (Fig. 9.2). The younger of the two peaks is recent enough that if due to polyploidy, it would most likely have left chromosomal evidence, and has yet to be explained (Young et al. 2011). The older *Medicago* peak, on the other hand, was estimated by Schleueter et al. (2004) to be around 54.6 MYA, very close to the 54 MY age estimated by Lavin et al. (2005) for the divergence of the soybean (millettioid) and *Medicago* (Hologalegina) lineages, and also similar to the age estimated for the older soybean peak (41.6 MYA). This raised the possibility that the two species shared an ancient WGD.

This hypothesis was tested by Pfeil et al. (2005) using a phylogenomic approach with 39 gene pairs chosen from among those used by Schleueter et al. (2004) to identify the K_s peak in soybean. Topologies of gene trees overwhelmingly favored the hypothesis that their common ancestor was polyploid. Comparisons of linkage relationships between *Medicago* and the other legume model species, *Lotus japonicus* (in the Hologalegina Loteae clade, Fig. 9.1), provided further support for this hypothesis, and also showed that the duplication was not found in poplar. Using the Lavin et al. (2005) date for the divergence of the millettioid (*Glycine*) and Hologalegina (*Lotus, Medicago*) clades, the WGD event had taken place by around 54 MYA. Thus, the common ancestor of the two major

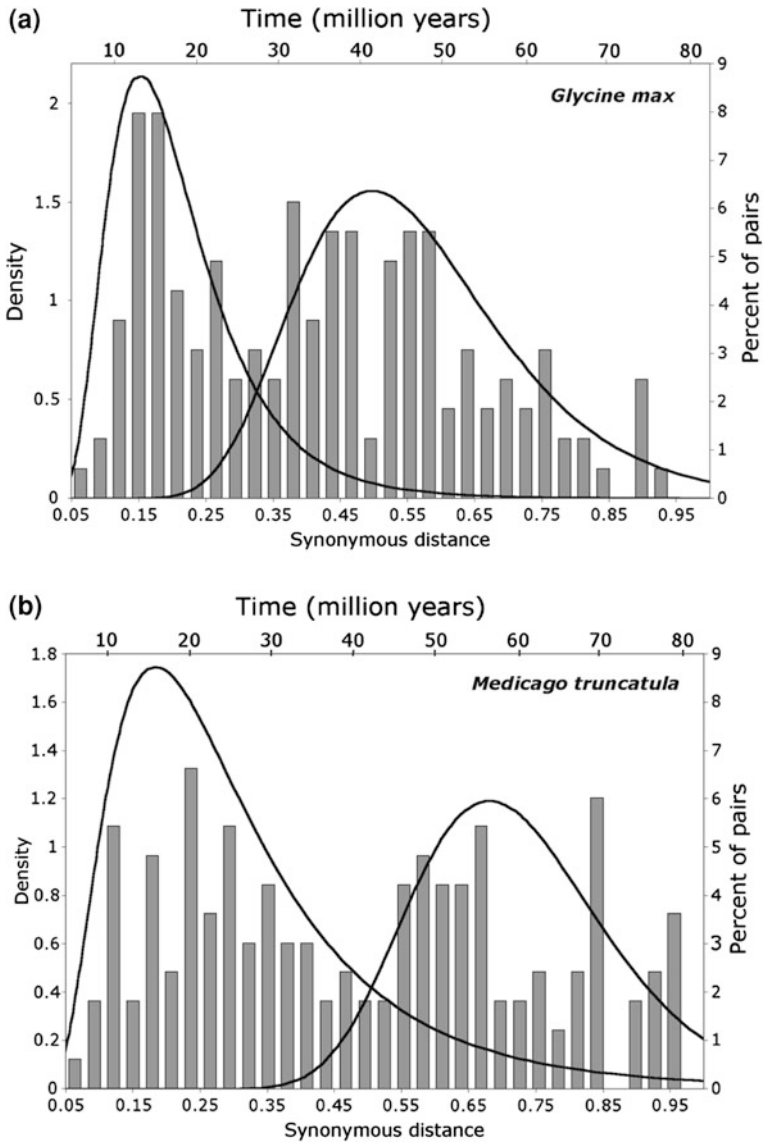

Fig. 9.2 Evidence for polyploidy in the genomes of: **a** *Glycine max* and **b** *Medicago truncatula*. The graphs plot the number of pairs of paralogous sequences ("density" or "percent of pairs") versus binned K_s ("synonymous distances") classes. Pairs with very low divergence (produced by ongoing recent duplications) were not plotted. Curves were fit to the binned divergence data and are interpreted as groups of genes duplicated simultaneously in large-scale genomic events such as polyploidy; modes of peaks provide a maximum age for allopolyploid events (Doyle and Egan 2010). Divergence time was estimated from synonymous distances using standard clock methods; note the different estimated ages (modes of curves) for the older event in the two species. Data are from Schleueter et al. (2004), who used expressed sequence tags (ESTs). Figure courtesy of Jessica Schlueter (UNC-Charlotte)

sister clades that comprise nearly 9,000 species—around 45 % of all legumes and 63 % of papilionoids—was polyploid.

Subsequently, Bertioli et al. (2009) studied *Arachis* (peanut) and showed that its genome also shows evidence of the >54 MYA event, indicating that the large dalbergioid clade is also fundamentally polyploid. Unpublished information reported by McClean (personal communication) at the 2009 International Conference of Legume Genetics and Genomics showed that *Lupinus*, and hence the genistoid clade, also shares this WGD. Thus, all of the lineages of the main radiation of the papilionoids share a polyploid ancestor.

The possibility that all legumes share this polyploidy event remained tenable until transcriptomic data from the caesalpinioid genus, *Chamaecrista*, became available (Singer et al. 2009). Cannon et al. (2010) produced and analyzed over 1,200 gene phylogenies from these data, which overwhelmingly supported the conclusion that the *Chamaecrista* genome shows no evidence of any polyploidy event subsequent to the prerosid triplication, and notably lacks the WGD found in core papilionoids (all but the early diverging lineages in Fig. 9.1). *Chamaecrista* belongs to the MCC clade, which is sister to the papilionoid clade. Therefore, the absence of the WGD in *Chamaecrista* indicates that the common ancestor of the MCC clade and papilionoids was not polyploid, placing the WGD within papilionoids (Fig. 9.1). Whether the WGD took place in the papilionoid common ancestor is still unknown, because the lineages that comprise the paraphyletic grade lacking the putative molecular synapomorphy for the major papilionoid radiation (chloroplast genome 50 kb inversion) remain to be sampled.

Thus, it is possible that this core papilionoid WGD facilitated the radiation of the most species-rich lineage of legumes, comprising 69 % of the species (13,390/19,327) and 59 % (438/741) of the genera of the third largest family of flowering plants. This is the group that is more uniformly characterized by the eponymous legume fruit, by the bilaterally symmetric papilionoid flower, and by the ability to nodulate. The early diverging grade of papilionoids does contain some genera with papilionoid flowers and legume fruits, but many lineages in this part of the tree are characterized by unusual, nonpapilionaceous corollas and drupaceous or samaroid fruits (Pennington et al. 2000); this grade also contains nearly all of the papilionoid genera that do not nodulate (Doyle 2011).

9.3.3 Polyploidy and Nodulation in Legumes

The correspondence between nodulation and polyploidy in the family is interesting. Core papilionoids nearly all nodulate, but this is also true of Mimosoideae, and *Chamaecrista* is among a handful of caesalpinioids known to be able to form a nodulation symbiosis (Sprent 2009). It remains unclear whether there was a single origin of nodulation in the common ancestor of the papilionoid and MCC clades, followed by many losses of nodulation, or whether there were multiple origins of

nodulation in the MCC clade and an independent origin in the ancestor of the core papilionoids (Doyle 2011).

The demonstration that *Chamaecrista* not only lacks the core papilionoid WGD, but also does not show any genomic evidence of other polyploidy events, indicates that polyploidy is not a prerequisite for nodulation in legumes as a whole (Cannon et al. 2010). Thus, in a model of a single origin of nodulation in the family, polyploidy would have played no role (Fig. 9.3). At the other extreme, polyploidy could not have been involved in an origin of nodulation unique to *Chamaecrista*, nor would it have been essential for the origin of nodulation in a model where the symbiosis evolved in a common ancestor of *Chamaecrista* and other members of the MCC clade (e.g., Mimosoideae).

This is not to say, of course, that polyploidy was not important in the origin or evolution of nodulation in core papilionoids or in other nodulating taxa whose genomes have yet to be explored, such as mimosoids. In the case of papilionoids, it is possible that nodulation and the WGD will be found to coincide, either in the ancestor of the main radiation of core papilionoids or at the first papilionoid ancestor, but better phylogenetic resolution is required before this can be tested (Pennington et al. 2001). In either case, refinement of the nodulation symbiosis in taxa such as *Medicago* may well have been facilitated by the availability of homoeologues produced in the core papilionoid WGD (Young et al. 2011).

9.3.4 Harmonizing Chromosomal and Genomic Evolution

Whatever the original basic chromosome number of the family, the earliest radiation from the common ancestor does not seem to have involved polyploidy, despite Goldblatt's (1981, p. 457) conclusion that the "… initial phase of polyploidy is probably very ancient and may have taken place in the late Cretaceous, when major groups of Leguminosae began differentiating and were probably evolving rapidly into new habitats." Evidence against Goldblatt's view of polyploidy in the ancestor of the entire family is the absence of any trace of polyploidy in gene families of *Chamaecrista* other than the prerosid whole-genome triplication (WGT; Jaillon et al. 2007). This indicates that the ancestor of the MCC and older ancestors back to the prerosid WGT did not experience polyploid duplications. The uniformity of base chromosome numbers in the major radiations of the family—Cercideae, detarioids, MCC, and probably papilionoids (see above) suggest that relatively high numbers ($2n = 24$–28) are plesiomorphic in the family.

The most parsimonious hypothesis for papilionoids is that the earliest papilionoid ancestor was also $2n = 28$. Shortly after the divergence of this ancestor from the MCC ancestor, the papilionoids radiated rapidly, and polyploidy occurred nearly simultaneously, no later than the divergence of the first major lineage to diverge in the core clade (genistoids, e.g., *Lupinus*). This WGD did not leave evidence in higher chromosome numbers; to the contrary, polyploidy is associated with chromosome number reduction in core papilionoids (Fig. 9.1). Goldblatt

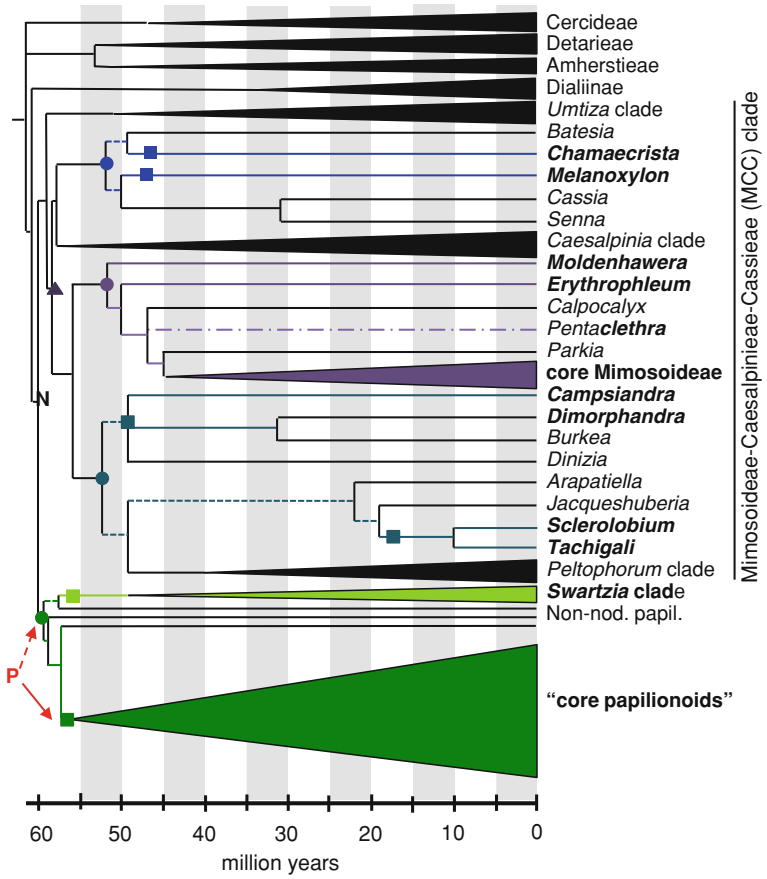

Fig. 9.3 Relationship between polyploidy and nodulation in *Leguminosae*. Genera known to nodulate are shown in bold face; *Pentaclethra* includes both nodulating species and species that apparently cannot nodulate. Possible origins of nodulation are indicated with symbols and colors. A single origin of nodulation for the entire family could have occurred in the common ancestor of the papilionoids and the MCC clade (indicated by "N"); this would have required many independent losses in the course of legume evolution. Independent origins within the MCC clade and Papilionoideae would require fewer losses. In the MCC clade, a single origin could have occurred in the common ancestor of all genera known to nodulate (*triangle*), or once in each major lineage of nodulating taxa (*circles*), or additional times within some clades (*squares*). Similarly, a single origin could be hypothesized for papilionoids (*green circle*), or twice (*squares*). The placement of the papilionoid polyploidy event (*red* "P") is indicated as in Fig. 9.1. Polyploidy is associated with nodulation only in the papilionoids and might not be directly associated there, given the uncertainty about the placement of both the polyploidy event and the origin(s) of nodulation. Figure adapted from Doyle (2011)

(1981, p. 457) notes that "the cytological history of legumes seems to involve some descending aneuploidy in every major evolutionary line but is most pronounced in Papilionoideae, in which most predominantly herbaceous tribes or

genera have achieved relatively low base numbers." The details of this process await elucidation of phylogenetic relationships of the early diverging papilionoid lineages, and sampling of these taxa for the presence of polyploidy. What is clear is that, at some point early in the history of Papilionoideae, chromosome numbers fell from as high as $n = 28$ to $n = 7$–11. This process could have begun in, or prior to, the common ancestor of the core papilionoids (Fig. 9.1). In any case, the initial reduction was rapid, taking place within 10 MY after the polyploid event, and perhaps within only 5 MY of the WGD. Detailed studies of synteny in the various core papilionoid lineages should help elucidate whether initial reorganization occurred in a common ancestor or was completely independent in the long-diverged major clades of the core papilionoids.

9.4 Polyploidy in Tribes and Genera of Legumes

Genomic and phylogenomic information currently is limited to the few genera discussed above. More information can be expected with the advent of new sequencing technologies, notably (at this writing) Illumina and 454. Twenty-five legume genera, representing all three subfamilies, are included in the 1,000 Transcriptomes project (1 kp, see http://www.onekp.com/angiosperms.html), but results from these species have not yet been analyzed (Steven Cannon, personal communication). Chromosome numbers provide a better guide at this level, but clearly should be interpreted with the caveat that high numbers are likely to indicate polyploidy, but low numbers cannot be assumed to be fundamentally diploid, given the potential for cryptic polyploidy, a phenomenon discussed further below.

Three sources of data were used for the following survey of polyploidy in legume clades: Goldblatt (1981); chromosome numbers provided for each genus listed in the tribal treatments in *Advances in Legume Systematics, Part* 1 (Polhill and Raven 1981); and the on-line Index of Plant Chromosome Numbers (IPCN; http://www.tropicos.org/projectwebportal.aspx?pagename=Home&projectid=9). IPCN was searched for all genera in Lewis et al. (2005) for which no data were available in Goldblatt (1981) or Polhill and Raven (1981), and for genera with polyploid counts, to identify infrageneric patterns of polyploidy, particularly when published phylogenies were available. For genera with evidence of polyploidy, BIOSIS was searched using the genus name in conjunction with either <polyploid*> and/or <phylogeny>. For some larger genera, IPCN was also consulted to search for polyploid counts published since 1981. There remain many gaps in our knowledge of legume chromosome numbers. Overall, only around 54 % of legume genera have counts reported (Table 9.1), and the percentage is far lower for tropical woody groups such as the caesalpinioid tribe Detarieae (30 %). This is perhaps not too surprising, given the size of the family, the large number of small genera, and the tropical distributions of many groups. It is also no doubt a commentary on how counting chromosomes has fallen from favor in this age of high-throughput science.

Table 9.1 Polyploidy in legume clades, based on chromosome numbers

Clade	Number of genera (species)[a]	Genera with chromosome counts	Number of genera with polyploidy	Percent genera with polyploidy (%)
Caesalpinioids[b]	171 (2251)	62	4	6
Mimosoideae	82 (3271)	39	10	26
Swartzieae + Sophoreae	40 (415)	15	2	13
Genistoids	83 (2354)	59	24	41
Dalbergioids	53 (1514)	32	11	34
Baphioids	7 (58)	6	1	17
Mirbelioids	32 (763)	22	5	23
Hologalegina: Robinioids + Loteae	34 (414)	23	9	39
Hologalegina: IRLC	54 (4351)	40	15	38
Millettioids: indigofereae	7 (768)	2	1	50
Millettioids: core Millettioids	56 (1104)	22	1	5
Millettioids: phaseoloids	112 (2064)	71	12	17
Total	731	393	95	24

[a] *Source*: Lewis et al. (2005)

[b] Summation of the following monophyletic groups: Cercideae, 12 genera total, 2 of 6 genera with chromosome counts have reports of polyploidy, 33 % (12, 2/6, 33 %); Dialiinae (17, 0/7, 0 %); Detarieae (82, 2/25, 8 %); MCC clade minus mimosoids (59, 0/24, 0 %)

Goldblatt (1981) made a distinction among genera that are exclusively polyploid relative to genera in the same tribe, and genera that include species with both "diploid" and "polyploid" chromosome numbers. In light of our current understanding of polyploidy, this is an artificial distinction, but exclusively polyploid genera are perhaps worth noting, because they represent lineages where polyploids may have replaced their diploid progenitors entirely. Such lineages would constitute evidence of polyploid "success" (e.g., Mayrose et al. 2011) in measures of diversity a million years from now.

The results of this survey show that polyploidy, as inferred solely from chromosome numbers, occurs in nearly a quarter of all legume genera, but varies widely in frequency among different lineages (Table 9.1). As Goldblatt (1981) noted, polyploidy is rare in woody, tropical groups such as caesalpinioids and early diverging papilionoid lineages.

9.4.1 Caesalpinioids and Mimosoideae

9.4.1.1 Clades of the Caesalpinioid Grade, Excluding the MCC Clade

Goldblatt (1981) stated that polyploidy (assuming a basic diploid number of $2n = 28$) is uncommon in caesalpinioid groups. As noted above, unpublished

genome size data do not support the hypothesis that *Bauhinia* (commonly $2n = 28$) is tetraploid relative to *Cercis* ($2n = 14$). However, this is negative evidence, and given the prevalence of genome downsizing (Leitch and Bennett 2004) in polyploids, it remains possible that *Bauhinia* (along with *Adenolobus* and *Griffonia*) is polyploid. More recent polyploidy occurs in the group, with *Tylosema* having $2n = 52$; *Tylosema* is nested within *Bauhinia* s. l. (Sinou et al. 2009). The large Detarieae s.l. clade is overwhelmingly $2n = 24$. The two detarioid genera having higher chromosome numbers (*Hardwickia*, $2n = 34$; *Colophospermum*, $2n = 36$) are strongly supported as sisters within the *Prioria* clade of Bruneau et al. (2008); they could represent independent aneuploid reduction from a polyploid ancestor. Goldblatt (1981) mentioned polyploidy in *Anthonotha* ($2n = 24$, 28, 72), but only $2n = 24$ and 28 are listed by Cowan and Polhill (1981), and no counts are listed in IPCN. All known counts from the Dialiinae clade are $2n = 24$ or 28.

9.4.1.2 The MCC Clade

Chromosome numbers of the caesalpinioid members of the MCC clade range from $2n = 20$–28, with the exception of *Chamaecrista*, in which numbers of $2n = 14$, 16, and 28 occur. Goldblatt (1981) raised the possibility that low chromosome numbers in this genus are not ancestral, but instead represent aneuploid reduction from ancestral $x = 14$ species. This makes good sense given the prevalence of $2n = 28$ counts throughout much of the MCC clade. A recent molecular phylogenetic analysis of *Chamaecrista* (Torres et al. 2011) supported the hypothesis that the $2n = 14$ species of sect. *Xerocalyx* form a monophyletic group nested within species having higher chromosome numbers. *Chamaecrista* is a genus of considerable interest as a potential model for nonpapilionoid legumes (Singer et al. 2009), making it an attractive system for exploring chromosome evolution in the caesalpinioids.

Mimosoideae comprise a monophyletic group embedded in the MCC clade, and the subfamily is dominated by taxa with $n = 13$. Goldblatt (1981) listed only three genera with base numbers higher than $n = 14$: *Schleinitzia* ($2n = 52$, 54), *Leucaena* ($2n = 52$, 56, 104, 112), and *Dichrostachys* ($2n = 50$, 56). These genera are all members of tribe Mimoseae and are relatively closely related within that tribe (Lewis et al. 2005), with the former two being members of the same clade and *Dichrostachys* being part of a sister clade (Luckow et al. 2003). However, *Schleinitzia* and *Leucaena* are not sisters within their clade (Hughes et al. 2003; Luckow et al. 2005), suggesting that polyploidy has originated independently in each case.

Leucaena itself has been fertile ground for systematic investigation. Boff and Schifino-Wittmann (2003) concluded that its species are segmental paleopolyploids. A series of studies has built a strong foundation for understanding the complex history of hybridization and polyploidy in the genus, and the impact of these phenomena on characters such as nrDNA ITS pseudogene evolution, and

their role in domestication (Hughes et al. 2002, 2007; Govindarajulu et al. 2011a, b). Govindarajulu et al. (2011b) concluded that "… a comprehensive picture of the complex evolutionary dynamics of polyploidy in *Leucaena* is emerging. This includes paleotetraploidization, diploidization of the last common ancestor to *Leucaena*, allopatric divergence among diploids, and recent allopolyploid origins for tetraploid species likely associated with human translocation of seed."

Acacia (sens. lat.) is reported to have $2n = 26$, 52, 76, and 104, though the majority of its species are diploid (Gallagher et al. 2011). Polyploidy has not been a focus of recent phylogenetic analyses (e.g., Brown et al. 2010; Murphy et al. 2010). Similarly, phylogenetic studies of *Prosopis* ($2n = 28$, 52, 56) or *Prosopidastrum* ($2n = 28$, 56) do not discuss polyploidy (e.g., Bessega et al. 2006; Catalano et al. 2008). No phylogenetic studies appear to have addressed polyploidy in *Neptunia* ($2n = 28$, 36, 54, 56, 72, 78), though Pandit et al. (2006) note that *N. plena*, an invasive species in Singapore, is a polyploid ($2n = 72$). No studies appear to exist on polyploidy and phylogeny of *Inga* ($2n = 26$, 52; the latter reported by Hanson 1995), *Albizia* ($2n = 26$, c. 78), or *Calliandra* ($2n = 16$, 22, 32, 44).

Mimosa ($2n = 24$, 26, 28, 40, 52) is a genus of around 500 species; a second genus whose species vary in ploidy, *Schrankia* ($2n = 16$, 22, 24, 26, 52), is deeply nested within *Mimosa* (Simon et al. 2011). Dahmer et al. (2011) concluded that the phylogenetic pattern "… suggests that duplication of chromosome numbers evolved several times in the genus and that polyploidy is not restricted to any particular clade within *Mimosa*. On the contrary, it seems that polyploids arose independently from ancestors with lower ploidy levels and are present in divergent lineages in the genus." Seijo and Fernandez (2001) reported chromosome numbers from the southern extreme of the range and discovered polyploidy within *M. balansae*. Chromosomal and morphological studies by Morales et al. (2010) clarified relationships in the *M. debilis-M. nuda* complex, demonstrating that hybridization and polyploidy are responsible for taxonomic complexity in the group.

9.4.2 Papilionoideae

Relatively few papilionoid genera appear to be exclusively polyploid based on chromosome number. As noted above, the early diverging papilionoid lineages have relatively high numbers, like the caesalpinioid and mimosoid groups. *Ateleia* (Swartzieae) is $2n = 40$, presumably representing a second polyploidy event followed by aneuploid reduction. Goldblatt listed *Dipteryx* (Dipterygeae) as being a polyploid genus but with a questionable count of $2n = 32$; this number was not reported in the treatment of the tribe by Polhill (1981a), nor is a count for the genus listed in IPCN.

9.4.2.1 Genistoids

The genistoid clade, which is weakly supported as sister to the remaining core papilionoids (Fig. 9.1), is one of the most complex groups with respect to polyploidy, and detailed discussion is beyond the scope of this review. The frequency of polyploidy in the genistoid clade is the highest for any well-sampled group of legumes (Table 9.1). Cusma-Velari and Feoli-Chiapella (2009) discussed cytology in "so-called 'primitive' genera of Genisteae" in light of molecular phylogenetic data. Genistoids clearly have a base number of $x = 9$ (Goldblatt 1981), with $2n = 18$ being common in most of its tribes.

Sophora s.l. has been divided into several segregates that vary in chromosome number. *Sophora* s.s. is part of the genistoid clade and is $x = 9$. Boatwright and van Wyk (2011) reported on the relationships of several of these based on nrDNA ITS sequences. They focused on the placement of the South African species, *S. ihambanensis*, which is polyploid ($2n = 36$); in their tree it is sister to *S. tomentosa*, a diploid, but they do not discuss origins of the polyploid. A count of $2n = 18$ is common in *Sophora* s.l., and several additional species are polyploids with $2n = 36$ (*S. alopecuroides, S. pachycarpa*, and *S. songarica*); *S. leachiana* is listed by ICPN as having $2n = 36$ and $2n = 54$ cytotypes. The small segregate, *Calia*, has $2n = 18$ and may be sister to the entire genistoid clade. The remainder of *Sophora* s.l. comprises $2n = 28$ species transferred to *Styphnolobium* in the early diverging papilionoid grade.

In Thermopsideae, *Thermopsis* has both diploid and polyploid species ($2n = 18, 36$). In Podalyrieae, two genera are exclusively polyploid: *Virgilia* ($2n = 54$) and *Cyclopia* ($2n = 36$); they are not supported as sisters in Boatwright et al. (2008). Crotalarieae are sister to Genisteae and include the exclusively polyploid *Buchenroedera* ($2n = 28$; Van Wyk and Schutte 1988), as well as polyploids within *Crotalaria* ($2n = 14, 16, 32$) and *Lotononis* ($2n = 18, 28, 36$).

Genisteae are by far the most cytologically complex group in the entire Leguminosae. Even genera with low numbers may be polyploid, such as *Anarthrophyllum* ($2n = 24$; Goldblatt 1981) and *Dichilus* ($2n = 28$). These genera, along with the polyploid *Polhillia* ($2n = 32$) and complex *Melolobium* ($2n = 18, 32$), were once placed in Crotalarieae. It is in the core Genisteae that polyploidy and aneuploidy have run rampant. The group includes *Argyrolobium* ($2n = 24, 26, 30, 32, 48$), *Adenocarpus* ($2n = 26, 46, 48, 52, 54$), *Laburnum* ($2n = 48, 50$), *Cytisophyllum* ($2n = 50, 52$), *Petteria* ($2n = 52$), *Argyrocytisus* ($2n = 50$), *Chamaecytisus* ($2n = 48, 96$), *Cytisus* ($2n = 22, 24, 46, 48, 92, 96$), *Calicotome* ($2n = 24, 48, 50$), *Erinacea* ($2n = 52$), *Spartium* ($2n = 48, 52, 54, 56$), *Retama* ($2n = 48$), *Genista* ($2n = 18, 22, 24, 26, 28, 30, 32, 36, 40, 42, 44, 46, 48, 50, 52, 56, 72, 80, 96$), *Echinospartium* ($2n = 44, 52$), *Stauracanthus* ($2n = 28, 48$, ca. 128), *Ulex* ($2n = 32, 64, 80, 96$), and *Lupinus* ($2n = 24, 30, 32, 34, 36, 38, 40, 42, 48, 50, 52, 96$).

Bisby (1981) considered the plethora of chromosome numbers attributed to individual genera to be partly a real phenomenon, but also due to difficulties in obtaining reliable counts given the small size and high numbers of chromosomes,

combined with the taxonomic complexity of the groups. Some of the taxonomic complexity is being resolved by molecular phylogenetic studies focused on *Cytisus* and *Genista* (Cubas et al. 2002; Pardo et al. 2004), but these studies do not discuss polyploidy per se. For *Lupinus*, three studies have documented the rapid radiation of the genus in its New World center of diversity (Hughes and Eastwood 2006; Drummond 2008; Drummond et al. 2012), and another phylogenetic study focused on the Old World species (Ainouche et al. 2004). Drummond (2008) noted that, "While a complex history of aneuploidy ($2n = 32, 34, 36, 38, 40, 42, 50, 52$) in the Old World and eastern New World … implies that allopolyploidy may have provided an additional mechanism for reproductive isolation and evolutionary divergence, chromosomal numbers in the western New World species ($2n = 48$ with occasional autopolyploids of $2n = 96$) are relatively stable." It is this western group that has radiated explosively, presumably driven by ecology and not due to polyploidy per se (Drummond 2008). Conterato and Schifino-Wittmann (2006) described chromosome numbers and meiotic behavior in diploid and polyploid American lupines, and noted consistencies with phylogenetic relationships in the genus.

The placement of polyploid former Crotalarieae in the same clade with core Genisteae may lend support to the idea that polyploidy arose early in the entire clade, as suggested by Goldblatt (1981), Lavin et al. (2005) dated the common ancestor of Crotalarieae and Genisteae at around 41 MYA. On the other hand, Goldblatt also noted (1981, p. 452) that "basic numbers for these genera (*Genista*, *Ulex*, *Cytisus*) are however in the diploid range and a basic number of $x = 12$ for the group as a whole and for several genera has been suggested …". Genomic data for members of Genisteae should eventually allow the determination of the number and relative timing of polyploid events in the group.

9.4.2.2 Dalbergioids

The dalbergioid s.l. clade is split into two major subclades, Amorpheae and a second clade comprising Adesmieae, Aeschynomeneae, and many members of the polyphyletic Dalbergieae (Lewis et al. 2005). The entire dalbergioid clade is dominated by $2n = 20$ species. Within Amorpheae, polyploidy occurs in genera from each of the major subclades described by McMahon (2005). In the daleoid clade, *Dalea* has $2n = 14$, 16, 28, and 42. Spellenberg (1981) hypothesized that tetraploids and hexaploids of *D. formosa* ($2n = 28$, 42) were autopolyploids derived from the diploid ($2n = 14$) cytotype. In the amorphoid clade, *Amorpha* includes both diploids and polyploids ($2n = 20$, 40), all native to the New World. The widespread *A. fruticosa* is exclusively polyploid and morphologically complex (Wilbur 1975). It has become an invasive weed in Europe (e.g., Hulina 2010), illustrating a common feature of polyploidy (e.g., Pandit et al. 2011; te Beest et al. 2011). Its relationships to other members of the genus appear to be complex, sharing chloroplast haplotypes with different sympatric diploids across its range (Straub and Doyle, unpublished data). Studies of the *A. georgiana* complex

identified mixed populations of diploids and polyploids in what had been assumed to be an exclusively diploid species; these included an apparent allopolyploid between *A. georgiana* and *A. herbacea* (Straub and Doyle 2009). Straub (unpublished data) has identified additional polyploid species in the genus and hypothesized their origins.

The core dalbergioid clade is split into the small *Adesmia* clade (six genera) and the much larger clade comprising the *Dalbergia* and *Pterocarpus* sister clades (Lewis et al. 2005). Polyploidy occurs in all three clades. In the *Adesmia* clade, *Adesmia* includes both diploids and polyploids ($2n = 20, 40$), and *Amicia* is exclusively polyploid ($2n = 38$).

In the *Dalbergia* clade, *Smithia* ($2n = 38$) is exclusively polyploid and is considered closely related to *Kotschya* (Lewis et al. 2005), a genus that includes species with chromosome numbers indicative of polyploidy and aneuploidy ($2n = 28, 30, 36, 40$). These two genera are grouped with *Aeschynomene* species (Lavin et al. 2001), a genus that also includes diploids and polyploids ($2n = 18, 20, 40$). Another dalbergioid-clade genus, *Ormocarpum* ($2n = 24, 26$), was not listed in Goldblatt's discussion of polyploid genera, but could potentially be a cryptic polyploid with aneuploid reduction. Information on chromosome numbers of other members of the *Ormocarpum* group (Thulin and Lavin 2001) would be useful in addressing this issue.

The *Pterocarpus* clade also includes several genera with both diploids (including presumed aneuploids) and polyploids scattered among its subclades: *Platymiscium* ($2n = 16, 18, 20, 32$), *Pterocarpus* ($2n = 22, 24, 44$), *Geoffroea* ($2n = 20, 60$), and *Arachis* ($2n = 20, 40$). *Arachis* includes the tetraploid peanut or groundnut (*A. hypogaea*), as well as three other tetraploid species, one of which (*A. glabrata*) is a tropical forage crop. Peanut is hypothesized to be an allopolyploid derived from A- and B-genome species (e.g., Burow et al. 2009). Seijo et al. (2007) provide a useful summary of hypotheses concerning the origin (or origins) of peanut; controversy exists concerning such issues as the exact progenitor species of both homoeologous genomes, mode and number of origins, and whether there was subsequent introgression from wild species into the cultigen. They identified likely diploid progenitors of *A. hypogaea* using GISH and studied meiotic behavior of two other tetraploids (Ortiz et al. 2011) and a spontaneous autotriploid of *A. pintoi* (Lavia et al. 2011).

9.4.2.3 Baphioids

This small group appears to be interesting from the standpoint of polyploidy. Chromosome numbers are known from six of its seven genera. Of these, four are listed as $2n = 22$, one (*Dalhousiea*) is $2n = 44$, and *Baphia* was reported in Polhill and Raven (1981) to have both numbers, though no reports for any of the 47 species of the genus exists in IPCN.

9.4.2.4 Mirbelioids

Polyploidy in the mirbeliod clade (23 %) is close to the average for the whole family (Table 9.1). *Isotropis* ($2n = 16, 18, 32$) is sister to the large "*Pultenaea* s.l. group" in Orthia et al. (2005), which includes *Oxylobium* ($2n = 16$). Chandler et al. (2001) sank both *Brachysema* ($2n = 16, 32$) and the monotypic *Jansonia* ($2n = 32$) in *Gastrolobium*, previously a genus with only $2n = 16$ species. Chandler et al. (2001) placed the single sampled species of *Jansonia* sister to *Brachysema celsianum* (not listed in IPCN), in a clade that also included *B. praemorsum*; that species is a diploid at $2n = 16$, as is at least one species in the sister clade, *Nemcia coriacea*. Thus, it is likely that polyploidy has arisen more than once just within *Gastrolobium* s.l., and another time in *Isotropis*. Among several "strongly paraphyletic" *Pultenaea* s.l. genera, nearly all of which are $2n = 16$, are two genera with polyploidy: *Pultenaea* s.s. ($2n = 8, 12, 14, 16, 18, 27, 32$) and *Chorizema* ($2n = 16, 32$). Sorting out how the various chromosome numbers in *Pultenaea* s.s. are related will be of considerable interest but will require more complete phylogenies than appear to be available at present. Of the 10 *Chorizema* species (out of 27 in the genus) listed in IPCN, polyploidy is only reported from *C. aciculare*, which has both diploid and tetraploid cytotypes.

Smaller genera in *Pultenaea* s.l. with known polyploidy are *Eutaxia* ($2n = 16, 32$) and *Dillwynia* ($2n = 14, 21, 28$), which are in the same weakly supported clade in Orthia et al. (2005). One of the two *Dillwynia* species (*D. phylicoides*) included in the Orthia et al. (2005) tree has both diploid and tetraploid cytotypes listed in IPCN.

9.4.2.5 Hologalegina

The Hologalegina clade includes robinioids and the Inverted Repeat Loss Clade (IRLC; named for the absence of a prominent feature of the chloroplast genome). Goldblatt (1981) concluded that "Species polyploidy is overwhelmingly concentrated in temperate to cool Eurasia" so it is not surprising that this largest clade of legumes, which includes many temperate genera, has a higher frequency of polyploidy than the family average, nearly 40 % in both of its major subclades (Table 9.1).

The robinioids comprise two clades: one with *Sesbania* plus Loteae (including Coronilleae), the other being Robinieae (s.s.). Diploid chromosome numbers vary considerably within Robinieae s.s., and Goldblatt (1981) suggested several possible base numbers, the most likely being $x = 10$ or 11. There is one apparently exclusively polyploid genus, *Poissonia*, which Goldblatt (1989) counted as ($2n =$ ca. 32). Although only numbers of $2n = 10$ and 11 were given for *Robinia* in Polhill and Sousa (1981), more recent counts of $2n = 30$ for *R. hispida* suggest polyploidy within this small genus.

Even with the limited sampling of Loteae in Wojciechowski et al. (2004), it is clear that numerous problems exist with the genera as circumscribed, notably that

Anthyllis ($2n = 10, 12, 14, 16, 28$) and *Ornithopus* ($2n = 14$) are nested within *Lotus* ($2n = 10, 12, 14, 24, 28$). Other genera with known polyploidy are *Coronilla* ($2n = 12, 20, 24$); *Hippocrepis* ($2n = 14, 28$); *Dorycnium* ($2n = 14, 28$; both cytotypes in *D. axilliflorum*), and *Scorpiurus* (*S. muricatus* has $2n = 14, 16, 28$; other species are $2n = 14$ or 28). Degtjareva et al. (2006) provided phylogenetic hypotheses for *Lotus*, with sampling of other genera, but did not discuss polyploidy. Rosello and Castro (2008) discussed polyploidy in the flora of the Balearic Isles, among which are species of *Anthyllis* and *Coronilla*.

The genus *Lotus* includes the genomic model legume, *L. japonicus* (Sato et al. 2008), which is part of the *L. corniculatus* (birdsfoot trefoil) polyploid complex. Grant and Small (1996) summarized many studies of this complex and concluded that it was a fertile topic for further study, particularly to identify the diploid progenitors of *L. corniculatus* itself, which they considered to be an allopolyploid. Gauthier et al. (1998a, b) discussed evolutionary patterns in the *L. corniculatus*/*L. alpinus* polyploid complex in the Alps of Europe; they described morphological and genetic consequences of autopolyploidy in *L. alpinus* and suggested introgression at the tetraploid level between it and *L. corniculatus*.

The majority of genera and species in the IRLC clade belong to two sister clades in Wojciechowski et al. (2004): one includes the Astragalean clade (*Astragalus* and allies) and Hedysareae, and the second includes the Vicioid clade. The remainder of the IRLC phylogeny, moving successively further from these clades, consists of a clade with *Wisteria* ($2n = 16$) and one species of *Callerya*, followed by a clade with *Glycyrrhiza* ($2n = 16$) and a second species of *Callerya*.

The Astragalean clade has extensive polyploidy. Perhaps most striking is the clade that includes the New Zealand endemic tribe Carmichaelieae plus the Australian *Swainsona* and an additional New Zealand genus, *Montigera*, all of which are polyploid (Wagstaff et al. 1999). The only exclusively polyploid genera that Goldblatt (1981) listed for the IRLC clade belong to this clade: *Swainsona* ($2n = 32$); *Clianthus* ($2n = 32$); *Carmichaelia* ($2n = 32$, ca. 96), *Chordospartium* ($2n = 32$); and *Corallospartium* ($2n = 32$), with the latter two subsumed in *Carmichaelia* in Lewis et al. (2005). Wagstaff et al. (1999) concluded that the New Zealand radiation was recent, involved an already polyploid colonizer, and may have been associated with orogeny and glaciation.

Elsewhere in the Astragalean clade are two large genera with extensive polyploidy and aneuploidy, *Oxytropis* (300–400 species: $2n = 16, 32, 48, 64, 96$) and *Astragalus* (ca. 2500 species: $2n = 16, 22, 24, 26, 28, 32, 44, 48, 64$). In *Astragalus*, polyploidy appears to be more common among Old World than among New World species. Wojciechowski (2005) summarized phylogenetic results for this huge genus, showing that aneuploid species form a clade. According to Gohil and Ashraf (2008), polyploidy occurs in around only 17 % of *Astragalus* species. However, *Astragalus* is one of the largest genera of plants with as many as 2500 species (Lewis et al. 2005), so if this percentage is correct, then there are over 400 polyploid species in the genus. There does not seem to be a comprehensive phylogeny that discusses origins of polyploidy in *Oxytropis*. However, a series of papers describe autopolyploidy, including multiple autopolyploid origins, in

Oxytropis chankaensis (e.g., Artyukova et al. 2011). Jorgensen et al. (2003) suggested "a scenario of multiple formations of polyploids, possibly including hybridization among diverged Alaskan *Oxytropis* populations."

Within the Hedysareae, phylogenetic studies of *Caragana* ($2n = 16, 24, 32, 48$) suggest that polyploidy is confined to a single group of species, and that triploids, tetraploids, and hexaploids may all be autopolyploid in origin (Zhang et al. 2009). Neither *Hedysarum* ($2n = 14, 16, 48$) nor *Onobrychis* ($2n = 14, 16\ 28, 32$) appears to be monophyletic on the basis of nrDNA ITS phylogenies (Ahangarian et al. 2007). Hejazi et al. (2010) discussed karyotypic evolution in diploid and polyploid species but did not provide a phylogenetic context or identify origins of polyploids. Based on IPCN listings, most polyploidy reported for *Hedysarum* in IPCN appears to involve multiple cytotypes within a single species (e.g., *H. arcticum* and *H. hedysaroides*, both $2n = 14, 28$; *H. dasycarpum* and *H. mackenziei*, both $2n = 16, 32$; *H. gmelinii*, $2n = 16, 28, 56$), but some species are exclusively polyploid (e.g., *H. inundatum*, $2n = 28$). Similarly, in *Onobrychis* there is variation within species (e.g., *O. aequidentata*, $2n = 14, 16, 28$; *O. arenaria*, *O. bobrovii*, $2n = 14, 28$; *O. crista-galli*, $2n = 16, 32$), with other species being exclusively polyploid (e.g., *O. biebersteinii*, *O. cyri*, *O. dielsii*, all $2n = 28$). There appear to be no phylogenies or evolutionary studies of polyploidy in *Alhagi* ($2n = 16, 28$; the latter number is not listed in IPCN).

The majority of the vicioid clade forms two sister clades, one with Fabeae (Vicieae) plus *Trifolium*, and a second comprising the remaining Trifolieae genera; polyploidy occurs in both clades. Successive sisters to this clade (*Cicer*, *Galega*, and *Parochetus*) are all $2n = 16$.

Vicia includes both diploids and tetraploids ($2n = 10, 12, 14, 24, 28$), but polyploidy was considered rare in the genus by Kupicha (1981). Indeed, *Vicia* is best known for its extensive non-polyploid variation in genome size (Chooi 1971; Neumann et al. 2006), which shows only weak correlation with ploidy: diploid ($2n = 14$) *V. peregrina* has a genome size of 9.48 pg/1C, double that of tetraploid ($2n = 24$) *V. tenuifolia* (4.73 pg/1C). Endo et al. (2008) did not discuss either issue in their phylogenetic study of New Wold *Vicia*. Travnicek et al. (2010) studied the history of polyploidy in *V. cracca*, determining the ploidy of over 6,500 individuals at more than 250 localities in Europe and mapping the distributions of diploids, triploids, and tetraploids; they noted the rarity of triploids, suggesting strong reproductive barriers between diploids and tetraploids.

Polyploidy is also noted to be rare in *Lathyrus* ($2n = 14, 28, 42$; Kupicha 1981). Gutierrez et al. (1994) hypothesized autopolyploid origins of *L. pratensis* and *L. palustris* from conspecific diploids, but an allopolyploid origin of *L. venosus* from two diploid species (*L. ochroleucus* and *L. palustris*). Only $2n = 14$ counts are listed for the closely related *Pisum* in IPCN, for which Kupicha (1981) listed polyploidy as "rare."

Turini et al. (2010) reconstructed nrDNA ITS and chloroplast phylogenies for 69 of the 86 species of *Ononis* ($2n = 16, 20, 30, 32, 60, 64$) and identified several well-supported clades. They concluded that, "Unfortunately, only limited information is available … on chromosome numbers to test support for these groups".

However, chromosome counts are available for nearly half of the species in their phylogeny in IPCN, and some conclusions can be drawn. For example, the clade that is strongly supported as sister to the remainder of the genus in their nrDNA ITS tree includes only polyploids (*O. tridentata* and *O. fruticosa* are both $2n = 30$; *O. rotundifolia* is $2n = 32$), suggesting that the genus as a whole could be polyploid. Only three species have low, potentially non-polyploid numbers in IPCN. These occur in different clades, and in two cases species with low numbers have higher numbers as well (*O. variegata*, $2n = 16, 30$; *O. ornithopodioides,* $2n = 20, 32$), raising the possibility that they are independent reductions from typical polyploid numbers. The exception, *O. adenotricha*, is only reported as $2n = 16$; its position varies between the nrDNA ITS and *trnL-F* trees of Turini et al. (2010), being sister to the *tridentata* clade in the *trnL-F* tree; however, this entire group is not resolved as sister to the remainder of the genus in that tree. Elsewhere in the genus, *O. spinosa* has multiple cytotypes ($2n = 30, 32, 60$), whereas *O. pendula* is only known at $2n = 64$. Kloda et al. (2008) studied patterns of genetic diversity in several diploid and polyploid species in England and concluded that gene flow was occurring within ploidy levels, but not between diploids and tetraploids.

Medicago includes the genomic model legume, *M. truncatula* (Young et al. 2011). Steele et al. (2010) provided a phylogeny for *Medicago* ($2n = 14, 16, 32, 48$), including multispecies sampling of its sister clade, which comprises the interdigitated species of the two paraphyletic genera *Melilotus* ($2n = 16, 24, 32$; though tetraploids are not reported in IPCN) and *Trigonella* ($2n = 16, 28, 32, 44$). Aneuploid change from $2n = 16$ to $2n = 14$ has occurred several times in *Medicago* (Steele et al. 2010). Polyploidy is concentrated in a clade that comprises most species of sect. *Medicago*, along with *M. arborea* (sect. *Dendrotelis*); an additional polyploid species, *M. scutellata*, occurs in the clade sister to this sect. *Medicago* clade. Rosato et al. (2008) used fluorescence in situ hybridization (FISH) to study relationships between polyploids and diploids in sect. *Dendrotelis*.

Some *Medicago* species are exclusively polyploid whereas others possess multiple cytotypes. The *M. sativa* complex, which includes cultivated autotetraploid alfalfa (*M. sativa* ssp. *sativa*) as well as other diploid ($2n = 16$) and autopolyploid ($2n = 32$) species and their hybrids, has been the focus of several recent studies (Sakiroglu et al. 2010; Havananda et al. 2010, 2011, and unpublished data). Two major autopolyploid pairs in the complex are: (1) *M. s. caerulea* and *M. s. sativa*, both with blue flowers and coiled pods, distinguishable by the larger size of the tetraploid (*M. s. sativa*) for several characters; and (2) *M. s. falcata*, a yellow-flowered taxon with falcate pods whose diploid and polyploid cytotypes are indistinguishable morphologically. Interestingly, although *M. s. sativa* and *M. s. caerulea* are undifferentiated for chloroplast haplotypes, the two cytotypes of *M. s. falcata* possess nearly mutually exclusive sets of haplotypes, with haplotypes in the tetraploid most likely derived by introgression from *M. prostrata*, a species from outside the complex (Havananda et al. 2011). Jenczewski et al. (1999) reported gene flow between wild and cultivated *M. sativa* populations; however, based on chloroplast data, there does not appear to be significant gene flow between blue- and yellow-flowered taxa in the complex either at the diploid or

tetraploid levels, despite the existence of morphologically intermediate hybrid subspecies (Havananda et al. 2011, and unpublished data). Much is known about the genetics of polyploidy in alfalfa, where unreduced gametes have received considerable study as a breeding tool (e.g., Bingham 1972; Veronesi et al. 1986; Tondini et al. 1993; Calderini and Mariani 1997).

Ellison et al. (2006) constructed a phylogeny of *Trifolium* ($2n = 10, 14, 16, 28, 32$) that included 218 of its ca. 255 species, as well as species from 11 genera of the vicioid clade. They showed that the genus is monophyletic; incongruence within the genus between nuclear and chloroplast markers suggests considerable hybridization. They also hypothesized a minimum of 19 shifts to aneuploidy and 22 instances of polyploidy from a base number of $2n = 16$. They identified the progenitors of two important species, both shown to be allopolyploids: the widespread weed, *T. dubium*, and the most commonly cultivated clover species, *T. repens* (Ellison et al. 2006).

9.4.2.6 Indigofereae

Schrire et al. (2009) provided a detailed phylogeny for this tribe, a monophyletic group that is sister to the millettioid clade. Schrire et al. (2009) did not comment on chromosomal variation or polyploidy, but numerous records are readily available in IPCN, and mapping these onto the phylogeny provides some insights into cytological evolution of the group.

The tribe is dominated by the very large genus *Indigofera* (ca. 700 species), which Goldblatt (1981) and Polhill (1981b) listed as having $2n = 14, 16, 32, 48$. The higher numbers thus would be interpreted as representing tetraploids and hexaploids. However, Frahm-Leliveld (1966), summarizing the cytotaxonomy of the tribe, cited two $x = 6$ species, *I. macrocalyx* ($2n = 12$) and *I. emarginella* ($2n = 24$), and concluded that "… the 48-chromosome Himalayan and East-Asiatic shrubby *Indigoferas* may not be hexaploids with base number $x = 8$, but octoploids in an $x = 6$ range." None of the species is listed in IPCN, but Reddy and Revathi (1993) reported $2n = 12$ for *I. anil*, confirming the presence of $x = 6$ in the genus.

The Schrire et al. (2009) phylogeny does not support the Frahm–Leliveld (1966) hypothesis. One of the two $2n = 12$ species, *I. macrocalyx*, is placed in the large Palaeotropical clade of Schrire et al. (2009) and is sister to a group of species that includes *I. pulchra* ($2n = 16$). All three sampled species with $2n = 48$ are in the Palaeotropical clade, but are placed nowhere near *I. macrocalyx*. In the Pantropical clade, *I. rhynchocarpa* ($2n = 16$) is sister to the clade that includes *I. emarginella*, which is on a long branch sister to several other species; the only other species counted from this subclade is also $2n = 16$. Thus, there is no evidence that $2n = 48$ species are derived from $x = 12$ species.

Tetraploids based on $x = 8$ are scattered throughout the phylogeny (Schrire et al. 2009), supporting the observation of Frahm-Leliveld (1966) that $2n = 32$ is common in the genus. In the Palaeotropical clade, *I. atriceps* ($2n = 32$) is in a

subclade that also includes diploids. In another subclade, *I. mysorensis* includes both $2n = 16$ and 32 cytotypes; other two members of its subclade are diploid. *Indigofera microcalyx*, in yet another subclade, is $2n = 32$; no other members of its subclade has been counted, but the only counts from its sister clade are diploid.

In the pantropical clade, a small subclade in Schrire et al. (2009) includes *I. koreana* ($2n = 32$), *I. grandiflora* ($2n = 32, 48$), *I. decora* ($2n = 48$ in Choi and Kim (1997) but not listed in IPCN), as well as *I. venulosa* (no count available) and *I. kirilowii* ($2n = 16$). Topologies differ between Schrire et al. (2009) and Choi and Kim (1997), who focused on this group of mostly Korean endemics. Choi and Kim (1997) listed *I. grandiflora* as $2n = 16$, and given this count their topology could suggest independent derivation of polyploidy in *I. koreana* (from *I. grandiflora*) and *I. decora* (from *I. venulosa* if it is diploid). An alternative explanation is a single derivation of polyploidy within this clade. Elsewhere in the pantropical clade, *I. heterantha* ($2n = 48$) is sister to *I. hebepetala* ($2n = 16$); the clade sister to these two species includes *I. amblyantha* ($2n = 48$) and *I. cassioides* ($2n = 16$). In a different subclade, *I. suffruticosa* is reported to have both $2n = 16$ and 32 cytotypes; the closest reported species to it is diploid. In the Tethyan clade, *I. sessiliflora* ($2n = 32$) is the only member of its subclade with a count in IPCN. *Indigofera hochstetteri* has both $2n = 16$ and 32 counts; its sister species, *I. arabica*, is diploid. *Indigofera angulosa* is $2n = 32$; no other species in its clade has counts in IPCN.

Thus, there appear to be no large clades composed exclusively of polyploids in *Indigofera*. Instead, as in other large legume genera, polyploidy is sporadic.

9.4.2.7 Millettioids

With the recognition that *Wisteria* and *Callerya* are part of the IRLC, and that *Cyclolobium* and *Poecilanthe* belong in the Brongniartieae, chromosome numbers in the Millettieae (Tephrosieae in Polhill and Raven (1981)) are mostly $2n = 20$ or 22, with $2n = 24$ in *Xeroderris*, though many genera have no reported counts in IPCN. Interestingly, *Xeroderris* is placed as sister to the remainder of the entire millettioid clade in Wojciechowski et al. (2004), suggesting that $2n = 20$ or 22 may be synapomorphic for the remainder of the millettioid clade (including phaseoloids; see below). Millettieae comprises the bulk of one of the two major millettioid clades (core millettioids), along with *Abrus* (Abreae), and much of the subtribe Diocleinae of tribe Phaseoleae. Both diploid and tetraploid cytotypes ($2n = 22, 44$) have been reported in three species of the ca. 40 IPCN records for the large (ca. 350 spp.) genus *Tephrosia* (e.g., Srivastav and Raina 1986). The low frequency of polyploidy in the core millettioid clade (5 %) is nearly identical to the frequency in caesalpinioids, both of which are largely woody, tropical groups.

The other large clade (phaseoloids) contains most of the tribe Phaseoleae as well as the tribes Desmodieae and Psoraleeae and is dominated by $2n = 20$ or 22 counts. Polyploidy is more frequent in the phaseoloid clade (17 %), but still less than half as common as in Hologalegina (Table 9.1). Within the phaseoloid clade,

chromosome numbers suggest that several genera are polyploid. The best known of these is *Glycine*, with around 30 species whose lowest chromosome numbers are $2n = 38$ and 40, in contrast to most of its phylogenetic neighbors (e.g., Doyle et al. 2003; Stefanovic et al. 2009) which are typical millettioids with $2n = 20$ or 22. As noted above, genomic data confirm the presence of two cycles of polyploidy in *G. max* (soybean) since the origin of the legumes. The more recent of these has resulted in homoeologous gene pairs that diverged around 10 MYA (Shoemaker et al. 2006; Egan and Doyle 2010), setting a maximum date for the polyploidy event, with the minimum date set by the earliest divergence of the various *Glycine* species around 5 MYA (Innes et al. 2008; Doyle and Egan 2010). Phylogenetic evidence is consistent either with autopolyploidy or with allopolyploidy from extinct species more closely related to one another than to any extant genera outside of *Glycine* (Straub et al. 2006). The presence of two classes of centromeric heterochromatin repeats suggests that *Glycine* could be an allopolyploid, with the two repeat types each derived from one of the diploid progenitor species (Gill et al. 2009). Such a hypothesis is difficult to test due to the extensive rearrangement of homoeologous segments in the soybean genome (Schmutz et al. 2010) and also requires complex patterns of concerted evolution among repeats on different chromosomes.

At least three other phaseoloid genera are likely to be exclusively polyploid based on chromosome number alone (Lackey 1981). In *Erythrina* (coral bean), all sampled species are $2n = 42$. The single count in IPCN for the small genus *Cologania* is $2n = 44$, and the monotypic *Teyleria* is also $2n = 44$ (Kumar and Hymowitz 1989). Goldblatt (1981) considered *Calopogonium*, with counts of $2n = 36$ and ca. 37 in *C. mucunoides*, to be a polyploid, presumably with aneuploid reduction from a base of $x = 10$; however, Gill and Husaini (1986) reported a count of $2n = 24$, which could suggest a more recent derivation of polyploidy within the genus. Similarly, counts of $2n = 28$—considered polyploid in *T. mollis* by Kumari and Bir (1990)—predominate in *Teramnus* species, though *T. labialis* is variously listed as $2n = 20$, 22, and 28. *Strongylodon* is also $2n = 28$.

Polyploidy appears to be rare within Phaseoleae genera. Even relatively large and well-studied genera such as *Rhynchosia* (ca. 230 species), *Phaseolus* (60–65 species), and *Vigna* (ca. 100 species) were reported in Polhill and Raven (1981) as being exclusively diploid, though Sen and Bhattacharya (1988) later reported a count of $2n = 44$ in *V. glabrescens*. Polyploidy has also been reported within species of *Amphicarpaea* and *Neonotonia* by Kumar and Hymowitz (1989; both $2n = 22$, 44). *Apios americana* includes both diploid and triploid cytotypes ($2n = 22$, 33); Joly and Bruneau (2004) reported multiple origins of autotriploidy and high heterozygosity in this species. *Glycine* not only is a relatively recent polyploid at the generic level (see above) but also includes several recently formed allopolyploid species whose genomic relationships to extant diploids have been worked out using molecular phylogenies (reviewed by Doyle et al. 2004), and which are the focus of physiological and transcriptomic studies (Coate and Doyle 2010; Ilut et al. (in press)).

Polyploidy occurs within at least one genus of Desmodieae, *Lespedeza*, which was listed in Polhill and Raven (1981) as $2n = 18, 20, 22, 36$. However, IPCN gives higher numbers, for example *L. bicolor* with both $2n = 22$ and 42 cytotypes, as well as *L. daurica* and *L. potaninii*, both exclusively $2n = 42$. Triploidy occurs in *Campylotropis polyantha* var. *leiocarpa* ($2n = 22, 33$) and possibly in the genus *Pseudarthria*, listed as $2n = 22, 26, 34$. Only diploid counts ($2n = 20, 22$) were reported by Ohashi et al. (1981) from the large (ca. 275 spp.) genus, *Desmodium*. However, additional counts are found in IPCN, both at the diploid ($2n = 24, 26$) and tetraploid levels, the latter in *D. styracifolium* ($2n = 42$) and *D. incanum* ($2n = 22, 44$).

9.5 Searching for Cryptic Polyploidy in the Phaseoloid Legumes

Clearly, all papilionoids are fundamentally polyploid, even those with low chromosome numbers. The tempo and mechanism(s) of chromosomal diploidization are unknown (Doyle et al. 2008; Soltis et al. 2010), and without that information it is difficult to estimate the prevalence of cryptic polyploidy. As noted above, consideration of the divergence times for major lineages suggests that the rate of chromosomal diploidization is rapid—likely 10 MY or less.

On the other hand, "polyploid" chromosome numbers have persisted for at least 5–10 MY in *Glycine* (Fig. 9.2). A cryptic polyploid papilionoid legume is a taxon that has experienced an additional polyploidy event subsequent to the ca. 50 MYA papilionoid WGD but has a low chromosome number typical of its clade. Thus, in the phaseoloid clade, cryptic polyploids would have chromosome numbers of $2n = 20$ or 22. Polyploids on the way to diploidization would have chromosome numbers between these numbers and $2n = 40-44$. As noted above, *Calopogonium* and *Teramnus* are candidates for this class; one perennial *Glycine* species with $2n = 38$ is likely at the first stages of this process.

We know from the paralog K_s profile of *Glycine* (Schleueter et al. 2004) that no additional polyploidy events took place between the two WGD episodes detectable in its genome. Therefore, we can infer that all of the ancestors of *Glycine* experienced only the papilionoid WGD; this includes the ancestors that form the backbone of the phaseoloid clade (Fig. 9.4), as well as the common ancestor of phaseoloids and Indigofereae, and also its common ancestor with the IRLC clade. Given these conditions, candidates for cryptic polyploidy include lineages that are connected to the phaseoloid backbone by branches longer than 5–10 MY. This includes several major groups, such as subtribes Phaseolinae and Cajaninae (Fig. 9.4). Initial sampling of one species of each lineage would provide information on another set of ancestors, and subsequent searches can then be focused on lineages connected to these ancestors by suitably long branches. This has now been done for *Cajanus cajan* (pigeonpea), which Varshney et al. (2011) have shown has no history of recent polyploidy. We are sampling other phaseoloids using 454

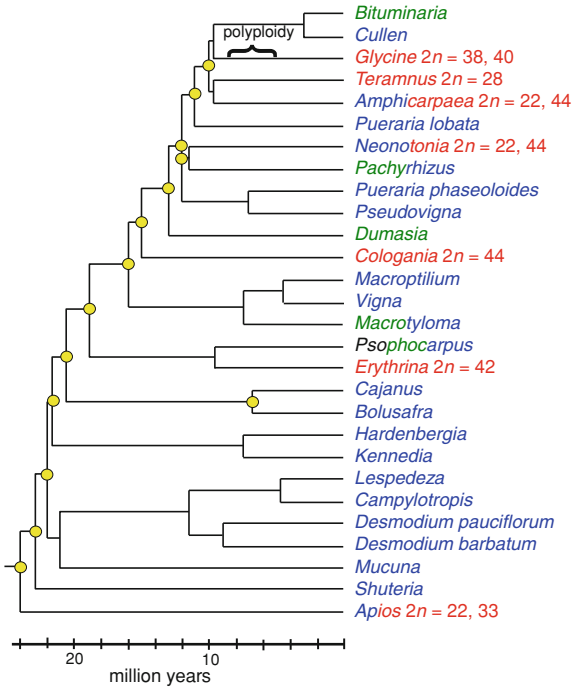

Fig. 9.4 Phylogeny and polyploidy in the phaseoloid clade. Topology and dates of the chronogram are taken from Stefanovic et al. (2009). Chromosome numbers of the species used in that study are color coded as follows: *Green*, $2n = 20$; *blue*, $2n = 22$; *black*, $2n = 18$; *red*, known or possible polyploid numbers, with the numbers shown following the taxon name. Species known to have multiple cytotypes are indicated by multiple colors corresponding to chromosome numbers listed above. The two numbers shown for *Glycine* are from different species, only one of which (*G. max*, $2n = 40$) was used in the Stefanovic et al. (2009) study. The range of dates for the polyploidy event in *Glycine* is indicated. *Yellow dots* indicate ancestral nodes that lacked any polyploidy event subsequent to the papilionoid WGD

transcriptome sequencing, including taxa with both low and high chromosome numbers; for the latter we wish to estimate maximum ages of polyploidy. Thus far we have not found examples of cryptic polyploidy, but have determined that polyploidy in *Erythrina* probably took place on roughly the same time scale as in *Glycine* (<10 MYA; Egan and Doyle, unpublished data).

9.6 Conclusions

Among the most persistent questions concerning polyploidy in plants are how successful the phenomenon is as an evolutionary mechanism. Is polyploidy a ticket to innovation, adaptation, invasiveness, survival in the face of global catastrophes,

or is it an evolutionary dead end … or both? Based on rates of polyploid formation and extinction in the phylogenetic record, Mayrose et al. (2011) conclude that "polyploidy is most often an evolutionary dead end, but the possibility remains that the expanded genomic potential of those polyploids that do persist drives long-term evolutionary success."

Legumes may illustrate both of these points. The most diverse and species-rich clade of this third largest family of flowering plants, the core papilionoids, is ancestrally polyploid. Clearly, the ancestor of this group of around 450 genera and 13,000 species, like the ancestor of seed plants and the ancestor of flowering plants (Jiao et al. 2011), was most emphatically not a "dead end." It remains to be determined whether there is a perfect correlation between the papilionoid polyploidy event and the origin of nodulation in core papilionoids, and it will take much more work to demonstrate that the two are causally related (Doyle 2011; Young et al. 2011). It is also clear that nodulation is not sufficient to explain the explosive radiation of papilionoid legumes, because other nodulating groups both in legumes and elsewhere in the rosids have not proliferated to the same extent as papilionoids (Doyle 2011).

Despite the obvious success of the core papilionoid lineage, the pattern of evolution within the core papilionoids suggests that polyploidy has not been a major feature in establishing new lineages, similar to the conclusion of Mayrose et al. (2011) for angiosperms generally. It is not that polyploidy is rare within the family—indeed, around a quarter of all legumes for which chromosome data are available have one or more species that are polyploid (Table 9.1). However, much of the polyploidy in the family occurs as single polyploid genera embedded within diploids, as scattered species within genera, or as multiple cytotypes within species. Two significant exceptions are the Genisteae, which may be entirely polyploid and within which nearly all genera show a propensity for polyploidy and aneuploidy, and the lineage that includes the IRLC tribe Carmichaelieae. The largest papilionoid genera, including *Astragalus*, are not fundamentally polyploid.

"Success" is a very ambiguous term and can be measured in many ways. Species with short evolutionary histories that have not been involved in subsequent speciation, yet have invaded extensive new territories and had major impact on the environment, certainly could be considered "successful." Many plant polyploids, including genera and species of legumes, fit this description. So does *Homo sapiens*.

Acknowledgments I thank many colleagues and lab members for discussions of polyploidy, and Jane Doyle for her support and encouragement. I also thank Jane Doyle, Sue Sherman-Broyles, Iben Sorensen, and Toby Pennington for critical reading of the manuscript, and Melissa Luckow for help with mimosoid systematics. I am grateful for many years of funding from the National Science Foundation for work on polyploidy, most recently grants DEB-0948800, IOS-0939423, IOS-0822258, and IOS-0744306. I thank Doug Soltis for helpful suggestions in review of the manuscript.

References

Ahangarian S, Osaloo SK, Maassoumi AA (2007) Molecular phylogeny of the tribe *Hedysareae* with special reference to *Onobrychis* (*Fabaceae*) as inferred from nrDNA ITS sequences. Iran J Bot 13:64–74

Ainouche A, Bayer RJ, Misset M- (2004) Molecular phylogeny, diversification and character evolution in *Lupinus* (*Fabaceae*) with special attention to Mediterranean and African lupines. Plant Syst Evol 246:211–222

Artyukova EV, Kozyrenko MM, Kholina AB, Zhuravlev YN (2011) High chloroplast haplotype diversity in the endemic legume *Oxytropis* chankaensis may result from independent polyploidization events. Genetica 139:221–232

Bell CD, Soltis DE, Soltis PS (2010) The age and diversification of the angiosperms re-revisited. Am J Bot 97:296–313

Bello MA, Bruneau A, Forest F, Hawkins JA (2009) Elusive relationships within order *Fabales*: phylogenetic analyses using matK and rbcL sequence data. Syst Bot 34:102–114

Bertioli D, Moretzsohn M, Madsen L, Sandal N, Leal-Bertioli S, Guimaraes P, Hougaard B, Fredslund J, Schauser L, Nielsen A, Sato S, Tabata S, Cannon S, Stougaard J (2009) An analysis of synteny of *Arachis* with *Lotus* and *Medicago* sheds new light on the structure, stability and evolution of legume genomes. BMC Genomics 10:45

Bessega C, Vilardi JC, Saidman BO (2006) Genetic relationships among American species of the genus *Prosopis* (*Mimosoideae*, *Leguminosae*) inferred from ITS sequences: evidence for long-distance dispersal. J Biogeogr 33:1905–1915

Bingham ET (1972) Sexual poly ploidy in *Medicago-Sativa-D*. Genetics 71:S5

Bisby FA (1981) Genisteae. In: Polhill RM, Raven PH (eds) Advances in legume systematics, Part 1. Royal Botanic Gardens, Kew, pp 409–425

Blanc G, Wolfe KH (2004) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell 16:1667–1678

Boatwright JS, Van Wyk B- (2011) The systematic position of *Sophora inhambanensis* (*Fabaceae*: *Sophoreae*). S Afr J Bot 77:249–250

Boatwright JS, Savolainen V, Van Wyk B, Schutte-Vlok AL, Forest F, van der Bank M (2008) Systematic position of the anomalous genus Cadia and the phylogeny of the tribe *Podalyrieae* (*Fabaceae*). Syst Bot 33:133–147

Boff T, Schifino-Wittmann MT (2003) Segmental allopolyploidy and paleopolyploidy in species of *Leucaena Benth*: evidence from meiotic behaviour analysis. Hereditas (Lund) 138:27–35

Brown GK, Clowes C, Murphy DJ, Ladiges PY (2010) Phylogenetic analysis based on nuclear DNA and morphology defines a clade of eastern Australian species of *Acacia* s.s. (section *Juliflorae*): the '*Acacia longifolia* group'. Aust Syst Bot 23:162–172

Bruneau A, Mercure M, Lewis GP, Herendeen PS (2008) Phylogenetic patterns and diversification in the caesalpinioid legumes. Botany-Botanique 86:697–718

Burow MD, Simpson CE, Faries MW, Starr JL, Paterson AH (2009) Molecular biogeographic study of recently described B- and A-genome *Arachis* species, also providing new insights into the origins of cultivated peanut. Genome 52:107–119

Calderini O, Mariani A (1997) Increasing 2n gamete production in diploid alfalfa by cycles of phenotypic recurrent selection. Euphytica 93:113–118

Cannon SB, Ilut D, Farmer AD, Maki SL, May GD, Singer SR, Doyle JJ (2010) Polyploidy did not predate the evolution of nodulation in all legumes. PLoS ONE 5:e11630

Catalano SA, Vilardi JC, Tosto D, Saidman BO (2008) Molecular phylogeny and diversification history of *Prosopis* (*Fabaceae*: *Mimosoideae*). Biol J Linn Soc 93:621–640

Chandler GT, Bayer RJ, Crisp MD (2001) A molecular phylogeny of the endemic Australian genus *Gastrolobium* (*Fabaceae*: *Mirbelieae*) and allied genera using chloroplast and nuclear markers. Am J Bot 88:1675–1687

Choi B, Kim J (1997) ITS sequences and speciation on far eastern *Indigofera* (*Leguminosae*). J Plant Res 110:339–346

Chooi WY (1971) Variation in nuclear DNA content in the genus *Vicia-D*. Genetics 68:195–211
Coate JE, Doyle JJ (2010) Quantifying whole transcriptome size, a prerequisite for understanding transcriptome evolution across species: an example from a plant allopolyploid. Genome Biol Evol 2:534–546
Conterato IF, Schifino-Wittmann MT (2006) New chromosome numbers, meiotic behaviour and pollen fertility in American taxa of *Lupinus* (*Leguminosae*): contributions to taxonomic and evolutionary studies. Bot J Linn Soc 150:229–240
Cowan RS, Polhill RM (1981) Amherstieae. In: Polhill RM, Raven PH (eds) Advances in legume systematics, Part 1. Royal Botanic Gardens, Kew, pp 135–142
Cubas P, Pardo C, Tahiri H (2002) Molecular approach to the phylogeny and systematics of cytisus (*Leguminosae*) and related genera based on nucleotide sequences of nrDNA (ITS region) and cpDNA (trnL-trnF intergenic spacer). Plant Syst Evol 233:223–242
Cusma-Velari T, Feoli-Chiapella L (2009) The so-called primitive genera of *Genisteae* (*Fabaceae*): systematic and phyletic considerations based on karyological data. Bot J Linn Soc 160:232–248
Dahmer N, Simon MF, Schifino-Wittmann MT, Hughes CE, Sfoggia Miotto ST, Giuliani JC (2011) Chromosome numbers in the genus *Mimosa L.*: cytotaxonomic and evolutionary implications. Plant Syst Evol 291:211–220
Degtjareva GV, Kramina TE, Sokoloff DD, Samigullin TH, Valiejo-Roman CM, Antonov AS (2006) Phylogeny of the genus *Lotus* (*Leguminosae*, *Loteae*): evidence from nrITS sequences and morphology. Can J Bot 84:813–830
Doyle JJ (2011) Phylogenetic perspectives on the origins of nodulation. Molec Plant Microbe Interact 24:1289–1295
Doyle JL, Rauscher JT, Brown AHD (2004) Diploid and polyploid reticulate evolution throughout the history of the perennial soybeans (Glycine …. New Phytologist)
Doyle JJ, Egan AN (2010) Dating the origins of polyploidy events. New Phytol 186:73–85
Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF (2008) Evolutionary genetics of genome merger and doubling in plants. Annu Rev Genet 42:443–461
Doyle JJ, Doyle JL, Harbison C (2003) Chloroplast-expressed glutamine synthetase in *Glycine* and related *Leguminosae*: phylogeny, gene duplication, and ancient polyploidy. Syst Bot 28:567–577
Doyle JJ, Luckow MA (2003) The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. Plant Physiol (Rockville) 131:900–910
Drummond CS (2008) Diversification of *Lupinus* (*Leguminosae*) in the western new world: derived evolution of perennial life history and colonization of montane habitats. Mol Phylogenet Evol 48:408–421
Drummond CS, Eastwood RJ, Miotto STS, Hughes CE (2012) Multiple continental radiations and correlates of diversification in lupinus (Leguminosae): testing for key innovation with incomplete taxon sampling. Syst Biol 61:443–460
Egan AN, Doyle J (2010) A comparison of global, gene-specific, and relaxed clock methods in a comparative genomics framework: dating the polyploid history of soybean (*Glycine max*). Syst Biol 59:534–547
Ellison NW, Liston A, Steiner JJ, Williams WM, Taylor NL (2006) Molecular phylogenetics of the clover genus (*Trifolium—Leguminosae*). Mol Phylogenet Evol 39:688–705
Endo Y, Choi B, Ohashi H, Delgado-Salinas A (2008) Phylogenetic relationships of *New World Vicia* (*Leguminosae*) inferred from nrDNA internal transcribed spacer sequences and floral characters. Syst Bot 33:356–363
Fawcett JA, Maere S, Van de Peer Y (2009) Plants with double genomes might have had a better chance to survive the cretaceous-tertiary extinction event. Proc Natl Acad Sci U S A 106:5737–5742
Frahm-Leliveld JA (1966) Cytotaxonomic notes on the genera *Indigofera L.* and *Cyamopsis DC*. [*Leguminosae*]. Genetica 37:403–426
Freeling M, Thomas BC (2006) Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. Genome Res 16:805–814

Gallagher RV, Leishman MR, Miller JT, Hui C, Richardson DM, Suda J, Travnicek P (2011) Invasiveness in introduced Australian *Acacias*: the role of species traits and geneome size. Divers Distrib 17:884–897

Gauthier P, Lumaret R, Bedecarrats A (1998a) Genetic variation and gene flow in Alpine diploid and tetraploid populations of *Lotus* (*L. alpinus* (*D.C.*) Schleicher/*L. corniculatus L.*). I. Insights from morphological and allozyme markers. Heredity 80:683–693

Gauthier P, Lumaret R, Bedecarrats A (1998b) Genetic variation and gene flow in Alpine diploid and tetraploid populations of *Lotus* (*L. alpinus* (*D.C.*) Schleicher/*L. corniculatus L.*). II. Insights from RFLP of chloroplast DNA. Heredity 80:694–701

Gill LS, Husaini (1986) Cytological observations in *Leguminosae* from southern Nigeria. Willdenowia 15:521–527

Gill N, Findley S, Walling JG, Hans C, Ma J, Doyle J, Stacey G, Jackson SA (2009) Molecular and chromosomal evidence for allopolyploidy in soybean. Plant Physiol 151:1167–1174

Gohil RN, Ashraf M (2008) Cytological parameters viz a viz probable modes of evolution in *Astragalus L.* Proc Nat Acad Sci India Sect B-Biol Sci 78:281–287

Goldblatt P (1989) Miscellaneous chromosome counts in *Asteraceae Bignoniaceae Proteaceae* and *Fabaceae*. Ann Mo Bot Gard 76:1186–1188

Goldblatt P (1981) Cytology and the phylogeny of leguminosae. In: Polhill RM, Raven PH (eds) Advances in legume systematics, Part 2. Royal Botanic Gardens, Kew, pp 427–464

Govindarajulu R, Hughes CE, Bailey D (2011a) Phylogenetic and population genetic analyses of diploid *Leucaena* (*Leguminosae-Mimosoideae*) reveal cryptic species diversity and patterns of divergent allopatric speciation. Am J Bot 98:2049–2063

Govindarajulu R, Hughes CE, Alexander P, Bailey D (2011b) The complex dynamics of ancient and recent polyploidy in *Leucaena* (*Leguminosae*). Am J Bot 98:2064–2076

Grant WF, Small E (1996) The origin of the *Lotus corniculatus* (*Fabaceae*) complex: a synthesis of diverse evidence. Can J Bot 74:975–989

Gutierrez JF, Vaquero F, Vences FJ (1994) Allopolyploid vs. autopolyploid origins in the genus *Lathyrus* (*Leguminosae*). Heredity 73:29–40

Hanson L (1995) Some new chromosome counts in the genus *Inga* (*Leguminosae*: *Mimosoideae*). Kew Bull 50:801–804

Havananda T, Brummer EC, Maureira-Butler IJ, Doyle JJ (2010) Relationships among diploid members of the *Medicago sativa* (*Fabaceae*) species complex based on chloroplast and mitochondrial DNA sequences. Syst Bot 35:140–150

Havananda T, Brummer EC, Doyle JJ (2011) Complex patterns of autopolyploid evolution in alfalfa and allies (*Medicago sativa*: Leguminosae). Am J Bot 98:1633-1646

Hejazi H, Mohsen S, Nasab MZ (2010) Cytotaxonomy of some *Onobrychis* (*Fabaceae*) species and populations in Iran. Caryologia 63:18–31

Hughes CE, Eastwood R (2006) Island radiation on a continental scale: exceptional rates of plant diversification after uplift of the Andes. Proc Natl Acad Sci U S A 103:10334–10339

Hughes CE, Govindarajulu R, Robertson A, Filer DL, Harris SA, Bailey CD (2007) Serendipitous backyard hybridization and the origin of crops. Proc Natl Acad Sci U S A 104:14389–14394

Hughes CE, Bailey CD, Harris SA (2002) Divergent and reticulate species relationships in *Leucaena* (*Fabaceae*) inferred from multiple data sources: insights into polyploid origins and nrDNA polymorphism. Am J Bot 89:1057–1073

Hughes CE, Bailey CD, Krosnick S, Luckow MA (2003) Relationships among genera of the informal *Dichrostachys* and *Leucaena* groups (*Mimosoideae*) inferred from nuclear ribosomal ITS sequences. In: Klitgaard B, Bruneau A (eds) Advances in legume systematics, Part 1.0. Royal Botanic Gardens, Kew, pp 221–238

Hulina N (2010) "Planta hortifuga" in flora of the continental part of Croatia. Agriculturae Conspectus Scientificus 75:57–65

Ilut DC, Coate JE, Luciano AK, Owens TG, May GD, Farmer A, Doyle JJ (2012) A comparative transcriptomic study of an allotetraploid and its diploid progenitors illustrates the unique advantages and challenges of RNA-Seq in plant species. Am J Bot 99:383–396

Innes RW, Ameline-Torregrosa C, Ashfield T, Cannon E, Cannon SB, Chacko B, Chen NWG, Couloux A, Dalwani A, Denny R, Deshpande S, Egan AN, Glover N, Hans CS, Howell S, Ilut D, Jackson S, Lai H, Mammadov J, del Campo SM, Metcalf M, Nguyen A, O'Bleness M, Pfeil BE, Podicheti R, Ratnaparkhe MB, Samain S, Sanders I, Segurens B, Sevignac M, Sherman-Broyles S, Thareau V, Tucker DM, Walling J, Wawrzynski A, Yi J, Doyle JJ, Geffroy V, Roe BA, Maroof MAS, Young ND (2008) Differential accumulation of retroelements and diversification of NB-LRR disease resistance genes in duplicated regions following polyploidy in the ancestor of soybean. Plant Physiol (Rockville) 148:1740–1759

Jaillon O, Aury J, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, Vezzi A, Legeai F, Hugueney P, Dasilva C, Horner D, Mica E, Jublot D, Poulain J, Bruyere C, Billault A, Segurens B, Gouyvenoux M, Ugarte E, Cattonaro F, Anthouard V, Vico V, Del Fabbro C, Alaux M, Di Gaspero G, Dumas V, Felice N, Paillard S, Juman I, Moroldo M, Scalabrin S, Canaguier A, Le Clainche I, Malacrida G, Durand E, Pesole G, Laucou V, Chatelet P, Merdinoglu D, Delledonne M, Pezzotti M, Lecharny A, Scarpelli C, Artiguenave F, Pe ME, Valle G, Morgante M, Caboche M, Adam-Blondon A, Weissenbach J, Quetier F, Wincker P, French-Italian Public (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature (London) 449:463

Jenczewski E, Prosperi J, Ronfort J (1999) Evidence for gene flow between wild and cultivated *Medicago sativa* (*Leguminosae*) based on allozyme markers and quantitative traits. Am J Bot 86:677–687

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, Soltis DE, Clifton SW, Schlarbaum SE, Schuster SC, Ma H, Leebens-Mack J, dePamphilis CW (2011) Ancestral polyploidy in seed plants and angiosperms. Nature 473:97–100

Joly S, Bruneau A (2004) Evolution of triploidy in *Apios americana* (*Leguminosae*) revealed by genealogical analysis of the histone H3-D gene. Evolution 58:284–295

Jorgensen JL, Stehlik I, Brochmann C, Conti E (2003) Implications of ITS sequences and RAPD markers for the taxonomy and biogeography of the *Oxytropis campestris* and *O. arctica* (*Fabaceae*) complexes in Alaska. Am J Bot 90:1470–1480

Kajita T, Ohashi H, Tateishi Y, Bailey CD, Doyle JJ (2001) rbcL and legume phylogeny, with particular reference to phaseoleae, millettieae, and allies. Syst Bot 26:515–536

Kloda JM, Dean PDG, Maddren C, MacDonald DW, Mayes S (2008) Using principle component analysis to compare genetic diversity across polyploidy levels within plant complexes: an example from British *Restharrows* (*Ononis spinosa* and *Ononis repens*). Heredity 100:253–260

Kumar PS, Hymowitz T (1989) Where are the diploid 2n equals 2x equals 20 genome donors of glycine willd. *Leguminosae Papilionoideae*. Euphytica 40:221–226

Kumari S, Bir SS (1990) Karyomorphological evolution in *Papilionaceae*. J Cytol Genet 25:173–219

Kupicha FK (1981) Vicieae. In: Polhill RM, Raven PH (eds) Advances in legume systematics, Part 1. Royal Botanic Gardens, Kew, pp 377–381

Lackey JA (1981) Phaseoleae. In: Polhill RM, Raven PH (eds) Advances in legume systematics, Part 1. Royal Botanic Gardens, Kew, pp 301–328

Lavia GI, Ortiz MA, Robledo G, Fernandez A, Seijo G (2011) Origin of triploid *Arachis pintoi* (*Leguminosae*) by autopolyploidy evidenced by FISH and meiotic behaviour. Ann Bot (London) 108:103–111

Lavin M, Herendeen PS, Wojciechowski MF (2005) Evolutionary rates analysis of *Leguminosae* implicates a rapid diversification of lineages during the tertiary. Syst Biol 54:575–594

Lavin M, Pennington RT, Klitgaard BB, Sprent JI, de Lima HC, Gasson PE (2001) The dalbergioid legumes (*Fabaceae*): delimitation of a pantropical monophyletic clade. Am J Bot 88:503–533

Leitch IJ, Bennett MD (2004) Genome downsizing in polyploid plants. Biol J Linn Soc 82:651–663

Lewis GP, Schrire B, MacKinder B, Lock M (2005) Legumes of the world. Royal Botanic Gardens, Kew

Luckow M, Fortunato RH, Sede S, Livshultz T (2005) The phylogenetic affinities of two mysterious monotypic mimosoids from southern South America. Syst Bot 30:585–602

Luckow M, Miller JT, Murphy DJ, Livshultz T (2003) A phylogenetic analysis of the *Mimosoideae* (*Leguminosae*) based on chloroplast DNA sequence data. In: Klitgaard B, Bruneau A (eds) Advances in legume systematics, Part 1.0. Royal Botanic Gardens, Kew, pp 197–220

Lynch M, Conery JS (2003) The origins of genome complexity. Science 302:1401–1404

Mayrose I, Zhan SH, Rothfels CJ, Magnuson-Ford K, Barker MS, Rieseberg LH, Otto SP (2011) Recently formed polyploid plants diversify at lower rates. Science (Washington DC) 333:1257

McMahon MM (2005) Phylogenetic relationships and floral evolution in the papilionoid legume clade *Amorpheae*. Brittonia 57:397–411

Morales M, Wulff AF, Fortunato RH, Poggio L (2010) Chromosome and morphological studies in the *Mimosa debilis* complex (*Mimosoideae*, *Leguminosae*) from southern South America. Aust J Bot 58:12–22

Murphy DJ, Brown GK, Miller JT, Ladiges PY (2010) Molecular phylogeny of Acacia Mill. (*Mimosoideae*: *Leguminosae*): evidence for major clades and informal classification. Taxon 59:7–19

Neumann P, Koblizkova A, Navratilova A, Macas J (2006) Significant expansion of *Vicia pannonica* genome size mediated by amplification of a single type of giant retroelement. Genetics 173:1047–1056

Ohashi H, Polhill RM, Schubert BG (1981) Desmodieae. In: Polhill RM, Raven PH (eds) Advances in legume systematics, Part 1. Royal Botanic Gardens, Kew, pp 292–300

Orthia LA, Cook LG, Crisp MD (2005) Generic delimitation and phylogenetic uncertainty: an example from a group that has undergone an explosive radiation. Aust Syst Bot 18:41–47

Ortiz MA, Guillermo Seijo J, Fernandez A, Lavia GI (2011) Meiotic behavior and pollen viability of tetraploid *Arachis glabrata* and *A. nitida* species (section *Rhizomatosae*, *Leguminosae*): implications concerning their polyploid nature and seed set production. Plant Syst Evol 292:73–83

Ossowski S, Schneeberger K, Lucas-Lledo JI, Warthmann N, Clark RM, Shaw RG, Weigel D, Lynch M (2010) The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana* RID E-2139-2011 RID C-1418-2008. Science 327:92–94

Pandit MK, Tan HTW, Bisht MS (2006) Polyploidy in invasive plant species of Singapore. Bot J Linn Soc 151:395–403

Pandit MK, Pocock MJO, Kunin WE (2011) Ploidy influences rarity and invasiveness in plants. J Ecol 99: 1108-1115

Pardo C, Cubas P, Tahiri H (2004) Molecular phylogeny and systematics of *Genista* (*Leguminosae*) and related genera based on nucleotide sequences of nrDNA (ITS region) and cpDNA (trnL-trnF intergenic spacer). Plant Syst Evol 244:93–119

Pennington, RT, Klitgaard BB, Ireland H, Lavin M (2000) New insights into floral evolution of basal Papilionoideae from molecular phylogenies. In: Herendeen PS Bruneau A (eds.) Advances in legume systematics Part 9. Royal Botanic Gardens, Kew p 233–248

Pennington, RT, Lavin M, Ireland H, Klitgaard BB, Preston J, Hu J-M (2001) Phylogenetic relationships of basal papilionoid legumes based upon sequences of the chloroplast trnL intron. Syst Bot 26:537–556

Pfeil BE, Schlueter JA, Shoemaker RC, Doyle JJ (2005) Placing paleopolyploidy in relation to taxon divergence: a phylogenetic analysis in legumes using 39 gene families. Syst Biol 54:441–454

Polhill RM (1981a) Dipteryxeae. In: Polhill RM, Raven PH (eds) Advances in legume systematics, Part 1. Royal Botanic Gardens, Kew, pp 231–232

Polhill RM (1981b) Indigofereae. In: Polhill RM, Raven PH (eds) Advances in legume systematics, Part 1. Royal Botanic Gardens, Kew, pp 289–291

Polhill RM, Raven PH (1981) Advances in legume systematics, Part 1. Royal Botanic Gardens, Kew

Polhill RM, Sousa M (1981) Robinieae. In: Polhill RM, Raven PH (eds) Advances in legume systematics, Part 1. Royal Botanic Gardens, Kew, pp 283–288

Reddy VRK, Revathi R (1993) Chemotaxonomic studies in the genus *Indigofera* Linn. J Econ Taxon Bot 17:115–120

Rosato M, Castro M, Rossello JA (2008) Relationships of the woody *Medicago* species (section Dendrotelis) assessed by molecular cytogenetic analyses. Ann Bot (London) 102:15–22

Rossello JA, Castro M (2008) Karyological evolution of the angiosperm endemic flora of the Balearic Islands. Taxon 57:259–273

Sakiroglu M, Doyle JJ, Brummer EC (2010) Inferring population structure and genetic diversity of broad range of wild diploid alfalfa (*Medicago sativa L.*) accessions using SSR markers. Theor Appl Genet 121:403–415

Sato S, Nakamura Y, Kaneko T, Asamizu E, Kato T, Nakao M, Sasamoto S, Watanabe A, Ono A, Kawashima K, Fujishiro T, Katoh M, Kohara M, Kishida Y, Minami C, Nakayama S, Nakazaki N, Shimizu Y, Shinpo S, Takahashi C, Wada T, Yamada M, Ohmido N, Hayashi M, Fukui K, Baba T, Nakamichi T, Mori H, Tabata S (2008) Genome structure of the legume, *Lotus japonicus*. DNA Res 15:227–239

Schleueter J, Dixon P, Granger C, Grant D, Clark L, Doyle JJ, Shoemaker RC (2004) Mining EST databases to resolve evolutionary events in major crop species. Genome 47:868–876

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the paleopolyploid soybean. Nature 463:178–183

Schrire BD, Lavin M, Barker NP, Forest F (2009) Phylogeny of the *Tribe Indigofereae* (*Leguminosae-Papilionoideae*): geographically structured more in succulent-rich and temperate settings than in grass-rich environments. Am J Bot 96:816–843, 844–852

Seijo G, Lavia GI, Fernandez A, Krapovickas A, Ducasse DA, Bertioli DJ, Moscone EA (2007) Genomic relationships between the cultivated peanut (*Arachis hypogaea*, *Leguminosae*) and its close relatives revealed by double GISH. Am J Bot 94:1963–1971

Seijo G, Fernandez A (2001) Chromosome numbers of some southernmost species of *Mimosa L.* (*Leguminosae*). Cytologia 66:19–23

Sen O, Bhattacharya S (1988) Cytomixis in vigna-glabrescens Ttk-1 wild. Cytologia 53:437–440

Shoemaker RC, Polzin K, Labate J, Specht J, Brummer EC, Olson T, Young N, Concibido V, Wilcox J, Tamulonis JP, Kochert G, Boerma HR (1996) Genome duplication in soybean (*Glycine* subgenus soja). Genetics 144:329–338

Shoemaker RC, Schlueter J, Doyle JJ (2006) Paleopolyploidy and gene duplication in soybean and other legumes. Curr Opin Plant Biol 9:104–109

Simon MF, Grether R, de Queiroz LP, Saerkinen TE, Dutra VF, Hughes CE (2011) The evolutionary history of *Mimosa* (*Leguminosae*): toward a phylogeny of the sensitive plants. Am J Bot 98:1201–1221

Singer SR, Maki SL, Farmer AD, Ilut D, May GD, Cannon SB, Doyle JJ (2009) Venturing beyond beans and peas: what can we learn from chamaecrista? Plant Physiol 151:1041–1047

Sinou C, Forest F, Lewis GP, Bruneau A (2009) The genus *Bauhinia* s.l. (*Leguminosae*): a phylogeny based on the plastid trnL-trnF region. Botany-Botanique 87:947–960

Soltis DE, Buggs RJA, Doyle JJ, Soltis PS (2010) What we still don't know about polyploidy. Taxon 59:1387–1403

Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, dePamphilis CW, Wall PK, Soltis PS (2009) Polyploidy and angiosperm diversification. Am J Bot 96:336–348

Spellenberg R (1981) Poly ploidy in *Dalea*-Formosa *Fabaceae* on the Chihuahuan desert. Brittonia 33:309–324
Sprent JI (2009) Legume nodulation: a global perspective. Wiley-Blackwell, Ames
Srivastav PK, Raina SN (1986) Cytogenetics of *Tephrosia* Vi. Meiotic systems in some taxa. Cytologia 51:359–374
Steele KP, Ickert-Bond SM, Zarre S, Wojciechowski MF (2010) Phylogeny and character evolution in *Medicago* (*Leguminosae*): evidence from analyses of plastid Trnk/matk and nuclear Ga3ox1 sequences. Am J Bot 97:1142–1155
Stefanovic S, Pfeil BE, Palmer JD, Doyle JJ (2009) Relationships among phaseoloid legumes based on sequences from eight chloroplast regions. Syst Bot 34:115–128
Straub SCK, Doyle JJ (2009) Conservation genetics of *Amorpha georgiana* (*Fabaceae*), an endangered legume of the Southeastern United States. Mol Ecol 18:4349–4365
Straub SCK, Pfeil BE, Doyle JJ (2006) Testing the polyploid past of soybean using a low-copy nuclear gene—is *Glycine* (*Fabaceae*: *Papilionoideae*) an auto- or allopolyploid? Mol Phylogenet Evol 39:580–584
te Beest M, Le Roux JJ, Richardson DM, Brysting AK, Suda J, Kubešová M, Pyšek P (2011) The more the better? The role of polyploidy in facilitating plant invasions. Ann Bot 109:19–45
Thulin M, Lavin M (2001) Phylogeny and biogeography of the *Ormocarpum* group (*Fabaceae*): a new genus *Zygocarpum* from the horn of Africa region. Syst Bot 26:299–317
Tondini F, Tavoletti S, Mariani A, Veronesi F (1993) A statistical approach to estimate the frequency of n, 2n and 4n pollen grains in diploid alfalfa. Euphytica 69:109–114
Torres DC, Matos Santos Lima JP, Fernandes AG, Nunes EP, Grangeiro TB (2011) Phylogenetic relationships within chamaecrista sect. Xerocalyx (*Leguminosae*, *Caesalpinioideae*) inferred from the cpDNA trnE- trnT intergenic spacer and nrDNA ITS sequences. Genet Mol Biol 34:244–251
Travnicek P, Eliasova A, Suda J (2010) The distribution of cytotypes of *Vicia cracca* in Central Europe: the changes that have occurred over the last four decades. Preslia (Prague) 82:149–163
Turini FG, Braeuchler C, Heubl G (2010) Phylogenetic relationships and evolution of morphological characters in *Ononis L.* (*Fabaceae*). Taxon 59:1077–1090
van Wyk BE, Schutte AL (1988) Chromosome numbers in *Lotononis* and *Buchenroedera* (*Fabaceae-Crotalarieae*). Ann Missouri Bot Gard 75:1603–1607
Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, Schlueter JA et al (2011) Draft genome sequence of *Pigeonpea* (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. Nat Biotech. doi:10.1038/nbt.2022
Veronesi F, Mariani A, Bingham ET (1986) Unreduced gametes in diploid medicago and their importance in alfalfa breeding. Theor Appl Genet 72:37–41
Wagstaff S, Heenan P, Sanderson M (1999) Classification, origins, and patterns of diversification in New Zealand *Carmichaelinae* (*Fabaceae*). Am J Bot 86:1346–1356
Wang H, Moore MJ, Soltis PS, Bell CD, Brockington SF, Alexandre R, Davis CC, Latvis M, Manchester SR, Soltis DE (2009) Rosid radiation and the rapid rise of angiosperm-dominated forests. Proc Nat Acad Sci 106:3853–3858
Wilbur RL (1975) A revision of the North American genus *Amorpha Leguminosae Psoraleae*. Rhodora 77:337–409
Wojciechowski MF, Lavin M, Sanderson MJ (2004) A phylogeny of legumes (*Legumenosae*) based on analyses of the plastid matK gene resolves many well-supported subclades within the family. Am J Bot 91:1846–1862
Wojciechowski M (2005) *Astragalus* (*Fabaceae*): a molecular phylogenetic perspective. Brittonia 57:382–396
Young N, Debellé F, Oldroyd G, Geurts R, Cannon SB et al (2011) The medicago genome provides insight into the evolution of rhizobial symbioses. Nature. doi:10.1038/nature10625
Zhang M, Fritsch PW, Cruz BC (2009) Phylogeny of *Caragana* (*Fabaceae*) based on DNA sequence data from rbcL, trnS-trnG, and ITS. Mol Phylogenet Evol 50:547–559

Chapter 10
Jeans, Genes, and Genomes: Cotton as a Model for Studying Polyploidy

Jonathan F. Wendel, Lex E. Flagel and Keith L. Adams

Abstract We present an overview of the cotton genus (*Gossypium*) as a model for the study of polyploidy. A synopsis of the origin and evolution of polyploid cotton is provided, offering an organismal framework and phylogenetic perspective that is critical for understanding modes and mechanisms of gene and genome evolution. Sequence data from thousands of genes implicate a mid-Pleistocene (1–2 mya) origin of polyploid cotton, following trans-oceanic dispersal of an Old World, A-genome diploid to the New World and subsequent hybridization with an indigenous D-genome diploid. This chance biological reunion, occurring after 5–10 million years of diploid evolution in isolation, has led to an array of molecular genetic interactions in the newly formed allopolyploid lineage, including nonreciprocal homoeologous recombination and perhaps other forms of interlocus concerted evolution, differential rates of genomic evolution, intergenomic spread of transposable elements, and myriad forms of alterations in duplicate expression relative to that experienced in the ancestral diploids. The latter include developmental, organ-, tissue-, and cell-specific biases in homoeologous gene expression, which can be sensitive to various forms of environmental perturbation and stress. The allopolyploid *Gossypium* transcriptome is exceptionally dynamic, with homoeolog expression ratios being subject to change even during development of the

J. F. Wendel (✉)
Department of Ecology, Evolution, and Organismal Biology,
Iowa State University, Ames, IA 50011, USA
e-mail: jfw@iastate.edu

L. E. Flagel
Department of Biology, Duke University, 90338 Durham, NC 27708, USA
e-mail: lex.flagel@duke.edu

K. L. Adams
Department of Botany, and UBC Botanical Garden and Centre for Plant Research,
University of British Columbia, 6268 University Blvd, Vancouver, BC V6T 1Z4, Canada
e-mail: keitha@mail.ubc.ca

P. S. Soltis and D. E. Soltis (eds.), *Polyploidy and Genome Evolution*,
DOI: 10.1007/978-3-642-31442-1_10,

single-celled cotton fiber. Expression evolution is temporally partitioned into changes accompanying genome merger (hybridization) at the diploid level, polyploidization, and longer term evolution at the allopolyploid level. Evidence indicates that allopolyploidy facilitated colonization of a new ecological niche for the genus and led to an enhanced capacity for developing agronomically superior cotton varieties. The myriad mechanisms that underlie genomic and regulatory evolution are suggested to have contributed to both ecological success and agronomic potential.

10.1 Introduction

Because of its economic importance, the cotton genus (*Gossypium* L.) has long attracted the attention of agricultural scientists, taxonomists, and biologists in multiple disciplines. Accordingly, a great deal is understood about the origin and diversification of the genus, its basic plant biology, and its properties as a crop plant (Paterson 2009; Stewart et al. 2010; Wendel et al. 2009). One of the most salient features of the genus is that its history is so encompassing in scope, involving a global (mostly austral) phylogenetic diversification and a repeated history of trans-oceanic dispersal. Superimposed on this natural diversification has been a many-thousand-year history of human manipulation, tracing to ancient human cultures on several continents who independently domesticated four species, two from the Americas, *G. hirsutum* and *G. barbadense,* and two from Africa-Asia, *G. arboreum* and *G. herbaceum.* In each of these four cases, aboriginal peoples discovered that the unique properties of cotton "fibers", which are unicellular epidermal seed trichomes, made them useful for ropes, textiles, and other applications. Each of these crop species has its own history of domestication, diversification, and current utilization (Brubaker et al. 1999a; Hutchinson 1951, 1954; Hutchinson et al. 1947; Percy and Wendel 1990; Wendel et al. 2009).

This rich history of scientific study has made *Gossypium* one of the best systems for studies of polyploidy in plants. As testified by the many papers in this volume, plant evolution has been characterized by repeated rounds of whole-genome doubling. Evolutionary footprints of paleopolyploidy in "diploid" ($n = 13$) cotton have long been evidenced using both classic and more modern techniques, including chromosome banding, comparative genome mapping, and analysis of synonymous substitution rate curves (Muravenko et al. 1998; Brubaker et al. 1999b; Reinisch et al. 1994; Paterson 2009; Lin and Paterson 2009; Lin et al. 2011). These more ancient events, which must have profoundly impacted morphological, ecological, and physiological diversification, predate a more recent allopolyploidy event that traces to the mid-Pleistocene (Wendel 1989; Wendel et al. 2009; Wendel and Cronn 2003). The most experimentally tractable polyploid event is this most recent one, resulting from the merger that led to the evolution of the five extant species of New World allopolyploid cotton ($2n = 4x = 52$), including the two

cultivated species *G. hirsutum* (Upland cotton) and *G. barbadense* (Pima cotton, Sea Island cotton). *Gossypium hirsutum* presently is responsible for over 90 % of the cotton crop internationally, having spread from its original home in Mesoamerica to over 50 countries in both eastern and western hemispheres. Accumulating evidence indicates that polyploidy *per se* has played a critical role in enabling the development of modern, agronomically elite varieties. Thus, it is of interest to explore what cotton has taught us about the evolutionary consequences of genome doubling, and in turn, what the process of genome doubling might reveal about phenotypic diversification and the goals of agronomic improvement.

We begin by reviewing our current understanding of the diversity of the genus and the origin of the *Gossypium* clade, which includes one of the classic examples of polyploidy. This phylogenetic and temporal perspective provides the organismal framework that serves as the foundation for all analyses of the consequences of polyploid evolution, including those focused on genomic, epigenomic, and phenotypic levels. We then provide a synopsis of the myriad genomic consequences that were set in motion by the evolutionary processes of genome merger, chromosome doubling, and subsequent evolutionary diversification which collectively gave rise to modern allopolyploid cottons, drawing attention to the relationships among evolutionary processes and temporal scale of divergence. Thus, we distinguish phenomena and processes that might characterize the earliest stages of polyploid formation from those that are responsible for longer term genomic and phenotypic changes. Finally, we summarize evidence that polyploidy enables the evolution of transgressive or novel phenotypes in cotton, as exemplified by the differences between modern cultivated diploid versus allopolyploid cotton.

10.1.1 Origin and Diversification of the Diploid Cottons

The cotton genus belongs to a small tribe, the *Gossypieae,* that includes only eight genera and ~120 species (Fryxell 1968, 1979). Four of these genera are either monotypic or contain only several species with restricted geographic distributions, *Lebronnecia* (Marquesas Islands), *Cephalohibiscus* (New Guinea, Solomon Islands), *Gossypioides* (east Africa, Madagascar), and *Kokia* (Hawaii). The tribe also includes four moderately sized genera with broader ranges: *Hampea*, with 21 neotropical species; *Cienfuegosia*, with 25 species from the neotropics and parts of Africa; *Thespesia*, with 17 tropical species; and last but not least, *Gossypium*, the largest and most widely distributed genus in the tribe with more than 50 species (Fryxell 1992).

Gossypium species collectively have achieved a nearly worldwide distribution, with several primary centers of diversity in the arid or seasonally arid tropics and subtropics. Species-rich regions include Australia, especially the Kimberley region in NW Australia, the Horn of Africa and southern Arabian Peninsula, and the western part of central and southern Mexico. Recognition of these groups of species reflects decades of accumulated understanding that emerged from basic plant exploration and taxonomic analysis (Fryxell 1979, 1992; Hutchinson et al.

1947; Saunders 1961; Watt 1907; Wendel et al. 2009). The genus is extraordinarily diverse; species morphologies range from fire-adapted, herbaceous perennials in NW Australia to trees in SW Mexico that escape the dry season by dropping their leaves. Corolla colors span a rainbow of blue to purple (*G. triphyllum*), mauves and pinks ("Sturt's Desert Rose", *G. sturtianum*, is the official floral emblem of the Northern Territory, Australia), whites and pale yellows (NW Australia, Mexico, Africa-Arabia), and even a deep sulfur-yellow (*G. tomentosum* from Hawaii). Seed coverings range from nearly glabrous (e.g., *G. klotzschianum* and *G. davidsonii*), to short stiff, dense, brown hairs that aid in wind-dispersal (*G. australe, G. nelsonii*), to the long, fine white fibers that characterize highly improved forms of the four cultivated species (Fig. 10.1). There are even seeds that produce fat bodies to facilitate ant-dispersal (Seelanan et al. 1999). At the other end of the ant coevolution spectrum is *G. tomentosum* from the Hawaiian Islands, which lost the foliar and extra-floral nectaries that are common in other *Gossypium* species, presumably in response to the absence of native ants.

The evolution of this morphological and geographic diversity was accompanied by a parallel diversification at the chromosomal level (Lin and Paterson 2009). Although all diploid species share the same chromosome number ($n = 13$), there is more than 3-fold variation in DNA content per genome (Hendrix and Stewart 2005). Chromosome morphology is similar among closely related species, as reflected in the ability of related species to form hybrids that display normal meiotic pairing and sometimes high F_1 fertility. In contrast, crosses among more distant relatives may be difficult to achieve, and those that are successful are characterized by meiotic abnormalities. The collective observations of pairing behavior, chromosome size, and relative fertility in interspecific hybrids led to the designation of single-letter genome symbols (Beasley 1941) for related clusters of species. Presently, eight diploid genome groups (A through G, plus K) are recognized (Endrizzi et al. 1985; Stewart 1995).

A genealogical framework for the genus is provided from analyses of multiple molecular phylogenetic investigations (reviewed in Wendel and Cronn 2003). A key phylogenetic conclusion has been the demonstration that the group of species recognized as belonging to *Gossypium* do, in fact, constitute a single natural lineage, despite their exceptionally broad geographic distribution and extraordinary morphological and cytogenetic diversity. A second important result is that the closest relative of *Gossypium* is the sister clade that includes the African-Madagascan genus *Gossypioides* and the Hawaiian endemic genus *Kokia*; these latter genera may thus be used as phylogenetic outgroups for studying evolutionary patterns and processes within *Gossypium*. A third phylogenetic conclusion is that each of the classically recognized genome groups comprises a monophyletic group. This information is summarized in a depiction of our present understanding of relationships (Fig. 10.2), which shows four major lineages of diploid species corresponding to three continents: Australia (C-, G-, K-genomes), the Americas (D-genome), and Africa/Arabia (two lineages: one comprising the A-, B-, and F-genomes, and a second containing the E-genome species). Embedded in this result is the observation that the earliest divergence in the genus separated the New

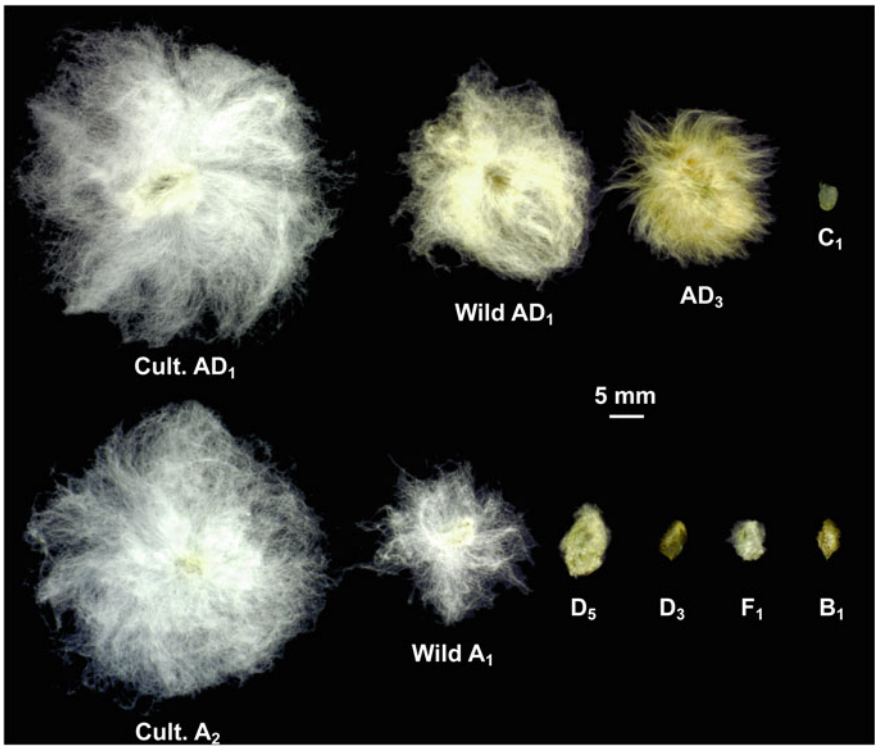

Fig. 10.1 Representative seed and trichome diversity in *Gossypium*. Seed and trichome size and morphology are exceedingly variable in the genus. Most wild species have relatively small seeds (<5 mm in any dimension) with equally *short* fibers. *Long* (spinnable) fiber evolved only once, in the ancestor of modern A-genome cottons, which subsequently donated this capacity to modern tetraploid species, including the commercially important *G. hirsutum* and *G. barbadense*, at the time of allopolyploid formation. Key to species: Cult. AD_1 = *G. hirsutum* TM1; Wild AD_1 = *G. hirsutum* Tx2094 from the Yucatan Peninsula; AD_3 = *G. tomentosum* WT936 from Hawaii; C_1 = *G. sturtianum* C_1-4 from Australia; Cult. A_2 = *G. arboreum* AKA8401; Wild A_1 = *G. herbaceum* subsp. *africanum* from Botswana; D_5 = *G. raimondii* from Peru; D_3 = *G. davidsonii* D_{3d}-32 from Baja California; F_1 = *G. longicalyx* F_1-3 from Tanzania; B_1 = G. *anomalum* B_1-1 from Africa

World D-genome lineage from the ancestor of all Old World taxa, and thus, that New World and Old World diploids are phylogenetic sister groups. Following this split in the genus, cottons comprising the Old World lineage divided into three groups, namely, the Australian cottons (C-, G-, and K-genome species), the African-Arabian E-genome species, and the African A-, B-, and F-genome cottons. The African F-genome clade, which consists of the sole species *G. longicalyx*, is diagnosed as sister to the A-genome species, an important realization in that this relationship identifies the wild forms most closely related to the clade (the A-genome) that "invented" long, or spinable, fiber since these two clades (A- and F-) shared a common ancestor.

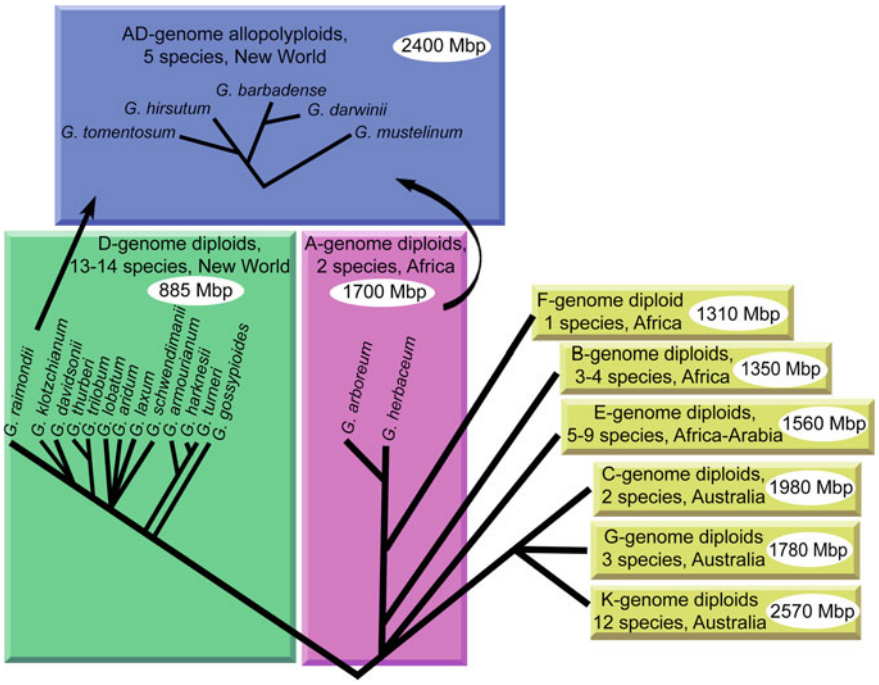

Fig. 10.2 Evolutionary history of *Gossypium*, as inferred from multiple molecular phylogenetic data sets. The closest relative of *Gossypium* is a lineage containing the African-Madagascan genus *Gossypioides* and the Hawaiian endemic genus *Kokia*. Following its likely origin 5–10 mya, *Gossypium* split into three major diploid lineages: the New World clade (D-genome); the African-Asian clade (A-, B-, E-, and F-genomes); and the Australian clade (C-, G-, and K-genomes). This global radiation involved several trans-oceanic dispersal events and was accompanied by morphological, ecological, and chromosomal differentiation (2C genome sizes shown in *white ellipses*). Allopolyploid cottons formed following trans-oceanic dispersal of an A-genome diploid to the Americas, where the immigrant underwent hybridization, as female, with a native D-genome diploid similar to modern *G. raimondii*. Polyploid cotton probably originated during the Pleistocene (1–2 mya), with the five modern species comprising the descendants of an early and rapid colonization of the New World tropics and subtropics

A temporal framework for the origin of *Gossypium* and its diversification is provided by sequence divergence data, which may serve as a proxy for time (Senchina et al. 2003). These analyses indicate that *Gossypium* diverged from its closest relatives during the Miocene, perhaps 10–15 mya, subsequently spreading around the world via trans-oceanic dispersal to acquire its modern geographic range. These early estimates of divergence, and hence times, have recently been supported by an enormous amount of new sequence data derived from a global assembly of ESTs (Flagel et al. 2012). Consideration of the phylogeny of Fig. 10.2 in a temporal context and in light of plate tectonic history leads to the inference that the history of *Gossypium* has entailed multiple episodes of trans-oceanic dispersal. These include at least one dispersal between Australia and Africa,

another to the Americas leading to the evolution of the D-genome diploids, and a second, much later colonization of the New World by the A-genome ancestor of the AD-genome allopolyploids (see below).

10.1.2 Origin and Diversification of the Polyploid Cottons

A rich body of cytogenetic and experimental evidence has demonstrated that the tetraploid species, which are entirely New World in their distribution, are allopolyploids containing two genomes, an A-genome from Africa or Asia, and a D-genome similar to those found in the American diploids (Endrizzi et al. 1985; Wendel et al. 2009; Wendel and Cronn 2003). The hemisphere-scale allopatry of these two diploid genome groups led to more than 50 years of mystery surrounding the timing of formation and parentage of the New World allopolyploids. With respect to the question of "when", gene sequence data convincingly demonstrate that allopolyploid *Gossypium* originated prior to the evolution of modern humans but relatively recently in geological terms, perhaps 1–2 mya, or in the mid-Pleistocene (Senchina et al. 2003; Wendel 1989; Flagel et al. 2012). With respect to the second part of the question, that of polyploid parentage, it is now clear that both extant A-genome species (*G. arboreum, G. herbaceum*) are equally divergent from the A-genome of allopolyploid cottons and that the closest living relative of the progenitor D-genome donor is *G. raimondii* (Endrizzi et al. 1985; Wendel et al. 2009; Wendel and Cronn 2003). Studies using nuclear (biparentally inherited) genes led to the same conclusion. Additionally, all allopolyploids contain an A-genome cytoplasm, as evidenced from analysis of both mitochondrial and plastid genomes (Galau and Wilkins 1989; Small and Wendel 1999; Wendel 1989). Finally, the studies mentioned above, and additional, extensive DNA sequence data from ongoing studies (Grover et al. unpublished), support a single origin for allopolyploid cotton.

Given a Pleistocene origin for allopolyploid cotton species, one may infer that their morphological diversification and spread must have been relatively rapid following polyploidization. At present, five allopolyploid species are widely recognized, although a sixth species (*G. ekmanianum*) was recently proposed (Krapovickas and Seijo 2008). *Gossypium darwinii* is native to the Galapagos Islands, where it may form large populations in some areas (Percy and Wendel 1990). *Gossypium tomentosum*, from the Hawaiian Islands, has a more diffuse population structure, occurring mostly as scattered individuals and small populations on several islands (DeJoode and Wendel 1992). A third allopolyploid, *G. mustelinum*, is restricted to a small region of northeast Brazil (Wendel et al. 1994). In addition to these three truly wild species, there are two cultivated species (*G. barbadense* and *G. hirsutum*), each of which has a large indigenous range, collectively encompassing a wealth of morphological forms that span the wild-to-domesticated continuum (Brubaker and Wendel 1993, 1994, 2001; Fryxell 1979; Hutchinson 1951; Percy and Wendel 1990). *Gossypium hirsutum* is widely

distributed in Central and northern South America, the Caribbean, and even reaches distant islands in the Pacific (eg., Solomon Islands, Marquesas). *Gossypium hirsutum* is thought to have a more northerly distribution than *G. barbadense,* with wild populations occurring as far north as Tampa Bay, Florida (27̊38'N) (Stewart, personal observation). *Gossypium barbadense* has a more southerly indigenous range, centered in the northern third of South America but with a large region of range overlap with *G. hirsutum* in the Caribbean.

Consideration of the distribution of the allopolyploid species suggests that polyploidy led to the invasion of a new ecological niche. Fryxell (1965, 1979) noted that in contrast to the majority of diploid species, allopolyploid species typically occur in coastal habitats, at least those forms that arguably are truly wild. Two species, both island endemics (*G. darwinii* and *G. tomentosum*), are restricted to near coastlines, and for two others (*G. barbadense* and *G. hirsutum*), wild forms occur in littoral habitats ringing the Gulf of Mexico, northwest South America, and distant Pacific Islands. Fryxell speculated that following initial formation, adaptation of the newly evolved allopolyploid to littoral habitats enabled it to exploit the fluctuating sea levels that characterized the Pleistocene. This ecological innovation is envisioned to have facilitated initial establishment of the new polyploid lineage and also may have provided a means for the rapid dispersal of the salt-water-tolerant seeds.

10.2 Evolution Following Genome Duplication

10.2.1 Chromosomal Stabilization and Structural Stasis

Classical cytogenetic evidence indicates that chromosomes of the A- and D-genomes of allopolyploid *Gossypium* are less able to pair with one another than are chromosomes of the living descendants of their diploid progenitors (reviewed in Endrizzi et al. 1985). For example, allopolyploid-derived haploids form an average of less than one bivalent per cell at meiotic metaphase, whereas chromosomes in hybrids between extant A- and D- genome diploids average 5.8 and 7.8 bivalents (reviewed in Endrizzi et al. 1985). These and similar observations indicate that natural selection has favored the evolution of mechanisms that promote exclusive bivalent formation in the allopolyploid. Neither the pace at which such mechanisms operate nor their nature are understood, but it seems rational to postulate that selection would be most intense in the first generations following allopolyploid formation, where the fitness cost of unbalanced gametes would be the greatest.

One hypothesis for this apparent lack of bivalent formation is that genome stabilization following polyploidization involved genomic reorganization of the two resident genomes such that they no longer are capable of homoeologous pairing. To evaluate the extent of structural change, genetic maps were generated and compared among interspecific F_2 progenies for diploid (both A- and D-) and allopolyploid cottons (Brubaker et al. 1999b; Rong et al. 2004). Comparisons of

gene order and synteny among the A- and D-genome maps, as well as those for both genomes of the allopolyploid (A vs. A_T and D vs. D_T), demonstrate that relatively few structural rearrangements have arisen in the 1–2 my since allopolyploid formation; conservation of collinear linkage groups is the rule rather than the exception. Thus, allopolyploidy in *Gossypium* has not been accompanied by extensive chromosomal rearrangement. This implies that structural rearrangement has not been a significant aspect of the process of polyploid genome stabilization in cotton. Additional support for this idea emerges from experiments involving synthetic allopolyploids; for example, Liu et al. (2001) used AFLP analysis to demonstrate almost exclusive fragment additivity for 22,000 genomic loci in nine sets of newly synthesized allotetraploid and allohexaploid *Gossypium*. Thus, and in contrast to some other plant models described in this volume (e.g., wheat, *Tragopogon*), the polyploid *Gossypium* genome appears to be relatively quiescent, at least with respect to the phenomenon of rapid genome change. A corollary is that the evolutionary enforcement of homologous pairing discussed above originated through means other than structural rearrangements.

10.2.2 Genome Sizes, Transposable Element Mobilization, and Genomic Downsizing

As shown in Fig. 10.2, genome sizes vary widely among diploid cotton species, from ~900 Mb in the D-genome diploids to ~2,600 Mb in the Australian diploids (Hendrix and Stewart 2005), reflecting primarily the differential and punctuated proliferation of various families of *copia* and *gypsy* transposable elements (TEs), as well as lineage-specific differences in the rate of deletions (Hawkins et al. 2008, 2006, 2009). The two progenitor genomes of allopolyploid cotton differ 2-fold in size, and moreover, they differ in their complement of resident TEs. Thus, allopolyploidization entailed the merger of two different complements of TEs, creating the potential for activation of TEs due to the generalized disruption of epigenetic suppression of TE activity following the merger of two diverged regulatory systems, a process commonly referred to as "genomic shock". To evaluate the possibility that polyploidization in *Gossypium* was accompanied by a transpositional burst, as in some other species (Kashkush et al. 2002; Shan et al. 2005; Ungerer et al. 2006), Hu et al. (2010) used phylogenetic and quantitative methods to identify changes in TE populations. These data showed that the major LTR retrotransposon classes in the AD genome phylogenetically clustered with either their A- or D-genome antecedent elements in a genome-specific fashion, with no evidence of an impressive, recent, TE burst. Thus, hybridization and polyploidy do not appear to have stimulated a massive TE proliferation in *Gossypium.*

Notwithstanding the relative TE quiescence indicated by these studies, evidence using FISH (Hanson et al. 1999, 1998) implicates at least a modest level of TE activity in allopolyploid cotton. These data show that a family of *copia*-like retrotransposable elements "horizontally" transferred across genomes following

allopolyploid formation. This result highlights the phenomenon of TE spread across previously separated genomes following polyploid formation, raising the possibility that this process has played a role in diversification and adaptation via novel TE insertions.

An attractive feature of the *Gossypium* model system is that species representing both progenitor genomes remain extant, and that they vary so dramatically (nearly 2-fold) in genome size yet retain collinearity with their orthologs in the allopolyploid genome. Although the *Gossypium* genome has yet to be sequenced, it is likely that a high-quality D-genome sequence (from *G. raimondii*) will be published by the time this volume is published, with perhaps the A-genome sequence not far behind. It is an exciting prospect to contemplate the availability of both diploid sequences as well as one or more from allopolyploid *Gossypium*. These data will offer a veritable gold mine for generating insight into the pace, patterns, and dynamics of genome evolution that accompany diploid divergence, allopolyploid formation, and subsequent evolution at the polyploid level. A glimpse of this promise is provided by current work focused on comparative sequencing of bacterial artificial chromosomes (BACs) from *Gossypium* in polyploid and diploid species. Two comparative BAC sequencing studies have been published, a region surrounding the *CesA1* gene (Grover et al. 2004) from both homoeologous genomes of *G. hirsutum*, and a region surrounding the *AdhA* gene (Grover et al. 2007) from *G. hirsutum* and diploids representing models of the two progenitor genomes, i.e., *G. arboreum* and *G. raimondii*. Data generated to date indicate that small deletions are more prevalent in the polyploid genomes (A_T and D_T) than in either diploid genome, illustrative of the general phenomenon of genomic downsizing in polyploid genomes (Bennett and Leitch 2005), while providing a glimpse into an underlying mechanism (illegitimate recombination, which is biased in the allopolyploid toward deletion). Extensions of this work into 20 regions and in more genomes, including a phylogenetic outgroup are underway (Grover, Wendel, and Paterson, unpubl.), offering an opportunity to learn more about features of genome evolution that distinguish polyploid cotton from its diploid progenitors.

10.2.3 Genic Evolution in Diploid Versus Allopolyploid Cotton

The most immediate and important genomic consequence of allopolyploid formation in *Gossypium* was simultaneous duplication of all nuclear genes. From a phylogenetic perspective, the various fates of gene duplication may partially be modeled as shown in Fig. 10.3. The null hypothesis for sequence evolution in allopolyploids derives from the organismal history; if both duplicated genes evolve independently following allopolyploid formation, then each homoeolog should be phylogenetically sister to its ortholog from its progenitor diploid, rather than to the other homoeolog. Similarly, if rates of sequence evolution are maintained between the diploid and allopolyploid level, branch lengths for the two A-genome sequences

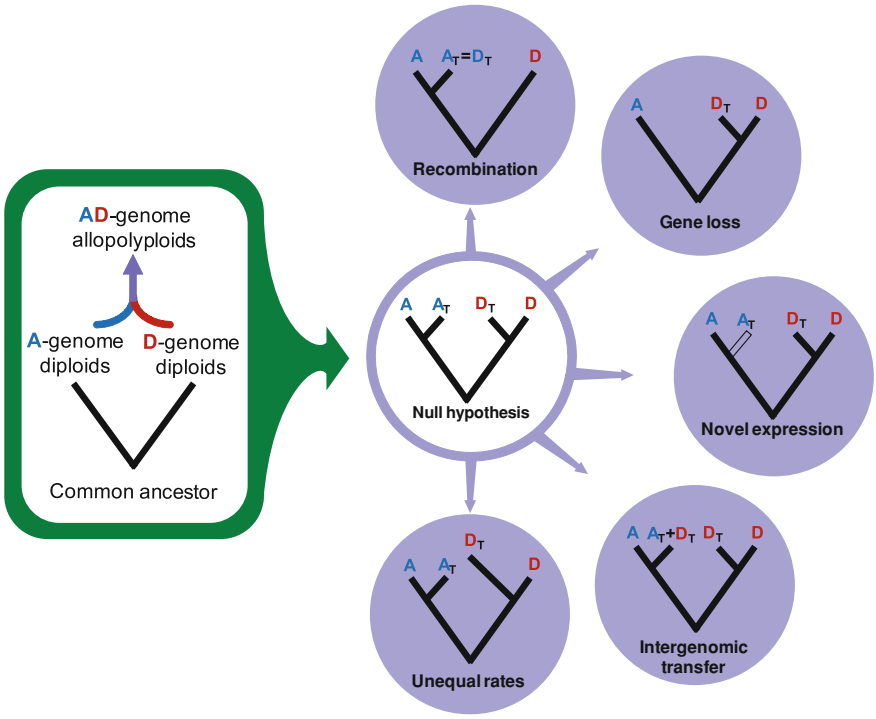

Fig. 10.3 A model of various possibilities for duplicate gene evolution after allopolyploidy in *Gossypium*. The null expectation (*center*) derives from the organismal history (*left*): if homoeologs evolve independently following allopolyploid formation, then each should be phylogenetically sister to its ortholog from the donor diploid, rather than to the other homoeolog. Similarly, if rates of sequence evolution are similar at the diploid and allopolyploid level, branch lengths for the two A-genome sequences (one from the diploid, "A", and the other from the allopolyploid, "A_T") should be similar, as they should for the two D-genome sequences ("D" and "D_T"). The utility of this null hypothesis lies in its falsification; if homoeologous sequences interact via concerted evolutionary forces or nonreciprocal homoeologous recombination, for example, a different tree may be recovered ("Recombination", *top center*), or if there is strong directional selection or pseudogenization, rate inequalities may become evident ("Unequal rates", *bottom center*). Additional possibilities include loss of one of the homeologs ("Gene loss", *top right*), replicative transfer of sequences from one genome to the other ("Intergenomic transfer", *bottom right*), and evolutionary divergence in duplicate gene expression domains or amounts ("Novel expression", *right middle*). This latter category, novel expression, encompasses multiple phenomena, including developmentally or environmentally regulated biases in homoeolog expression ratios, organ- or tissue- or cell-specific homoeolog silencing, novel expression domains, and transgressive (*higher* or *lower* than either progenitor diploid) expression amounts

(one from the diploid, the other from the allopolyploid) should be similar, as they should for the two D-genome sequences. The utility of the null hypothesis lies in its falsification; if homoeologous sequences interact, for example, a different tree may be recovered, or if there is strong directional selection or pseudogenization, rate

inequalities may become evident. Additional possibilities include silencing or loss of one of the duplicated copies, "horizontal transfer" of sequences from one genome to the other, and novel expression, the latter encompassing a variety of phenomena discussed under *Duplicate gene expression* (Sect. 10.4).

10.2.3.1 Molecular Evolution of Homoeologs and Orthologs

Early work (Cronn et al. 1999; Senchina et al. 2003) demonstrated that duplicated genes typically are both retained even in natural (1–2 mya -ld) allopolyploids, and that they evolve essentially as modeled in the central panel of Fig. 10.3, i.e., equally, additively, and at equivalent rates. A more recent analysis (Flagel et al. 2012), based on a global assembly of 5 million Sanger and 454 ESTs supplemented by ~150 million 82-bp Illumina reads, has provided a vastly expanded and detailed view of genic evolution in diploid and allopolyploid cotton, perhaps the most extensive yet for any plant genus. These data, representing analysis of ~10,000 genes in each comparison, show that rates of synonymous substitution (*Ks*) between A- and D-genome orthologs (0.036) are nearly identical to that experienced by their homoeologous descendants in the allopolyploid genome (0.037), with non-synonymous substitution rates being identical in the two contrasts (0.009). These results demonstrate unequivocally that allopolyploidy has not, in general, been accompanied by an enhanced rate of nucleotide substitution in coding regions, as might be expected from an assumption of rapid decay of "redundant" duplicated copies.

An additional perspective provided by Flagel et al. (2012) is that the data also provide an extraordinarily accurate depiction of the ancestry of polyploid cotton. By using thousands of genic alignments, the relative distances of the A- and D-genome diploids to their counterparts in the allopolyploid (A_T, D_T) can be calculated and effectively translated into relative divergence times because of the demonstration of rate homogeneity discussed above. The results show that the A genome diploid has a mean *Ks* of 0.009 from the A_T-genome, whereas the comparable figure for the D genome comparison is 0.015. From this we infer that modern A-genome diploids are a better model (by about 50 %) of the actual genome donor of allopolyploid cotton than are extant D-genome diploids, consistent with previous suggestions based on diverse data sources (Wendel et al. 2009; Wendel and Cronn 2003; Senchina et al. 2003). These new data provide deep insight into the cotton model system, furthering its value for comparative analyses.

A general expectation of molecular evolution in allopolyploids is that mutations occur randomly among homoeologs, and hence that evolutionary rates will be equivalent for homoeologs (Fig. 10.3). A corollary expectation is that duplicated gene copies will accumulate diversity, within and among populations, at equivalent rates. Small et al. (1999) tested this hypothesis of rate equivalence among homoeologs using the gene *AdhA*, for which both homoeologs were sequenced for

22 accessions (44 alleles per genome) of *G. hirsutum* and for five accessions (10 alleles per genome) of *G. barbadense*. In both allopolyploids, estimates of nucleotide diversity were higher for *AdhA* from the D-genome than from the A-genome. In a follow-up study (Small and Wendel 2002) using a second ADH gene (*AdhC*), this conclusion was even more strongly supported; here 24 different alleles were detected for the D-genome homoeolog versus only 7 for the A-genome homoeolog, with a similar increase in allelic diversity for the D-homoeolog in *G. barbadense*. These observations indicate that at least for some genes, homoeologs may accumulate synonymous substitutions at vastly different rates. At present, the responsible forces and underlying molecular mechanisms are obscure, but a logical suggestion is that they are causally connected to the nearly 2-fold difference in genome size between the co-resident genomes in the allopolyploid nucleus (the A_T genome is approximately twice the size of the D_T genome).

At present, little information exists for *Gossypium* that enables a thorough analysis of the relative rates of pseudogenization between diploids and allopolyploids, or between the two homoeologous genomes of allopolyploid cotton. The ongoing extensions of the comparative BAC sequencing approach taken by Grover and colleagues (Grover et al. 2004, 2007) are promising in this respect, as are the ongoing genome sequencing projects. Similarly, relatively little information exists on rates of gene loss in cotton, though evidence to date based on comparative BAC sequencing (Grover et al. 2004, 2007), comparative mapping analyses (Rong et al. 2004; Brubaker et al. 1999b), EST collections (Flagel et al. 2012), and AFLP studies (Liu et al. 2001) suggests that rates of gene loss are neither high nor particularly biased with respect to genomic origin. A recent study using Southern hybridization, however, detected three losses (of 27 genes studied) of D-homoeologs from allopolyploid cotton and no losses of the A-homoeolog, suggesting a possible bias (Rong et al. 2010). Again, key data likely will be generated soon that will permit these speculations to be evaluated.

10.2.3.2 Interaction Among Homoeologs at the DNA Level

One of the first indications that duplicated genes could behave in an evolutionarily dependent fashion was the study of Wendel et al. (1995), who demonstrated interaction among the 18S–26S ribosomal genes that exist at multiple loci in the A- and D-genomes. Instead of evolving independently, as expected if homoeologous repeats did not interact, repeats at the different arrays in allopolyploid cotton have been "homogenized" to the same sequence (either "A-like" or "D-like") by one or more processes of concerted evolution (reviewed by Elder and Turner 1995). In four of the five allopolyploid species, interlocus homogenization has created exclusively D-genome like rDNAs, whereas in *G. mustelinum* nearly all rDNA repeats have been homogenized to an A-like form. This example showed that since polyploid formation 1–2 mya, some 3,800 repeats, each approximately 10 kb in length, were "overwritten" with the alternative form originating from the

other parental genome, probably through unequal crossing over or gene conversion, and that this phenomenon operated bidirectionally, in different directions in different allopolyploid lineages. Interlocus concerted evolution of rDNA repeats has since been documented in many other plant polyploids, including *Nicotiana* and *Tragopogon* (see this Chaps. 10 and 14, this volume). From a mechanistic perspective, it seems likely that the homogenization has resulted from unequal crossing over within and among arrays and nonreciprocal homoeologous recombination events, with homoeolog exchanges being facilitated by the sub-telomeric location of rDNAs in many plant lineages, including *Gossypium* (Wendel 2000).

This demonstration that some repeated sequences could interact across genomes in the allopolyploid nucleus led to additional investigations of the scope of the phenomenon. In an analogous study, Cronn et al. (1996) showed that in contrast to 18S–26S arrays, 5S rDNA genes are not homogenized by concerted evolutionary forces in the allopolyploid. Similarly, and as noted above, early studies demonstrated apparent independent evolution of homoeologs (Cronn et al. 1999; Senchina et al. 2003), consistent with the null hypothesis of Fig. 10.3. Thus, until recently it was thought that most nuclear genes duplicated by allopolyploidy largely evolve independently of one another in the polyploid nucleus, a proposition that seemed entirely reasonable given the absence of evident cytogenetic interactions among homoeologs.

This presumption of homoeolog independence turns out to be incorrect; however, and for a surprisingly large percentage of duplicated genes. Using ESTs generated from both A- and D-genome diploids and AD-genome allopolyploids, Salmon et al. (2010) sought evidence of small "gene conversion" or nonreciprocal exchanges among homoeologs, based on the presence in allopolyploid cotton of genes that have genome-diagnostic (A or D) SNPs (single nucleotide polymorphisms) that occurred in patterns suggestive of genic interactions (e.g., a single expressed sequence containing diagnostic SNPs in the following order, AADDAA, implicating conversion of the middle section of an A_T homoeolog by the D_T homoeolog). These bioinformatic inferences were validated by de novo PCR and sequencing. Results were convincing and compelling, demonstrating that about 2 % of contigs in *G. hirsutum* have experienced nonreciprocal homoeologous exchanges since the origin of polyploid cotton 1–2 mya. Moreover, when a sampling of these homoeologous interactions was studied throughout the polyploid clade using a phylogenetic approach, nonreciprocal homoeologous exchanges were shown to have occurred throughout polyploid divergence and speciation, as opposed to being concentrated at the root of the polyploid tree. Among six homoeologous interactions, five occurred in only one species, with the sixth event being shared. This result refutes a logical a priori prediction, namely that interactions among homoeologs should be most frequent early in allopolyploid evolution (and hence shared among species), prior to reinforcement of apparently strict bivalent pairing. Finally, some genomic regions showed multiple patterns of homoeologous recombination among species, suggesting that some regions or genes may be "hot-spots" for nonreciprocal homoeologous exchanges.

This demonstration of non-independent evolution of homoeologs has recently been confirmed and extended to tens of thousands of genes using an expanded EST set (Flagel et al. 2012). This analysis resulted in an even higher percentage (5 %) of contigs in the assembly that show evidence of homoeolog contact in at least one of the two polyploid species studied (*G. hirsutum* and *G. barbadense*), as well as confirm the suggestion of hot-spots for homoeologous recombination. This high frequency of homoeologous contact is an astonishing result given the absence of prior cytogenetic or other observations that would have suggested this possibility, and the apparently complete bivalent formation at meiosis. Flagel et al. (2012) also detected 50 % more homoeologous exchanges in *G. hirsutum* than in *G. barbadense*, suggesting that following polyploid formation, rates of nonreciprocal homoeologous recombination may diverge, even among closely related species that share nearly all evident ecological and life-history characteristics. These results are fascinating and lead to a number of questions. For example, what is the genomic distribution of genes subject to nonreciprocal recombination events, and does this information suggest a causal mechanism? Are nonreciprocal recombination events random with respect to outcome (A vs D), or alternatively, is the nuclear genome of allopolyploid cotton slowly becoming more "A-like" or D-like"? Do any of the detected "gene conversion" events have physiological consequences and hence possible adaptive significance? Insight into these and related questions in cotton and in other allopolyploids is likely to emerge in the next several years.

10.3 Duplicate Gene Expression

In addition to evolutionary changes in *gene and genome structure*, a key component of polyploid evolution concerns the consequences of genome doubling on *gene expression*. The biological reunion in the allopolyploid nucleus of two regulatory systems that had evolved in isolation for 5–10 my (A- and D-genome), in only one of the two parental cytoplasms (A-genome), might be expected to lead to violations of the equal and additive expression modeled in Fig. 10.3. In principle, both genomes in an allotetraploid could contribute equally to the transcriptome, for any pair of homoeologs or overall, or alternatively, there may be preferential transcription of one genome due to intergenomic interactions that could bias the transcription machinery. Many aspects of this problem are of interest, including the scope and scale of preferential homoeolog expression, its tissue and organ specificity, the level of genomic bias, the mechanistic underpinnings that result in regulatory responses, the temporal scale at which alterations in homoeolog expression evolve, and ultimately, its physiological and evolutionary relevance.

Steps toward answering these questions for *Gossypium* have been taken using many different experimental approaches over the last decade. We summarize these results below, focusing on lessons regarding (1) developmental and environmental

effects on gene expression; (2) global analyses of genomic interactions; and (3) the temporal scale at which expression evolution arises.

10.3.1 Developmental and Environmental Effects on Duplicate Gene Expression

The first indication that polyploidy in *Gossypium* is accompanied by extensive organ-specific changes in duplicate gene expression emerged nearly a decade ago in a study of 40 homoeologous gene pairs in different organs of *G. hirsutum* using SSCP-cDNA (Adams et al. 2003). Almost one-third of the genes revealed bias toward one homoeolog or the other, or only expression of one homoeolog, in at least one organ. Transcript levels for the two members of each gene pair varied by gene and, unexpectedly, by organ type. Especially noteworthy were genes that showed organ-specific, reciprocal silencing of alternative homoeologs; that is, one member of a duplicated gene pair displayed minimal to no transcription in some organs, whereas a reciprocal pattern was exhibited by the alternative homoeolog in other organ(s). In particular, floral organs showed dramatic expression patterns in this regard, with major differences among petals, stamens, and stigmas/styles (Adams et al. 2003, 2004). Organ-specific expression of homoeologs was assayed more extensively by Chaudhary et al. (2009), who employed a novel high-resolution mass-spectrometry technology (Sequenom) to investigate relative expression levels of each homoeolog for 63 gene pairs in 24 tissues. Results from over 2,000 assays demonstrated that 40 % of homoeologs are transcriptionally biased in at least one stage of cotton development, that genome merger *per se* has a large effect on relative expression of homeologs (see section on temporal partitioning, below), and that the majority of alterations are caused by *cis*-regulatory divergence between diploid progenitors. The study also revealed 15 cases of probable regulatory neofunctionalization among 8 tissues, perhaps the first such demonstration in allopolyploid plants.

These surprising indications of partitioning of duplicate gene expression during development have since been found in many other allopolyploid plants (see also Chap. 14, this volume, and, e.g., Buggs et al. 2010; Bottley and Koebner 2008; Bottley et al. 2006; Wang et al. 2006), opening the experimental floodgates on the study of duplicate gene expression. Because differences in duplicate gene expression were found between different organ types, the question arose as to patterns of duplicate gene expression during development of any single organ. Liu and Adams (2007) examined the expression ratio of homoeologous *AdhA* genes in eight developmental stages of hypocotyls, cotyledons, and roots, and 11 developmental stages of ovary walls and ovules. They showed that expression ratios of the two homoeologs changed considerably when comparing some stages of organ development, indicating that differential expression of homoeologous genes is developmentally regulated and that determining the extent of homoeologous gene

expression changes requires analysis of multiple developmental stages and organs. An important extension of this line of analysis was to ask whether biased expression of homoeologs can be cell type-specific. In this respect, cotton offers an excellent opportunity because the single-celled fibers are easily accessible. To address this question, Hovav et al. (2008a, b) used homoeolog-specific microarrays to assay homoeologous gene expression during fiber development. In these remarkable studies, discussed further below, a full range of duplicate gene expression was observed among the approximately 1,400 genes, from balanced expression of homoeologs, to varying homoeolog ratios that shifted during development. Notably, four genes showed complete reciprocal silencing of alternative homoeologs during fiber development. Collectively, the foregoing studies show that variation in homoeologous gene expression during development is the rule rather than the exception, and that it occurs even during development of a single cell type.

An additional twist on duplicate gene expression has been the demonstration that homoeolog expression patterns are not simply developmentally fixed, but instead are responsive to environmental conditions. Liu and Adams (2007) examined expression of *AdhA* homoeologs in response to abiotic stress conditions, with a focus on cold and water submersion (flooding simulation), at different developmental stages. They found that some stress treatments significantly altered A_T to D_T expression ratios, including a case of reciprocal silencing of homoeologs in response to abiotic stress, an unprecedented finding. In a more recent study, Dong and Adams (2011) extended their analysis to 30 gene pairs using three organs (leaves, roots, and cotyledons) and five abiotic stress treatments (heat, cold, drought, high salt, and water submersion). Over 70 % of the genes showed stress-induced changes in the relative expression levels of the duplicates under one or more stress treatments, and 12 sets of homoeologs showed opposite changes in expression levels in response to different abiotic stress treatments. These results indicate that abiotic stress conditions can have considerable effects on duplicate gene expression in a polyploid, with the effects varying by gene, stress, and organ type. Together with the profound developmental partitioning noted above, differential expression in response to environmental stresses or cues may all be factors that contribute to the preservation of duplicated genes in polyploids.

10.3.2 Global Biases in Duplicate Gene Expression

To provide a more global perspective on homologous gene expression in allopolyploids, Udall et al. (2006) developed homoeolog-specific microarrays that utilized genome-diagnostic SNPs from ESTs generated from the two genomes of allopolyploid cotton and the diploids *G. arboreum* (A-genome) and *G. raimondii* (D-genome). Using leaf RNA, they found that 199 of 461 gene pairs (43 %) deviated from equal expression, thereby providing an initial quantitative perspective on the scale of biased homoeolog expression. This microarray

approach was refined and extended to a larger set of gene pairs ($n = 1383$) by Flagel et al. (2008), who reported that 70 % of the genes in petals have biased homoeolog expression ratios, and that more of these genes are D-genome (39.5 %) than A-genome (30.5 %) biased. In addition, they found that the D-genome copies of 69 genes and the A-genome copies of 46 genes were silenced, collectively representing about 8 % of all genes. Similarly, Hovav et al. (2008a, b) used the same SNP-specific microarray technology on RNAs extracted from trichomes harvested from three developmental time points in wild and modern accessions of two independently domesticated cotton species, *G. hirsutum* and *G. barbadense*. Among these species 25–37 % of genes were significantly biased toward one of the two genomes at each developmental stage, but these biases were not random with respect to genome-of-origin; instead, duplicate gene expression was biased toward the D-genome at all three time points studied, accounting for 63–76 % of the biased genes. Finally, and to place homoeolog expression evolution in a phylogenetic context, Flagel and Wendel (2010) extended their analysis of duplicate gene expression in petals to all five extant allopolyploid species. Several aspects of this study are notable, including the demonstration that all five species display an overall preference for D-genome expression (D-genome bias accounting for 54–60 % of genes with biased homoeolog expression), that the percentage of duplicate genes that are biased varies widely among species (from 48 to 88 %), and that the overall *magnitude* of bias (as opposed to simply whether or not a gene exhibits bias) similarly varies widely among species.

The foregoing synopsis shows that gene expression is massively altered in polyploid cotton relative to its diploid progenitors. Unequal expression of one of the two homoeologs likely is the rule rather than the exception, when integrated across organs and tissues, with gene silencing representing only the endpoint in a continuum of biased homoeolog expression. A global bias in gene expression from one of two co-resident genomes (the D-genome; see, however, Yang et al. 2006) is evident in allopolyploid cotton. At present, little evidence connects these observations to physiology or metabolism, and so connections to function and ecological relevance remain obscure (see, however, section on fiber below). Recent data (Hu et al. 2011) have extended the concept of bias to the proteomic level, where allopolyploid cotton seeds have been shown to preferentially accumulate proteins from the D-genome diploid parent, suggesting the possibility of a functional connection. A promising direction for future research is targeted, functional studies of specific gene families, protein complexes, and metabolic pathways and networks, aiming to place homoeolog expression into a context that permits insight into the biological significance of expression bias.

An emerging phenomenon in the study of polyploid gene expression on a large scale is the concept of *genomic dominance* in expression. This phenomenon, originally discovered and elaborated in cotton (Flagel and Wendel 2010; Rapp et al. 2009), and just recently detected in *Coffea* allopolyploids (Bardil et al. 2011), is defined as the state where total expression of a homoeolog pair mimics the expression level of one of two diploid parents of an allopolyploid. That is, if parent A is upregulated relative to parent B, the allopolyploid would exhibit the

expression state of parent A; in the case where parent A is downregulated relative to parent B, so is the allopolyploid. In this example, the A parent is genomically dominant with respect to gene expression. Using two different genomic combinations in synthetic allopolyploid *Gossypium*, Rapp et al. (2009) demonstrated that there exists a quantitative bias in genomic dominance toward one of the two parents. In the case of the synthetic AD allopolyploid, more than 10,000 genes display this dominant, quantitative expression phenotype. At present, the functional significance of genomic dominance is unknown.

10.3.3 Temporal Dynamics of Duplicate Gene Expression Evolution

One of the early (Adams et al. 2003) and subsequently further characterized findings about duplicate gene expression in cotton is that expression evolution has a temporal dimension; phenomena that typify genome merger and doubling need not be the same, qualitatively or quantitatively, as those that characterize natural allopolyploids that are 1–2 million years distant from polyploidization (Fig. 10.4). In some cases, such as organ-specific silencing of particular homoeologs, it has been shown that this can arise immediately upon allopolyploid formation (Adams et al. 2003, 2004; Chaudhary et al. 2009), leading to the notion (Adams et al. 2003; Rapp and Wendel 2005) that homoeolog silencing may be significant in duplicate gene retention and that it may actually represent a form of instantaneous subfunctionalization. These studies of individual genes have demonstrated that some of the changes detected in synthetic allopolyploids are mirrored in natural allopolyploids, whereas others are distinct between the synthetic and natural allopolyploids (see examples in Adams et al. 2004; Chaudhary et al. 2009). One possibility raised by these observations is that in some cases there is an immediate epigenetic response to genome merger and doubling that becomes either epigenetically stable or genetically fixed, for 1–2 million years.

With respect to the phenomena of genomic bias, the evolution of novel quantitative (transgressive) expression states, and genomic dominance, the studies of Flagel et al. (2008, 2010) have been particularly illuminating with respect to this temporal dimension to expression evolution. For example, whereas only about one-third of genes exhibit biased homoeolog expression in petals of AxD-genome, diploid hybrids, half or more do in petals of natural allopolyploids. Moreover, many of the genes that display bias in the initial F_1 retain or even magnify the degree of bias after 1–2 million years. Also, whereas allopolyploid formation appears to result in relatively few genes that exhibit quantitatively transgressive expression levels, relative to their diploid parents, this number rises sharply in natural allopolyploids, suggesting novel opportunities for exploring gene expression space. Finally, a comparison of genomic dominance in all five natural allopolyploids (Flagel and Wendel 2010) to that observed in synthetic allopolyploids (Rapp et al. 2009) demonstrated that the magnitude of genomic dominance remains high after

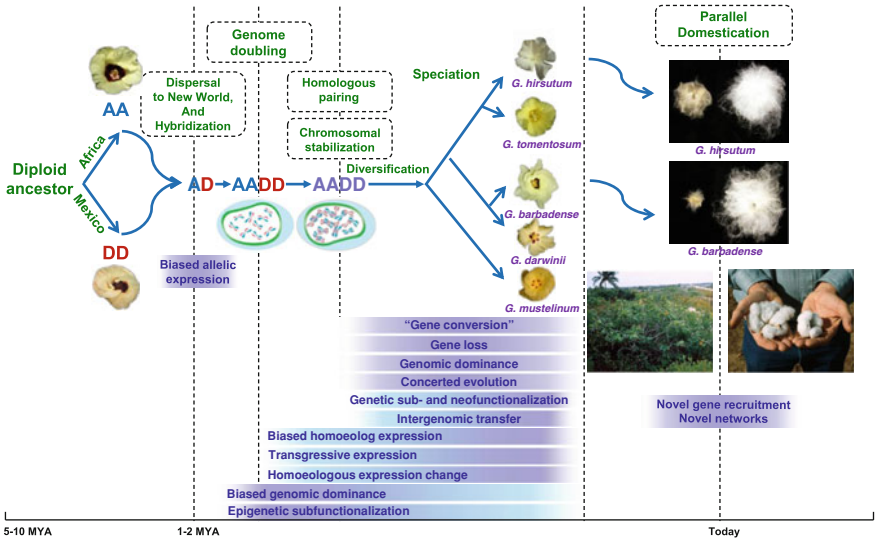

Fig. 10.4 A temporal depiction of phenomena that characterize polyploid evolution in *Gossypium*. Shown are the formation and diversification of the allopolyploid clade, several accompanying phenomena (dispersal, chromosomal stabilization), and the parallel domestication of *G. hirsutum* and *G. barbadense*. Images to the *right* represent wild and domesticated states, with close-ups of single seeds and the transformations in morphology accompanying domestication and crop improvement. Genetic and epigenetic phenomena associated with polyploidy in cotton are shown below, with a timeline (*bottom*), illustrating the temporal context for each phenomenon and whether there are increases or decreases (where known) in magnitude over time (*blue* to *purple* gradation indicates increase, the reverse indicates decrease)

1–2 my of evolution, but that the bias in its direction dissipates (shifting from predominantly D-genome dominance in the synthetics to similar levels of A- and D-genome dominance in the natural allopolyploids). Thus, it appears that natural allopolyploids have adjusted to more equally utilize the transcriptomes of the two co-resident genomes, notwithstanding the residual D-genome homoeolog bias described above. An intriguing possibility is that this D-genome homoeolog bias is connected causally to the biased D-genome expression dominance that arises following genome merger; to the extent that it is, it leads to the suggestion that homoeolog bias in other allopolyploids may be predicted by the initial conditions established by genomic merger in the distant past, and which may be experimentally mirrored in many systems through the use of synthetic hybrids and allopolyploids.

The partitioning of gene expression evolution into its temporal components leads to the suggestion that these different components may entail different or at least complementary mechanisms. The first, involving rapid or instantaneous gene expression alteration as a consequence of genome merger and doubling, reflects the myriad novel interactions accompanying a biological reunion of two differentiated genomes into a common nucleus. The precise nature of these interactions

is not known, but probably includes disruptions in gene dosage balance, stoichiometric changes resulting from differences in competition for transcription factors, differences in microRNA expression, and a host of novel *cis* and *trans* interactions (Birchler et al. 2005; Birchler and Veitia 2010; Chen 2007; Osborn et al. 2003; Veitia 2005) (see also Chaps. 1, 2, and 4, this volume). Superimposed on these rapid evolutionary responses to polyploidy are those that arise more slowly during the stabilization of the new polyploid genome and during evolution and speciation over much longer periods of time. Thus, the presence of duplicated genomes would seem to provide evolutionary opportunity and consequences immediately on polyploid formation and for millions of years thereafter.

10.4 Polyploidy and Ecological Novelty

The foregoing discussion of the genomic and genetic attributes of allopolyploid cotton demonstrates that polyploid formation has led to a diverse array of genetic and genomic responses, including non-Mendelian transmission and various forms of non-additive gene expression. The question naturally arises as to whether allopolyploidy has stimulated novel adaptation or physiological capability. A voluminous literature in plants documents the frequency of polyploids in various habitats, their morphological and physiological attributes, and their ecological success relative to diploids (Grant 1981; Stebbins 1947, 1950; Soltis and Soltis 2000). One generalization that has emerged is that polyploidy often is associated with broader ecological amplitude and novel evolutionary opportunity, often suggested to be mediated by the increased "buffering" capacity afforded by duplicated genes and the enhanced vigor resulting from the "fixed heterozygosity" of their duplicated genomes. We might now rephrase or expand these suggested mechanistic explanations to encompass a network perspective, and the vastly increased combinatorial possibilities for regulation and evolution enabled by a suddenly duplicated complement of genes and merged regulatory systems.

With respect to *Gossypium*, allopolyploidy led to the apparent invasion of a new ecological niche. In considering the Pleistocene origin of allopolyploid cotton, Fryxell (1965, 1979) noted that in contrast to the majority of diploid species, which occur mostly inland in various arid to seasonally arid environments, allopolyploid species typically occur in coastal habitats, at least those forms that arguably are truly wild (see also Brubaker and Wendel 1994). Thus, among the five allopolyploid species, two are completely restricted to near coastlines, in that they are island endemics (*G. darwinii* and *G. tomentosum*), and for two others (*G. barbadense* and *G. hirsutum*), wild forms occur disparately in littoral habitats ringing the Gulf of Mexico, northwest South America, and distant Pacific Islands. The capacity for oceanic dispersal in *Gossypium* (Fryxell 1965, 1979; Stephens 1958, 1966) was associated at the allopolyploid level with specialization for establishment in coastal communities. Fryxell (1965, 1979) forwarded the tantalizing suggestion that following initial formation, adaptation of the newly evolved

allopolyploid to littoral habitats enabled it to exploit the fluctuating sea levels that characterized the Pleistocene. This ecological innovation not only is envisioned to have permitted the initial establishment of the nascent polyploid lineage, but is also suggested to have provided a means for the rapid dispersal of the salt-water-tolerant seeds. By this means, perhaps, the mobile shorelines of the Pleistocene facilitated exploitation of a new ecological niche, and hence colonization of the New World tropics.

10.5 Polyploidy and Fiber

Finally, we would be remiss if we failed to consider the consequences of polyploidy on the development of agronomically advanced cultivars of allopolyploid cotton. Although four separate species of *Gossypium* were independently domesticated for their seed hairs, the characteristic that attracted the attention of the earliest domesticators, the seed "lint" itself, evolved only once in the progenitor of the A-genome diploids (Hovav et al. 2008c; Wendel et al. 2009). Applequist et al. (2001) generated growth curves for trichomes from cultivated and wild diploid and allopolyploid species and demonstrated that the evolution of an extended primary wall elongation occurred in the ancestor of wild A-genome cotton prior to domestication and in Africa. Follow-up comparative expression profiling experiments (Hovav et al. 2008a, b; Rapp et al. 2010) further identify some of the metabolic pathways that were modified to enable this evolution in fiber properties. These results led to the fascinating implication that domestication of the New World allopolyploid cottons (which contain an A-genome, in addition to a D-genome) that presently dominate cotton agriculture worldwide was first precipitated by developmental and physiological transformations that occurred hundreds of thousands of years ago in a different hemisphere.

Because fibers from all D-genome diploids are short and non-spinnable, it is particularly interesting that fiber from the cultivated (New World) allopolyploids is agronomically superior to that of the cultivated A-genome diploids; in this sense, the cultivated allotetraploid fiber morphology is "non-additive", or perhaps "heterotic". A number of studies have noted this point (Jiang et al. 1998; Paterson 2005; Wright et al. 1998), suggesting that allopolyploidization provided novel opportunities for agronomic improvement. A recent meta-analysis (Rong et al. 2007) of a large number of QTL studies in allopolyploid cotton leads to a general picture consistent with this interpretation; more loci affecting fiber yield and quality traits are found in the D_T ($n = 221$) than the A_T ($n = 184$) genome, possibly explaining the superiority of the lint of the allopolyploids relative to the A-genome diploids. Support for this speculation that "recruitment" of D-genome genes has been important in enabling the development of advanced allotetraploid cultivars is also emerging from comparative expression profiling studies (Hovav et al. 2008a, b; Rapp et al. 2010), which reveal in exquisite detail the thousands of gene expression differences that distinguish wild from domesticated cotton fiber

development, as well as a bias toward preferential expression of D-genome homoeologs. Finally, similar implications emanate from genetic mapping experiments, where it has been shown that for 535 genes implicated in cotton fiber development, more transcription factors were from D_T than A_T genome, whereas the reverse was true for fiber development genes (Xu et al. 2010). These data are interpreted to suggest that the D-genome ancestor provided key transcription factors that regulate the expression of fiber genes donated by the ancestral A-genome parent. Taken together, these studies may provide actual genetic evidence for a speculation forwarded 75 years ago by Harland (1936), who stated *If as a consequence of polyploidy a large number of genes become duplicated, and the characters governed by such genes are of importance to the species, one of the members may mutate, leaving the character unimpaired, with the further possibility that the mutation may be of benefit to the species*. An exciting prospect is that in the near future we will develop a deeper understanding of the nature of these genes, the molecular genetic meaning of Harland's invocation of the word "mutation", and their effects on the developmental networks that underlie altered morphology and agronomic improvement.

Acknowledgments Research in the Wendel lab has largely been funded by the NSF Plant Genome Program, with additional support from other NSF programs, the USDA NRI, and Cotton Incorporated. Research in the Adams lab has been supported by the Natural Science and Engineering Research Council of Canada and by the USDA NRI program. We gratefully acknowledge all of these sources of support.

References

Adams KL, Cronn R, Percifield R, Wendel JF (2003) Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proc Nat Acad Sci USA 100:4649–4654

Adams KL, Percifield R, Wendel JF (2004) Organ-specific silencing of duplicated genes in a newly synthesized cotton allotetraploid. Genetics 168:2217–2226

Applequist WL, Cronn RC, Wendel JF (2001) Comparative development of fiber in wild and cultivated cotton. Evol Dev 3:3–17

Bardil A, Dantas de Almeida J, Combes MC, Lashermes P, Bertrand B (2011) Genomic expression dominance in the natural allopolyploid *Coffea arabica* is massively affected by growth temperature. New Phytol (in press)

Beasley JO (1941) Hybridization, cytology, and polyploidy of *Gossypium*. Chronica Bot 6:394–395

Bennett MD, Leitch IJ (2005) Genome size evolution in plants. In: Gregory TR (ed) The evolution of the genome. Elsevier, San Diego, pp 89–162

Birchler JA, Riddle NC, Auger DL, Veitia RA (2005) Dosage balance in gene regulation: biological implications. Trends Genet 21:219–226

Birchler JA, Veitia RA (2010) The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. New Phytol 186:54–62

Bottley A, Koebner RM (2008) Variation for homoeologous gene silencing in hexaploid wheat. Plant J 56:297–302

Bottley A, Xia G, Koebner R (2006) Homoeologous gene silencing in hexaploid wheat. Plant J 47:897–906

Brubaker CL, Bourland FM, Wendel JF (1999a) The origin and domestication of cotton. In: Smith CW, Cothren JT (eds) Cotton; origin, history, technology and production. Wiley, New York, pp 3–31

Brubaker CL, Paterson AH, Wendel JF (1999b) Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. Genome 42:184–203

Brubaker CL, Wendel JF (1993) On the specific status of *Gossypium lanceolatum* Todaro. Genet Res Crop Evol 40:165–170

Brubaker CL, Wendel JF (1994) Reevaluating the origin of domesticated cotton (*Gossypium hirsutum*; Malvaceae) using nuclear restriction fragment length polymorphisms (RFLPs). Am J Bot 81:1309–1326

Brubaker CL, Wendel JF (2001) RFLP diversity in cotton. In: Jenkins JN, Saha S (eds) Genetic improvement of cotton: emerging technologies. Science Publishers, Inc., Enfield, pp 81–102

Buggs RJA, Elliott NM, Zhang L, Koh J, Viccini LF, Soltis DE, Soltis PS (2010) Tissue-specific silencing of homoeologs in natural populations of the recent allopolyploid Tragopogon mirus. New Phytol 186:175–183

Chaudhary B, Flagel L, Stupar RM, Udall JA, Verma N, Springer NM, Wendel JF (2009) Reciprocal silencing, transcriptional bias and functional divergence of homeologs in polyploid cotton (*Gossypium*). Genetics 182:503–517

Chen ZJ (2007) Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. Annu Rev Plant Biol 58:377–406

Cronn R, Small RL, Wendel JF (1999) Duplicated genes evolve independently following polyploid formation in cotton. Proc Nat Acad Sci USA 96:14406–14411

Cronn RC, Zhao X, Paterson AH, Wendel JF (1996) Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. J Mol Evol 42(6):685–705

DeJoode DR, Wendel JF (1992) Genetic diversity and origin of the Hawaiian Islands cotton, *Gossypium tomentosum*. Amer J Bot 79:1311–1319

Dong S, Adams KL (2011) Differential contributions to the transcriptome of duplicated genes in response to abiotic stresses in natural and synthetic polyploids. New Phytol 190(4):1045–1057

Elder JF, Turner BJ (1995) Concerted evolution of repetitive DNA sequences in eukaryotes. Quart Rev Biol 70:297–320

Endrizzi JE, Turcotte EL, Kohel RJ (1985) Genetics, cytology, and evolution of *Gossypium*. Adv Genet 23:271–375

Flagel L, Udall J, Nettleton D, Wendel J (2008) Duplicate gene expression in allopolyploid *Gossypium* reveals two temporally distinct phases of expression evolution. BMC Biol 6:11

Flagel LE, Wendel JF (2010) Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. New Phytol 186:184–193

Flagel LE, Wendel JF, Udall JA (2012) Duplicate gene evolution, homoeologous recombination, and transcriptome characterization in allopolyploid cotton. BMC Genomics (in press)

Fryxell PA (1965) Stages in the evolution of *Gossypium*. Adv Frontiers Plant Sci 10:31–56

Fryxell PA (1968) A redefinition of the tribe *Gossypieae*. Bot Gaz 129:296–308

Fryxell PA (1979) The natural history of the cotton tribe Texas. A&M University Press College Station, College Station

Fryxell PA (1992) A revised taxonomic interpretation of *Gossypium* L. (Malvaceae). Rheedea 2:108–165

Galau GA, Wilkins TA (1989) Alloplasmic male sterility in AD allotetraploid *Gossypium hirsutum* upon replacement of its resident a cytoplasm with that of D species *G.* harknessii. Theor Appl Genet 78:23–30

Grant V (1981) Plant speciation. Columbia University Press, New York

Grover CE, Kim H, Wing RA, Paterson AH, Wendel JF (2004) Incongruent patterns of local and global genome size evolution in cotton. Genome Res 14:1474–1482

Grover CE, Kim H, Wing RA, Paterson AH, Wendel JF (2007) Microcolinearity and genome evolution in the AdhA region of diploid and polyploid cotton (*Gossypium*). Plant J 50: 995–1006

Hanson RE, Islam-Faridi MN, Crane CF, Zwick MS, Czeschin DG, Wendel JF, Mcknight TD, Price HJ, Stelly DM (1999) Ty1-copia-retrotransposon behavior in a polyploid cotton. Chromosome Res 8:73–76

Hanson RE, Zhao X-P, Islam-Faridi MN, Paterson AH, Zwick MS, Crane CF, McKnight TD, Stelly DM, Price HJ (1998) Evolution of interspersed repetitive elements in *Gossypium* (Malvaceae). Am J Bot 85:1364–1368

Harland SC (1936) The genetical conception of the species. Cambridge Philos Soc Biol Rev 11:83–112

Hawkins JS, Hu G, Rapp RA, Grafenberg JL, Wendel JF (2008) Phylogenetic determination of the pace of transposable element proliferation in plants: copia and LINE-like elements in *Gossypium*. Genome 51:11–18

Hawkins JS, Kim H, Nason JD, Wing RA, Wendel JF (2006) Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. Genome Res 16(10):1252–1261

Hawkins JS, Proulx SR, Rapp RA, Wendel JF (2009) Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants. Proc Nat Acad Sci USA 106(42):17811–17816

Hendrix B, Stewart JM (2005) Estimation of the nuclear DNA content of *Gossypium* species. Ann Bot 95:789–797

Hovav R, Chaudhary B, Udall JA, Flagel L, Wendel JF (2008a) Parallel domestication, convergent evolution and duplicated gene recruitment in allopolyploid cotton. Genetics 179(3):1725–1733

Hovav R, Udall J, Chaudhary B, Flagel L, Rapp R, Wendel J (2008b) Partitioned expression of duplicated genes during development and evolution of a single cell in a polyploid plant. Proc Nat Acad Sci USA 105:6191

Hovav R, Udall JA, Chaudhary B, Hovav E, Flagel L, Hu G, Wendel JF (2008c) The evolution of spinnable cotton fiber entailed prolonged development and a novel metabolism. PLoS Genet 4:e25

Hu G, Hawkins JS, Grover CE, Wendel JF (2010) The history and disposition of transposable elements in polyploid *Gossypium*. Genome 53:599–607

Hu G, Houston NL, Pathak D, Schmidt L, Thelan JJ, Wendel JF (2011) Genomically biased accumulation of seed storage proteins in allopolyploid cotton. Genetics 189:1103–1115

Hutchinson JB (1951) Intra-specific differentiation in *Gossypium hirsutum*. Heredity 5:161–n193

Hutchinson JB (1954) New evidence on the origin of the old world cottons. Heredity 8: 225–241

Hutchinson JB, Silow RA, Stephens SG (1947) The evolution of *Gossypium* and the differentiation of the cultivated cottons. Oxford University Press, London

Jiang C, Wright R, El-Zik K, Paterson A (1998) Polyploid formation created unique avenues for response to selection in *Gossypium* (cotton). Proc Nat Acad Sci USA 95:4419–4424

Kashkush K, Feldman M, Levy AA (2002) Gene loss, silencing, and activation in a newly synthesized wheat allotetraploid. Genetics 160:1651–1659

Krapovickas A, Seijo G (2008) *Gossypium ekmanianum* (Malvaceae), algodon silvestre de la Republica Dominicana. Bonplandia 17:55–63

Lin L, Paterson AH (2009) Physical composition and organization of the *Gossypium* genomes. In: Paterson AH (ed) Genomics of cotton, plant genetics and genomics, crops and models 3. Springer, New York, pp 141–156

Lin LF, Tang HB, Compton RO, Lemke C, Rainville LK, Wang XY, Rong JK, Rana MK, Paterson AH (2011) Comparative analysis of *Gossypium* and Vitis genomes indicates genome duplication specific to the *Gossypium* lineage. Genomics 97:313–320

Liu B, Brubaker CL, Mergeai G, Cronn RC, Wendel JF (2001) Polyploid formation in cotton is not accompanied by rapid genomic changes. Genome 44:321–330

Liu Z, Adams KL (2007) Expression partitioning between genes duplicated by polyploidy under abiotic stress and during organ development. Curr Biol 17:1669–1674

Muravenko O, Fedotov AR, Punina EO, Federova LI, Grif VG, Zelenin AV (1998) Comparison of chromosome BrdU-hoechst-giemsa banding patterns of the A_1 and $(AD)_2$ genomes of cotton. Genome 41:616–625

Osborn TC, Chris Pires J, Birchler JA, Auger DL, Jeffery Chen Z, Lee H-S, Comai L, Madlung A, Doerge RW, Colot V, Martienssen RA (2003) Understanding mechanisms of novel gene expression in polyploids. Trends Genet 19(3):141–147

Paterson AH (2005) Polyploidy, evolutionary opportunity, and crop adaptation. Genetica 123:191

Paterson AH (2009) Genomics of cotton, plant genetics and genomics, crops and models 3. Springer, New York

Percy RG, Wendel JF (1990) Allozyme evidence for the origin and diversification of *Gossypium barbadense* L. Theor Appl Genet 79:529–542

Rapp R, Haigler C, Flagel L, Hovav R, Udall J, Wendel J (2010) Gene expression in developing fibres of Upland cotton (*Gossypium hirsutum* L.) was massively altered by domestication. BMC Biol 8:139

Rapp R, Wendel J (2005) Epigenetics and plant evolution. New Phytol 168:81

Rapp RA, Udall JA, Wendel JF (2009) Genomic expression dominance in allopolyploids. BMC Biol 7:18

Reinisch AJ, Dong J, Brubaker CL, Stelly DM, Wendel JF, Paterson AH (1994) A detailed RFLP map of cotton, *Gossypium hirsutum* x *G*. barbadense: chromosome organization and evolution in a disomic polyploid genome. Genetics 138:829–847

Rong J, Abbey C, Bowers JE, Brubaker CL, Chang C, Chee PW, Delmonte TA, Ding X, Garza JJ, Marler BS, Park C, Pierce GJ, Rainey KM, Rastogi VK, Schulze SR, Trolinder NL, Wendel JF, Wilkins TA, Williams-Coplin TD, Wing RA, Wright RJ, Zhao X, Zhu L, Paterson AH (2004) A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). Genetics 166:389–417

Rong J, Feltus EA, Waghmare VN, Pierce GJ, Chee PW, Draye X, Saranga Y, Wright RJ, Wilkins TA, May OL, Smith CW, Gannaway JR, Wendel JR, Paterson AH (2007) Meta-analysis of polyploid cotton QTL shows unequal contributions of subgenomes to a complex network of genes and gene clusters implicated in lint fiber development. Genetics 176(4):2577–2588

Rong J, Feltus FA, Liu L, Lin L, Paterson AH (2010) Gene copy number evolution during tetraploid cotton radiation. Heredity 105(5):463–472

Salmon A, Flagel L, Ying B, Udall JA, Wendel JF (2010) Homoeologous nonreciprocal recombination in polyploid cotton. New Phytol 186:123–134

Saunders JH (1961) The wild species of *Gossypium* and their evolutionary history. Oxford University Press, London

Seelanan T, Brubaker CL, Stewart JM, Craven LA, Wendel JF (1999) Molecular systematics of Australian *Gossypium* section grandicalyx (*Malvaceae*). Syst Bot 24:183–208

Senchina DS, Alvarez I, Cronn RC, Liu B, Rong JK, Noyes RD, Paterson AH, Wing RA, Wilkins TA, Wendel JF (2003) Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. Mol Biol Evol 20:633–643

Shan X, Liu Z, Dong Z, Wang Y, Chen Y, Lin X, Long L, Han F, Dong Y, Liu B (2005) Mobilization of the active MITE transposons mPing and Pong in rice by introgression from wild rice (Zizania latifolia Griseb.). Mol Biol Evol 22:976–990

Small RL, Ryburn JA, Wendel JF (1999) Low levels of nucleotide diversity at homoeologous Adh loci in allotetraploid cotton (*Gossypium* L.). Mol Biol Evol 16:491–501

Small RL, Wendel JF (1999) The mitochondrial genome of allotetraploid cotton (*Gossypium* L.). J Hered 90:251–253

Small RL, Wendel JF (2002) Differential evolutionary dynamics of duplicated paralogous Adh loci in allotetraploid cotton (*Gossypium*). Mol Biol Evol 19:597–607

Soltis PS, Soltis DE (2000) The role of genetic and genomic attributes in the success of polyploids. Proc Nat Acad Sci USA 97:7051–7057

Stebbins GL (1947) Types of polyploids: their classification and significance. Adv Genet 1: 403–429
Stebbins GL (1950) Variation and evolution in plants. Columbia University Press, New York
Stephens SG (1958) Salt water tolerance of seeds of *Gossypium* species as a possible factor in seed dispersal. Amer Nat 92:83–92
Stephens SG (1966) The potential for long range oceanic dispersal of cotton seeds. Amer Nat 100:199–210
Stewart JM (1995) Potential for crop improvement with exotic germplasm and genetic engineering. In: Constable GA, Forrester NW (eds) Challenging the future: proceedings of the world cotton research, CSIRO, Melbourne, pp 313–327
Stewart JM, Oosterhuis D, Heithholt JJ, Mauney JR (2010) Physiology of cotton. Springer, The Netherlands
Udall JA, Swanson JM, Nettleton D, Percifield RJ, Wendel JF (2006) A novel approach for characterizing expression levels of genes duplicated by polyploidy. Genetics 173(3): 1823–1827
Ungerer MC, Strakosh SC, Zhen Y (2006) Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. Curr Biol 16R:872–873
Veitia RA (2005) Gene dosage balance: deletions, duplications and dominance. Trends Genet 21:33
Wang JL, Tian L, Lee HS, Wei NE, Jiang HM, Watson B, Madlung A, Osborn TC, Doerge RW, Comai L, Chen ZJ (2006) Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. Genetics 172:507–517
Watt G (1907) The wild and cultivated cotton plants of the world. Longmans, Green and Co, London
Wendel JF (1989) New world tetraploid cottons contain old world cytoplasm. Proc Nat Acad Sci USA 86:4132–4136
Wendel JF (2000) Genome evolution in polyploids. Plant Mol Biol 42:225–249
Wendel JF, Brubaker CL, Alvarez JP, Cronn RC, Stewart JM (2009) Evolution and natural history of the cotton genus. In: Paterson AH (ed) Genomics of cotton, plant genetics and genomics, crops and models 3. Springer, New York, pp 3–22
Wendel JF, Cronn RC (2003) Polyploidy and the evolutionary history of cotton. Adv Agron 78:139–186
Wendel JF, Rowley R, Stewart JM (1994) Genetic diversity in and phylogenetic relationships of the Brazilian endemic cotton, *Gossypium mustelinum* (*Malvaceae*). Pl Syst Evol 192:49–59
Wendel JF, Schnabel A, Seelanan T (1995) Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). Proc Nat Acad Sci USA 92:280–284
Wright RJ, Thaxton PM, El-Zik KM, Paterson AH (1998) D-subgenome bias of Xcm resistance genes in tetraploid *Gossypium* (cotton) suggests that polyploid formation has created novel avenues for evolution. Genetics 149:1987–1996
Xu Z, Yu JZ, Cho J, Yu J, Kohel RJ, Percy RG (2010) Polyploidization altered gene functions in cotton (*Gossypium* spp.). PLoS ONE 5:e14351
Yang SS, Cheung F, Lee JJ, Ha M, Wei NE, Sze SH, Stelly DM, Thaxton P, Triplett B, Town CD, Chen ZJ (2006) Accumulation of genome-specific transcripts, transcription factors and phytohormonal regulators during early stages of fiber cell development in allotetraploid cotton. Plant J 47:761–775

Chapter 11
Evolutionary Implications of Genome and Karyotype Restructuring in *Nicotiana tabacum* L.

Ales Kovarik, Simon Renny-Byfield, Marie-Angèle Grandbastien and Andrew Leitch

Abstract *Nicotiana tabacum* is an allopolyploid that formed within the last 200,000 years from relatives of the extant diploids *N. sylvestris* and *N. tomentosiformis*, the donors of the S- and T-genomes, respectively. Here we review progress in our understanding of the divergence of *N. tabacum* subsequent to its formation, by comparing the *N. tabacum* genome with those of its diploid progenitors. We also review the data from synthetic *N. tabacum*, where there is evidence for much genetic change in early generations, including various chromosomal translocations, allopolyploid-induced retroelement mobility and loss, and reductions in the copy numbers of some tandem repeats. These observations are similar to patterns found in natural *N. tabacum*, suggesting that rapid genetic divergence is induced by allopolyploidy. The T-genome of *N. tabacum* shows the greatest number of genetic changes and appears to be less stable than the S-genome. We describe possible mechanisms that may have stimulated these genetic changes and propose that these can lead to enhanced fertility, more regular chromosome pairing, and the evolution of disomic inheritance.

A. Kovarik
Institute of Biophysics, Brno, Czech Republic

S. Renny-Byfield · A. Leitch (✉)
School of Biological and Chemical Sciences, Queen Mary University of London, London, UK
e-mail: a.r.leitch@qmul.ac.uk

M.-A. Grandbastien
Institute Jean-Pierre Bourgin, Versailles, France

P. S. Soltis and D. E. Soltis (eds.), *Polyploidy and Genome Evolution*,
DOI: 10.1007/978-3-642-31442-1_11,

11.1 Introduction to Nicotiana

The genus *Nicotiana* is the sixth largest genus in the family *Solanaceae* and includes 76 species found naturally in the Americas, Australia and surrounding islands, and with a single species in Namibia (Goodspeed 1954; Knapp et al. 2004). However, several species, including *N. tabacum* (tobacco), now occur more widely. The genus has received much recent attention because of evidence for recurrent polyploidy, resulting in nearly half the species in the genus being chromosomally polyploid. It is likely there have been six independent polyploidy events within the genus: at about 10 million years ago (mya) (section *Suaveolentes*, *c*. 24 species), 5 mya (section *Repandae*, 4 species), 1–2 mya (section *Polydicleae*, 2 species), and within the last 200,000 years (*N. rustica*, *N. arentsii*, and *N. tabacum*) (Lim et al. 2004a). There is also reticulation at the diploid level, with at least three species likely to be of homoploid hybrid origin (*N. glauca*, *N. linearis*, and *N. spegazzinii*) (Goodspeed 1954; Kelly et al. 2010).

N. tabacum is an autotetraploid (2n = 4x = 48) derived from diploid species that most closely resemble *N. sylvestris* (2n = 2x = 24), the maternal S-genome donor, and *N. tomentosiformis* (2n = 2x = 24, section *Tomentosae*), the paternal T-genome donor. *Nicotiana sylvestris* is the only species in section *Sylvestres*, and detailed genetic analysis of multiple accessions indicates that there is little genetic diversity in this species (4.2 % polymorphisms in AFLPs—Amplified Fragment Length Polymorphisms, Petit et al. 2007). However, there is considerably more genetic diversity among accessions of *N. tomentosiformis* (31.7 % AFLP polymorphism), which form two distinct groups, one of which most closely resembles the T-genome of *N. tabacum* (Murad et al. 2002; Petit et al. 2007). These data indicate that *N. tabacum* formed subsequent to the divergence of the two groups of *N. tomentosiformis*.

Molecular clock estimates of plastid and internally transcribed spacer sequences (ITS) of nuclear ribosomal DNA (rDNA) suggest that *N. tabacum* formed less than 200,000 years ago. These estimates were calibrated using likely maximum ages of some endemic species occurring on islands of known geological ages (Clarkson et al. 2005; Leitch et al. 2008). However, given that only feral populations of *N. tabacum* have been found and no truly "wild" population exists (Knapp personal communication), it is quite possible that *N. tabacum* was born subsequent to the origin of human agriculture, within the last 10,000 years.

The purpose of this paper is to review current understanding of the genetic consequences of polyploidy in *N. tabacum*. We show that genetic change can occur rapidly within a few generations, perhaps as a consequence of the "genomic shock" of allopolyploidy (McClintock 1984), a process that generates variants from which selection favors those with enhanced fertility. This unstable phase is presumably transient; however, there remains evidence for genome dynamism over time scales of thousands of years, including the replacement of multiple copies of rDNA sequences, the turnover of retroelements, and the loss of sequences targeted at the T-genome. We discuss how such changes may promote fertility and lead to the fixation of rearranged karyotypes during polyploid species establishment.

11.2 *Nicotiana tabacum* Genome Structure

Kenton et al. (1993) were the first to use genomic in situ hybridization (GISH) to study the chromosomes of *N. tabacum*. The fluorescent probes derived from total genomic DNA of *N. sylvestris* and *N. tomentosiformis* hybridized to a separate subset of 24 chromosomes, corresponding to the chromosomes of the S- and T-genomes, respectively. However, the efficacy of the T-genome labeling was inferior (Lim et al. 2000b), likely because of a loss of T-genome sequences (Renny-Byfield et al. 2011). GISH also revealed 4-9 intergenomic translocations in all *N. tabacum* accessions (Kenton et al. 1993; Lim et al. 2004a; Moscone et al. 1996). Similar translocations (up to 3) were also observed in some synthetic *N. tabacum* lines that were only a few generations old (Skalicka et al. 2005). Perhaps these translocations have arisen as a consequence of multivalent formation (see Chap. 7), or they may represent hotspots of recombination. While some translocations are fixed (i.e. occur in all varieties), others are specific for particular accessions. There are more translocations of S-genome origin chromatin to T-genome chromosomes (T/s chromosome) than the reverse (S/t chromosome). Kenton et al. (1993) hypothesized that this could be caused by selection against S/t chromosomes. Alternatively, some of the "translocations" may actually be S-genome subtelomeric satellite sequences that now occur on T-genome chromosomes, perhaps arising in their new location via recombination-based homogenization processes (Koukalova et al. 2010). Next-generation sequencing (NGS) may enable us to distinguish among these alternative hypotheses.

As with most other plant species (Heslop-Harrison and Schwarzacher 2011), a large fraction of *Nicotiana* genomes is composed of various types of retroelements. Genome sampling using NGS revealed that the major component of the genome of *N. tabacum* and its diploid relatives comprises long terminal repeat (LTR)-retrotransposons, with at least 17.1–22.5 % of the genome being *Ty3/gypsy* elements and 2.2–3.4 % being *Ty1/copia* elements. DNA transposons comprise around 1.7 % of the genome of these species (Renny-Byfield et al. 2011).

In addition, there are several unrelated satellite repeat families that have been characterized (the distributions of some are shown Fig. 11.1, Lim et al. 2000b). These repeats include: (1) The HRS60 family, which is the best-characterized tandem repeat family in *Nicotiana* and includes an interstitial repeat in *N. tomentosiformis* and *N. tabacum* (called GRS, Gazdova et al. 1995) and predominantly subtelomeric repeats in *N. sylvestris* and *N. tabacum* (called HRS60 and NSYL2, Koukalova et al. 2010; Koukalova et al. 1989); (2) NTS9 in *N. sylvestris* and *N. tabacum* (Jakowitsch et al. 1998); (3) NTRS in *N. tomentosiformis* and *N. tabacum* (Fig. 11.1, Matyasek et al. 1997), and (4) *Nic*CL3, a long 2.2 kb tandem repeat comprising ~2 % of the *N. tomentosiformis* genome and in lower abundance in *N. tabacum* (Fig. 11.2, Renny-Byfield et al. 2011 and Renny-Byfield et al. 2012). There are also a number of satellite repeat families of known origin; these are: (5) A1/A2 satellite repeats derived from the intergenic spacer (IGS) of rDNA, which have transposed and amplified to multiple locations across the genome (Lim et al. 2004b); (6) tandem repeats of

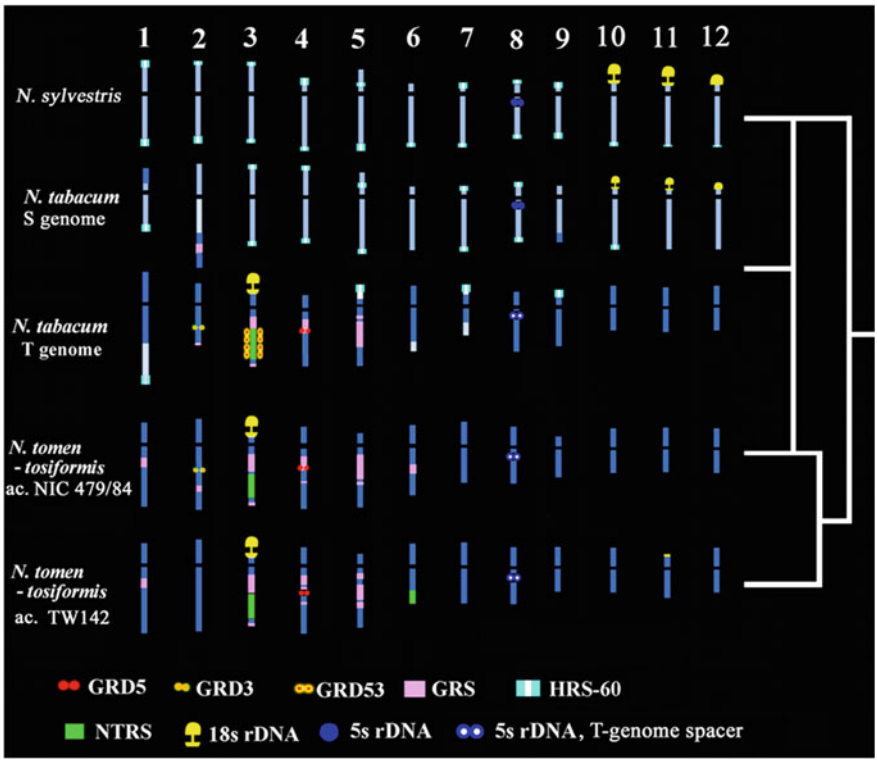

Fig. 11.1 Ideograms showing the distribution of tandem repeat sequences in *N. tabacum* and its diploid progenitors, *N. sylvestris* and *N. tomentosiformis*. Taken with permission from Murad et al. (2002). See text for a description of the repeats shown

geminivirus-related DNA (GRD), occurring in two distinct families in some accessions of *N. tomentosiformis* and inherited in *N. tabacum* (Murad et al. 2004; Murad et al. 2002), the first endogenous viruses discovered in plants (Bejarano et al. 1996); (7) sequences of pararetroviral origin (Matzke et al. 2004), including distinct variants found in *N. tomentosiformis* (NtoEPRV) and *N. sylvestris* (NsEPRV), both of which have been inherited in *N. tabacum*.

11.3 Retroelement Response to Allopolyploid "Genomic Shock"

Nicotiana hybrids were among the first in which chromosomal changes following hybridization were demonstrated. Gerstel and Burns (1967) reported that *N. otophora* ×*N. tabacum* hybrids have genetic instabilities of two kinds. Firstly, the heterochromatin from *N. otophora* undergoes breakage causing chromatin loss; this

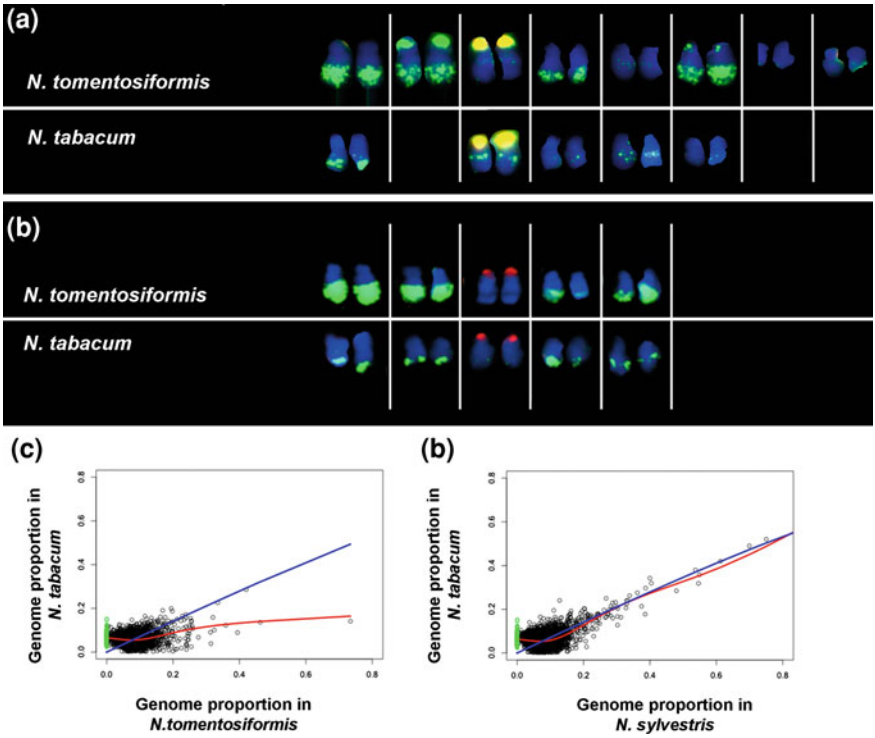

Fig. 11.2 Reduction in copy number of T-genome repeats in *N. tabacum* from expectation given their abundance in *N. tomentosiformis*. **a** Fluorescence in situ hybridization (FISH) showing the distribution of dispersed A1/A2 repeats (*green fluorescence*). The 18-5.8-26S rDNA locus on chromosome 3 (*N. tomentosiformis*) and T3 (*N. tabacum*) is labeled with *yellow fluorescence* due to the overlap of *red* and *green* rDNA fluorescence. Note that there is much reduced A1/A2 signal in *N. tabacum* (for further information see Lim et al. 2004b). **b** FISH showing the distribution of *Nic*CL3 in both species (*green fluorescence*). Note the reduced abundance of this repeat in *N. tabacum*. The 18-5.8-26S rDNA locus on chromosome 3 (*N. tomentosiformis*) and T3 (*N. tabacum*) is labeled with *red fluorescence*. For further information see Renny-Byfield et al. (2012). **c**, **d**. NGS analysis of repeat clusters in *N. tabacum* and its diploid progenitors. The graphs show the genome proportions (arcsine transformed) of repeats uniquely inherited in *N. tabacum* from **c** *N. tomentosiformis* and **d** *N. sylvestris*. If *N. tabacum* had faithfully inherited the repeats as found in the diploids, then all repeats would fall on the *blue lines*. However, in *N. tomentosiformis* **c** most repeats at higher genome proportions are underrepresented in *N. tabacum*. This is not the case for *N. sylvestris* **d**, where repeats are more or less in their expected abundance. This trend is reflected in the abundance of repeats that are biparentally inherited, i.e., the more abundant they are in *N. tomentosiformis*, the more likely they will be underrepresented in *N. tabacum*. For further information see Renny-Byfield et al. (2011)

can be observed cytologically and also phenotypically by the leaf variegation it causes. Secondly, in some cells heterochromatic blocks from *N. otophora* proliferate enormously to many times their normal length, forming "megachromosomes". This pioneering work revealed that interspecific hybridization can induce chromosomal rearrangements and rapid sequence losses and gains.

Melayah et al. (2001) were the first to show, using *N. tabacum*, that environmental stresses can induce retrotransposition. There is also evidence that allopolyploidy can stimulate transposition in *N. tabacum* and synthetic mimics of *N. tabacum* made from the diploid progenitors (Petit et al. 2007, 2010). These authors analyzed four populations of Ty1/copia LTR-retrotransposons (Tnt1-ol16, Tnt1-ol13, Tnt2, and Tto1) using sequence-specific amplification polymorphism (SSAP). They showed that the parental diploid species share essentially similar classes of retroelements. In *N. tabacum* there is evidence of retrotransposon diversification subsequent to allopolyploidy, with sequence losses concomitant with gains. Losses of retroelements were more frequent from the T-genome, while novel insertions of some populations, such as Tnt2, were shown to preferentially locate to the S-genome. However, each retrotransposon population seems to behave differently, with some populations undergoing rapid turnover, while others display relative stasis. There was little or no diversification of retrotransposons among individual accessions of natural *N. tabacum*, indicating that retroelement diversification occurred before the accessions analyzed had diverged, perhaps early in the species'evolution and arising as a consequence of allopolyploidy and induced "genomic shock" (McClintock 1984).

To test the hypothesis that allopolyploidy can induce retroelement diversification, Petit et al. (2010) compared insertion patterns of the Tnt1 family in the synthetic *N. tabacum* Th37 (Burk 1973) with its diploid progenitors. In some Th37 individuals in the fourth synthetic generation (S4), there was evidence of new Tnt1 insertions. Newly, transposed copies were amplified from elements located on the *N. sylvestris*-derived genome and were highly similar to the Tnt1A tobacco copies amplified in response to microbial factors (Grandbastien et al. 2005). Furthermore, a high proportion of parental SSAP bands was lost in Th37, particularly those from the *N. tomentosiformis*-derived genome, again as observed in natural *N. tabacum*. Together, these data indicate that retrotransposon amplification and molecular restructuring in, or around, insertion sites occur rapidly in response to allopolyploidy, and that similar sequences have responded in the same way to allopolyploidy in natural and synthetic *N. tabacum*.

11.4 Loss of DNA: Targeting the T-Genome

As with many other allopolyploid species (Leitch and Bennett 2004), *N. tabacum* has a lower genome size than is expected from the sum of parental genomes (Leitch et al. 2008, *N. tabacum* genome size, 1C = 5110 Mbp; *N. sylvestris*, 1C = 2636 Mbp; *N. tomentosiformis*, 1C = 2683 Mbp (http://data.kew.org/cvalues/). Recently, (Renny-Byfield et al. 2011) used NGS analysis of *N. tabacum* and its two diploid progenitors and showed that repeats derived from *N. tomentosiformis* was underrepresented in *N. tabacum*, a trend that was not observed for repeats from *N. sylvestris* (Fig. 11.2). These observations lead to the conclusion that the T-genome of

N. tabacum has undergone extensive sequence losses, and that genetic changes have occurred more rapidly to the paternally derived T-genome than the S-genome.

N. tabacum harbors two families of endogenous pararetrovirus DNA, one from each progenitor (Jakowitsch et al. 1999). These are NsEPRV from *N. sylvestris*, which in *N. tabacum* is at the expected abundance, and NtoEPRV from *N. tomentosiformis*, which occurs in *N. tabacum* with reduced copy number. Likewise, the A1/A2 repeats (Lim et al. 2004b) and *Nic*CL3 (Renny-Byfield et al. 2012), which are found in multiple locations around the genome in *N. tomentosiformis* (Fig. 11.2), are also underrepresented in *N. tabacum*. Similarly, NGS data show that all families of *Ty3/gypsy* elements have undergone sequence loss in *N. tabacum* (Renny-Byfield et al. 2011), and SSAP analysis of Tnt retroelements (*Ty1/copia*) reveals more deletion events influencing loci from the T-genome than the S-genome of *N. tabacum* (Petit et al. 2007). There is evidence of rapid and targeted loss of repeats from the T-genome even from the earliest generations, since in synthetic *N. tabacum* lines, many individuals already have reduced copy numbers of NtoEPRV, NTRS repeats (Skalicka et al. 2005), and *Nic*CL3 (Renny-Byfield et al. 2012), whilst most losses of Tnt1 retroelements were from the T-genome (Petit et al. 2010). All these data suggest rapid, targeted changes early in allopolyploid evolution. The targeting of repeat losses from the T-genome may support the nuclear-cytoplasmic interaction hypothesis, whereby the paternally derived genome functions in a maternal cytoplasm (Gill 1991; Leitch et al. 2006), which in allopolyploids may lead to incompatibilities that preferentially destabilize the paternally derived genome.

11.5 Ribosomal DNA Homogenization is Rapid and Ongoing

In most eukaryotes, 5S and 18–5.8–26S rDNA units occur in tandem arrays at one or several loci. Each large rDNA unit contains the 18S, 5.8S, and 26S ribosomal RNA (rRNA) genes, ITS, and IGS sequences. Whilst these sequences are vital for cell functioning, they can also be highly recombinogenic and labile sequences influencing genome stability (Kobayashi 2011). The genes themselves are highly conserved; however, their spacers diverge at suitable rates for resolving species relationships within most genera, including *Nicotiana* (Chase et al. 2003). Early studies revealed that the restriction fragment length polymorphism (RFLP) patterns of the *N. tabacum* IGS is not additive of the diploid progenitors (Kovarik et al. 1996), but that *N. tabacum* has evolved its own distinct rRNA gene family(ies). Sequence analysis revealed that the tobacco-specific units arose by reorganization of *N. tomentosiformis*-inherited units followed by their subsequent amplification (Volkov et al. 1999). The sequence changes mainly involved amplification and reduction of subrepeats upstream and downstream of the transcription start site within the IGS. However, the newly evolved units still occur at the four rDNA loci that *N. tabacum* inherited from its parents (see Fig. 11.1). Thus, it is likely that the parental units of S-genome origin were overwritten by the newly amplified

N.tabacum-specific units (Lim et al. 2000a), although a few units of S genome-origin remain intact, perhaps because they are methylated and inactive (Kovarik et al. 2008). In contrast, the 5S rDNA follows a different evolutionary pattern in *N. tabacum*, with no evidence for intergenomic homogenization, although shifts in copy numbers of parental gene families have been observed (Fulnecek et al. 2002).

11.6 The Fate of Duplicated Genes

Although there has been substantial restructuring of the repetitive fraction of the *N. tabacum* genome since its formation, most genic sequences analyzed have remained in duplicate copies, e.g., genes for drug resistance (Schenke et al. 2003), putrescine N-methytransferases (Riechers and Timko 1999), a family of small GTP-binding proteins (Takumi et al. 2002), lignin forming peroxidase (Matassi et al. 1991), nitrite reductase (Kronenberger et al. 1993), nitrate reductase (Vaucheret et al. 1989), glutamine synthetase (Clarkson et al. 2010), phytochrome A (Intrieri et al. 2008), the developmental gene LEAFY (McCarthy 2010), and families of DNA methyltransferases (Fulnecek et al. 2009). The only current exception is a family of *N. tabacum* glucan endo-1,3-beta-glucosidase genes that appeared to be recombinants of both ancestral sequences (Sperisen et al. 1991). Locus additivity does not always seem to be a rule even in recently formed synthetic allopolyploids. For example, frequent deletions of homoeologs were found among 70 protein-coding loci in *Tragopogon miscellus* that formed within the last 80 years (Buggs et al. 2009, 2012; see also Chap. 14, this volume). These changes may not be random, as Buggs et al. (2012) showed that clusters of genes are repeatedly lost or retained and the likelihood of retention reflects gene ontology categories or their predicted levels of dosage sensitivity. Similarly, synthetic lines of *Brassica napus* seem to eliminate much of the parental DNA, although it is currently unknown whether coding or noncoding sequences are preferentially targeted (Song et al. 1995; Szadkowski et al. 2011). The occurrence of both homoeologs in *N. tabacum* may reflect limited sampling, or high levels of retention. Additivity is observed in synthetic lines of *Gossypium* (Liu et al. 2002; see also Chap. 10, this volume) and in synthetic wheat allopolyploids, although in the latter case there remains some controversy (Mestiri et al. 2010; Feldman and Levy 2009; see also Chap. 7, this volume). Additivity has also been demonstrated at the expression level. Alleles of phytochrome A, lignin forming peroxidases, nitrite reductase, and DNA methyltransferases are expressed from both parental homeologues. There are also reports of epigenetic silencing of one subset of the parental alleles, including rRNA genes that have escaped homogenization (Kovarik et al. 2008) and the CYP82E4 locus (Chakrabarti et al. 2007) involved in the alkaloid biosynthesis pathway.

Duplicate copies arising through polyploidy can be retained through a number of mechanisms (Doyle et al. 2008): (1) one copy can evolve a new function (neofunctionalization) that can become fixed through a selective advantage

(Lynch et al. 2001); (2) duplicate copies can diverge through complementary loss of function at a particular point in the development or in particular tissues (subfunctionalization), a process that can occur through the action of drift alone; and (3) selection can occur against gene loss because that loss would compromise appropriate levels of gene product in relation to another gene(s) (gene balance hypothesis, cf. Birchler and Veitia 2010). In *N. tabacum*, a mechanism termed "nonfunctionalization" has been proposed which considers a combination of degenerative mutation and epigenetic silencing (Chakrabarti et al. 2007). Here, "non-functionalization" has altered alkaloid metabolism through reduced conversion of nicotine to nornicotine. Perhaps this mutation has been important to the "success" of *N. tabacum* as a recreational drug.

11.7 Genome Revolution and Allopolyploid Establishment

Until the early 1990s, it was commonly thought that chromosomal changes establish isolation barriers among populations because of reduced fertility in heterozygotes. Rearranged karyotypes may, then, become fixed through inbreeding or meiotic drive, particularly in small populations (King 1993). However, it is difficult to distinguish between karyotype changes that drive speciation events and those arising by genetic drift after speciation. Certainly, many rearrangements are deleterious and can cause embryo lethality or hybrid breakdown prior to reproduction (Rieseberg 1997). Only neutral or advantageous recombinants will pass the bottleneck of selection during the early phase of speciation. Furthermore, evolutionary geneticists pointed out that genetic mutations at the DNA level are more common than karyotype change and are therefore likely to be more important in speciation processes (Butlin 1993). However, recent evidence from a number of sources, e.g., *Anopheles gambiae* (Turner and Hahn 2010) and *Helianthus* homoploid hybrids (Strasburg et al. 2009), indicates that chromosomal inversions can, indeed, have a role in establishing barriers between populations and drive speciation. Strasburg et al. (2009) suggested that inversions may become foci for the accumulation of adaptive genes, particularly at the junctions between collinear and inverted regions, where gene flow is likely to be most impeded. Likewise, computer modeling has shown that when recombination is eliminated at an inverted region, and in the presence of strong selection and minimal or no gene flow, then species isolation can be driven by such a rearrangement (Feder and Nosil 2009).

In the context of early polyploid evolution, karyotype rearrangement may drive enhanced fertility and potentially be fundamental to allopolyploid species establishment. Certainly, high levels of chromosomal change have been observed in young polyploids, e.g., karyotype variability (dysploidy and intergenomic translocations) within and among populations of *Tragopogon* polyploids that formed within the last 80 years (Chester et al. 2010, 2012; Lim et al. 2008), and large-scale chromosomal deletions occurring in individual plants of early generation

synthetic allopolyploids of *B. napus* (Gaeta et al. 2007). Typically, a major hurdle that a newly formed polyploid must overcome is reduced fertility, often arising through multivalent formation, where chromosomes can segregate aberrantly leading to aneuploidy and reduction in fertility. We can envisage several allopolyploid-induced processes that may act to reduce multivalent formation and be favored by selection, these are:

- *Reduced chiasma frequency*. Selection may favor the formation of fewer and/or focussed chiasma at recombination hotspots. This would result in quadrivalents formed in prophase I falling into two bivalents in metaphase I and then segregating normally (Fig. 11.3a). Computer modeling has shown that reduced chiasma frequency results in the evolution of disomic inheritance (Le Comber et al. 2010), although we are unaware of empirical evidence to support this assertion.
- *Structural and epigenetic divergence of homeologues*. Allopolyploidy can trigger extensive karyotype and molecular restructuring in early generations (Gaeta et al. 2007; Renny-Byfield et al. 2011). This may result in genetic divergence of the parental genomes, e.g., through the loss or rearrangement of sequences shared by homeologues and recognized by the homolog recognition machinery in meiosis (Fig. 11.3b). Such sequences are unlikely to be those that are highly repeated across the genome since they are not chromosome specific. Nevertheless, genetic changes may introduce regions among homeologues that do not recombine, forming "islands of divergence", as in the establishment of homoploid hybrid species (Strasburg et al. 2009). Without recombination, the local regions around these islands will continue to diverge, leading to regular homolog pairing, and bivalent formation. Similarly, allopolyploidy can trigger epigenetic changes across the genome (Parisod et al. 2009). Epigenetic changes, which are thought to alter patterns of ectopic recombination (Colot et al. 1995), may also influence patterns of meiotic recombination and initiate islands of divergence. In addition, newly formed epi-alleles may establish tissue or temporal patterns of gene expression (subfunctionalization, Adams and Wendel 2005). Subfunctionalization is likely to favor the evolution of regular disomic inheritance and to select against multivalent formation. This is because multivalents result in the inappropriate segregation of homeoalleles and the generation of individuals with reduced fitness, because they have aberrant patterns of gene expression (Le Comber et al. 2010).
- *Intergenomic translocations*. In the context of a diploid, translocations among heterologues can promote quadrivalent formation, leading to chromosome loss, and reduced fertility. Similarly, in the context of polyploids, translocations and aneuploidy are associated with low fertility, as observed in synthetic *B. napus* lines (Xiong et al. 2011). However, it is possible that a more complex dynamic can occur (Fig. 11.3c). If homeologues already form multivalents, then it is conceivable that the gain of genetic material from a heterologous chromosome may promote bivalent formation, disomic inheritance, and increased fertility. Alternatively, complex rearrangements can result through cascades of induced

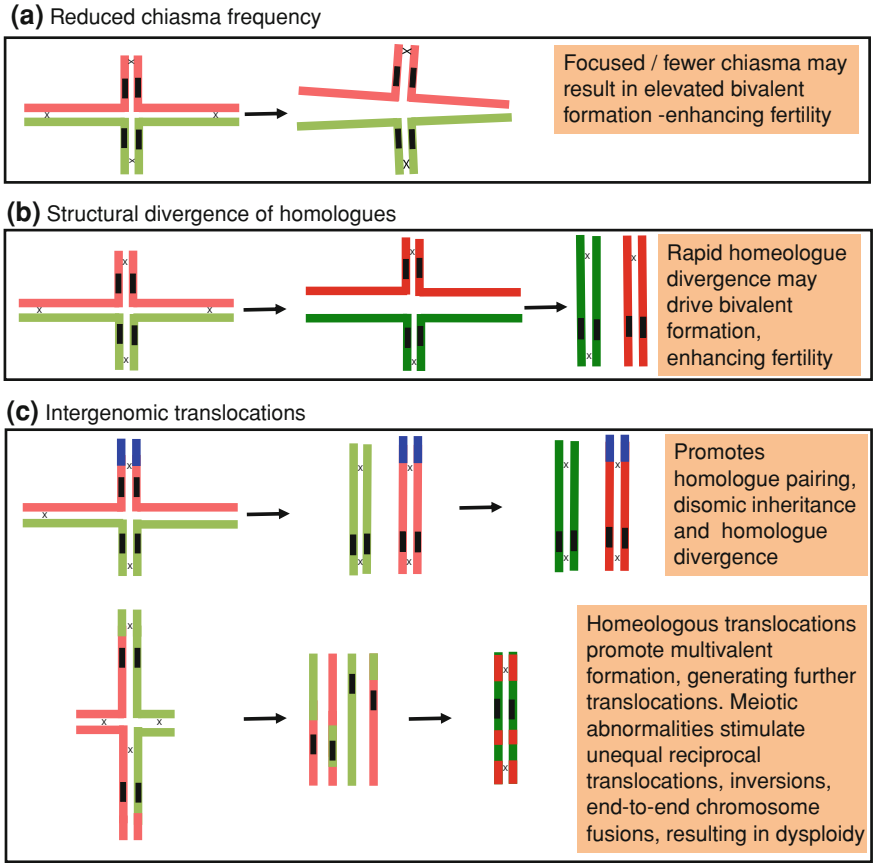

Fig. 11. 3 Fast molecular and chromosome evolution may enhance bivalent formation, driving disomic inheritance, and fertility in young polyploids. Homologs are shown in the same color (*pink* or *pale green*). Homeologues diverge over time, this is reflected in their colors becoming more distinct (*red* and *green*). A translocation from a heterologue is shown in *blue*. Crossovers among chromosomes are shown with crosses

pairing problems resulting from recurrent multivalent formation (unequal reciprocal translocations, inversions, end-to-end chromosome fusions), potentially generating novel chromosomes with segments from multiple ancestral chromosomes. Such a process may be associated with a reduction in chromosome number to a diploid-like level, as reported in the divergence of *Arabidopsis* karyotypes (Mandakova et al. 2010).

From *Nicotiana* allopolyploids, we have evidence for several of the proposed modes of chromosome evolution. First, in synthetic *N. tabacum*, we have observed plants that were homozygous for an intergenomic translocation that is similar to translocations observed in natural *N. tabacum* (Lim et al. 2004a). Perhaps this translocation is selected because it enhances fertility, and that in the synthetic

N. tabacum evolution is repeating itself. Second, the allotetraploids of sections *Polydicliae* and *Repandae*, which formed *c*. 1 mya and 5 mya, respectively, have chromosome numbers that are additive of the progenitor diploids ($2n = 4x = 48$). However, the sequence organization along their chromosomes has diverged considerably, particularly in the latter. Indeed, Lim et al. (2007) suggested that in some species of section *Repandae* there has been near complete genome turnover. Perhaps these changes contributed to the establishment of disomic inheritance. Finally, in section *Suaveolentes*, several species show evidence of chromosome number reduction from expectation (chromosome numbers range from $2n = 4x = 32$ to 46, depending on species, Goodspeed 1954; Knapp et al. 2004). We anticipate that this reduction is associated with karyotype restructuring, as in the polyploids of *Arabidopsis* (Mandakova et al. 2010).

11.8 Advantages of the *Nicotiana* System and Future Perspectives

Nicotiana provides many opportunities for studying allopolyploid genome divergence and addressing key questions facing polyploidy researchers [outlined in Soltis et al. (2010)]. It is important, even vital for the future of humankind, to know how allopolyploid genomes interact together. This is because many of our most important crop species are recognizably polyploid based on their chromosome numbers, and recent evidence is emerging that all seed-bearing plants have undergone at least one round of polyploidy in their ancestry (Jiao et al. 2011). The allopolyploid species of *Nicotiana* and the synthetic allopolyploids provide a unique opportunity to study snap shots of polyploid divergence over 10 million years of evolution.

Acknowledgments We thank NERC and the Czech Science Foundation for support.

References

Adams KL, Wendel JF (2005) Allele-specific, bidirectional silencing of an alcohol dehydrogenase gene in different organs of interspecific diploid cotton hybrids. Genetics 171:2139–2142

Bejarano ER, Khashoggi A, Witty M, Lichtenstein C (1996) Integration of multiple repeats of geminiviral DNA into the nuclear genome of tobacco during evolution. Proc Nat Acad Sci USA 93:759–764

Birchler JA, Veitia RA (2010) The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. New Phytol 186:54–62

Buggs RJ, Doust AN, Tate JA, Koh J, Soltis K, Feltus FA, Paterson AH, Soltis PS, Soltis DE (2009) Gene loss and silencing in *Tragopogon miscellus* (*Asteraceae*): comparison of natural and synthetic allotetraploids. Heredity 103:73–81

Buggs RJA, Chamala S, Wu W, Tate JA, Schnable PS, Soltis DE, Soltis PS, Barbazuk WB (2012) Rapid, repeated, and clustered loss of duplicate genes in allopolyploid plant populations of independent origin. Curr Biol 22:248–252

Burk LG (1973) Partial self-fertility in a theoretical amphiploid progenitor of *N. tabacum*. J Hered 64:348–350

Butlin RK (1993) Species evolution—the role of chromosome change—King, M. Nature 366:27

Chakrabarti M, Meekins KM, Gavilano LB, Siminszky B (2007) Inactivation of the cytochrome P450 gene CYP82E2 by degenerative mutations was a key event in the evolution of the alkaloid profile of modern tobacco. New Phytol 175:565–574

Chase MW, Knapp S, Cox AV, Clarkson JJ, Butsko Y, Joseph J, Savolainen V, Parokonny AS (2003) Molecular systematics, GISH and the origin of hybrid taxa in *Nicotiana* (*Solanaceae*). Ann Bot 92:107–127

Chester M, Gallagher JP, Symonds VV, Cruz da Silva AV, Mavrodiev EV, Leitch AR, Soltis PS and Soltis DE (2012) Extensive chromosomal variation in a recently formed natural allopolyploid species, Tragopogon miscellus (Asteraceae). Proceedings of the National Academy of Sciences of the United States of America, 109:1176-1181

Chester M, Sykorova E, Fajkus J, Leitch AR (2010) Single integration and spread of a Copia-like sequence nested in rDNA intergenic spacers of *Allium cernuum* (*Alliaceae*). Cytogenet Genome Res 129:35–46

Clarkson JJ, Kelly LJ, Leitch AR, Knapp S, Chase MW (2010) Nuclear glutamine synthetase evolution in *Nicotiana*: phylogenetics and the origins of allotetraploid and homoploid (diploid) hybrids. Mol Phylogenet Evol 55:99–112

Clarkson JJ, Lim KY, Kovarik A, Chase MW, Knapp S, Leitch AR (2005) Long-term genome diploidization in allopolyploid *Nicotiana* section *Repandae* (*Solanaceae*). New Phytol 168:241–252

Colot V, Goyon C, Faugeron G, Rossignol JL (1995) Methylation of repeated DNA sequences and genome stability in Ascobolus immersus. Can J Bot-Rev Canadienne De Botanique 73:S221–S225

Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF (2008) Evolutionary genetics of genome merger and doubling in plants. Ann Rev Genet 443–461

Feder JL, Nosil P (2009) Chromosomal inversions and species differences: when are geens affecting adaptive divergence and reproductive isolation expected to reside within inversions? Evolution 63:3061–3075

Feldman M, Levy AA (2009) Genome evolution in allopolyploid wheat-a revolutionary reprogramming followed by gradual changes. J Genet Genomics 36:511–518

Fulnecek J, Lim KY, Leitch AR, Kovarik A, Matyasek R (2002) Evolution and structure of 5S rDNA loci in allotetraploid *Nicotiana tabacum* and its putative parental species. Heredity 88:19–25

Fulnecek J, Matyasek R, Kovarik A (2009) Faithful inheritance of cytosine methylation patterns in repeated sequences of the allotetraploid tobacco correlates with the expression of DNA methyltransferase gene families from both parental genomes. Mol Genet Genomics 281:407–420

Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC (2007) Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. Plant Cell 19:3403–3417

Gazdova B, Siroky J, Fajkus J, Brzobohaty B, Kenton A, Parokonny A, Heslop-Harrison JS, Palme K, Bezdek M (1995) Characterization of a new family of tobacco highly repetitive DNA, GRS, specific for the *Nicotiana tomentosiformis* genomic component. Chromosome Res 3:245–254

Gerstel DU, Burns JA (1967) Phenotypic and chromosomal abnormalities associated with the introduction of heterochromatin from *Nicotiana otophora* into *N. tabacum*. Genetics 56:483–502

Gill BS (1991) Nucleocytoplasmic interaction (NCI) hypothesis of genome evolution and speciation in polyploid plants. In: Sasakuma T, KinoshitaT (ed.) Proceedings of the Kihara

memorial international symposium on cytoplasmic engineering in wheat, Kihara Memorial Foundation, Yokohama, Japan, pp 48–53
Goodspeed TH (1954) The genus *Nicotiana* Massachusetts. Chronica Botanica Company, USA
Grandbastien M, Audeon C, Bonnivard E, Casacuberta JM, Chalhoub B, Costa APP, Le QH, Melayah D, Petit M, Poncet C, Tam SM, Van Sluys MA, Mhiri C (2005) Stress activation and genomic impact of Tnt1 retrotransposons in *Solanaceae*. Cytogenetic Genome Res 110:229–241
Heslop-Harrison JS, Schwarzacher T (2011) Organisation of the plant genome in chromosomes. Plant J 66:18–33
Intrieri MC, Muleo R, Buiatti M (2008) *Phytochrome* A as a functional marker of phyletic relationships in *Nicotiana* genus. Biol Plant 52:36–41
Jakowitsch J, Mette MF, van der Winden J, Matzke MA, Matzke AJM (1999) Integrated pararetroviral sequences define a unique class of dispersed repetitive DNA in plants. Proc Nat Acad Sci USA 96:13241–13246
Jakowitsch J, Papp I, Matzke MA, Matzke AJM (1998) Identification of a new family of highly repetitive DNA, NTS9, that is located predominantly on the S9 chromosome of tobacco. Chromosome Res 6:649–659
Jiao YN, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang HY, Soltis PS, Soltis DE, Clifton SW, Schlarbaum SE, Schuster SC, Ma H, Leebens-Mack J, dePamphilis CW (2011) Ancestral polyploidy in seed plants and angiosperms. Nature 473:97–113
Kelly LJ, Leitch AR, Clarkson JJ, Hunter RB, Knapp S, Chase MW (2010) Intragenic recombination events and evidence for hybrid speciation in *Nicotiana* (*Solanaceae*). Mol Biol Evol 27:781–799
Kenton A, Parokonny AS, Gleba YY, Bennett MD (1993) Characterization of the *Nicotiana tabacum* L. genome by molecular cytogenetics. Mol Gen Genet 240:159–169
King M (1993) Species evolution the role of chromosome change. Cambridge University Press, Cambridge
Knapp S, Chase MW, Clarkson JJ (2004) Nomenclatural changes and a new sectional classification in *Nicotiana* (*Solanaceae*). Taxon 53:73–82
Kobayashi T (2011) Regulation of ribosomal RNA gene copy number and its role in modulating genome integrity and evolutionary adaptability in yeast. Cell Mol Life Sci 68:1395–1403
Koukalova B, Moraes AP, Renny-Byfield S, Matyasek R, Leitch AR, Kovarik A (2010) Fall and rise of satellite repeats in allopolyploids of *Nicotiana* over c. 5 million years. New Phytol 186:148–160
Koukalova B, Reich J, Matyasek R, Kuhrova V, Bezdek M (1989) A BamHI family of highly repeated DNA sequecnes of *Nicotiana tabacum*. Theor Appl Genet 78:77–80
Kovarik A, Dadejova M, Lim K, Chase M, JJ C, Knapp S, Leitch A (2008) Evolution of rDNA in Nicotiana allopolyploids: a potential link between rDNA homogenization and epigentics. Ann Bot 101:815–823
Kovarik A, Fajkus J, Koukalova B, Bezdek M (1996) Species-specific evolution of telomeric and rDNA repeats in the tobacco composite genome. Theor Appl Genet 92:1108–1111
Kronenberger J, Lepingle A, Caboche M, Vaucheret H (1993) Cloning and expession of distinct nitrite reductases in tobacco-leaves and roots. Mol Gen Genet 236:203–208
Le Comber SC, Ainouche ML, Kovarik A, Leitch AR (2010) Making a functional diploid: from polysomic to disomic inheritance. New Phytol 186:113–122
Leitch AR, Lim KY, Skalicka K, Kovarik A (2006) Nuclear cytoplasmic interaction hypothesis and the role of translocations in *Nicotiana allopolyploids*. In: Cigna AAD, M. Yerevan (eds.) Radiation risk estimates in normal and emergency situations book series: NATO security through science series B: Physics and Biophysics (ARMENIA) pp 319–326
Leitch IJ, Bennett MD (2004) Genome downsizing in polyploid plants. Biol J Linn Soc 82:651–663

Leitch IJ, Hanson L, Lim KY, Kovarik A, Chase MW, Clarkson JJ, Leitch AR (2008) The ups and downs of genome size evolution in polyploid species of *Nicotiana* (*Solanaceae*). Ann Bot 101:805–814

Lim KY, Kovarik A, Matyasek R, Bezdek M, Lichtenstein CP, Leitch AR (2000a) Gene conversion of ribosomal DNA in *Nicotiana tabacum* is associated with undermethylated, decondensed and probably active gene units. Chromosoma 109:161–172

Lim KY, Kovarik A, Matyasek R, Chase MW, Clarkson JJ, Grandbastien MA, Leitch AR (2007) Sequence of events leading to near-complete genome turnover in allopolyploid *Nicotiana* within five million years. New Phytol 175:756–763

Lim KY, Matyasek R, Kovarik A, Leitch AR (2004a) Genome evolution in allotetraploid *Nicotiana*. Biol J Linn Soc 82:599–606

Lim KY, Matyasek R, Lichtenstein CP, Leitch AR (2000b) Molecular cytogenetic analyses and phylogenetic studies in the *Nicotiana* section Tomentosae. Chromosoma 109:245–258

Lim KY, Skalicka K, Koukalova B, Volkov RA, Matyasek R, Hemleben V, Leitch AR, Kovarik A (2004b) Dynamic changes in the distribution of a satellite homologous to intergenic 26-18S rDNA spacer in the evolution of *Nicotiana*. Genetics 166:1935–1946

Lim KY, Soltis DE, Soltis PS, Tate J, Matyasek R, Srubarova H, Kovarik A, Pires JC, Xiong ZY, Leitch AR (2008) Rapid chromosome evolution in recently formed polyploids in *Tragopogon* (*Asteraceae*). Plos One, 3

Liu B, Brubaker CL, Mergeai G, Cronn RC, Wendel JF (2002) Polyploid formation in cotton is not accompanied by rapid genomic changes. Genome 44:321–330

Lynch M, O'Hely M, Walsh B, Force A (2001) The probability of preservation of a newly arisen gene duplicate. Genetics 159:1789–1804

Mandakova T, Joly S, Krzywinski M, Mummenhoff K, Lysak MA (2010) Fast diploidization in close mesopolyploid relatives of *Arabidopsis*. Plant Cell 22:2277–2290

Matassi G, Melis R, Macaya G, Bernardi G (1991) Compositional bimodality of the nuclear genome of tobacco. Nucleic Acids Res 19:5561–5567

Matyasek R, Gazdova B, Fajkus J, Bezdek M (1997) NTRS, a new family of highly repetitive DNAs specific for the T1 chromosome of tobacco. Chromosoma 106:369–379

Matzke M, Gregor W, Mette MF, Aufsatz W, Kanno T, Jakowitsch J, Matzke AJM (2004) Endogenous paragretroviruses of polyploid *Nicotiana tabacum* and its diploid progenitors, *N. sylvestris* and *N. tomentosiformis*. Biol J Linn Soc 82:627–638

McCarthy EW (2010) The effect of molecular and spectral reflectance evolution on *Nicotiana* polyploids of different ages. In School of Biological and Chemical Sciences. London: Ph.D. thesis from Queen Mary University of London, p 259

McClintock B (1984) The significance of responses of the genome to challenge. Science 226:792–801

Melayah D, Bonnivard E, Chalhoub B, Audeon C, Grandbastien MA (2001) The mobility of the tobacco Tnt1 retrotransposon correlates with its transcriptional activation by fungal factors. Plant Journal, 28:159-168

Mestiri I, Chague V, Tanguy AM, Huneau C, Huteau V, Belcram H, Coriton O, Chalhoub B, Jahier J (2010) Newly synthesized wheat allohexaploids display progenitor-dependent meiotic stability and aneuploidy but structural genomic additivity. New Phytol 186:101–186

Moscone EA, Matzke MA, Matzke AJM (1996) The use of combined FISH/GISH in conjunction with DAPI counterstaining to identify chromosomes containing transgene inserts in amphidiploid tobacco. Chromosoma 105:231–236

Murad L, Bielawski JP, Matyasek R, Kovarik A, Nichols RA, Leitch AR, Lichtenstein CP (2004) The origin and evolution of *Geminivirus*-related DNA sequences in *Nicotiana*. Heredity 92:352–358

Murad L, Lim KY, Christopodulou V, Matyasek R, Lichtenstein CP, Kovarik A, Leitch AR (2002) The origin of tobacco's T genome is traced to a particular lineage within *Nicotiana tomentosiformis* (*Solanaceae*). Am J Bot 89:921–928

Parisod C, Salmon A, Zerjal T, Tenaillon M, Grandbastien MA, Ainouche M (2009) Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in *Spartina*. New Phytol 184:1003–1015

Petit M, Guidat C, Daniel J, Denis E, Montoriol E, Bui QT, Lim KY, Kovarik A, Leitch AR, Grandbastien MA, Mhiri C (2010) Mobilization of retrotransposons in synthetic allotetraploid tobacco. New Phytol 186:135–147

Petit M, Lim KY, Julio E, Poncet C, de Borne FD, Kovarik A, Leitch AR, Grandbastien MA, Mhiri C (2007) Differential impact of retrotransposon populations on the genome of allotetraploid tobacco (*Nicotiana tabacum*). Mol Genet Genomics 278:1–15

Renny-Byfield S, Chester M, Kovarik A, Le Comber S, Grandbastien M, Deloger M, Nichols R, Macas J, Novak P, Chase M, Leitch A (2011) Next generation sequencing reveals genome downsizing in allotetraploid *Nicotiana tabacum*, predominantly through the elimination of paternally derived repetitive DNAs Molecular Biology and Evolution 28:2843–2854

Renny-Byfield S, Kovarik A, Chester M, Nichols RA, Macas J, Novak P and Leitch AR (2012) Independent, rapid and targeted loss of highly repetitive DNA in natural and synthetic allopolyploids of *Nicotiana tabacum*. Plos One, 7:e36963

Riechers DE, Timko MP (1999) Structure and expression of the gene family encoding putrescine N-methyltransferase in *Nicotiana tabacum*: new clues to the evolutionary origin of cultivated tobacco. Plant Mol Biol 41:387–401

Rieseberg LH (1997) Hybrid origins of plant species. Annu Rev Ecol Syst 28:359–389

Schenke D, Sasabe M, Toyoda K, Inagaki Y, Shiraishi T, Ichinose Y (2003) Genomic structure of the NtPDR1 gene, harboring the two miniature inverted-repeat transposable elements, NtToya1 and NtStowaway101. Genes Genet Syst 78:409–418

Skalicka K, Lim KY, Matyasek R, Matzke M, Leitch AR, Kovarik A (2005) Preferential elimination of repeated DNA sequences from the paternal, *Nicotiana tomentosiformis* genome donor of a synthetic, allotetraploid tobacco. New Phytol 166:291–303

Soltis DE, Buggs RJA, Doyle JJ, Soltis PS (2010) What we still don't know about polyploidy. Taxon 59:1387–1403

Song KM, Lu P, Tang KL and Osborn TC (1995) Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. Proceedings of the National Academy of Sciences of the United States of America, 92:7719-7723

Sperisen C, Ryals J, Meins F (1991) Comparison of cloned genes provides evidence for intergenomic exchange of DNA in the evolution of a tobacco glucan endo-1,3 beta-glucosidase gene family. Proc Nat Acad Sci USA 88:1820–1824

Strasburg JL, Scotti-Saintagne C, Scotti I, Lai Z, Rieseberg LH (2009) Genomic patterns of adaptive divergence between chromosomally differentiated sunflower species. Mol Biol Evol 26:1341–1355

Szadkowski E, Eber F, Huteau V, Lode M, Coriton O, Jenczewski E and Chevre AM (2011) Polyploid formation pathways have an impact on genetic rearrangements in resynthesized *Brassica napus*. New Phytologist, 191:884-894

Takumi S, Ida M, Haisa Y, Ando S, Nakamura C (2002) Genomic structure and homoeologous relationship of the two alpha-subunit genes of a heterotrimeric GTP-binding protein in tobacco. Genome 45:626–633

Turner TL, Hahn MW (2010) Genomic islands of speciation or genomic islands and speciation? Mol Ecol 19:848–850

Vaucheret H, Vincentz M, Kronenberger J, Caboche M, Rouze P (1989) Molecular cloning and characterization of the 2 homologous genes coding for nitrate reductase in tobacco. Mol Gen Genet 216:10–15

Volkov RA, Borisjuk NV, Panchuk II, Schweizer D, Hemleben V (1999) Elimination and rearrangement of parental rDNA in the allotetraploid *Nicotiana tabacum*. Mol Biol Evol 16:311–320

Xiong ZY, Gaeta RT, Pires JC (2011) Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. Proc Nat Acad Sci USA 108:7908–7913

Chapter 12
Polyploid Evolution in *Spartina*: Dealing with Highly Redundant Hybrid Genomes

M. Ainouche, H. Chelaifa, J. Ferreira, S. Bellot, A. Ainouche and A. Salmon

Abstract Polyploidy and recurrent interspecific hybridization represent major features of *Spartina* evolution, resulting in several superimposed divergent genomes that coexist in the currently living species. This chapter summarizes what we presently know about *Spartina* history, emphasizing the recent hybridization and polyploidization events that have important ecological and evolutionary consequences. Particular attention is devoted to the recent formation of the allododecaploid invasive *Spartina anglica*, a salt-marsh "ecosystem engineer" that resulted from hybridization between the hexaploid *S. alterniflora* (introduced from North America) and tetraploid *S. maritima* (a European native) and subsequent genome duplication of the F_1 hybrid *S. x townsendii* during the nineteenth century in Western Europe. Allopolyploidy was not accompanied by substantial restructuring of the parental genomes, as observed in some other allopolyploid systems. The major evolutionary events affect the regulatory systems controlling gene expression (including epigenetic regulation), which appear to have been profoundly altered by the merger of different genomes. Methodological challenges in exploring non-model, highly redundant genomes resulting from superimposed events of polyploidization (such as those encountered in *Spartina*) and the contribution of the new massive parallel sequencing technologies are discussed.

An erratum to this chapter is available at 10.1007/978-3-642-31442-1_19.

M. Ainouche (✉) · H. Chelaifa · J. Ferreira · S. Bellot · A. Ainouche · A. Salmon
University of Rennes 1, UMR CNRS 6553 Ecobio, Bât. 14A, Campus Scientifique de Beaulieu, 35042, Rennes Cedex, France
e-mail: malika.ainouche@univ-rennes1.fr

P. S. Soltis and D. E. Soltis (eds.), *Polyploidy and Genome Evolution*,
DOI: 10.1007/978-3-642-31442-1_12, © Springer-Verlag Berlin Heidelberg 2012

12.1 Introduction

Polyploidy and recurrent interspecific hybridization represent major features of *Spartina* evolution, resulting in several superimposed divergent genomes that coexist in the currently living species. Particularly fascinating is the rapid range expansion of the recently formed allododecaploid species *S. anglica* Hubbard that formed in Western Europe during the end of the nineteenth century. The ecological impact and genetic determinants of the spectacular propagation of this invasive species that is now (deliberately or accidentally) introduced on several continents generated an abundant literature (reviewed in Gray et al. 1990, Ainouche et al. 2004a, Triplet and Gallicé 2008, Ainouche et al. 2009) and much interest, as illustrated by the successive International Invasive *Spartina* Conferences or Forums held in 1990 (Seattle WA, USA), 1997 (Olympia, WA, USA), 2004 (San Francisco, CA, USA), and 2011 (Oakland, CA, USA). In the field of evolutionary biology, *Spartina* has long represented a textbook example of recent allopolyploid speciation in which the historical context of species formation is particularly well documented (Huskin 1930; Stebbins 1950).

In this chapter, we summarize the current state of knowledge about *Spartina* evolutionary history, including recent insights from evolutionary genetics and genomic approaches. We also examine how genome evolution following natural interspecific hybridization and polyploidization has contributed to diversification and adaptation. Methodological challenges in exploring highly redundant genomes resulting from superimposed events of polyploidization will be discussed.

12.2 Recurrent Reticulate Evolution and Polyploidy in *Spartina*

The grass genus *Spartina* ("cordgrasses") belongs to the Chloridoideae subfamily (Fig. 12.1), one of the most poorly understood lineages of the Poaceae. Divergence between *Spartina* and various grass model species is currently estimated as 35–40 MYA with *Sorghum*—maize–sugar cane (subfamily Panicoideae) and 50 MYA with rice (subfamily Erhartoideae) (Christin et al. 2008, but see Prasad et al. 2011). Phylogenetic relationships among genera of Chloridoideae are not fully resolved and still are under debate (Hilu and Alice 2001, GPWG 2001). In recent molecular phylogenies, *Spartina* appears closely related to *Sporobolus*, *Calamovilfa* and *Zoysia* (Columbus et al. 2007, Fortune et al. 2007). Species from these genera share a C_4-type photosynthetic system that evolved in the Chloridoideae 25–32 MYA (Christin et al. 2008). C_4 photosynthesis is generally considered an adaptation conferring higher productivity under warm temperatures. However, species of *Spartina* exhibit geographic distributions that cover a range of climatic conditions, from temperate to tropical–subtropical regions; the species exhibit salt and/or drought tolerance on coastal or inland marshes or sand dunes.

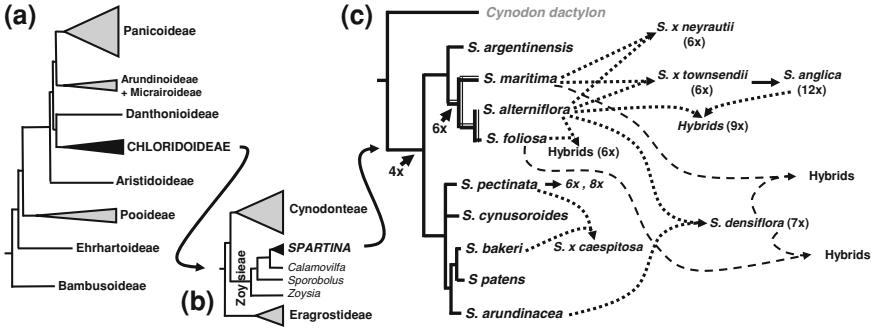

Fig. 12.1 Phylogeny of *Spartina* (see references in the text) **a** position within the grass family **b** position within the Chloridoideae subfamily **c** phylogenetic relationships among *Spartina* species, recurrent hybridization and polyploidy

Table 12.1 Chromosome numbers reported for *Spartina*. Additional chromosome numbers resulting from interspecific hybridization and/or backcrosses are presented in the text

Taxa	2n	References
S. patens (Aiton) Muhl.	40	Marchant (1968a)
S. cynusoroides (L.) Roth	40	Marchant (1968a)
S. backeri Merr.	40	Marchant (1968a)
S. gracilis Trin.	40	Marchant (1968a)
S. arundinacea (Thouars) Carmich.	40	Marchant. (1968a)
S. pectinata Link	40	Marchant (1968a)
	40, 60, 80	Kim *et al*. (2010)
S. densiflora Brongn.	70	Ayres *et al*. (2008), Fortune *et al*. . (2008)
S. maritima (Curtis) Fern.	60	Marchant. (1968b)
S. foliosa Trin.	62	Ayres *et al*. (2008)
S. alterniflora Loisel.	62	Marchant (1968b)
S. x townsendii H & J Groves	62	Marchant (1968b)
S. x neyrautii Foucaud	62	Marchant (1977)
S. anglica C.E. Hubbard	120, 122, 124	Marchant. (1968b)

Spartina is composed of about 15 perennial species that have mostly diversified in the New World (Mobberley 1956). Accidental or deliberate introduction of species outside their native range over the past 150 years has accelerated diversification by facilitating hybridization with native species, introgression or speciation. The basic (haploid) chromosome number in this lineage is considered to be $x = 10$ (Marchant 1968a), and all *Spartina* species recorded to date are polyploid, ranging from tetraploids to dodecaploids (Table 12.1).

Molecular phylogenies based on nuclear and chloroplast DNA sequences indicate that *Spartina* comprises two main lineages that include tetraploid and hexaploid species, respectively (Baumel et al. 2002a). The tetraploid lineage is composed of species native to the New World; these species colonize coastal or

inland salt marshes from either North America (*S. patens*, *S. bakeri, S. gracilis*, *S. cynusoroides, S. pectinata*) or South America (*S. ciliata, S. arundinacea*). *S. argentinensis* (syn. *S. spartinae*), which has a disjunct distribution in Central America and South America, is sister to the hexaploid lineage (Fig. 12.1c). This hexaploid clade is composed of *S. maritima*, *S. alterniflora*, and *S. foliosa,* all colonizing low-marsh zones. *S. maritima,* native to the Atlantic coasts of Western Europe and Africa was, until the nineteenth century, the only Old World native species, if we exclude recent taxa of hybrid origin. *S. foliosa*, a species limited to the Pacific coast of North America (California and Mexico), is a weakly supported sister species to *S. alterniflora* (Baumel et al. 2002a; Ainouche et al. 2004b) and is distributed along the east coast of North and South America from Canada to Argentina. This species has now one of the largest geographic distributions in the genus, as it has been introduced to several continents (Europe, Asia).

The nature of polyploidy (auto- vs. allopolyploidy) and the origin of the hexaploid clade are not fully understood. Up to three different duplicated (homoeologous) genes were distinguished in hexaploids for the low-copy nuclear gene *Waxy*, with substitution rates ranging from 2.18 to 4.79 % among homoeologs (Fortune et al. 2007). The presence of three different homoeologous copies of *Waxy* could support a hybrid (allopolyploid) origin of this lineage (Fortune et al. 2007). These hexaploid species exhibit regular bivalent pairing (Marchant 1968b). Although allopolyploidy is the currently favored hypothesis for the origin of the hexaploid clade, and is supported by the propensity for interspecific hybridization in this genus, an autopolyploid origin cannot be ruled out. It has been shown that genetic diploidization via disomic inheritance may occur rapidly following autopolyploid formation (Le Comber et al. 2010), leading to a progressive divergence of genes duplicated by polyploidy, and to a gene topology that becomes similar to what would be expected for allopolyploids (Staub et al. 2003).

Dating the origin of a polyploid species from (biparentally inherited) nuclear gene divergence data is difficult in that (selective and neutral) population-level evolutionary processes and differential evolution of genes duplicated by polyploidy may give different estimates of the date of origin at different loci. Moreover, in allopolyploids, the divergence of homoeologs actually reflects the divergence between the parental species, thus leading to an overestimated age of the polyploid (Doyle and Negan 2009). We recently attempted to estimate the divergence times between different *Spartina* lineages from maternally inherited chloroplast sequences. Using reconstruction of the *Spartina* chloroplast genome from 454 Roche GS-FLX (Titanium) pyrosequencing (Bellot S. and Ainouche M., unpublished), we were able to analyze three coding (*matK*, *rbcL*, *ndhF*) and eight non-coding (intergenic spacers or introns) chloroplast sequences. Using a Bayesian phylogenetic analysis calibrated with known divergence times in the grass family (Wolfe et al. 1989; Gaut 2002; Prasad et al. 2005; Chalupska et al. 2008; Christin et al. 2008), we estimated the divergence between the tetraploid and hexaploid *Spartina* maternal genomes at 5 MYA, and the divergence time between the two hexaploid *S. alterniflora* and *S. maritima* chloroplast genomes at approximately 3 MYA.

Several additional recent hybridization events between *Spartina* species have led to the formation of homoploid hybrids or new allopolyploids. These events not only have involved parental species of the same ploidal level within the tetraploid or the hexaploid clades, but also involved crosses between tetraploid and hexaploid species (Fig. 12.1c). In the tetraploid lineage, early morphological analyses suggested that *S.* x *caespitosa* was a hybrid between *S. patens* (Eastern North America) and *S. pectinata*, which has a wide distribution across much of North America (Mobberley 1956). Moreover, a recent survey of *S. pectinata* populations across North America has revealed hexaploid and octoploid cytotypes that are morphologically indistinguishable from the tetraploids (Kim et al. 2010). The origin of these octoploids is unknown, and the current hypothesis that the hexaploids derive from backcrosses between octoploids and tetraploids has to be verified. Genetic and genomic resources are currently being developed for this species, which is considered a good candidate for bioenergy due to its high biomass production (Gedye et al. 2010).

The South American species *S. densiflora* illustrates well the genetic exchanges occurring between the tetraploid and hexaploid lineages. The history of this species was only recently elucidated: this vigorous, high marsh species originated from southeastern coast of South America, but has been introduced to Chile, California and Spain (Bortolus 2006). Genetic analyses and molecular phylogenies (Baumel et al. 2002a; Ayres et al. 2008; Fortune et al. 2008) have revealed that this species is heptaploid ($2n = 70$), with a hybrid origin from a tetraploid maternal species closely related to *S. arundinacea* (with which it shares high chloroplast sequence similarity), which occurs in Southern Hemisphere islands (Amsterdam, St Paul and Tristan de Cunha islands), and a hexaploid paternal species related to *S. alterniflora* that also occurs in Argentina. Interestingly, *S. densiflora* appears to have hybridized with each species from the hexaploid clade in both its native area and in more recently colonized regions where it came in contact with native species: (1) At the mouth of the Rio de la Plata in Argentina and Uruguay, morphologically intermediate plants named *Spartina longispica* were early recognized as hybrids between *S. densiflora* and *S. alterniflora* (Saint-Yves 1932; Mobberley 1956). (2) *S. densiflora* was accidentally introduced to California, probably from Chile, as suggested by geographic, historical, and molecular data (Bortolus 2006; Fortune et al. 2008). In the San Francisco Bay, Ayres et al. (2008) have reported hybrids between *S. densiflora* and the native *S. foliosa* ($2n = 62$). Different chromosome numbers ($2n = 66$, 94–96) suggest that the F_1 hybrids deriving from a heptaploid and a hexaploid parent might also have backcrossed with the parental species. (3) On the Atlantic coast of the Iberian Peninsula, Castillo et al. (2010) have recently detected hybrids between introduced high-marsh *S. densiflora* and the native low-marsh hexaploid *S. maritima* ($2n = 60$). As in California, different chromosome numbers ($2n = 64$–66, $2n =$ ca. 94) were recorded for the hybrids, with either *S. densiflora* or *S. maritima* as maternal genome donors. No genome doubling (allopolyploid speciation) is recorded to date in these perennial (apparently sterile) hybrids, as occurred for *S. anglica.*

Within the hexaploid clade, the outcomes of interspecific hybridization are consistent with the phylogenetic relationships and genetic divergence among species (Baumel et al. 2002a; Ainouche et al. 2004b): production of fertile invasive hybrids between closely related taxa on the one hand, and formation of sterile hybrids (followed by genome doubling) between more divergent species on the other (Fig. 12.1c). These hybridizations resulted from recent introductions (nineteenth and twentieth centuries) of *S. alterniflora* from the eastern coast of the Americas to the Pacific coast of North America and to the Atlantic coast of Western Europe.

In California, plants of *S. alterniflora* spread rapidly (Daehler and Strong 1997; Civille et al. 2005). Hybridization with its sister species *S. foliosa* (native to California) has been shown to occur in both directions (Ayres et al. 1999). The greater male fitness of *S. alterniflora* and recurrent backcrosses has resulted in hybrid swarms that progressively replace original *S. foliosa* plants (Antilla et al. 2000; Ayres et al. 2007). Adding an additional layer of complexity, some of these plants might also be involved in the formation of new hybrid genotypes with the introduced *S. densiflora* (Ayres et al. 2008).

In Europe, *S. alterniflora* was accidentally introduced by ship ballast in southern England and western France, where it hybridized with *S. maritima*. In England, hybridization recorded in 1870 gave rise to *S. x townsendii*, a perennial, sterile hybrid (Groves and Groves 1880). Another sterile hybrid between *S. alterniflora* and *S. maritima* was discovered in 1892 in southwestern France in the Bidassoa Estuary (Foucaud 1897) and named *Spartina x neyrautii* (Jovet 1941). Because of their different morphology, some authors suggested that *S. x neyrautii* and *S. x townsendii* might result from reciprocal crosses; however, molecular data revealed that both hybrids share the same chloroplast genome of *S. alterniflora*, identifying it as the maternal parent of both hybrids (Baumel et al. 2003).

After 1890 in England, fertile plants were recorded that appeared to have resulted from chromosome doubling in *S. x townsendii* (Marchant 1963), thus leading to the formation of a new allododecaploid species named *S. anglica* (Hubbard 1968), with chromosome numbers $2n = 120$, 122, or 124 (Marchant et al. 1968b) suggesting aneuploidy. The vigorous plants (Fig. 12.2) have rapidly colonized Western European salt marshes (Raybould et al. 1991a; Thompson 1991; Genegou and Levasseur 1993). Robust shoots, rhizomes and root systems enable this new species to accumulate large volumes of tidal sediments. For this reason, *S. anglica* was deliberately introduced in several parts of the world (northern Europe, Australia, New Zealand, China, North America) for land reclamation and marsh restoration purposes. *S. anglica* has rapidly expanded in its introduced range and now has a worldwide distribution (Ainouche et al. 2009). The rapid spread of the introduced populations has led to various attempts to control or eradicate the species; in fact, it is now listed among the 100 "world's worst" invaders (IUCN 2000). A recent survey of the ploidal levels in the original population (Southampton area, UK) where *S. anglica* and *S. x townsendii* still coexist, has revealed that the perennial, sterile F_1 hybrids represent more than 90 % of the population (Renny-Byfield et al. 2010). Using genomic in situ

Fig. 12.2 *Spartina anglica* in the Baie des Veys (Cotentin), first colonized site (1906) in France (Corbière 1926)

hybridization (GISH), this study also confirmed the existence of nonaploid plants ($2n$ = ca. 90), most likely resulting from backcrosses between *S. anglica* and its maternal parent *S. alterniflora*.

The recurrent and continuing hybridization and genome duplication in *Spartina* (Fig. 12.1) make it a particularly useful model with which to explore the consequences of these processes at various evolutionary time scales. Recent events allow ecological, phenotypic, and genetic comparisons between the newly formed hybrids or polyploids and their parents that are still extant in natural populations, a situation that is only met in a few biological systems (see also Chaps. 13 and 14, this volume).

12.3 Ecological and Adaptive Consequences of Hybridization and Polyploidy in *Spartina*

Spartina species play an important ecological role in the sedimentary dynamics of salt marshes, where the plants are considered to be "ecosystem engineers"(Crooks 2002). The ecological range of *S. anglica* along the shore is larger than either of its parents. *Spartina anglica* tolerates several hours of immersion at high tides and thus is able to occupy a vacant niche as a pioneer species in the low-tide zone. This

species may accrete large volumes of tidal sediments, making the habitat more terrestrial, and allowing colonization by other salt marsh plant species, which modifies the physical structure of intertidal coastal zones.

Recently formed *Spartina* hybrids and allopolyploids display hybrid vigor and rapid expansion in their native range and invasive abilities when introduced, with important implications for ecosystem management (Lambrinos 2008). *S. alterniflora x foliosa* hybrids have rapidly invaded the San Francisco Bay (Ayres et al. 2007). Several generations of introgressive hybridization make it difficult to differentiate invasive "cryptic hybrids" from "pure" native *S. foliosa* plants, and this complicates eradication plans (The Invasive *Spartina* Project, www.spartina.org). The heptaploid South American *S. densiflora* has also rapidly colonized Californian marshes where it increased fivefold in distribution during the last 25 years following its introduction (Ayres et al. 2004). In Spain, the hybrid *S. densiflora x maritima* exhibits greater ecological amplitude than either parental species: hybrids are able to survive both in lower elevations in the intertidal zone where *S. maritima* naturally grows and also in high marshes where *S. densiflora* invades (Castillo et al. 2010).

Hybridization and polyploidy in *Spartina* are accompanied by various biological changes that have influenced important adaptive traits such as breeding system, physiology, and morphology. The Californian *S. densiflora x foliosa* hybrids, derived from self-incompatible, outcrossing parents, have evolved self-fertility that has contributed substantially to their rapid spread (Sloop et al. 2009). This breakdown in self-incompatibility, also observed following allopolyploidy in *Brassica* and *Arabidopsis*, is most likely triggered by epigenetic mechanisms (Nasrallah et al. 2007). Several transgressive traits in height and biomass, vegetative growth rates, intertidal amplitude, and salinity tolerance are also reported in the *Spartina* hybrids from California (Ayres et al. 2007).

Hybridization between *S. alterniflora* and *S. maritima* had very different morphological consequences in the two independent events that occurred in England (*S. x townsendii*) and France (*S. x neyrautii*), even though these hybridization events involved crosses in the same direction (*S. alterniflora* being the maternal genome donor) and similar parental genotypes (Baumel et al. 2003; Yannic et al. 2004; Salmon et al. 2005). *S. x neyrautii* has shorter spikelets and is distinctly more slender than *S. x townsendii*, which has longer fleshy leaves, resembling more closely the maternal parent *S. alterniflora,* whereas *S. x townsendii* has intermediate morphological features between *S. maritima* and *S. alterniflora* (Mobberley 1956). The phenotypic differences between these two F_1 hybrids of similar genetic origin are puzzling and most likely represent differential effects of the "genomic shock" resulting from the merger of divergent genomes. *S. x townsendii* is almost indistinguishable from its allopolyploid derivative *S. anglica*; moreover, the latter species exhibits larger phenotypic plasticity (Thompson et al. 1991).

Physiological and anatomic adaptations are important components of *Spartina* ecology and distribution (Maricle et al. 2006, 2009). As observed in many polyploids (Otto 2007), stomatal cell size increases with ploidy level in *Spartina* (Marchant 1967; Kim et al. 2010), which may affect photosynthetic rates (Warner

and Edwards 1993). The larger ecological amplitude of the allopolyploid *S. anglica* compared to its parents has to be related to increased tolerance to highly reducing and sulfidic sediment conditions. This increased tolerance may explain the ability of *S. anglica* to colonize low-marsh zones (Maricle et al. 2006). Survival of *S. anglica* in anoxic sediments likely is facilitated by its particular ability to develop aerenchyma systems that supply the submerged plants with atmospheric oxygen and efficiently transport oxygen to the roots (Maricle and Lee 2002). *S. anglica* displays enhanced mechanisms to transport O_2 and exhibits five times greater H_2S removal than its progenitor species *S. alterniflora* (the other parental species, *S. maritima*, was not investigated; Lee 2003).

An interesting function seems to have accompanied the formation of the hexaploid lineage of *Spartina*: the ability to produce dimethylsulfoniopropionate (DMSP), an osmoprotectant and anti-stress molecule (Larher et al. 1977). DMSP is environmentally important as the main biogenic precursor of atmospheric dimethyl sulfide (DMS), which has roles in the biogeochemical sulfur cycle, in cloud formation and in acid precipitation (Kocsis et al. 1998). DMSP is commonly produced by many marine algae, but this capacity is rare in angiosperms, where it has been found only in three genera, one in the Asteraceae (*Wollastonia*) and two in the Poaceae (*Saccharum* and *Spartina*) (Otte et al. 2004). Asteraceae and Poaceae have independently developed different metabolic pathways to achieve this synthesis (Kocsis et al. 1998; Kocsis and Hanson 2000).

High DMSP concentrations are found in leaves of the hexaploids *S. alterniflora*, *S. foliosa* and *S. maritima* (Otte et al. 2004), providing these plants with a characteristic "unpleasant sulphurous odor" noticed by early taxonomists (Mobberley 1956). In contrast, no DMSP is detected in the tetraploid species analyzed to date. Although the evolutionary steps giving rise to this physiological novel capacity are yet to be elucidated, the comparative data generated to date indicate that it may have been enabled by the transition to the hexaploid condition. If so, this would represent an example of a presumably important ecological adaptation arising from polyploidy.

Physiological (and more generally phenotypic) evolution following hybridization and polyploidy is a direct outcome of the genomic consequences of genome merger and duplication. Although we are still far from obtaining an exhaustive knowledge of the complex gene and regulatory networks involved in most phenotypic traits of adaptive importance, significant progress has been made in recent years regarding genetic and genomic processes accompanying polyploid evolution.

12.4 Genome Evolution Following Hybridization and Allopolyploid Speciation in *Spartina*

It is now well established that hybrid and allopolyploid genomes are not simply additive with respect to their parental genomes and that myriad novel interactions at both the structural and functional levels may lead to rapid evolution and

evolutionary novelties. It also may be that these phenomena play a critical role in the evolutionary success of newly formed species in both short-term and long-term evolutionary time (e.g., Wendel 2000; Chen 2007; Doyle et al. 2008; Van de Peer et al. 2009).

The genetic context of the hybrid or allopolyploid species formation is an important parameter to consider for subsequent evolution. Recurrent hybridization events in natural populations provide new lineages with an expanded genetic base, as observed in the introgressant populations resulting from multiple hybridization and recurrent backcrosses involving different *S. alterniflora* and S. *foliosa* genotypes in California (Ayres et al. 2007). In contrast, low genetic diversity is encountered in the parental populations of *S. maritima* and *S. alterniflora* that hybridized in southern England and southern France (Baumel et al. 2003; Yannic et al. 2004; Ainouche et al. Ainouche et al. 2004a). As a result, the F_1 hybrids *S. x townsendii* and *S. x neyrautii* share very similar parental genotypes and exhibit similar, additive genetic composition. A strong genetic bottleneck seems to have affected the new allopolyploid *S. anglica* as a result of a unique origin from *S. x townsendii*: a very low interindividual genetic diversity is encountered within and among populations in both its native and introduced ranges (Guénégou 1988; Raybould et al. 1991b; Baumel et al. 2001; Ainouche et al. 2004a), although some genetic variants may be encountered (Ayres et al. 2001).

Contrasting with most other young or experimentally resynthesized allopolyploid systems that exhibit rapid genome structural evolution (Ozkan et al. 2001; Skalika et al. 2005; Gaeta et al. 2007; Lim et al. 2008; Tate et al. 2009; Szadowski et al. 2010; Buggs et al. 2012; Chester et al. 2012), the new allododecaploid *S. anglica* exhibits relative genome stability (Baumel et al. 2002b; Ainouche et al. 2004a). Most of the early evolutionary changes following allopolyploid speciation in *Spartina* seem to affect the regulation of gene expression, including epigenetic and transcriptome changes (Fig. 12.3). DNA methylation alterations revealed by Methylation Sensitive AFLP (MSAP) appear triggered by hybridization in both *S. x townsendii* and *S. x neyrautii* (Salmon et al. 2005). Genome duplication does not entail significant additional changes, as *S. anglica* has inherited most of the changes observed in *S. x towsendii* but exhibits few specific methylation alterations. Parisod et al. (2009) have shown that an important fraction of these methylation changes affect regions flanking transposable elements, which agrees with the general view of methylation having evolved to control transposable elements in eukaryotic genomes (Slotkin and Martienssen 2007) and with the fact that no burst of transposition was detected following allopolyploid speciation (Baumel et al. 2002b; Parisod et al. 2010). The hexaploid parental species *S. maritima* and *S. alterniflora* have a genome size of 2C = 3.8 pg and 4.3 pg, respectively (Fortune et al. 2008), which suggest that the basic haploid genome ranges between 600 and 700 MB in these species. A preliminary investigation from 454 pyrosequencing of genomic DNA in *S. maritima* revealed that about 27 % of the analyzed sequences were recognized as repetitive, with *Gypsy*-like elements being mostly represented.

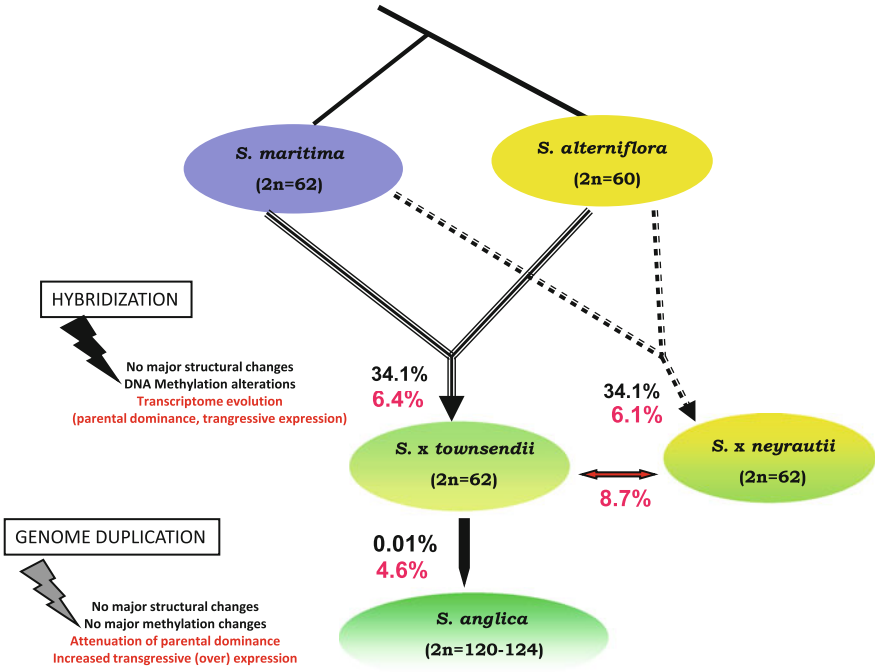

Fig. 12.3 Genome evolution following recent hybridization and allopolyploidy in *Spartina*. Percentages in *red* represent transcriptome changes evaluated using microarrays as described in Chelaifa et al. (2010a, b). Effects of hybridization were estimated by comparing expression profiles in the two independently formed natural F_1 hybrids to a theoretical mid-parent value representing additive parental expression. Effects of genome duplication were evaluated by comparing *S. anglica* to *S.* x *townsendii*. Percentages in *black* represent DNA methylation alterations evaluated using methylation sensitive AFLP, following the procedure employed by Salmon et al. (2005) and Parisod et al. (2009)

Transcriptome evolution was first investigated in *Spartina* using oligo-microarrays. As no Expressed Sequence Tag (EST) database was then available for *Spartina*, heterologous hybridization was performed using the related model species *Oryza sativa* (Chelaifa et al. 2010a, b). The hexaploid species *S. maritima* and *S. alterniflora* that exhibit high exon sequence identity (94.0–99.7 %) at homologous loci displayed 1,247 differentially expressed genes in leaves from plants grown in the same controlled conditions (Chelaifa et al. 2010a). Most of these genes were found to be up-regulated in *S. alterniflora*. Similar levels of non-additive parental patterns of gene expression were observed in both of the hybrids *S. x townsendii* and *S. x neyrautii* (6.1 and 6.4 % of the analyzed genes, respectively, Fig. 12.3), including parental (mostly maternal) gene expression dominance and transgressively expressed genes. However, the dominance of maternal expression appeared more pronounced in *S. x townsendii* than in *S. x neyrautii.* About 8.7 % of the analyzed genes were found differentially expressed between these two F_1 hybrids (Fig. 12.3), and interestingly, most transgressively expressed

genes were different, with genes up-regulated in *S. x townsendii* being related to transport, detoxification and stress, and genes up-regulated in *S. x neyrautii* being related to cellular growth and development. The two independent hybridization events involving the same parental species appear to have generated differential consequences in terms of gene expression. The functions of these differentially expressed genes are consistent with the phenotypic differences previously mentioned between the two hybrids. Genome duplication in *S. anglica* entailed additional transcriptome changes (Fig. 12.3), consisting of the attenuation of the maternal dominance observed in the F_1 hybrid and an increased number of transgressively overexpressed genes (Chelaifa et al. 2010b). Thus, both hybridization and genome duplication appear to have important, though different, effects on the *Spartina* transcriptome, occurring shortly after genome merger and polyploidization. For the first time, these decoupled effects were analyzed during the allopolyploid speciation process, by comparing the actual (naturally formed) F_1 hybrid to its immediately derived allopolyploid that formed and survived in natural conditions. Interestingly, our findings seem to parallel the conclusions emerging from similar comparisons involving natural, more or less recent allopolyploids and/or synthetic F_1 hybrids (Hegarty et al. 2006; Flagel et al. 2008; Flagel and Wendel 2010; Buggs et al. 2011).

The microarray analyses allowed the first large-scale investigation of the *Spartina* transcriptome. However, it should be kept in mind that the expression changes reported in these studies are most likely underestimated: only global expression of the genes that hybridized on the rice microarrays were examined, and homoeologous gene expression could not be distinguished. Similar levels of expression might be attained via biased parental expression, which represents an important component of the functional plasticity of polyploid genomes (Adams et al. 2003; Chaudhary et al. 2009; Flagel et al. 2009; Buggs et al. 2011).

Distinction of homoeologous gene expression has been studied in several allopolyploid models where diploid representatives of the parental species are identified (e.g. cotton: Udall et al. 2006; Flagel et al. 2008; *Arabidopis*: Chang et al. 2010; *Tragopogon*: Buggs et al. 2010; Buggs et al. 2011). In *Spartina,* this task is particularly challenging as the parents of the young allopolyploid *S. anglica* are hexaploids (expected to have retained up to three more-or-less divergent duplicated homoeologs per locus) and because no diploid species is known in the genus (Fig. 12.4a).

Next-Generation Sequencing (NGS) technologies offer unique avenues to distinguish homoeologous copies in highly redundant genomes from natural, non-model species that have experienced successive polyploidization events. This procedure is being developed for *Spartina* as follows (Fig. 12.4b): (1) construction of a *Spartina* reference transcriptome for the hexaploid parental species using 454 Roche GS-FLX pyrosequencing; (2) Single Nucleotide Polymorphism (SNP) detection among reads per annotated contig; (3) haplotype assembly to discriminate homoeoalleles; and (4) Illumina RNA-Sequencing (RNA-Seq) to explore variation of homoeolog expression in the parental species.

Using the procedure outlined above, 38,000 contigs representing $\sim$17,000 unigenes were annotated for *S. maritima* and *S. alterniflora* from leaf and root

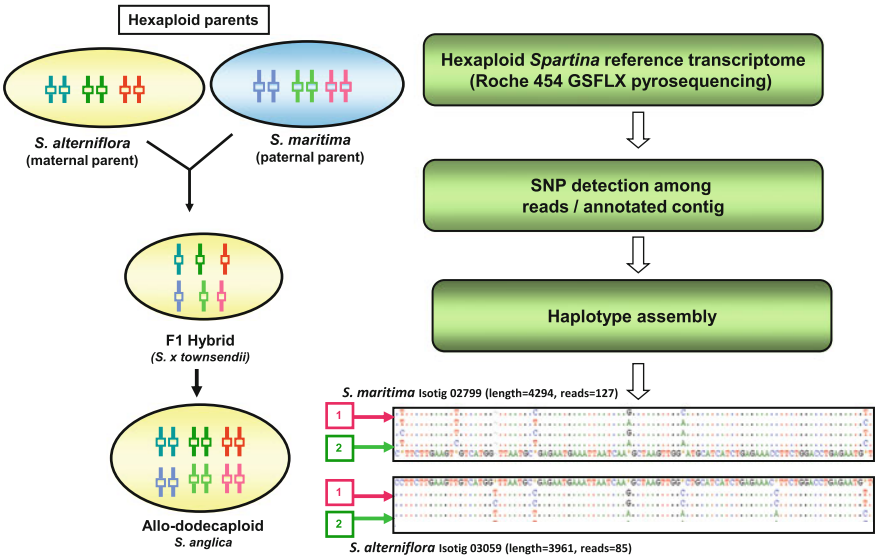

Fig. 12.4 **a** Expected nuclear homoeologous gene copies per locus in the hexaploid parents, F_1 hybrid and allododecaploid *Spartina* species **b** Procedure for sequence heterogeneity detection at homologous nuclear coding loci; example from two homologous contigs in the hexaploid *S. maritima* (contig length = 4,294, number of read = 127) and *S. alterniflora* (contig length = 3,961, number of reads = 85) assembled from 454 Roche pyrosequencing of cDNA libraries. Arrows of similar color (*green* or *pink*) represent divergent alleles (1 and 2) present in the two hexaploids and detected from shared polymorphic sites between the two species, in the central zone of the compared region. Other haplotypes represent slight variants of these two copies

cDNA libraries, which represent the first reference transcriptome for the hexaploid *Spartina* species (Ferreira et al. 2013). An example of SNP detection among reads is presented in Fig. 12.4b for an aligned portion of two homologous contigs annotated in *S. maritima* and *S. alterniflora* as HECT-domain-containing protein (*Oryza sativa* annotation LOC_Os12g24080.1|13112.m02448|cDNA). Reads were assembled using a 95 % identity threshold, to avoid potential comparisons involving paralogs. Six polymorphic sites are detected in this region, including four polymorphic sites shared between *S. maritima* and *S. alterniflora* and two species-specific polymorphic sites. The shared polymorphisms allow distinction of two divergent haplotypes (1 and 2, Fig. 12.4b) present in both hexaploids, and one (in *S. maritima*) or two (in *S. alterniflora*) additional minor variants. Screening of a larger number of polymorphic sites and loci will provide information about the number and divergence of (homoeo) alleles encountered in the hexaploids and shed light on the evolutionary history of this lineage. A variable number of retained copies per homologous locus may be expected. For instance, Fortune et al. (2007) analyzed the low-copy nuclear gene *Waxy* that is present in two paralogs (A and B) in *Spartina*. Only one B copy was encountered in *S. maritima*, whereas three distinct B copies were found in *S. alterniflora*. These two species have

apparently lost the A copy that is still retained in the hexaploid *S. foliosa* (Fortune et al. 2007). The large-scale detection of homoeologs from massive parallel sequencing will provide a genome-wide view of the retention-loss process at various evolutionary time scales (in the hexaploid parents and the nascent allododecaploid). Transcript loss might result from either homoeolog silencing as observed in the various cases of subfunctionalization reported in allopolyploids (reviewed in Osborn et al. 2003; Chen 2007; Doyle et al. 2008), physical loss of the duplicated copies that may occur more or less rapidly following polyploid speciation (e.g. Gaeta et al. 2007; Tate et al. 2009; Koh et al. 2010; Buggs et al. 2012) or from homoeologous recombination (Cifuentes et al. 2010; Salmon et al. 2010; Gaeta et al. 2010). High-throughput sequencing of genomic DNA and targeted sequencing (e.g. Grover et al. 2012) offer new possibilities to differentiate the effects of these alternatives. Finally, distinction among *Spartina* homoeologs will allow more accurate analysis of homoeologous expression in the F_1 hybrids and the allododecaploid using RNA-Seq data for various organs from plants collected in different ecological conditions.

In conclusion, the well-established framework now available for *Spartina* offers important opportunities to elucidate the phenotypic, ecological, and genomic consequences of recurrent hybridization and polyploidy. Based on the dramatic increase in knowledge that has accumulated for various polyploid systems in recent years (Ainouche and Jenczewski 2010), it has become clear that hybridization and polyploidy generate a range of possible responses that vary among genera. Recent allopolyploidy in *S. anglica* was not accompanied by rapid restructuring of the parental genomes as has occurred in other polyploid systems. The major evolutionary events in *S. anglica* appear to affect the regulation of gene expression (including epigenetic regulation); these appear profoundly altered by the merger of different genomes. The rapid advances of NGS technologies will allow more exhaustive exploration of highly redundant genomes (which until now suffered from severe technical limitations) such as those of *Spartina*. These will provide a better understanding of the genetic and epigenetic mechanisms underlying expression plasticity, and their effect on adaptive and ecologically relevant functions.

Acknowledgments This work benefited from the financial support of the Centre National de la Recherche Scientifique (CNRS), University of Rennes 1, the French National Research Agency (ANR), Region Bretagne, Genoscope, and from the Biogenouest (Transcriptomics and Environmental Genomics) platform facilities. P. Wincker, J. Poulain, Corinne Da Silva, O. Lima, S. Coudouel, D. Naquin, A. Deillhy are thanked for their contribution to the 454 pyrosequencing

References

Adams KL, Cronn R, Percifield R, Wendel JF (2003) Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proc Natl Acad Sc USA 100(8):4649–4654

Ainouche ML, Jenczewski E (2010) Focus on polyploidy. New Phytol 186:1–4

Ainouche ML, Baumel A, Salmon A (2004a) *Spartina anglica* Schreb. A natural model system for analysing early evolutionary changes that affect allopolyploid genomes. Biol J Linn Soc 82:475–484

Ainouche ML, Baumel A, Salmon A, Yannic G (2004b) Hybridization, polyploidy and speciation in *Spartina* Schreb (Poaceae). New Phytol 161:165–172

Ainouche ML, Fortune PM, Salmon A, Parisod C, Grandbastien M-A, Fukunaga K, Ricou M, Misset M-T (2009) Hybridization, polyploidy and invasion: lessons from *Spartina* (Poaceae). Biol Invasion. doi:10.1007s10530-0089383-2

Antilla CK, King RA, Ferris C, Ayres DR, Strong DR (2000) Reciprocal hybrid formation of *Spartina* in San Francisco Bay. Mol Ecol 9:765–770

Ayres DR, Strong DR (2001) Origin and genetic diversity of *Spartina anglica* (Poaceae) using nuclear DNA markers. Am J Bot 88:1863–1867

Ayres DR, Garcia-Rossi D, Davis HG, Strong DR (1999) Extent and degree of hybridization between exotic (*Spartina* alterniflora) and native (S. foliosa) cordgrass (Poaceae) in California, USA determined by randomly amplified polymorphic DNA (RAPDs). Mol Ecol 8:1179–1186

Ayres DR, Smith DL, Zaremba K, Klohr S, Strong DR (2004) Spread of exotic cordgrass and hybrids (*Spartina sp*) in the tidal marshes of San-Francisco Bay CA, USA. Biol Invasions 6:221–231

Ayres DA, Zaremba K, Sloop CM, Strong DR (2007) Sexual reproduction of cordgrass hybrids (Spartina foliosa × alterniflora) invading tidal marshes in San Francisco Bay. Divers Distrib 14:187–195

Ayres DR, Grotkopp E, Zaremba C, Sloop CM, Bloom MJ, Bailey JP, Anttila CK, Strong DR (2008) Hybridization between invasive *Spartina densiflora* (Poaceae) and native *S. foliosa* in San Francisco Bay. Am J Bot 95:713–719

Baumel A, Ainouche ML, Levasseur JE (2001) Molecular investigations in populations of *Spartina anglica* C.E. Hubbard (Poaceae) invading coastal Brittany (France). Mol Ecol 10:1689–1701

Baumel A, Ainouche ML, Bayer RJ, Ainouche AK, Misset M-T (2002a) Molecular phylogeny of hybridizing species from the genus *Spartina Schreb. (Poaceae)*. Mol Phylogenet Evol 22:303–314

Baumel A, Ainouche ML, Kalendar R, Schulman AH (2002b) Retrotransposons and genomic stability in populations of the young allopolyploid species *Spartina anglica* C.E. Hubbard (Poaceae). Mol Biol Evol 19:1218–1227

Baumel A, Ainouche ML, Misset MT, Gourret JP, Bayer RJ (2003) Genetic evidence for hybridization between the native *Spartina maritima* and the introduced *Spartina alterniflora* (Poaceae) in South-West France: *Spartina x neyrautii* re-examined. Plant Syst Evol 237:87–97

Bortolus A (2006) The austral cordgrass *Spartina densiflora* Brong.: its taxonomy, biogeography and natural history. J Biogeogr 33:158–168

Buggs RJA, Chamala S, Wu W, Gao L, May GD, Schnable PS, et al. (2010) Characterization of duplicate gene evolution in the recent natural allopolyploid *Tragopogon miscellus* by next generation sequencing and Sequenom iPLEX MassARRAY genotyping. Mol Ecol 19:132–146

Buggs RJA, Zhang L, Miles N, Gao L, Wu W, Schnable P, Barbazuk WB, Soltis PS, Soltis DE (2011) Transcriptomic shock generates evolutionary novelty in a newly formed, natural allopolyploid plant. Curr Biol 21:551–556

Buggs RJA, Chamala S, Wu W, Tate JA, Schnable PS, Soltis DS, Soltis PS, Barbazuk WB (2012) Rapid, repeated, and clustered loss of duplicated genes in allopolyploid *Tragopogon* populations of independent origin. Curr Biol 22:248–252

Castillo JM, Ayres DR, Leira-Doce1 P, Bailey J, Blum M, Strong DR, Luque T and Figueroa E (2010) The production of hybrids with high ecological amplitude between exotic *Spartina densiflora* and native S. maritima in the Iberian Peninsula. Diversity and Distributions, 1–12 doi: 10.1111/j.1472-4642.2010.00673.x

Chalupska D, Lee HY, Faris JD, Evrard A, Chalhoub B, Kaselkorn R, Gorniki P (2008) Acc Homeoloci and the evolution of wheat genomes. Proc Natl Acad Sci U S A 105:9691–9696

Chang PL, Dilkes B, McMahon M, Comai L, Nuzhdin SV (2010) Homoeolog-specific retention and use in allotetraploid *Arabidopsis suecica* depends on parent of origin and network partners. Genome Biol 11:R125. doi:10.1186/gb-2010-11-12-r125

Chaudhary B, Flagel L, Stupar RM, Udall JA, Verma N, Springer NM, Wendel JF (2009) Reciprocal silencing, transcriptional bias and functional divergence of homeologs in polyploid cotton (Gossypium). Genetics 182: 503–517

Chelaifa H, Mahe F, Ainouche M (2010a) Transcriptome divergence between the hexaploid salt-marsh sister species *Spartina maritima* and *Spartina alterniflora (Poaceae)*. Mol Ecol 19:2050–2063

Chelaifa H, Monnier A, Ainouche M (2010b) Transcriptomic changes following recent natural hybridization and allopolyploidy in the salt marsh species *Spartina x townsendii* and *Spartina anglica (Poaceae)*. New Phytol 186:161–174

Chen ZJ (2007) Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. Annu Rev Plant Biol 58:377–406

Chester M, Gallagher JP, Symonds VV, Veruska Cruz da Silva A, da Silva A, Mavrodiev EV, Leitch AR, Soltis PS, Soltis DE (2012) Extensive chromosomal variation generated in a recently formed natural allopolyploid species, Tragopogon miscellus (Asteraceae). Proc Natl Acad Sci U S A 109:1176–1181

Christin P-A, Besnard G, Samaritani E, Duvall MR, Hodkinson TR, Savolainend V, Salamini N (2008) Oligocene CO2 Decline Promoted C4 Photosynthesis in Grasses. Curr Biol 18:37–43

Cifuentes M, Grandont L, Moore G, Chèvre AM, Jenczewski E (2010) Genetic regulation of meiosis in polyploid species: new insights into an old question. New Phytol 186:37–45

Civille JC, Sayce K, Smith SD, Strong DR (2005) Reconstructing a century of *Spartina alterniflora* invasion with historical records and contemporary remote sensing. Ecoscience 12:330–338

Columbus JT, Cerros-Tlatilpa R, Kinney MS, Siqueiros-Delgado M-E, Bell HL, Griffith MP, Refulio-Rodrigez NF (2007) Phylogenetics of Chloridoideae (Gramineae): a preliminary study based on nuclear internal transcribed spacer and chloroplast trnL-F sequences. Aliso 23:565–579

Corbière L (1926) La Spartine de Townsend en Normandie. Bulletin de la société Linéenne de Normandie, 7^{e} série 9:92–117

Crooks JA (2002) Characterizing ecosystem-level consequences of biological invasions: the role of ecosystem engineers. Oikos 97:153–166

Daehler CC, Strong DR (1997) Hybridization between introduced smooth cordgrass (*Spartina alterniflora*; Poaceae) and native California cordgrass (*S. foliosa*) in San Francisco Bay, California. U S A Am J Bot 81:307–313

Doyle JJ, Negan EN (2009) Dating the origins of polyploidy events. New Phytol 186:73–85

Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF (2008) Evolutionary genetics of genome merger and doubling in plants. Annu Rev Genet 42:443–461

Ferreira de Carvalho J, Poulain J, Da Silva C, Wincker P, Michon-Coudouel S, Dheilly A, Naquin D, Boutte J, Salmon A, Ainouche M. (2013) Transcriptome de novo assembly from Next-Generation Sequencing and comparative analyses in the hexaploid salt marsh species *Spartina maritima* and *Spartina alterniflora* (Poaceae) Heredity, (in press)

Flagel LE, Wendel JF (2010) Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. New Phytol 186:184–193

Flagel L, Udall J, Nettleton D, Wendel JF (2008) Duplicate gene expression in allopolyploid Gossypium reveals two temporally distinct phases of expression evolution. BMC Biol 6:16

Flagel LE, Chen L, Chaudhary B, Wendel JF (2009) Coordinated and fine-scale control of homoeologous gene expression in allotetraploid cotton. J Hered 100:487–490

Fortuné PM, Schierenbeck K, Ainouche A, Jacquemin J, Wendel JF, Ainouche ML (2007) Evolutionary dynamics of *waxy* and the origin of hexaploid *Spartina* species. Mol Phylogenet Evol 43:1040–1055

Fortuné PM, Schierenbeck K, Ayres D, Bortolus A, Clatrice O, Ainouche ML (2008) The enigmatic invasive *Spartina densiflora*: a history of hybridizations in a polyploidy context. Mol Ecol 17:4304–4316

Foucaud (1897) Un *Spartina* inédit. Ann Soc Sci Nat Char Inf 32:220–222
Gaeta RT, Pires J (2010) Homoeologous recombination in allopolyploids: the polyploidy rachet. New Phytol 186:18–27
Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC (2007) Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. Plant Cell 19:3403–3417
Gaut BS (2002) Evolutionary dynamics of grass genomes. New Phytol 154:15–28
Gedye K, Gonzalez-Hernandez J, Ban Y, Thimmapuram J, Sun F, Wright C, Ali S, Boe A, Owens V (2010) Investigation of the transcriptome of prairie cordgrass, a new cellulosic biomass crop. The Plant Genome 3, 2:69–80
Grass Phylogeny Working Group (GPWG) (2001) Phylogeny and subfamilial classification of the grasses (Poaceae). Ann Missouri Bot Gard 88:373–457
Gray AJ, Benham PEM, Raybould AF (1990) *Spartina anglica*-the evolutionary and ecological background. In: Gray AJ, Benham PEM (eds) *Spartina anglica*-a research review. Institute of terrestrial ecology, Natural environment research council, pp 5–10
Grover CE, Salmon A, and Wendel JF (2012) Targeted sequence capture as a powerful tool for evolutionary analysis. Am J Bot (in press)
Groves H, Groves J (1880) *Spartina townsendii nobis*. Rep Bot Soc Exch Club Br Id 1:37
Guénégou MC, Levasseur JE (1993) La nouvelle espèce amphidiplo *Spartina anglica* C.E. Hubbard: son origine, argumentation et implications. Biogeographica 69:125–133
Guénégou MC, Citharel J, Levasseur JE (1988) The hybrid status of *Spartina anglica* (Poaceae). Enzymatic analysis of the species and the presumed parents. Can J Bot 66:1830–1833
Hegarty MJ, Barker GL, Wilson ID, Abbott RJ, Edwards KJ, Hiscock SJ (2006) Transcriptome shock after Interspecific hybridization in *Senecio* is ameliorated by genome duplication. Curr Biol 16:1652–1659
Hilu KW, Alice LA (2001) A phylogeny of Chloridoideae (Poaceae) based on *matK* sequences. Syst Bot 26:386–405
Hubbard JCE (1968) Grasses, 2nd edn. Penguin Books, London
Huskin CL (1930) The origin of *S. x townsendii*. Genetica 12:531–538
IUCN (2000) World's worst invasive alien species. In: IUCN The World Conservation Union. http://iucn.org
Jovet P (1941) Notes systématiques et écologiques sur les Spartines du Sud-Ouest. Bull Soc Bot Fr 88:115–123
Kim S, Rayburn AL, Lee DK (2010) Genome size and chromosome analyses in prairie cordgrass. Crop Sci 50:2277–2282
Kocsis MG, Hanson AD (2000) Biochemical evidence for two novel enzymes in the biosynthesis of 3-dimethylsulphoniopropionate in *Spartina alterniflora*. Plant Physiol 123:1153–1161
Kocsis MG, Nolte KD, Rhodes D, Shen TL, Gage DA, Hanson AD (1998) Dimethylsulfoniopropionate biosynthesis in *Spartina alterniflora*. Plant Physiol 117:273–281
Koh J, Soltis P, Soltis DE (2010) Homoeolog loss and expression changes in natural populations of the recently formed allotetraploid *Tragopogon mirus* (Asteraceae). BMC Genomics 11:97
Lambrinos G (2008) Managing invasive ecosystem engineers: the case of *Spartina* in Pacific estuaries. Theor Ecol Ser 4:299–322
Larher F, Hamelin J, Steward GR (1977) L'acide diméthylsulphonium-3-propano de *Spartina anglica*. Phytochemistry 16:2019–2020
Le Comber SC, Ainouche ML, Kovarik A, Leitch AR (2010) Making a functional diploid: from polysomic to disomic inheritance. New Phytol 186:113–122
Lee RW (2003) Physiological adaptations of the invasive cordgrass *Spartina anglica* to reducing sediments: rhizome metabolic gas fluxes and enhanced O2 and H2S transport. Mar Biol 143:9–15
Lim KY, Soltis DE, Soltis PS, Tate J, Matyasek R, Srubarova H, Kovarik A, Pires JC, Xiong Z, Leitch AR (2008) Rapid chromosome evolution in recently formed polyploids in *Tragopogon* (Asteraceae). PLoS ONE 3(10):e3353. doi:10.1371/journal.pone.0003353

Marchant CJ (1963) Corrected chromosome numbers for *Spartina x townsendii* and its parent species. Nature 199:929
Marchant CJ (1967) Evolution in *Spartina* (Gramineae): I. the history and morphology of the genus in Britain. Bot J Linn Soc 60(381):1–24
Marchant CJ (1968a) Evolution in *Spartina* (Graminae). III species chromosome numbers and their taxonomic signifiance. Bot J Linn Soc 60(383):411–417
Marchant CJ (1968b) Evolution in *Spartina* (Graminae) II. chromosome basic relationships and the problem of *S. x towsendii agg*. Bot J Linn Soc 60(383):381–409
Marchant CJ (1977) Hybrid characteristics in *Spartina x neyrautii* Fouc., a taxon rediscovered in northern Spain. Bot J Lin Soc. 74:289–296
Maricle BR, Lee RW (2002) Aerenchyma development and oxygen transport in the estuarine cordgrasses *Spartina alterniflora* and *S. anglica*. Aquat Bot 74:109–120
Maricle BR, Crosier JJ, Bussiere BC, Lee RW (2006) Respiratory enzyme activities correlate with anoxia tolerance in saltmarsh grasses. J Exp Mar Biol Ecol 337:30–37
Maricle BR, Koteyeva NK, Voznesenskaya EV, Thomasson JR, Edwards GE (2009) Diversity in leaf anatomy, and stomatal distribution and conductance, between salt marsh and freshwater species in the C4 genus *Spartina* (Poaceae). New Phytol 184:216–233
Mobberley DG (1956) Taxonomy and distribution of the genus *Spartina*. Iowa State Coll J Sci 30:471–574
Nasrallah JB, Liu P, Sherman-Broyles S, Schmidt R, Nasrallah ME (2007) Epigenetic mechanisms for breakdown of self-incompatibility in interspecific hybrids. Genetics 175:1965–1973
Osborn TC, Pires JC, Birchler JA, Auger DL, Chen ZJ, Lee HS, Comai L, Madlung A, Doerge RW, Colot V, Martienssen RA (2003) Understanding mechanisms of novel gene expression in polyploids. Trends Genet 19:141–147
Otte ML, Wilson G, Morris JT, Moran BM (2004) Dimethylsulphoniopropionate (DMSP) and related compounds in higher plants. J Exp Bot 55:919–925
Otto SP (2007) The evolutionary consequences of polyploidy. Cell 131:452–462
Ozkan H, Levy AA, Feldman M (2001) Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. Plant Cell 13:1735–1747
Parisod C, Salmon A, Zerjal T, Tenaillon M, Grandbastien M-A, Ainouche M (2009) Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in *Spartina*. New Phytol 184:1003–1015
Parisod C, Alix K, Just J, Petit M, Sarilar V, Mhiri C, Ainouche M, Chalhoub B, Grandbastien MA (2010) Impact of transposable elements on the organization and function of allopolyploid genomes. New Phytol 186:37–45
Prasad V, Stroemberg CAE, Alimohammadian H, Sahni A (2005) Dinosaur coprolites and the early evolution of grasses and grazers. Science 310:1177–1180
Prasad V, Stroemberg CAE, Leaché AD, Samant B, Patnaik R, Tang L, Mohabey DM, Ge S, Sahni A (2011) Late Cretaceous origin of the rice tribe provides evidence for early diversification in Poaceae. Nat Communication 2:480. doi:10.1038/ncomms1482
Raybould AF, Gray AJ, Lawrence MJ, Marshall DF (1991a) The evolution of *Spartina anglica* C.E. Hubbard (Gramineae): genetic variation and status of the parental species in Britain. Biol J Linn Soc 44:369–380
Raybould AF, Gray AJ, Lawrence MJ, Marshall DF (1991b) The evolution of *S. anglica* C.E. Hubbard (Gramineae): origin and genetic variability. Biol J Linnean Soc 43:111–126
Renny-Byfield S, Ainouche M, Leitch IJ, Lim KY, Le Comber SC, Leitch AR (2010) Flow cytometry and GISH reveal mixed ploidy populations and *Spartina* nonaploids with genomes of *S. alterniflora* and *S. maritima* origin. Ann Bot 105:527–533
Saint-Yves A (1932) Monographia Spartinarum. Cand 5:19–100
Salmon A, Ainouche ML, Wendel JF (2005) Genetic and epigenetic consequences of recent hybridization and polyploidy in *Spartina* (Poaceae). Mol Ecol 14:1163–1175
Salmon A, Flagel L, Ying B, Udall JA, Wendel JF (2010) Homoeologous nonreciprocal recombination in polyploid cotton. New Phytol 186:123–134

Skalická K, Lim KY, Matyasek R, Matzke M, Leitch AR, Kovarik A (2005) Preferential elimination of repeated DNA sequences from the paternal, *Nicotiana tomentosiformis* genome donor of a synthetic, allotetraploid tobacco. New Phytol 166:291–303

Sloop C, Ayres DR, Strong DR (2009) The rapid evolution of self-fertility in *Spartina* hybrids (*Spartina alterniflora* × *foliosa*) invading San Francisco Bay, CA. Biol Invasions 11:1131–1144

Slotkin K, Martienssen R (2007) Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet 8:272–285

Soltis DE et al (2013) Polyploidy and genome evolution. In: Soltis PS, Soltis DE (eds) The early stages of polyploidy: rapid and repeated evolution in Tragopogon. Springer, Heidelberg

Stebbins GL (1950) Variation and evolution in plants. Columbia University Press, New York

Straub SCK, Pfeil BE, Doyle JJ (2003) Testing the polyploid past of soybean using a low-copy nuclear gene-Is Glycine (Fabaceae: Papilionoideae) an auto- or allopolyploid? Mol Phylogenet Evol 39:580–584

Szadkowski E, Eber F, Huteau V, Lode M, Huneau C, Belcram H, Coriton O, Manzanares-Dauleux MJ, Delourme R, King GJ et al (2010) The first meiosis of resynthesized *Brassica napus*, a genome blender. New Phytol 186:102–112

Tate JA, Joshi P, Soltis KA, Soltis PS, Soltis DE (2009) On the road to diploidization? Homoeolog loss in independently formed populations of the allopolyploid *Tragopogon miscellus* (Asteraceae). BMC Plant Biol 9:80. doi:10.1186/1471-2229-9-80

Thompson JD, McNeilly T, Gray AJ (1991) Population variation in *Spartina anglica* C. E. Hubbard. I. Evidence from a common garden experiment. New Phytol 117:115–128

Triplet P, Gallicé A (2008) Les plantes envahissantes du littoral atlantique: le cas de la Spartine anglaise (*Spartina anglica*). Aestuaria 13 Aestuarium– Le Forum des Marais Atlantiques eds

Udall JA, Swanson JM, Nettleton D, Percifield RJ, Wendel JF (2006) A novel approach for characterizing expression levels of genes duplicated by polyploidy. Genetics 173:1823–1827

Van de Peer Y, Maere S, Meyer A (2009) OPINION the evolutionary significance of ancient genome duplications. Nat Rev Genet 10:725–732

Warner DA, Edwards GE (1993) Effects of polyploidy on photosynthesis. Photosynthetis research 35:135–147

Wendel JF (2000) Genome evolution in polyploids. Plant Mol Biol 42:225–249

Wolfe KH, Gouy M, Yang Y-W, Sharpt PM, Li W-H (1989) Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. Proc Natl Acad Sci U S A 86:6201–6205

Yannic G, Baumel A, Ainouche ML (2004) Uniformity of the nuclear and chloroplast genomes of *Spartina maritima* (Poaceae) a salt marshes species in decline along the Western European Coast. Heredity 93:182–188

Chapter 13
Allopolyploid Speciation in Action: The Origins and Evolution of *Senecio cambrensis*

Matthew J. Hegarty, Richard J. Abbott and Simon J. Hiscock

Abstract *Senecio cambrensis* is one of a few allopolyploid plant species known to have originated in the recent past and, therefore, provides excellent material for analysing allopolyploid speciation. This allohexaploid species originated in the UK within the last 100 years following hybridization between diploid *S. squalidus* and tetraploid *S. vulgaris*. In this chapter, we first describe the events leading up to hybridization between these two species, focusing mainly on the origin and spread of *S. squalidus* in the UK. We then consider alternative pathways by which *S. cambrensis* might have originated and conclude that current evidence suggests an origin via formation of the triploid hybrid (*S. x baxteri*) followed by chromosome doubling. We next review our investigations into levels of genetic diversity and also changes to gene expression and the possible causes of this (epigenetic effects) during the origin of *S. cambrensis*. High levels of genetic diversity, assessed by surveys of allozyme and AFLP variation, have been recorded in *S. cambrensis*, and it is likely that intergenomic recombination was an important generator of this diversity. Our studies of 'resynthesized' *S. cambrensis* have shown that the initial genome merger (hybridization) producing *S. x baxteri* generates genome-wide, non-additive alterations to parental patterns of gene expression and DNA

M. J. Hegarty (✉)
Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Penglais Campus, Aberystwyth, Ceredigion SY23 3DA, UK
e-mail: ayh@aber.ac.uk

R. J. Abbott
School of Biology, University of St. Andrews, Harold Mitchell Building, St. Andrews, Fife KY16 9TH, UK
e-mail: rja@st-andrews.ac.uk

S. J. Hiscock
School of Biological Sciences, University of Bristol, Woodland Road, Bristol BS8 1UG, UK
e-mail: simon.hiscock@bristol.ac.uk

P. S. Soltis and D. E. Soltis (eds.), *Polyploidy and Genome Evolution*,
DOI: 10.1007/978-3-642-31442-1_13,

methylation, with genome duplication resulting in a secondary burst of both transcriptional and epigenetic modification. In synthetic allohexaploid lines of *S. cambrensis* phenotypic changes become apparent from the second to fifth generations, possibly as a consequence of recombination or epigenetic effects; these include changes in ray flower form and emergence of self-incompatible individuals. We conclude by considering the future of *S. cambrensis* from the standpoint of it being a model species for further study of allopolyploid speciation, and second its long-term success in the wild. Ongoing work to produce a draft reference genome for *S. squalidus* will underpin future research in *S. cambrensis*, enabling a more thorough survey of changes to DNA methylation, small RNA activity and promoter binding in the hybrids, as well as comparison with the related allotetraploid *S. eboracensis* to determine the effects of genome dosage. The future of the species in the wild is currently uncertain. The population in Edinburgh that represented a separate origin of the species in the wild during the 1970s is now extinct, and there has been a marked decline in the number of populations and individuals of the species in its heartland, North Wales, since the 1980s. An analysis of how its ecology compares with those of its parents is lacking. However, it appears to share the same habitats in the wild with its parents, which might have contributed to its decline. Although *S. cambrensis* may become extinct in the wild in the near future, the potential will remain for it to originate again in the UK providing that conditions prevail for its parents to hybridize.

13.1 Introduction

13.1.1 General Introduction

Polyploidization (genome duplication) is an important evolutionary process in plants (Grant 1981) that appears to have accompanied major transitions in land plant evolution, including the evolution of the seed habit and the evolution of angiosperms (Jiao et al. 2011). All angiosperms are thought to have a polyploid ancestry (Jiao et al. 2011) but most of these are paleopolyploids that now function essentially as diploids. Nevertheless, there are also numerous examples of recently formed polyploids (Adams and Wendel 2005; Wood et al. 2009). Most recent polyploids have formed in association with interspecific hybridization (Grant 1981; Leitch and Bennett 1997; Soltis and Soltis 1999; Otto and Whitton 2000; Leitch and Leitch 2008; Hegarty and Hiscock 2008)—allopolyploidy, which is now recognised as perhaps the most important mechanism of abrupt speciation in plants (Grant 1981; Leitch and Leitch 2008; Hegarty and Hiscock 2008). Allopolyploidy confers rapid fertility to hybrids because duplicated parental chromosomes pair 'normally' during meiosis; it also confers reproductive isolation of the hybrids from their parental species because of aberrant or failed chromosome pairing (Soltis and Soltis 1999; Rieseberg et al. 2003;

Rieseberg and Willis 2007). Numerous examples of allopolyploid speciation have been described in the literature, including at least six new plant species that have arisen in the last century: *Tragopogon mirus* and *T. miscellus* (Novak et al. 1991, Chap. 14, this volume); *Spartina anglica* (Gray et al. 1991, Chap. 12, this volume), *Senecio cambrensis* and *S. eboracensis* (Ashton and Abbott 1992, Abbott and Lowe 1996, 2004) and *Cardamine schulzii* (Urbanska et al. 1997). In this chapter, we review the origins of the *Senecio* allopolyploid species *S. cambrensis* (Welsh groundsel). We show how resynthesised forms of this allopolyploid can reveal important insights into the genetic and genomic consequences of allopolyploidization and how parental traits, both morphological and physiological, recombine in neopolyploids.

13.1.2 *The Genus* Senecio *and the Origins of UK Allopolyploids*

The genus *Senecio* (Asteraceae), which includes ragworts and groundsels, is one of the largest and most morphologically diverse genera of flowering plants. With between 1000 and 3000 species, the genus has a worldwide distribution with species described from almost every land mass on Earth (Vincent 1996). Recent revisions, following molecular phylogenetic analysis, of the genus have assigned many former *Senecio* species to other genera leaving a conservative 1200 species of *Senecio sensu stricto* (Pelser et al. 2007). Within the genus there are numerous examples of hybridization and polyploidy, and many species have been proposed to have an allopolyploid origin (Abbott and Lowe 1996; Coleman et al. 2001; Kadereit et al. 2006; Pelser et al. 2012). In the UK, three new polyploid taxa have arisen within the last 100 years as a consequence of hybridization between native tetraploid *Senecio vulgaris* (common groundsel) and the introduced invasive diploid species *S. squalidus*. The origins of these new polyploid taxa (see below) provide one of the best examples of 'evolution in action' (Abbott and Lowe 1996, Hegarty and Hiscock 2008). The introduction and rapid spread of alien *S. squalidus* across the UK was the main catalyst for this burst of hybrid speciation, which was further facilitated by the obligate outcrossing mating system of *S. squalidus* increasing the frequency of interspecies pollinations (Brennan and Hiscock 2010).

13.1.3 *Oxford Ragwort in the British Isles: Introduction and Spread*

Senecio squalidus ($2n = 2x = 20$), commonly known as Oxford ragwort, is itself a recently evolved (homoploid) hybrid species (James and Abbott 2005; Abbott et al. 2010; Brennan et al. 2012) that originated on Mount Etna, Sicily, as a result of hybridization between *S. aethnensis* ($2n = 2x = 20$, a Mount Etna endemic of higher altitudes) and *S. chrysanthemifolius* ($2n = 2x = 20$, a native Sicilian

species of lower altitudes). At mid-altitudes on the volcano the distribution of these species overlaps, leading to the formation of a stable hybrid zone (Brennan et al. 2009). Material from this hybrid zone was introduced to the Oxford Botanic Garden in the early 1700s, but records of the collection locality or localities, methods of cultivation, and numbers of plants cultivated in the garden have been lost. Plants subsequently escaped from the Botanic Garden and colonised the masonry of the old college walls from the end of the eighteenth century (Harris 2002). During the industrial revolution of the nineteenth and early twentieth centuries *S. squalidus* moved rapidly out of Oxford by colonising the clinker beds of the expanding UK railway, of which Oxford was a key hub. The chronology of this famous plant invasion is meticulously documented in a set of papers by Kent (reviewed in Abbott et al. 2009), which record that it began to spread northwards in the late nineteenth century, reaching different parts of northern England during the middle of the twentieth century, before becoming established in the Central Belt of Scotland by the mid-1950s. Today, *S. squalidus* continues to spread north in Scotland and across Northern Ireland.

The rapidity of this invasion is intriguing because *S. squalidus* is self-incompatible (Abbott and Forbes 1993; Hiscock 2000a, b, and obligate outcrossers are generally thought not to make good colonizers or invasives (Baker 1967). Most invasive species tend to have uniparental reproduction, either sexual (selfing) or asexual (apomixis or vegetative reproduction), although there are exceptions, most notably other species of Asteraceae, such as yellow star thistle, *Centaurea solstitialis* (Sun and Ritland 1998). According to Baker's 'rule' (Baker 1967), successful colonising and invasive plants are usually self-fertile (self-compatible [SC]) (Stebbins 1957; Baker 1967). Studies of the mating system of *S. squalidus,* however, have shown that individuals exhibit strong self-incompatibility (SI) across its entire British range, and, as in other species of Asteraceae, this SI is regulated sporophytically by a single polymorphic *S* locus (Hiscock 2000a, b; Brennan et al. 2002, 2005, 2006). The finding of strong SI in *S. squalidus* is intriguing, particularly in the light of the extreme population bottleneck that its ancestors must have experienced during its introduction and early colonisation. Following a population bottleneck, allelic diversity at the *S* locus will be lowered and opportunities for mating (between individuals carrying different *S* alleles) correspondingly reduced (Hiscock 2000b; Brennan et al. 2002). An extensive survey of SI in *S. squalidus* across the UK showed that a combination of substantial between-population sharing of the seven *S* alleles contained in the entire UK population and low levels of selfing ('pseudo-self-compatibility') were the most likely cause of *S. squalidus'* reproductive success as a colonizer (Brennan et al. 2005, 2006).

13.1.4 The Origins of Senecio cambrensis

During its spread across the UK, *S. squalidus* ($2n = 2x = 20$) hybridized with the self-compatible native groundsel, *S. vulgaris* ($2n = 4x = 40$), resulting in the recent

Fig. 13.1 The neoallohexaploid *Senecio cambrensis* (*centre*) flanked by its parents, tetraploid *S. vulgaris* (*left*) and diploid *S. squalidus* (*right*). Here, *S. cambrensis* has flower heads (capitula) of intermediate type to its parents; however, its ray florets can vary in length, and occasionally non-radiate forms (lacking ray florets) are also found in the wild

origin of three hybrid taxa. These are the allohexaploid *S. cambrensis* ($2n = 6x = 60$), the recombinant tetraploid *S. eboracensis* ($2n = 4x = 40$), and the stabilized introgressant radiate form of *S. vulgaris, S. vulgaris* var. *hibernicus* ($2n = 4x = 40$). Interestingly, all three new hybrid taxa are self-compatible, suggesting that this SC mating system, inherited from *S. vulgaris*, is 'dominant' over the SI mating system present in *S. squalidus*. Detailed descriptions of these new taxa and what is known about their origins are presented elsewhere (see Abbott et al. 1992; Lowe and Abbott 2003; Abbott and Lowe, 2004; Kim et al. 2008).

Here, we briefly summarise the information available on the origin of *S. cambrensis* (Fig. 13.1) focussing particularly on the possible pathways of its origin. Knowing the pathway of origin of a polyploid taxon is helpful in regard to understanding the species' potential to generate genetic diversity during its initial stages of development, and also for accurate production of synthetic forms of the polyploid used to study possible genetic and epigenetic changes that occurred in the taxon immediately following its origin in the wild (Hegarty et al. 2006, 2008, 2011; Lukens et al. 2006; Buggs et al. 2009).

Senecio cambrensis was described by Rosser (1955) from material provided by H.E. Green, who first observed the plant growing at Ffrith and Ceffn-y-bedd, North Wales, UK, in 1948. The plant was described as an annual or short-lived perennial herb that was hexaploid with flower heads (capitula) containing ray florets having short ligules ($\sim$4.8 mm in length). Rosser (in Crisp 1972) later determined a herbarium specimen, collected in Denbigh, North Wales in 1925 and originally named as *S. squalidus* x *S. vulgaris*, to be *S. cambrensis*. However, in the absence of a chromosome count there remains some doubt as to whether this specimen is *S. cambrensis* or, alternatively, a fertile hybrid of *S. squalidus* and *S. vulgaris* (see below). Fertile, hexaploid plants with similar morphology to the wild form of

S. cambrensis can be produced by treating synthetic triploid hybrids between *S. squalidus* and *S. vulgaris* with colchicine (Harland 1955, Weir and Ingram 1980, Hegarty et al. 2005). On this basis, Rosser (1955) concluded that *S. cambrensis* was a new species that originated by hybridization between native *S. vulgaris* and introduced *S. squalidus* followed by chromosome doubling. The species is likely to have originated shortly before it was first recorded in North Wales and after *S. squalidus* had spread to the region in the early part of the twentieth century (Kent 1963).

In 1982, *S. cambrensis* was found growing in Edinburgh, UK (Abbott et al. 1983), and subsequent molecular analysis involving surveys of allozyme and chloroplast DNA variation showed that it had originated independently in Edinburgh rather than being dispersed there from North Wales (Ashton and Abbott 1992; Harris and Ingram 1992). Herbarium records indicate that the Edinburgh lineage may have existed since at least 1974; however, it is now thought to be extinct as the species has not been recorded in the Edinburgh area or nearby since 1993 (Abbott and Forbes 2002).

Because *S. cambrensis* is readily synthesised by treating the triploid hybrid between *S. squalidus* and *S. vulgaris* with colchicine, it has been assumed that chromosome doubling of the triploid hybrid was the likely pathway of origin of the allopolyploid species in the wild (Rosser 1955). In theory, however, the species could have originated along several possible pathways (Table 13.1) with the first step involving formation of a triploid, tetraploid, pentaploid or hexaploid hybrid. Of these alternatives, the formation of a triploid hybrid is more likely in that it results from fusion of normal haploid gametes produced by each parent. In contrast, formation of higher ploidy hybrids relies on the production of unreduced gametes, which will be generated at a much lower frequency in each parent species.

There are many records of the triploid hybrid (*Senecio x baxteri* Druce) occurring in the wild (Crisp 1972; Benoit et al. 1975; Marshall and Abbott 1980). It is easily recognised because of its intermediate morphology and its almost complete seed sterility. Progeny tests of *S. vulgaris* plants have shown that this hybrid is generated regularly but at very low frequencies in the wild where *S. vulgaris* and *S. squalidus* co-occur (Marshall and Abbott 1980). In contrast, the tetraploid hybrid has never been reported unequivocally in the wild, although Crisp (1972) described a plant likely to have been such a hybrid based on morphology and an analysis of its offspring. Although no chromosome count was made of the plant, all of its offspring were approximately tetraploid and segregated for a range of morphological, reproductive and disease resistance traits. Crisp (1972) suggested that such offspring could become stabilized in the wild to form distinct taxa, and it is feasible that the tetraploid *S. eboracensis* originated in this way (Lowe and Abbott 2000). Whether such a hybrid might have contributed to the origin of *S. cambrensis* as detailed in Table 13.1 remains unknown, but seems less likely than an origin involving the triploid hybrid, given the apparent rarity of the tetraploid hybrid in the wild. Similarly, origins involving the formation of either a pentaploid or hexaploid hybrid are less parsimonious than one involving the triploid hybrid, although cannot be ruled out entirely. Further support for the hypothesis that formation of a triploid hybrid was the first step in the origin of

Table 13.1 Some possible pathways of origin for *Senecio cambrensis* in the wild

(1) Via Triploid hybrid (2*n*=30)
• Step 1 - Formation of triploid hybrid (2*n*=30) through fusion of haploid gametes produced by each parent species.
• Step 2 - Chromosome doubling of triploid hybrid by: (i) fusion of 'unreduced' triploid gametes (*n*=30); or (ii) doubling of chromosome number of a somatic cell ancestral to a floret or flower head producing 'reduced' triploid gametes (*n*=30).
(2) Via Tetraploid hybrid (2*n*=40)
• Step 1 - Formation of tetraploid hybrid (2*n*=40) by: (i) fusion of a haploid gamete (n=20) of *S. vulgaris* and a diploid (unreduced) gamete (*n*=20) of *S. squalidus*; or (ii) fusion of an 'unreduced' triploid gamete of triploid hybrid (n=30) and haploid gamete of *S. squalidus* (*n*=10).
• Step 2 - Production of hexaploid hybrid by: (i) fusion of an 'unreduced' tetraploid gamete (*n*=40) generated by tetraploid hybrid and a 'reduced' diploid gamete (*n*=20) of same hybrid or of *S. vulgaris* or an unreduced gamete of *S. squalidus*; or (ii) fusion of 'balanced' triploid gametes (*n*=30) produced by same hybrid.
(3) Via Pentaploid hybrid (2*n*=50)
• Step 1 - Formation of pentaploid hybrid by: (i) fusion of a haploid gamete of *S. squalidus* (*n*=10) and an unreduced gamete of *S. vulgaris* (*n*=40); or (ii) fusion of unreduced gamete of triploid hybrid (*n*=30) with haploid gamete of *S. vulgaris* (*n*=20) or unreduced gamete of *S. squalidus* (*n*=20).
• Step 2 - Production of hexaploid hybrid by: (i) fusion of gametes with same or different 'balanced' chromosome numbers (i.e. *n*=10, *n*=20, *n*=30, *n*=40, *n*=50) generated by pentaploid hybrid such that the zygote produced is hexaploid (2*n*=60); (ii) fusion of tetraploid 'balanced' gamete (2*n*=40) produced by pentaploid hybrid with reduced gamete (*n*=20) of *S. vulgaris* or unreduced gamete of *S. squalidus*. (iii) fusion of diploid 'balanced' gamete (*n*=20) produced by pentaploid hybrid with unreduced gamete (*n*=40) of *S. vulgaris*.
4) Direct formation of Hexaploid hybrid (2*n*=60)
• Step 1 - Formation of hexaploid hybrid by fusion of unreduced gametes from both *S. squalidus* (*n*=20) and *S. vulgaris* (*n*=40).

S. cambrensis comes from reports by Vosa (in Crisp 1972) and Ingram (1978) that rare allohexaploid offspring were produced spontaneously by natural selfing of the synthetic triploid hybrids they made. However, Weir and Ingram (1980) also reported the production of an allohexaploid plant directly from a cross between *S. vulgaris* and *S. squalidus*. This could have been formed by fusion of unreduced gametes or alternatively by chromosome doubling of a triploid hybrid early in its development. We shall never know exactly how the different lineages of *S. cambrensis* originated in the UK, but given that the triploid hybrid is regularly encountered in the wild and is capable of producing allohexaploid offspring spontaneously, an origin involving doubling of the chromosome number of a triploid hybrid seems the most likely route of origin.

Because *S. cambrensis* is self-fertile, one newly formed individual of the species would have been able to reproduce sexually and successfully following the species' origin, i.e. without need of a mate. Moreover, if the species had originated through chromosome doubling of a triploid hybrid, it would be expected initially to be homozygous at all loci within its parental genomes, but to exhibit frequent fixed heterozygosity at duplicated loci among parental genomes. Somewhat

surprisingly, however, the species has been shown through surveys of allozyme variation (Ashton and Abbott 1992) and particularly AFLP variation (Abbott et al. 2007) to contain high levels of genetic diversity, indicating that it rapidly generated this diversity following its origin. Abbott et al. (2007) considered the ways in which such genetic diversity was produced and concluded that intergenomic recombination would most likely have been an important mechanism, although other mechanisms such as aneuploidy, gene conversion, activation of transposons and retroelements, other forms of mutation and gene flow from parental species could not be ruled out. The occurrence of radiate and non-radiate forms of *S. cambrensis* as well as variation in the ligule length of ray florets have been attributed to intergenomic recombination (Ingram and Noltie 1984), while the observation of multivalent formation occurring at low frequency in meiotic cells of the species (Ingram and Noltie, 1989) provides a mechanism for such recombination to occur. Although not reported by Ingram and Noltie (1989), Crisp (1972) observed up to eight chromosomes with subterminal centromeres in the somatic complement of *S. cambrensis* plants (based on root tip squashes). He pointed out that as neither parent species possessed such chromosomes they were probably the products of chromosome rearrangements following meiotic abnormalities. In addition to generating genetic diversity, intergenomic recombination resulting in chromosome rearrangements could lead to the formation of reproductive barriers between divergent lineages of *S. cambrensis*. Clearly, further work on the frequency of intergenomic recombination and its possible effects in *S. cambrensis* would be worthwhile.

13.2 Consequences of Hybridization and Polyploidy in Natural and Resynthesised *Senecio cambrensis*

13.2.1 Transcriptome Shock

In common with many other allopolyploid species, the merger of two divergent genomes during the formation of *Senecio cambrensis* has had a dramatic impact at the level of gene expression. As part of our investigation of the allopolyploid origins of *S. cambrensis*, we conducted gene expression analysis using a custom cDNA microarray platform (Hegarty et al. 2005) to survey the transcript levels of floral genes in both the intermediate triploid hybrid *S. x baxteri* and wild *S. cambrensis*, relative to their progenitor species. The experimental design of this comparison is shown in Fig. 13.2. This experiment revealed an initial large change in floral gene expression in *S. x baxteri*, with approximately 475 cDNA clones showing up- or down-regulation relative to its parental taxa or, also importantly, relative to natural *S. cambrensis*, from which it differs primarily by a change in ploidal level (Hegarty et al. 2005). Thus, the greatest changes in gene expression relative to the parents appeared to be associated with the hybridization step to form

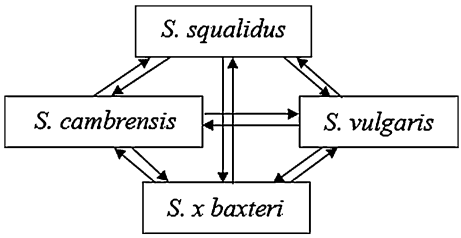

Fig. 13.2 Experimental design employed in microarray comparisons of gene expression between the allopolyploid *S. cambrensis* and its progenitor taxa, *S. vulgaris* and *S. squalidus*, and their sterile triploid F1 hybrid, *S. x baxteri*. Experimental details can be found in Hegarty et al. (2005), but, briefly, mature flower bud tissue was harvested from a mixed population of approximately 30 plants and pooled prior to RNA extraction to create an 'average' for each taxon. Labelled cDNA for each taxon was hybridized to a custom floral cDNA microarray. Two taxa were differentially labelled and compared per array hybridization (with 10 replicate hybridizations performed per comparison) using dye swaps to account for any bias in labelling efficiency. Each taxon was compared with the other three, for a total of 30 array hybridizations per taxon. Raw expression data for each taxon were extracted from these 30 replicates and imported separately into the GENESPRING microarray analysis software (Silicon Genetics) to enable comparison between all four taxa. Figure reproduced from Hegarty et al. (2008)

S. x baxteri. This initial burst of altered gene expression we termed "transcriptome shock", after the phenomenon of "genome shock" described by McClintock in her seminal work on transposable elements in plant hybrids (McClintock 1984). The 'transcriptome shock' effect in *S. x baxteri* was confirmed in our later analysis of resynthesised *S. cambrensis*, which further showed that the polyploidization event (here induced by colchicine) had an immediate calming (ameliorating) effect on altered patterns of gene expression detected in *S. x baxteri* (Hegarty et al. 2006). Importantly, this altered pattern of gene expression, apparent in first-generation allopolyploids, was preserved in four successive generations of the synthetic allopolyploids and in wild *S. cambrensis* (Hegarty et al. 2006). Previous research in resynthesised wheat (Feldman and Levy 2005) identified separate effects of hybridization and polyploidization on the genome and transcriptome, but our findings in *S. cambrensis* represented one of the first indications that these changes in gene expression were genome-wide. Interestingly, the putative functional classes of genes affected by hybridization and allopolyploidization were remarkably similar, with no functional class of genes being overly affected by hybridization or allopolyploidization (Fig. 13.3). However, perhaps not surprisingly, when compared with functional classes of genes not affected by either process, there was a greater representation of genes potentially involved in flower/inflorescence and pollen developments, which may reflect the transitions in floral phenotypes observed after hybridization and allopolyploidization.

We later reassessed the data (Hegarty et al. 2008) in light of a new approach used by Wang et al. (2006a) in their studies of allotetraploid *Arabidopsis suecica*. In this study, they focused on the identification of genes whose expression in hybrids differed from the additive expression midpoint of the two different parental

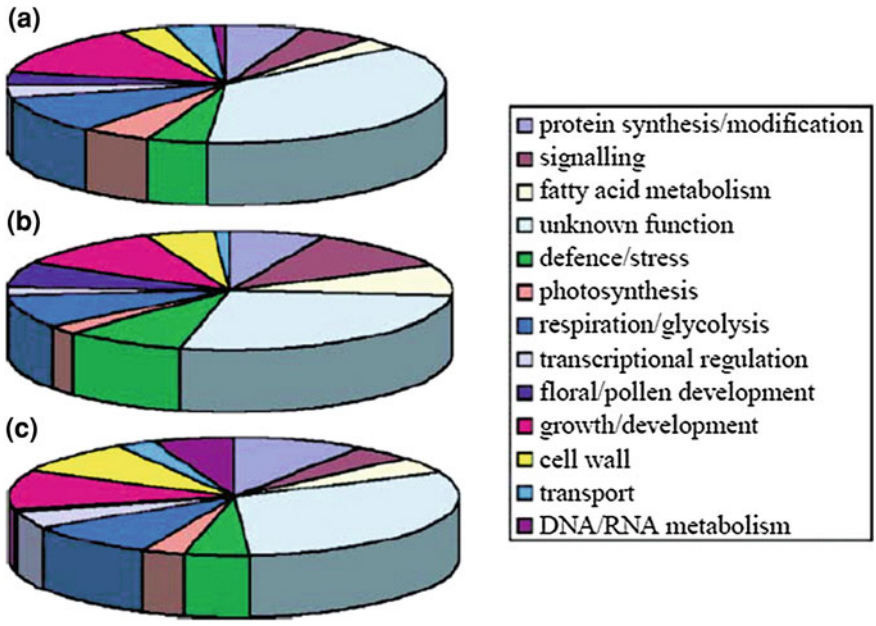

Fig. 13.3 Functional classes of genes affected by allopolyploidization and hybridization. Basic gene ontologies for cDNA clones displaying **a** conserved expression changes in both wild and synthetic *Senecio cambrensis* relative to *S. x baxteri* (genes affected by allopolyploidy, 540 clones), **b** expression changes relative to the parental taxa *S. squalidus* and *S. vulgaris* in both hybrid taxa (genes affected by hybridization, 99 clones) and **c** genes showing no expression difference between the parental and hybrid taxa (unaffected by hybridization or polyploidy, 289 clones). With the exception of a higher proportion of floral/pollen-related genes in (**a, b**) compared with (**c**), there are no substantial differences between the classes of affected genes (adapted with permission from Hegarty et al. 2006)

gene copies. A similar approach had also been used to analyse gene expression change in maize hybrids (Stupar et al. 2007). Such an approach provides a consistent and unified methodology for identifying genes affected by hybridization and/or polyploidization in different model study systems. We therefore reanalysed the microarray data from our original study to identify specific genes and classes of genes affected by hybridization and polyploidization. Using methods similar to Stupar et al. (2007), we tested whether changes in gene expression observed in synthetic *S. x baxteri* and wild allohexaploid *S. cambrensis* were additive or non-additive (Hegarty et al. 2008). By averaging the parental expression values for each feature on the array showing differential expression in the hybrids, a parental midpoint expression value (MPV) was obtained. The derived midpoint values were then used to calculate a ratio of hybrid and parental expression values compared to the MPV for each array feature. A ratio of −0.33 indicates additive gene expression, whilst ratios below −1 or above 1 represent expression in the hybrid outside the range of either parent. Statistical analysis of differentially expressed genes from our previous microarray experiment (Hegarty et al. 2005) showed that,

for both hybrids, the median ratio was significantly different from −0.33, allowing us to reject the null hypothesis of largely additive gene expression changes. Instead, for both hybrids, the majority of the data were skewed towards one of the parents; in the case of *S. x baxteri*, expression was skewed towards that of the lower expressing parent, whereas in *S. cambrensis* it was skewed towards that of the higher expressing parent. Further analysis of the data showed that for both hybrids *S. vulgaris* was the lower expressing parent in 70 % of cases. Expression outside the parental range was observed in a substantial proportion of cases in both hybrids: 7.42 % in *S. x baxteri* and 3.03 % *S. cambrensis* (Hegarty et al. 2008).

Having identified a pool of cDNA clones displaying non-additive changes to gene expression in both hybrid taxa, we then tested these clones for evidence of expression beyond the parental ranges, i.e. transgressive gene expression. In *S. x baxteri*, 80.4 % of non-additively expressed clones differed from the MPV by >1.5-fold, with 42.2 % of clones in *S. cambrensis* showing the same effect. Within both of these groups, the majority of cases involved upregulation compared with the MPV (66.9 and 70.4 % in *S. x baxteri* and *S. cambrensis*, respectively). Aside from the genes for which no functional class could be ascribed (49.2 %), the major functional groups affected in *S. x baxteri* were genes involved in development (6.6 %), nucleotide binding (6.1 %), mitochondrial activity (4.76 %) and cell wall function (3.97 %). Within the development category, a high proportion of clones (32 %) were found to encode tubulins, profilins or senescence-associated proteins. Of the clones involved in nucleotide binding, 34 % were transcription factors. In *S. cambrensis*, the majority of clones (58 %) could not be assigned to a functional category. Of the remainder, the largest categories were defense (11.1 %) and cell wall-related genes (6.17 %).

Our reanalysis of the microarray data therefore revealed a relatively high proportion of non-additive gene expression change in the hybrids relative to their parental expression levels (Hegarty et al. 2008). In addition, the degree of non-additive gene expression was lower in allohexaploid *S. cambrensis* compared with its triploid intermediate *S. x baxteri*. This finding was consistent with our previous observation that the "transcriptome shock" resulting from allopolyploidization is largely due to hybridization, with polyploidization resulting in a distinct secondary shift (amelioration) in gene expression (Hegarty et al. 2006). The fairly diverse nature of the genes affected was consistent with other findings in *Arabidopsis* (Wang et al. 2006b), cotton (Adams et al. 2004) and maize (Stupar et al. 2007) that non-additive changes to gene expression are genome-wide.

Interestingly, similar functional classes of genes were affected by hybridization in *Senecio*, *Arabidopsis* and maize (Fig. 13.4), suggesting that certain gene networks may be particularly susceptible to perturbation by hybridization; the functional categories of nucleotide binding, defense and mitochondria being good examples. In terms of the classes of genes affected in *Senecio*, it is also noteworthy that one of the major affected groups in *S. x baxteri* was nucleotide binding. In addition to a number of (primarily down-regulated) transcription factors that have shown similar alterations in expression pattern in the polyploid *Arabidopsis suecica* (Wang et al. 2006b), this category also contained clones encoding cytidine

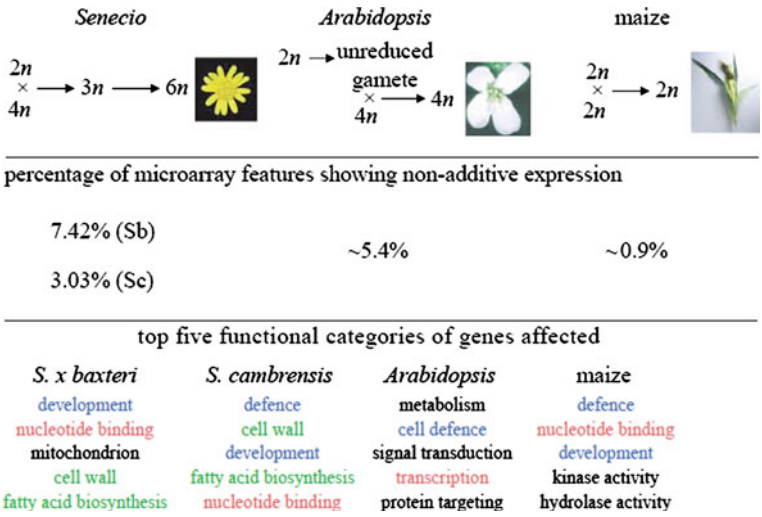

Fig. 13.4 Comparison of non-additive expression changes resulting from hybridization in maize and hybridization/polyploidy in *Senecio* and *Arabidopsis*. The formation of the hybrids is shown in each case, together with the level of non-additive gene expression in the hybrids expressed as a percentage of the features on the microarray platform used. Finally, the top five functional gene classes affected (ignoring unknowns) for each hybrid are displayed for comparison. *Red* indicates a functional gene class affected in all four hybrid systems, *blue* indicates a functional class affected in at least one of the *Senecio* hybrids and one of the other two hybrid taxa, and *green* indicates a functional class affected in both *Senecio* hybrids but not in either *Arabidopsis* or maize. For *Arabidopsis suecica*, gene function data were taken from Wang et al. (2006b), and for maize from an extrapolation of the supplementary data given in Stupar et al. (2007). Sb, *S. x baxteri*; Sc, *S. cambrensis*. Reproduced from Hegarty et al. (2008)

deaminase (CDA) and 8-oxoG-DNA glycosylase (OGG1). OGG1 has been implicated in DNA base excision repair (García-Ortiz et al. 2001), while CDA has been suspected to be involved in RNA editing, although it now appears that pentatricopeptide repeat proteins that contain CDA-like domains are the more likely candidates (Salone et al. 2007). These genes were of interest, given that we also observed a relatively high number of clones encoding proteins involved in either DNA modification or cell division. In addition to cytidine deaminase and OGG1 (both of which were upregulated compared with the parental midpoint in *S. x baxteri*), there was also up-regulation of adenosylhomocysteinase and adenosyl kinase, the genes involved in S-adenosylmethionine (SAM) dependent methylation (Moffatt et al. 2002; Mull et al. 2006). The expression of SAM synthetase was also increased relative to both parents. SAM-dependent methylation is used in gene silencing and also in pectin methylation (Pereira et al. 2006), and, indeed, we observed an increase in the expression of pectin methylesterase as well as another SAM-dependent enzyme, caffeic acid 3-O-methyltransferase (Hegarty et al. 2008).

It is clear from these analyses that hybridization and polyploidy have separate, distinct effects on gene expression in *Senecio*. These changes in gene expression

affect a wide spectrum of transcripts from a number of functional groups and are non-additive in nature. A good proportion of these changes are also transgressive—outside the range of either parent species—representing a possible source of heterotic success in interspecific hybrids. To investigate possible mechanisms for these changes in gene expression, we focused our attention on epigenetic modification in *Senecio* hybrids.

13.2.2 Epigenetic Modification

While we were unable to demonstrate complete silencing of any clones in our microarray analyses, it is important to note that the cDNA-based microarrays used could not distinguish between different parental homeologues. Therefore, in the cases where a hybrid showed down-regulation of a gene relative to its parental taxa, it may have been due to either homeologue loss or silencing as a consequence of DNA methylation. Both phenomena were observed using cleaved amplified polymorphic sequence (CAPS) analysis in natural and resynthesised allotetraploid *Tragopogon miscellus* (Buggs et al. 2009) and suggested that homeologue loss was not an immediate consequence of hybridization or polyploidy but occurs slightly later due to recombinational events. Silencing was also not observed to occur immediately in *Tragopogon* synthetics, but observations in natural populations suggest that silencing of one homoeologue is more prevalent than sequence loss. Studies in other allopolyploid systems such as wheat, *Arabidopsis* and *Spartina* have demonstrated that methylation status in F1 hybrids and first-generation allopolyploids displays similar changes to those observed for gene expression. To determine whether methylation in *Senecio x baxteri* and *S. cambrensis* was affected in a similar manner, we undertook a methylation-sensitive AFLP (MSAP) analysis of the parental taxa, three triploid lines and their S_0–S_1 allohexaploid derivatives (Hegarty et al. 2011). The MSAP technique (Xiong et al. 1999) involves digestion of genomic DNA with a standard, rare-cutter enzyme (i.e., *Eco*RI) and one of a pair of isoschizomeric enzymes that share a restriction site but are either sensitive or insensitive to cytosine methylation. In this case, we employed *Hpa*II, which is sensitive to methylation of either cytosine in its recognition site (CCGG) and *Msp*I, which is sensitive only to methylation of the external cytosine. Comparing the two AFLP profiles produced from each isoschizomer enables identification of the methylation status of a given locus, as described in Table 13.2. Five selective primer combinations were used to screen 33 plants as detailed in Table 13.3.

We successfully amplified MSAP products from all 33 individuals and obtained a total of 408 reliable MSAP loci, which were then surveyed to determine levels of non-additivity in the hybrids. Of the 408 loci, 264 (64.7 %) showed polymorphisms between the parental taxa *S. squalidus* and *S. vulgaris*. In the remaining 144 loci that were monomorphic in the parents, 75.7 % were monomorphic across all hybrid samples tested. Surveying the polymorphic loci, it was found that the

Table 13.2 HpaII% MspI banding patterns and locus methylation state (reproduced from Hegarty et al, 2011)

HpaII	MspI	Methylation status
+	-	CCGG (unmethylated)
-	+	C^{m}**C**GG (methylation of internal cytosine)
-	-	^{m}**C**CGG or ^{m}**C**^{m}**C**GG (methylation of external cytosine)
+	-	^{hm}**C**CGG (hemimethylation of external cytosine)

Note: In the case of methylation of the external cytosine, it is not possible from MSAP alone to determine whether the internal cytosine is also methylated

Table 13.3 Oligonucleotides for methylation-sensitive AFLP and number of loci (reproduced from Hegarty et al, 2011)

Oligo	Sequence (5'-3'	Primer Combination	# Loci
EcoRI adaptor 1	CTCGTAGACTGCGTACC	Eco+AGC/Hpa+CTG	118
EcoRI adaptor 2	AATTGGTACGCAGTC	Eco+AAC/Hpa+CTT	88
HpaII adaptor 1	GATCATGAGTCCTGCT	Eco+AGC/Hpa+CTT	80
HpaII adaptor 2	CGAGCAGGACTCATGA	Eco+AAC/Hpa+AAG	Failed
Eco+A	GACTGCGTACCAATTCA	Eco+ACG/Hpa+AAG	69
Hpa	ATCATGAGTCCTGCTCGG	Eco+AAC/Hpa+CTT	53
Eco+AAC (NED)	GACTGCGTACCAATTCAAC		
Eco+ACG (FAM)	GACTGCGTACCAATTCACG		
Hpa+CTG	ATCATGAGTCCTGCTCGGCTG		
Hpa+CTT	ATCATGAGTCCTGCTCGGCTT		
Hpa+AAG	ATCATGAGTCCTGCTCGGAAG		

triploid hybrids each displayed similar overall methylation patterns (Table 13.4), with a strong bias in favour of the *S. vulgaris* (maternal) methylation state for each locus (an average of 57.4 % of loci between the three triploid lines). In addition, the triploids also displayed approximately equivalent levels of non-additive methylation (13.4 % on average). To determine whether methylation patterns were maintained following genome duplication, the methylation status of the S_0 allohexaploids was compared to the triploid lines from which they were derived. In the vast majority of cases, the synthetic allohexaploids retained the methylation state observed in the triploid. On average, 78.2 % of these cases involved additive methylation, whereas 9.7 % of loci involved retention of a non-additive methylation profile. An average of 10.1 % of loci displayed a shift relative to the triploid in the specific parental methylation state favoured (4.5 % shift to *S. squalidus*, 3.8 % shift to *S. vulgaris*, 1.8 % shift to both). Finally, an average of 2 % of loci showed novel non-additive methylation not observed in the triploids.

The analysis was then extended to the S_1 allohexaploids, comparing them to both their triploid ancestors and the preceding S_0 generation. As with the S_0 allohexaploids, overall methylation state was highly similar among the three lines. Again, the most common result was retention of methylation status compared to both the triploid and the S_0 allohexaploid, with an average of 70.3 % of loci showing retention of additive methylation patterns and 7.3 % of non-additive

Table 13.4 Summarised methylation status of hybrid lines (reproduced from Hegarty et al. 2011)

Methylation	Percentage of Loci (triploids)		
State	Line 1	Line 2	Line 3
Additive (SS)	24.58%	24.58%	22.26%
Additive (SV)	56.15%	57.14%	58.80%
Additive (monomorphic)	4.65%	6.98%	4.65%
Nonadditive	14.62%	11.30%	14.29%
	Percentage of Loci (S_0 allohexaploids)		
Same as triploid (additive)	76.17%	78.86%	79.53%
Same as triploid (nonadditive)	9.73%	9.73%	9.73%
Differs from triploid (additive SS)	5.03%	4.36%	4.03%
Differs from triploid (additive SV)	5.37%	3.69%	2.35%
Differs from triploid (monomorphic)	2.01%	0.67%	2.68%
Novel nonadditive methylation	1.68%	2.68%	1.68%
	Percentage of Loci (S_1 allohexaploids)		
Same as triploid+S_0 (additive)	72.06%	72.38%	66.35%
Same as triploid+S_0 (nonadditive)	7.62%	7.94%	6.35%
Same as S_0 not triploid (additive)	7.62%	5.71%	3.49%
Same as S_0 not triploid (nonadditive)	0.32%	1.27%	0.63%
Same as triploid not S_0 (additive)	3.81%	3.17%	1.90%
Same as triploid not S_0 (nonadditive)	0.95%	0.00%	2.54%
Differs from triploid+S_0 (additive SS)	2.54%	4.76%	7.62%
Differs from triploid+S_0 (additive SV)	3.49%	2.22%	7.62%
Novel nonadditive methylation	1.59%	2.54%	3.49%

Note: SS = *S. squalidus*, SV = *S. vulgaris*. "Monomorphic" refers to loci with a common methlyation state in the parents, rather than across hybrid lines

methylation. A further 5.6 % of loci showed retention of additive methylation in the two allohexaploid generations, whereas the triploid had displayed non-additivity. Methylation status was not always consistent between the two allohexaploid generations; however, in 4.1 % of cases, loci displayed a shift in methylation state to that seen in the triploid but not in the S_0 allohexaploid. Furthermore, the S_1 generation showed a return to additivity in 9.4 % of cases, where both the triploid and the S_0 allohexaploids had been non-additive. As with the S_0 generation, though, there was a small (2.5 % average) degree of novel non-additivity.

The results were in accordance with our previous studies of gene expression (Hegarty et al. 2006, 2008): we found that, while cytosine methylation in both hybrid taxa was largely additive between the two parental patterns, a significant degree of non-additivity also exists. Overall methylation status was well conserved between different hybrid lines; while individual loci displayed differences, the global percentages of different methylation states were highly similar between lines (Hegarty et al. 2011). In all three triploid lines, approximately 13.4 % of loci showed non-additive methylation, although the precise type of methylation was not identical between lines in all cases. Levels of non-additive methylation observed in other allopolyploid systems are variable: 8.3 % in *Arabidopsis*

(Madlung et al. 2002), 9 % in *Brassica* (Gaeta et al. 2007), 13–20 % in wheat (Dong et al. 2005; Pumphrey et al. 2009) and as high as 30 % in *Spartina* (Salmon et al. 2005). It has been speculated that the higher genome copy number in wheat and *Spartina* might explain their greater levels of methylation, although Doyle et al. (2008) point out that both species are monocots, which tend to possess a higher GC content (and thus greater potential for methylation) than eudicots. The fact that conserved methylation changes between the *Senecio* hybrids are more on a par with the levels seen in *Arabidopsis* and *Brassica* suggests that this latter hypothesis may be correct, as *S. cambrensis* exhibits the same ploidy as wheat. We should note, however, that the wheat genome is significantly larger than that of *Senecio* (2C genome sizes of 33.96 Gbp wheat; 5.05 Gbp *S. cambrensis*) and is known to contain a significant amount of repetitive DNA including large numbers of retroelements. It is therefore probable that alterations to methylation are more necessary to prevent widespread activation of these genetic regions in wheat and similar polyploid monocots. Indeed, studies of methylation in *Spartina* (Parisod et al. 2010) showed that such changes frequently occur in the vicinity of transposable elements and, perhaps as a result, no transposition burst was detected in the *Spartina* hybrids analysed. Methylation change therefore appears to play a frequent role in genome mergers, but there are exceptions: despite significant differential gene expression in the allotetraploid *Gossypium hirsutum*, almost no differences in methylation could be observed between the hybrid and its parental taxa (Liu et al. 2001) nor do the parental genomes undergo any significant rearrangement. In this situation, it appears that subfunctionalization of the two genomes is the primary cause of phenotypic variation (Adams et al. 2004; Liu and Adams 2007), although a recent study by Chaudhary et al. (2009) demonstrated that neofunctionalization and divergence in parental *cis*-regulatory sequences also play a significant role. Exactly what factors determine the response of the parental genomes to hybridization are largely unknown, although the degree of parental divergence is speculated to play a large role (Chapman and Burke 2007; Buggs et al. 2008; Paun et al. 2009).

In further accordance with our expression analyses, we observed that non-additive methylation in *S. x baxteri* triploids was maintained, on average, in only 73.6 and 55.6 % of cases in the S_0 and S_1 allohexaploids, respectively (Hegarty et al. 2011). In approximately 73 % of cases observed in our microarray studies, the resynthesised allohexaploid lines (S_0–S_4) displayed either an immediate or gradual shift towards an expression pattern similar to that of wild *S. cambrensis* (Hegarty et al. 2006). It seems likely, therefore, that our previous observation that non-additivity results from hybridization but can be partially reduced by genome duplication, also holds true when applied to DNA methylation. This again matches observations from MSAP analysis of *Spartina* allopolyploids (Salmon et al. 2005), which showed that the allopolyploid *S. anglica* retained 71.4 % of the non-additive methylation patterns observed in the F_1 hybrid. These findings were again confirmed when assessing methylation changes associated with transposable elements (Parisod et al. 2010), with the additional observation that many of the changes seen in the F_1 hybrid involved loss of parental markers (usually in the maternal

genome), indicating that such changes involved structural rearrangements to the parental genomes. A similar process may be at play in *S. x baxteri*, because MSAP markers also detect structural changes: indeed, wild populations of *S. cambrensis* show evidence of intergenomic recombination (Abbott et al. 2007, and above), again favouring the *S. vulgaris* genome as with our triploid lines here. By contrast, most of the differences observed in allopolyploid *Spartina* involved alterations to methylation status, rather than structural changes. Similar findings have also been identified in *Brassica*, where most of the methylation changes identified in the S_0 allotetraploid were maintained in S_5 lines, but with a number of revertants and novel changes present as well (Gaeta et al. 2007).

A key finding from our analysis was that the global patterns of DNA methylation change observed in our experiments strongly mirror the global changes in gene expression observed in our earlier microarray analyses, indicating a possible underlying causation. Whilst further investigation of specific loci showing methylation differences is required to make a definitive case, the similarities between changes in gene expression and DNA methylation are nevertheless striking. For example, we noted that a number of loci displayed novel non-additivity in both the S_0 and S_1 allohexaploids (2.01 % on average in the S_0 lines, 2.54 % in the S_1), again a point of consistency between the methylation study and our earlier microarray expression analysis (Hegarty et al. 2008). The overall level of non-additive methylation may therefore not actually decrease as a consequence of genome duplication, but instead the level of conserved methylation may be lessened. A proportion of loci also displayed unstable methylation patterns across generations in the hexaploids, with an average of 16.08 % of loci showing differences between the S_0 and S_1 lines (including the aforementioned novel non-additive methylation). Of these, the majority consisted of cases where the S_1 allohexaploids revert to additivity or favoured a different parental methylation state to the S_0 line. This reversion to an additive profile was also observed between the triploids and the S_0 plants and agreed with observations from the microarray data that wild *S. cambrensis* often showed an opposing expression pattern to *S. x baxteri*.

However, approximately one-quarter of loci also displayed a shift relative to the S_0 allohexaploids to favour the methylation state seen in the triploid. These findings suggest that the methylation state of some loci may vary as a consequence of segregation. This may explain the novel changes observed by Gaeta et al. (2007) in their S_5 allopolyploids of *Brassica napus*. Similarly, an analysis of natural populations of the allopolyploid *Tragopogon miscellus* (Buggs et al. 2009), where hybridization occurred at least 40 generations ago, identified a random loss of one parental homoeologue at a rate of 3.2 % across 10 loci in 57 natural hybrids from five populations. In addition, a further 6.8 % of loci showed evidence of gene silencing in one parental copy. The loci lost/silenced were not consistent across populations or individuals, although within populations, there was some conservation in the loci affected. Whilst Buggs et al. (2009) did not note any homoeologue loss/silencing in resynthesised S_0 hybrids, the variability in silencing after such an extended period of time since hybrid formation suggests that

independently formed hybrids can still display significant levels of epigenetic variation. This study, as well as that of Gaeta et al. (2007), was based on a survey of a limited number of loci using cDNA-AFLP or CAPS assay. A later study by Buggs et al. (2011) used the Sequenom MassARRAY allelotyping technology to survey a much larger set of loci in *Tragopogon miscellus* and confirmed that natural hybrids displayed altered patterns of tissue-specific gene expression, whilst resynthesised hybrids demonstrated relaxed control of tissue-specificity. This latter finding suggests a possible mechanism for the 'transcriptome shock' effect we observed in *Senecio x baxteri* and *S. cambrensis*, resulting from a loss of tissue-specific expression patterns seen in the parent taxa. Further work will be needed to confirm if this is the case in *Senecio*. The Sequenom assay used by Buggs et al. (2011) shows the benefits of new molecular tools for studies of allopolyploid systems: with the advent of new technologies for global analysis of DNA methylation such as MSAP and next-generation sequencing (Salmon and Ainouche 2010), it would also be interesting to analyse our later-generation allohexaploid derivatives at a global scale to investigate the longer term changes in methylation as hybrid genomes undergo recombination and adaptation. Such studies may therefore provide further insights into which epigenetic changes are mandated by hybridity, and which may vary between populations and serve as a source of novelty upon which selection may act.

13.2.3 Phenotypic Change

Senecio x baxteri F_1 hybrids generated by crossing *S. vulgaris*, as female, with *S. squalidus* were all self-sterile in contrast to previous studies which reported some self-fertility (Crisp 1972; Ingram 1978). Treating shoots of *S. x baxteri* plants with colchicine produced 'chimeric' plants with allohexaploid branches that produced flower heads that were fully self-fertile. Seed from these flower heads was then used to found the synthetic *S. cambrensis* lines (S_0–S_5) used in the transcriptomic and epigenetic analyses described above. The first wholly allohexaploid plants generated from this seed (S_0 lines), and their progeny (S_1 lines), showed similar ray flower structure and self-compatibility (Hiscock and Hegarty, unpublished). However, in subsequent lines, from the S_2 onwards, variation in ray flower form was detected between individuals within and between the nine independent lineages of synthetic *S. cambrensis*. Some individuals were observed with no ray flowers, some had short partially tubular ray flowers, while in others ray flowers were observed of different lengths and number (Fig. 13.5.) Observations on the progeny of these different individuals showed that these various forms of ray flower are heritable (Hiscock, unpublished). Comparable variation in *S. cambrensis* ray flower form was previously attributed to intergenomic recombination (Ingram and Noltie 1984), but here we suggest another possibility for such abrupt changes to ray flower phenotype—epigenetic effects associated with the epigenetic instability observed in early-generation synthetic *S. cambrensis*

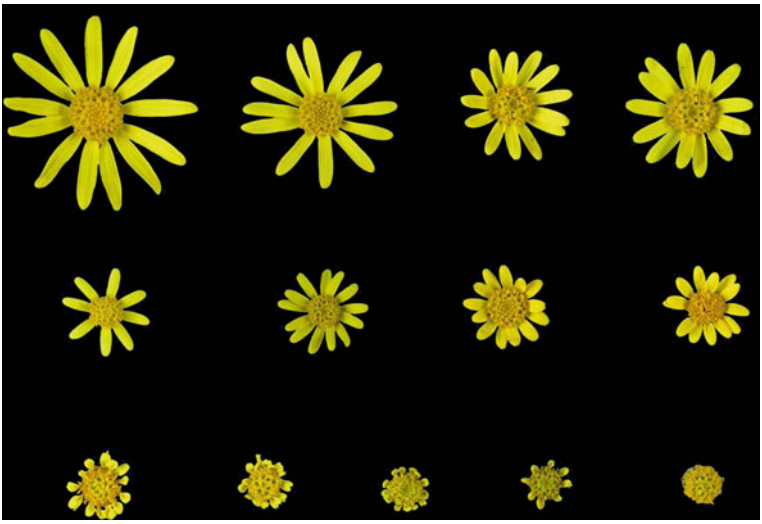

Fig. 13.5 Variation in ray flower morphology observed in individuals of synthetic lines of allohexaploid *S. cambrensis* from the S2 to S5 generation

(Hegarty et al. 2011). Genes shown to be involved in ray flower development in *Senecio* (*RAY1* and *RAY2*) are orthologues of *CYCLOIDEA* (*CYC*) (Kim et al. 2008), which has been shown to occur as a stable mutant (hypermethylated) epiallele that in *Linaria* (toadflax, Plantaginaceae) manifests itself in a change in flower symmetry from bilateral to radial symmetry (Cubas et al. 1999). We therefore hypothesise that the observed changes in ray flower form may in part be associated with epigenetic modifications to *RAY1* and/or *RAY2*. Testing this hypothesis will be a focus for future work.

Another unexpected observation in the synthetic *S. cambrensis* lines was the appearance of self-sterile individuals in otherwise SC lines, again from the S_2 generation onwards. Subsequent analyses using controlled self- and cross-pollinations confirmed that these self-sterile individuals possessed functional sporophytic self-incompatibility (Brennan and Hiscock 2010). This is the first time that sporophytic SI has been shown to be inherited and expressed in an allopolyploid and raises intriguing questions about the mechanism regulating this important trait. All S_0 and S_1 lines of synthetic *S. cambrensis* were highly self-fertile (SC) indicating that the SI system, present in parental *S. squalidus*, was repressed in these allopolyploids, only to be reactivated/derepressed later. The emergence of SI individuals may be a consequence of recombination or might also be associated with epigenetic changes observed in the early-generation synthetic allopolyploids. Most observations of wild *S. cambrensis* have reported it to be SC (Abbott and Lowe 2004), but the finding of SI in synthetic *S. cambrensis* prompted a re-examination of the mating system of wild *S. cambrensis.* An analysis of selfing rates in 41 wild *S. cambrensis* individuals from Edinburgh and North Wales

identified five SI individuals (Brennan and Hiscock 2010) implying, albeit from a relatively small sample size, that SI may be present in wild *S. cambrensis* at a frequency of ~12 %. This important finding means that *S. cambrensis* should now be considered as possessing a mixed mating system that has the potential to evolve towards either outcrossing or selfing.

13.3 Future Prospects

13.3.1 Next-Generation Approaches to Studying Evolution in S. cambrensis

We are currently engaged in the generation of a draft reference sequence of the *Senecio squalidus* "gene-space" (low-copy, non-methylated regions of the genome). Next-generation sequencing (NGS) platforms enable a variety of potential experiments to examine the consequences of polyploidy and hybridization in *Senecio* further. For example, we intend to identify promoter regions using chromatin-immunoprecipitation sequencing (ChIPseq) to enrich for DNA fragments bound by enzymes involved in transcription. We can then determine whether hybrids and polyploids display alterations in promoter sequence/binding that may explain the altered patterns of expression observed in our microarray experiments. Further, once a reference sequence is available, we can consider bisulphite sequencing to identify genomic regions which show differential methylation in hybrids relative to their progenitors. One key target for bisulphite sequencing will be the *RAY1* and *RAY2* genes (Kim et al. 2009) which we suspect may show differential methylation associated with the observed variation in ray flower morphology that appears in synthetic *S. cambrensis* lines. Identification of small interfering RNAs (siRNAs) and their targets will enable analysis of changes to epigenetic regulation of gene expression in hybrids. At the structural level, resequencing of hybrid "gene-space" and comparative sequence analysis may allow us to detect genomic rearrangements and sequence loss, plus the activity of transposable elements. The increasing capability of genotyping-by-sequencing (GBS) approaches such as restriction-associated DNA (RAD) sequencing (Baird et al. 2008) may also prove useful in detecting structural rearrangements in hybrid genomes. Finally, comparative sequencing of *Senecio cambrensis* and the two other hybrid derivatives of *S. vulgaris* and *S. squalidus*, i.e. *S. vulgaris* var. *hibernicus* and *S. eboracensis,* may enable identification of dosage effects, since the hybrids share parental genomes but differ in the specific combinations thereof.

13.3.2 S. cambrensis in the Wild

Senecio cambrensis now exists in the wild only in North Wales, UK, following extinction of the Edinburgh population in 1993 (Abbott and Forbes 2002).

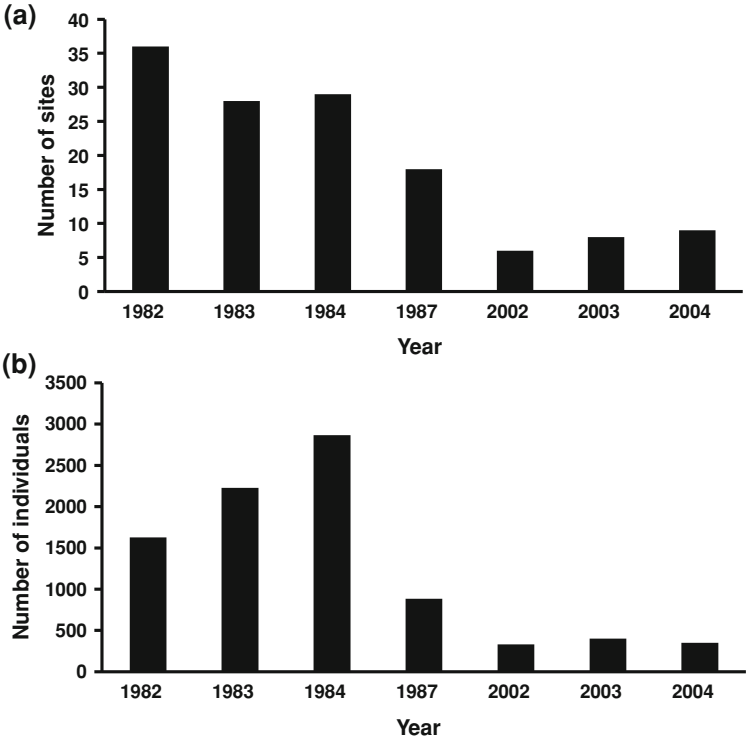

Fig. 13.6 Decline in the number of **a** sites and **b** flowering individuals across sites, recorded for *S. cambrensis* in North Wales in different years from 1982 to 2004. Records are from Abbott et al. (2007) and were made in June each year except in 1987 when they were collected in May

Although the species was recorded at numerous sites and in relatively high numbers in North Wales in the early 1980s (Ingram and Noltie 1995; Fig. 13.6), it has undergone a significant decline in this region since then. In the most recent census of the species undertaken in 2004, it was found at only nine sites in North Wales with a total of 349 individuals recorded across all sites (Abbott et al. 2007; Fig. 13.6). It has been speculated that this decline in numbers has been caused by a reduction of available sites for colonisation by the species (e.g., waste ground) and also to increased use of herbicide to control weed populations (Abbott et al. 2007). Its disappearance from the thin layer of soil that collects along roads between the road edge and verge, where it was found often in the past, is most likely due to increased use of herbicides on plants growing along road sides. However, it has also been noted that plants of the species are frequently heavily infected with the rust *Puccinia lagenophorae*. Although this rust also infects both parent species, it is possible that its effects on *S. cambrensis* are more dramatic in terms of production of offspring for colonising new sites, given that numbers of *S. squalidus* and *S. vulgaris* that reproduce each year in North Wales and elsewhere in the UK are vast relative to those of *S. cambrensis*. Although no detailed analysis has been

undertaken on whether *S. cambrensis* is ecologically divergent from its two parents, it tends to grow in sympatry with one or both parent species in the wild. Thus, there is likely competition between the three species for occupation of available open sites, and this might place *S. cambrensis* at a disadvantage if the seed it produces is less numerous relative to that of *S. vulgaris* and/or *S. squalidus* occurring in the same area. Whatever the cause of its marked decline in numbers in North Wales over the last 25 years or so, it is clear that *S. cambrensis* has reached the stage where its presence in the wild is endangered and that a conservation plan is required to prevent it from possibly becoming extinct in the near future.

Given the decline in numbers of *S. cambrensis* in recent years it is important to gain a better understanding of the nature of its mating system and mating dynamics, especially in the light of our findings of SI in wild populations (Brennan and Hiscock 2010). If there has been a recent shift in mating system from predominantly SC towards SI, it is possible that mating potential has become compromised due to obligate outcrossing. Given the possibility of a single hybrid origin for *S. cambrensis* in Wales, it is likely that wild populations possess very few *S* alleles. Whilst a small number of shared *S* alleles are not necessarily a problem when populations are large, it becomes a problem when populations are in decline and when stochastic effects can lead to *S* allele loss (Brennan et al. 2003; Pickup and Young 2008). This can then exacerbate decline, leading to an uncontrollable spiral of extinction. A reappraisal of the mating system of *S. cambrensis* is thus urgently needed.

Acknowledgments Our research on *Senecio cambrensis* and its parents has been supported by grants from the NERC, BBSRC and Leverhulme Trust to whom we are grateful. We would also like to acknowledge our long-term collaboration with Keith Edwards and the considerable efforts of former research students, technicians and postdocs who contributed to the research over the past 25 years. These include Alexandra Allen, Paul Ashton, Garry Barker, Tom Batstone, Adrian Brennan, Mark Chapman, David Forbes, Amanda Gillies, Helen Ireland, Juliet James, Joanna Jones, Andrew Lowe, David Tabah, and Ian Wilson.

References

Abbott RJ, Forbes DG (1993) Outcrossing rate and self-incompatibility in the colonizing species *Senecio squalidus* L. Heredity 71:155–159

Abbott RJ, Forbes DG (2002) Extinction of the edinburgh lineage of the allopolyploid neospecies, *Senecio cambrensis* Rosser (Asteraceae). Heredity 88:267–269

Abbott RJ, Lowe AJ (1996) A review of hybridization and evolution in British *Senecio* In:Hind DJN, Beentje HJ (eds.) Compositae: systematics. Proceedings of the international compositae conference Kew 1994, Royal Botanic Gardens, Kew, pp 679–689

Abbott RJ, Lowe AJ (2004) Origins, establishment and evolution of new polyploid species: *Senecio cambrensis* and *S. eboracensis* in the British isles. Biol J Linn Soc 82:467–474

Abbott RJ, Ashton PA, Forbes DG (1992) Introgressive origin of the radiate groundsel *Senecio vulgaris* L. var. *hibernicus* Syme: *Aat-3* evidence. Heredity 68:425–435

Abbott RJ, Ingram R, Noltie HJ (1983) Discovery of *Senecio cambrensis* Rosser in edinburgh. Watsonia 14:407–408

Abbott RJ, Ireland HE, Rogers HJ (2007) Population decline despite high genetic diversity in the new allopolyploid species *Senecio cambrensis* (Asteraceae). Mol Ecol 16:1023–1033

Abbott RJ, Brennan AC, James JK, Forbes DG, Hegarty MJ, Hiscock SJ (2009) Recent hybrid origin and invasion of the British isles by a self-incompatible species, Oxford ragwort (*Senecio squalidus* L., Asteraceae). Biol Invasions 11:1145–1158

Abbott RJ, Hegarty MJ, Hiscock SJ, Brennan AC (2010) Homoploid hybrid speciation in action. Taxon 59:1375–1386

Adams KL, Percifield R, Wendel JF (2004) Organ-specific silencing of genes in a newly synthesized cotton allotetraploid. Genetics 168:2217–2226

Adams KL, Wendel JF (2005) Polyploidy and genome evolution in plants. Curr Opin Plant Biol 8:135–141

Ashton PA, Abbott RJ (1992) Multiple origins and genetic diversity in the newly arisen allopolyploid species, *Senecio cambrensis* Rosser (Compositae). Heredity 68:25–32

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker WA, Cresko WA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3:e3376

Baker HG (1967) Support for Baker's' Law—as a rule. Evolution 21:853–856

Benoit PM, Crisp PC, Jones BMG (1975) *Senecio L.* In: Stace CA (ed) Hybridization and the flora of the British isles. Academic, London, pp 404–410

Brennan AC, Hiscock SJ (2010) Expression and inheritance of sporophytic self-incompatibility in synthetic allohexaploid *Senecio cambrensis* (Asteraceae). New Phytol 186:251–261

Brennan ACE, Harris SA, Tabah DA, Hiscock SJ (2002) The population genetics of sporophytic self-incompatibility in *Senecio squalidus* L. (Asteraceae) I: *S* allele diversity in a natural population. Heredity 89:430–438

Brennan ACE, Harris SA, Hiscock SJ (2003) The population genetics of sporophytic self-incompatibility in *Senecio squalidus* L. (Asteraceae): avoidance of mating constraints imposed by low *S*-allele number. Phil Trans R. Soc Lond B 358:1047–1050

Brennan AC, Bridle JR, Wang A-L, Hiscock SJ, Abbott RJ (2009) Adaptation and selection in the *Senecio* (Asteraceae) hybrid zone on Mount Etna, Sicily. New Phytol 183:702–717

Brennan ACE, Harris SA, Hiscock SJ (2006) The population genetics of sporophytic self-incompatibility in *Senecio squalidus* L. (Asteraceae): *S* allele diversity across the British range. Evolution 60:213–224

Brennan ACE, Harris SA, Hiscock SJ (2005) Modes and rates of selfing and associated inbreeding depression in the self-incompatible plant *Senecio squalidus* (Asteraceae): a successful colonizing species in the British Isles. New Phytol 168:475–486

Brennan AC, Barker D, Hiscock SJ, Abbott RJ (2012) Molecular genetic and quantitative trait divergence associated with recent homoploid hybrid speciation: a study of *Senecio squalidus* (Asteraceae). Heredity 109:87–95

Buggs RJA, Soltis PS, Mavrodiev EV, Symonds VV, Soltis DE (2008) Does phylogenetic distance between parental genomes govern the success of polyploids? Castanea 73:74–93

Buggs RJA, Doust AN, Tate JA, Koh J, Soltis K, Feltus FA, Paterson A, Soltis PS, Soltis DE (2009) Gene loss and silencing in *Tragopogon miscellus* (Asteraceae): comparison of natural and synthetic allotetraploids. Heredity 103:73–81

Buggs RJA, Zhang L, Miles N, Tate JA, Gao L, Wei W, Schnable PS, Barbazuk WB, Soltis PS, Soltis DE (2011) Transcriptomic shock generates evolutionary novelty in a newly formed natural allopolyploid plant. Curr Biol 21:551–556

Chaudhary B, Flagel L, Stupar RM, Udall JA, Verma N, Springer NM, Wendel JH (2009) Reciprocal silencing, transcriptional bias and functional divergence of homeologs in polyploid cotton (gossypium). Genetics 182:503–517

Chapman MA, Burke JM (2007) Genetic divergence and hybrid speciation. Evolution 61:1773–1780

Coleman M, Forbes DG, Abbott RJ (2001) A new subspecies of *Senecio mohavensis* (Compositae) reveals an old–new World species disjunction. Edinburgh J Botany 58:384–403

Crisp PC (1972) Cytotaxonomic studies in the section Annui of Senecio. Ph.D Thesis, University of London

Cubas P, Vincent C, Coen E (1999) An epigenetic mutation responsible for natural variation in floral symmetry. Nature 401:157–161

Dong YZ, Liu ZL, Shan XH, Qiu T, He MY, Liu B (2005) Allopolyploidy in wheat induces rapid and heritable alterations in DNA methylation patterns of cellular genes and mobile elements. Genetika 41:1089–1095

Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF (2008) Evolutionary genetics of genome merger and doubling in plants. Annu Rev Genet 42:443–461

Feldman M, Levy AA (2005) Allopolyploidy—a shaping force in the evolution of wheat genomes. Cytogenet Genome Res 109:250–258

Garcı′a-Ortiz MV, Ariza RR, Rolda′n-Arjona T (2001) An OGG1 orthologue encoding a functional 8-oxoguanine DNA glycosylase/lyase in *Arabidopsis thaliana*. Plant Mol Biol 47:795–804

Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC (2007) Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. Plant Cell 19:3403–3417

Gray AJ, Marshall DF, Raybould AF (1991) A century of evolution in *Spartina anglica*. Adv Ecol Res 21:1–62

Grant V (1981) Plant speciation. Columbia University Press, New York

Harland SC (1955) The experimental approach to the species problem. In: Lousley JE (ed) Species studies in the British flora. Botanical Society of the British Isles, London, pp 16–20

Harris SA (2002) Introduction of Oxford ragwort, *Senecio squalidus* L. (Asteraceae), to the United Kingdom. Watsonia 24:31–43

Harris SA, Ingram R (1992) Molecular systematics of the genus *Senecio* L. I. hybridization in a British polyploid complex. Heredity 69:1–10

Hegarty MJ, Hiscock SJ (2008) Genomic clues to the evolutionary success of polyploid plants. Curr Biol 18:R435–R444

Hegarty MJ, Jones JM, Wilson ID, Barker GL, Coghill JA, Sanchez-Baracaldo P, Liu G, Buggs RJA, Abbott RJ, Edwards KJ, Hiscock SJ (2005) Development of anonymous cDNA microarrays to study changes to the *Senecio* floral transcriptome during hybrid speciation. Mol Ecol 14:2493–2510

Hegarty MJ, Barker GL, Wilson ID, Abbott RJ, Edwards KJ, Hiscock SJ (2006) Transcriptome shock after interspecific hybridization in *Senecio* is ameliorated by genome duplication. Curr Biol 16:1652–1659

Hegarty MJ, Barker GL, Brennan AC, Edwards KJ, Abbott RJ, Hiscock SJ (2008) Changes to gene expression associated with hybrid speciation in plants: further insights from transcriptomic studies in *Senecio*. Philos Trans R Soc Lond B 363:3055–3069

Hegarty MJ, Batstone T, Barker GL, Edwards KJ, Abbott RJ, Hiscock SJ (2011) Nonadditive changes to cytosine methylation as a consequence of hybridization and genome duplication in *Senecio* (Asteraceae). Mol Ecol 20:105–113

Hiscock SJ (2000a) Genetic control of self-incompatibility in *Senecio squalidus* L. (Asteraceae) – a successful colonising species. Heredity 84:10–19

Hiscock SJ (2000b) Self-incompatibility in *Senecio squalidus* L. (Asteraceae). Ann Bot 85(Supplement A):181–190

Ingram R (1978) The genomic relationship of *Senecio squalidus* L. and *Senecio vulgaris* L. and the significance of genomic balance in their hybrid *S. x baxteri* Druce. Heredity 40:459–462

Ingram R, Noltie HJ (1984) Ray floret morphology and the origin of variability in *Senecio cambrensis* Rosser, a recently established allopolyploid species. New Phytol 96:601–607

Ingram R, Noltie HJ (1989) Early adjustment of patterns of metaphase association in the evolution of polyploid species. Genetica 78:21–24

Ingram R, Noltie HJ (1995) Biological flora of the British isles: *Senecio cambrensis* Rosser. J Ecol 83:537–546

James JK, Abbott RJ (2005) Recent, allopatric, homoploid hybrid speciation: the origin of *Senecio squalidus* (Asteraceae) in the British Isles from a hybrid zone on Mount Etna, Sicily. Evolution 59:2533–2547

Jiao L, Wickett NJ, Ayyampalayam S, Chanderbali et al (2011) Ancestral polyploidy in seed plants and angiosperms. Nature 473:97–100

Kadereit JW, Uribe-Convers S, Westberg E, Comes HP (2006) Reciprocal hybridization at different times between *Senecio flavus* and *Senecio glaucus* gave rise to two polyploidy species in north Africa and south-west Asia. New Phytol 169:431–441

Kent DH (1963) *Senecio squalidus* L. in the British isles. 7 Wales. Nat Wales 8:175–178

Kim M, Cui M-L, Cubas P, Gillies A, Lee K, Chapman MA, Abbott RJ, Coen E (2008) Regulatory genes control a key morphological and ecological trait transferred between species. Science 322:1116–1119

Leitch IJ, Bennett MD (1997) Polyploid in angiosperms. Trends Plant Sci 2:470–476

Leitch AR, Leitch IJ (2008) Genomic plasticity and diversity of polyploid plants. Science 320:481–483

Liu Z, Adams KL (2007) Expression partitioning between genes duplicated by polyploidy under abiotic stress and during organ development. Curr Biol 17:1669–1674

Liu B, Brubaker CL, Mergeai G, Cronn RC, Wendel JF (2001) Polyploid formation in cotton is not accompanied by rapid genomic changes. Genome 44:321–330

Lowe AJ, Abbott RJ (2000) Routes of origin of two recently evolved hybrid taxa: *Senecio vulgaris* var. *hibernicus* and York radiate groundsel (Asteraceae). Am J Bot 87:1159–1167

Lowe AJ, Abbott RJ (2003) A new British species, *Senecio eboracensis* (Asteraceae), another hybrid derivative of *S. vulgaris* L. and *S. squalidus* L. Watsonia 24:375–388

Lukens LN, Pires JC, Leon E, Vogelzang R, Oslach L, Osborn T (2006) Patterns of sequence loss and cytosine methylation within a population of newly resynthesized *Brassica napus* allopolyploids. Plant Physiol 140:336–348

McClintock B (1984) The significance of responses of the genome to challenge. Science 226:792–801

Madlung A, Masuelli RW, Watson B, Reynolds SH, Davison J, Comai L (2002) Remodeling of DNA methylation and phenotypic and transcriptional changes in synthetic *Arabidopsis* allotetraploids. Plant Physiol 129:733–746

Marshall DF, Abbott RJ (1980) On the frequency of introgression of the radiate (*Tr*) allele from *Senecio squalidus* L. into *Senecio vulgaris* L. Heredity 45:133–135

Moffatt BA, Allen M, Snider S, Pereira LA, Todorova M, Summers PS, Weretilnyk EA, Martin-McCaffrey L, Wagner C (2002) Adenosine kinase deficiency is associated with developmental abnormalities and reduced transmethylation. Plant Physiol 128:812–821

Mull L, Ebbs ML, Bender J (2006) A histone methylation-dependent DNA methylation pathway is uniquely impaired by deficiency in *Arabidopsis* S-adenosylhomocysteine hydrolase. Genetics 174:1161–1171

Novak SJ, Soltis DE, Soltis PS (1991) Ownbey's Tragopogons: 40 years later. Am J Bot 78:1586–1600

Otto SP, Whitton J (2000) Polyploidy: incidence and evolution. Annu Rev Genet 34:401–437

Parisod C, Salmon A, Zerjal T, Tenaillon M, Grandbastien MA, Ainouche M (2010) Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in *Spartina*. New Phytol 184:1003–1015

Paun O, Forest F, Fay MF, Chase MW (2009) Hybrid speciation in angiosperms: parental divergence drives ploidy. New Phytol 182:507–518

Pelser PB, Nordenstam B, Kadereit JW, Watson LE (2007) An ITS phylogeny of tribe Senecioneae (Asteraceae) and a new delimitation of *Senecio* L. Taxon 56:1062–1077

Pelser PB, Abbott RJ, Comes HP, Milton JJ, Möller M, Looseley ME, Cron GV, Barcelona JF, Kennedy AH, Watson LE, Barone R, Hernández F, Kadereit JW (2012) The genetic ghost of an invasion past: colonization and extinction revealed by historical hybridization in *Senecio*. Mol Ecol 21:369–387

Pereira LA, Schoor S, Goubet F, Dupree P, Moffatt BA (2006) Deficiency of adenosine kinase activity affects the degree of pectin methyl-esterification in cell walls of *Arabidopsis thaliana*. Planta 224:1401–1414

Pickup M, Young AG (2008) Population size, self-incompatibility and genetic rescue in diploid and tetraploid races of *Rutidosis leptorrhynchoides* (Asteraceae). Heredity 100:268–274

Pumphrey M, Bai J, Laudencia-Chingcuanco D, Anderson O, Gill BS (2009) Nonadditive expression of homoeologous genes is established upon polyploidization in hexaploid wheat. Genetics 181:1147–1157

Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K, Nakazato T, Durphy JL, Schwarzbach AE, Donovan LA, and Lexer C (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. Science 301:1211–1216

Rieseberg LH, Willis JH (2007) Plant speciation. Science 317:910–914

Rieseberg LH, Archer MA, Wayne RK (1999) Transgressive segregation, adaptation and speciation. Heredity 83:363–372

Rosser EM (1955) A new British species of *Senecio*. Watsonia 3:228–232

Salone V, Rudinger M, Polsakiewicz M, Hoffmann B, Groth-Malonek M, Szurek B, Small I, Knoop V, Lurin C (2007) A hypothesis on the identification of the editing enzyme in plant organelles. FEBS Lett 581:4132–4138

Salmon A, Ainouche ML (2010) Polyploidy and DNA methylation: new tools available. Mol Ecol 19:213–215

Salmon A, Ainouche ML, Wendel JF (2005) Genetic and epigenetic consequences of recent hybridization and polyploidy in *Spartina* (Poaceae). Mol Ecol 14:1163–1175

Soltis PS, Soltis DE (1999) Polyploidy: recurrent formation and genome evolution. Trends Ecol Evol 14:348–352

Stebbins GL (1957) Self fertilization and population variability in higher plants. Am Nat 91:337–354

Stupar RM, Hermanson PJ, Springer NM (2007) Nonadditive expression and parent-of-origin effects identified by microarray and allele-specific expression profiling of maize endosperm. Plant Physiol 145:411–425

Sun M, Ritland K (1998) Mating system of yellow starthistle (*Centaurea solstitialis*), a successful colonizer in North America. Heredity 80:225–232

Urbanska KM, Hurka H, Landolt E, Neuffer B, Mummenhoff K (1997) Hybridization and evolution in *Cardamine* (Brassicaceae) at Urnerboden, central Switzerland: biosystematics and molecular evidence. Plant Syst Evol 204:233–256

Vincent PLD (1996) Progress on clarifying the generic concept of *Senecio* based on an extensive world-wide sample of taxa. In Hind DJN, Beentje HJ (eds) Compositae: systematics. proceedings of the international compositae conference Kew 1994, vol 1, Royal Botanic Gardens, Kew, pp 597–611

Wang J, Tian L, Lee H-Y, Chen ZJ (2006a) Nonadditive regulation of FRI and FLC loci mediates flowering-time variation in *Arabidopsis* allopolyploids. Genetics 173:965–974

Wang J, Tian L, Lee H-S, Wei NE, Jiang H, Watson B, Madlung A, Osborn TC, Doerge RW, Comai L, Chen ZJ (2006b) Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. Genetics 172:507–517

Weir J, Ingram R (1980) Ray morphology and cytological investigations of *Senecio cambrensis* Rosser. Heredity 86:237–241

Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH (2009) The frequency of polyploid speciation in flowering plants. Proc Nat Acad Sci USA 106:13875–13879

Xiong LZ, Xu CG, Saghai Maroof MA, Zhang Q (1999) Patterns of cytosine methylation in an elite rice hybrid and its parental lines, detected by a methylation-sensitive amplification polymorphism technique. Mol Gen Genet 261:439–446

Chapter 14
The Early Stages of Polyploidy: Rapid and Repeated Evolution in *Tragopogon*

Douglas E. Soltis, Richard J. A. Buggs, W. Brad Barbazuk, Srikar Chamala, Michael Chester, Joseph P. Gallagher, Patrick S. Schnable and Pamela S. Soltis

Abstract Elucidating the causes and consequences of polyploidy (whole-genome duplication; WGD) is arguably central to understanding the evolution of most eukaryotic lineages. However, much of what we know about these processes is derived from the study of crops and synthetic polyploids. *Tragopogon* provides the unique opportunity to investigate the genetic and genomic changes that occur across an evolutionary series from F_1 hybrids, synthetic allopolyploids, independently formed natural populations of *T. mirus* and *T. miscellus* that are 60–80 years post-formation, to older Eurasian polyploids that are dated by molecular clocks at several million years old, and finally to a putative ancient polyploidization thought to have occurred prior to or early in the history of the Asteraceae (40–43 mya). *Tragopogon* joins other well-studied natural polyploid systems (e.g., *Glycine*, *Nicotiana*, *Gossypium*, *Spartina*, *Senecio*), but presents a range of research possibilities that is not available in any other

D. E. Soltis (✉) · R. J. A. Buggs · W. B. Barbazuk · S. Chamala · M. Chester · J. P. Gallagher
Department of Biology, University of Florida, Gainesville, FL 32611, USA
e-mail: dsoltis@botany.ufl.edu

Present Address:
J. P. Gallagher
Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA 50011, USA

R. J. A. Buggs
School of Biological and Chemical Sciences, Queen Mary University of London, London, E1 4NS, UK

P. S. Schnable
Center for Plant Genomics, Iowa State University, Ames, IA 50011, USA

P. S. Soltis
Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA

P. S. Soltis and D. E. Soltis (eds.), *Polyploidy and Genome Evolution*,
DOI: 10.1007/978-3-642-31442-1_14,

system. We have shown in *T. mirus* and *T. miscellus* that upon allopolyploidization, massive gene loss occurs in patterns that are repeated across populations of independent origin and with a bias against genes derived from *T. dubius*, the diploid parent shared by both new allotetraploids. We have also shown significant changes in gene expression (transcriptomic shock) in the early generations of allopolyploidy in these species. Massive and repeated patterns of chromosomal variation (intergenomic translocations and aneuploidy) have been revealed by fluorescence in situ hybridization. Aneuploidy results in substitutions between homeologous chromosomes, through reciprocal monosomy-trisomy (1:3 copies) or nullisomy-tetrasomy (0:4 copies). We propose that substantial chromosomal instability results in karyotype restructuring, a likely common process following WGD and a driver of allopolyploid speciation, which has largely unexplored implications for gene losses, gains, and expression patterns. But gene loss and expression changes as well as karyotypic changes are ongoing in *T. mirus* and *T. miscellus*, in that no population is fixed for any of these events; thus, we have literally caught evolution in the act.

14.1 Introduction

14.1.1 General Introduction

Polyploidy, or whole-genome duplication (WGD), is currently recognized as a major evolutionary force in eukaryotes (e.g., Mable 2003; Gregory and Mable 2005; Mable et al. 2011). Polyploidy generally results in instant speciation, increasing biodiversity and providing new genetic material for evolution (e.g., Levin 1983, 2002). Some evidence suggests that two polyploid events occurred in ancestors of vertebrates (Ohno 1970; Panapoulou and Poustka 2005; see Chap. 16 of this volume), with subsequent polyploidy in amphibians and fish (Mable et al. 2011; see Chaps. 18 and 17, respectively, this volume). The genomes of yeast and other *Saccharomyces* also appear to be anciently duplicated (Wolfe and Shields 1997; Kellis et al. 2004; Dujon et al. 2004; see Chap. 15 of this volume). Researchers have long recognized that polyploidy is an inseparable part of angiosperm biology; the polyploidy process has, in fact, been studied in plants for a little over a century. Early reviews of polyploidy in plants include now classic papers by Müntzing (1936), Darlington (1937), Clausen et al. (1945), Löve and Löve (1949), and Stebbins (1950, 1971). However, these seminal works did not anticipate the huge role for polyploidy in evolution that genomic studies now suggest.

During the past decade there has been a tremendous resurgence of interest in polyploidy, stimulated in large part by the development of increasingly powerful genetic and genomic tools. The result has been numerous new insights into the genomic and genetic consequences of polyploidy (other chapters of this volume). Recent discoveries have dramatically reshaped traditional views and concomitantly revealed that polyploidy is a highly dynamic and ubiquitous process. For example, studies of many duplicated genes across genomes suggest that all

angiosperms have undergone at least one round of genome doubling. Significantly, genomic and phylogenetic analyses also associate polyploidy with major diversifications (e.g., within Poales, Solanaceae, Fabaceae; Soltis et al. 2009a), and the origin of angiosperms and seed plants (Jiao et al. 2011).

As a result of a diverse array of studies, we have learned a great deal about the interactions that occur among the diploid genomes forced together via allopolyploidy. Because of its apparent prevalence, elucidating the causes and consequences of polyploidy is arguably central to understanding the origin and diversification of most major lineages of eukaryotes. Significantly, however, most of what we know about the genetic and genomic consequences of polyploidy is derived from the study of synthetic polyploids, crops (e.g., cotton, wheat, *Brassica*, *Nicotiana*), and genetic models (*Arabidopsis*) (see refs above, Chen et al. 2004; Soltis and Soltis 2009). To understand better how polyploidization impacts genome evolution and gene function in natural populations, we must extend from a few crops, genetic models, and synthetics to naturally occurring polyploids.

Three systems are known that permit insights into the early stages of polyploidy in nature: *Spartina anglica* (Ainouche et al. 2004, 2009; Salmon et al. 2005; see Chap. 12 of this volume), *Senecio cambrensis* (Ashton and Abbott 1992; Abbott and Lowe 2004; Hegarty et al. 2005, 2006; see Chap. 13, this volume), and *Tragopogon* (*T. mirus*, *T. miscellus*, Soltis et al. , 2009b). Research has progressed on all, and they are complementary. In addition, a2004s stressed here, *Tragopogon* also affords unique opportunities to investigate polyploidy over a continuum of ages, as well as the consequences of recent and frequently repeated polyploidy.

14.1.2 Introduction to Tragopogon

Tragopogon now provides a well-known, textbook example of recent allopolyploid speciation; two new allotetraploid species originated within the last 80 years (Soltis et al. 2004, 2009b) (Fig. 14.1). *Tragopogon mirus* and *T. miscellus* (Fig. 14.1) formed repeatedly following the introduction of three diploids from Europe into the Palouse region of North America in the early 1900s (Ownbey 1950; Soltis et al. 2004); the tetraploids have not formed in Europe. The parentage of *T. mirus* (*T. dubius* and *T. porrifolius*) and *T. miscellus* (*T. dubius* and *T. pratensis*) is well documented (Figs. 14.1 and 14.2) and confirmed with multiple markers and approaches (Soltis et al. 1995, 2004, 2009b).

Ownbey (1950) described the few populations of the newly formed allotetraploids (each consisting of fewer than 100 individuals) as "small and precarious", but noted that they had "attained a degree of success" and they appeared to be "competing successfully" with their diploid parents. He also stated that it would be "important to follow the ecological development of the newly formed polyploids" through time. Significantly, both tetraploids have been highly successful since their formation. Novak et al. (1991) conducted a survey to determine the distributions of the two polyploids 40 years after Ownbey's discovery. One or

both polyploids were found in most towns of the Palouse with populations ranging from small (fewer than 100 individuals) to many thousands of individuals. *Tragopogon miscellus* is now one of the most common weeds in and around Spokane, WA, as well as in Moscow, ID, and Spangle, WA. Populations of *T. mirus* and *T. miscellus* often form dense stands and are, in fact, displacing their parents, particularly *T. pratensis* and *T. porrifolius*.

14.1.3 Tragopogon *as a Unique Evolutionary Model*

For many reasons *Tragopogon* is a novel system that affords the opportunity to examine the early stages of polyploidization in nature. The natural populations are approximately 80 years (40 generations in these biennials) old—this time frame and the fact that they have experienced natural selection provide a window into polyploidization that cannot be matched via the study of synthetic polyploids (e.g., crops and genetic models). Furthermore, molecular studies suggest that the diploid parents diverged ~2.5–5 MYA and that there may be as many as 21 lineages of separate origin of *T. miscellus* and 11 of *T. mirus* just in the Palouse (Soltis et al. 1995, 2004, 2009b; Symonds et al. 2010); the polyploids have also formed in Arizona, Oregon, Wyoming, and Montana (Soltis et al. 2012; Ownbey unpublished data). Recent studies employing microsatellite markers (Symonds et al. 2010) reveal multiple origins on a small geographic scale (see Sect. 14.1.4). Given that multiple polyploidizations are common in plants (Soltis and Soltis 1993, 1999, 2000, 2009), *Tragopogon* represents in microcosm what occurs in other polyploids over much larger geographic areas and longer time frames. These repeated origins in a small geographic area and a narrow time frame also provide the unique opportunity to ask if evolution repeats itself across these many lineages.

Adding to the utility of the *Tragopogon* system as an evolutionary model is the recent production of multiple synthetic lines of both *T. mirus* and *T. miscellus* (Tate et al. 2009a), providing the added opportunity of examining both species from polyploidization onward. Morphologically, the synthetics resemble the natural polyploids with short- and long-liguled forms of *T. miscellus* resulting when *T. pratensis* and *T. porrifolius* are reciprocally crossed (Tate et al. 2009a). In nature, all formations of *T. mirus* have *T. porrifolius* as the maternal parent and *T. dubius* as the paternal parent, but we have synthesized *T. mirus* reciprocally. We also produced allotetraploids between *T. porrifolius* and *T. pratensis*, which are not known from nature (Fig. 14.1). All of these synthetic lines are now in the fourth generation and offer the unique opportunity for comparative study of repeated formations of both natural and synthetic polyploids.

Further adding to the allure of *Tragopogon* are Old World polyploids that are much older than the recently formed New World polyploids. *Tragopogon* comprises ~150 species, 12 of which are Eurasian polyploids (Mavrodiev et al. 2008a). Of several Eurasian polyploids for which we clarified parentage (Mavrodiev et al. 2008a, b, c), *T. castellanus* ($2n = 24$) from Spain (Blanca and Díaz de la Guardia 1996) has emerged as a promising new model. Estimating the

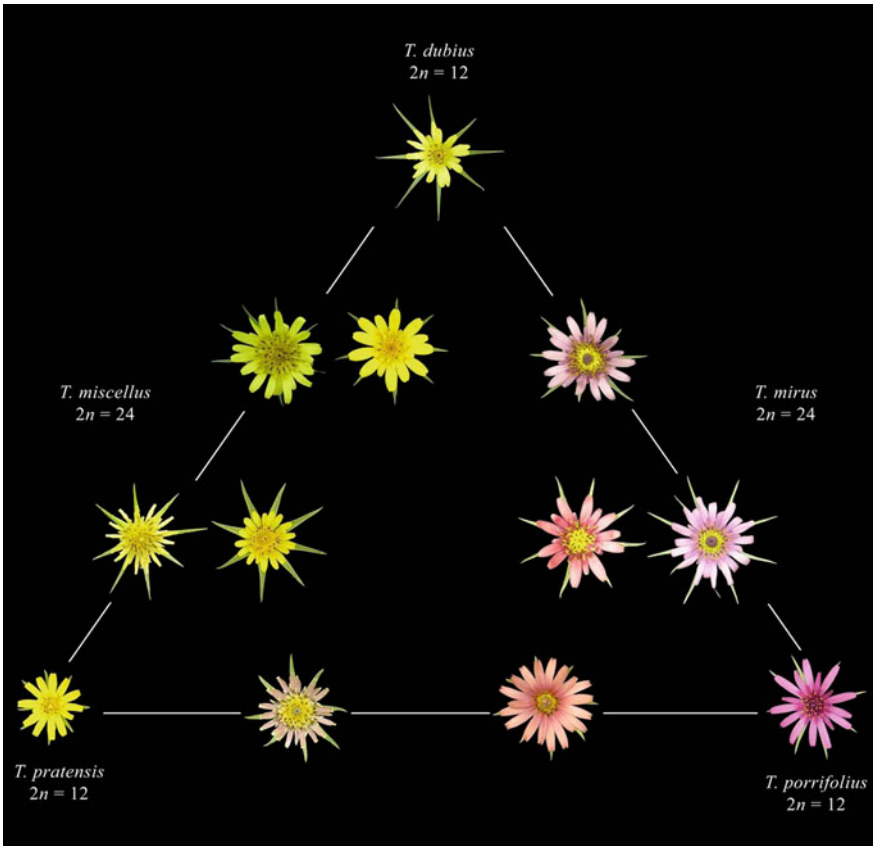

Fig. 14.1 Summary of parentage of tetraploid *Tragopogon* species comparing natural with synthetic allopolyploids. The diploid parents (with $2n = 12$) are at the corners of the *triangle*; polyploids ($2n = 24$) are in between the corners. Synthetic polyploids are on the outside of the *triangle* (connected by *white lines*); those polyploids forming naturally are to the inside of the *triangle*. Polyploids are placed closer to the maternal parent. In nature, *T. miscellus* has formed reciprocally, and *T. mirus* has formed only with *T. porrifolius* as the maternal parent. However, we have made reciprocal synthetic lines of both and have also made reciprocal polyploids of *T. pratensis* × *T. porrifolius*; this polyploid has not formed in nature. Note that populations of *T. miscellus* of reciprocal origin differ in morphology. Those with *T. pratensis* as the maternal parent have short ligules, and those with *T. dubius* as the maternal parent have long ligules

date of origin of a polyploid is difficult, but given well-known caveats, we used DNA sequence data to estimate the age of *T. castellanus* as 0.8–2.8 million years (Mavrodiev et al. 2008a, b, c). This agrees with other evidence in that *T. castellanus* occurs in well-known Pleistocene glacial refugia that harbor other paleoendemics (e.g., Petit et al. 2003). Thus, in *Tragopogon* we have the opportunity to extend our analyses to older Eurasian polyploids, providing a continuum of ages from F_1 hybrids and raw synthetics, to 80-year-old natural polyploids, to natural polyploids that are perhaps several million years old.

14.1.4 Origins of Species

The *Tragopogon* system is comparable in some ways to an island biogeography scenario. The diploids and the derivative polyploids only occur in small towns that are scattered across the Palouse area of eastern Washington and adjacent Idaho and not in the intervening areas, which are large tracts of agricultural land. This begged the question—how did the polyploids spread so quickly to many towns? Certainly, seed dispersal is one possibility given the wind-dispersed nature of the achenes. Molecular data indicate instead that repeated formation has played the major role in range expansion. This was suspected by Ownbey and McCollum (1953), who surmised that *T. miscellus* had formed reciprocally with the long-liguled form (found only in Pullman, WA) having *T. dubius* as the maternal parent and all other populations having *T. pratensis* as the maternal parent; this reciprocal parentage results in distinctive morphologies (Fig. 14.1) and was later confirmed by molecular methods (Soltis and Soltis 1989).

A suite of molecular markers including allozymes, AFLPs, sequence data, and microsatellites has now revealed that most populations of the allotetraploids are of distinct origin (Soltis et al. 2004; Symonds et al. 2010). Furthermore, microsatellites have documented multiple origins on a fine geographic scale, revealing in several cases that distinct populations in the same town separated by only 1–2 km are of separate origin (Symonds et al. 2010). Interestingly, microsatellite data reveal that of the many genotypes of *T. dubius* currently in nature, only three general types appear to have contributed to the repeated formations of both polyploids, and there are no exact matches to present-day *T. dubius* genotypes. Hence, the genotypes detected in the two polyploid species appear to represent a snapshot of the historical population structure in the diploid progenitors, rather than modern diploid genotypes.

14.2 Does Evolution Repeat Itself?

Evolutionary biologists have long wondered if evolution would repeat itself, given the chance. Gould (1994) suggested that, on a broad evolutionary scale, if we could replay the evolutionary tape of life on Earth, it would play differently—"history involves too much chaos," and too many chance events are involved for the evolutionary process to be repetitive. He stated that "chains of historical events are so intricate, so imbued with random and chaotic elements, so unrepeatable in encompassing such a multitude of unique objects, that standard models of simple prediction and replication do not apply." In contrast, other researchers have argued that "within certain limits the outcome of evolutionary processes might be rather predictable" (Morris 1998; see also Stern et al. 2009). However, is this true on a finer scale? Are certain aspects of the polyploidy process actually "hard-wired"? Preservation of duplicated gene copies following genome duplication appears far from random, with specific functional categories preferentially retained (Blanc and

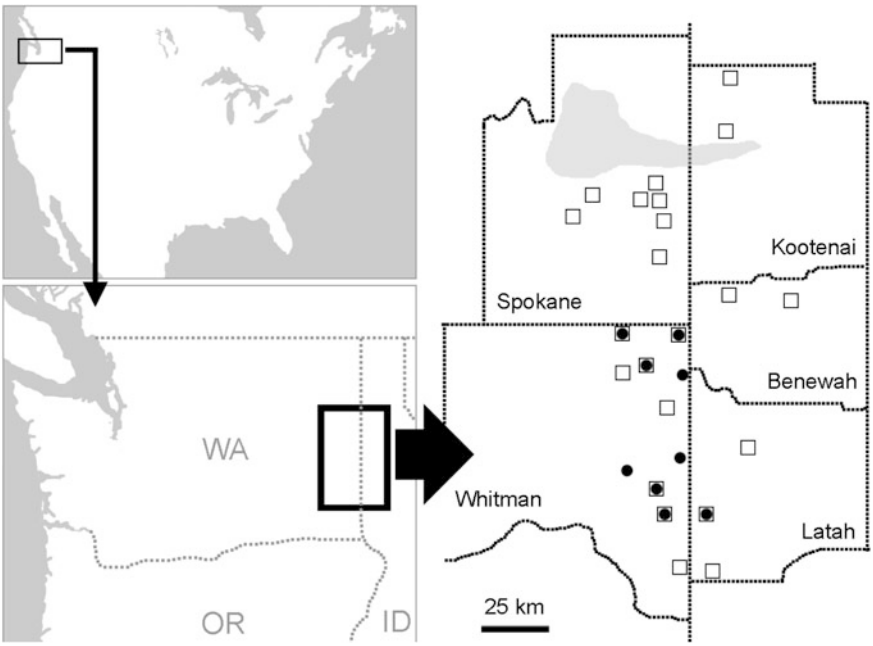

Fig. 14.2 Maps showing the location of *Tragopogon* allotetraploid populations that are found within the Palouse region of the Pacific Northwest of the U.S.A. Populations of *T. miscellus* (*squares*) and *T. mirus* (*circles*) located in towns are indicated in the relevant counties of Washington and Idaho. A contiguous urbanized area containing multiple *T. miscellus* populations is indicated (*shaded*)

Wolfe 2004; Freeling 2009) and reduplicated in subsequent polyploidizations (Paterson et al. 2006). Independent WGDs in the ancestors of *Arabidopsis*, *Oryza* (rice), *Saccharomyces* (yeast), and *Tetraodon* (pufferfish) have been followed by convergent fates of many gene families (Paterson et al. 2006). Collectively, these observations indicate that at deep timescales there *may* exist certain "principles" that govern the fates of gene and genome duplications. But is this true in the early stages following polyploid formation?

In the sections that follow we review data from different sets of molecular markers as well as chromosomal data across polyploid populations of separate origin and ask if aspects of polyploid evolution are indeed hard-wired or if stochastic processes prevail.

14.2.1 rDNA Loci/Concerted and Repeated Evolution

Concerted evolution, which results in the homogenization of gene sequences to one type, is a common feature of ribosomal RNA genes (e.g., Zimmer et al. 1980). In both *T. mirus* and *T. miscellus*, concerted evolution is ongoing, but incomplete

(Kovarik et al. 2005). In contrast to 80-year-old natural polyploids, F_1 hybrids have equal contributions of the diploid parents, as do raw (S_0) synthetic polyploids, as well as the earliest natural populations of *T. mirus* and *T. miscellus* (based on DNA from herbarium specimens). But in all modern-day natural populations except one, each representing a distinct origin, the rDNA type of *T. dubius* is consistently in very low abundance, with either the *T. pratensis* rDNA type (in *T. miscellus*) or *T. porrifolius* rDNA type (in *T. mirus*) in much greater abundance. In only one population of *T. mirus* are the parental contributions balanced (Malinska et al. 2011). Thus, concerted evolution has consistently occurred in these new polyploid lines of separate origin, and it has repeatedly operated "against" *T. dubius*, homogenizing those copies in the direction of the other parent. Surprisingly, despite being the least abundant in terms of rDNA gene copy number, *T. dubius* is by far the most abundant transcript in natural polyploid populations (Matyasek et al. 2007).

This bias against the rDNA cistron in natural polyploid populations is already apparent in the S_1 generation of synthetic polyploids (Malinska et al. 2010, 2011). For example, in four lines of synthetic *T. miscellus*, only three individuals (4 %) had balanced parental gene ratios while 65 individuals (92 %) inherited more *T. pratensis*-origin units than would be expected under additivity. In seven lines of synthetic *T. mirus*, 32 individuals (29 %) exhibited balanced rDNA genotypes, 69 individuals (63 %) showed more 35S rDNA of *T. porrifolius* origin than expected, and only 9 plants (8 %) had more *T. dubius*-origin rDNA (Malinska et al. 2010, 2011).

14.2.2 Homeolog Loss and Gene Silencing

14.2.2.1 One Gene at a Time

Although *Tragopogon* affords unique opportunities for evolutionary study, it is not a genetic model organism; hence, until recently, genetic resources have not been available (but see Sect. 14.2.2.2), which has slowed research progress. As a result, genetic and genomic changes in the newly formed tetraploids were initially examined one gene at a time for upto 29 loci (Tate et al. 2006, 2009b; Buggs et al. 2009, 2010b; Koh et al. 2010). Initially, we used AFLP-cDNA display to screen plants of *T. miscellus*, *T. mirus*, and parental diploids to look for promising candidate genes—that is, fragments that did not show additivity in the allopolyploids as would be expected (Tate et al. 2006; Koh et al. 2010). Additional genes were surveyed because they were orthologous to genes that were singletons in other Asteraceae species (Buggs et al. 2009; Koh et al. 2010); the fate of such genes seemed of particular interest in new polyploids.

The results of these one-gene-at-a-time surveys are presented in detail elsewhere (Tate et al. 2006, 2009b; Buggs et al. 2009, 2010b; Koh et al. 2010). Significantly, most of the changes observed in populations of both young polyploids are homeolog loss events, which far outnumber gene-silencing events in all

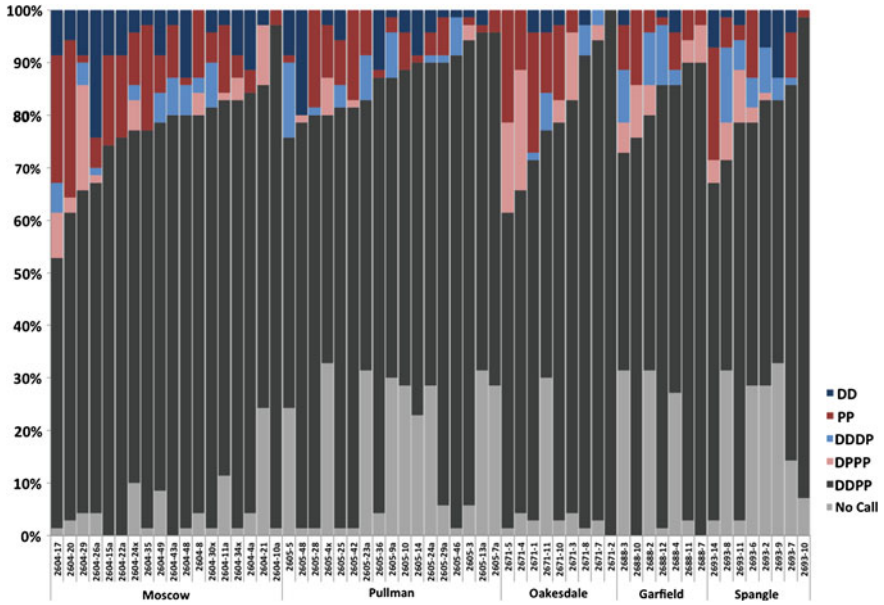

Fig. 14.3 Considerable variation from plant to plant within and among populations in the amount of homeolog loss detected in natural populations of *T. miscellus*. Population and plant designations are given on the *x*-axis; proportion of loci on the *y*-axis. Colors designate genotype as indicated in the figure. D = *T. dubius* allele; P = *T. pratensis* allele

populations examined of both polyploids. Furthermore, most of the homeolog losses in both polyploids were from *T. dubius*, the diploid parent that is shared by both *T. mirus* and *T. miscellus*.

It is also noteworthy that the same suite of genes consistently shows additivity of the parental gene copies (no loss or silencing) in polyploid populations of separate origin, whereas some of the genes analyzed consistently show some evidence of loss across at least some of the populations surveyed. Thus, these early analyses of *Tragopogon* polyploids are also in agreement with the hypothesis that there may be some underlying "principles" to polyploidization at the genetic or biochemical level.

But stochasticity is operating as well in these young polyploids. Although homeolog loss is present in the polyploid populations, the process is ongoing, and appears to be stochastic within individual populations. In no population examined has silencing or loss been complete (i.e., observed in all individuals of a population). The amount of gene loss within populations varies (Fig. 14.3). Furthermore, these losses and gene silencing events were not detected in F_1 hybrids or early-generation synthetic lines (S_1). Hence, loss of homeologs and gene silencing are not immediate consequences of hybridization or polyploidization in *Tragopogon*, but appear to occur several generations after polyploid formation and for certain genes.

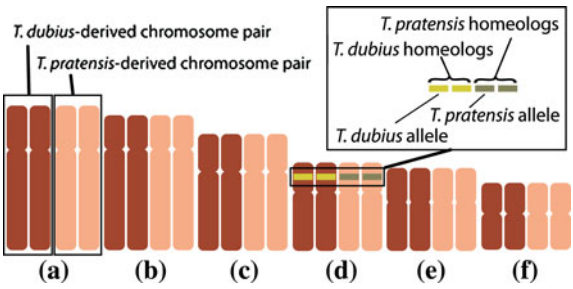

Fig. 14.4 Chromosome diagram for the alloetraploid *T. miscellus* (representative of other allotetraploids) illustrating terminology used throughout this chapter—chromosome, homeolog, and allele

14.2.2.2 *Tragopogon* Goes Genomic

Although important insights were obtained in *Tragopogon*, as well as other non-model systems, using the methods described above, these approaches have shortcomings. These analyses were based on examination of specific homeolog pairs using cleaved amplified polymorphic sequence analysis (CAPS). This approach uses restriction enzymes and hence has limitations, as discussed by Gaeta and Pires (2010). In addition, such surveys are slow and labor-intensive (reviewed in Soltis et al. 2009b; Buggs et al. 2009). Because of these limitations, we have sought ways to quickly and inexpensively build a framework for addressing genome-scale questions in *Tragopogon*.

Newly developed genomic resources quickly facilitated the use of *Tragopogon* as a model for the study of recent and repeated polyploidization. Using a combination of 454 and Illumina sequencing of genomic and cDNA, we identified SNPs (single-nucleotide polymorphisms) between the parental species *T. dubius* and *T. pratensis* for analysis of the allotetraploids (Buggs et al. 2010a, b). Thousands of SNPs distinguish the three parental diploids. Following SNP discovery, we designed primers for Sequenom analysis (see Buggs et al. 2010a). Briefly, Sequenom iPLEx genotyping uses mass spectrometry to carry out high-throughput and highly accurate genotyping with multiple SNPs multiplexed in one reaction (Jurinke et al. 2005). The assays permitted detection of gene loss and changes in allele number ('allele' is used here to designate alleles at both homologous and homeologous loci; Fig. 14.4) within polyploids versus changes in gene expression (see Sect. 14.2.3 below).

14.2.2.3 Genomic Insights

Using Sequenom genotyping, we examined patterns of homeolog presence/absence in 70 sets of homeologs in 59 plants from five independently formed populations of allotetraploid *T. miscellus* (Buggs et al. 2012). Extensive gene loss occurred in the $\leq$40 generations since polyploidization in *T. miscellus*. An average of $\sim$20 % of the 70 loci investigated in each plant of *T. miscellus* was missing one or both alleles of one homeolog (excluding assays where neither homeolog was

detected). Approximately one-third of these cases were single-allele absences. There was an overall bias toward loss of paternally derived homeologs. Assuming that immediately after polyploidization each individual had two alleles from *T. dubius* and two from *T. pratensis* at each pair of homeologous loci, on average at least 7.7 % of the original 280 allele copies have been lost in an individual plant's lineage since polyploidization occurred. In only one case did we find that absence of a homeolog was fixed in a population. The fact that few gene losses were fixed in populations, and many individuals were missing single alleles, indicates that allele loss is ongoing in these populations (Buggs et al. 2012).

These genomic data confirmed and extended our previous results and provided more insight into the processes involved. Certain loci were repeatedly missing in independently formed populations; the loci studied could be grouped into 12 clusters that followed recurrent patterns of presence/absence in populations with unique origins. Therefore, evolution is repeated in separate lineages of independent origin.

Homeolog loss is also repeated at deeper phylogenetic scales. We compared patterns of gene loss found in *T. miscellus* with patterns found in 12 other species of Asteraceae that are considered ancient polyploids (Barker et al. 2008). Eighteen genes in our study had GO categories that tended to be lost in Asteraceae (Barker et al. 2008), and these had median homeolog absence of 7.0 % in *T. miscellus*. A comparison of these 18 genes versus all other genes (i.e., those that tend to be retained in Asteraceae, plus those with no preferential loss or retention: 45 genes, with a median absence of 4.0 %) showed significantly higher homeolog absence in the 18 genes. Thus, gene loss in *T. miscellus*, a very young Asteraceae polyploid, appears to repeat patterns occurring in older Asteraceae polyploids.

One hypothesis for the repeated patterns of duplicate gene retention is the gene balance hypothesis (Freeling 2009; Freeling and Thomas 2006; Birchler et al. 2005; Papp et al. 2003). This hypothesis maintains that genes coding for products that are highly connected (i.e., within protein complexes or biochemical pathways) are dosage-sensitive in that they must be present in the nucleus in the same number of copies as the genes for products with which they interact. Thus, "connected" genes are hypothesized to be retained together as duplicate copies so as to preserve stoichiometry rather than reverting to singleton status one by one over time. In contrast, genes whose products are less connected are dosage-insensitive and are expected to revert gradually to singleton status. Our analyses so far suggest that dosage sensitivity may in fact be playing a role in *Tragopogon* (Buggs et al. 2012).

14.2.3 Tissue-Specific Silencing

Divergence of duplicate gene expression patterns among tissues has been suggested as a precursor of future evolution (Ohno 1970). Expression of a gene duplicate in a tissue where the progenitor copy was not expressed may indicate

neofunctionalization (Ohno 1970; Duarte et al. 2006; see Chap. 1 of this volume), while division of ancestral patterns of tissue-specific expression among duplicates suggests subfunctionalization (Lynch and Conery 2000; Rodin and Riggs 2003; Duarte et al. 2006) and silencing of a gene duplicate in all tissues points to nonfunctionalization (Duarte et al. 2006). Both neofunctionalization and subfunctionalization will lead to long-term retention of duplicated genes, whereas silencing/nonfunctionalization will generally lead to loss of a duplicate. Tissue-specific expression of duplicated genes has been studied in older gene duplicates in model organisms and crops (Adams et al. 2003; Duarte et al. 2006; Ganko et al. 2007; Semon and Wolfe 2008; Chaudhary et al. 2009), but in these species the ancestral patterns of gene expression and the ages of duplicates are not known precisely. Cases of tissue-specific expression patterns of very young gene duplicates are restricted to a few synthetic polyploids (Adams et al. 2003, 2004; Wang et al. 2004; Chaudhary et al. 2009).

We examined the expression of 13 homeolog pairs in seven tissues of 10 plants of *T. mirus* from two natural populations of independent origin (Buggs et al. 2010b). Of the 910 assays in *T. mirus*, 63 % showed expression of both homeologs, 7 % showed no expression of either homeolog, 20 % showed non-expression of one homeolog across all tissues of a plant, and 8 % showed non-expression of a homeolog in a particular tissue within a plant. We found two cases of reciprocal tissue-specific expression between homeologs, potentially indicative of subfunctionalization. This study therefore showed that tissue-specific silencing, and even apparent subfunctionalization, can arise rapidly in the early generations of natural allopolyploidy. Similar results were found for 18 homeolog pairs using the same approaches in *T. miscellus* populations (Buggs et al. 2011b).

In *T. miscellus* we also examined tissue-specific gene expression using the previously described Sequenom assays (Buggs et al. 2011b). Tissue-specific expression of 144 homeolog pairs in two natural populations was compared with patterns of allelic expression in both in vitro ''hybrids'' and hand-crossed F_1 hybrids between the parental diploids *T. dubius* and *T. pratensis*, and with patterns of homeolog expression in synthetic (S_1) allotetraploids. Tissue-specific homeolog expression was frequent in natural allopolyploids, but F_1 hybrids and S_1 allopolyploids showed less tissue-specific homeolog expression than the natural allopolyploids and the in vitro ''hybrids'' of diploid parents. These results suggest that "transcriptomic shock" upon hybridization (McClintock 1984) includes the activation of allele/homeolog expression in all tissues, causing a loss of tissue-specific expression patterns seen in the diploid parents. Such activation has seldom been considered in terms of the tissue-specific activation of protein-coding genes. Activation of homeologs has also been found in cotton F_1 hybrids and allopolyploids (Chaudhary et al. 2009), who termed it "transcriptional neofunctionalization". We showed this to be widespread in *Tragopogon*. This may fit a scenario in which activity of small interfering RNA molecules, which influence gene expression, is temporarily lost in F_1 hybrids and early allopolyploids, but restored subsequently.

14.2.4 Cytogenetic Insights

Despite enormous progress in our understanding of many aspects of polyploidy, little attention has been paid to chromosomal constitution, structure, and organization. The New World *Tragopogon* allotetraploids illustrate the valuable role of cytology in examining hybridization and polyploidy in plants. Our recent studies have revealed high levels of chromosomal variation in ~80-year-old natural populations, as well as the newly synthesized allotetraploids *T. mirus* and *T. miscellus*. The earliest studies of *T. mirus* and *T. miscellus* identified mitotic complements of $2n = 24$ and typically 12 bivalents at meiosis (Ownbey 1950; Ownbey and McCollum 1954). This supported the expectation that allotetraploids were chromosomally additive of the diploid progenitors, which are both $2n = 12$.

Lim et al. (2008) used genomic and fluorescence in situ hybridization (GISH/FISH) in a preliminary survey of several *T. mirus* and *T. miscellus* plants and found a few plants of each species that were not chromosomally additive of the diploid parents. Several natural *T. mirus* and *T. miscellus* plants were found to be aneuploid, with intergenomic translocations. Although most of the aneuploid individuals examined were $2n = 24$, not all chromosomes were present in two copies as expected. GISH conducted on synthetic *T. mirus* pollen mother cells at the diplotene stage of meiosis showed allosyndetic pairing within multivalents. Meiotic instability, in the form of anaphase bridges and lagging chromosomes, was also observed by Tate et al. (2009a) in the first synthetic polyploid generation (S_1) in *T. mirus* and *T. miscellus* lines. Thus, gametes with aneuploid and/or rearranged chromosome complements can be potentially generated as early as the first meiosis following genome doubling, which has been shown to be the case for synthetic allotetraploid *B. napus* (Szadkowski et al. 2010).

The preliminary work of Lim et al. (2008) prompted a detailed examination using GISH and FISH of the chromosomal variation generated in six *T. miscellus* populations of independent origin (Chester et al. 2012). In all six populations, both aneuploidy and translocations were common (Fig. 14.5). Only 3 of the 58 plants exhibited the expected additivity of the diploid parental karyotypes (with neither aneuploidy nor translocations). Although approximately 70 % of polyploid plants were aneuploid for one or more chromosomes, variation in copy number appears to be constrained. Most plants were $2n = 24$, and the total copy number for each homeologous group of chromosomes was typically four as a result of aneuploidy being reciprocal between homeologous chromosomes (Fig. 14.5). Thus, most deviations from disomy were in the form of monosomy-trisomy or nullisomy-tetrasomy, between homeologous chromosomes. This pattern of extensive aneuploidy while maintaining the overall copy number closely resembles cytological changes in synthetic neoallotetraploid *Brassica napus* (Xiong et al. 2011). Gene dosage has been implicated as a major factor constraining chromosomal changes such that imbalances, which arise, require compensation by chromosomes (compensatory aneuploidy) or homeologous segments (compensatory translocations).

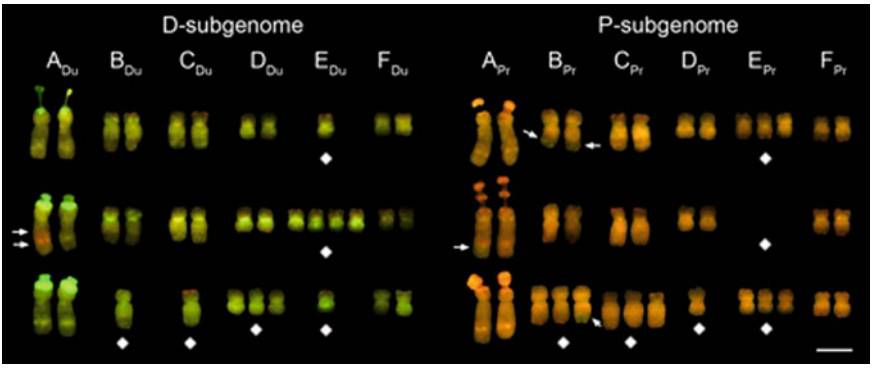

Fig. 14.5 Mitotic GISH karyotypes of three aneuploid *T. miscellus* individuals. Chromosomes derived from each parental genome are shown, i.e., the *T. dubius* genome (D-subgenome) and the *T. pratensis* genome (P-subgenome). Chromosomes deviating from disomy are as follows: *top*, monosomy: trisomy for E chromosomes; *middle*, nullisomy: tetrasomy for E chromosomes; *bottom*, four cases of monosomy: trisomy for chromosomes B, C, D, and E. All three individuals are $2n = 24$ and have four copies of each homeologous group (A–F). Scale bar: 5 μm

In *T. miscellus*, genome rearrangements via intergenomic translocations were common; 76 % of the plants showed evidence of at least one translocation. The two largest chromosomes (groups A and B) showed the highest incidence of translocations. Across all *T. miscellus* plants at least six different translocation breakpoint positions were observed along the A group chromosomes. In one *T. miscellus* population in Spokane, WA, an A-chromosome intergenomic translocation appeared close to fixation, being homozygous in 8 of the 10 individuals analyzed. Individuals from six populations also appear to have undergone reciprocal translocations on the B-group chromosomes, but they mostly occur at a similar position, near the end of the long arm.

The observed chromosomal changes in the recently formed *Tragopogon* allotetraploids provide two possible explanations for the loss of DNA that has been observed using molecular SNP-based assays. (1) If a translocation was in a non-reciprocal state and homozygous, this could lead to the loss of DNA from one of the parental diploids in the translocated region. (2) Nullisomy would lead to the complete loss of DNA from one of the parental chromosomes. This may have contributed to the clustering of patterns of gene loss and retention found using Sequenom assays (see above). However, the diverse patterns of homeolog losses that have been detected are not readily explainable by only these large-scale chromosomal changes detectable with GISH.

14.3 Comparing *Tragopogon* to Other Well-Studied Systems

Genome evolution in other well-studied polyploid systems (e.g., *Gossypium*, *Triticum*, *Nicotiana*, *Arabidopsis*, *Brassica*, *Senecio*, *Spartina*) exhibits important similarities and differences. *Tragopogon* is noteworthy in that initial studies show

that while true expression changes play a major role, homeolog losses are also very prominent in these young polyploids (Tate et al. 2006, 2009b; Koh et al. 2010; Buggs et al. 2009, 2010a, b, 2011b, 2012). Of the initial 23 genes analyzed in *T. miscellus*, 15 showed homeolog loss in one or more plants from nature, and 8 showed true expression changes (results for *T. mirus* are comparable); these patterns were confirmed via genomic analyses of many more genes. In contrast, in synthetic wheat (*Triticum*) and synthetic *Arabidopsis thaliana* and *A. suecica* polyploids, as well as in tetraploid cotton (*Gossypium*), expression changes dominate (Adams et al. 2003, 2004; Hovav et al. 2008a, b; Udall et al. 2006; Flagel et al. 2008; Madlung et al. 2004; Kashkush et al. 2002; Wang et al. 2006).

Tragopogon may be most similar to the allotetraploid *B. napus*, in which most of the apparent gene silencing events observed in later generations were due to gene losses, most likely resulting from genomic rearrangements (Song et al. 1995; Gaeta et al. 2007). Across 50 lines of *B. napus*, genetic changes are equally distributed between the parental diploid genomes (Gaeta et al. 2007). The system is dynamic—some lines become more "*oleracea* like" and others more "*rapa* like" in terms of losses and corresponding expression differences. In *Arabidopsis suecica* allopolyploids, silencing of homeologs from one parent (*A. thaliana*) was observed more frequently than the silencing of homeologs from the other parent, *Arabidopsis arenosa* (Wang et al. 2006). In synthetic polyploids in *Triticale*, the contribution of the rye genome (*Secale cereale*) is preferentially silenced (Ma et al. 2004; Ma et al. 2006). In cotton, tissue-specific subfunctionalization occurs for some loci (Adams et al. 2003, 2004; Adams and Wendel 2004), but overall gene expression is biased toward one parent (Udall et al. 2006; Flagel et al. 2008; Rapp et al. 2009; Flagel and Wendel 2010). In both *T. miscellus* and *T. mirus*, homeologs of one diploid genome (*T. dubius*) are more often lost or not expressed; that is, *T. dubius* is often the "loser genome" based on the set of genes surveyed to date.

Tragopogon also exhibits cytogenetic similarities to *Brassica*. Both *T. mirus* and *T. miscellus* exhibit numerous translocations as well as extensive aneuploidy while maintaining the overall copy number expected in an allotetraploid (reciprocal monosomy: trisomy and nullisomy: tetrasomy); this closely resembles cytological changes in synthetic neoallotetraploid *B. napus* (Xiong et al. 2011).

It is also now possible to compare results for *Tragopogon* with several of the other recently formed natural polyploids. Great research progress has now been made on four of the five polyploids known to have formed in the past two centuries. In addition to *T. mirus* and *T. miscellus*, numerous insights into recent polyploidy have been obtained for *Senecio cambrensis* (also in the Asteraceae) and *Spartina anglica* (Poaceae). Both these model systems are reviewed elsewhere in this volume (see Chaps. 13 and 12, respectively). Interestingly, all of these recent polyploids have formed following introductions of one or both progenitors into a completely new geographic area. Some of these new polyploids have been successful in nature, particularly *Spartina anglica*, which is now distributed worldwide with major ecological impact (Ainouche et al. 2004, 2009; see Chap. 12 of this volume). The *Tragopogon* polyploids are now major weeds in the Palouse region of northwestern USA. In contrast, following initial range expansion,

Senecio cambrensis now seems to be disappearing from parts of its original range (Abbott et al. 2007; see Chap. 13 of this volume).

Senecio cambrensis formed from parents of different ploidal levels (one parent is itself a polyploid), whereas in *Spartina anglica* both parents are polyploids but of the same ploidal level. In *Senecio cambrensis* there is an additional level of complexity in that the diploid parent is itself of homoploid hybrid origin. In contrast, both parents of the recent *Tragopogon* polyploids are diploid. The multiple layers of recent hybridization and polyploidization in *Senecio* and *Spartina* could influence the genetic and expression changes detected in the recent polyploids in those genera.

Molecular studies have revealed striking genetic and genomic changes in all of these recently formed polyploids. Transcriptomic shock has now been shown following hybridization in both *Senecio* (Hegarty et al. 2005, 2006) and *Tragopogon* (Buggs et al. 2011b), and substantial expression changes have also been reported in F_1 hybrids and polyploids in *Spartina* (Cheilafa et al. 2010a, b) Thus, in recent polyploids in all three genera, hybridization seems to result in a major burst of altered gene expression.

Changes in gene expression have been shown to be important in all of the systems, but have been best characterized in *Spartina* (Ainouche et al. 2009, Chap. 12 of this volume) and *Senecio* (Hegarty et al. 2008, Chap. 13 of this volume). Methylation alterations have been detected in both *Spartina* (Salmon et al. 2005) and *Senecio* (Hegarty et al. 2011), with epigenetic changes now considered to play a major role in these new polyploids (Parisod et al. 2009; Ainouche et al. 2009; Hegarty et al. 2011). However, epigenetic changes have not yet been analyzed in *Tragopogon*. Dramatic changes in gene expression have been documented in *Spartina anglica*, *Senecio cambrensis*, and the recently formed *Tragopogon* polyploids. However, homeolog loss seems to have played a more prominent role in young *Tragopogon* polyploids than in these other recent polyploids.

In both the recently formed *Tragopogon* polyploids and in *S. cambrensis*, genomic studies reveal that one parental genome predominates over the other in terms of global patterns of expression and gene retention. In the allohexaploid *S. cambrensis*, gene expression is more similar to that of the tetraploid parent (*S. vulgaris*) than to the diploid parent, *S. squalidus* (Hegarty et al. 2006). In the young *Tragopogon* polyploids, homeologs of the diploid parent *T. dubius* are more often lost or not expressed. That is, following polyploidy there are clear winner and loser parental genomes, and these patterns are established quickly and repeatedly.

Chromosomal changes have been investigated in detail in both *Tragopogon* and *Spartina* using FISH/GISH; recent polyploids in both genera exhibit substantial change, but different types of variation are evident. In *Spartina*, nonaploid plants ($2n$ = ca. 90) were detected (rather than the expected $2n$ = 120–124), most likely resulting from backcrosses between *S. anglica* and its maternal parent *S. alterniflora* (Renny-Byfield et al. 2010). In *Tragopogon*, frequent translocations were detected as well as frequent reciprocal monosomy: trisomy and nullisomy: tetrasomy—these have not been previously reported from nature (Chester et al. 2012). Chromosomal pairing abnormalities (multivalents) have been detected in both *Tragopogon*

allopolyploids as well as in synthetic polyploids (Ownbey and McCollum 1953; Tate et al. 2009a). FISH/GISH has been problematic in *Senecio*, but traditional cytogenetic analysis suggests possible structural chromosomal changes, as well as some meiotic pairing irregularities (Crisp 1972). In addition to generating genetic diversity, intergenomic recombination resulting in chromosome rearrangements could lead to the formation of reproductive barriers between lineages of separate or independent origin, which is under investigation in *Tragopogon*.

14.4 Conclusions

A diverse array of experimental approaches has helped us to understand the evolution of young, natural *Tragopogon* polyploids at the genomic and transcriptomic levels. These have demonstrated: (1) the allopolyploids have multiple origins; (2) substantial gene loss occurs within the first 40 generations in nature, and is still ongoing; (3) patterns of gene loss are repeated among independent origins of the allopolyploids; (4) transcriptomic shock occurs upon allopolyploidization, involving a reduction in tissue-specific expression; (5) major chromosomal changes have occurred and are ongoing in the natural allopolyploids.

The dynamic evolutionary processes that appear to be underway in *Tragopogon* polyploids may be representative of those that have occurred in other groups, and after past polyploidization events. Care is needed when extrapolating from a single system, because patterns of change may be influenced by the genetic background of the polyploid, such as the level of genetic differentiation between the two parental species of an allopolyploid (reviewed in

Buggs et al. 2011a). Studies of polyploids in many plant families, of different ages and of different parental combinations, are needed to provide a comprehensive understanding of the evolution of allopolyploids.

Tragopogon has great potential for further development as a model system for polyploidy. Older Eurasian polyploids will allow placement of another time point in the evolutionary series. Research is also underway on natural hybrids between *T. pratensis* and *T. porrifolius* (*T.* × *mirabilis*) that occur in the United Kingdom. Further development of genetic resources—sequencing of the *T. dubius* genome and production of a genetic map—will allow new questions to be addressed.

In recent years, progress in understanding the genetic consequences of polyploidy has outstripped progress in understanding the ecological background in which evolution occurs (but see Ramsey 2011). As a genus that occurs almost exclusively in the wild, *Tragopogon* affords a system in which relevant ecological studies can be carried out, comparing the fitness of polyploids and diploids, and the different chromosomal variants of the polyploids. Such studies will contribute to a comprehensive view of polyploid evolution.

Acknowledgments Funding for this research was provided by the University of Florida and NSF grants MCB-0346437, DEB-0614421, DEB-0919254, DEB-0922003, and DEB-0919348. R.J.A.B. has been supported since March 2010 by NERC Fellowship NE/G01504X/1.

References

Abbott RJ, Ireland HE, Rogers HJ (2007) Population decline despite high genetic diversity in the new allopolyploid species *Senecio cambrensis* (Asteraceae). Mol Ecol 16:1023–1033

Abbott RJ, Lowe AJ (2004) Origins, establishment, and evolution of new polyploids species: *Senecio cambrensis* and *S. eboracensis* in the British Isles. Biol J Linn Soc 82:467–474

Adams KL, Cronn R, Percifield R, Wendel JF (2003) Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proc Nat Acad Sci USA 100:4649–4654

Adams KL, Percifield R, Wendel JF (2004) Organ-specific silencing of duplicated genes in a newly synthesized cotton allotetraploid. Genetics 168:2217–2226

Adams KL, Wendel JF (2004) Exploring the genomic mysteries of polyploidy in cotton. Biol J Linn Soc 82:573–581

Ainouche ML, Baumel A, Salmon A (2004) *Spartina anglia* C. E. Hubbard: a natural model system for analyzing early evolutionary changes that affect allopolyploid genomes. Biol J Linn Soc 82:475–484

Ainouche ML, Fortune PM, Salmon A, Parisod C, Grandbastien M-A, Ricou K, Fukunaga M, Misset M-T (2009) Hybridization, polyploidy and invasion: Lessons from *Spartina* (Poaceae). Biol Invasion. doi:10.1007s10530-0089383-2

Ashton PA, Abbott RJ (1992) Multiple origins and genetic diversity in the newly arisen allopolyploid species *Senecio cambrensis* Rosser (Compositae). Heredity 68:25–32

Barker MS, Kane NC, Matvienko M, Kozik A, Michelmore RW, Knapp SJ, Rieseberg LH (2008) Multiple Paleopolyploidizations during the evolution of the compositae reveal parallel patterns of duplicate gene retention after millions of years. Mol Biol Evol 25:2445–2455

Birchler JA, Riddle NC, Auger DL, Veitia R (2005) Dosage balance in gene regulation: biological implications. Trends Genet 21:219–226

Blanc G, Wolfe KH (2004) Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* divergence. Plant Cell 16:1679–1691

Blanca G, Díaz de la Guardia C (1996) Sinopsis del género *Tragopogon* L. (Asteraceae) en la Peninsula Ibérica. Anales del Jardín Botánico de Madrid 54:358–363

Buggs RJA, Doust AN, Tate JA, Koh J, Soltis K, Feltus FA, Paterson AH, Soltis PS, Soltis DE (2009) Gene loss and silencing in *Tragopogon miscellus* (Asteraceae): comparison of natural and synthetic allotetraploids. Heredity 103:73–81

Buggs RJA, Chamala S, Wu W, Gao L, May GD, Schnable PS, Soltis DE, Soltis PS, Barbazuk WB (2010a) Characterization of duplicate gene evolution in the recent natural allopolyploid *Tragopogon miscellus* by next-generation sequencing and Sequenom iPLEX genotyping. Mol Ecol 19(1):1–15

Buggs RJA, Elliott NM, Zhang L, Koh J, Viccini LF, Soltis DE, Soltis PS (2010b) Tissue-specific silencing of homoeologs in natural populations of the recent allopolyploid *Tragopogon mirus*. New Phytol 186:175–183

Buggs RJA, Soltis PS, Soltis DE (2011a) Biosystematic relationships and the formation of polyploids. Taxon 60:324–332

Buggs RJA, Zhang L, Miles N, Tate JA, Gao L, Schnable PS, Barbazuk WB, Soltis PS, Soltis DE (2011b) Genomic and transcriptomic shock generate evolutionary novelty in a newly formed, natural allopolyploid plant. Curr Biol 21:1–6

Buggs RJA, Gao L, Wu W, Chamala S, Tate JA, Schnable PS, Soltis DE, Soltis PS, Barbazuk WB (2012) Rapid and repeated gene loss in a young polyploidy species. Curr Biol 22:248–252

Chaudhary B, Flagel L, Stupar RM, Udall JA, Verma N, Springer NM, Wendel JF (2009) Reciprocal silencing, transcriptional bias and functional divergence of homoeologs in polyploid cotton (*Gossypium*). Genetics 182:503–517

Chelaifa H, Mahe F, Ainouche M (2010a) Transcriptome divergence between the hexaploid salt-marsh sister species *Spartina maritima* and *Spartina alterniflora* (Poaceae). Mol Ecol 19:2050–2063

Chelaifa H, Monnier A, Ainouche M (2010b) Transcriptomic changes following recent natural hybridization and allopolyploidy in the salt marsh species *Spartina* x *townsendii* and *Spartina anglica* (Poaceae). New Phytol 186:161–174

Chen ZJ, Wang J, Tian L, Lee HS, Wang JJ, Chen M, Lee JJ, Josefsson C, Madlung A, Watson B, Pires JC, Lippman Z, Vaughn M, Colot V, Birchler JA, Doerge RW, Martienssen RA, Comai L, Osborn TC (2004) The development of an *Arabidopsis* model system for genome-wide analysis of polyploidy effects. Biol J Linn Soc 82:689–700

Chester M, Gallagher JP, Symonds VV, da Veruska Cruz Silva A, Mavrodiev EV, Leitch AR, Soltis PS, Soltis DE (2012) Extensive and repeated patterns of chromosomal variation in natural populations of a recently formed polyploid plant species. Proc Nat Acad Sci USA 109:1176–1181

Clausen J, Keck DD, Hiesey WM. 1945. Experimental studies on the nature of species II. Plant evolution through amphiploidy and autopolyploidy, with examples from the Madiinae. Publication 564, Carnegie Institute of Washington, Washington, DC

Crisp PC (1972) *Cytotaxonomic studies in the section* Annui *of* Senecio. Ph. D Thesis, University of London

Darlington CD (1937) Recent advances in cytology, 2nd edn. The Blakiston Company, Philadelphia

Duarte JM, Cui L, Wall PK, Zhang Q, Zhang X, Leebens-Mack J, Ma H, Altman N, dePamphilis CW (2006) Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of Arabidopsis. Mol Biol Evol 23:469–478

Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, De Montigny J, Marck C, Neuvéglise C, Talla E et al (2004) Genome evolution in yeasts. Nature 430:35–44

Flagel L, Udall J, Nettleton D, Wendel J (2008) Duplicate gene expression in allopolyploid *Gossypium* reveals two temporally distinct phases of expression evolution. BMC Biol 6:11

Flagel LE, Wendel JF (2010) Evolutionary rate variation, genomic dominance and duplicate gene expression during allotetraploid cotton speciation. New Phytol 186:184–193

Freeling M (2009) Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. Annu Rev Plant Biol 60:433–453

Freeling M, Thomas BC (2006) Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. Genome Res 16:805–814

Gaeta RT, Pires JC (2010) Homoeologous recombination in allopolyploids: the polyploid ratchet. New Phytol 186:18–28

Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC (2007) Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. Plant Cell 19:3403–3417

Ganko EW, Meyers BC, Vision TJ (2007) Divergence in expression between duplicated genes in *Arabidopsis*. Mol Biol Evol 24:2298–2309

Gould SJ (1994) The evolution of life on Earth. Sci Am 271:85–86

Gregory TR, Mable BK (2005) Polyploidy in animals. In: Gregory TR (ed) The evolution of the Genome. Elsevier/Academic, San Diego, pp 428–501

Hegarty MJ, Jones JM, Wilson ID, Barker GL, Coghill JA, Sanchez-Baracaldo P, Liu G, Buggs RJA, Abbott RJ, Edwards KJ, Hiscock SJ (2005) Development of anonymous cDNA microarrays to study changes to the *Senecio* floral transcriptome during hybrid speciation. Mol Ecol 14:2493–2510

Hegarty MJ, Barker GL, Wilson ID, Abbott RJ, Edwards KJ, Hiscock SJ (2006) Transcriptome shock after interspecific hybridization in *Senecio* is ameliorated by genome duplication. Curr Biol 16:1652–1659

Hegarty MJ, Barker GL, Brennan AC, Edwards KJ, Abbott RJ, Hiscock SJ (2008) Changes to gene expression associated with hybrid speciation in plants: further insights from transcriptomic studies in *Senecio*. Philos Trans R Soc London B series 363:3055–3069

Hegarty MJ, Batstone T, Barker GL, Edwards KJ, Abbott RJ, Hiscock SJ (2011) Nonadditive changes to cytosine methylation as a consequence of hybridization and genome duplication in *Senecio* (Asteraceae). Mol Ecol 20:105–113

Hovav R, Udall J, Chaudhary B, Flagel L, Rapp R, Wendel J (2008a) Partitioned expression of duplicated genes during development and evolution of a single cell in a polyploid plant. Proc Nat Acad Sci USA 105:6191

Hovav R, Udall JA, Chaudhary B, Hovav E, Flagel L, Hu G, Wendel JF (2008b) The evolution of spinnable cotton fiber entailed prolonged development and a novel metabolism. PLoS Genet 4:e2

Jiao Y, Wickett N, Ayyampalayam S, Chanderbali A, Landherr L, Ralph PE, Soltis PS, Soltis DE, Clifton SE, Ma H, Leebens-Mack J, dePamphilis CW (2011) Phylogenomic analysis reveals ancient genome duplications in seed plant and angiosperm history. Nature 473:97–100

Jurinke C, Denissenko MF, Oeth P, Ehrich M, van den Boom D, Cantor CR (2005) A single nucleotide polymorphism based approach for the identification and characterization of gene expression modulation using MassARRAY. Mutat Res 573:83–95

Kashkush K, Feldman M, Levy AA (2002) Gene loss, silencing, and activation in a newly synthesized wheat allotetraploid. Genetics 160:1651–1659

Kellis M, Birren BW, Lander ES (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. Nature 428:617–624

Koh J, Tate JA, Soltis PS, Soltis DE (2010) Genomic and expression differences in natural populations of the recently formed allotetraploid *Tragopogon mirus* (Asteraceae). BMC Genomics 11:97

Kovarik A, Pires JC, Leitch AR, Lim KY, Sherwood A, Matyasek R, Rocca J, Soltis DE, Soltis PS (2005) Rapid concerted evolution in two allopolyploids of recent and recurrent origin. Genetics 169:931–944

Levin DA (1983) Polyploidy and novelty in flowering plants. Am Nat 122:1–25

Levin DA (2002) The role of chromosomal change in plant evolution. Oxford University Press, New York

Lim KY, Soltis DE, Soltis PS, Tate JA, Matyasek R, Srubarova H, Kovarik A, Pires JC, Xiong ZY, Leitch AR (2008) Rapid chromosome evolution in recently formed polyploids in *Tragopogon* (Asteraceae). PLoS One 3:e3353

Löve A, Löve D (1949) The geobotanical significance of polyploidy. I. Polyploidy and latitude. Portugaliae Acta Biologica Serie A, Suppl vol. pp 273–352

Lynch M, Connery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290:1151–1155

Ma XF, Fang P, Gustafson JP (2004) Polyploidization-induced genome variation in *Triticale*. Genome 47:839–848

Ma XF, Fang P, Gustafson JP (2006) Timing and rate of genome variation in *Triticale* following allopolyploidization. Genome 49:950–958

Mable B (2003) Breaking down taxonomic barriers in polyploidy research. Trends Plant Sci 8:582–590

Mable BK, Alexandrou MA, Taylor MI (2011) Genome duplications in amphibians and fish: an extended synthesis. J Zool 284:151–182

Madlung A, Comai L (2004) The effect of stress on genome regulation and structure. Ann Bot 94:481 495

Malinska H, Tate JA, Matyasek R, Leitch AR, Soltis DE, Soltis PS, Kovarik A (2010) Similar patterns of rDNA evolution in synthetic and recently formed natural populations of *Tragopogon* (Asteraceae) allotetraploids. BMC Evol Biol 10:291

Malinska H, Tate JA, Mavrodiev E, Matyasek R, Lim KY, Leitch AR, Soltis DE, Soltis PS, Kovarik A (2011) Ribosomal RNA genes evolution in *Tragopogon*: A story of New and Old World allotetraploids and synthetic lines. Taxon 60:348–354

Matyasek R, Tate J, Lim YK, Srubaraova H, Koh J, Leitch A, Soltis DE, Soltis PS (2007) Concerted evolution of rDNA in recently formed *Tragopogon* allotetraploids is typically associated with an inverse correlation between gene copy number and expression. Genetics 176:2509–2519

Mavrodiev E, Soltis PS, Soltis DE (2008a) Parentage of six Old World polyploids in *Tragopogon* L. (Asteraceae: Scorzonerinae) based on ITS. ETS and plastid sequence data. Taxon 57:1217–1232

Mavrodiev E, Nawchoo I, Soltis DE, Soltis PS (2008b) Molecular data reveal that the tetraploid *Tragopogon kashmirianus* Singh, a narrow endemic of Kashmir, is distinct from the North American *T. mirus* M. Ownbey. Bot J Linn Soc 158:391–398

Mavrodiev EV, Albach DC, Speranza P (2008c) A new polyploid species of the genus *Tragopogon* (Asteraceae, Cichorieae) from Russia. Novon 18:229–232

McClintock B (1984) The significance of responses of the genome to challenge. Science 226:792–801

Morris SC (1998) The crucible of creation: the Burgess shale and the rise of animals. Oxford University Press, Oxford

Müntzing A (1936) The evolutionary significance of autopolyploidy. Hereditas 21:263–378

Novak SJ, Soltis DE, Soltis PS (1991) Ownbey *Tragopogons*—40 Years later. Am J Bot 78:1586–1600

Ohno S (1970) Evolution by gene duplication. Springer, Berlin

Ownbey M (1950) Natural hybridization and amphiploidy in the genus *Tragopogon*. Am J Bot 37:487–499

Ownbey M, McCollum G (1954) The chromosomes of *Tragopogon*. Rhodora 56:7–21

Ownbey M, McCollum GD (1953) Cytoplasmic inheritance and reciprocal amphiploidy in Tragopogon. Am J Bot 40:788–796

Panopoulou G, Poustka AJ (2005) Timing and mechanism of ancient vertebrate genome duplications—the adventure of a hypothesis. Trends Genet 10:559–567

Papp B, Pal C, Hurst LD (2003) Dosage sensitivity and the evolution of gene families in yeast. Nature 424:194–197

Parisod C, Salmon A, Zerjal T, Tenaillon M, Grandbastien M-A, Ainouche M (2009) Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in *Spartina*. New Phytol 184:1003–1015

Paterson AH, Chapman BA, Kissinger J, Bowers JE, Feltus FA, Estill JC, Marler BS (2006) Many gene and domain families have convergent fates following independent whole-genome duplication events in *Arabidopsis*, *Oryza*, *Saccharomyces* and *Tetraodon*. Trends Genet 22:597–602

Petit RJ, Aguinagalde, JL de Beaulieu, C Bittkau, S Brewer, R Cheddadi, R Ennos, S Fineschi, D Grivet, M Lascoux, A Mohanty, G Müller-Starck, B Musch, A Palmé, S Rendell, GG. Vendramin (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. Science 300:1563–1565

Ramsey J (2011) Polyploidy and ecological adaptation in wild yarrow. Proc Nat Acad Sci USA 108:6697–6698

Rapp RA, Udall JA, Wendel JF (2009) Genomic expression dominance in allopolyploids. BMC Biol. 7:18

Renny-Byfield S, Ainouche M, Leitch IJ, Lim KY, Le Comber SC, Leitch AR (2010) Flow cytometry and GISH reveal mixed ploidy populations and *Spartina* nonaploids with genomes of *S. alterniflora* and *S. maritima* origin. Ann Bot 105:527–533

Rodin SN, Riggs AD (2003) Epigenetic silencing may aid evolution by gene duplication. J Mol Evol 56:718–729

Salmon A, Ainouche ML, Wendel JF (2005) Genetic and epigenetic consequences of recent hybridization and polyploidy in *Spartina* (Poaceae). Mol Ecol 14:1163–1175

Semon M, Wolfe KH (2008) Preferential subfunctionalization of slow-evolving genes after allopolyploidization in *Xenopus laevis*. Proc Nat Acad Sci USA 105:8333–8338

Soltis DE, Soltis PS (1995) The dynamic nature of polyploid genomes. Proc Nat Acad Sci USA 92:8089–8091

Soltis DE, Soltis PS, Pires JC, Kovarik A, Tate JA, Mavrodiev E (2004) Recent and recurrent polyploidy in *Tragopogon* (Asteraceae): cytogenetic, genomic and genetic comparisons. Biol J Linn Soc 82:485–501

Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson A, Zheng C, Sankoff D, Wall PK, Soltis PS (2009a) Polyploidy and angiosperm diversification. Am J Bot 96:336–348

Soltis DE, Buggs RJA, Barbazuk WB, Schnable PS, Soltis PS (2009b) On the origins of species: does evolution repeat itself in polyploid populations of independent origin? Cold spring harbor symposia on quantitative biology, Vol. LXXIV

Soltis DE, Mavrodiev EV, Meyers, SC, Severns PM, Zhang L, Gitzendanner MA, Ayers T, Chester M, Soltis PS (2012) Additional origins of Ownbey's *Tragopogon mirus*. Bot Linn Soc 169:297–311

Soltis DE, Soltis PS (1989) Allopolyploid speciation in Tragopogon: Insights from chloroplast DNA. Am J Bot. 76:1119–1124

Soltis DE, Soltis PS (1993) Molecular data and the dynamic nature of polyploidy. Crit Rev Plant Sci 12:243–273

Soltis DE, Soltis PS (1999) Polyploidy: recurrent formation and genome evolution. Trends Ecol Evol 14:348–352

Soltis PS, Soltis DE (2000) The role of genetic and genomic attributes in the success of polyploids. Proc Nat Acad Sci (USA) 97:7051–7057

Soltis PS, Soltis DE (2009) The role of hybridization in plant speciation. Annu Rev Plant Biol 60:561–588

Soltis PS, Plunkett GM, Novak SJ, Soltis DE (1995) Genetic variation in *Tragopogon* species: additional origins of the allotetraploids *T. mirus* and *T. miscellus* (Compositae). Am J Bot 82:1329–1341

Song KM, Lu P, Tang KL, Osborn TC (1995) Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. Proc Nat Acad Sci USA 92:7719–7723

Stebbins GL (1950) Variation and evolution in plants. Columbia, New York

Stebbins GL (1971) Chromosomal evolution in higher plants. Addison-Wesley, London

Stern DL, Orgogozo V (2009) Is genetic evolution predictable? Science 323:746–751

Symonds VV, Soltis PS, Soltis DE (2010) Dynamics of polyploid formation in *Tragopogon* (Asteraceae): recurrent formation, gene flow, and population structure. Evolution 64:1984–2003

Szadkowski E, Eber F, Huteau V, Lod M, Huneau C, Belcram H, Coriton O, Manzanares-Dauleux M, Delourme R, King G (2010) The first meiosis of resynthesized *Brassica napus*, a genome blender. New Phytol 186:102–112

Tate JA, Ni Z, Scheen AC, Koh J, Gilbert CA, Lefkowitz D, Chen ZJ, Soltis PS, Soltis DE (2006) Evolution and expression of homoeologous loci in *Tragopogon miscellus* (Asteraceae), a recent and reciprocally formed allopolyploid. Genetics 173:1599–1611

Tate JA, Symonds VV, Doust AN, Buggs RJA, Mavrodiev EV, Soltis PS, Soltis DE (2009a) Synthetic polyploids of *Tragopogon miscellus* and *T. mirus* (Asteraceae): 50 + years after Ownbey's discovery. Am J Bot 96:979–988

Tate JA, Joshi P, Soltis K, Soltis PS, Soltis DE (2009b) On the road to diploidization? Homoeolog loss in independently formed populations of the allopolyploid *Tragopogon miscellus* (Asteraceae). BMC Plant Biol 9:80

Udall JA, Swanson JM, Nettleton D, Percifield RJ, Wendel JF (2006) A novel approach for characterizing expression levels of genes duplicated by polyploidy. Genetics 173(3):1823–1827

Xiong Z, Gaeta RT, Pires JC (2011) Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. Proc Nat Acad Sci USA 108:7908–7913

Wang JL, Tian L, Madlung A, Lee HS, Chen M, Lee JJ, Watson B, Kagochi T, Comai L, Chen ZJ (2004) Stochastic and epigenetic changes of gene expression in Arabidopsis polyploids. Genetics 167:1961–1973

Wang JL, Tian L, Lee HS, Wei NE, Jiang HM, Watson B, Madlung A, Osborn TC, Doerge RW, Comai L, Chen ZJ (2006) Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. Genetics 172:507–517

Wolfe KH, Shields DC (1997) Molecular evidence for an ancient duplication of the entire yeast genome. Nature 387:708–713

Zimmer EA, Martin SL, Beverley SM, Kan YW, Wilson AC (1980) Rapid duplication and loss of genes coding for the ∂ chains of hemoglobin. Proc Nat Acad Sci USA 77:2158–2162

Chapter 15
Yeast as a Window into Changes in Genome Complexity Due to Polyploidization

Corey M. Hudson and Gavin C. Conant

Abstract Due to the long history of genetic analyses in yeasts and their experimental tractability, the yeast genome duplication provides important perspectives on the genome and population-level processes that follow whole-genome duplication (WGD). We discuss the history of the discovery of the *Saccharomyces cerevisiae* WGD, with special emphasis on the role of comparative genomics in its analysis. We then explore models of population and species divergence, both at a gene level (e.g., Dobzhansky-Muller incompatibility) and from the perspective of recent work on secondary allopolyploidy in *Saccharomyces pastorianus*. Finally, we explore the selective forces that act on the WGD-produced paralogs and shape their patterns of loss and retention. In addition to discussing the dosage balance hypothesis as it applies to the yeast WGD, we explore the role of the WGD in shaping several complex metabolic and regulatory phenotypes.

15.1 Introduction

Researchers have found remnants of ancient whole-genome duplications (WGDs) preserved in the genomes of many and diverse eukaryotes. This book, in fact, is a testament to that diversity, illustrating the sheer number of independent events in plants as well as the evolutionarily basal events in vertebrates and other, more recent WGDs in teleost fishes and frogs. Although we have still not fully validated

C. M. Hudson · G. C. Conant
MU Informatics Institute, University of Missouri, Columbia, MO 65211, USA

G. C. Conant (✉)
Division of Animal Sciences, University of Missouri, Columbia, MO, USA
e-mail: conantg@missouri.edu

P. S. Soltis and D. E. Soltis (eds.), *Polyploidy and Genome Evolution*,
DOI: 10.1007/978-3-642-31442-1_15,

Susumo Ohno's claim for the primacy of polyploidization in the generation of new adaptations (Ohno 1970), it is clear that WGD events have had a massive influence on the content and structure of the genomes of their possessors. The next step in exploring Ohno's hypothesis is to link genome evolution to known changes in function. This goal, however, remains challenging, primarily because our knowledge of how genotype links to phenotype remains woefully incomplete (Pigliucci 2010). However, one group of organisms in which we can at least begin to make such associations is in the polyploid yeasts. Our knowledge of the functional genomics of yeast is drawn primarily from *Saccharomyces cerevisiae*, which has a well-annotated genome, decades of biochemical, genetic, and cell biology research, a relatively small genome, and a life cycle that lends itself to scalable laboratory analyses. Taken together, these facts have allowed yeast researchers not only to understand the structure of the genome following WGD but also to experimentally evaluate hypotheses regarding the evolution of particular complex phenotypes. As we will stress throughout this chapter, one of the themes that emerges from all of these analyses is the degree to which the outcome of a WGD depends as much on the interactions between genes as on the role of any particular locus. Of equal importance evolutionarily, we now also have data from other polyploid yeast species, which are valuable both as a point of comparison to *S. cerevisiae* and for their own sakes. This wealth of data affords us insight into the mechanisms that drive the preferential loss and preservation of gene duplicates after polyploidy, lead to the functional divergence of genes, and are behind the evolutionary origins of complex phenotypes.

15.2 Evidence for WGD in Yeast

Although early analyses of genes potentially created by the vertebrate 2R events used phylogenetic approaches (Hughes 1999; Furlong and Holland 2002), most current studies of WGD rely on one or both of two methods: (1) finding numerous blocks of paralogous genes in multiple chromosomes with similar gene orders and (2) clustering homologs into groups by measuring the rate of synonymous substitutions (K_s or dS). This second method assumes that the gene pairs created by WGD cluster about some mean K_s value (Lynch and Conery 2000). The simultaneous application of both methods has been used to group multiple WGD events within species (e.g., *Arabidopsis thaliana* and *Tetraodon nigroviridis*; Jaillon et al. 2004; Van de Peer et al. 2009). However, the yeast genomes present an interesting challenge in this respect because the synonymous substitutions between yeast paralogs produced by WGD (hereafter ohnologs; Wolfe 2000) are often saturated (Byrne and Wolfe 2007). In other words, identical synonymous positions between two ohnologs occur almost as often due to repeated convergent substitutions as due to common ancestry, a fact pointed out by Smith (1987), who attempted to date histone gene duplicates in yeast. While the genomic structure of the core histone genes suggested that they were all duplicated simultaneously, these genes show

considerable variation in the numbers of synonymous substitutions separating them. This observation led Smith (1987) to hypothesize that *S. cerevisiae* underwent a WGD ancient enough that the duplicates surviving from it had saturated. However, it was not until the genome sequence of *S. cerevisiae* became available that this speculation could be confirmed (see below). Another similar but subtler problem in using K_s as a means of dating duplicate genes is the issue of gene conversion. Gene conversion was presumed to be quite common in yeast (Petes and Hill 1988), even prompting some authors (Gao and Innan 2004) to suggest that estimates of duplication rates based on duplicate divergences were inapplicable due to the homogenization of duplicate loci by conversion. Fortunately, although gene conversion is very common among yeast ribosomal proteins, it does not appear to be a general characteristic of the genome (Evangelisti and Conant 2010). Nonetheless, these various issues collectively meant that comparisons of paralogous sequence divergence were deemed unhelpful as a means to detect WGD in yeast.

15.2.1 Synteny-Based Evidence for WGD in Saccharomyces cerevisiae

Given that paralogous sequence comparisons were generally unhelpful in finding WGD relics, another tactic was to consider gene order. In fact, even before the *S. cerevisiae* genome sequence was completed in 1996, it was clear to many researchers that it contained numerous, long, homologous clusters of ordered genes (Goffeau et al. 1996). Melnick and Sherman (1993) found ordered homologous gene clusters in chromosomes V and X covering 7.5 kb. Lalo et al. (1993) similarly found ordered homologous gene clusters in chromosomes XIV and III covering 15 kb. When the genome was sequenced, researchers found 18 ordered homologous genes in chromosomes IV and II that covered 120 and 170 kb, respectively (Goffeau et al. 1996). Just how to interpret these redundant regions remained a challenge at that time (Goffeau et al. 1996; Oliver 1996), and, in spite of Smith's prior hypothesis (Smith 1987), few, if any, of the contemporaneous explanations included an ancient WGD.

However, opinions changed the next year when Wolfe and Shields (1997) presented a thorough, genome sequence-based, analysis that gave strong evidence for WGD in *S. cerevisiae*. To find syntenic regions, they conducted a BLASTP search of amino acid sequences throughout the yeast genome and made a dot plot of the results. They then created gene blocks from these data, where each block was required to have at least three homologous pairs with intergenic distances $\leq$ 50 kb and conservation of gene order and orientation. This analysis yielded 55 duplicated regions containing a total of 376 pairs of ohnologs. The large number of duplicated regions led Wolfe and Shields to posit two explanations: (1) successive independent gene duplications, and (2) a single duplication of the entire genome,

followed by massive gene loss. There were two lines of evidence discounting the first possibility. First, 90 % (50/55) of the gene regions shared the same orientation with respect to the centromeres of the duplicated regions when we would expect independent duplications to be instead randomly distributed about the centromeres. Second, there were no examples of triplicated regions in the *S. cerevisiae* genome. If the duplications involved several distinct events separated in time, such a pattern would be highly unlikely, because it would require that later duplication events never overlapped with prior ones. Given these arguments, Wolfe and Shields (1997) argued for a single ancient WGD, which they dated to be hundreds of millions of years old (note that attempts to conclusively date this event have been difficult, due to a lack of fossils and the previously mentioned saturation of substitutions; see Taylor and Berbee 2006; Rolland and Dujon 2011).

15.2.2 Comparative Genomics and Proof of WGD in S. cerevisiae

A number of researchers disputed the claims of Wolfe and Shields (1997), arguing that, because the syntenic regions identified made up only a small part of the genome, independent duplications better explained the genomic structure of *S. cerevisiae* (Coissac et al. 1997; Mewes et al. 1997; Hughes et al. 2000; Llorente et al. 2000a; Llorente et al. 2000b; Friedman and Hughes 2001; Piskur 2001; Koszul et al. 2004). However, this independent duplication hypothesis became untenable following the genome sequencing of other yeasts that proved to lack these syntenic paralog blocks. These sequences were described by three independent groups. The comparison of *S. cerevisiae* with *Kluyveromyces waltii* (Kellis et al. 2004) and the comparison of *S. cerevisiae* with *Ashbya gossypii* (Dietrich et al. 2004) involved different genomes, but effectively made the same argument: that the 2:1 mapping of blocks of paralogs from *S. cerevisiae* to homologous single-copy genes in *K. waltii/A.gossypii* could best be explained by WGD. This explanation was particularly striking because the doubly conserved synteny blocks cover 90 % of the genome in *K. waltii* (Kellis et al. 2004) and 96 % of that in *A. gossypii* (Dietrich et al. 2004). Furthermore, both studies found a large number of 2:1 pairing of centromeres in the species-respective chromosomes. There were 16:8 such pairings between *S. cerevisiae* and *K. waltii* and 14:7 between *S. cerevisiae* and *A. gossypii* with a subsequent break at the expected centromere position in *S. cerevisiae* chromosomes X and XII that are syntenic with regions in *A. gossypii* chromosomes I and III. Finally, and perhaps most strikingly, both groups also showed that the single-copy orthologs of genes from *A. gossypii* or *K. waltii* in the genome of *S. cerevisiae* are interleaved between two paralogous chromosomes in *S. cerevisiae* that nonetheless retain the relative gene order of the single chromosome in the non-WGD yeast (see Fig. 15.1). Such a pattern is only explicable under the hypothesis of a WGD event followed by massive gene losses.

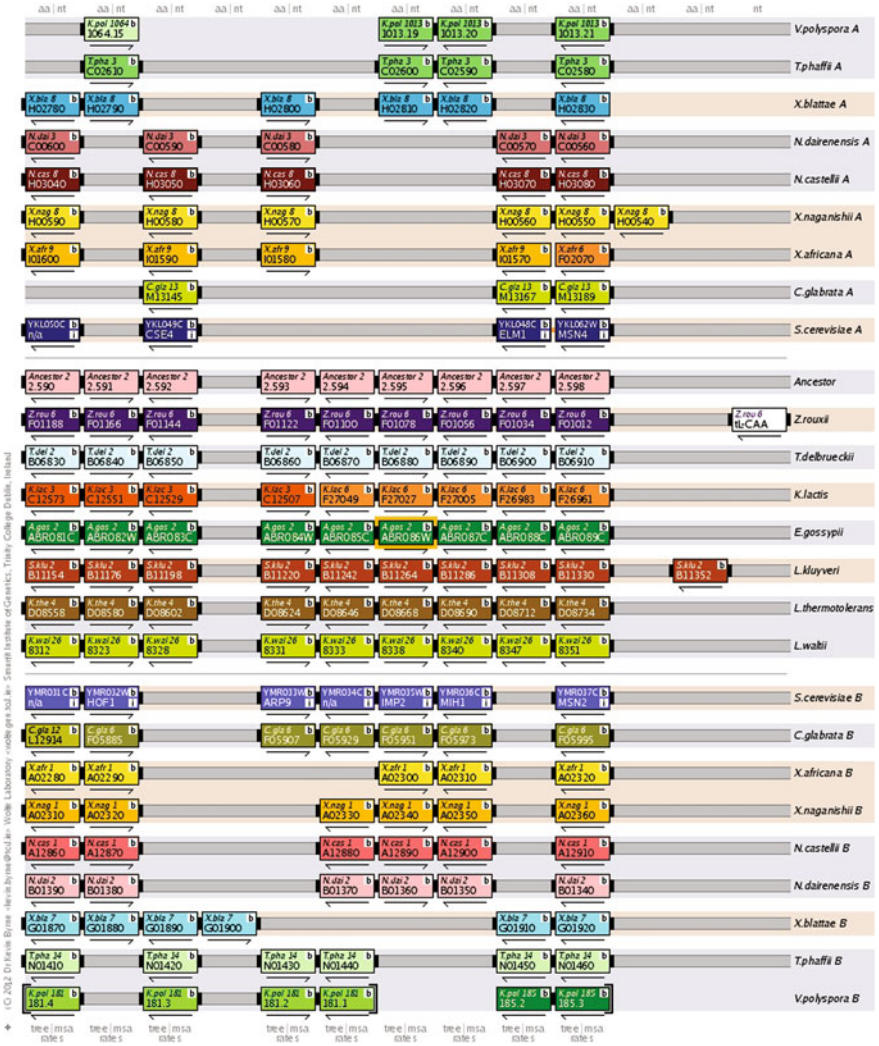

Fig. 15.1 Yeast gene order browser (YGOB) screenshots with a window size of six. Each box represents a gene; each color, a chromosome. The gene in focus, the *A. gossypii* gene *ABR086W*, is highlighted by an *orange* border. Each vertical column ("pillar") represents a single gene prior to the WGD (hence, all genes in a column are homologs, and the paired upper and lower genes, when present, are paralogs). The ancestral order of these genes (*pink* boxes) just prior to the WGD has also been exhaustively inferred (Gordon et al. 2011). Connectors join nearby genes: a solid bar for adjacent genes, two bars for loci less than five genes apart, and one bar for loci <20 genes apart. The connectors are extended in gray over intervening space. The end of a chromosome or contig is denoted by a *brace*. *Arrows* denote transcriptional orientation. The browser also includes a control panel that allows users to select the window size and the gene to focus on. This panel also has buttons for running BLAST searches against YGOB's database, outputting YGOB data in tabular format, obtaining pairwise K_a and K_s values among genes, and computing multiple sequence alignments and phylogenetic trees of individual pillars. Species names for each track are labeled at *right* (Byrne and Wolfe 2005)

The argument of Dujon et al. (2004) is subtly different. They sequenced and analyzed four other genomes. One genome that of *Candida glabrata*, shares the genome duplication with *S. cerevisiae*. This was determined by comparing syntenic blocks in *S. cerevisiae* and *C. glabrata* with the other three sequenced genomes, *Kluveromyces lactis*, *Debaryomyces hansenii*, and *Yarrowia lipolytica*. Dujon et al. (2004) found 20 distinct blocks of paralogs shared by both *S. cerevisiae* and *C. glabrata*. These blocks allowed them to map the WGD onto a phylogeny, rather than do a simple pairwise comparison. Mapping this WGD phylogenetically creates distinct hypotheses as to where in the tree we expect to find polyploid yeasts (c.f., Fig. 15.1), predictions that have been confirmed with each of the subsequently sequenced genomes of known phylogenetic position (Wapinski et al. 2007; Scannell et al. 2011).

15.2.3 Yeast Gene Order Browser

One of the major benefits of studying the yeast WGD is that the relatively slow rates of gene order change in yeast genomes and the compactness of their genomes means that an exhaustive enumeration of all WGD-produced ohnologs is possible. Just such a project was carried out, with the results presented as the web-based Yeast Gene Order Browser (YGOB) (Byrne and Wolfe 2005), which illustrates a number of non- and post-WGD yeasts in a graphical framework (Fig. 15.2). This work has been followed by a reconstruction of the set of genes and their relative orders that existed just prior to the WGD (Gordon et al. 2009) and by a likelihood-based model of post-WGD duplicate loss that attempts to quantify the orthology inferences made by YGOB (Conant and Wolfe 2008a). On the basis of these three projects, the post-WGD evolutionary history of virtually every locus in the *S. cerevisiae* genome can be traced (Fig. 15.1 is thus illustrative of the predominant pattern seen across the genome).

*15.2.4 Additional Non-*Saccharomyces*-Specific WGDs*

In addition to the ancient WGD that characterizes the *Saccharomyces* clade (Fig. 15.2), several cases of allopolyploidy have been discovered in yeasts. Some of these occur in species within the *Saccharomyces sensu stricto* clade (Scannell et al. 2011), while others are independent.

15.2.4.1 Secondary Allopolyploidy in *Saccharomyces pastorianus*

Several cases of allopolyploidy are known from within *S. sensu stricto* (Dequin and Casaregola 2011). One of the most well studied is that of the lager yeast, *S. pastorianus* (syn. *S. carlsbergensis*). It has long been known that the polyploid *S. pastorianus* and other members of the complex of related lager yeasts are

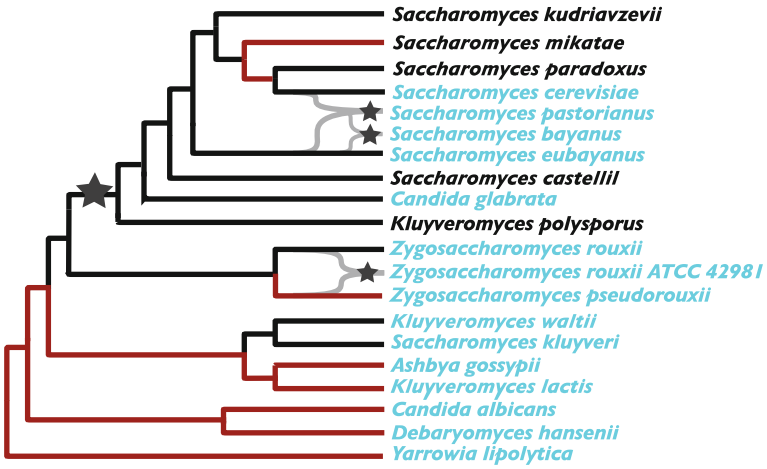

Fig. 15.2 Consensus view of the evolutionary relationships between the yeast taxa discussed. Black branches indicate relationships described by both Kurtzman and Robnett (2003) and Fitzpatrick et al. (2006). *Red* branches indicate conflicts between the two phylogenies, in which case Fitzpatrick et al. (2006) is presented. *Curved grey* branches illustrate the allopolyploidy events between two species (*S. cerevisiae* and *S. bayanus, Z. rouxii,* and *Z. pseudorouxii*). Taxa in *blue* are reported in text (e.g., *S. pastorianus* and *Z. rouxii ATCC 42981*). *Stars* mark whole-genome duplications. Note that genus names are an imperfect guide to the relationships

allotetraploids of diploid *S. cerevisiae* and some other unknown diploid species (Martini and Kurtzman 1985; Kielland-Brandt et al. 1995). However, aside from the general difficulties facing anyone interested in identifying the origins of hybrid genomes, the debate surrounding the origin of the second parental diploid species was further complicated by a difficulty in delimiting species within these groups (Rainieri et al. 2006). The tetraploid *S. pastorianus* belongs to a group of yeast species which, until recently, was represented as a phylogenetically unresolved species complex including *S. pastorianus*, *S. monacensis* (*S. pastorianus* strain CBS 1503), *S. bayanus*, and *S. bayanus* var. *uvarum* (Casaregola et al. 2001; Rainieri et al. 2006). This taxonomic confusion has recently been partially resolved through the sequencing of the genomes of both *S. pastorianus* and one of its presumed parental diploid species, *S. eubayanus* (Nakao et al. 2009). The genome history that has emerged is a complicated story of allopolyploidy followed by genomic transformation forming the related species *S. bayanus* (Libkind et al. 2011). As summarized by Libkind et al. (2011), *S. cerevisiae* hybridized with *S. eubayanus* (a species recently recovered in Patagonia) with subsequent genome doubling producing the allotetraploid progenitor of modern *S. pastorianus*. Following domestication, smaller regions of the *S. pastorianus* genome were then apparently transferred into the genome of the diploid parent *S. eubayanus* (which is nonetheless a descendant of the ancient polyploidy). This hybrid form of *S. eubayanus*, with contributions from *S. pastorianus,* then proceeded to interbreed with diploid *S. uvarum* to produce the modern, diploid, *S. bayanus* (Fig. 15.3).

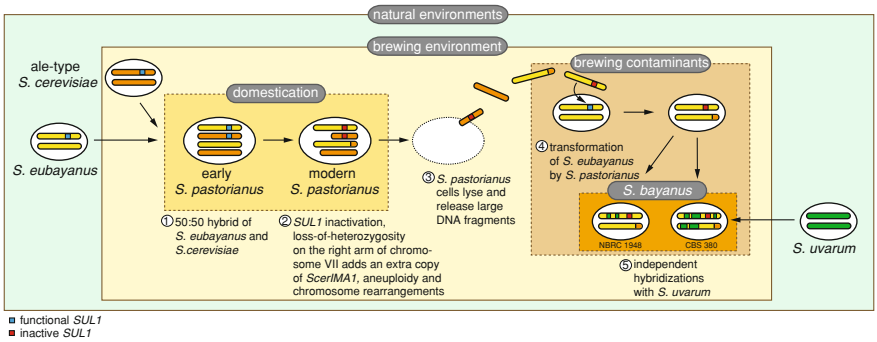

Fig. 15.3 Genome evolution in *S. pastorianus*. A model of the formation of *S. pastorianus* and the hybrid strains of *S. bayanus*. First, wild *S. eubayanus* and ale-type *S. cerevisiae* hybridized to form an allotetraploid that became the ancestor of the modern (doubly paleopolyploid) *S. pastorianus*. Second, domestication imposed strong selective pressure for strains with the most desirable brewing properties. Third, in the brewing vats with high densities of *S. pastorianus*, cell lysis releases large DNA fragments that occasionally transform, fourth, contaminating wild strains of *S. eubayanus* (which possesses only the ancient WGD shared with *S. cerevisiae*) because of the lack of pure culture techniques. Fifth, multiple hybridization events between *S. eubayanus* and wild strains of *S. uvarum* gave rise to CBS 380T and NBRC 1948. This model does not exclude prior or parallel involvement of *S. uvarum* in brewing or contamination. *Reprinted from Libkind et al.* (2011)

15.2.4.2 Allopolyploidy in *Zygosaccharomyces rouxii*

The spoilage agent and industrial yeast *Zygosaccharomyces rouxii* strain ATCC 42981 was identified as another allopolyploid by James et al. (2005) and Gordon and Wolfe (2008). This hybridization/polyploidy event is significant for two reasons. First, unlike all of the previous examples, it occurs outside of *S. sensu stricto*. Second, Gordon and Wolfe (2008) determined that most of the paralogs produced by WGD are still present, presumably due to the recentness of the event. Thus, while other yeast genome duplications are ancient and show considerable gene loss and rearrangement (Wolfe and Shields 1997), the *Z. rouxii* genome retains most of the "new" genes produced by its WGD. Since the survival time of ohnologs has been modeled to follow a power law, most of the duplicates are expected to be lost very rapidly (Maere et al. 2005), suggesting that *Z. rouxii* represents an example of the early features of genome evolution following WGD.

15.3 WGD and Speciation

An important potential outcome of polyploidy is in altering patterns of speciation. This change can happen in at least two ways. First, the WGD can relax selective constraints resulting in an adaptive radiation by means of ecological speciation. Another, more neutral mechanism, is a special case of the Dobzhansky-Muller (DM) process of speciation, in which species lose reciprocal paralogs following

some period of isolation (Lynch and Force 2000; Werth and Windham 1991; see also Chap. 1, this volume). WGD potentially increases the probability of this simply by increasing the number of paired genes in a genome. The fertility of hybrids is 0.75^n, where n is the number of reciprocal losses of essential genes among populations (Werth and Windham 1991). Clearly, for any significant number of reciprocal losses (such as occur after WGD), the number of viable, fertile offspring of a crossing of two such populations is negligible. Both phylogenetic and experimental studies of the DM process after WGD have been carried out in yeast. Scannell et al. (2006) showed that the number of reciprocal gene losses in several species of yeast sharing the *S. cerevisiae* WGD was sufficient to induce such inviability. This observation suggests that a DM mechanism was partly responsible for the multiple speciation events among the *Saccharomyces* species (e.g., *S. cerevisiae*, *S. bayanus*, and *C. glabrata*) following WGD.

An advantage of studying the DM process in yeast is the ability to experimentally create and cross artificial polyploids. This possibility has been highlighted in experimental studies of reproductive isolation. Polyploid yeasts have been allowed to evolve in different selective environments (Dettman et al. 2007) and in neutral environments subject to random mutagenesis (Maclean and Greig 2011). These two experiments have shown that moderate reproductive isolation, coupled with reciprocal gene loss, results in a clear loss of fitness when independently derived polyploids are crossed. Similarly, Lee et al. (2008) showed that hybrids of *S. cerevisiae* and *S. bayanus* were less fit than their parental phenotypes, due primarily to incompatibility between their nuclear and mitochondrial genomes. Chou et al. (2010) extended this analysis, providing another pair of mitochondrial and nuclear genes and posited nuclear-mitochondrial incompatibility as a common mechanism in species formation. In another twist, Anderson et al. (2010) demonstrated the existence of alleles with depressed hybrid fitness in low-glucose environments, which argues for a model in which neutral changes in paired genes are followed by strong selection, a sequence of events that promotes rapid reproductive isolation. Kao et al. (2010), however, argue against the existence of a small number of so-called *speciation genes*, instead claiming that genome scans provide no evidence of any single paired dominant or recessive genic incompatibilities. They instead argue that following WGD, many changes in loci of little effect resulted in lowered fitness due, in part, to the rewiring of transcriptional and metabolic networks.

Another debate that has emerged in this field is whether these changes are due primarily to the decrease in the fertility of hybrids (Xu and He 2011) or a decrease in their viability (Greig 2008). This question ultimately amounts to a debate about what stage in the yeast life cycle the genetic incompatibilities occur—sporulation or clonal growth, and whether the decrease in fitness is the result of competition for resources or offspring. The discontinuity between these ideas likely represents an opportunity to explain speciation as a process across different genomic and temporal scales, and we would speculate that the process of DM incompatibility induces selection for the evolution of some form of prezygotic barrier.

15.4 Changes in Genome Content and Complexity Post-WGD

Duplicate retention and evolutionary models. In addition to such population processes as speciation, WGD also altered many other aspects of the *S. cerevisiae* lifestyle. For instance, several pairs of ohnologs have been shown to have undergone various types of functional divergence, allowing the study of some of the proposed mechanisms of duplicate divergence after duplication (Conant and Wolfe 2008b). In an elegant series of experiments, van Hoof (2005) showed that two ohnologs, *ORC1* and *SIR3*, have distinct and non-overlapping functions (in DNA replication and gene silencing, respectively). Strikingly, however, the mutual ortholog of these genes from the non-WGD yeast *S. kluyveri* is able to complement both functions, constituting a clear example of subfunctionalization. An apparently similar case, involving the *S. cerevisiae* ohnolog pair *GAL1* and *GAL3*, which presently functions, respectively, as an enzyme and as a transcriptional regulator, was complicated by the discovery of an adaptive conflict between the shared regulator and enzymatic function of their ortholog in the non-WGD *K. lactis*. Thus, although the *K. lactis GAL1* gene does indeed serve the functions of both *GAL1* and *GAL3* in *S. cerevisiae*, it does so in a suboptimal way, being unable to tune its expression to both roles simultaneously (Hittinger and Carroll 2007). This conflict illustrates an important point about subfunctionalization, namely that the original neutral model of subfunctionalization proposed by Force and coauthors (Force et al. 1999) is not the only possible mechanism for such functional partitioning (Des Marais and Rausher 2008). Other examples of divergence among ohnologs where the mechanism of that divergence is less clear include ribosomal proteins (Ni and Snyder 2001; Komili et al. 2007; Kim et al. 2009), glucose sensors (Özcan et al. 1998), and glycolysis enzymes (Boles et al. 1997).

The dosage balance hypothesis (DBH). In addition to facilitating the above work, the wealth of functional data from *S. cerevisiae* also provides an excellent opportunity to test hypotheses explaining the differences in gene retention patterns after WGD and small-scale duplications (hereafter SSD). Chief among these is probably the DBH (Papp et al. 2003; Freeling and Thomas 2006; Birchler and Veitia 2007; Freeling 2009), which states that, in eukaryotes, there is selection operating to disfavor duplications of central network genes due to the imbalance in network stoichiometry that results. This situation is reversed for WGD because in that case, the loss of a second copy of a gene introduces imbalances relative to the remaining duplicated genes. In keeping with the DBH, several classes of genes are over-retained after several evolutionarily ancient WGD events, including that in yeast. They include ribosomal proteins, protein kinases and transcription factors (Seoighe and Wolfe 1999; Blanc and Wolfe 2004; Maere et al. 2005; Aury et al. 2006; Conant and Wolfe 2008a). Similarly, genes that tend to have been fixed by WGD are less likely to have undergone SSD in other yeast species (Wapinski et al. 2007). However, duplicates produced by WGD have more protein interactions (Guan et al. 2007; Hakes et al. 2007), more phosphorylation sites (Amoutzias et al. 2010), and tend to be highly expressed (Seoighe and Wolfe 1999) than those from

SSD. Although genes retained in duplicate after WGD are rarely essential on an individual basis (Guan et al. 2007), this dispensability appears to be due to functional compensation by the other ohnolog (DeLuna et al. 2008). Thus, it appears that while ohnologs are less likely to be essential than their SSD counterparts today, their ancestral genes were actually at least as essential as current single-copy genes (DeLuna et al. 2008).

System-level changes produced by WGD. Of course, one of the unique features of polyploidy relative to SSD is the possibility of coordinated changes among multiple sets of ohnologs. At the simplest level, we have previously illustrated examples of what appears to be network subfunctionalization where a number of ohnologs collectively divided two expression domains among themselves (Conant and Wolfe 2006). A more complex and interesting example is the role of the WGD (Piškur et al. 2006) in shaping *S. cerevisiae's* propensity for aerobic glucose fermentation (the Crabtree effect; Geladé et al. 2003; Johnston and Kim 2005), a novel and somewhat paradoxical phenotype. There is a general association between the presence of the WGD and the Crabtree effect across yeast species (Merico et al. 2007). As a result, we and others have argued that dosage effects among the glycolysis enzymes post-WGD helped to increase flux through glycolysis (Blank et al. 2005; Kuepfer et al. 2005; Conant and Wolfe 2007; Merico et al. 2007; van Hoek and Hogeweg 2009). Such increased flux likely could only be accommodated through fermentation pathways, given the complex spatial organization of the competing respiratory pathway (Conant and Wolfe 2007). Supporting this hypothesis is an elegant computational analysis by van Hoek and Hogeweg (2009) showing that future WGD events in modern *S. cerevisiae* could also be expected to provide a selective advantage in glucose-rich environments through the preferential retention of duplicated glycolysis enzymes. Note that the apparently "wasteful" fermentation can actually be selectively advantageous in the context of rich but ephemeral resource patches (Pfeiffer et al. 2001; Pfeiffer and Schuster 2005), a phenomenon that has been experimentally confirmed in yeast (MacLean and Gudelj 2006). Such a change in the yeast lifestyle likely led to other, later changes in the genome. One suggestive example concerns the decoupling of cytosolic and mitochondrial ribosomal protein expression post-WGD (Ihmels et al. 2005). Prior to WGD, bakers' yeast was likely similar to other yeasts in having a strong association in the expression of the two types of ribosomal proteins. After WGD, however, *cis*-regulatory element evolution diverged in the two groups of genes (Ihmels et al. 2005), allowing *S. cerevisiae* to express only cytosolic proteins at high levels during fermentation, an important refinement in a fermentative lifestyle.

Connecting the DBH to large-scale evolutionary changes following WGD, Conant (2010) and Fusco et al. (2010) found transcriptional regulatory motifs to be over-retained in ohnologs. Modeling network evolution after WGD, these authors find the network enriched for transcription factors and particular network motifs. Duplicated transcription factors still show some relics of the WGD, being more likely to share targets than are random transcription factors, but on the whole show considerable divergence post-WGD (Conant 2010). Given this rapid regulatory evolution, it may not be easy to ascertain the role of WGD in the evolution of

the modern *S. cerevisiae* regulatory network. Nonetheless, the retention of many transcription factors that have acquired distinct sets of target genes may imply that the WGD served to "relax" the regulatory complexity of this organism, which may have implications for its future ability to adapt (as seen for the *GAL1/GAL3* example).

15.5 Conclusions

The *S. cerevisiae* WGD has been implicated in a number of evolutionarily complex events. At a minimum, a set of duplicated genes of identical age is a powerful system for exploring duplicate gene evolution (van Hoof 2005; Conant and Wolfe 2006; Fares et al. 2006; Kim and Yi 2006). However, we also suggest that, as with the *GAL1/GAL3* example, we will not fully understand the biology of *S. cerevisiae* until we account for how the WGD has altered both the individual roles of particular genes and their relationships to each other. We have outlined some of the areas of yeast biology that we think were altered by this genome-doubling event: there remain others yet to be discovered. Similarly, the presence of other WGD events, of varying ages, allows us to study how these events unfold over various timescales, including, potentially, on the timescale of laboratory experiments in evolution.

Acknowledgments We would like to thank Michaël Bekaert, Patrick Edger, and Chris Pires for helpful discussions. This work was supported by the National Library of Medicine Biomedical and Health Informatics Training Fellowship [LM007089-19] (CMH) and the Reproductive Biology Group of the Food for the twenty-first century program at the University of Missouri (GCC).

References

Amoutzias GD, He Y, Gordon J, Mossialos D, Oliver SG, Van de Peer Y (2010) Posttranslational regulation impacts the fate of duplicated genes. Proc Natl acad sci U S A 107:2967–2971

Anderson JB, Funt J, Thompson DA et al (2010) Determinants of divergent adaptation and Dobzhansky-Muller interaction in experimental yeast populations. Curr Biol 20:1383–1388

Aury JM, Jaillon O, Duret L et al (2006) Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. Nature 444:171–178

Birchler JA, Veitia RA (2007) The gene balance hypothesis: from classical genetics to modern genomics. Plant Cell 19:395–402

Blanc G, Wolfe KH (2004) Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. Plant Cell 16:1679–1691

Blank LM, Lehmbeck F, Sauer U (2005) Metabolic-flux and network analysis of fourteen hemiascomycetous yeasts. FEMS Yeast Res 5:545–558

Boles E, Schulte F, Miosga T, Freidel K, Schlüter E, Zimmermann FK, Hollenberg CP, Heinisch JJ (1997) Characterization of a glucose-repressed pyruvate kinase (Pyk2p) in *Saccharomyces cerevisiae* that is catalytically insensitive to fructose-1-6-biphosphate. J Bacteriol 179:2987–2993

Byrne KP, Wolfe KH (2005) The yeast gene order browser: combining curated homology and syntenic context reveals gene fate in polyploid species. Genome Res 15:1456–1461

Byrne KP, Wolfe KH (2007) Consistent patterns of rate asymmetry and gene loss indicate widespread neofunctionalization of yeast genes after whole-genome duplication. Genetics 175:1341–1350

Casaregola S, Nguyen HV, Lapathitis G, Kotyk A, Gaillardin C (2001) Analysis of the constitution of the beer yeast genome by PCR, sequencing and subtelomeric sequence hybridization. Int J Syst Evol Microbiol 51:1607–1618

Chou J-Y, Hung Y-S, Lin K-H, Lee H-Y, Leu J-Y (2010) Multiple molecular mechanisms cause reproductive isolation between three yeast species. PLoS Biol 8:e1000432

Coissac E, Maillier E, Netter P (1997) A comparative study of duplications in bacteria and eukaryotes: the importance of telomeres. Mol Biol Evol 14:1062–1074

Conant GC (2010) Rapid reorganization of the transcriptional regulatory network after genome duplication in yeast. Proc R Soc B 277:869–876

Conant GC, Wolfe KH (2006) Functional partitioning of yeast co-expression networks after genome duplication. PLoS Biol 4:e109

Conant GC, Wolfe KH (2007) Increased glycolytic flux as an outcome of whole-genome duplication in yeast. Mol Syst Biol 3:129

Conant GC, Wolfe KH (2008a) Probabilistic cross-species inference of orthologous genomic regions created by whole-genome duplication in yeast. Genetics 179:1681–1692

Conant GC, Wolfe KH (2008b) Turning a hobby into a job: how duplicated genes find new functions. Nat Rev Genet 9:938–950

DeLuna A, Vetsigian K, Shoresh N, Hegreness M, Colón-González M, Chao S, Kishony R (2008) Exposing the fitness contribution of duplicated genes. Nat Genet 40:676–681

Dequin S, Casaregola S (2011) The genomes of fermentative Saccharomyces. CR Biol 334:687–693

Des Marais DL, Rausher MD (2008) Escape from adaptive conflict after duplication in an anthocyanin pathway gene. Nature 454:762–765

Dettman JR, Sirjusingh C, Kohn LM, Anderson JB (2007) Incipient speciation by divergent adaptation and antagonistic epistasis in yeast. Nature 447:585–588

Dietrich FS, Voegeli S, Brachat S et al (2004) The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. Science 304:304–307

Dujon B, Sherman D, Fischer G et al (2004) Genome evolution in yeasts. Nature 430:35–44

Evangelisti AM, Conant GC (2010) Nonrandom survival of gene conversions among yeast ribosomal proteins duplicated through genome doubling. Genome Biol Evol 2:826–834

Fares MA, Byrne KP, Wolfe KH (2006) Rate asymmetry after genome duplication causes substantial long-branch attraction artifacts in the phylogeny of *Saccharomyces* species. Mol Biol Evol 23:245–253

Fitzpatrick D, Logue M, Stajich J, Butler G (2006) A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. BMC Evol Biol 6:99

Force A, Lynch M, Pickett FB, Amores A, Yan Y, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531–1545

Freeling M (2009) Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. Annu Rev Plant Biol 60:433–453

Freeling M, Thomas BC (2006) Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. Genome Res 16:805–814

Friedman R, Hughes AL (2001) Gene duplication and the structure of eukaryotic genomes. Genome Res 11:373–381

Furlong RF, Holland PWH (2002) Were vertebrates octoploid? Philos Trans R Soc Lond B 357:531–544

Fusco D, Grassi L, Bassetti B, Caselle M, Lagomarsino MC (2010) Ordered structure of the transcription network inherited from the yeast whole-genome duplication. BMC Syst Biol 4:77

Gao LZ, Innan H (2004) Very low gene duplication rate in the yeast genome. Science 306: 1367–1370

Geladé R, Van De Velde S, Van Dijck P, Thevelein JM (2003) Multi-level response of the yeast genome to glucose. Genome Biol 4:233

Goffeau A, Barrell B, Russey H et al (1996) Life with 6000 genes. Science 274:562–567

Gordon JL, Byrne KP, Wolfe KH (2009) Additions, losses and rearrangements on the evolutionary route from a reconstructed ancestor to the modern *Saccharomyces cerevisiae* genome. PLoS Genet 5:e1000485

Gordon JL, Byrne KP, Wolfe KH (2011) Mechanisms of chromosome number evolution in yeast. PLoS Genet 7:e1002190

Gordon JL, Wolfe KH (2008) Recent allopolyploid origin of *Zygosaccharomyces rouxii* strain ATCC 42981. Yeast 25:449–456

Greig D (2008) Reproductive isolation in *Saccharomyces*. Heredity 102:39–44

Guan Y, Dunham MJ, Troyanskaya OG (2007) Functional analysis of gene duplications in *Saccharomyces cerevisiae*. Genetics 175:933–943

Hakes L, Pinney JW, Lovell SC, Oliver SG, Robertson DL (2007) All duplicates are not equal: the difference between small-scale and genome duplication. Genome Biol 8:R209

Hittinger CT, Carroll SB (2007) Gene duplication and the adaptive evolution of a classic genetic switch. Nature 449:677–681

Hughes AL (1999) Phylogenies of developmentally important proteins do not support the hypothesis of two rounds of genome duplication early in vertebrate history. J Mol Evol 48:565–576

Hughes TR, Roberts CJ, Dai H et al (2000) Widespread aneuploidy revealed by DNA microarray expression profiling. Nat Genet 25:333–337

Ihmels J, Bergmann S, Gerami-Nejad M, Yanai I, McClellan M, Berman J, Barkai N (2005) Rewiring of the yeast transcriptional network through the evolution of motif usage. Science 309:938–940

Jaillon O, Aury J-M, Brunet F et al (2004) Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. Nature 431:946–957

James SA, Bond CJ, Stratford M, Roberts IN (2005) Molecular evidence for the existence of natural hybrids in the genus *Zygosaccharomyces*. FEMS Yeast Res 5:747–755

Johnston M, Kim J-H (2005) Glucose as a hormone: Receptor-mediated glucose sensing in the yeast *Saccharomyces cerevisiae*. Biochem Soc Trans 33:247–252

Kao KC, Schwartz K, Sherlock G (2010) A genome-wide analysis reveals no nuclear Dobzhansky-Muller pairs of determinants of speciation between *S. cerevisiae* and *S. paradoxus*, but suggests more complex incompatibilities. PLoS Genet 6:e1001038

Kellis M, Birren BW, Lander ES (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. Nature 428:617–624

Kielland-Brandt MC, Nilsson-Tillgren T, Gjermansen C, Holmberg S, Pedersen MB (1995) Genetics of brewing yeasts. In: Rose AH, Wheals AE, Harrison JS (eds) The Yeasts, vol 6, 2nd edn. Academic, London, pp 223–254

Kim S-H, Yi SV (2006) Correlated asymmetry of sequence and functional divergence between duplicate proteins of *Saccharomyces cerevisiae*. Mol Biol Evol 23:1068–1075

Kim T-Y, Ha CW, Huh W-K (2009) Differential subcellular localization of ribosomal protein L7 paralogs in *Saccharomyces cerevisiae*. Mol Cells 27:539–546

Komili S, Farny NG, Roth FP, Silver PA (2007) Functional specificity among ribosomal proteins regulates gene expression. Cell 131:557–571

Koszul R, Caburet S, Dujon B, Fischer G (2004) Eucaryotic genome evolution through the spontaneous duplication of large chromosomal segments. EMBO J 23:234–243

Kuepfer L, Sauer U, Blank LM (2005) Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. Genome Res 15:1421–1430

Kurtzman C, Robnett C (2003) Phylogenetic relationships among yeasts of the '*Saccharomyces* complex' determined from multigene sequence analyses. FEMS Yeast Res 3:417–432

Lalo D, Stettler S, Mariotte S, Slominski PP, Thuriaux P (1993) Two yeast chromosomes are related by a fossil duplication of their centromeric regions. C R Acad Sci 316:367–373

Lee H-Y, Chou J-Y, Cheong L, Chang N-H, Yang S-Y, Leu J-Y (2008) Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. Cell 135:1065–1073

Libkind D, Hittinger CT, Valério E, Gonçalves C, Dover J, Johnston M, Gonçalves P, Sampaio JP (2011) Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. Proc Nat Acad Sci 108:14539–14544

Llorente B, Durrens P, Malpertuy A et al (2000a) Genomic exploration of the hemiascomycetous yeasts: 20. Evolution of gene redundancy compared to *Saccharomyces cerevisiae*. FEBS Lett 487:122–133

Llorente B, Malpertuy A, Neuvéglise C et al (2000b) Genomic exploration of the hemiascomycetous yeasts: 18. Comparative analysis of chromosome maps and synteny with *Saccharomyces cerevisiae*. FEBS Lett 487:101–112

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290:1151–1155

Lynch M, Force AG (2000) The origin of interspecific genomic incompatibility via gene duplication. Am Nat 156:590–605

Maclean CJ, Greig D (2011) Reciprocal gene loss following experimental whole-genome duplication causes reproductive isolation in yeast. Evolution 65:932–945

MacLean RC, Gudelj I (2006) Resource competition and social conflict in experimental populations of yeast. Nature 441:498–501

Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer. Y (2005) Modeling gene and genome duplications in eukaryotes. Proc Nat Acad Sci U S A 102:5454–5459

Martini AV, Kurtzman CP (1985) Deoxyribonucleic acid relatedness among species of the genus *Saccharomyces sensu stricto*. Int J Syst Bacteriol 35:508–511

Melnick L, Sherman F (1993) The gene clusters ARC and COR on chromosomes 5 and 10, respectively, of *Saccharomyces cerevisiae* share a common ancestry. J Mol Biol 233:372–388

Merico A, Sulo P, Piškur J, Compagno C (2007) Fermentative lifestyle in yeasts belonging to the *Saccharomyces* complex. FEBS J 274:976–989

Mewes H, Albermann K, Bähr M et al (1997) Overview of the yeast genome. Nature 387:7–65

Nakao Y, Kanamori T, Itoh T, Kodama Y, Rainieri S, Nakamura N, Shimonaga T, Hattori M, Ashikari T (2009) Genome sequence of the lager brewing yeast, an interspecies hybrid. DNA Res 16:115–129

Ni L, Snyder M (2001) A genomic study of the bipolar bud site selection pattern in *Saccharomyces cerevisiae*. Mol Biol Cell 12:2147–2170

Ohno S (1970) Evolution by gene duplication. Springer, New York

Oliver SG (1996) From DNA sequence to biological function. Nature 379:597–600

Özcan S, Dover J, Johnston M (1998) Glucose sensing and signaling by two glucose receptors in the yeast *Saccharomyces cerevisiae*. EMBO J 17:2566–2573

Papp B, Pal C, Hurst LD (2003) Evolution of cis-regulatory elements in duplicated genes of yeast. Trends Genet 19:417–422

Petes T, Hill C (1988) Recombination between repeated genes in microorganisms. Annu Rev Genet 22:147–168

Pfeiffer T, Schuster S (2005) Game-theoretical approaches to studying the evolution of biochemical systems. Trends Biochem Sci 30:20–25

Pfeiffer T, Schuster S, Bonhoeffer S (2001) Cooperation and competition in the evolution of ATP-producing pathways. Science 292:504–507

Pigliucci M (2010) Genotype-phenotype mapping and the end of the 'genes as blueprint' metaphor. Philos Trans R Soc Lond B 365:557–566

Piskur J (2001) Origin of the duplicated regions in the yeast genomes. Trends Genet 17:302–303

Piškur J, Rozpedowska E, Polakova S, Merico A, Compagno C (2006) How did *Saccharomyces* evolve to become a good brewer? Trends Genet 22:183–186

Rainieri S, Kodama Y, Kaneko Y, Mikata K, Nakao Y, Ashikari T (2006) Pure and mixed genetic lines of *Saccharomyces bayanus* and *Saccharomyces pastorianus* and their contribution to the lager brewing strain genome. Appl Environ Microbiol 72:3968–3974

Rolland T, Dujon B (2011) Yeasty clocks: dating genomic changes in yeasts. CR Biol 334:620–628

Scannell DR, Byrne KP, Gordon JL, Wong S, Wolfe KH (2006) Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. Nature 440:341–345

Scannell DR, Zill OA, Rokas A, Payen C, Dunham MJ, Eisen MB, Rine J, Johnston M, Hittinger CT (2011) The awesome power of yeast evolutionary genetics: new genome sequences and strain resources for the *Saccharomyces sensu stricto* genus. G3: Genes, Genomes, Genetics 1:11–25

Seoighe C, Wolfe KH (1999) Yeast genome evolution in the post-genome era. Curr Opin Microbiol 2:548–554

Smith M (1987) Molecular evolution of the *Saccharomyces cerevisiae* histone gene loci. J Mol Evol 24:252–259

Taylor JW, Berbee ML (2006) Dating divergences in the fungal tree of life: review and new analyses. Mycologia 98:838–849

Van de Peer Y, Fawcett J, Proost S, Sterk L, Vandepoele K (2009) The flowering world: a tale of duplications. Trends Plant Sci 14:680–688

van Hoek MJ, Hogeweg P (2009) Metabolic adaptation after whole genome duplication. Mol Biol Evol 26:2441–2453

van Hoof A (2005) Conserved functions of yeast genes support the duplication, degeneration and complementation model for gene duplication. Genetics 171:1455–1461

Wapinski I, Pfeffer A, Friedman N, Regev A (2007) Natural history and evolutionary principles of gene duplication in fungi. Nature 449:54–61

Werth CR, Windham MD (1991) A model for divergent, allopatric speciation of polyploid pteridophytes resulting from silencing of duplicate-gene expression. Am Nat 137:515–526

Wolfe KH (2000) Robustness-it's not where you think it is. Nat Genet 25:3–4

Wolfe KH, Shields D (1997) Molecular evidence for an ancient duplication of the entire yeast genome. Nature 387:708–713

Xu M, He X (2011) Genetic incompatibility dampens hybrid fertility more than hybrid viability: yeast as a case study. PLoS ONE 6:e18341

Chapter 16
Two Rounds of Whole-Genome Duplication: Evidence and Impact on the Evolution of Vertebrate Innovations

Cristian Cañestro

Abstract The origin and evolution of the vertebrates have been linked to the study of genome duplications since Susumo Ohno ventured the 2R-hypothesis, suggesting that the successful diversification of complex vertebrates was facilitated by polyploidization in the stem vertebrate ancestor due to two rounds of whole-genome duplication (2R-WGD). This chapter first reviews evidence supporting Ohno's 2R-hypothesis and gathers information about the timing and mechanisms underlying the 2R-WGD. Second, this chapter describes the impact of the 2R-WGD on the evolution of the vertebrate genome structure, gene number, and the evolutionary dynamics of the functional fate of vertebrate ohnologs (paralogous genes that originated by WGD) in comparison with non-vertebrate chordate gene homologs. Finally, this review discusses the functional consequences of the 2R-WGD on the origin and evolution of vertebrate innovations compared with urochordates and cephalochordates, paying special attention to the origin of neural crest cells, placodes, and the big complex brain, key features that probably facilitated the transition from ancestral filter-feeding non-vertebrate chordates to voracious vertebrate predators. Currently available data, however, seem to suggest that these putative key features were present to at least some extent in stem Olfactores; hence, the impact of the 2R-WGD may not be related to the immediate origin of vertebrate innovations, but to the subsequent diversification of a wide variety of complex structures that facilitated the successful radiation of vertebrates.

C. Cañestro (✉)
Departament de Genètica, Facultat de Biologia, Universitat de Barcelona,
Av. Diagonal, 643, edifici Prevosti, 2a planta, 08028 Barcelona, Spain
e-mail: canestro@ub.edu

P. S. Soltis and D. E. Soltis (eds.), *Polyploidy and Genome Evolution*,
DOI: 10.1007/978-3-642-31442-1_16,

16.1 Introduction

The study of the origin and evolution of vertebrates, the subphylum to which we belong, has stood at the crossroad between genome evolution and molecular developmental biology since the late 1960s, when Susumo Ohno published his famous work on *Evolution by Gene Duplication* and proposed his hypothesis about the pivotal role of genome duplication in the origin of vertebrates and their diversification (Ohno et al. 1968; Ohno 1970). Vertebrates comprise all animals that have a backbone and include mammals, birds, reptiles, amphibians, fishes, and agnathans—the jawless lampreys and hagfishes. Vertebrates together with urochordates (tunicates) form the Olfactores, which together with cephalochordates (amphioxus or lancelets) constitute the chordates (Fig. 16.1). All chordates share a common basic body plan at least during the larval stage of their life cycle, consisting of a notochord running through a post-anal tail, with a dorsal hollow nerve cord, longitudinal blocks of muscle along the notochord, and ciliated pharyngeal gill slits (Brusca and Brusca 2002). Recent phylogenomic analyses have dethroned cephalochordates from the long-assumed position as sister group of the vertebrates; this position is now occupied by urochordates, which include ascidians, larvaceans and thaliaceans (Wada et al. 2006; Oda et al. 2002; Bourlat et al. 2006; Delsuc et al. 2006; Putnam et al. 2008) (Fig. 16.1).

Ohno's 2R-hypothesis was based on comparative analyses of genome sizes and isozyme complexity among chordate taxa. Ohno found that basally divergent chordate subphyla had smaller genomes and less isozyme complexity than vertebrate lineages. This observation led him to suggest that the combination of tandem gene duplication and in particular an octoploidization event involving two rounds of whole-genome duplication were key to the invertebrate–vertebrate transition, and for the subsequent successful vertebrate diversification (Ohno et al. 1968).

Ohno was one of the pioneers in conceiving the evolutionary significance of whole-genome duplication (see also Chap. 1, this volume). Ohno emphasized the importance of gene duplication as probably the main source of raw genetic material for the evolution of new gene functions (reviewed in Taylor and Raes 2004). In Ohno's classical model, one of the duplicated genes retains the original function whereas its duplicate either disappears by accumulation of detrimental mutations (called pseudogenization or nonfunctionalization) or it is preserved after gaining advantageous mutations that confer positively selected novel functions (neofunctionalization) (Ohno 1970; Nowak et al. 1997; Force et al. 1999). The duplication, degeneration, complementation hypothesis (or DDC model) suggests a third possibility for duplicate gene preservation: subfunctionalization, the complementary partitioning of ancestral structural and/or regulatory subfunctions between two duplicate genes, so that the sum of their functions provides at least that of the original pre-duplication gene (Force et al. 1999). The DDC model predicts that subfunctionalized genes will have lower pleiotropy than the original pre-duplicated gene and lower evolutionary constraints, and thereby will be more permissive to the

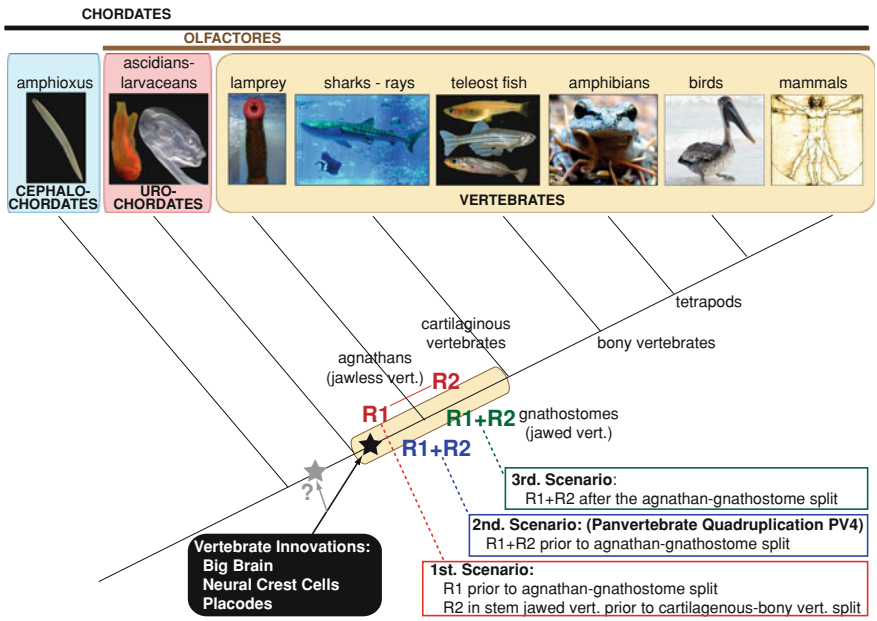

Fig. 16.1 Evolutionary tree of chordates representing the three possible scenarios for the timing of the 2R-WGD and the origins for vertebrate innovations. Evidence suggests that the 2R-WGD might have had a significant impact on the diversification (*black star*) of vertebrate innovations, including structures derived from neural crest cells and placodes, as well as the development of a big complex brain in the stem vertebrate. But whether the 2R-WGD was crucial for the evolutionary origin of these structures remains unclear, and that the hypothesis that the origin of these innovations dated back to at least stem olfactores (*gray star*) cannot be dismissed. (Lamprey picture courtesy of Juan Pascual-Anaya)

accumulation of mutations that might confer novel functions. The acquisition of new functions is favored if the duplication affects the entire genome at once, as opposed to multiple individual gene duplications, because when entire gene networks are duplicated, gene stoichiometry is maintained, and therefore deleterious gene dosage effects can be counteracted (Birchler and Veitia 2007, 2010; Van de Peer et al. 2009; Makino and McLysaght 2010; see also Chap. 2, this volume). Genes that originated by gene duplication are called paralogs. Genes that have been duplicated via genome duplication, however, are a special type of paralogs referred as *ohnologs*, a term suggested by Wolfe (2000) in honor of Ohno's contribution. This term is useful because of the special properties that ohnologs possess at their birth compared to duplications that arise by other local mechanisms such as unequal crossing-over, tandem gene duplication, or retrotransposition.

The sudden creation and evolution of ohnologs by the 2R-WGD that occurred in the stem vertebrate lineage has been suggested as one of the potential key events underlying the increase of morphological complexity, facilitating the acquisition of genetic and developmental innovations of vertebrates (Shimeld and Holland 2000; Aburomia et al. 2003). Genome duplication doubles the number of genes,

many of which have the chance of evolving 'novel' functions that might provide new selectable advantages promoting species diversification (Lynch et al. 2001; Van de Peer et al. 2009). Some vertebrate species that have undergone recent polyploidization, such as the frog *Xenopus laevis* that experienced tetraploidization ~40 million years ago (Hellsten et al. 2007), show a higher adaptability to a variety of different environments, such as drought, salt, cold, and disease resistance, than closely related diploid species, such as *Silurana tropicalis* (for further details on frog polyploidization see also Chap. 18, this volume). Interestingly, only a limited number of retained ohnologs present evidence of neofunctionalization or subfunctionalization in *X. laevis*, suggesting that additional selective mechanisms might act on preserving gene duplicates that could promote species diversification (Chain and Evans 2006; Semon and Wolfe 2008).

In addition to the evolutionary significance of novel fates of duplicate genes on species biology, another mechanism that might contribute to species diversification after genome duplication is reciprocal ohnolog loss between different populations, a concept known as 'divergent resolution' that can lead to reproductive isolation (Lynch and Force 2000). Reciprocal ohnolog loss is likely to occur in the period of relaxed selection that duplicate genes experience while they are functionally redundant (Werth and Windham 1991; Lynch and Conery 2000; Lynch and Force 2000; Scannell et al. 2006; Taylor et al. 2001; Semon and Wolfe 2007). The divergent resolution of gene redundancies, such that one population loses one ohnolog copy while the second population loses the other ohnolog copy, leads to chromosomal restructuring such that gametes produced by hybrid individuals can be completely lacking in functional genes for a duplicate pair. In addition to the isolation due to reciprocal gene losses, this model can be further expanded to isolation due to independent processes of gene duplicate subfunctionalization between two populations, in which hybrids will lack one or more subfunctions (Force et al. 1999; Lynch and Force 2000). Hence, large-scale reciprocal ohnolog loss and independent subfunctionalization of ohnologs can be the cause of reproductive isolation of two populations after polyploidization, favoring genetic divergence of these newly incipient future species. This hypothesis is supported by the analyses of both fish and angiosperm lineages that have undergone polyploidization and include more species diversity (e.g. salmonids, catostomids, eudicots, grasses) than their sister groups that did not go through polyploidization and include a lower number of species (Nelson 1994; Ferris et al. 1979; Soltis et al. 2009). Recent integrated approaches of comparative genomics and gene expression analyses in teleosts, however, provide limited evidence supporting the significance of differential ohnolog loss in reproductive isolation and diversification (Kassahn et al. 2009) (see also Chap. 17, this volume).

Many studies have tackled the central question of whether or not the 2R-WGD had a significant impact on the origin of vertebrate innovations and their subsequent diversification. This chapter first reviews evidence supporting the 2R-hypothesis and information regarding the timing and potential mechanisms underlying the 2R-WGD in vertebrates. Second, this chapter examines the impact that the 2R-WGD may have had on the evolution of vertebrate genome structure, number of

genes, and the fate of retained ohnologs in comparison with non-vertebrate chordate paralogs. Finally, this chapter discusses how the 2R-WGD might have affected the origin and evolution of vertebrate innovations, with special emphasis on the vertebrate big complex brain, and structures derived from neural crest cells and placodes. Recent data, however, suggest that neural crest cells and placodes could already have been present in stem chordates. Hence, the impact of the 2R-WGD cannot be related to the origin of neural crest cells and placodes, but it could be related to their subsequent diversification and development of a wide variety of complex structures.

16.2 Supporting Evidence for the 2R-Hypothesis

If two rounds of genome duplication (2R-WGD) have occurred, we would expect the presence of many paralogs (ohnologs) in conserved, syntenic genomic regions, which are known as paralogons (Coulier et al. 2000) (or ohnologons (Gout et al. 2009)). Paralogons, therefore, consist of series of linked (but frequently functionally and phylogenetically unrelated) genes on one chromosome region, many of which have linked paralogs on at least one other chromosome region. The discovery of paralogy groups made of four paralogons in the genome of human and mouse was interpreted as remnants of the two events of tetraploidization that occurred early during vertebrate evolution and therefore provided the earliest strong evidence supporting Ohno's 2R-hypothesis (Lundin 1979, 1993; Pebusque et al. 1998). Possibly one of the best and first examples of a paralogy group supporting the 2R-hypothesis is the case of the four *Hox*-bearing regions on human chromosomes *Hsa*2, *Hsa*7, *Hsa*12, and *Hsa*17 (Fig. 16.2) (Kappen et al. 1989; Bailey et al. 1997; Larhammar et al. 2002; Lundin et al. 2003). In contrast, only a single *Hox* cluster is present in the cephalochordate amphioxus (Garcia-Fernàndez and Holland 1994). This 4:1 ratio is consistent with Ohno's hypothesis of two tetraploidization events after the cephalochordate–vertebrate split (Holland et al. 1994; Sidow 1996; Garcia-Fernandez 2005). In addition to the *Hox* paralogy group, several other similar examples have been identified (e.g. *MHC* (Katsanis et al. 1996), *Tbx* (Ruvinsky and Silver 1997), *G-protein-coupled* receptors (Fredriksson et al. 2003), *ParaHox* clusters (Ferrier et al. 2005), linked receptor *tyrosine kinases* (Siegel et al. 2007), *endothelin* ligands and receptors (Braasch et al. 2009), *Fox* cluster (Wotton and Shimeld 2006), and the *EGF* ligand paralogons (Laisney et al. 2010)).

Several databases of conserved syntenic chromosomal regions between different species are available, and these provide additional evidence of two rounds of genome duplication. Popovici et al. (2001), for instance, identified 14 paralogons containing more than 1600 genes assembled in a human genome paralogy map (http://u119.marseille.inserm.fr/Db/paralogy.html). The ParaDB (http://abi.marseille.inserm.fr/paradb) predicted that the human genome far exceeds 1000 paralogons that contain more than three pairs of duplicated genes (Leveugle et al. 2003).

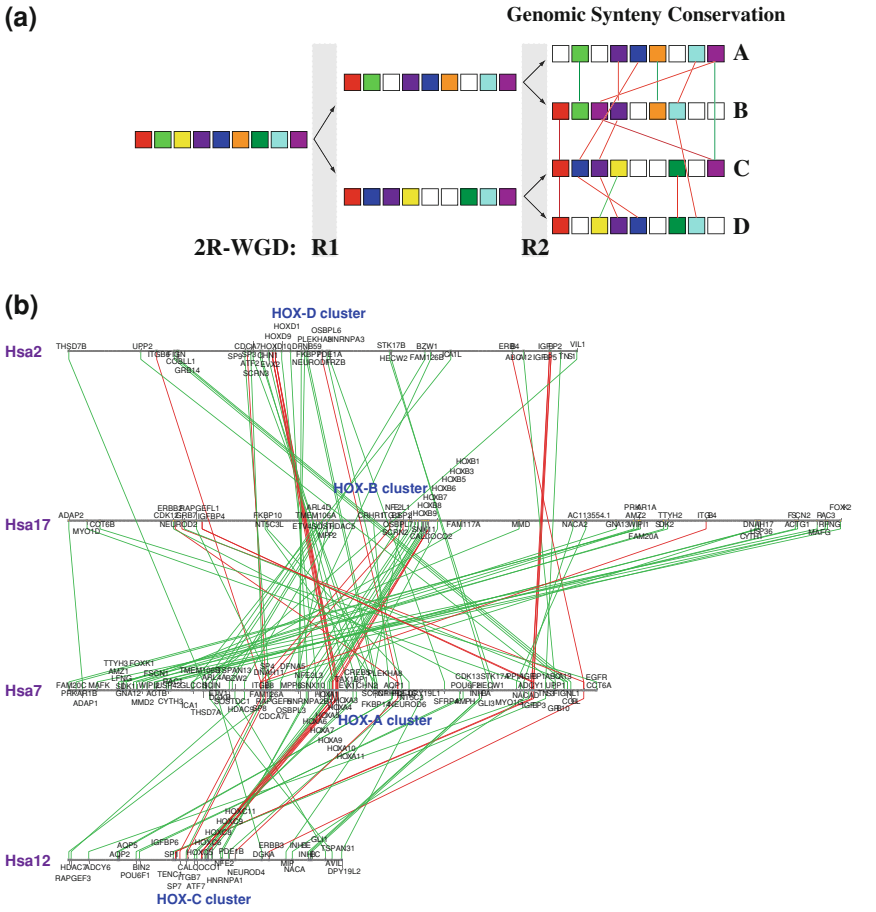

Fig. 16.2 Conserved synteny in the vertebrate genome generated by the 2R-WGD. **a** Simplified representation of a genomic region that has been amplified by the 2R-WGD (R1 and R2) showing conserved genes (*colored boxes*) in four syntenic regions (**a**–**d**), which have suffered genomic rearrangements and gene loss (*white boxes*) and different degrees of conservation (*green and red lines* label ohnologs preserved in two or more than two regions, respectively). **b** Representation of the four human paralogons containing *Hox* A-D clusters in chromosomes *Hsa*2, *Hsa*7, *Hsa*12, and *Hsa*17, displaying high amounts of conserved synteny between ohnologs in two (*green lines*) or at least three (*red lines*) different paralogons. This representation has been generated using a 100-gene sliding window in the Synteny Database (Catchen et al. 2009)

The Genomicus database v60.01 (http://www.dyogen.ens.fr/genomicus) predicted that 18,228 ancestral vertebrate genes were grouped in 2,642 conserved ancestral synteny blocks with a median N50 size of 5 genes (Muffato et al. 2010). The Synteny Database (http://teleost.cs.uoregon.edu/synteny_db) predicted 231 paralogy clusters with more than 5 genes, and 102 paralogy clusters with more than 10 genes, in a

more rigorous count using a 100-gene sliding window and taking amphioxus genes as the outgroup for paralogy assignment in human (Catchen et al. 2009, 2011).

Until recently, however, available evidence did not permit us to discard the possibility that these groups of paralogous genes originated by multiple independent block duplications, rather than two duplications of the entire genome (Skrabanek and Wolfe 1998; Wolfe 2001; Larhammar et al. 2002). An initial hypothesis, based on extensive phylogenetic analysis and dating of the duplications that produced hundreds of vertebrate gene families, proposed a 'big-bang mode' of sudden large-scale gene origin resulting from two waves of gene duplications, rather than the alternative hypothesis of a constant generation by small-scale duplications (Gu et al. 2002). Wave-I was suggested to consist of tandem or segmental duplications that occurred after the mammalian radiation, and wave-II was interpreted as a rapid increase of paralogs in the early stage of vertebrate evolution after their split from non-vertebrate chordates, consistent with one round of whole-genome duplication (Gu et al. 2002). The first analyses of the human genome draft led to the conclusion that the most parsimonious explanation of the current structure of the human paralogons was a 'big-bang' expansion event by a paleopolyploidy that included the whole genome or substantial sections of it. However, no specific evidence was found for two rounds of polyploidy as opposed to one (Venter et al. 2001; McLysaght et al. 2002; Panopoulou et al. 2003).

Recently, however, several analyses have provided definitive support for the 2R-hypothesis (reviewed in Kasahara 2007). Dehal and Boore (2005) developed an elegant, compelling approach to test the 2R-hypothesis by plotting the genomic map position of only those genes that were duplicated prior to the fish–tetrapod split, which rendered a clear global physical pattern of four-way paralogon organization covering most of the human genome. Dehal and Boore's work therefore provided unmistakable evidence of two distinct rounds of genome duplication during early vertebrate evolution.

Furthermore, the recent sequencing of the whole genome of the chordate amphioxus *Branchiostoma floridae* (sister to all other chordates) provided even more indisputable evidence supporting the 2R-hypothesis (Putnam et al. 2008). Despite the fact that small-region comparison between human, chicken, teleost fish, and amphioxus genomes revealed low gene-order conservation at the local level (microsynteny), striking extensive gene linkage conservation was observed when entire chromosomes were considered (macrosynteny). Syntenic analysis reconstructed 17 chordate linkage groups (CLG) that might represent the protochromosomes of the last common chordate ancestor (Putnam et al. 2008). Exhaustive evaluation of the 17 CLGs revealed that most of the human genome (112 segments spanning 2.68 Gb, which is the equivalent of 95 % of the euchromatic genome) was affected by large-scale duplication events that occurred on the stem vertebrate lineage before the teleost/tetrapod split. Analysis of the distribution of the human segments among the 17 CLGs showed that nearly all ancient chordate chromosomes were quadruplicated (Putnam et al. 2008) (Fig. 16.3). This result robustly demonstrated the occurrence of two rounds of genome duplication, corroborating previous lines of evidence based on analysis of

specific regions of interest, such as the *Hox*-bearing regions (Garcia-Fernàndez and Holland 1994; Larhammar et al. 2002) and the major histocompatibility regions (Vienne et al. 2003; Danchin and Pontarotti 2004). Spring (1997) proposed the term "tetralogs" to refer groups of quadruplicated vertebrate genes at four different chromosomal locations formed by the 2R-WGD corresponding to a single invertebrate gene, with all four more similar to each other than to members of the other tetralogy group.

16.2.1 Timing of the Vertebrate 2R-WGD

Analyses of the completely sequenced genome of the cephalochordate amphioxus (Putnam et al. 2008; Holland et al. 2008) and the genomes from various urochordates (Dehal et al. 2002; Small et al. 2007; Denoeud et al. 2010) validated Ohno's hypothesized lower-bound timing for the 2R-WGD as after the split between vertebrates and non-vertebrate chordates. Regarding the upper-bound timing, extensive analysis of gene duplicates (Robinson-Rechavi et al. 2004) and the identification of the four clusters in the genome of the elephant shark suggested that the 2R-WGD took place before the cartilaginous/bony vertebrate split (Venkatesh et al. 2007; Ravi et al. 2009).

Within this time window, the most prevalent hypothesis suggests a scenario in which the first round (R1) occurred before the split between gnathostome and jawless vertebrates, and the second (R2) occurred in the stem jawed vertebrates after their divergence from jawless vertebrates (Fig. 16.1). However, a second scenario proposes that both rounds (R1 + R2) of genome duplication took place before the split between gnathostome and jawless vertebrates (pan-vertebrate quadruplication (PV4) hypothesis (Kuraku et al. 2009)) (Fig. 16.1). Comparative analysis of 55 gene families revealed a common expansion in both jawless and jawed vertebrates, which has been interpreted as evidence supporting this second scenario (Kuraku et al. 2009). Available information from sea lampreys and hagfish does not permit us to discern between these two hypothetical scenarios, because these organisms also appear to have suffered lineage-specific duplications and reciprocal gene losses compared to vertebrates, which together obscure the assessments of orthology/paralogy (reviewed in Kuraku 2008, 2010). For instance, multiple *Hox* gene surveys in different species of sea lampreys and hagfish suggested that extensive independent duplications of *Hox* genes might have occurred during the evolution of jawless vertebrates (Pendleton et al. 1993; Sharman and Holland 1998; Takio et al. 2004; Force et al. 2002; Irvine et al. 2002; Fried et al. 2003; Stadler et al. 2004; Kuraku et al. 2009). Some of the jawless *Hox* clusters might have disintegrated, casting doubt as to the usefulness of *Hox* genes as reliable markers to trace duplications during genome evolution in stem vertebrates (Kuraku 2011). Finally, recent phylogenetic analysis of the degenerated *ParaHox* cluster in hagfish has opened the possibility of a third scenario, in which both rounds (R1 + R2) occurred in stem jawed vertebrates after their divergence

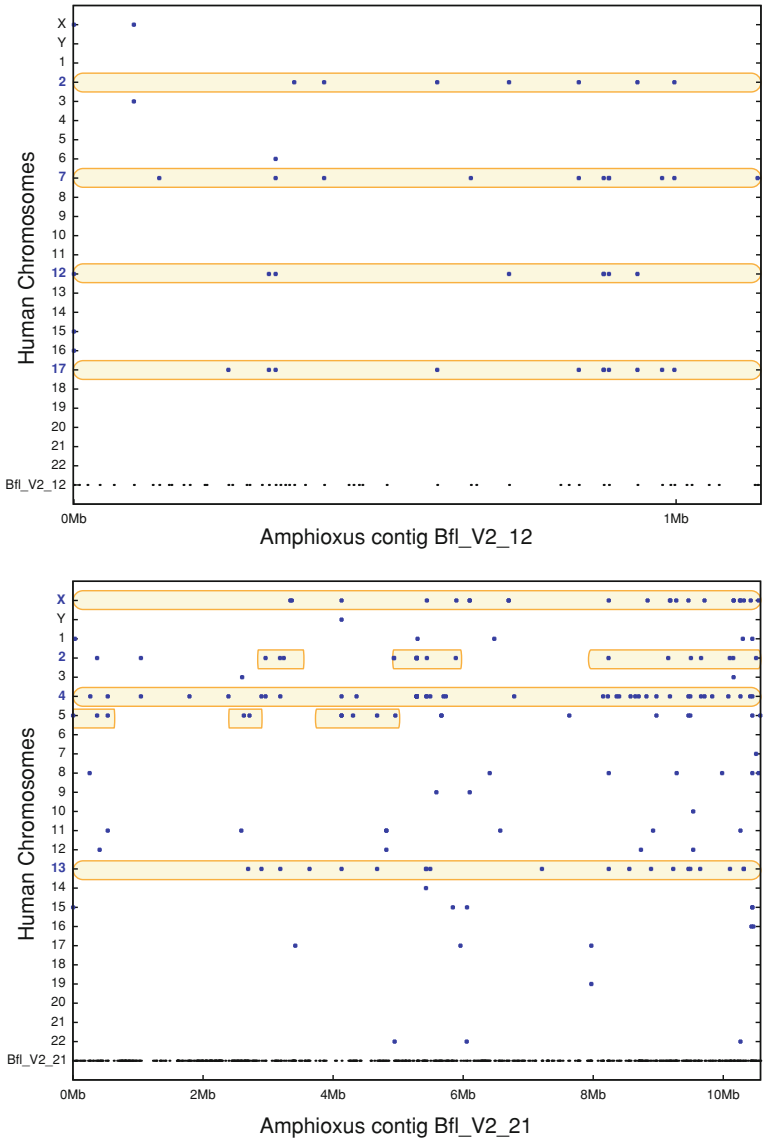

Fig. 16.3 Quadruplicated conserved syntenic pattern between the amphioxus and the human genome as a result of the 2R-WGD. *Dot-plots* display the distribution throughout human chromosomes (y-axes) of human orthologs (*blue dots*) of amphioxus genes (*black dots*) located in two arbitrarily selected genomic regions of approximately 1 Mb (**a**) and 10 Mb (**b**) (x-axes). The *dot-plots* reveal four major human chromosomes (*yellow shadow*) of conserved synteny as the product of the two rounds of whole-genome duplication. In panel (**a**), the four paralogons coincide with the *Hox*-cluster bearing chromosomes 2, 7, 12, and 17, whereas in panel B the four paralogons coincide with the endothelin receptors and *ParaHox*-cluster bearing chromosomes 2/5, 4, 13, and X. The *dot-plots* were generated as described in Canestro et al. (2009) using the Synteny Database (Catchen et al. 2009)

from jawless vertebrates (Furlong et al. 2007) (Fig. 16.1). The validation of this third scenario could have a significant impact on our understanding of vertebrate evolution, because it would imply that the 2R-WGD would have not been important for the origin of vertebrate innovations (i.e. big brain, neural crest cells, and placodes, which clearly exist in jawless vertebrates (Kuratani and Ota 2008; Kuratani 2009)). According to this third scenario, however, the 2R-WGD would have been important for the radiation of gnathostomes into cartilaginous fish, bony fishes, and tetrapods. A solid answer about the timing of the 2R-WGD may have to wait until larger-scale comparisons of the whole-genome organization of hagfish and lampreys are available.

16.2.2 Mechanisms Underlying the Vertebrate 2R-WGD

A question that still remains is how did the stem vertebrate genome become octoploid by two rounds of tetraploidization. There are two main mechanisms of tetraploidization observed in many species of plants and animals (Van de Peer et al. 2009). The first mechanism is allotetraploidy, which occurs when two related but not identical genomes are combined by hybridization of closely related species and associated (often subsequent) genome duplication. In the case of allotetraploidy, the pairs of distinct 'homologous' chromosomes that are sufficiently different due to their separated origin are called *homeologs*. The second mechanism is autotetraploidy, which occurs when the genomes are not sufficiently diverged into homeologous sets; autotetraploidy therefore ranges from the combination of genomes of two conspecific individuals (perhaps from different populations) to the combination of identical genomes from a single individual. The genetic attributes of allo- and autotetraploids differ and may have substantial effects at individual, population, and species levels (see also Chap. 2, this volume). Both allotetraploidy and autotetraploidy could be generated by several processes such as: (i) an abnormal non-disjunction of sister chromatids at meiosis; (ii) the uncoupling of mitotic DNA replication and cell division during early development of the germ line (this process, for instance, occurs normally during the endoreplication of megakaryocytic bone-marrow precursors of blood platelets, or during the development of the oikoblastic epithelia that secrete the house in basal urochordate larvaceans); (iii) potential cell fusion during early embryo development or in germline precursors in syncytial gametogenesis (cell fusion is observed naturally, for instance, in skeletal muscle cells and placenta) (reviewed in Storchova and Pellman 2004; Shemer and Podbilewicz 2000).

In the case of allotetraploids, each pair of homologous chromosomes should segregate normally during meiosis, and genetic interchange between homeologous chromosomes is rare. If two consecutive events of allotetraploidization occurred in stem vertebrates, we would predict that in the ideal situation in the absence of gene losses, a phylogenetic tree of homeologs will render a symmetrical (A,B) (C,D) topology (Furlong and Holland 2002).

In autopolyploids, however, meiotic pairing might occur between any of the four identical chromosomes at meiosis I, facilitating genetic interchanges freely among the four alleles, and leading to 'tetrasomic inheritance'. Eventually the alleles, and chromosomes, might diverge, starting a process of diploidization that reestablishes diploidy. Randomly one of the chromosomes will diverge first and no longer form homologous structures, while the other three will keep pairing until another further divergence. Hence, if two consecutive events of autotetraploidy occurred in quick succession (pseudo-octoploidy) in stem vertebrates, we would predict that in the absence of gene losses, gene family phylogenetic trees will likely render asymmetrical (((A,B),C),D) topologies (Furlong and Holland 2002). Because many vertebrate gene families do render asymmetrical tree topologies (Friedman and Hughes 2001; Hughes 1999; Hughes and Friedman 2003), two quick consecutive events of autotetraploidy have been considered a likely mechanism for the 2R-WGD in stem vertebrates (Furlong and Holland 2002; Lynch and Wagner 2009).

16.3 Consequences of the 2R-WGD on the Evolution of Vertebrate Genome Structure

It has been suggested that polyplodization events, at least in plants, can trigger genomic stress associated with major genomic rearrangements, in many cases mediated by a burst of mobilization of transposable elements (Matzke and Matzke 1998; Comai 2000). Transposable elements can be substrates for unequal and illegitimate recombination and can be responsible for a variety of genome reorganizations associated with the transposition, including chromosomal insertions, deletions, inversions, translocations, and duplications. Lineage-specific genome rearrangements mediated by transposable elements might facilitate rapid evolution, reproductive isolation of different populations, and consequently species diversification (Parisod et al. 2010).

Contrary to possible genome reorganization after polyploidization, as noted above for plants, in stem vertebrates, recent work based on proximate gene pair methods and measurement of syntenic clustering conservation found that the 2R-WGD in vertebrates were not followed by an increase of genome rearrangement (Hufton et al. 2008). Unexpectedly, this work measured massive genome rearrangements prior to the 2R-WGD, which has been interpreted as a pre-existing 'disposition' toward genomic structural change (Hufton et al. 2008). Interestingly, in contrast to the archetypal condition that has been described in the organization of particular genomic regions (e.g. *Hox*-cluster region (Garcia-Fernàndez and Holland 1994)), the amphioxus genome structure is not exceptionally well conserved, evolving its own particular type of repetitive elements (e.g. '*mirage*' minisatellites (Cañestro et al. 2002b; Ebner et al. 2010)), undergoing extensive local tandem gene duplications (see section below), and experiencing a moderate rate of synteny loss similar to that of sea urchin or sea

anemone (Hufton et al. 2008). Therefore, the amphioxus genome structure cannot be considered a fossil genome representing the pre-duplication condition, at least in terms of genome structure (Garcia-Fernàndez et al. 2001; Hufton et al. 2008), although it is far less divergent from the vertebrate genome structure than is any known urochordate genome (Dehal et al. 2002; Denoeud et al. 2010; Louis et al. 2012).

There have been several attempts to infer the karyotype and genome structure from common chordate ancestors and to reconstruct the evolutionary history leading to present chromosome structures. The first comparisons of conserved syntenic associations in different vertebrate karyotypes, using an in silico chromosome painting approach, allowed reconstructions of the ancestral vertebrate genome containing 10–13 ancestral proto-chromosomes (Kohn et al. 2006; Nakatani et al. 2007). Recently, the sequencing of the amphioxus genome has allowed researchers to reconstruct the ancestral chordate genome as consisting of 17 conserved syntenic blocks, which might represent the ancient chordate proto-chromosomes (Putnam et al. 2008).

After the 2R-WGD, under the naive assumption of absence of loss or fusions of chromosomes, we would expect 68 (17×4) proto-vertebrate segments, but parsimonious reconstruction of chromosome history revealed that numerous chromosomal fusions and translocations have occurred. These reconstructions predict at least 20 fusions that led to 37–49 chromosomes in the bony vertebrate ancestor, which became 12 chromosomes in the stem teleost ancestor due to many additional fusions, and 33–45 chromosomes in the stem tetrapod ancestor due to at least 4 fusions shared between human and chicken genomes (Putnam et al. 2008; Naruse et al. 2004; Nakatani et al. 2007). An excellent example of chromosomal rearrangement after the 2R-WGD has been recently provided by a phylogenetic analysis of members of the four *Hox* paralogons that resulted in a (B(A(C,D))) topology. These results suggest that two chromosomal rearrangements between protochromosomes 11 and 4, and 7 and 5 occurred after the clusters duplicated but before the diversification of extant vertebrates 450 million years ago (Lynch and Wagner 2009). These chromosomal rearrangements resolve conflicting data regarding the order of linked genes and support the hypothesis that the 2R-WGD occurred by two consecutive events of autotetraploidy, and thereby the ancestral vertebrate might have been “pseudo-octoploid”. Interestingly, the asymmetrical (B(A(C,D))) topology of the vertebrate *Hox* cluster (Lynch and Wagner 2009) contrasts with the symmetrical (A,B) (C,D) topology inferred from the cartilaginous elephant shark using the amphioxus *Hox* cluster as the outgroup (Ravi et al. 2009). Further extensive analyses including *HoxA-D* clusters from a broader representation of cartilaginous and bony vertebrates will be required to resolve these conflicting topologies, which could suggest that the *Hox*-cluster rearrangement took place after the cartilaginous/bony vertebrate split and not immediately subsequent to the 2R-WGD.

16.4 Consequences of the 2R-WGD on the Evolution of Vertebrate Gene Fate

16.4.1 Function of Gene Duplicates After 2R-WGD

After polyploidization, a period of transilience may follow in which genes might enjoy extra 'degrees of freedom' to mutate without selective penalty (reviewed in Soltis and Soltis 1999; Otto 2007). Understanding the processes by which genome duplication might influence the fate of duplicated genes is crucial to evaluate how the 2R-WGD might have impacted the evolution of vertebrate innovations. Neofunctionalization and subfunctionalization are the two main processes driving the functional fate of newly generated ohnologs after the 2R-WGD and have been extensively discussed in the literature (Hughes 1994; Force et al. 1999; Lynch and Conery 2000; Durand 2003; Postlethwait et al. 2004; Hoekstra and Coyne 2007; Conant and Wolfe 2008; Semon and Wolfe 2007, 2008; Jimenez-Delgado et al. 2009). A prominent example of neofunctionalization related to the 2R-WGD occurred during the expansion of the vertebrate retinoic acid receptor (RAR) family, which acquired new functions in both their expression domains and in their structural protein activities (Escriva et al. 2006). There are also examples, however, in which neofunctionalization and subfunctionalization are related to both 2R-WGD and local tandem duplications (e.g. the expansion of the vertebrate globin superfamily, which promotes the vertebrate innovation related to oxygen transport and storage (Hoffmann et al. 2011)), or merely related to local tandem duplications, and not the 2R-WGD, such as the expansion of the vertebrate Alcohol Dehydrogenase (*Adh*) family, which promotes the acquisition of new enzymes for the synthesis of retinoic acid (Cañestro et al. 2000, 2002a, 2003b). Therefore, not all vertebrate innovations can be exclusively attributed to the 2R-WGD, and the global weight of the impact of the 2R-WGD on the evolution of vertebrate gene functions remains unknown.

16.4.2 Gene Network Rewiring by Tranposons After 2R-WGD

In many cases, neofunctionalization and subfunctionalization can be due to alterations in *cis*-regulatory elements that might lead to adaptative changes in duplicated genes (Force et al. 1999). Many of the *cis*-regulatory elements appear to be embedded in distinct repeat families, especially in transposable elements (TE) (Thornburg et al. 2006; Polak and Domany 2006; Bourque et al. 2008). Analysis of the distribution of 10,000 TEs in the human genome, for instance, revealed that most TEs are concentrated under strong purifying selection near regulatory and developmental genes (Lowe et al. 2007). Most of the described examples of TE mobilization and rewiring of gene regulatory networks have been associated with relatively recent events of TE mobilizations. For instance, TE-mediated rewiring

for neofunctionalization after gene duplication has been recently described for the sex-determining gene *dmrt1bY* in medaka fish, in which a novel regulatory element driving a negative feedback on *dmrt1bY* has been acquired due to the insertion of an *Izanagi* transposon (Herpin et al. 2010), or for the origin of a novel gene regulatory network dedicated to pregnancy in placental mammals, which was due to a transposition of the MER20 TE (Lynch et al. 2011).

A massive expansion of TEs appears, therefore, as a powerful mechanism that could boost a vast redeployment of *cis*-regulatory elements into new gene regulatory networks (Feschotte 2008), promoting large-scale events of neofunctionalization and subfunctionalization (van de Lagemaat et al. 2003; Bennetzen 2005; Bejerano et al. 2006). Polyploidization can trigger the mobilization of transposable elements (Matzke and Matzke 1998; Parisod et al. 2010), because recently duplicated genomes contain many redundant genes and substantial repetitive DNA, which serve as buffer against TE insertional mutagenesis (Matzke et al. 2000). According to this expectation, bursts of TE mobilization have been described after polyploidization in different organisms (Matzke and Matzke 1998; Comai 2000; SanMiguel et al. 1996, 1998).

A question that remains unclear is whether there was or was not a massive TE mobilization after the 2R-WGD that could have favored a significant redeployment of *cis*-regulatory elements into new gene regulatory networks in the stem vertebrate lineage. Recent comparison of the diversity and content of TEs between vertebrates and amphioxus has provided some insights that might help to answer this question (Canestro and Albalat 2012). The dynamics of the TE content within a genome follows a competition model in which the expansion of a particular TE might cause the reduction of other types of TEs, consequently reducing the TE diversity, until a new equilibrium that preserves the functionality of the genome is reached (Abrusán and Krambeck 2006). According to this model, if a massive expansion of TEs occurred after the 2R-WGD, we expect that the diversity of TEs shared among vertebrates should be smaller than in cephalochordates. Consistent with this prediction, a recent comparative study reveals that the shared TE diversity of vertebrates (14 superfamilies in lampreys, 28 in ray-finned fishes, 20 in amphibians, 14 in reptiles, 10 in birds, and 15 in mammals) is lower than the TE diversity in amphioxus (33 superfamilies), which makes plausible the hypothesis that a TE burst could have occurred after the 2R-WGD in the stem vertebrate lineage (Canestro and Albalat 2012). Further comparative genomic analysis between different vertebrates and cephalochordates will be required to test this hypothetical burst of TEs, and especially to evaluate its putative impact on the evolution of gene functions after the 2R-WGD.

16.4.3 Ohnologs Gone Missing After 2R-WGD and Impact on Surviving Ohnologs

While several studies focus on the functional fate of retained gene duplicates, less attention has been paid to how losses of paralogs or ohnologs might impact the evolution of the functions of other genes (reviewed in Cañestro et al. 2007). Loss of one copy of two fully redundant gene duplicates should not usually have significant impact, but loss of one of the paralogs after functional divergence likely has evolutionary consequences. Recent analyses of gene losses by comparative genomics have led to the unexpected finding that significant components of the developmental toolkit might be lost without major changes to the body plan (Cañestro and Postlethwait 2007; Holland 2007), which suggests the presence of compensatory mechanisms or the acquisition of innovations that have preserved unaltered the ancestral condition (Cañestro et al. 2007). Tracing the evolution of gene families throughout ancestral proto-chromosomes using blocs of conserved synteny has become a powerful tool to clarify uncertain phylogenies, to detect ohnologs gone missing (OGM) (Postlethwait 2007; Catchen et al. 2009, 2011), to provide robust assessments of orthology and paralogy between different species, and to discern evolutionary innovations from losses of ancestral features in sister lineages (Canestro et al. 2009).

There are cases in which different ohnologs in different species acquire the same expression pattern, which has been called function shuffling (McClintock et al. 2001) and synfunctionalization (Gitelman 2007), and in some cases the convergence of expression patterns between paralogs can be related to OGM (Postlethwait 2007; Canestro et al. 2009). The evolution of the vertebrate retinaldehyde dehydrogenease *Aldh*1*a* family provides a paradigmatic example of how uncovering the evolution of gene family members through the 2R-WGD has been fundamental to illuminating how gene functions evolve among newly generated paralogs after genome duplications in the face of loss of ohnologs (Canestro et al. 2009). For instance, analysis of conserved synteny revealed that the presence of *Aldh*1*a*1 in tetrapods and its absence in teleost fish was not due to a tetrapod innovation, but to an OGM in the teleost stem lineage, which was accompanied by a re-acquisition of ancestral functions by surviving paralogs (Canestro et al. 2009). Medaka provides a more radical example in which *aldh*1*a*2, the only survivor of the *aldh*1*a* family in this species, recapitulates the expression pattern of all other *aldh*1*a* paralogs that have been lost in medaka. This result is in agreement with a model of functional evolution in which surviving genes re-acquire ancestral gene family roles in the face of loss of ohnologs. Other examples that illustrate the importance of identifying OGMs ohnologs are shown in the *endothelin* and *agouti* systems, in which the exclusive presence of *endothelin* 4 (*edn*4) and the *agouti-signaling protein* 2 genes (*asip*2*a*/*b*) in teleost fish was not due to a fish innovation related to the teleost-specific whole-genome duplication, but instead to a loss of ohnologs that originated in the 2R-WGD in the tetrapod lineage (Braasch et al. 2009; Braasch and Postlethwait 2011). To understand acquisition of functions of vertebrate ohnologs that were generated by

the 2R-WGD, both the impact of the retention of neo- or subfunctionalized ohnologs, as well as the impact of OGM, on the functions of other survivor gene family members should be studied.

16.5 Consequences of the 2R-WGD on Vertebrate Gene Number and Functional Evolution

How many of the genes that were part of the original fourfold increase in genes generated by 2R-WGD in the stem vertebrate have actually survived nonfunctionalization? And importantly, how significant have the functional consequences of those retained genes been for promoting the origin and evolution of vertebrate innovations? Estimates on gene retention in other organisms that have experienced a WGD have reported ~13 % retention over ~100 million years (MY) in yeast (Wolfe and Shields 1997), ~72 % in maize over ~11 MY (Ahn and Tanksley 1993; Gaut and Doebley 1997), and ~77 % in *Xenopus* over ~40 MY (Hellsten et al. 2007). In vertebrates, a ~33 % retention of divergent functional genes after the 2R-WGD over ~500 MY was inferred initially based on theoretical models applied to 270 gene families of the human genome (Nadeau and Sankoff 1997). More recent and broader analyses based in the complete catalog of human ohnologs estimated a rate of retention between 20 and 30 % (Putnam et al. 2008; Huminiecki and Heldin 2010; Makino and McLysaght 2010).

But how can we assess the impact of the 2R-WGD on the origin of vertebrate complex features? A naive approach to estimating this impact could be to perform a comparison of the total number of retained paralogs and their distribution among functional categories in vertebrates and non-vertebrate chordates that did not undergo any WGD since their split from our last common chordate ancestor. Comparison of the gene catalog of the three chordate subphyla (i.e. cephalochordates, urochordates, and vertebrates) has allowed us to identify a lower bound of 8,437 gene families with members that descend from a single gene in the last common chordate ancestor (Putnam et al. 2008). Through subsequent genome or local duplication, these families account for 13,610 amphioxus genes, 13,401 human genes, and 7,216 ascidian genes, the latter being a significantly lower number due to the extensive gene losses that have occurred in urochordate lineages (Dehal et al. 2002; Cañestro et al. 2003a; Edvardsen et al. 2005; Denoeud et al. 2010). Although it is frequently true that the multiple ohnologs of a vertebrate gene family are represented by a single gene in amphioxus, the total number of paralogs derived from a single-copy gene in the last common ancestor is surprisingly similar between amphioxus (13,610) and human (13,401) (Putnam et al. 2008). Therefore the mere total numbers of retained genes after the 2R-WGD duplicates might not be the key to explain the gain of complexity during the evolution of the vertebrate lineage in comparison with amphioxus.

In vertebrates, analysis of the functional categories of the gene families that have expanded after the 2R-WGD revealed that cell signalers and transcriptional

regulators of developmental pathways are generally retained as multiple ohnologs (Roux and Robinson-Rechavi 2008; Putnam et al. 2008; Hufton et al. 2008; Huminiecki and Heldin 2010). Genes associated with basic cellular functions (i.e. translation, replication, splicing, and recombination, with the important exception of cell cycle), however, have been less successfully retained after the 2R-WGD (Huminiecki and Heldin 2010) (although see Gout et al. (2009) for different results in other organisms that have also undergone WGD). Analysis of the human genome reveals that dosage-balance constraints act on the retention of ohnologs, resulting in an enrichment of dosage-balanced genes, an observation predicted following WGD (Birchler and Veitia 2007, 2010) and also reported for other vertebrates, plants, and yeast (e.g., Paterson et al. 2006). Interestingly, many of these retained ohnologs in humans are refractory to copy number variation, have rarely experienced subsequent small-scale duplication, and are frequently associated with diseases related to dosage-imbalance such as down syndrome (Makino and McLysaght 2010). Analysis of retained genes that have originated in vertebrates by local duplications revealed a strong underrepresentation of genes related to cell communication, cell cycle, and embryo development (Huminiecki and Heldin 2010).

In amphioxus, although a thorough analysis of amphioxus-specific gene family expansions has not been performed, Table 16.1 shows an extensive list of amphioxus-specific duplicated genes reported in the literature (this list is probably not complete, and may be biased toward the research with which I am most familiar). This list shows numerous retained duplicates from a broad array of functional categories, including metabolic enzymes, members of transduction and signaling cascades, members of the immunity system, as well as pivotal transcription factors of developmental pathways. Awaiting a more exhaustive analysis, including different amphioxus species to infer the ancestral cephalochordate condition, the list in Table 16.1 shows no obvious bias toward any particular functional category, although it is noticeable that duplicated developmental transcription factors do not account for more than two paralogs (with the exception of the eight hairy amphioxus paralogs (Minguillon et al. 2003)).

Remarkably, the main difference between the newly acquired paralogs in amphioxus and ohnologs in vertebrates is the mechanism of duplication. While approximately 25 % of the ancestral chordate gene families have two or more ancient vertebrate ohnologs generated by the 2R-WGD, there is strong evidence that most amphioxus paralogs originated by local tandem duplications rather than large-scale chromosomal duplications (Table 16.1). Therefore, considering the functional bias of retention of genes duplicated by WGD or local duplication, it is reasonable to speculate that the key influence of the 2R-WGD promoting the successful diversification of vertebrate features resides in the fact that whole networks were duplicated, in contrast to local duplications such as those that occurred in amphioxus, an organism that seems to have maintained morphological and genetic stasis during the last 200 million years (Garcia-Fernàndez and Holland 1994; Cañestro et al. 2002a; Somorjai et al. 2008; Canestro and Albalat 2012; Paps et al. 2012). Duplication of whole gene networks is dosage-balanced and increases the

Table 16.1 List of paralogs originated independently in the amphioxus lineage

Functional category	Gene	Number of amphioxus paralogs	References
Metabolic proteins	*Aldh1a*	5	Cañestro et al. (2006)
	Cyp26	3	Albalat (2009)
	Hsd11β1	22	Albalat et al. (2011)
	Rdh cluster	12	
	Bdh1	14	
	Rdh11/12	22	
	Hsd17β8	6	
	ApoD	5	
	Cyp2	10	
	Cyp11/24/27	8	
	Hsd3β	7	
	Fabp	8	
	Nos	3	Andreakis et al. (2011)
Transduction and signaling cascades	*Opsins*	20	Holland et al. (2008)
	GPR54	18	
	GnRH	3	
	CRHR	5	
	Somatostatine receptors	12	
	LGR7/8	6	
	PTHR	4	
	Secretin receptors	5	
	Estrogen receptors	2	Bridgham et al. (2008)
	NOK	22	D'Aniello et al. (2008)
	ACK	3	
	TIE	7	
	MARTK	8	
	EXTK	47	
	HNK-Ras	2	Bertrand et al. (2009)
	Calmoduline-like	3	Karabinos and Bhattacharya (2000)
Immunity system	*Toll-like receptors*	28	Holland et al. (2008)
	C1q-like	39	
	LRR-containing gene models	406	
Developmental transcription factors	*Hairy*	8	Minguillon et al. (2003)
	Brachyury	2	Holland et al. (1995)
	bHLH	2	Araki et al. (1996)
	Emx	2	Minguillon et al. (2002)
	Hox13–14	2	Ferrier et al. (2000)
	Evx	2	Ferrier et al. (2001)
	Mnx	2	Holland et al. (2008)
	Uncx	2	
	Pou3	2	
	Lhx2/9	2	
	Iro	2	

evolvability to generate novel functions, which in the case of the vertebrate 2R-WGD could have led to an increase in complexity of the signaling and developmental regulatory networks that facilitated the acquisition of innovations.

In addition to the evolutionary role of coding genes in the acquisition of innovations, microRNAs (miRNAs) also play crucial roles during development and have been postulated as important players for the evolution of organismal complexity (Lee et al. 2007; Sempere et al. 2006). Analysis of miRNAs in chordate species showed that the 2R-WGD has increased the diversity of the inventory of miRNAs in vertebrates, which correlated with the increase of complex patterns of tissue specificity of miRNAs (Heimberg et al. 2008; Campo-Paysaa et al. 2011). However, the finding of 41 vertebrate-specific miRNA families, absent in non-vertebrate chordates, suggests that their origin must have occurred in stem vertebrates after their separation from urochordates and is not explained by the 2R-WGD (Heimberg et al. 2008). The appearance of these 41 vertebrate-specific miRNA families has been proposed as a potential key evolutionary force lying behind the dramatic increase of vertebrate complexity (Heimberg et al. 2008). Future exhaustive analysis of the expression patterns of the members of these 41 vertebrate-specific families, and an understanding of their roles, will allow a reevaluation of the importance that this innovation could have had on the origin of vertebrate features.

16.6 Consequences of the 2R-WGD on the Innovation of Vertebrate Features

A major question not resolved yet is the precise impact of the 2R-WGD on the innovation of particular vertebrate features. Three vertebrate features are perhaps the most prominent innovations: derivatives from neural crest cells, sensory organs concentrated in the head derived from ectodermal placodes, and a big complex brain. When taken together, these features probably allowed the transition from ancestral, peaceful, filter-feeding, non-vertebrate chordates to active, voracious, vertebrate predators (Northcutt and Gans 1983; Gans and Northcutt 1983); reviewed in Yu et al. 2008; Holland 2009) (Fig. 16.1).

Vertebrate neural crest cells are a transient population of developmental cells that delaminate at the border of the neural plate through an epithelial–mesenchymal transition, migrate, and differentiate at their final destination into a variety of structures such as sensory neurons, glial cells, peripheral nervous system, pigment cells, smooth muscle cells, connective tissue, cranio-facial cartilage, skeletal bones, and teeth (Weston 1970). Vertebrate crest development depends on four crucial sets of genes that form what is called the neural crest gene regulatory network (NC-GRN): (1) patterning signal genes establish the expression of (2) neural plate border specifier genes, which activate (3) crest specifier genes, which turn on (4) neural crest effector genes that provide differentiated products (Meulemans and Bronner-Fraser 2004, 2005; Ota and Kuratani 2007; Sauka-Spengler et al. 2007). Analysis of the amphioxus genome has revealed the presence of cephalochordate

orthologs from all of these four sets of genes, including (1): *Fgf*, *Wnt*, *Bmp*, *Notch Dlx, AP*2, *SoxB, Zic*, and *islet*; (2): *Pax*3/7, *Msx*, *Dlx*5, and *Zic*; (3): *Snail*, *SoxE, AP*2, *Twist, Id, FoxD*, and *Myc*; and (4): *Rho, cRet, Erbb*3, *Mitf, tyrosinase*, and *tyrosinase-related* genes, with the remarkable exception of the tyrosine kinase *c-Kit* essential for migration and survival of crest cells, and the gene for myelin protein P0, consistent with the notion that the glial myelin sheath is a vertebrate innovation (Meulemans and Bronner-Fraser 2007; Holland et al. 2008; Holland and Short 2008; Nikitina et al. 2009).

The fact that most of the specifier genes are present as single copy in amphioxus, but multiple paralogs in vertebrates, presumably due to the 2R-WGD, has led to the hypothesis that neofunctionalization and subfunctionalization of paralogs may have facilitated the co-option of ancestral genes into the NC-GRN (Sauka-Spengler et al. 2007; Meulemans and Bronner-Fraser 2007; Holland et al. 2008; Holland and Short 2008). Gene ontology (GO) analysis estimates that 91 % of the neural crest genes in vertebrates have been co-opted from genes already present in basal metazoans, while the remaining 9 % of the neural crest genes are vertebrate innovations (Martinez-Morales et al. 2007), including the assembly of new signaling pathways like the endothelin system (Braasch et al. 2009). The evolution of the vertebrate NC-GRN, therefore, appears as the result of a combination of ancestral gene co-option, newly evolving genes, and amplification of these components by the 2R-WGDs (Braasch et al. 2009).

Interestingly, however, the recent description in urochordates of neural crest-like cells that express typical vertebrate crest marker orthologs, migrate, and differentiate into pigments challenges the idea that neural crest cells are a vertebrate innovation (Jeffery et al. 2004, 2008; Jeffery 2007). Thus, it cannot be discounted that some types of neural crest cells might have been present in the last common ancestor of olfactores (urochordates + vertebrates), followed by losses during the significant morphological and genetic simplification suffered by urochordate lineages (Seo et al. 2004; Cañestro et al. 2005; Cañestro and Postlethwait 2007; Holland 2007). Therefore, it seems plausible to consider that the 2R-WGD might have not been crucial for the origin of the neural crest cells, but the 2R-WGD might have been important for increasing the evolvability of the NC-GRN and the diversification of derivative structures.

Similar conclusions have been reached through studies of the gene regulatory network underlying placode and brain development. Analysis of placode-marking genes (e.g. *Eye*, *Pitx*, *Six*, and *Pax*) in ascidian and larvacean urochordates suggested that the last common olfactore ancestor already presented multiple placode derivatives, such as olfactory and adenohypophyseal. Additional and independent proliferation and loss of a variety of placodes probably occurred in both urochordate and vertebrate lineages (Bassham and Postlethwait 2005; Mazet et al. 2005), in some cases recruiting paralogs that had been independently duplicated in both urochordates and vertebrates (Bassham et al. 2008).

Despite the fact that non-vertebrate chordates have a simple brain lacking a midbrain and a midbrain–hindbrain organizer (MHB), most brain-making gene orthologs are present in non-vertebrate chordates, suggesting that vertebrate brain

features were built on a foundation already present in the ancestral chordate probably facilitated by the new ohnologs created by the vertebrate 2R-WGD (reviewed in Holland 2009). Recent analysis of developmental genes in the ascidian brain revealed that the expression of *Fgf*8 can reorganize the expression of other brain genes and transform hindbrain structures into an expanded mesencephelon, recapitulating the organizing activity of the vertebrate MHB and therefore suggesting that the MHB was already present at least in the last common ancestor of olfactores (Imai et al. 2009). Analysis of urochordate genomes revealed that important genes (i.e. *Gbx*) for the positioning of the MHB have been lost in stem urochordates (Cañestro et al. 2005), as has the retinoic acid dependent anterior–posterior axial patterning of the central nervous system (Cañestro et al. 2006), making plausible the hypothesis that the absence of midbrain in urochordates is not due to a vertebrate innovation of a midbrain, but a simplification in urochordates of an ancestral tripartite brain structure (Cañestro and Postlethwait 2007; Cañestro et al. 2007). Evolutionary analysis of the origin of the complex Nova-regulated splice variants of the vertebrate brain genes revealed that many of these variants were already present in the last common olfactore ancestor (Irimia et al. 2011). It is possible that the 2R-WGD promoted the increase of the complexity of Nova-dependent splice variants in the vertebrate brain, although a simplification of this system during urochordate evolution cannot be discarded.

In conclusion, it is likely that the origin of vertebrate features such as neural crest cells, placodes, and a complex tripartite brain are not related to the 2R-WGD, but that these features were already present to some extent in stem non-vertebrate chordates (Fig. 16.1) (reviewed in Donoghue et al. 2008). However, it is likely that the subsequent evolution of these three features has been strongly influenced by the new ohnologs that originated after the 2R-WGD, due to processes of neofunctionalization, subfunction partitioning and subsequent refinement, recruitment of *cis*-regulatory elements driven by genome rearrangement and transposable element activity, inventions of novel miRNA families, and evolution of novel splice variants, which overall increased the complexity of duplicated developmental gene regulatory networks after the 2R-WGD. Future integrative analysis of comparative genomics, functional evo-devo, and examinations of gene regulatory networks in a wide variety of non-vertebrate chordates as well as basally divergent jawless vertebrates will help to narrow down the precise timing of the 2R-WGD and evaluate its actual impact on the origin and evolution of vertebrate features.

Probably the new '2R, or not 2R' question (Hughes and Friedman 2003) is now to ascertain whether the origins of vertebrate innovations were, or were not, the consequence of the 2R-WGD, and to understand the mechanisms by which the 2R-WGD increased the evolvability of developmental gene regulatory networks that facilitated the diversification of complex vertebrate features.

Acknowledgments This work has been supported by grant BFU2010-14875 from the Ministerio de Ciencia e Innovación (Spain). I would like to thank Julian Catchen for his generous support on the Synteny Database, and Ingo Braasch, John H. Postlethwait, Ricard Albalat, and Adriana Rodriguez for their valuable comments on the chapter, and cheerful and endless discussions on "2R, or not 2R, that is the question … on vertebrate innovations".

References

Abrusán G, Krambeck H-J (2006) Competition may determine the diversity of transposable elements. Theor Popul Biol 70(3):364–375

Aburomia R, Khaner O, Sidow A (2003) Functional evolution in the ancestral lineage of vertebrates or when genomic complexity was wagging its morphological tail. J Struct Funct Genomics 3(1–4):45–52

Ahn S, Tanksley SD (1993) Comparative linkage maps of the rice and maize genomes. Proc Natl Acad Sci U S A 90(17):7980–7984

Albalat R (2009) The retinoic acid machinery in invertebrates: ancestral elements and vertebrate innovations. Mol Cell Endocrinol 313(1–2):23–35

Albalat R, Brunet F, Laudet V, Schubert M (2011) Evolution of retinoid and steroid signaling: vertebrate diversification from an amphioxus perspective. Genome Biol Evol 3:985–1005

Andreakis N, D'Aniello S, Albalat R, Patti FP, Garcia-Fernandez J, Procaccini G, Sordino P, Palumbo A (2011) Evolution of the nitric oxide synthase family in metazoans. Mol Biol Evol 28(1):163–179

Araki I, Terazawa K, Satoh N (1996) Duplication of an amphioxus myogenic bHLH gene is independent of vertebrate myogenic bHLH gene duplication. Gene 171(2):231–236

Bailey W, Kim J, Wagner G, Ruddle F (1997) Phylogenetic reconstruction of vertebrate *Hox* cluster duplications. Mol Biol Evol 14:843–853

Bassham S, Cañestro C, Postlethwait JH (2008) Evolution of developmental roles of Pax2/5/8 paralogs after independent duplication in urochordate and vertebrate lineages. BMC Biol 6:35

Bassham S, Postlethwait JH (2005) The evolutionary history of placodes: a molecular genetic investigation of the larvacean urochordate *Oikopleura dioica*. Development 132(19):4259–4272

Bejerano G, Lowe CB, Ahituv N, King B, Siepel A, Salama SR, Rubin EM, Kent WJ, Haussler D (2006) A distal enhancer and an ultra conserved exon are derived from a novel retroposon. Nature 441(7089):87–90

Bennetzen JL (2005) Transposable elements, gene creation and genome rearrangement in flowering plants. Curr Opin Genet Dev 15(6):621–627

Bertrand S, Campo-Paysaa F, Camasses A, Garcia-Fernandez J, Escriva H (2009) Actors of the tyrosine kinase receptor downstream signaling pathways in amphioxus. Evol Dev 11(1):13–26

Birchler JA, Veitia RA (2007) The gene balance hypothesis: from classical genetics to modern genomics. Plant Cell 19(2):395–402

Birchler JA, Veitia RA (2010) The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. New Phytol 186(1):54–62

Bourlat SJ, Juliusdottir T, Lowe CJ, Freeman R, Aronowicz J, Kirschner M, Lander ES, Thorndyke M, Nakano H, Kohn AB, Heyland A, Moroz LL, Copley RR, Telford MJ (2006) Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. Nature 444(7115):85–88

Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG, Chew JL, Ruan Y, Wei CL, Ng HH, Liu ET (2008) Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Res 18(11):1752–1762

Braasch I, Postlethwait JH (2011) The teleost agouti-related protein 2 gene is an ohnolog gone missing from the tetrapod genome. Proc Natl Acad Sci U S A 108(13):E47–E48

Braasch I, Volff JN, Schartl M (2009) The endothelin system: evolution of vertebrate-specific ligand–receptor interactions by three rounds of genome duplication. Mol Biol Evol 26(4):783–799

Bridgham JT, Brown JE, Rodriguez-Mari A, Catchen JM, Thornton JW (2008) Evolution of a new function by degenerative mutation in cephalochordate steroid receptors. PLoS Genet 4(9):e1000191

Brusca RC, Brusca GJ (2002) Invertebrates. Sinauer Associates, Sunderland

Campo-Paysaa F, Semon M, Cameron RA, Peterson KJ, Schubert M (2011) microRNA complements in deuterostomes: origin and evolution of microRNAs. Evol Dev 13(1):15–27

Cañestro C, Albalat R (2012) Transposon diversity is higher in amphioxus than in vertebrates: functional and evolutionary inferences. Brief Funct Genomics 11(2):131–141

Cañestro C, Albalat R, Hjelmqvist L, Godoy L, Jornvall H, Gonzalez-Duarte R (2002a) Ascidian and amphioxus Adh genes correlate functional and molecular features of the ADH family expansion during vertebrate evolution. J Mol Evol 54(1):81–89

Cañestro C, Bassham S, Postlethwait JH (2003a) Seeing chordate evolution through the *Ciona* genome sequence. Genome Biol 4(3):208–211

Cañestro C, Bassham S, Postlethwait JH (2005) Development of the central nervous system in the larvacean *Oikopleura dioica* and the evolution of the chordate brain. Dev Biol 285(2):298–315

Cañestro C, Catchen JM, Rodriguez-Mari A, Yokoi H, Postlethwait JH (2009) Consequences of lineage-specific gene loss on functional evolution of surviving paralogs: ALDH1A and retinoic acid signaling in vertebrate genomes. PLoS Genet 5(5):e1000496

Cañestro C, Godoy L, Gonzàlez-Duarte R, Albalat R (2003b) Comparative expression analysis of Adh3 during arthropod, urochordate, cephalochordate and vertebrate development challenges its predicted housekeeping role. Evol Dev 5(2):157–162

Cañestro C, Gonzàlez-Duarte R, Albalat R (2002b) Minisatellite instability at the Adh locus reveals somatic polymorphism in amphioxus. Nucleic Acids Res 30(13):2871–2876

Cañestro C, Hjelmqvist L, Albalat R, Garcia-Fernàndez J, Gonzàlez-Duarte R, Jörnvall H (2000) Amphioxus alcohol dehydrogenase is a class 3 form of single type and of structural conservation but with unique developmental expression. Eur J Biochem 267:6511–6518

Cañestro C, Postlethwait JH (2007) Development of a chordate anterior–posterior axis without classical retinoic acid signaling. Dev Biol 305(2):522–538

Cañestro C, Postlethwait JH, Gonzàlez-Duarte R, Albalat R (2006) Is retinoic acid genetic machinery a chordate innovation? Evol Dev 8(5):394–406

Cañestro C, Yokoi H, Postlethwait JH (2007) Evolutionary developmental biology and genomics. Nat Rev Genet 8(12):932–942

Catchen JM, Braasch I, Postlethwait JH (2011) Conserved synteny and the zebrafish genome. Methods Cell Biol 104:259–285

Catchen JM, Conery JS, Postlethwait JH (2009) Automated identification of conserved synteny after whole-genome duplication. Genome Res 19(8):1497–1505

Chain FJ, Evans BJ (2006) Multiple mechanisms promote the retained expression of gene duplicates in the tetraploid frog *Xenopus laevis*. PLoS Genet 2(4):e56

Comai L (2000) Genetic and epigenetic interactions in allopolyploid plants. Plant Mol Biol 43 (2–3):387–399

Conant GC, Wolfe KH (2008) Turning a hobby into a job: how duplicated genes find new functions. Nat Rev Genet 9(12):938–950

Coulier F, Popovici C, Villet R, Birnbaum D (2000) MetaHox gene clusters. J Exp Zool 288(4):345–351

D'Aniello S, Irimia M, Maeso I, Pascual-Anaya J, Jimenez-Delgado S, Bertrand S, Garcia-Fernandez J (2008) Gene expansion and retention leads to a diverse tyrosine kinase superfamily in amphioxus. Mol Biol Evol 25(9):1841–1854

Danchin EG, Pontarotti P (2004) Towards the reconstruction of the bilaterian ancestral pre-MHC region. Trends Genet 20(12):587–591

Dehal P, Boore JL (2005) Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol 3(10):e314

Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, De Tomoso A, Davidson B, Di Gregorio A, Gelpke M, Goodstein DM, Harafuji N, Hastings KEM, Ho I, Hotta K, Huang W, Kawashima T, Lemaire P, Martinez D, Meinertzhagen IA, Necula S, Nonaka M, Putnam N, Rash S, Saiga H, Satake M, Terry A, Yamada L, Wang H-G, Awazu S, Azumi K, Boore J, Branno M, Chin-bow S, DeSantis R, Doyle S, Francino P, Keys DN, Haga S, Hayashi H, Hino K, Imai KS, Inaba K, Kano S, Kobayashi K, Kobayashi M, Lee B-I, Makabe KW, Manohar C, Matassi G, Medina M, Mochizuki Y, Mount S, Morishita T, Miura S, Nakayama A, Nishizaka S, Nomoto H, Ohta F, Oishi K, Rigoutsos I, Sano M, Sasaki A, Sasakura Y, Shoguchi E, Shin-i T, Spagnuolo A, Stainier D, Suzuki MM, Tassy O, Takatori N, Tokuoka M, Yagi K, Yoshizaki F, Wada S, Zhang C, Hyatt PD, Larimer F, Detter C, Doggett N, Glavina T, Hawkins T, Richardson P, Lucas S, Kohara Y, Levine M, Satoh N, Rokhsar DS (2002) The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. Science 298(5601):2157–2167

Delsuc F, Brinkmann H, Chourrout D, Philippe H (2006) Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature 439(7079):965–968

Denoeud F, Henriet S, Mungpakdee S, Aury JM, Da Silva C, Brinkmann H, Mikhaleva J, Olsen LC, Jubin C, Cañestro C, Bouquet JM, Danks G, Poulain J, Campsteijn C, Adamski M, Cross I, Yadetie F, Muffato M, Louis A, Butcher S, Tsagkogeorga G, Konrad A, Singh S, Jensen MF, Cong EH, Eikeseth-Otteraa H, Noel B, Anthouard V, Porcel BM, Kachouri-Lafond R, Nishino A, Ugolini M, Chourrout P, Nishida H, Aasland R, Huzurbazar S, Westhof E, Delsuc F, Lehrach H, Reinhardt R, Weissenbach J, Roy SW, Artiguenave F, Postlethwait JH, Manak JR, Thompson EM, Jaillon O, Du Pasquier L, Boudinot P, Liberles DA, Volff JN, Philippe H, Lenhard B, Roest Crollius H, Wincker P, Chourrout D (2010) Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate. Science 330(6009):1381–1385

Donoghue PC, Graham A, Kelsh RN (2008) The origin and evolution of the neural crest. BioEssays 30(6):530–541

Durand D (2003) Vertebrate evolution: doubling and shuffling with a full deck. Trends Genet 19(1):2–5

Ebner B, Panopoulou G, Vinogradov SN, Kiger L, Marden MC, Burmester T, Hankeln T (2010) The globin gene family of the cephalochordate amphioxus: implications for chordate globin evolution. BMC Evol Biol 10:370

Edvardsen RB, Seo HC, Jensen MF, Mialon A, Mikhaleva J, Bjordal M, Cartry J, Reinhardt R, Weissenbach J, Wincker P, Chourrout D (2005) Remodelling of the homeobox gene complement in the tunicate *Oikopleura dioica*. Curr Biol 15(1):R12–R13

Escriva H, Bertrand S, Germain P, Robinson-Rechavi M, Umbhauer M, Cartry J, Duffraisse M, Holland L, Gronemeyer H, Laudet V (2006) Neofunctionalization in vertebrates: the example of retinoic acid receptors. PLoS Genet 2(7):e102

Ferrier DE, Dewar K, Cook A, Chang JL, Hill-Force A, Amemiya C (2005) The chordate *ParaHox* cluster. Curr Biol 15(20):R820–R822

Ferrier DE, Minguillon C, Holland PW, Garcia-Fernandez J (2000) The amphioxus *Hox* cluster: deuterostome posterior flexibility and *Hox*14. Evol Dev 2(5):284–293

Ferrier DEK, Minguillón C, Cebrián C, Garcia-Fernàndez J (2001) Amphioxus *Evx* genes: implications for the evolution of the midbrain–hindbrain boundary and the chordate tailbud. Dev Biol 237(2):270–281

Ferris SD, Portnoy SL, Whitt GS (1979) The roles of speciation and divergence time in the loss of duplicate gene expression. Theor Popul Biol 15(1):114–139

Feschotte C (2008) Transposable elements and the evolution of regulatory networks. Nat Rev Genet 9(5):397–405

Force A, Amores A, Postlethwait JH (2002) *Hox* cluster organization in the jawless vertebrate *Petromyzon marinus*. J Exp Zool 294(1):30–46

Force A, Lynch M, Pickett FB, Amores A, Yan Y-L, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151(4):1531–1545

Fredriksson R, Lagerstrom MC, Lundin LG, Schioth HB (2003) The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. Mol Pharmacol 63(6):1256–1272

Fried C, Prohaska SJ, Stadler PF (2003) Independent *Hox*-cluster duplications in lampreys. J Exp Zoolog Part B Mol Dev Evol 299(1):18–25

Friedman R, Hughes AL (2001) Pattern and timing of gene duplication in animal genomes. Genome Res 11(11):1842–1847

Furlong R, Holland P (2002) Were vertebrates octoploid? Philos T Roy Soc B 357:531–544

Furlong RF, Younger R, Kasahara M, Reinhardt R, Thorndyke M, Holland PW (2007) A degenerate *ParaHox* gene cluster in a degenerate vertebrate. Mol Biol Evol 24(12): 2681–2686

Gans C, Northcutt RG (1983) Neural crest and the origin of vertebrates: a new head. Science 220:268–274

Garcia-Fernandez J (2005) The genesis and evolution of homeobox gene clusters. Nat Rev Genet 6(12):881–892

Garcia-Fernàndez J, Ferrier DE, Minguillón C, Cebrián C (2001) The amphioxus genome in evo-devo: archetype or "cul de sac". Int J Dev Biol 45(S1):S137–S138

Garcia-Fernàndez J, Holland PW (1994) Archetypal organization of the amphioxus *Hox* gene cluster. Nature 370(6490):563–566

Gaut BS, Doebley JF (1997) DNA sequence evidence for the segmental allotetraploid origin of maize. Proc Natl Acad Sci U S A 94(13):6809–6814

Gitelman I (2007) Evolution of the vertebrate twist family and synfunctionalization: a mechanism for differential gene loss through merging of expression domains. Mol Biol Evol 24(9): 1912–1925

Gout JF, Duret L, Kahn D (2009) Differential retention of metabolic genes following whole-genome duplication. Mol Biol Evol 26(5):1067–1072

Gu X, Wang Y, J G (2002) Age distribution of human gene families shows significant roles of both large- and small-scale duplications in vertebrate evolution. Nat Genet 31:205–209

Heimberg AM, Sempere LF, Moy VN, Donoghue PC, Peterson KJ (2008) microRNAs and the advent of vertebrate morphological complexity. Proc Natl Acad Sci U S A 105(8):2946–2950

Hellsten U, Khokha MK, Grammer TC, Harland RM, Richardson P, Rokhsar DS (2007) Accelerated gene evolution and subfunctionalization in the pseudotetraploid frog *Xenopus laevis*. BMC Biol 5:31

Herpin A, Braasch I, Kraeussling M, Schmidt C, Thoma EC, Nakamura S, Tanaka M, Schartl M (2010) Transcriptional rewiring of the sex determining dmrt1 gene duplicate by transposable elements. PLoS Genet 6(2):e1000844

Hoekstra HE, Coyne JA (2007) The locus of evolution: evo devo and the genetics of adaptation. Evolution 61(5):995–1016

Hoffmann FG, Opazo JC, Storz JF (2012) Whole-genome duplications spurred the functional diversification of the globin gene superfamily in vertebrates. Mol Biol Evol 29:303−312

Holland LZ (2007) Developmental biology: a chordate with a difference. Nature 447(7141):153–155

Holland LZ (2009) Chordate roots of the vertebrate nervous system: expanding the molecular toolkit. Nat Rev Neurosci 10(10):736–746

Holland LZ, Albalat R, Azumi K, Benito-Gutierrez E, Blow MJ, Bronner-Fraser M, Brunet F, Butts T, Candiani S, Dishaw LJ, Ferrier DE, Garcia-Fernandez J, Gibson-Brown JJ, Gissi C, Godzik A, Hallbook F, Hirose D, Hosomichi K, Ikuta T, Inoko H, Kasahara M, Kasamatsu J, Kawashima T, Kimura A, Kobayashi M, Kozmik Z, Kubokawa K, Laudet V, Litman GW, McHardy AC, Meulemans D, Nonaka M, Olinski RP, Pancer Z, Pennacchio LA, Pestarino M, Rast JP, Rigoutsos I, Robinson-Rechavi M, Roch G, Saiga H, Sasakura Y, Satake M, Satou Y, Schubert M, Sherwood N, Shiina T, Takatori N, Tello J, Vopalensky P, Wada S, Xu A, Ye Y, Yoshida K, Yoshizaki F, Yu JK, Zhang Q, Zmasek CM, de Jong PJ, Osoegawa K, Putnam

NH, Rokhsar DS, Satoh N, Holland PW (2008) The amphioxus genome illuminates vertebrate origins and cephalochordate biology. Genome Res 18(7):1100–1111
Holland LZ, Short S (2008) Gene duplication, co-option and recruitment during the origin of the vertebrate brain from the invertebrate chordate brain. Brain Behav Evol 72(2):91–105
Holland PWH, Garcia-Fernàndez J, Williams NA, Sidow A (1994) Gene duplications and the origins of vertebrate development. Development (Suppl.):125–133
Holland PWH, Koschorz B, Holland LZ, Herrmann BG (1995) Conservation of *Brachyury* (*T*) genes in amphioxus and vertebrates: developmental and evolutionary implications. Development 121:4283–4291
Hufton AL, Groth D, Vingron M, Lehrach H, Poustka AJ, Panopoulou G (2008) Early vertebrate whole genome duplications were predated by a period of intense genome rearrangement. Genome Res 18(10):1582–1591
Hughes AL (1994) The evolution of functionally novel proteins after gene duplication. Proc Roy Soc Lond B 256:119–124
Hughes AL (1999) Phylogenies of developmentally important proteins do not support the hypothesis of two rounds of genome duplication early in vertebrate history. J Mol Evol 48(5):565–576
Hughes AL, Friedman R (2003) 2R or not 2R: testing hypotheses of genome duplication in early vertebrates. J Struct Funct Genomics 3(1–4):85–93
Huminiecki L, Heldin CH (2010) 2R and remodeling of vertebrate signal transduction engine. BMC Biol 8:146
Imai KS, Stolfi A, Levine M, Satou Y (2009) Gene regulatory networks underlying the compartmentalization of the Ciona central nervous system. Development 136(2):285–293
Irimia M, Denuc A, Burguera D, Somorjai I, Martin-Duran JM, Genikhovich G, Jimenez-Delgado S, Technau U, Roy SW, Marfany G, Garcia-Fernandez J (2011) Stepwise assembly of the Nova-regulated alternative splicing network in the vertebrate brain. Proc Natl Acad Sci U S A 108(13):5319–5324
Irvine SQ, Carr JL, Bailey WJ, Kawasaki K, Shimizu N, Amemiya CT, Ruddle FH (2002) Genomic analysis of *Hox* clusters in the sea lamprey *Petromyzon marinus*. J Exp Zool 294:47–62
Jeffery WR (2007) Chordate ancestry of the neural crest: new insights from ascidians. Semin Cell Dev Biol 18:481−491
Jeffery WR, Chiba T, Krajka FR, Deyts C, Satoh N, Joly JS (2008) Trunk lateral cells are neural crest-like cells in the ascidian *Ciona intestinalis*: insights into the ancestry and evolution of the neural crest. Dev Biol 324(1):152–160
Jeffery WR, Strickler AG, Yamamoto Y (2004) Migratory neural crest-like cells form body pigmentation in a urochordate embryo. Nature 431(7009):696–699
Jimenez-Delgado S, Pascual-Anaya J, Garcia-Fernandez J (2009) Implications of duplicated cis-regulatory elements in the evolution of metazoans: the DDI model or how simplicity begets novelty. Brief Funct Genomic Proteomic 8(4):266–275
Kappen C, Schughart K, Ruddle FH (1989) Two steps in the evolution of Antennapedia-class vertebrate homeobox genes. Proc Natl Acad Sci U S A 86(14):5459–5463
Karabinos A, Bhattacharya D (2000) Molecular evolution of calmodulin and calmodulin-like genes in the cephalochordate *Branchiostoma*. J Mol Evol 51(2):141–148
Kasahara M (2007) The 2R hypothesis: an update. Curr Opin Immunol 19(5):547–552
Kassahn KS, Dang VT, Wilkins SJ, Perkins AC, Ragan MA (2009) Evolution of gene function and regulatory control after whole-genome duplication: comparative analyses in vertebrates. Genome Res 19(8):1404–1418
Katsanis N, Fitzgibbon J, Fisher EMC (1996) Paralogy mapping: identification of a region in the human MHC triplicated onto human chromosomes 1 and 9 allows the prediction and isolation of novel *PBX* and *NOTCH* loci. Genomics 35:101–108
Kohn M, Hogel J, Vogel W, Minich P, Kehrer-Sawatzki H, Graves JA, Hameister H (2006) Reconstruction of a 450-My-old ancestral vertebrate protokaryotype. Trends Genet 22(4):203–210

Kuraku S (2008) Insights into cyclostome phylogenomics: pre-2R or post-2R. Zoolog Sci 25(10):960–968
Kuraku S (2010) Palaeophylogenomics of the vertebrate ancestor—impact of hidden paralogy on hagfish and lamprey gene phylogeny. Integr Comp Biol 50(1):124–129
Kuraku S (2011) *Hox* gene clusters of early vertebrates: do they serve as reliable markers for genome evolution? Genomics Proteomics Bioinformatics 9(3):97–103
Kuraku S, Meyer A, Kuratani S (2009) Timing of genome duplications relative to the origin of the vertebrates: did cyclostomes diverge before or after? Mol Biol Evol 26(1):47–59
Kuratani S (2009) Insights into neural crest migration and differentiation from experimental embryology. Development 136(10):1585–1589
Kuratani S, Ota KG (2008) Hagfish (cyclostomata, vertebrata): searching for the ancestral developmental plan of vertebrates. BioEssays 30(2):167–172
Laisney JA, Braasch I, Walter RB, Meierjohann S, Schartl M (2010) Lineage-specific co-evolution of the Egf receptor/ligand signaling system. BMC Evol Biol 10:27
Larhammar D, Lundin L, Hallbook F (2002) The human *Hox*-bearing chromosome regions did arise by block or chromosome (or even genome) duplications. Genome Res 12(12): 1910–1920
Lee CT, Risom T, Strauss WM (2007) Evolutionary conservation of microRNA regulatory circuits: an examination of microRNA gene complexity and conserved microRNA-target interactions through metazoan phylogeny. DNA Cell Biol 26(4):209–218
Leveugle M, Prat K, Perrier N, Birnbaum D, Coulier F (2003) ParaDB: a tool for paralogy mapping in vertebrate genomes. Nucleic Acids Res 31(1):63–67
Louis A, Roest Crollius H, Robinson-Rechavi M (2012) How much does the amphioxus genome represent the ancestor of chordates? Brief Funct Genomics 11(2):89–95
Lowe CB, Bejerano G, Haussler D (2007) Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. Proc Natl Acad Sci U S A 104(19):8005–8010
Lundin LG (1979) Evolutionary conservation of large chromosomal segments reflected in mammalian gene maps. Clin Genet 16(2):72–81
Lundin LG (1993) Evolution of the vertebrate genome as relected in paralogous chromosomal regions in man and the house mouse. Genomics 16:1–19
Lundin LG, Larhammar D, Hallbook F (2003) Numerous groups of chromosomal regional paralogies strongly indicate two genome doublings at the root of the vertebrates. J Struct Funct Genomics 3(1–4):53–63
Lynch M, Conery J (2000) The evolutionary fate and consequences of gene duplication. Science 290(5494):1151–1155
Lynch M, Force A (2000) The origin of interspecific genomic incompatibility via gene duplication. Am Nat 156:590–605
Lynch M, O'Hely M, Walsh B, Force A (2001) The probability of preservation of a newly arisen gene duplicate. Genetics 159(4):1789–1804
Lynch VJ, Leclerc RD, May G, Wagner GP (2011) Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. Nat Genet 43:1154−1159
Lynch VJ, Wagner GP (2009) Multiple chromosomal rearrangements structured the ancestral vertebrate *Hox*-bearing protochromosomes. PLoS Genet 5(1):e1000349
Makino T, McLysaght A (2010) Ohnologs in the human genome are dosage balanced and frequently associated with disease. Proc Natl Acad Sci U S A 107(20):9270–9274
Martinez-Morales JR, Henrich T, Ramialison M, Wittbrodt J (2007) New genes in the evolution of the neural crest differentiation program. Genome Biol 8(3):R36
Matzke MA, Matzke AJ (1998) Polyploidy and transposons. Trends Ecol Evol 13(6):241
Matzke MA, Mette MF, Matzke AJ (2000) Transgene silencing by the host genome defense: implications for the evolution of epigenetic control mechanisms in plants and vertebrates. Plant Mol Biol 43(2–3):401–415

Mazet F, Hutt JA, Milloz J, Millard J, Graham A, Shimeld SM (2005) Molecular evidence from *Ciona intestinalis* for the evolutionary origin of vertebrate sensory placodes. Dev Biol 282(2):494–508

McClintock JM, Carlson R, Mann DM, Prince VE (2001) Consequences of *Hox* gene duplication in the vertebrates: an investigation of the zebrafish *Hox* paralogue group 1 genes. Development 128(13):2471–2484

McLysaght A, Hokamp K, Wolfe KH (2002) Extensive genomic duplication during early chordate evolution. Nat Genet 31:200–204

Meulemans D, Bronner-Fraser M (2004) Gene-regulatory interactions in neural crest evolution and development. Dev Cell 7(3):291–299

Meulemans D, Bronner-Fraser M (2005) Central role of gene cooption in neural crest evolution. J Exp Zoolog B Mol Dev Evol 304(4):298–303

Meulemans D, Bronner-Fraser M (2007) Insights from amphioxus into the evolution of vertebrate cartilage. PLoS One 2(8):e787

Minguillon C, Ferrier DE, Cebrian C, Garcia-Fernandez J (2002) Gene duplications in the prototypical cephalochordate amphioxus. Gene 287(1–2):121–128

Minguillon C, Jimenez-Delgado S, Panopoulou G, Garcia-Fernandez J (2003) The amphioxus Hairy family: differential fate after duplication. Development 130(24):5903–5914

Muffato M, Louis A, Poisnel CE, Roest Crollius H (2010) Genomicus: a database and a browser to study gene synteny in modern and ancestral genomes. Bioinformatics 26(8):1119–1121

Nadeau JH, Sankoff D (1997) Comparable rates of gene loss and functional divergence after genome duplications early in vertebrate evolution. Genetics 147(3):1259–1266

Nakatani Y, Takeda H, Kohara Y, Morishita S (2007) Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. Genome Res 17(9):1254–1265

Naruse K, Tanaka M, Mita K, Shima A, Postlethwait J, Mitani H (2004) A medaka gene map: the trace of ancestral vertebrate proto-chromosomes revealed by comparative gene mapping. Genome Res 14(5):820–828

Nelson JS (1994) Fishes of the world, 3rd edn. Wiley-Interscience, New York

Nikitina N, Sauka-Spengler T, Bronner-Fraser M (2009) Chapter 1. Gene regulatory networks in neural crest development and evolution. Curr Top Dev Biol 86:1–14

Northcutt RG, Gans C (1983) The genesis of neural crest and epidermal placodes: a reinterpretation of vertebrate origins. Quart Rev Biol 58:1–28

Nowak MA, Boerlijst MC, Cooke J, Smith JM (1997) Evolution of genetic redundancy. Nature 388(6638):167–171

Oda H, Wada H, Tagawa K, Akiyama-Oda Y, Satoh N, Humphreys T, Zhang S, Tsukita S (2002) A novel amphioxus cadherin that localizes to epithelial adherens junctions has an unusual domain organization with implications for chordate phylogeny. Evol Dev 4(6):426–434

Ohno S (1970) Evolution by gene duplication. Springer, New York

Ohno S, Wolf U, Atkins NB (1968) Evolution from fish to mammals by gene duplication. Hereditas 59(1):169–187

Ota KG, Kuratani S (2007) Cyclostome embryology and early evolutionary history of vertebrates. Integr Comp Biol 47(3):329–337

Otto SP (2007) The evolutionary consequences of polyploidy. Cell 131(3):452–462

Panopoulou G, Hennig S, Groth D, Krause A, Poustka AJ, Herwig R, Vingron M, Lehrach H (2003) New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes. Genome Res 13(6A):1056–1066

Paps J, Holland PW, Shimeld SM (2012) A genome-wide view of transcription factor gene diversity in chordate evolution: less gene loss in amphioxus? Brief Funct Genomics 11(2):177–186

Parisod C, Alix K, Just J, Petit M, Sarilar V, Mhiri C, Ainouche M, Chalhoub B, Grandbastien MA (2010) Impact of transposable elements on the organization and function of allopolyploid genomes. New Phytol 186(1):37–45

Paterson AH, Chapman BA, Kissinger JC, Bowers JE, Feltus FA, Estill JC (2006) Many gene and domain families have convergent fates following independent whole-genome duplication events in Arabidopsis, Oryza, Saccharomyces and Tetraodon. Trends Genet 22(11):597–602

Pebusque MJ, Coulier F, Birnbaum D, Pontarotti P (1998) Ancient large-scale genome duplications: phylogenetic and linkage analyses shed light on chordate genome evolution. Mol Biol Evol 15(9):1145–1159

Pendleton JW, Nagai BK, Murtha MT, Ruddle FH (1993) Expansion of the *Hox* gene family and the evolution of chordates. Proc Natl Acad Sci U S A 90:6300–6304

Polak P, Domany E (2006) Alu elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes. BMC Genomics 7:133

Popovici C, Leveugle M, Birnbaum D, Coulier F (2001) Coparalogy: physical and functional clusterings in the human genome. Biochem Biophys Res Commun 288(2):362–370

Postlethwait J, Amores A, Cresko W, Singer A, Yan YL (2004) Subfunction partitioning, the teleost radiation and the annotation of the human genome. Trends Genet 20(10):481–490

Postlethwait JH (2007) The zebrafish genome in context: ohnologs gone missing. J Exp Zoolog B Mol Dev Evol 308(5):563–577

Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu JK, Benito-Gutierrez EL, Dubchak I, Garcia-Fernandez J, Gibson-Brown JJ, Grigoriev IV, Horton AC, de Jong PJ, Jurka J, Kapitonov VV, Kohara Y, Kuroki Y, Lindquist E, Lucas S, Osoegawa K, Pennacchio LA, Salamov AA, Satou Y, Sauka-Spengler T, Schmutz J, Shin IT, Toyoda A, Bronner-Fraser M, Fujiyama A, Holland LZ, Holland PW, Satoh N, Rokhsar DS (2008) The amphioxus genome and the evolution of the chordate karyotype. Nature 453(7198):1064–1071

Ravi V, Lam K, Tay BH, Tay A, Brenner S, Venkatesh B (2009) Elephant shark (*Callorhinchus milii*) provides insights into the evolution of *Hox* gene clusters in gnathostomes. Proc Natl Acad Sci U S A 106(38):16327–16332

Robinson-Rechavi M, Boussau B, Laudet V (2004) Phylogenetic dating and characterization of gene duplications in vertebrates: the cartilaginous fish reference. Mol Biol Evol 21(3): 580–586

Roux J, Robinson-Rechavi M (2008) Developmental constraints on vertebrate genome evolution. PLoS Genet 4(12):e1000311

Ruvinsky I, Silver LM (1997) Newly identified paralogous groups on mouse chromosomes 5 and 11 reveal the age of a T-box cluster duplication. Genomics 40(2):262–266

SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL (1998) The paleontology of intergene retrotransposons of maize. Nat Genet 20(1):43–45

SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL (1996) Nested retrotransposons in the intergenic regions of the maize genome. Science 274(5288):765–768

Sauka-Spengler T, Meulemans D, Jones M, Bronner-Fraser M (2007) Ancient evolutionary origin of the neural crest gene regulatory network. Dev Cell 13(3):405–420

Scannell DR, Byrne KP, Gordon JL, Wong S, Wolfe KH (2006) Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. Nature 440(7082):341–345

Semon M, Wolfe KH (2007) Consequences of genome duplication. Curr Opin Genet Dev 17(6):505–512

Semon M, Wolfe KH (2008) Preferential subfunctionalization of slow-evolving genes after allopolyploidization in *Xenopus laevis*. Proc Natl Acad Sci U S A 105(24):8333–8338

Sempere LF, Cole CN, McPeek MA, Peterson KJ (2006) The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. J Exp Zool B Mol Dev Evol 306(6):575–588

Seo HC, Edvardsen RB, Maeland AD, Bjordal M, Jensen MF, Hansen A, Flaat M, Weissenbach J, Lehrach H, Wincker P, Reinhardt R, Chourrout D (2004) *Hox* cluster disintegration with persistent anteroposterior order of expression in *Oikopleura dioica*. Nature 431(7004):67–71

Sharman AC, Holland PWH (1998) Estimation of *hox* gene cluster number in lampreys. Int J Dev Biol 42:617–620

Shemer G, Podbilewicz B (2000) Fusomorphogenesis: cell fusion in organ formation. Dev Dyn 218(1):30–51

Shimeld SM, Holland PW (2000) Vertebrate innovations. Proc Natl Acad Sci U S A 97(9): 4449–4452

Sidow A (1996) Gen(om)e duplications in the evolution of early vertebrates. Curr Opin Genet Dev 6:715–722

Siegel N, Hoegg S, Salzburger W, Braasch I, Meyer A (2007) Comparative genomics of *ParaHox* clusters of teleost fishes: gene cluster breakup and the retention of gene sets following whole genome duplications. BMC Genomics 8:312

Skrabanek L, Wolfe KH (1998) Eukaryote genome duplication—where's the evidence? Curr Opin Genet Dev 8:694–700

Small KS, Brudno M, Hill MM, Sidow A (2007) A haplome alignment and reference sequence of the highly polymorphic *Ciona savignyi* genome. Genome Biol 8(3):R41

Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, Depamphilis CW, Wall PK, Soltis PS (2009) Polyploidy and angiosperm diversification. Am J Bot 96(1):336–348

Soltis DE, Soltis PS (1999) Polyploidy: recurrent formation and genome evolution. Trends Ecol Evol 14(9):348–352

Somorjai I, Bertrand S, Camasses A, Haguenauer A, Escriva H (2008) Evidence for stasis and not genetic piracy in developmental expression patterns of *Branchiostoma lanceolatum* and *Branchiostoma floridae*, two amphioxus species that have evolved independently over the course of 200 Myr. Dev Genes Evol 218(11–12):703–713

Spring J (1997) Vertebrate evolution by interspecific hybridization—are we polyploid? Fed Eur Biol Soc Lett 400:2–8

Stadler PF, Fried C, Prohaska SJ, Bailey WJ, Misof BY, Ruddle FH, Wagner GP (2004) Evidence for independent *Hox* gene duplications in the hagfish lineage: a PCR-based gene inventory of *Eptatretus stoutii*. Mol Phylogenet Evol 32(3):686–694

Storchova Z, Pellman D (2004) From polyploidy to aneuploidy, genome instability and cancer. Nat Rev Mol Cell Biol 5(1):45–54

Takio Y, Pasqualetti M, Kuraku S, Hirano S, Rijli FM, Kuratani S (2004) Evolutionary biology: lamprey *Hox* genes and the evolution of jaws. Nature 429(6989):1–262

Taylor JS, Raes J (2004) Duplication and divergence: the evolution of new genes and old ideas. Annu Rev Genet 38:615–643

Taylor JS, van de Peer Y, Meyer M (2001) Genome duplication, divergent resolution and speciation. Trends Genet 17:299–301

Thornburg BG, Gotea V, Makalowski W (2006) Transposable elements as a significant source of transcription regulating signals. Gene 365:104–110

van de Lagemaat LN, Landry JR, Mager DL, Medstrand P (2003) Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. Trends Genet 19(10):530–536

Van de Peer Y, Maere S, Meyer A (2009) The evolutionary significance of ancient genome duplications. Nat Rev Genet 10(10):725–732

Venkatesh B, Kirkness EF, Loh YH, Halpern AL, Lee AP, Johnson J, Dandona N, Viswanathan LD, Tay A, Venter JC, Strausberg RL, Brenner S (2007) Survey sequencing and comparative analysis of the elephant shark (*Callorhinchus milii*) genome. PLoS Biol 5(4):e101

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA et al (2001) The sequence of the human genome. Science 291(5507):1304–1351

Vienne A, Shiina T, Abi-Rached L, Danchin E, Vitiello V, Cartault F, Inoko H, Pontarotti P (2003) Evolution of the proto-MHC ancestral region: more evidence for the plesiomorphic organisation of human chromosome 9q34 region. Immunogenetics 55(7):429–436

Wada H, Okuyama M, Satoh N, Zhang S (2006) Molecular evolution of fibrillar collagen in chordates, with implications for the evolution of vertebrate skeletons and chordate phylogeny. Evol Dev 8(4):370–377

Werth CR, Windham MD (1991) A model for divergent, allopatric speciation of polypoid pteridophytes resulting from silencing of duplicate-gene expression. Am Nat 137:515–526
Weston JA (1970) The migration and differentiation of neural crest cells. Adv Morphog 8:41–114
Wolfe K (2000) Robustness—it's not where you think it is. Nat Genet 25(1):3–4
Wolfe KH (2001) Yesterday's polyploids and the mystery of diploidization. Nat Rev Genet 2(5):333–341
Wolfe KH, Shields DC (1997) Molecular evidence for an ancient duplication of the entire yeast genome. Nature 387(6634):708–713
Wotton KR, Shimeld SM (2006) Comparative genomics of vertebrate Fox cluster loci. BMC Genomics 7:271
Yu JK, Meulemans D, McKeown SJ, Bronner-Fraser M (2008) Insights from the amphioxus genome on the origin of vertebrate neural crest. Genome Res 18(7):1127–1132

Chapter 17
Polyploidy in Fish and the Teleost Genome Duplication

Ingo Braasch and John H. Postlethwait

Abstract Multiple rounds of whole-genome duplications (WGDs) punctuated the evolution of rayfin fish, a species-rich group comprising about half of all vertebrates. Rayfin fish, along with lobefin vertebrates including humans, derive from early vertebrate ancestors that evolved through two rounds of polyploidization (the first and second rounds of vertebrate genome duplication, VGD1 and VGD2) at the dawn of the vertebrate lineage. Furthermore, teleost fish underwent an additional round of polyploidization in their stem lineage, the teleost genome duplication (TGD). Additional WGD events occurred independently in numerous species and higher level taxa of teleosts and other rayfin fish, for example in salmonids, carp, and sturgeon, so that some fish lineages experienced at least four rounds of WGD since the origin of vertebrates. This chapter provides an overview of these polyploidization events in the fish lineage and focuses on the impact these genome duplications (GD) had on genome evolution in selected fish taxa. We then review evidence for the TGD and discuss its consequences for the evolution of gene content, order, and functions in the teleost lineage. We argue that, although evidence remains sparse, the TGD may have had a profound influence on the evolutionary success and the biodiversity of teleosts. Importantly, an in-depth understanding of the causes and consequences of the TGD and other teleost GD events will help to inform us about the evolution of our own paleopolyploid genome.

I. Braasch · J. H. Postlethwait (✉)
Institute of Neuroscience, University of Oregon, Eugene, OR 97403, USA
e-mail: jpostle@uoneuro.uoregon.edu

I. Braasch
e-mail: ibraasch@uoneuro.uoregon.edu

P. S. Soltis and D. E. Soltis (eds.), *Polyploidy and Genome Evolution*,
DOI: 10.1007/978-3-642-31442-1_17,

The problems of this world are only truly solved in two ways: by extinction or duplication.

Susan Sontag

In their great numbers and degree of anatomical diversity, the modern ray-finned fishes may be considered the most successful of all vertebrates.

Robert L. Carroll

17.1 A Brief Introduction to "Fish"

In common usage, "fish" applies to any aquatic vertebrate that possesses gills and fins (if any appendages) (Nelson 2006). Following this phenotypic definition, "fish" include jawless fish such as lampreys and hagfish, cartilaginous fish, such as sharks and rays, lobefin fish including coelacanth and lungfish, and rayfin fish such as teleosts. "Fish", however, are a paraphyletic assemblage because it excludes tetrapod vertebrates, which share a more recent common ancestor with lungfish than lungfish share with teleost fish. Until we tend to call tetrapods (including ourselves) lobefin fish and accept that all living vertebrates are fish, the term "fish" should be used with caution (see Fig. 17.1).

This chapter focuses on the monophyletic group of rayfin fish, or Actinopterygii, and, in more detail, their largest subgroup, teleost fish. The jawless, cartilaginous, and lobefin fish, however, are important for understanding evolution of the vertebrate genome by whole-genome duplications (WGDs) at the dawn of the vertebrate lineage and are thus featured in Chap. 12 (this volume).

Among actinopterygian fish, the teleosts are by far the largest subgroup in terms of extant species numbers. Nelson (2006) lists 26,891 living rayfin fish species (453 families), of which only 51 species (5 families) are not teleosts. The teleosts, in contrast, comprise 26,840 living species (448 families), an impressive 99.8 % of all rayfins. Even more impressive, adding all living vertebrates to the calculation gives around 50 % of all vertebrates being teleost fish.

Draft genome assemblies of ten rayfin species (all of them clupeocephalan teleosts) are currently publicly available (Fig. 17.1): zebrafish (*Danio rerio*), three-spined stickleback (*Gasterosteus aculeatus*), medaka (*Oryzias latipes*) (Kasahara et al. 2007), fugu (*Takifugu rubripes*) (Aparicio et al. 2002), spotted green pufferfish (*Tetraodon nigroviridis*) (Jaillon et al. 2004), Atlantic cod (*Gadus morhua*) (Star et al. 2011), and four species of East African cichlids, including tilapia (*Oreochromis niloticus*). More species such as platyfish (*Xiphophorus maculatus*), Atlantic salmon (*Salmo salar*), rainbow trout (*Oncorhynchus mykiss*), and the non-teleost spotted gar (*Lepisosteus oculatus*; see below) are soon to be added to this list (Fig. 17.1), and numerous additional rayfins will be sequenced as part of the GENOME 10K Project (Haussler et al. 2009).

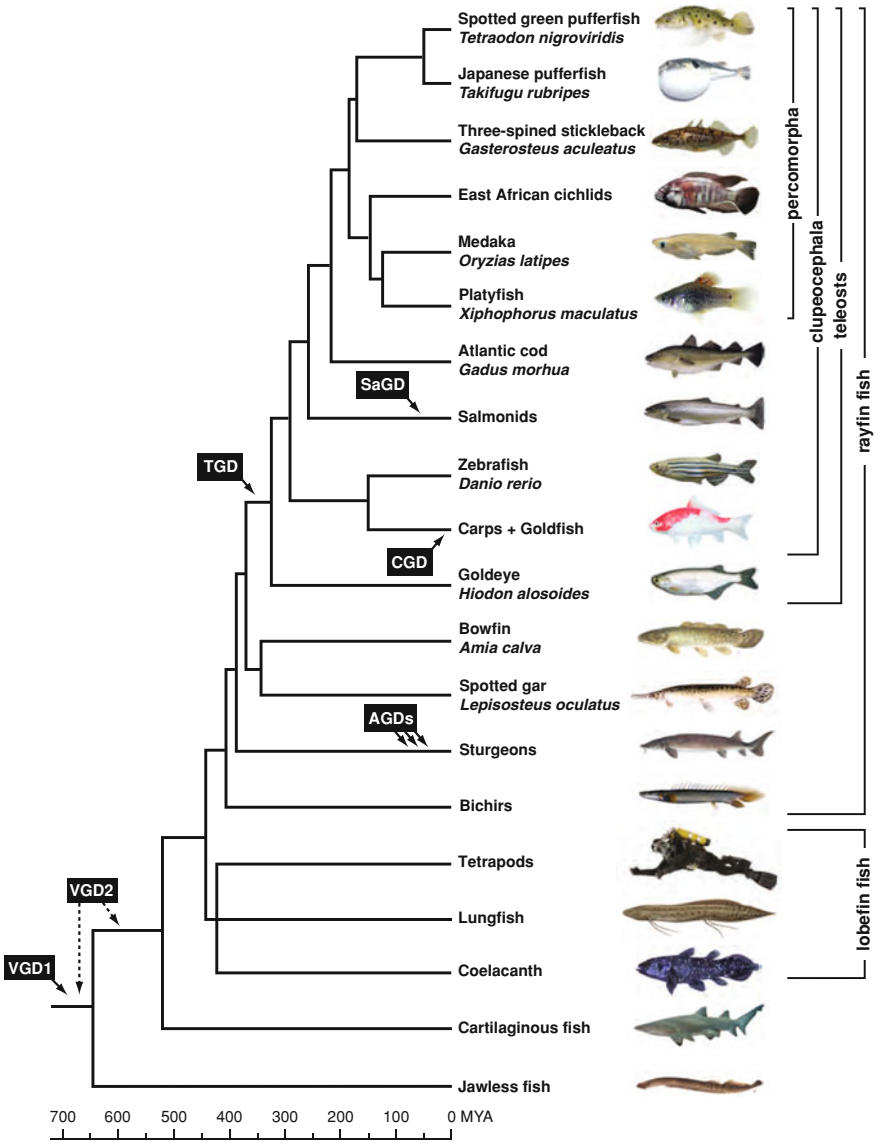

Fig. 17.1 A phylogeny of all kinds of "fish". Tree topology and timing of divergence events follows data from Setiamarga et al. (2009) and www.timetree.org. The figure shows the timing of genome duplication (*GD*) events (*VGD*1/*VGD*2 the vertebrate GDs, *TGD* the teleost GD, *AGDs* acipenserid GDs, *SaGD* the salmonid GD, *CGD* the carp GD). Whether the second vertebrate genome duplication (*VGD*2) occurred before or after the divergence of jawless fish remains controversial. *MYA* million years ago

The species richness of rayfin and teleost fish poses the question of the secret of their evolutionary success. A recurring pattern of genome evolution in rayfin fish is their propensity to polyploidization. In line with this idea, a positive correlation of genome size and species richness has been reported among actinopterygians as well as for specific teleost clades (Mank and Avise 2006a).

17.2 Duplications in the Crown: Polyploidy in Rayfin Fish

In comparison to plants, polyploidizations are rare in the animal kingdom (Muller 1925), yet the reasons for this genomic discrepancy between fauna and flora remain controversial (Mable 2004). In the vertebrate lineage, polyploids are most often found among amphibians and rayfin fish (Mable et al. 2011; Otto 2007; Comai 2005). Factors that may favor the proclivity of rayfin fish to generate surviving polyploid lineages compared to other groups of vertebrates (e.g., high production rate of unreduced gametes, their propensity for hybridization, genomic flexibility, etc.) have been extensively discussed recently elsewhere, yet a clear explanation remains elusive (Mable et al. 2011). Given that polyploid events also appear to have occurred in some jawless, cartilaginous, and lobefin fish (see Leggatt and Iwama 2003 and references therein), it may be worth considering that a paucity of genome duplication (GD) is an amniote vertebrate-specific phenomenon. Several works provide comprehensive lists of known cases of polyploidy in fish (Otto and Whitton 2000; Leggatt and Iwama 2003; Le Comber and Smith 2004; Mable et al. 2011).

Here, we give an overview of the few general trends that can be inferred from the phylogenetic distribution of polyploidization events and will then discuss several specific cases in more detail.

17.2.1 Twigs and Leaves: Polyploidization Events in Fish

Despite large variations in genome size, rayfin fish have surprisingly conserved karyotypes, with the majority of diploid genomes having 48 or 50 chromosomes (Ohno et al. 1968; Mank and Avise 2006b); thus, chromosome counts have been used as indicators of polyploidizations in fish. A study based on chromosome counts of 615 rayfin species concluded that at least 7–20 polyploidization events occurred in the evolution of extant rayfins (Mank and Avise 2006b); this estimate is likely an underestimate because less than 3 % of all living rayfin species were included in this study.

Polyploidization events in fish are phylogenetically restricted, i.e., are unevenly distributed across the actinopterygian tree. Some groups, especially those occupying early diverging branches in rayfin phylogeny, and among teleosts particularly the Ostariophysi (including carps and suckers and many others), seem to be

particularly prone to polyploidization in their lineage, while other groups, including more derived lineages such as the percomorphs (perch and relatives, like pufferfish, stickleback, cichlids and many more), have few or no GDs after the teleost genome duplication (TGD) (Leggatt and Iwama 2003; Mable et al. 2011).

Polyploidizations have occurred in many individual fish genera, species, and populations (i.e., the "leaves" of the rayfin tree), as for example in multiple species of barbs (genus *Barbus*, family Cyprinidae) (Chenuil et al. 1999), in the cyprinid *Squalius alburnoides* species complex (Alves et al. 2001), in the pond loach (*Misgurnus anguillicaudatus*) (Li et al. 2011) and other loach species (Cobitidae, order Cypriniformes) (Ferris and Whitt 1977a), and in the stinging catfish (*Heteropneustes fossilis,* order Siluriformes) (Pandian and Koteeswaran 1999). On the other hand, piscine GD events have also occurred in the last common ancestor of higher level taxa (some of the tree's "twigs"). All members are polyploid in the three families Salmonidae (salmonids) (Allendorf and Thorgaard 1984), Callichthyidae (armored catfish) (Oliveira et al. 1992), and Catostomidae (suckers) (Ferris 1984; Uyeno and Smith 1972). Common carp and goldfish (subfamily Cyprininae) also arise from a shared tetraploid origin (Ohno et al. 1967; David et al. 2003; Schultz 1980; Larhammar and Risinger 1994). Among rayfins diverging before teleosts, the Acipenseridae (sturgeons) are famous for their multiple ploidy levels, up to at least hexadecaploid (Birstein et al. 1997; Ludwig et al. 2001). In addition, artificial polyploids, particularly triploids, are routinely produced for multiple commercially important fish species to induce sterility and thereby improve somatic growth rate (Piferrer et al. 2009).

17.2.2 Piscine Polyploidy and the Evolution of Genome Function: Salmon, Carp, and Calandinos

Traditional studies of polyploidy in fish generally involve the analysis of karyotypes and allozymes. Only recently, with the advent of genome sequencing techniques, we have begun to understand some of the evolutionary footprints left by polyploidization events in fish genomes.

17.2.2.1 Salmonids

Salmonids (salmon, trout, whitefish, grayling) are among the fish of highest economic importance (Davidson et al. 2010). Salmonids have been suspected to be polyploids since the 1940s (Svärdson 1945; Kupka 1948), but it was several decades until it was generally accepted that all living salmonids are of autotetraploid origin due to a GD event that occurred in their stem lineage 25–100 million years ago (Ohno et al. 1968; Allendorf and Thorgaard 1984). This WGD is unlikely to have occurred by segmental duplications because all segments of each

chromosome appear to be in duplicate (although of course not every gene in each segment is still present in duplicate); if segmental duplications had been responsible, then some chromosome segments would be present in three or five copies. Salmonids still show multivalent chromosomes during meiosis as well as tetrasomic inheritance of some loci (Allendorf and Thorgaard 1984; Phillips and Rab 2001) and are thus considered pseudotetraploids, i.e., the process of diploidization has not yet concluded in these fish (Danzmann et al. 2008).

The autotetraplodization in salmonids is sometimes referred to as "4R", because it is the fourth round of WGD in this lineage since the rise of the vertebrate lineage (see later this chapter and Chap. 16, this volume). This numerating terminology, however, imposes difficulties to distinguish independent GD events in various lineages of rayfins as discussed above. Therefore, we suggest using the term *salmonid genome duplication* (*SaGD*). Recent progress in salmonid genomics (Davidson et al. 2010; Miller et al. 2011) has provided early insights into the evolutionary dynamics of the salmonid genome after the SaGD.

In line with tetraploidization in a salmonid ancestor, the diploid genomes of most salmonids contain 96–104 chromosome arms. The Atlantic salmon is a special case among salmonids, because its more derived karyotype has been secondarily reduced to 72–74 chromosome arms by chromosome fusions (Phillips and Rab 2001; Phillips et al. 2009). Conserved synteny analysis comparing genetic linkage maps of Atlantic salmon and rainbow trout (Danzmann et al. 2008; Lien et al. 2011) and analyzing the genomic distribution of conserved non-coding elements (CNEs) (Moghadam et al. 2009) show first that gene order in salmonids generally reflects the inferred ancestral rayfin karyotypes (Nakatani et al. 2007; Kasahara et al. 2007), and second that two salmonid homeologous chromosomes share conserved synteny with one teleost outgroup chromosome (e.g., medaka, zebrafish, and stickleback), as would be expected from an additional round of GD.

The duplication of *hox* gene clusters, generally a good indicator for GDs in animals (see Chap. 16, this volume), has been studied in detail in salmonids. Salmonid genomes contain at least 13 *hox* clusters, nearly twice as many as most diploid teleost genomes (Moghadam et al. 2005a, b; Mungpakdee et al. 2008a). Given that gene loss (nonfunctionalization, or pseudogenization) is the most frequent fate of gene duplicates (Lynch and Conery 2000), it is surprising that the salmonid *hox* cluster repertoire, as judged by collapsing duplicated clusters, shows no loss of *hox* paralogy groups between the ancestral teleost and today's Atlantic salmon genome, which has retained more *hox* paralogy groups than any other studied teleost (Mungpakdee et al. 2008a). After the SaGD, several *hox* paralogs evolved asymmetrically, as manifested by pseudogenization, elevated rates of molecular evolution, and/or divergence of non-coding regions in one of the two *hox* gene duplicates. This asymmetry is also apparent over entire *hox* clusters (Mungpakdee et al. 2008a). Overall, however, expression domains of duplicated Atlantic salmon *hox* genes are conserved with their unduplicated orthologs in other teleosts, although salmon *hox* paralogs show evidence for quantitative subfunctionalization (Mungpakdee et al. 2008b), a predicted type of functional evolution of duplicate genes (Force et al. 1999).

Overall, around 50 % of SaGD-duplicated loci have been retained in salmonids (Allendorf and Thorgaard 1984). Although purifying selection appears to be the prevalent mode of salmonid gene duplicate evolution, asymmetric divergence of salmonid gene duplicates was also observed in an analysis of EST sequences when compared to the diploid Northern pike (*Esox lucius*) as outgroup (Koop et al. 2008; Leong et al. 2010). The asymmetric pattern of molecular evolution among paralogs, observed particularly often for DNA binding proteins, is thought to be caused by relaxed functional constraints on one of the two paralogs after the SaGD (Leong et al. 2010; Mungpakdee et al. 2008a) as for GDs in general.

17.2.2.2 Carp and Goldfish: Facilitated Domestication by Genome Duplication?

An allotetraploidization occurred in a shared ancestor of common carp (*Cyprinus carpus*; $2n = 100$) and goldfish (*Cyprinus carpio*; $2n = 100$). The carp genome duplication (CaGD) occurred between 11 and 21 million years ago (Risinger and Larhammar 1993; Ohno et al. 1967; Larhammar and Risinger 1994; David et al. 2003; Yuan et al. 2010). Evidence for the hybrid (i.e., allotetraploid) origin of these polyploids comes from the absence of quadrivalents (Ohno et al. 1967) and the disomic inheritance of genetic markers in view of the short time period since the GD event (David et al. 2003). Independent chromosomal rearrangements took place in goldfish and carp after the allotetraploidization (Ohno et al. 1967). Around 52–60 % of genomic loci of the common carp are still duplicated (David et al. 2003; Ferris and Whitt 1977b), but duplicate retention seems to be considerably lower in goldfish (Woods and Buth 1984). As in salmonids, several duplicates of *hox* cluster genes have become pseudogenes after the CaGD (Luo et al. 2007; Yuan et al. 2010).

Like salmonids, the common carp and its relatives are important economic species in aquaculture as food sources and as ornamental species. Common carp and goldfish are the oldest domesticated fish species, bred in central Europe for $\sim$2,000 years and in China for $\sim$1,000 years, respectively (reviewed in Balon 2004). During domestication of carp and goldfish, significant morphological change has been fixed in populations compared to their wild ancestors in a short period of time, as exemplified by the multiple, independent loss of scales in the carp (Balon 2004), the color morphs of koi carp (David et al. 2004), and the selection for monstrosities such as telescope eyes, lion-head, or fin loss in ornamental goldfish (Komiyama et al. 2009).

The molecular basis of the strong response to artificial selection in carp and goldfish has rarely been explored. Based on the multigenic inheritance of coloration in ornamental koi carp, it has been suggested that the propensity to generate the many color morphs may be related to the tetraploidization of carps (David et al. 2004). Is it possible that allopolyploidization provided carp and goldfish with a high degree of genomic flexibility similar to domestic polyploid plants (see Chaps. 7 and 10, this volume)—made them particularly responsive to morphological selection

and domestication? A recent study of the common carp suggests that the duplication of the *fgfr*1*a* gene during the course of the CaGD may have permitted the breeding of "mirror" forms that lack almost all scales (Rohner et al. 2009). Importantly, two different domesticated forms of mirror carp have been bred by selection of two different loss-of-function mutations in the same paralog, *fgfr*1*a*1 (Rohner et al. 2009).

17.2.2.3 Gene Expression Regulation in Triploid Calandino

Although polyploidization events in salmonids and carp are fairly recent events on the evolutionary timescale, they are nevertheless too old to study changes in gene expression pattern following soon after polyploidization and hybridization in fish and the interplay of these two factors. The relationship of polyploidization and hybridization, however, has been analyzed in naturally occurring populations of an Iberian cyprinid, the calandino (*S. alburnoides*).

The allopolyploid *S. alburnoides* species complex consists of di-, tri-, and tetraploid fish of different genomic compositions derived from interspecific hybridizations of a paternal ancestor (the *A* genome) and different, geographically separated, maternal genome contributors, the southern *S. pyraneicus* (*P* genome) or the northern *S. caroliterii* (*C* genome) (Alves et al. 2001; Pala et al. 2010). Expression studies of seven genes in different adult tissues revealed that dosage compensation occurs in polyploids, so that overall gene expression is reduced to the diploid level (Pala et al. 2008, 2010). In *S. alburnoides*, gene dosage compensation is accomplished by a complex pattern of gene copy silencing in a gene-specific and tissue-specific manner: for example, in southern triploids of the *PAA* genome composition, some genes are expressed only from the *A* genome, others from *P* and *A* genomes; some tissues express only *A* genes, while others express both *P* and *A* genes; some genes are expressed from *A* in one tissue, but from *P* and *A* in another tissue (Pala et al. 2008, 2010).

The gene expression patterns also depend on genomic composition: while the *A* genome allele is dominant in southern individuals, *C* and *A* alleles are co-dominant in northern polyploids (*CAA*, *CCA*, *CCAA*, or *C*-*A*-) (Pala et al. 2010). Although the exact molecular mechanism leading to dosage compensation in the calandino remains elusive at this point, these studies illustrate the potential these animals provide for the study of gene regulation in diverging polyploid populations.

17.3 The Doubled Trunk: The Teleost Genome Duplication

Although recent polyploidization events appear to be comparatively common in rayfin fish compared to other vertebrate clades, the most important polyploidization event for genome evolution in fish took place on the branch leading to the teleosts, the largest clade of rayfin fish. Initial arguments for and against the

occurrence of such a third round of vertebrate WGD (alias 3R, fish-specific genome duplication, FSGD, or teleost genome duplication, TGD) have been overcome by sequencing the genomes of several teleost species.

17.3.1 Evidence for the Teleost Genome Duplication

17.3.1.1 Expansion of Vertebrate Gene Families in Rayfin Fish

Initially, the branch leading to the teleosts did not raise suspicion of polyploidization because the ancestral vertebrate karyotype, like that of teleosts, was initially inferred to be 48 chromosomes (Ohno et al. 1968).

A general observation since the rise of allozyme data, however, was that many enzyme loci with a single locus in tetrapods appeared to have multiple copies in fish (see Morizot 1990 and references therein). With the advent of the zebrafish and other teleosts as model organisms for developmental and genetic studies in the 1990s and the accompanying cloning of teleost DNA sequences, this impression of "more genes in fish" (Wittbrodt et al. 1998) was substantially reinforced.

Three alternative explanations were initially considered to explain these observations: (1) higher frequencies of individual or local or tandem gene duplication in teleosts than in tetrapods; (2) higher retention of gene duplicates from the ancestral vertebrate genome duplications (VGD) in teleosts than in tetrapods; or (3) a WGD in the rayfin fish lineage after the divergence from tetrapods (Morizot 1990; Postlethwait et al. 1998; Wittbrodt et al. 1998; Meyer 1998; Robinson-Rechavi et al. 2001).

17.3.1.2 Teleost *Hox* Gene Clusters

Sequencing the *hox* gene clusters, which had already suggested the earlier VGDs (see Chap. 16, this volume), settled the argument toward the teleost-specific GD hypothesis: since the cloning of the first teleost *hox* genes in the late 1980s (Eiken et al. 1987; Njolstad et al. 1988), it became apparent that *hox* genes were no exception to the generalization of "more genes in fish" and additional paralogs of *hox* genes were found in different teleosts (e.g., Misof and Wagner 1996; Aparicio et al. 1997; Prince et al. 1998). The genomes of tetrapods and other non-teleost vertebrates generally have four *hox* gene clusters (Graham et al. 1989). The discovery that *hox* genes are actually organized in seven gene clusters in the zebrafish, that two copies of the *hoxA* cluster were present in fugu, and that each duplicated *hox* cluster in fugu was orthologous to a single duplicated copy in zebrafish provided the first significant support that a GD had occurred in the lineage leading to zebrafish and that the event was shared by fugu (Amores et al. 1998; Aparicio et al. 1997). Importantly, the *hox* clusters and genes closely linked to the *hox* clusters were found to be distributed over eight different chromosomes

in zebrafish, as would be expected if a WGD increased the *hox* cluster number from four to eight, followed by the loss of one of the seven *hox* complements (Amores et al. 1998). Seven or eight *hox* gene clusters have since been identified in several other teleosts (see e.g. Prohaska and Stadler 2004; Hoegg and Meyer 2005; Hoegg et al. 2007, and references therein).

17.3.1.3 More than just *Hox*: Duplication of Other Multigene Families in Fish

Although the genomic location of *hox* gene clusters and linked gene families could be best explained by a WGD in the lineage leading to teleosts (Meyer and Schartl 1999), several alternative possibilities were discussed, including: (1) that the seven *hox* clusters in zebrafish were due to polyploidization specific to the zebrafish lineage (Stellwag 1999); (2) that *hox* cluster-containing chromosomal segments had been amplified in teleosts by local duplications (Elgar et al. 1999); and/or (3) that teleosts had an unusually high gene duplication rate (Robinson-Rechavi et al. 2001).

Phylogenetic analyses of numerous duplicated zebrafish genes, however, including those not related or linked to the *hox* genes and those distributed all over the genome, provided evidence for their origin after the divergence of zebrafish and tetrapod lineages (Taylor et al. 2001; Van de Peer et al. 2001). Additional work on specific gene families was also in line with this hypothesis [e.g. *mitf* (Altschmied et al. 2002), midkines (Winkler et al. 2003), *sox*9 (Cresko et al. 2003), *egfr* (Gomez et al. 2004), *pomc* (de Souza et al. 2005), and receptor tyrosine kinases (Braasch et al. 2006)].

At the same time, putting more and more genes on the genetic maps of zebrafish (Gates et al. 1999; Postlethwait et al. 2000; Woods et al. 2000, 2005) and medaka (Naruse et al. 2004) provided additional evidence for a teleost WGD beyond just the duplication of *hox* cluster-bearing chromosomes.

Importantly, the inclusion of gene sequences from the Japanese pufferfish (*T. rubripes*) revealed that these fish-specific gene duplications within multigene families usually date back to a point during rayfin fish evolution before the divergence of zebrafish and pufferfish (Amores et al. 1998; Postlethwait et al. 2002; Taylor et al. 2003). Therefore, it became apparent that the GD shared by these teleost fish most likely occurred somewhere along the branch to the teleost fish; but did it occur after the origin of teleosts? Did it include non-teleost rayfin fish? Or did it include all teleosts and only teleosts?

17.3.1.4 Global Evidence from Teleost Genome Projects

The sequence of the fugu (*T. rubripes*) genome was the second vertebrate and the first teleost fish draft genome assembly (Aparicio et al. 2002). The authors were cautious about making conclusions from the *Takifugu* genome assembly with respect to the TGD hypothesis, stating that the distribution of gene duplicates in

the *Takifugu* genome indicated segmental or large-scale, but not tandem, duplications (Aparicio et al. 2002).

Two follow-up large-scale analyses, however, came to the conclusion that the distribution and age of gene duplicates in the *Takifugu* genome were indeed consistent with the TGD hypothesis (Vandepoele et al. 2004; Christoffels et al. 2004). Both studies found that the majority of gene duplicates in *Takifugu* were located in paralogons, i.e., chromosomal blocks that share paralogous syntenies within the genome. Also, molecular clock-based analyses of gene families containing fish-specific gene duplicates helped to date the duplication event and hinted at their occurrence around the origin of the teleost lineage.

In contrast to the genome assembly of *Takifugu*, the genome assembly of the spotted green pufferfish (*T. nigroviridis*) was anchored onto chromosomes, which made possible a more detailed analysis of conserved synteny blocks within a teleost genome (Jaillon et al. 2004). The publication of the *Tetraodon* genome assembly put the TGD beyond doubt, by showing that internal paralogy and chromosomal blocks of doubled conserved synteny extended genome-wide within this pufferfish genome. Doubled conserved synteny refers to a human chromosomal region sharing conserved synteny with two pufferfish paralogons (see Sect. 17.3.2.2). These patterns of conserved synteny also led to reconstructions of ancestral karyotypes (Jaillon et al. 2004).

With the publication of the medaka genome assembly, patterns of conserved synteny within and between the genomes of medaka, pufferfish, and zebrafish, using human as an outgroup, allowed more refined reconstruction of ancestral, pre-WGD karyotypes (see Sect. 17.3.2.5) (Kasahara et al. 2007).

Formal publications of the zebrafish and stickleback genome assemblies are pending, but analyses of these two species (as exemplified in Sect. 17.3.2) overall confirm observations made from the genomes of pufferfish and medaka: the occurrence of a WGD in the lineage leading to the teleosts is the most parsimonious explanation of all available genome data.

17.3.1.5 Phylogenetic Timing of the Teleost Genome Duplication

Genome analyses of zebrafish, pufferfish, medaka, and stickleback pointed to a GD event somewhere in the rayfin fish lineage prior to the divergence of these teleosts, but when exactly did the event occur? Is it, for example, a trait shared only by clupeocephalan teleosts, to which the aforementioned sequenced species belong (Fig. 17.1)? Or did it happen in an ancestor of all teleosts? Or even earlier, before the rise of the teleost lineage and as such represents a trait shared with earlier diverging rayfin branches, such as those leading to bowfin, gar, sturgeon, or bichir? Of course, the phylogenetic timing of the TGD is essential for understanding any possible causal relationships between the TGD and the teleost radiation. This question is further complicated by difficulties in establishing historical relationships of various rayfin fish lineages diverging basal to the teleosts as well as of the

basal relationships within teleosts themselves based on morphological and molecular data.

The sequencing of four *hox* clusters from the bichir (*Polypterus senegalus*), a representative of the most basally branching extant lineage of rayfin fish, the polypteriforms, indicated that the GD did not occur in an ancestor of all rayfins (Chiu et al. 2004; Raincrow et al. 2011). Subsequently, several studies cloned and sequenced homeobox genes and several other nuclear markers from different species representing major phylogenetic lineages of rayfins (Hoegg et al. 2004; Crow et al. 2006; Mulley et al. 2006; Hurley et al. 2007). These studies did not find any evidence for the presence of paralogs derived from the TGD in the genomes of extant nonteleosts, i.e., Lepisosteiformes (gars), Amiiformes (bowfin), and Acipenseriformes (sturgeons and paddlefish, which have, however, experienced their own lineage-specific polyploidizations; see above). In contrast, paralogs were found in all teleost lineages analyzed, including representatives of the early branching teleost lineages, i.e., Osteoglossomorpha (bonytongues and mooneyes), Elopomorpha (tarpons and eels), and Clupeomorpha (herrings and shads) (see Fig. 17.1). Finally, Amores et al. (2011) recently presented the first genome-wide synteny comparison between a non-teleost and teleost rayfins through the generation of a high-density genetic map for the spotted gar (*L. oculatus*). A clear pattern of double conserved synteny became apparent between the gar and teleost genomes showing that the gar lineage diverged from the teleost lineage before the GD shared among teleosts (Amores et al. 2011).

To summarize, the GD initially detected in zebrafish, medaka, and pufferfish was an event shared by all teleosts but not by non-teleost rayfin fish, and it occurred (depending on the use of different molecular clocks) around 226–350 million years ago (Vandepoele et al. 2004; Christoffels et al. 2004; Hoegg et al. 2004; Hurley et al. 2007) (Fig. 17.1). It thus seems better to call this tetraploidization event the TGD (The often-used term Fish-Specific Genome Duplication, or FSGD, is ambiguous due to the many GD events in multiple fish lineages, such as salmonids, sturgeons, and cyprinids, and the paraphyletic term "fish".).

Reviewing morphological characters defining the teleost lineage, de Pinna (1996) articulated: *Intriguingly, molecular data have yet to provide consistent support for teleostean monophyly.* Fifteen years later, apparently it is the TGD that provides the best synapomorphy, a molecular character supporting the monophyly of teleosts and probably the best we will ever obtain from extant molecules.

17.3.2 The TGD and Gen(om)e Evolution in Teleosts

17.3.2.1 The TGD and Conserved Syntenies

A pair of genes that are syntenic are on the same chromosome in one species. Conserved syntenies are situations in which a pair of genes that are syntenic in one species have orthologs that are syntenic in another species. People often use the

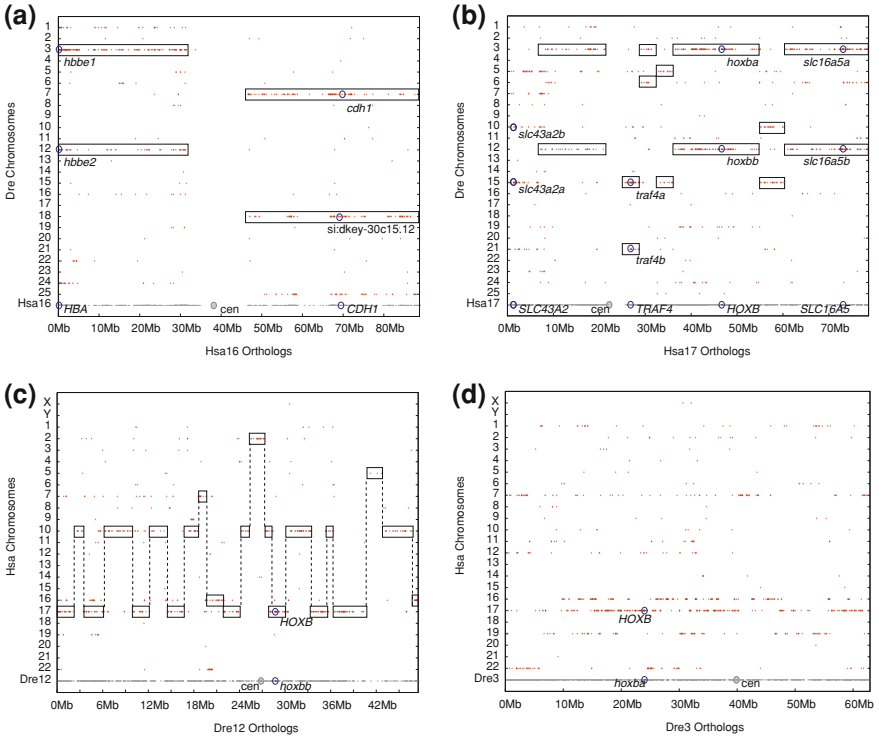

Fig. 17.2 Conserved syntenies are as predicted by GD. **a** Genes on human chromosome 16 are aligned along the horizontal axis and their zebrafish orthologs and co-orthologs are marked directly above each human gene on the zebrafish chromosome on which it resides. Results reveal duplicate ohnologons on Dre3 and Dre12 for Hsa16p and Dre7 and Dre18 for Hsa16q. A *gray circle* marks the location of the centromere on Hsa16. **b** Zebrafish ohnologons for Hsa17 are mainly on Dre3 and Dre12. **c** Human orthologs of Dre12 genes occupy single portions of human chromosomes, mainly Hsa10 and Hsa17. **d** Human orthologs of Dre12 are widely scattered on mostly Hsa16 and Hsa17. Individual genes are located on each plot

term 'syntenic' erroneously to mean 'conserved synteny'. Conserved syntenies give strong evidence for the TGD. The Synteny Database (http://teleost.cs.uoregon.edu/synteny_db/) uses BLAST scores to identify groups of paralogous genes in a genome rather than a single pair of genes showing the ''best hit''; the algorithm then anchors paralogy groups to an ortholog in another genome, and finally plots orthologs and paralogs along chromosomes (Catchen et al. 2009, 2011). Results show that segments of human chromosomes generally have orthologs and co-orthologs in paralogons on two zebrafish chromosomes. For example, the horizontal axis of Fig. 17.2 displays genes ordered along human chromosome 16 (Hsa16) and on the vertical axis shows the zebrafish orthologs and co-orthologs of each Hsa16 gene plotted on the appropriate linkage groups directly above the human gene. The short arm of Hsa16 (Hsa16p) has orthologs and co-orthologs mainly on zebrafish chromosomes Dre3 and Dre12, including duplicated hemoglobin loci, and the long

arm (Hsa16q) has orthologs and co-orthologs on Dre7 and Dre18, including co-orthologs of *CDH*1 and many other genes. Other zebrafish chromosomes also have strings of genes that appear on the plot, including Dre24 and Dre25, but most of these genes are more distantly related paralogs resulting from the VGDs.

The content of other human chromosomes appears to have involved more rearrangements than Hsa16 since the divergence of the human and zebrafish lineages, either in the rayfin or the lobefin lineage or both. For example, in zebrafish, the genetic content now on Hsa17 occupies duplicates of at least seven distinct large chromosome segments, mostly on Dre3 and Dre12 (Fig. 17.2b), which indicates several major translocations with respect to human, some of which occurred before the TGD because they are shared by both copies, and some after the TGD, because they differ between the two paralogons (or *ohnologons*, paralogons arising from GD).

Comparisons in the other direction, zebrafish to human, generally display a one-to-one relationship with a substantial number of translocations and inversions (Fig. 17.2c). For example, about half of the human orthologs of Dre12 genes are located on Hsa17, and about half on Hsa10, each with several rearrangements; minor numbers are located on four other zebrafish chromosomes. Other chromosomes, like Dre3, appear to have experienced an even greater number of chromosome rearrangements (Fig. 17.2d).

17.3.2.2 Doubled Conserved Synteny

The most frequent fate of a pair of ohnologs after WGD is nonfunctionalization (Lynch and Conery 2000); hence, ohnologons often experience reciprocal gene loss, like the regions containing two zebrafish co-orthologs of *SLC16A 5* on Dre3 and Dre12 (Figs. 17.2b and 17.3). Of the string of four genes on Hsa17 including human *SLC16A 5* , different sets are missing from each duplicated zebrafish segment, so that the content of both zebrafish ohnologons must be summed to get the orthologous human gene content in this region (Fig. 17.3). In the extreme, all WGD duplicates will have resolved to singletons, and so conserved syntenies will be maintained but with no co-orthologs to anchor the segments in a doubled conserved synteny.

17.3.2.3 An Example: The *Hox* Clusters

Circle plots can display conserved paralogs on a genome-wide scale. For example, the Synteny Database shows that paralogs of genes within 10 Mb of the four human *HOX* clusters preferentially occupy regions near each of the other clusters (Fig. 17.4a) (For nomenclature, see https://wiki.zfin.org/display/general/ZFIN+Zebrafish+Nomenclature+Guidelines.). Note stronger links between the pairs *HOXA*/*HOXB* and *HOXC*/*HOXD*, which may reflect the order of their origin, with

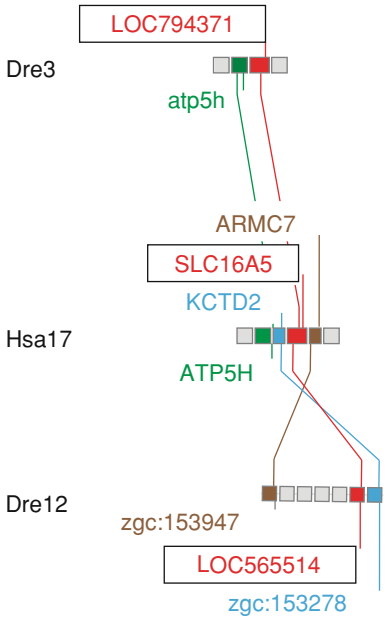

Fig. 17.3 Doubled conserved synteny. The region at about 73 Mb on Hsa17 including *SLC16A5* shares syntenies with small regions of Dre3 and Dre12. Zebrafish has two co-orthologs of *SLC16A5* (*boxed*), but some of its neighbors were reciprocally lost from one or the other ohnologon. In addition, an inversion on Dre12 reversed gene orders and caused other genes to disrupt the ancestral group. This pattern of doubled conserved synteny with disrupting inversions is quite common

VGD1 giving rise to *HOXA/B* and *HOXC/D* clusters (Amores et al. 1998) and VGD2 providing all four clusters. A circle plot of orthologs and paralogs of genes surrounding zebrafish *hox* clusters (Fig. 17.4b) shows that zebrafish has eight *hox* cluster-containing ohnologons even though only seven of them contain protein-coding *hox* genes (Amores et al. 1998). While zebrafish has two copies of the *hoxa, hoxb,* and *hoxc* clusters, it has a single *hoxd* cluster that contains protein-coding genes. The *hoxdb* cluster, while lacking protein-coding genes, retains a *microRNA*-10 paralog, which is embedded in *Hox* clusters from flies to fish to humans (Woltering and Durston 2006), as well as paralogs of many surrounding genes easily detected in the plot (Fig. 17.4b). Likewise, stickleback lacks a second *hoxc* cluster, but the location where it 'should' go is apparent from the circle plot (Fig. 17.4c).

17.3.2.4 Symmetry of Gene Loss Between Members of Ohnologons

Immediately after GD, both ohnologons have identical gene compositions. As nonfunctionalization events occur, ohnologons begin to diverge from each other. Gene loss events raise the question: Do both members of a pair of ohnologons tend to lose genes at the same rate? Or do genes tend to disappear more frequently from one duplicated chromosome segment than the other? According to one hypothesis, gene losses would be purely stochastic. According to an alternative hypothesis, expression of a number of genes in the same neighborhood might be regulated by a central element, for example, an element that helped to regulate local chromosome

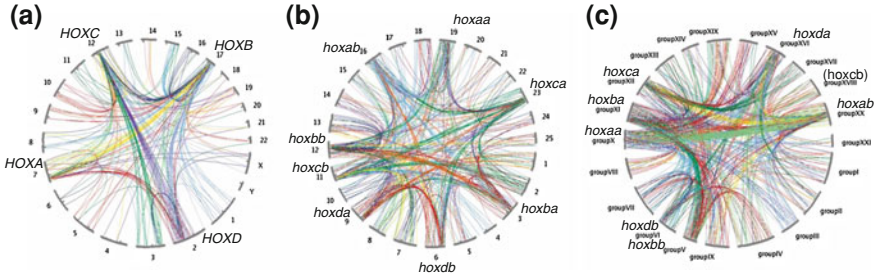

Fig. 17.4 *Hox* cluster ohnologons. **a** Paralogs within 10 Mb of human *HOX* clusters mapped on the human genome. **b** Paralogs within 5 Mb of zebrafish *hox* clusters mapped on the zebrafish genome. Although only a microRNA gene remains in the *hoxdb* cluster, the paralogy of this segment and its relationship to *hoxda* is evident. **c** Paralogs within 3 Mb of stickleback *hox* clusters mapped on the stickleback genome

structure, and loss of that element would then lead to rapid loss of the several genes that element controls and hence asymmetric gene loss. Analysis of a few individual chromosome regions suggests that asymmetries often occur in teleosts (Canestro et al. 2009; Braasch et al. 2006; Siegel et al. 2007), but this question has not been explored sufficiently on a genome-wide scale.

To study this question, we used the Synteny Database to plot the zebrafish and stickleback chromosomes on which orthologs, co-orthologs, and paralogs of each human gene lie for all human chromosomes. Figure 17.5 displays results for two human chromosomes (Hsa7 and Hsa18) for both zebrafish and stickleback. Figure 17.5a shows genes on Hsa7 as gray dots along the bottom with the zebrafish orthologs and paralogs directly above; note that gene order in these plots reflects gene arrangements in the human genome, not the fish genome.

Results revealed several principles (Fig. 17.5a–d). (1) Many regions of ohnology are clear, as marked by rectangles. For example, the region from 18 to 38 Mb on Hsa7 has two clear ohnologons in both zebrafish and stickleback, with orthologous chromosome segments marked by color (Fig. 17.5a, b). (2) In contrast, many regions have ambiguous ohnologons, such as the region from the left telomere of Hsa7 to about 18 Mb for both zebrafish and stickleback. (3) Some ohnolog pairs are clear in one species but ambiguous in the other; for example, from 38 to 64 Mb of Hsa7, where three or four zebrafish chromosomes have strings of 'hits' compared to two clear regions in stickleback. Regions like these (also note for stickleback region 101–142 Mb the 'ghost' paralogons on groupXII and groupXVIII) are often due to paralogs arising in the VGD1 and VGD2 events rather than the TGD. (4) Most chromosome translocations occurred before the divergence of the zebrafish and stickleback lineages. Evidence supporting this conclusion is that, with respect to the human chromosome, orthologous blocks tend to have the same termini in both fish species. And (5), a minority of translocations occurred in one or the other fish lineage after divergence; for example, the region from 77 to 88 Mb and the region from 101 to 142 Mb are syntenic (on the same chromosome) in human (Hsa7) and in both ohnologons in stickleback

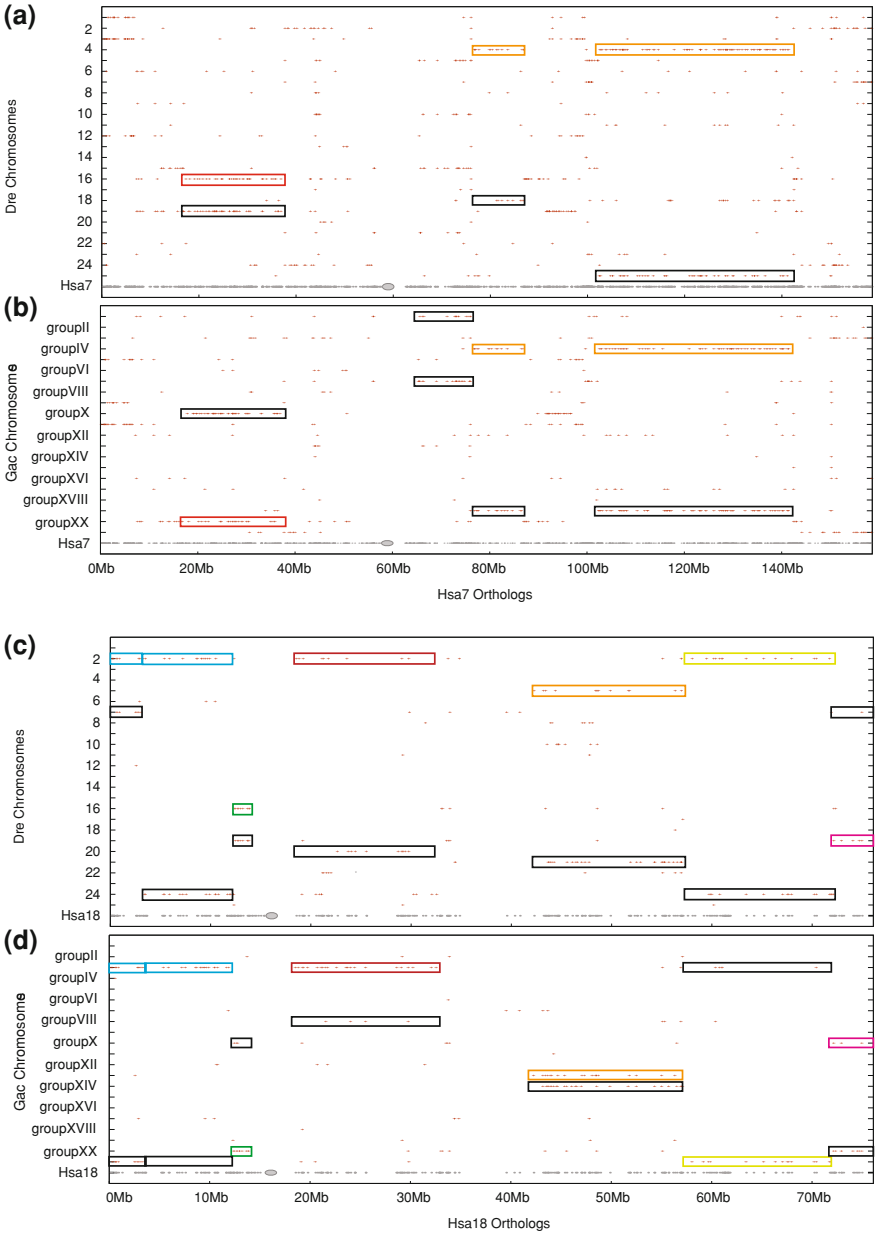

Fig. 17.5 Ohnologons in zebrafish (**a**, **c**) and stickleback (**b**, **d**) for Hsa7 (**a**, **b**) and Hsa18 (**c**, **d**). These plots show doubled conserved synteny as well as the pattern of chromosome translocations in various lineages

(groupIV and groupXX)—thus representing the ancestral condition—but for zebrafish, these two regions are syntenic for only one of the copies (Dre4) but not the other (Dre18 and Dre25). These results demonstrate a translocation in the zebrafish lineage after it diverged from the stickleback lineage (Fig. 17.5a, b). Note that the effects of inversions in the fish lineage are invisible in these plots because gene orders in the fish genomes are displayed according to their order in the human genome.

Having identified ohnologons, we can now pose the question of gene loss asymmetries. Over the human genome, we identified 49 unambiguous ohnologon pairs in zebrafish or stickleback or both, and counted the number of fish orthologs and paralogs of human genes occupying each segment. For zebrafish, 7 of 39 (17.9 %), and for stickleback, 13 of 44 (29.5 %) ohnologon pairs had significantly different number of genes [χ^2 test, 1 df, $p < 0.01$; 10/39 (25.6 %), which increased to 18/44 (40.9 %) if $p < 0.05$]. We conclude: (1) that most ohnologons appear to lose genes approximately randomly between the two duplicated copies, but (2) that about 20–40 % of ohnologons in these fish species have resolved TGD ohnologs asymmetrically between the two duplicated chromosome segments, with one member of the pair retaining significantly more genes than the other. (3) Of 13 segments with significance at the 0.01 level in one or the other fish, 6 are significant at 0.01 or 0.05 level in both species, which diverged deeply in teleost phylogeny (Fig. 17.1), and for all of those pairs, chromosome segments retaining the most genes in zebrafish are orthologs of those retaining the most in stickleback. This result is consistent with the idea that the basis for most asymmetric gene losses occurred before the divergence of zebrafish and stickleback lineages. A question that remains is the mechanism that led to asymmetric resolution of gene duplicates after the TGD. One hypothesis is that, if a long-range regulatory function that controls the expression of many genes in a neighborhood disappears from one ohnologon, then nonfunctionalization of the genes that the regulatory function controls can follow without penalty given the maintenance of that long-range function in the sister ohnologon. That hypothesis is yet to be seriously investigated. An example of resolution asymmetries is shown in Fig. 17.6 for Hsa15 in zebrafish and stickleback, along with chromosome diagrams comparing genes in their order along a segment of Hsa15 and their fish orthologs and co-orthologs in zebrafish (52 ortholog pairs in one ohnologon vs. 4 ortholog pairs in the other) and stickleback (54 ortholog pairs vs. 16 ortholog pairs) (Fig. 17.6c, d).

A related question is whether ohnologs residing in ohnologons that suffer asymmetric rates of gene loss also experience asymmetric rates of gene evolution. Asymmetric rates of molecular evolution have been found for a substantial number of TGD paralogs (Van de Peer et al. 2001; Steinke et al. 2006; Brunet et al. 2006), sometimes correlating with asymmetric loss of genes from ohnologons (Braasch et al. 2006; Siegel et al. 2007). Available evidence from other WGDs, however, suggests that gene evolutionary rate asymmetry does not strongly depend on the conservation of syntenies (Bu et al. 2011).

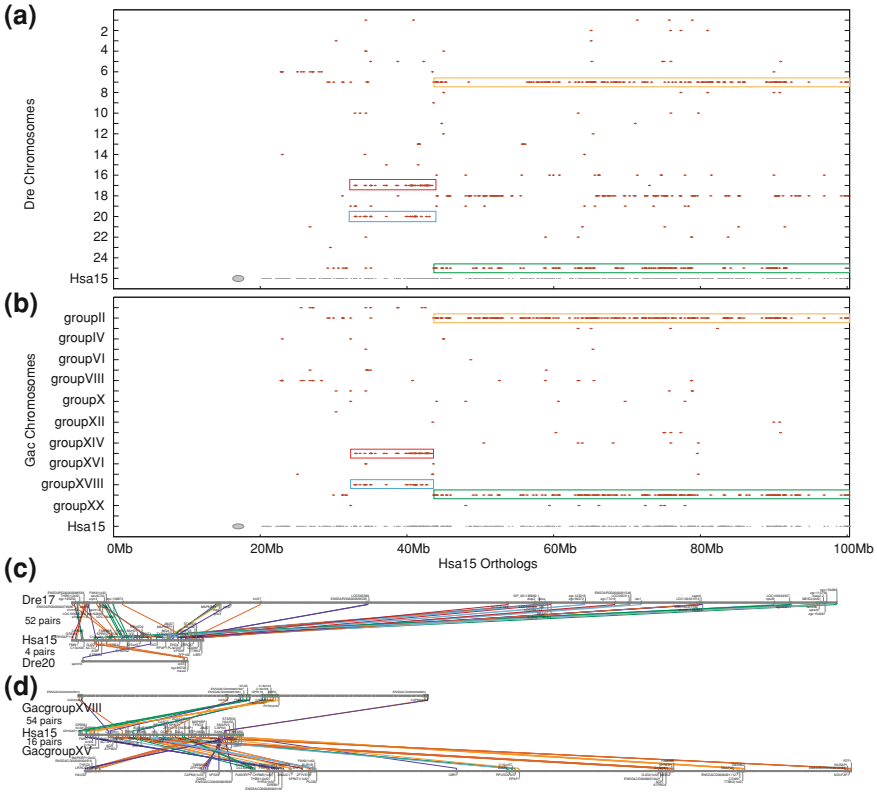

Fig. 17.6 Asymmetrically resolved ohnologons for Hsa15 in zebrafish (**a**) and stickleback (**b**). Chromosome contents displayed with proper gene orders for zebrafish duplicates (**c**, 52 vs. 4 ortholog pairs) and stickleback (**d**, 54 vs. 16 pairs)

17.3.2.5 Reconstruction of the Ancestral Karyotype and Teleost Chromosome Rearrangements

As we discussed earlier, the ancestral haploid teleost (post-TGD) genome was inferred to contain 24 chromosomes based on karyotype data (Ohno et al. 1968; Mank and Avise 2006b). Using several teleost genetic maps as well as the genome assemblies of *Tetraodon* and medaka, several attempts have been made to reconstruct the ancestral pre-TGD protokaryotype (Naruse et al. 2004; Jaillon et al. 2004; Woods et al. 2005; Kohn et al. 2006; Kasahara et al. 2007; Nakatani et al. 2007). All studies largely agree that the teleost karyotype is derived from 11 to 12 pre-TGD protochromosomes, which is in agreement with the inferred number of 24 post-TGD chromosomes (Mank and Avise 2006b). There are, however, some differences in the assignment of chromosomal blocks from different extant teleosts to the protochromosomes (Woods et al. 2000; Kasahara et al. 2007). The reconstruction of the ancestral teleost karyotype has so far relied on using tetrapod

genomes as outgroups, but the genome sequence of gar or bowfin, rayfin fish that can serve as outgroup to the TGD, will help to provide a more detailed picture of the ancestral teleost genome.

Nevertheless, the current state of karyotype reconstruction is helpful to compare and cross-refer different types of data on duplicated genes in teleosts. As an example, for zebrafish and stickleback we downloaded all pairs of predicted duplicated genes and their genomic locations from the EnsemblCompara GeneTree set (Vilella et al. 2009) and parsed them for the 'Clupeocephala' duplication node, which indicates a duplication on the branch leading to teleosts, after divergence from tetrapods, but before the divergence of teleosts. We then plotted the location of these pairs as Oxford grids for zebrafish and stickleback genomes (Fig. 17.7) and overlaid them with the 12 ancestral pre-TGD protochromosomes *a–m* inferred from the analysis of the medaka genome (Kasahara et al. 2007). This analysis clearly shows that paralogs are nonrandomly distributed over the genome; for example, of the 228 genes in the analysis that reside on zebrafish chromosome Dre3 and have TGD duplicates, the vast majority, 105 genes, are on Dre12 and a sizable minority on Dre1, and of the 167 genes on stickleback groupXIV that are duplicated and assigned to chromosomes, 152 are on groupXIII. These results show that the vast majority of 'Clupeocephala-duplicated' genes obtained from the tree-only method used by Ensembl supports a TGD origin of the paralogs and is in line with conserved synteny data and ancestral karyotype reconstructions.

Teleost genomes are rearranged with respect to tetrapod genomes, which may be due to chromosome rearrangements that were facilitated by the TGD, for example through illegitimate recombination between homeologous (paralogous) chromosomes. This hypothesis, however, is controversial (Comai 2005; Semon and Wolfe 2007a; Hufton et al. 2008), because chromosome restructuring may have likewise occurred before the TGD, on the long branch separating teleosts and tetrapods after the divergence of rayfin fish from lobefin fish.

Using the genetic map of the spotted gar, which contains nearly 1,000 coding markers, Amores et al. (2011) recently showed unexpectedly high conservation of synteny between human and gar when compared to zebrafish and stickleback. This suggests that chromosome rearrangements and the loss of ancestral syntenies accelerated after the TGD, but before the divergence of stickleback from zebrafish, and supports the hypothesis that WGD can facilitate syntenic rearrangements.

The zebrafish genome appears to be more rearranged than the genome of percomorph teleosts (stickleback, medaka, pufferfish). Chromosomal blocks of the inferred ancestral pre-TGD karyotype (Kasahara et al. 2007; Nakatani et al. 2007) are generally distributed over many more chromosomes in zebrafish than in percomorphs. For example, as shown in Fig. 17.7, chromosomal blocks of the ancestral pre-TGD chromosome *m* are found on four stickleback chromosomes (GacI, GacIII, GacVIII, GacXXI), but on six chromosomes in zebrafish (Dre2, Dre6, Dre8, Dre11, Dre22, Dre24); pieces of ancestral chromosome *j* are found on two chromosomes in stickleback (GacII, GacXIX), but on three in zebrafish (Dre7, Dre18, Dre25), and so forth. A genome sequence of another representative of the ostariophysians, which, after percomorphs, is the next most species-rich group of

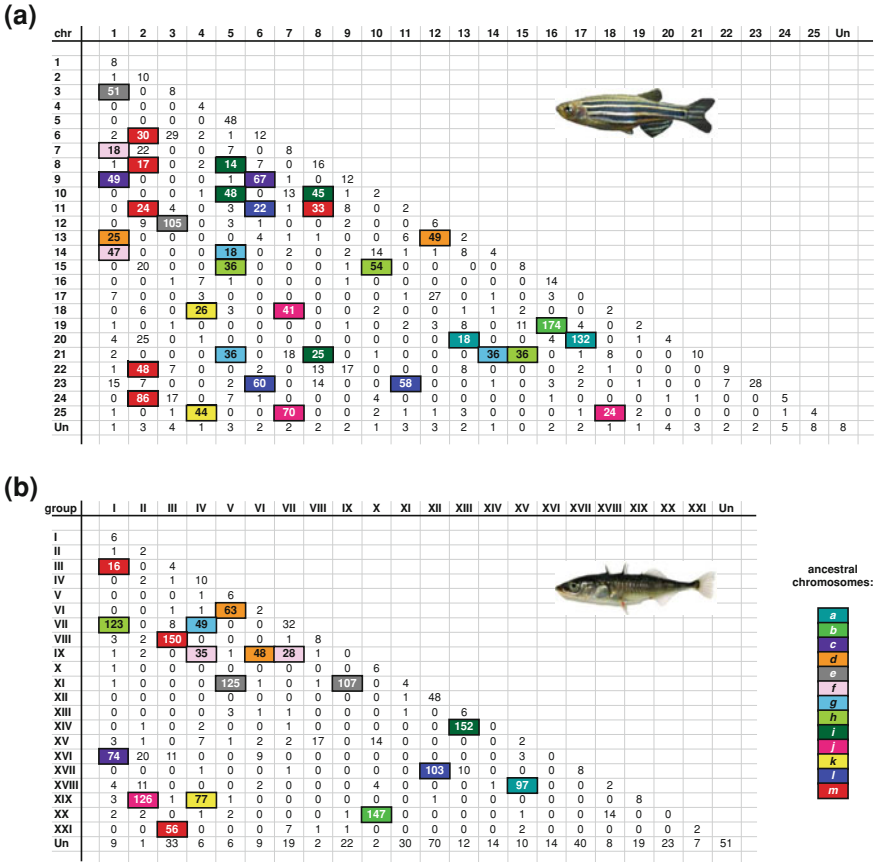

Fig. 17.7 Oxford grid of duplicated genes in zebrafish (**a**) and stickleback (**b**). Chromosomal location of paralogs determined by EnsemblCompara GeneTrees (Vilella et al. 2009) to be generated at the 'Clupeocephala' node. Each *box* in the grid shows the number of duplicated genes shared between the indicated chromosomes. The distribution of duplicated genes is nonrandom as expected from the TGD. Most clusters of duplicated genes can be traced back to the ancestral protochromosome (Kasahara et al. 2007), from which they are derived (see color code)

teleosts and includes zebrafish and carps (Nelson 2006), will be necessary to clarify whether this enhanced rearrangement rate is specific to the zebrafish genome or a more general phenomenon on this major branch of teleost diversity.

17.3.2.6 Global Patterns of Gene Retention and Loss After the TGD

The TGD obviously had a major impact on the evolution of genome structure in teleosts. We still do not know, however, how many paralogs from the TGD have escaped nonfunctionalization and have been retained in duplicate in extant

teleosts. Genome-wide estimates range between 12 and 24 % depending on the approach and dataset (Postlethwait et al. 2000; Jaillon et al. 2004; Woods et al. 2005; Brunet et al. 2006; Kassahn et al. 2009) and do not differ significantly among sequenced teleost genomes (Kassahn et al. 2009). Between ~1,800 and ~2,200 TGD-derived pairs of gene duplicates can be identified in the five sequenced teleost genomes (I. Braasch, "unpublished data"). Loss of paralogs after the TGD appears to be time-dependent, with most losses occurring soon after the GD event (Braasch et al. 2009a; Sato et al. 2009). A similar trend can be observed when comparing paralog loss after different GD events in teleosts, from the CaGD ~11–21 million years ago (mya) with a paralog retention rate of ~60 % (David et al. 2003), over the sucker genome duplication (SuGD) and SaGD, which occurred ~50 mya and 25–100 mya, respectively, and which both have estimated paralog retention rates of ~50 % (Ferris 1984; Allendorf and Thorgaard 1984), to the TGD that happened ~226–350 mya and that has an estimated paralog retention of ~12–24 % (see above).

Retained paralogs from the TGD are enriched for transcription factors, developmental genes, and cell communication proteins (Brunet et al. 2006; Kassahn et al. 2009), similar to results for the earlier rounds of WGD in vertebrates (see Putnam et al. 2008; Huminiecki and Heldin 2010; Chap. 16, this volume). Some gene classes appear to have a high rate of TGD paralog retention, such as genes encoding components of pigmentation pathways (Braasch et al. 2007, 2009a) and of signaling pathways involved in long-term potentiation of synaptic transmission, and olfactory and taste transduction (Sato et al. 2009). However, a more detailed analysis of gene family evolution after the TGD awaits the genome assembly of a rayfin outgroup to the TGD, the spotted gar (see below).

17.3.3 Functional Divergence of Gene Duplicates: Examples in Fish

After a polyploidization event, ohnolog pairs are initially identical, both in regulatory elements and in protein-coding areas, so they are fully redundant. Ohnologs, however, soon begin to accumulate mutations that become fixed in populations and distinguish the *a* copy from the *b* copy; these mutations can occur in both regulatory and protein coding regions.

Some mutations can cause one copy of a duplicated pair of genes to become nonfunctional, either because they inactivate the protein by premature stop mutation or destructive amino acid substitution, or because they eliminate gene expression. Nonfunctionalization is the most frequent fate of gene duplicates (Lynch and Conery 2000). For genes that are not dosage-sensitive, nonfunctionalization of one copy is likely to carry little selective penalty (and may provide in some cases an advantage), reflecting the fact that most mutations are recessive in a diploid.

Besides nonfunctionalization of one member of a duplicate pair, reciprocal changes to both members of the pair cause both ohnologs to become essential. Reciprocal changes can be of two main sorts, either reciprocal loss of ancestral gene subfunctions that occur purely by neutral processes (subfunctionalization), or by mutation in one or both copies that result in both ohnologs being positively selected for new functions (neofunctionalization) (Force et al. 1999). Mutations that lead to the partitioning of subfunctions between ohnologs do not preclude the later origin of novel subfunctions [called subneofunctionalization (He and Zhang 2005; Rastogi and Liberles 2005)] and vice versa. Here, we note a few examples for each of several fates of TGD ohnologs.

17.3.3.1 Subfunctionalization

The classic example that led to the idea of subfunctionalization—duplicate gene preservation by the reciprocal loss of ancestral gene subfunctions—is the zebrafish pair of *engrailed*-1 co-orthologs (Force et al. 1999). The mouse *En*1 gene is expressed in both the pectoral appendage bud and in a specific set of hindbrain interneurons (Joyner and Martin 1987; Davis et al. 1991; Gardner and Barald 1992). In contrast, the zebrafish co-orthologs *eng*1*a* and *eng*1*b* partitioned these two expression domains between them, with *eng*1*a* expressed in the pectoral appendage bud and, reciprocally, *eng*1*b* expressed in the hindbrain interneurons, a case of *spatial subfunctionalization* (Force et al. 1999). The reciprocal loss of ancestral regulatory domains like this, which can occur purely by neutral evolution, makes both copies essential, with *eng*1*a* essential to provide Engrailed function to the pectoral fin bud and *eng*1*b* necessary to deliver Engrailed protein to interneurons.

Many genes evolved like *eng*1 genes to show *spatial subfunctionalization*, but some genes evolved by *temporal subfunctionalization*, duplicate gene expression in the same tissue but at multiple developmental stages. For example, *Sox*1 is expressed in tetrapods at several developmental stages in the lens, but after the TGD, *sox*1*a* came to be expressed several hours earlier in the lens than is *sox*1*b* (Okuda et al. 2006). Subfunctionalization is pervasive: a global analysis of TGD ohnologs showed that nearly all that were examined had diverged in their patterns of spatial and/or temporal expression during embryogenesis and that such regulatory changes were more frequent than changes at the level of protein function (Kassahn et al. 2009). Tetrapod enhancers, which represent a type of regulatory subfunction, and CNEs, many of which may be enhancers, evolved divergent sequences in teleosts by altering a few specific bases that underlie divergent functions, as for medaka and fugu ohnologs derived from *Hoxa*2 (Tumpel et al. 2006) or zebrafish and stickleback *Fgf*8 co-orthologs (Canestro et al. 2007); this result would be expected to occur by subfunctionalization. Cases of preserved ohnologs that have no documented differences in embryonic expression domains may display differences at developmental stages or tissues that have not yet been examined.

Besides spatial or temporal differences in expression domains, some ohnologs may achieve *quantitative subfunctionalization*, the accumulation of activity-reducing mutations in both ohnologs, so that neither gene by itself can provide sufficient product to achieve normal function (Force et al. 1999; Stoltzfus 1999). Mutations that partially reduce gene function should be relatively common, and if a population fixes a partial loss-of-function allele at one copy, then it is unlikely to fix a null allele in the gene's ohnolog. Although quantitative subfunctionalization may be common after WGD in yeast (Scannell and Wolfe 2008), few examples exist from the TGD, probably because of the lack of appropriate analyses. We do know that quantitative subfunctionalization occurred in many of the *hox* gene duplicates after the SaGD (Mungpakdee et al. 2008b). The finding that most pairs of CNEs in teleost ohnologs experience substantial degeneration in element length might also reflect cases of quantitative subfunctionalization (Woolfe and Elgar 2007). The evolutionary significance of quantitative subfunctionalization may be that it preserves both duplicates long enough to increase the opportunity for later mutations that might secondarily increase gene activity, as appears to have happened for the medaka copy of *aldh*1*a*2, a gene originating in VGD2 (Canestro et al. 2009).

Protein structure evolution: Individual tetrapod genes can encode proteins with multiple functions. In some cases, a single protein sequence performs several functions; in other cases, a single gene encodes different protein sequences, often with different functions, due to alternative splicing or start sites.

Amino acid substitutions can cause proteins encoded by a pair of ohnologs to diverge in sequence in functionally important ways. For example, the chemokine Cxcl12 guides gastrulation and the migration of primordial germ cells (PGCs) in several vertebrates (Nair and Schilling 2008; Richardson and Lehmann 2010). Zebrafish has co-orthologs of Cxcl12 that appear to have arisen in the TGD (Doitsidou et al. 2002). Cxcl12 functions by binding Cxcr4 receptor, which is also duplicated in zebrafish. Cxcl12a, but not Cxcl12b, guides PGCs to their target, the gonad, and a single amino acid substitution causes the relative affinity of Cxcl12 ligands to switch from one of the duplicated Cxcr4 receptors to the other, thereby controlling whether the Cxcl12:Cxcr4 ligand:receptor system controls PGC migration or gastrulation (Boldajipour et al. 2011). These experiments are a particularly deeply studied case study supporting the concept of protein subfunctionalization.

Alternative splicing can generate different proteins with variant functions. For example, the human *STAT*1 gene, which mediates interferon signaling, makes two different proteins by alternative splicing: STAT1-alpha, which contains a TAZ2-binding domain, and STAT1-beta, which lacks that domain and may act as a negative regulator of STAT1-alpha (Bromberg et al. 1996). The TGD provided teleosts with duplicate *stat*1 genes, and, due to reciprocal mutations that affect transcript splicing, the zebrafish *stat*1*a* gene came to encode a protein that contains the TAZ2-binding domain like the human STAT1-alpha isoform, while *stat*1*b* evolved to encode a protein that lacks this domain like STAT1-beta (Song et al. 2011). Thus, due to subfunctionalization of splicing signals, two genes in

zebrafish encode two protein variants that are encoded by a single gene in human. Another example is the synapsin gene in fugu (Yu et al. 2003).

Alternative start sites can also result in alternative exon usage. The human *MITF* gene controls pigment production both in melanocytes and in the pigmented retina, with different transcription start sites employed by the two different cell types (Udono et al. 2000). Zebrafish and other teleosts have two *MITF* co-orthologs, with *mitfa* expressed in melanophores and *mitfb* in the pigmented retina; correspondingly, evolution of the two ohnologs resulted in *mitfa* using one transcription start site and *mitfb* using the other (Lister et al. 2001; Altschmied et al. 2002). Again, two different tissue-specific proteins encoded by one gene in mammals are formed from sister ohnologs in teleosts due to subfunctionalization.

17.3.3.2 Neofunctionalization

In the race between nonfunctionalization, subfunctionalization, and neofunctionalization, nonfunctionalization generally rides the fastest horse. Subfunctionalization appears to place second, and neofunctionalization—the preservation of ohnologs by the origin of new functions (Ohno 1970)—shows in third. As with subfunctionalization, neofunctionalization can occur at the level of either gene regulation or protein structure. Preservation by neofunctionalization differs from subfunctionalization in two main ways: first, it involves positive selection rather than neutral events, and second, it involves the origin of gene subfunctions that did not exist in the ancestral gene rather than the partitioning of ancestral subfunctions. Genome-wide studies suggest that the acquisition of novel protein domains might have occurred in a quarter of duplicate pairs after the TGD and happened more frequently in duplicates than in single-copy genes (Kassahn et al. 2009). Although several specific cases of neofunctionalization after the TGD have been suggested (Yao and Ge 2010; Sha et al. 2008; Howarth et al. 2008), conclusions are often difficult due to assumptions about the ancestral state, which is usually based on tetrapods, species that are far removed from the teleost lineage.

To distinguish neofunctionalization from subfunctionalization, one must accurately identify the ancestral state. Inferences about the ancestral state can be difficult with WGD but are often easier with tandem duplications, which have led to neofunctionalization of vitellogenin and aquaporin genes in teleost egg hydration (Finn and Kristoffersen 2007; Zapater et al. 2011; Cerda 2009), the evolution of antifreeze proteins by Antarctic teleosts (Deng et al. 2010), and the preservation of fatty acid-binding protein genes (Karanth et al. 2009).

Several examples of neofunctionalization after the TGD have been suggested. (1) Phosphoglucose isomerase (PGI) ohnologs from the TGD, after the partitioning of tissue-specific regulatory elements for expression, diverged so that the copy expressed in organs of the body cavity became more negative in charge while the muscle isoform became more positively charged, although the functional significance of these changes is not yet well understood (Sato and Nishida 2007). (2) After the TGD, cyprinid fish, like zebrafish, retained only one copy of the

androgen receptor gene (*ar*), but percomorph fish, like pufferfish and medaka, retained both ohnologs, one of which (*arb*) experienced substantial sequence alterations that are consistent with neofunctionalization; further, these changes occurred after the divergence of basal teleost lineages, including eels, but before the radiation of percomorphs (Douard et al. 2008). (3) Some TGD duplicates show elevated evolutionary rates, like *csf*1*rb* and its neighbor *pdgfrbb* in cichlids, which could be due either to neofunctionalization or to relaxed selection; it has been suggested that such behavior could be a causal factor in the evolution of the spectacular coloration patterns of some teleost fish (Braasch et al. 2006). To learn whether neofunctionalization is at play in these examples and many others, the duplicates must be compared to each other by function in addition to sequence and compared to the structure and function of the ancestral genes in a rayfin fish that diverged before the TGD.

In two cases of neofunctionalization—a *hox* gene and a sodium channel gene—TGD duplicates have been investigated functionally in a phylogenetic context that leaves little doubt about the functions of the preduplication gene. Zebrafish has co-orthologs of *Hoxa*13 derived from the TGD (Amores et al. 1998), and, as determined by a broad phylogenetic study, these ohnologs began to diverge asymmetrically long after their initial preservation subsequent to the TGD (Crow et al. 2009). Ohnolog knockdown experiments in zebrafish, coupled with rescue experiments, showed that *hoxa*13*a,* the faster evolving ohnolog, contains many derived *a* clade-specific amino acid replacements that are necessary for development of the yolk sac extension (the hind yolk), a derived feature in zebrafish and its relatives (Crow et al. 2009).

A great example of neofunctionalization after the TGD occurred in electric fish. The distantly related African elephantfish (Mormyridae; Osteoglossiformes) and South American knifefish (Gymnotideae; Gymnotiformes) independently evolved a muscle-derived electric organ, which facilitates electrical communication and utilizes the voltage-gated sodium channel subunit Scn4a. Teleosts have two *scn*4*a* genes from the TGD, both expressed in skeletal muscle in most teleosts; in two lineages of electric fish, however, *scn*4*aa* independently lost its expression in skeletal muscle and gained expression in the electric organ (Arnegard et al. 2010). Phylogenetic analyses showed that this neofunctionalization event occurred more than 100 million years after the TGD with a signal of positive selection (Arnegard et al. 2010). Meanwhile, *scn*4*ab* maintained the ancestral expression in skeletal muscle. These results for Scn4a and Hoxa13 both show that, long after the preservation of ohnologs by subfunctionalization, one copy can still provide the genetic variation necessary to lead to the innovation of new morphologies and functions.

17.3.3.3 Spotted Gar as a Rayfin Outgroup for Functional Studies of TGD Paralogs

As pointed out above, knowledge of the functions of the preduplication gene is essential for inferring subfunctionalization versus neofunctionalization. Investigators often infer the properties of the pre-TGD gene from the phenotype of well-

studied tetrapods, especially mouse. The tetrapod lineage, however, diverged from the teleost lineage about 450 million years ago and involved the evolution of substantial differences in morphologies from ancestral bony fish. In addition, rayfin fish followed an independent lineage for about 200 million years before the TGD; thus, what appears to involve neofunctionalization when comparing teleost co-orthologs to tetrapod genes (Fig. 17.8a) could instead be due to the tetrapod lineage having lost a subfunction originally present in the single-copy gene in the last common ancestor of teleosts and tetrapods (Fig. 17.8b). A third alternative is the reciprocal loss of ancestral VGD ohnologs, which could also mimic neofunctionalization, pointing out the importance of careful distinction between orthologs and ohnologs (Fig. 17.8c). Finally, the novel function thought to have arisen after the TGD might have actually originated in ancestral rayfin fish, pointing out the importance of identifying a fish representing this node (Fig. 17.8d). In addition, distinguishing among the possibilities shown in Fig. 17.8b–d requires examination of an outgroup diverging basal to the diagrammed tree, such as a cartilaginous fish.

Recently, the spotted gar *L. oculatus* has been identified as an unduplicated rayfin outgroup to the TGD (see Sect. 17.3.1.5 and Amores et al. 2011). The spotted gar occupies a lineage of 'ancient fish' that includes not only lepisosteids (gars), but also bowfin (*Amia calva*) and possibly acipenseriforms (sturgeons, e.g. *Acipenser transmontanus*, and paddlefish, e.g. *Polyodon spathula*) (Inoue et al. 2003; Kikugawa et al. 2004; and Fig. 17.1). Among these 'ancient fish', spotted gar is the species that has the smallest genome and can most readily be fertilized in vitro, providing embryos that are amenable to in situ hybridization studies and that grow to adults in laboratory aquaria with little care. Furthermore, no lineage-specific polyplodization has occurred in gar, in contrast to acipenseriforms. Thus, the spotted gar makes an ideal outgroup for the investigation of the mechanisms of evolution of gene function after GD.

17.3.4 Mind the Gap: The TGD and the Teleost Radiation—Is There a Connection?

The preceding paragraphs demonstrate that rayfin species experienced many WGD events and that teleosts are the most species-rich group of vertebrates. Importantly, many lineages experiencing WGD, such as barbs, armored catfish, and salmonids, appear to be particularly species-rich (Le Comber and Smith 2004). These considerations raise the question of whether WGD facilitates lineage diversification in fish. Ever since its discovery, the TGD has been suggested to have had a major impact on the radiation and biodiversity of the teleost lineage as a whole (Amores et al. 1998; Wittbrodt et al. 1998; Meyer and Schartl 1999; Taylor et al. 2001, 2003; Postlethwait et al. 2004; Volff 2005; Meyer and Van de Peer 2005; Froschauer et al. 2006; Ravi and Venkatesh 2008; Volff et al. 2011; Christoffels

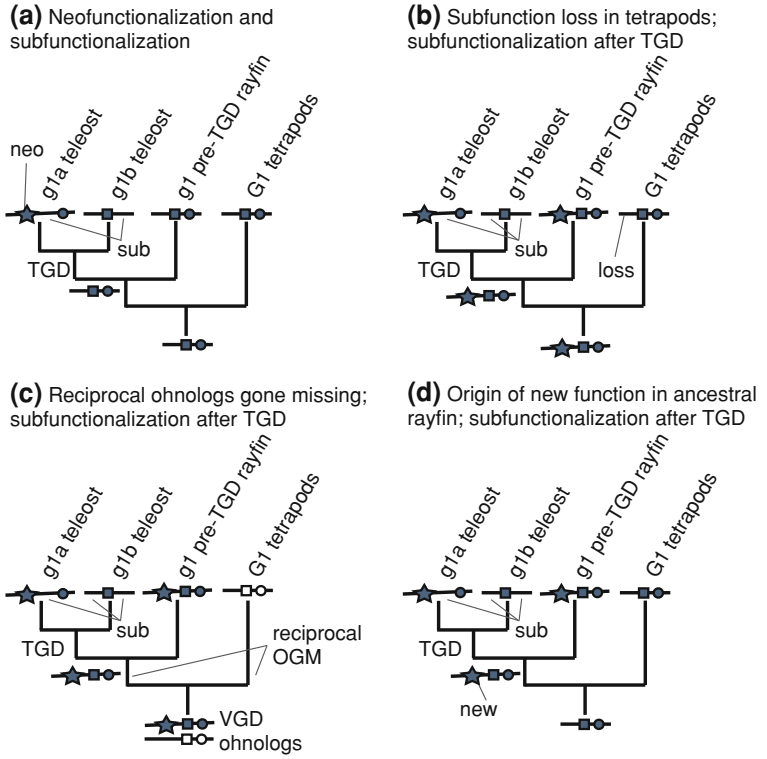

Fig. 17.8 Functional analysis of a pre-TGD diverging rayfin fish is important to distinguish neofunctionalization from subfunctionalization. **a** Preservation of one ohnolog by neofunctionalization (the function symbolized by a star) and preservation of the other by subfunctionalization. **b** Preservation of teleost ohnologs by subfunctionalization; the 'star' function was ancestral, not newly originated after the TGD, but was lost in the tetrapod lineage, as demonstrated by the pre-TGD-diverging rayfin outgroup. **c** The 'star' function that appears to be newly originated after the TGD was ancestral in a VGD ohnolog that was lost in the tetrapod lineage, while the rayfin lineage reciprocally lost the other ohnolog. **d**. A novel function not present in the ancestral gene could have originated in stem rayfins and then have been lost in one TGD ohnolog. Distinguishing among **b**, **c** and **d** requires an additional outgroup, such as a cartilaginous fish

et al. 2004; Van de Peer et al. 2009), but likewise, this has been questioned by others based on the fossil record and speciation rate analyses (Donoghue and Purnell 2005; Hurley et al. 2007; Santini et al. 2009).

17.3.4.1 Testing Speciation Models in the Fish World

Genetic principles argue that a WGD event that is followed by population isolation, lineage-specific reciprocal nonfunctionalization events, and later hybridization would tend to promote speciation (Lynch and Force 2000; Werth and

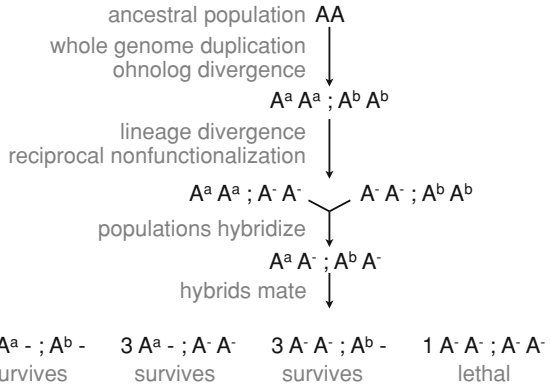

Fig. 17.9 How WGD could facilitate lineage diversification. 'A' indicates the original gene. 'A^{a}' and 'A^{b}' indicate functional but diverged ohnologs. 'A^{-}' indicates a non-functionalized allele. A '-' indicates an allele that could be either functional or nonfunctional. The analysis assumes that non-functional alleles are recessive to functional alleles

Windham 1991). As Fig. 17.9 shows, the fertility of hybrids would be expected to be somewhat reduced if even a single duplicated gene had reciprocal nonfunctionalization, but if reciprocal nonfunctionalization (or reciprocal gene loss) occurred at least once on several or many chromosomes, which is likely, then hybrids would have greatly reduced fertility, thus strengthening lineage divergence.

Regarding the TGD, while one study estimated ~7 % reciprocal gene loss rate between zebrafish and *Tetraodon*, which would amount to ~1,700 genes (Semon and Wolfe 2007b), more recent work including more teleost species found little support for reciprocal gene loss coinciding with speciation events in teleosts (Kassahn et al. 2009). The latter is more in line with our observation that differences in asymmetric gene loss from paralogons are rare between zebrafish and stickleback (see Sect. 17.3.2.4).

Similar to reciprocal nonfunctionalization, speciation after WGD may also be promoted by reciprocal subfunctionalization that results in genomic incompatibilities in post-WGD hybrids (Lynch and Force 2000; Werth and Windham 1991). Differential functional evolution of TGD paralogs in divergent teleost lineages is being found in more and more cases (see below), yet associations with genetic incompatibilities and speciation in teleosts remain to be shown.

Finally, it has also been suggested that the TGD may have caused reduced probability of extinction in teleosts (Crow and Wagner Crow and Wagner 2006). The extinction rate of teleost lineages, however, is apparently *higher* than in non-teleost rayfin lineages (Santini et al. 2009).

17.3.4.2 TGD-Based Morphological Evolution in Teleosts

Besides their species richness, teleosts are also extremely diverse in terms of morphology, physiology, behavior, ecology, and biogeography. If the TGD laid the groundwork for divergence in such traits, one would expect to find differential function of TGD paralogs in various teleost lineages. Testing this proposition

involves experimentally challenging studies in different fish species, and only a handful of examples are currently available. For teleost *sox*9 co-orthologs, several ancient expression domains are differentially retained in zebrafish, stickleback, and medaka, indicating lineage-specific subfunctionalization (Cresko et al. 2003; Kluver et al. 2005; Yokoi et al. 2009); similar lineage-specific partitioning of expression domains have been reported for *fgf*8 (Jovelin et al. 2007) and *tyrp*1 (Braasch et al. 2009b) co-orthologs.

A better way to compare gene functions across teleosts is to compare mutant phenotypes. Unfortunately, few genes have yet been mutated in teleosts other than zebrafish and medaka, and still fewer have been mutated in both species. Most teleosts have two TGD ohnologs encoding FGF receptor-1; in zebrafish and carp, both cyprinids, mutation of *fgfr*1*a* causes a scaleless phenotype (Rohner et al. 2009). In contrast, mutation of *fgfr*1*a* in medaka, a perciform fish, deletes the trunk and tail (the *headfish* mutant) (Yokoi et al. 2007; Shimada et al. 2008). This finding shows that subfunction partitioning in *fgfr*1 ohnologs occurred after the divergence of medaka and zebrafish lineages. Such divergence in subfunction partitioning after the TGD may provide the genomic basis for some morphological divergence in teleosts.

And what about the TGD and morphological complexity? The identification of the seven *hox* clusters in teleosts, generally not considered to be more complex than other vertebrate groups, called into question the idea that morphological complexity along the body axis and the number of *hox* clusters and genes may be directly linked (Amores et al. 1998; Prince 2002). The high retention rate of TGD paralogs for developmentally important genes (Brunet et al. 2006; Kassahn et al. 2009), however, argues for an involvement of at least some TGD paralogs in teleost morphological evolution. Also, some aspects of teleost physiology and morphology are unique among vertebrates. To these belong, for example, the complexity and diversity of teleost body pigmentation and color patterning (Braasch et al. 2008); some functional modules of pigment cell development and differentiation have been retained in large part in duplicate in teleosts after the TGD (Braasch et al. 2009a).

Another question that has so far been rarely addressed but will need to be investigated in more detail is the possible involvement of TGD duplicates in morphological and physiological novelties. For example, is functional divergence of TGD paralogs causing some of the synapomorphies of the teleost lineage (de Pinna 1996), such as the truly symmetric (homocercal) tail fin? Or are functional shifts in TGD paralogs involved in lineage-specific key innovations? We met one intriguing case earlier, the parallel, independent gain of electric organs used for electrical communication in knifefish and elephantfish by neofunctionalization of the *scn*4*aa* paralog (Arnegard et al. 2010). A more comprehensive picture, however, can only be gathered by studying TGD paralogs in more than just a few teleost representatives and adding an appropriate rayfin outgroup to the analyses.

17.3.4.3 The Significance of the TGD for the Teleost Radiation

Based on the fossil record, a major temporal delay of about 150 million years separated the TGD event and the major radiation of the teleost lineage in the percomorphs (Donoghue and Purnell 2005; Hurley et al. 2007). Recent diversification rate shift analyses lead to the recognition that three incidents of accelerated speciation rates occurred within rayfin fish: on the branches leading (1) to the teleosts, (2) to the ostariophysians (zebrafish and relatives), and (3) to the percomorphs (perch-related fish, including medaka, stickleback, and pufferfish), respectively (Santini et al. 2009; Alfaro et al. 2009). Diversification rate shifts shortly after the TGD accounted for about 10 % of species diversity in teleosts (Santini et al. 2009). As noted above, however, TGD duplicates can alter their functions hundreds of millions of years after the GD itself. Whether individual fish taxa with additional WGD events (salmonids, barbs, etc.) are actually more species rich than diploid taxa has not been critically tested.

For plants, it has recently been shown that recent polyploids had lower rates of speciation and higher rates of extinction than diploid control lineages (Mayrose et al. 2011). In contrast to recent polyploids, the number of ancient polyploidization events that provided today's flora is higher than expected if diversification rates of diploids and polyploids were equal (Mayrose et al. 2011). Furthermore, ancient WGDs appear to be associated with several major radiations in the angiosperms (Soltis et al. 2009). In contrast, the ratio of species death rate (extinction) to species birth rate is 1.5 times higher in teleosts than in non-teleost rayfin fish (Santini et al. 2009). Thus, crucial questions are, *when* and *under which circumstances* do we expect a polyploidization event to deploy its full potential for lineage diversification and radiation? Given the deep influence of the TGD on genome evolution in teleosts illustrated in this chapter and given the described potential for ongoing non-, sub-, and neofunctionalization long after the TGD event, its major impact may not be apparent immediately after the GD event. Obviously, the TGD event was not an evolutionary dead-end, but quite the contrary should be considered a genomic exaptation that generated countless "spandrels" (Gould and Lewontin 1979) in the teleost genome upon which selection could act long after the actual polyploidization event, when "opportunity meets preparation".

17.4 Outlook

As we have seen—although recognized since early studies of vertebrate genomes—the apparent propensity for polyploidization in the rayfin fish lineage, and among them, of cyprinid teleosts in particular, still requires an explanation. Since the first genomic evidence was published in 1998, the TGD hypothesis has overcome initial criticism, and the TGD is so far the best-studied GD in "fish". We are confident that many questions regarding the evolution of genome structure,

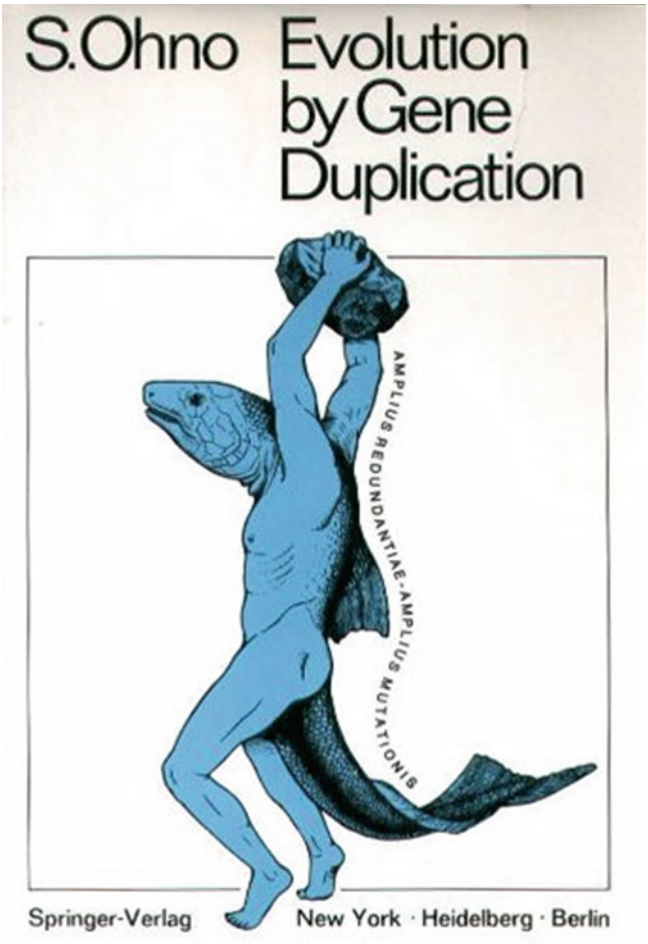

Fig. 17.10 The cover of Susumo Ohno's book (1970) underlines the importance of fish for understanding our own past

gene families, and regulatory circuitry after the TGD can be solved by adding the sequences of basal rayfins (such as the spotted gar) and of basal teleosts to the genomic pond. The even harder question about the significance of the TGD for the teleost radiation will require extensive work on gene function evolution at all levels of teleost diversity, including numerous non-model fish species. The advent of next-generation sequencing techniques and the application of developmental biology studies to an increasing arsenal of fish species are promising for the acceleration of important research opportunities in this direction. Yet other questions, for example whether the TGD was an auto- or an allopolyploidization event, may never be solved due to the millions of years that elapsed since the event.

Studying other, more recent GDs in rayfins, such as the salmonid and carp polyloidizations, will help to reveal in more detail the immediate epigenetic and genomic changes that follow a polyploidization. Luckily, the multitude and distribution of piscine GD events over evolutionary time will contribute to studies of the gradual evolution of vertebrate animals after different types of GD.

Finally, as illustrated so marvelously by the cover of Ohno's book (Ohno 1970) (Fig. 17.10), the study of polyploidizations in fish will not only allow us to better understand our own fishy heritage, but importantly, will help to inform us about the evolutionary processes that have shaped the vertebrate genome after VGD1 and VGD2, which had such a profound influence on our own, human evolution.

Acknowledgments We would like to thank Cristian Canestro for extensive discussions as well as Irene Pala for discussions of gene expression in the calandino. This work was supported by a grant from the Volkswagen Foundation Germany (IB) and National Institutes of Health grant R01 RR020833 (JHP).

References

Alfaro ME, Santini F, Brock C, Alamillo H, Dornburg A, Rabosky DL, Carnevale G, Harmon LJ (2009) Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. Proc Nat Acad Sci USA 106(32):13410–13414. doi:10.1073/pnas.0811087106 0811087106 [pii]

Allendorf FW, Thorgaard GH (1984) Tetraploidy and the evolution of salmonid fishes. In: Turner BT (ed) Evolutionary genetics of fishes. Plenum Press, New York, pp 1–53

Altschmied J, Delfgaauw J, Wilde B, Duschl J, Bouneau L, Volff JN, Schartl M (2002) Subfunctionalization of duplicate *mitf* genes associated with differential degeneration of alternative exons in fish. Genetics 161(1):259–267

Alves MJ, Coelho MM, Collares-Pereira MJ (2001) Evolution in action through hybridisation and polyploidy in an Iberian freshwater fish: a genetic review. Genetica 111(1–3):375–385

Amores A, Force A, Yan YL, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang YL, Westerfield M, Ekker M, Postlethwait JH (1998) Zebrafish *hox* clusters and vertebrate genome evolution. Science 282(5394):1711–1714

Amores A, Catchen J, Ferrara A, Fontenot Q, Postlethwait JH (2011) Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication. Genetics 188(4):799–808. doi:10.1534/genetics.111.127324 188/4/799 [pii]

Aparicio S, Hawker K, Cottage A, Mikawa Y, Zuo L, Venkatesh B, Chen E, Krumlauf R, Brenner S (1997) Organization of the *Fugu rubripes hox* clusters: evidence for continuing evolution of vertebrate *hox* complexes. Nat Genet 16(1):79–83. doi:10.1038/ng0597-79

Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A, Gelpke MD, Roach J, Oh T, Ho IY, Wong M, Detter C, Verhoef F, Predki P, Tay A, Lucas S, Richardson P, Smith SF, Clark MS, Edwards YJ, Doggett N, Zharkikh A, Tavtigian SV, Pruss D, Barnstead M, Evans C, Baden H, Powell J, Glusman G, Rowen L, Hood L, Tan YH, Elgar G, Hawkins T, Venkatesh B, Rokhsar D, Brenner S (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. Science 297(5585):1301–1310. doi:10.1126/science.1072104 1072104[pii]

Arnegard ME, Zwickl DJ, Lu Y, Zakon HH (2010) Old gene duplication facilitates origin and diversification of an innovative communication system-twice. P Nat Acad Sci USA 107(51):22172–22177. doi:10.1073/Pnas.1011803107

Balon EK (2004) About the oldest domesticates among fishes. J Fish Biol 65:1–27. doi:10.1111/j.1095-8649.2004.00563.x

Birstein VJ, Hanner R, DeSalle R (1997) Phylogeny of the Acipenseriformes: cytogenetic and molecular approaches. Environ Biol Fish 48(1–4):127–156

Boldajipour B, Doitsidou M, Tarbashevich K, Laguri C, Yu SR, Ries J, Dumstrei K, Thelen S, Dorries J, Messerschmidt EM, Thelen M, Schwille P, Brand M, Lortat-Jacob H, Raz E (2011) *Cxcl12* evolution—subfunctionalization of a ligand through altered interaction with the chemokine receptor. Development 138(14):2909–2914. doi:10.1242/Dev.068379

Braasch I, Salzburger W, Meyer A (2006) Asymmetric evolution in two fish-specifically duplicated receptor tyrosine kinase paralogons involved in teleost coloration. Mol Biol Evol 23(6):1192–1202. doi:10.1093/molbev/msk003 msk003 [pii]

Braasch I, Schartl M, Volff JN (2007) Evolution of pigment synthesis pathways by gene and genome duplication in fish. BMC Evol Biol 7:74. doi:10.1186/1471-2148-7-74 1471-2148-7-74 [pii]

Braasch I, Volff JN, Schartl M (2008) The evolution of teleost pigmentation and the fish-specific genome duplication. J Fish Biol 73(8):1891–1918. doi:10.1111/J.1095-8649.2008.02011.X

Braasch I, Brunet F, Volff JN, Schartl M (2009a) Pigmentation pathway evolution after whole-genome duplication in fish. Genome Biol Evol 1:479–493. doi:10.1093/gbe/evp050

Braasch I, Liedtke D, Volff JN, Schartl M (2009b) Pigmentary function and evolution of *tyrp1* gene duplicates in fish. Pigment Cell Melanoma Res 22(6):839–850. doi:10.1111/j.1755-148X.2009.00614.x PCR614 [pii]

Bromberg JF, Horvath CM, Wen ZL, Schreiber RD, Darnell JE (1996) Transcriptionally active Stat1 is required for the antiproliferative effects of both interferon alpha and interferon gamma. P Nat Acad Sci USA 93(15):7673–7678

Brunet FG, Crollius HR, Paris M, Aury JM, Gibert P, Jaillon O, Laudet V, Robinson-Rechavi M (2006) Gene loss and evolutionary rates following whole-genome duplication in teleost fishes. Mol Biol Evol 23(9):1808–1816. doi:10.1093/Molbev/Mls049

Bu L, Bergthorsson U, Katju V (2011) Local synteny and codon usage contribute to asymmetric sequence divergence of *Saccharomyces cerevisiae* gene duplicates. BMC Evol Biol 11:279. doi:10.1186/1471-2148-11-279 1471-2148-11-279 [pii]

Canestro C, Yokoi H, Postlethwait JH (2007) Evolutionary developmental biology and genomics. Nat Rev Genet 8(12):932–942. doi:10.1038/nrg2226 nrg2226 [pii]

Canestro C, Catchen JM, Rodriguez-Mari A, Yokoi H, Postlethwait JH (2009) Consequences of lineage-specific gene loss on functional evolution of surviving paralogs: ALDH1A and retinoic acid signaling in vertebrate genomes. PLoS Genet 5(5):e1000496. doi:10.1371/journal.pgen.1000496

Catchen JM, Conery JS, Postlethwait JH (2009) Automated identification of conserved synteny after whole-genome duplication. Genome Res 19(8):1497–1505. doi:10.1101/Gr.090480.108

Catchen JM, Braasch I, Postlethwait JH (2011) Conserved synteny and the zebrafish genome. Method Cell Biol 104:259–285. doi:10.1016/B978-0-12-374814-0.00015-X

Cerda J (2009) Molecular pathways during marine fish egg hydration: the role of aquaporins. J Fish Biol 75(9):2175–2196. doi:10.1111/j.1095-8649.2009.02397.x JFB2397 [pii]

Chenuil A, Galtier N, Berrebi P (1999) A test of the hypothesis of an autopolyploid vs. allopolyploid origin for a tetraploid lineage: application to the genus *Barbus* (Cyprinidae). Heredity (Edinb) 82(Pt 4):373–380. doi:her489 [pii]

Chiu CH, Dewar K, Wagner GP, Takahashi K, Ruddle F, Ledje C, Bartsch P, Scemama JL, Stellwag E, Fried C, Prohaska SJ, Stadler PF, Amemiya CT (2004) Bichir *HoxA* cluster sequence reveals surprising trends in ray-finned fish genomic evolution. Genome Res 14(1):11–17. doi:10.1101/gr.1712904 14/1/11[pii]

Christoffels A, Koh EG, Chia JM, Brenner S, Aparicio S, Venkatesh B (2004) Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. Mol Biol Evol 21(6):1146–1151. doi:10.1093/molbev/msh114 msh114 [pii]

Comai L (2005) The advantages and disadvantages of being polyploid. Nat Rev Genet 6(11):836–846. doi:10.1038/nrg1711

Cresko WA, Yan YL, Baltrus DA, Amores A, Singer A, Rodriguez-Mari A, Postlethwait JH (2003) Genome duplication, subfunction partitioning, and lineage divergence: *sox9* in stickleback and zebrafish. Dev Dyn 228(3):480–489. doi:10.1002/dvdy.10424

Crow KD, Wagner GP (2006) Proceedings of the SMBE tri-national young investigators' workshop. What is the role of genome duplication in the evolution of complexity and diversity? Mol Biol Evol 23(5):887–892. doi:10.1093/molbev/msj083 msj083 [pii]

Crow KD, Stadler PF, Lynch VJ, Amemiya C, Wagner GP (2006) The "fish-specific" Hox cluster duplication is coincident with the origin of teleosts. Mol Biol Evol 23(1):121–136. doi:10.1093/molbev/msj020 msj020 [pii]

Crow KD, Amemiya CT, Roth J, Wagner GP (2009) Hypermutability of *HoxA13A* and functional divergence from its paralog are associated with the origin of a novel developmental feature in zebrafish and related taxa (cypriniformes). Evolution 63(6):1574–1592. doi:10.1111/j.1558-5646.2009.00657.x EVO657 [pii]

Danzmann RG, Davidson EA, Ferguson MM, Gharbi K, Koop BF, Hoyheim B, Lien S, Lubieniecki KP, Moghadam HK, Park J, Phillips RB, Davidson WS (2008) Distribution of ancestral proto-actinopterygian chromosome arms within the genomes of 4R-derivative salmonid fishes (rainbow trout and Atlantic salmon). BMC Genomics 9:557. doi:10.1186/1471-2164-9-557

David L, Blum S, Feldman MW, Lavi U, Hillel J (2003) Recent duplication of the, common carp (*Cyprinus carpio* L.) genome as revealed by analyses of microsatellite loci. Mol Biol Evol 20(9):1425–1434. doi:10.1093/molbev/msg173

David L, Rothbard S, Rubinstein I, Katzman H, Hulata G, Hillel J, Lavi U (2004) Aspects of red and black color inheritance in the Japanese ornamental (Koi) carp (*Cyprinus carpio* L.). Aquaculture 233(1–4):129–147. doi:10.1016/j.aquaculture.2003.10.033

Davidson WS, Koop BF, Jones SJM, Iturra P, Vidal R, Maass A, Jonassen I, Lien S, Omholt SW (2010) Sequencing the genome of the Atlantic salmon (*Salmo salar*). Genome Biol 11(9):403. doi:10.1186/gb-2010-11-9-403

Davis CA, Holmyard DP, Millen KJ, Joyner AL (1991) Examining pattern formation in mouse, chicken and frog embryos with an en-specific antiserum. Development 111(2):287–298

de Pinna MCC (1996) Teleostean monophyly. In: Stiassny MLJ, Parenti LR, Johnson GD (eds) Interrelationships of fishes. Academic, San Diego, pp 147–162

de Souza FS, Bumaschny VF, Low MJ, Rubinstein M (2005) Subfunctionalization of expression and peptide domains following the ancient duplication of the proopiomelanocortin gene in teleost fishes. Mol Biol Evol 22(12):2417–2427. doi:10.1093/molbev/msi236 msi236 [pii]

Deng C, Cheng CH, Ye H, He X, Chen L (2010) Evolution of an antifreeze protein by neofunctionalization under escape from adaptive conflict. Proc Nat Acad Sci USA 107(50):21593–21598. doi:10.1073/pnas.1007883107 1007883107 [pii]

Doitsidou M, Reichman-Fried M, Stebler J, Koprunner M, Dorries J, Meyer D, Esguerra CV, Leung T, Raz E (2002) Guidance of primordial germ cell migration by the chemokine SDF-1. Cell 111(5):647–659 S0092867402011352[pii]

Donoghue PC, Purnell MA (2005) Genome duplication, extinction and vertebrate evolution. Trends Ecol Evol 20(6):312–319. doi:10.1016/j.tree.2005.04.008 S0169-5347(05)00108-4 [pii]

Douard V, Brunet F, Boussau B, Ahrens-Fath I, Vlaeminck-Guillem V, Haendler B, Laudet V, Guiguen Y (2008) The fate of the duplicated androgen receptor in fishes: a late neofunctionalization event? BMC Evol Biol 8:336. doi:10.1186/1471-2148-8-336 Artn 336

Eiken HG, Njolstad PR, Molven A, Fjose A (1987) A zebrafish homeobox-containing gene with embryonic transcription. Biochem Biophys Res Commun 149(3):1165–1171

Elgar G, Clark MS, Meek S, Smith S, Warner S, Edwards YJ, Bouchireb N, Cottage A, Yeo GS, Umrania Y, Williams G, Brenner S (1999) Generation and analysis of 25 Mb of genomic DNA from the pufferfish *Fugu rubripes* by sequence scanning. Genome Res 9(10):960–971

Ferris SD (1984) Tetraploidy and the evolution of the catostomid fishes. In: Turner BT (ed) Evolutionary genetics of fishes. Plenus Press, New York, pp 54–93

Ferris SD, Whitt GS (1977a) Duplicate gene expression in diploid and tetraploid loaches (cypriniformes, Cobitidae). Biochem Genet 15(11–12):1097–1112

Ferris SD, Whitt GS (1977b) Evolution of duplicate gene-expression in carp (*Cyprinus carpio*). Experientia 33(10):1299–1301

Finn RN, Kristoffersen BA (2007) Vertebrate vitellogenin gene duplication in relation to the "3R hypothesis": correlation to the pelagic egg and the oceanic radiation of teleosts. PLoS One 2(1):e169. doi:10.1371/journal.pone.0000169

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151(4):1531–1545

Froschauer A, Braasch I, Volff JN (2006) Fish genomes, comparative genomics and vertebrate evolution. Curr Genomics 7(1):43–57. doi:10.2174/138920206776389766

Gardner CA, Barald KF (1992) Expression patterns of engrailed-like proteins in the chick embryo. Dev Dyn 193(4):370–388. doi:10.1002/aja.1001930410

Gates MA, Kim L, Egan ES, Cardozo T, Sirotkin HI, Dougan ST, Lashkari D, Abagyan R, Schier AF, Talbot WS (1999) A genetic linkage map for zebrafish: comparative analysis and localization of genes and expressed sequences. Genome Res 9(4):334–347

Gomez A, Volff JN, Hornung U, Schartl M, Wellbrock C (2004) Identification of a second *egfr* gene in *Xiphophorus* uncovers an expansion of the epidermal growth factor receptor family in fish. Mol Biol Evol 21(2):266–275. doi:10.1093/molbev/msh017 msh017[pii]

Gould SJ, Lewontin RC (1979) Spandrels of San-Marco and the Panglossian paradigm—a critique of the adaptationist program. Proc R Soc Lond B Biol Sci 205(1161):581–598

Graham A, Papalopulu N, Krumlauf R (1989) The murine and *Drosophila* homeobox gene complexes have common features of organization and expression. Cell 57(3):367–378. doi:0092-8674(89)90912-4 [pii]

Haussler D, O'Brien SJ, Ryder OA, Barker FK, Clamp M, Crawford AJ, Hanner R, Hanotte O, Johnson WE, McGuire JA, Miller W, Murphy RW, Murphy WJ, Sheldon FH, Sinervo B, Venkatesh B, Wiley EO, Allendorf FW, Amato G, Baker CS, Bauer A, Beja-Pereira A, Bermingham E, Bernardi G, Bonvicino CR, Brenner S, Burke T, Cracraft J, Diekhans M, Edwards S, Ericson PGP, Estes J, Fjelsda J, Flesness N, Gamble T, Gaubert P, Graphodatsky AS, Graves JAM, Green ED, Green RE, Hackett S, Hebert P, Helgen KM, Joseph L, Kessing B, Kingsley DM, Lewin HA, Luikart G, Martelli P, Moreira MAM, Nguyen N, Orti G, Pike BL, Rawson DM, Schuster SC, Seuanez HN, Shaffer HB, Springer MS, Stuart JM, Sumner J, Teeling E, Vrijenhoek RC, Ward RD, Warren WC, Wayne R, Williams TM, Wolfe ND, Zhang YP, Graph-Odatsky A, Johnson WE, Felsenfeld A, Turner S, Scientists GKC, Grp M, Grp B, Grp AR, Grp F, Grp GP, Grp A (2009) Genome 10 K: a proposal to obtain whole-genome sequence for 10,000 vertebrate species. J Hered 100(6):659–674. doi:10.1093/jhered/esp086

He X, Zhang J (2005) Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. Genetics 169(2):1157–1164. doi:10.1534/genetics.104.037051 genetics.104.037051 [pii]

Hoegg S, Meyer A (2005) Hox clusters as models for vertebrate genome evolution. Trends Genet 21(8):421–424. doi:10.1016/j.tig.2005.06.004 S0168-9525(05)00165-4 [pii]

Hoegg S, Brinkmann H, Taylor JS, Meyer A (2004) Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. J Mol Evol 59(2):190–203. doi:10.1007/s00239-004-2613-z

Hoegg S, Boore JL, Kuehl JV, Meyer A (2007) Comparative phylogenomic analyses of teleost fish hox gene clusters: lessons from the cichlid fish *Astatotilapia burtoni*. BMC Genomics 8:317. doi:10.1186/1471-2164-8-317 1471-2164-8-317 [pii]

Howarth DL, Law SHW, Barnes B, Hall JM, Hinton DE, Moore L, Maglich JM, Moore JT, Kullman SW (2008) Paralogous vitamin D receptors in teleosts: transition of nuclear receptor function. Endocrinology 149(5):2411–2422. doi:10.1210/En.2007-1256

Hufton AL, Groth D, Vingron M, Lehrach H, Poustka AJ, Panopoulou G (2008) Early vertebrate whole genome duplications were predated by a period of intense genome rearrangement. Genome Res 18(10):1582–1591. doi:10.1101/gr.080119.108 gr.080119.108 [pii]

Huminiecki L, Heldin CH (2010) 2R and remodeling of vertebrate signal transduction engine. BMC Biol 8:146. doi:10.1186/1741-7007-8-146 1741-7007-8-146 [pii]

Hurley IA, Mueller RL, Dunn KA, Schmidt EJ, Friedman M, Ho RK, Prince VE, Yang Z, Thomas MG, Coates MI (2007) A new time-scale for ray-finned fish evolution. Proc Biol Sci 274(1609):489–498

Inoue JG, Miya M, Tsukamoto K, Nishida M (2003) Basal actinopterygian relationships: a mitogenomic perspective on the phylogeny of the "ancient fish". Mol Phylogenet Evol 26(1):110–120

Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, Nicaud S, Jaffe D, Fisher S, Lutfalla G, Dossat C, Segurens B, Dasilva C, Salanoubat M, Levy M, Boudet N, Castellano S, Anthouard V, Jubin C, Castelli V, Katinka M, Vacherie B, Biemont C, Skalli Z, Cattolico L, Poulain J, De Berardinis V, Cruaud C, Duprat S, Brottier P, Coutanceau JP, Gouzy J, Parra G, Lardier G, Chapple C, McKernan KJ, McEwan P, Bosak S, Kellis M, Volff JN, Guigo R, Zody MC, Mesirov J, Lindblad-Toh K, Birren B, Nusbaum C, Kahn D, Robinson-Rechavi M, Laudet V, Schachter V, Quetier F, Saurin W, Scarpelli C, Wincker P, Lander ES, Weissenbach J, Roest Crollius H (2004) Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. Nature 431(7011):946–957. doi:10.1038/nature03025 nature03025 [pii]

Jovelin R, He X, Amores A, Yan YL, Shi R, Qin B, Roe B, Cresko WA, Postlethwait JH (2007) Duplication and divergence of *fgf8* functions in teleost development and evolution. J Exp Zool B Mol Dev Evol 308(6):730–743. doi:10.1002/jez.b.21193

Joyner AL, Martin GR (1987) *En-1* and *En-2*, two mouse genes with sequence homology to the *Drosophila engrailed* gene: expression during embryogenesis. Genes Dev 1(1):29–38

Karanth S, Lall SP, Denovan-Wright EM, Wright JM (2009) Differential transcriptional modulation of duplicated fatty acid-binding protein genes by dietary fatty acids in zebrafish (*Danio rerio*): evidence for subfunctionalization or neofunctionalization of duplicated genes. BMC Evol Biol 9:219. doi:10.1186/1471-2148-9-219 1471-2148-9-219 [pii]

Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, Nagayasu Y, Doi K, Kasai Y, Jindo T, Kobayashi D, Shimada A, Toyoda A, Kuroki Y, Fujiyama A, Sasaki T, Shimizu A, Asakawa S, Shimizu N, Hashimoto S, Yang J, Lee Y, Matsushima K, Sugano S, Sakaizumi M, Narita T, Ohishi K, Haga S, Ohta F, Nomoto H, Nogata K, Morishita T, Endo T, Shin IT, Takeda H, Morishita S, Kohara Y (2007) The medaka draft genome and insights into vertebrate genome evolution. Nature 447(7145):714–719. doi:10.1038/nature05846 nature05846 [pii]

Kassahn KS, Dang VT, Wilkins SJ, Perkins AC, Ragan MA (2009) Evolution of gene function and regulatory control after whole-genome duplication: comparative analyses in vertebrates. Genome Res 19(8):1404–1418. doi:10.1101/Gr.086827.108

Kikugawa K, Katoh K, Kuraku S, Sakurai H, Ishida O, Iwabe N, Miyata T (2004) Basal jawed vertebrate phylogeny inferred from multiple nuclear DNA-coded genes. BMC Biol 2:3. doi:10.1186/1741-7007-2-3 1741-7007-2-3 [pii]

Kluver N, Kondo M, Herpin A, Mitani H, Schartl M (2005) Divergent expression patterns of *Sox9* duplicates in teleosts indicate a lineage specific subfunctionalization. Dev Genes Evol 215(6):297–305. doi:10.1007/s00427-005-0477-x

Kohn M, Hogel J, Vogel W, Minich P, Kehrer-Sawatzki H, Graves JA, Hameister H (2006) Reconstruction of a 450-My-old ancestral vertebrate protokaryotype. Trends Genet 22(4):203–210. doi:10.1016/j.tig.2006.02.008 S0168-9525(06)00063-1 [pii]

Komiyama T, Kobayashi H, Tateno Y, Inoko H, Gojobori T, Ikeo K (2009) An evolutionary origin and selection process of goldfish. Gene 430(1–2):5–11. doi:10.1016/j.gene.2008.10.019 S0378-1119(08)00543-X [pii]

Koop BF, von Schalburg KR, Leong J, Walker N, Lieph R, Cooper GA, Robb A, Beetz-Sargent M, Holt RA, Moore R, Brahmbhatt S, Rosner J, Rexroad CE, McGowan CR, Davidson WS (2008) A salmonid EST genomic study: genes, duplications, phylogeny and microarrays. BMC Genomics 9:545. doi:10.1186/1471-2164-9-545

Kupka E (1948) Chromosomale Verschiedenheiten bei schweizerischen Coregonen (Felchen). Rev Suisse Zool 55:293–295

Larhammar D, Risinger C (1994) Molecular genetic aspects of tetraploidy in the common carp *Cyprinus carpio*. Mol Phylogenet Evol 3(1):59–68. doi:10.1006/mpev.1994.1007 S1055-7903(84)71007-4 [pii]

Le Comber SC, Smith C (2004) Polyploidy in fishes: patterns and processes. Biol J Linn Soc 82(4):431–442

Leggatt RA, Iwama GK (2003) Occurrence of polyploidy in the fishes. Rev Fish Biol Fisher 13(3):237–246

Leong JS, Jantzen SG, von Schalburg KR, Cooper GA, Messmer AM, Liao NY, Munro S, Moore R, Holt RA, Jones SJM, Davidson WS, Koop BF (2010) *Salmo salar* and *Esox lucius* full-length cDNA sequences reveal changes in evolutionary pressures on a post-tetraploidization genome. BMC Genomics 11:279. doi:10.1186/1471-2164-11-279

Li YJ, Yu Z, Zhang MZ, Qian C, Abe S, Arai K (2011) The origin of natural tetraploid loach *Misgurnus anguillicaudatus* (Teleostei: Cobitidae) inferred from meiotic chromosome configurations. Genetica 139(6):805–811. doi:10.1007/s10709-011-9585-x

Lien S, Gidskehaug L, Moen T, Hayes BJ, Berg PR, Davidson WS, Omholt SW, Kent MP (2011) A dense SNP-based linkage map for Atlantic salmon (*Salmo salar*) reveals extended chromosome homeologies and striking differences in sex-specific recombination patterns. BMC Genomics 12(1):615. doi:10.1186/1471-2164-12-615 1471-2164-12-615 [pii]

Lister JA, Close J, Raible DW (2001) Duplicate *mitf* genes in zebrafish: complementary expression and conservation of melanogenic potential. Dev Biol 237(2):333–344

Ludwig A, Belfiore NM, Pitra C, Svirsky V, Jenneckens I (2001) Genome duplication events and functional reduction of ploidy levels in sturgeon (*Acipenser*, *Huso* and *Scaphirhynchus*). Genetics 158(3):1203–1215

Luo J, Stadler PF, He S, Meyer A (2007) PCR survey of *Hox* genes in the goldfish *Carassius auratus auratus*. J Exp Zool B Mol Dev Evol 308(3):250–258. doi:10.1002/jez.b.21144

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290(5494):1151–1155 8976 [pii]

Lynch M, Force AG (2000) The origin of interspecific genomic incompatibility via gene duplication. Am Nat 156(6):590–605

Mable BK (2004) 'Why polyploidy is rarer in animals than in plants': myths and mechanisms. Biol J Linn Soc 82(4):453–466

Mable BK, Alexandrou MA, Taylor MI (2011) Genome duplication in amphibians and fish: an extended synthesis. J Zool 284(3):151–182. doi:10.1111/j.1469-7998.2011.00829.x

Mank JE, Avise JC (2006a) Cladogenetic correlates of genomic expansions in the recent evolution of actinopterygiian fishes. Proc R Soc Lond B Biol Sci 273(1582):33–38. doi:10.1098/rspb.2005.3295

Mank JE, Avise JC (2006b) Phylogenetic conservation of chromosome numbers in actinopterygiian fishes. Genetica 127(1–3):321–327. doi:10.1007/s10709-005-5248-0

Mayrose I, Zhan SH, Rothfels CJ, Magnuson-Ford K, Barker MS, Rieseberg LH, Otto SP (2011) Recently formed polyploid plants diversify at lower rates. Science 333(6047):1257. doi:10.1126/Science.1207205

Meyer A (1998) Hox gene variation and evolution. Nature 391 (6664):225, 227–228. doi:10.1038/34530

Meyer A, Schartl M (1999) Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. Curr Opin Cell Biol 11(6):699–704. doi:S0955-0674(99)00039-3 [pii]

Meyer A, Van de Peer Y (2005) From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). BioEssays 27(9):937–945. doi:10.1002/bies.20293

Miller MR, Brunelli JP, Wheeler PA, Liu S, Rexroad CE 3rd, Palti Y, Doe CQ, Thorgaard GH (2011) A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. Mol Ecol. doi:10.1111/j.1365-294X.2011.05305.x

Misof BY, Wagner GP (1996) Evidence for four hox clusters in the killifish *Fundulus heteroclitus* (teleostei). Mol Phylogenet Evol 5(2):309–322. doi:10.1006/mpev.1996.0026 S1055-7903(96)90026-3 [pii]

Moghadam HK, Ferguson MM, Danzmann RG (2005a) Evidence for *Hox* gene duplication in rainbow trout (*Oncorhynchus mykiss*): a tetraploid model species. J Mol Evol 61(6):804–818. doi:10.1007/s00239-004-0230-5

Moghadam HK, Ferguson MM, Danzmann RG (2005b) Evolution of *Hox* clusters in Salmonidae: a comparative analysis between Atlantic salmon (*Salmo salar*) and rainbow trout (*Oncorhynchus mykiss*). J Mol Evol 61(5):636–649. doi:10.1007/s00239-004-0338-7

Moghadam HK, Ferguson MM, Danzmann RG (2009) Comparative genomics and evolution of conserved noncoding elements (CNE) in rainbow trout. BMC Genomics 10:278. doi:10.1186/1471-2164-10-278 Artn 278

Morizot DC (1990) Use of fish gene maps to predict ancestral vertebrate genome organization. In: Ogita Z-I, Markert CL (eds) Isozymes: structure, function, and use in biology, and medicine. Wiley-Liss, Inc., New York, pp 207–234

Muller HJ (1925) Why polyploidy is rarer in animals than in plants. Am Nat 59(663):346–353

Mulley JF, Chiu CH, Holland PW (2006) Breakup of a homeobox cluster after genome duplication in teleosts. Proc Nat Acad Sci USA 103(27):10369–10372. doi:10.1073/pnas.0600341103 0600341103 [pii]

Mungpakdee S, Seo HC, Angotzi AR, Dong XJ, Akalin A, Chourrout D (2008a) Differential evolution of the 13 Atlantic salmon *Hox* clusters. Mol Biol Evol 25(7):1333–1343. doi:10.1093/molbev/msn097

Mungpakdee S, Seo HC, Chourrout D (2008b) Spatio-temporal expression patterns of anterior *Hox* genes in Atlantic salmon (*Salmo salar*). Gene Expr Patterns 8(7–8):508–514. doi:10.1016/j.gep.2008.06.004

Nair S, Schilling TF (2008) Chemokine signaling controls endodermal migration during zebrafish gastrulation. Science 322(5898):89–92. doi:10.1126/science.1160038 1160038 [pii]

Nakatani Y, Takeda H, Kohara Y, Morishita S (2007) Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. Genome Res 17(9):1254–1265. doi:10.1101/gr.6316407

Naruse K, Tanaka M, Mita K, Shima A, Postlethwait J, Mitani H (2004) A medaka gene map: the trace of ancestral vertebrate proto-chromosomes revealed by comparative gene mapping. Genome Res 14(5):820–828. doi:10.1101/gr.2004004 2004004 [pii]

Nelson JS (2006) Fishes of the world, 4th edn. Wiley, Hoboken

Njolstad PR, Molven A, Hordvik I, Apold J, Fjose A (1988) Primary structure, developmentally regulated expression and potential duplication of the zebrafish homeobox gene ZF-21. Nucleic Acids Res 16(19):9097–9111

Ohno S (1970) Evolution by gene duplication. Springer, Berlin

Ohno S, Muramoto J, Christia L, Atkin NB (1967) Diploid-tetraploid relationship among old-world members of fish family Cyprinidae. Chromosoma 23(1):1

Ohno S, Wolf U, Atkin NB (1968) Evolution from fish to mammals by gene duplication. Hereditas-Genetisk A 59(1):169 &

Okuda Y, Yoda H, Uchikawa M, Furutani-Seiki M, Takeda H, Kondoh H, Kamachi Y (2006) Comparative genomic and expression analysis of group B1 *sox* genes in zebrafish indicates their diversification during vertebrate evolution. Dev Dyn 235(3):811–825. doi:10.1002/dvdy.20678

Oliveira C, Almeidatoledo LF, Mori L, Toledofilho SA (1992) Extensive chromosomal rearrangements and nuclear-DNA content changes in the evolution of the armored catfishes genus *Corydoras* (pisces, siluriformes, callichthyidae). J Fish Biol 40(3):419–431

Otto SP (2007) The evolutionary consequences of polyploidy. Cell 131(3):452–462. doi:10.1016/j.cell.2007.10.022

Otto SP, Whitton J (2000) Polyploid incidence and evolution. Annu Rev Genet 34:401–437. doi:10.1146/annurev.genet.34.1.401 34/1/401 [pii]

Pala I, Coelho MM, Schartl M (2008) Dosage compensation by gene-copy silencing in a triploid hybrid fish. Curr Biol 18(17):1344–1348. doi:10.1016/j.cub.2008.07.096

Pala I, Schartl M, Brito M, Vacas JM, Coelho MM (2010) Gene expression regulation and lineage evolution: the north and south tale of the hybrid polyploid *Squalius alburnoides* complex. Proc R Soc B Biol Sci 277(1699):3519–3525. doi:10.1098/rspb.2010.1071

Pandian TJ, Koteeswaran R (1999) Natural occurrence of monoploids and polyploids in the Indian catfish, *Heteropneustes fossilis*. Curr Sci India 76(8):1134–1137

Phillips R, Rab P (2001) Chromosome evolution in the Salmonidae (pisces): an update. Biol Rev Camb Philos Soc 76(1):1–25

Phillips RB, Keatley KA, Morasch MR, Ventura AB, Lubieniecki KP, Koop BF, Danzmann RG, Davidson WS (2009) Assignment of Atlantic salmon (*Salmo salar*) linkage groups to specific chromosomes: conservation of large syntenic blocks corresponding to whole chromosome arms in rainbow trout (*Oncorhynchus mykiss*). BMC Genet 10:46. doi:10.1186/1471-2156-10-46

Piferrer F, Beaumont A, Falguiere JC, Flajshans M, Haffray P, Colombo L (2009) Polyploid fish and shellfish: production, biology and applications to aquaculture for performance improvement and genetic containment. Aquaculture 293(3–4):125–156. doi:10.1016/j.aquaculture.2009.04.036

Postlethwait JH, Yan YL, Gates MA, Horne S, Amores A, Brownlie A, Donovan A, Egan ES, Force A, Gong Z, Goutel C, Fritz A, Kelsh R, Knapik E, Liao E, Paw B, Ransom D, Singer A, Thomson M, Abduljabbar TS, Yelick P, Beier D, Joly JS, Larhammar D, Rosa F, Westerfield M, Zon LI, Johnson SL, Talbot WS (1998) Vertebrate genome evolution and the zebrafish gene map. Nat Genet 18(4):345–349. doi:10.1038/ng0498-345

Postlethwait JH, Woods IG, Ngo-Hazelett P, Yan YL, Kelly PD, Chu F, Huang H, Hill-Force A, Talbot WS (2000) Zebrafish comparative genomics and the origins of vertebrate chromosomes. Genome Res 10(12):1890–1902

Postlethwait JH, Amores A, Yan YL, Austin CA (2002) Duplication of a portion of human chromosome 20q containing topoisomerase (top1) and snail genes provides evidence on genome expansion and the radiation of teleost fish. In: Shimizu N, Aoki T, Hirono I, Takashima F (eds) Aquatic genomics. Springer, Tokyo

Postlethwait J, Amores A, Cresko W, Singer A, Yan YL (2004) Subfunction partitioning, the teleost radiation and the annotation of the human genome. Trends Genet 20(10):481–490. doi:10.1016/j.tig.2004.08.001 S0168-9525(04)00213-6 [pii]

Prince VE (2002) The *Hox* paradox: more complex(es) than imagined. Dev Biol 249(1):1–15. doi:10.1006/Dbio.2002.0745

Prince VE, Joly L, Ekker M, Ho RK (1998) Zebrafish *Hox* genes: genomic organization and modified colinear expression patterns in the trunk. Development 125(3):407–420

Prohaska SJ, Stadler PF (2004) The duplication of the *Hox* gene clusters in teleost fishes. Theory Biosci 123(1):89–110. doi:10.1016/j.thbio.2004.03.004

Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu JK, Benito-Gutierrez EL, Dubchak I, Garcia-Fernandez J, Gibson-Brown JJ, Grigoriev IV, Horton AC, de Jong PJ, Jurka J, Kapitonov VV, Kohara Y, Kuroki Y, Lindquist E, Lucas S, Osoegawa K, Pennacchio LA, Salamov AA, Satou Y, Sauka-Spengler T, Schmutz J, Shin IT, Toyoda A, Bronner-Fraser M, Fujiyama A, Holland LZ, Holland PW, Satoh N, Rokhsar DS (2008) The amphioxus genome and the evolution of the chordate karyotype. Nature 453(7198):1064–1071. doi:10.1038/nature06967 nature06967 [pii]

Raincrow JD, Dewar K, Stocsits C, Prohaska SJ, Amemiya CT, Stadler PF, Chiu CH (2011) *Hox* clusters of the bichir (Actinopterygii, *Polypterus senegalus*) highlight unique patterns of sequence evolution in gnathostome phylogeny. J Exp Zool B Mol Dev Evol 316(6):451–464. doi:10.1002/jez.b.21420

Rastogi S, Liberles DA (2005) Subfunctionalization of duplicated genes as a transition state to neofunctionalization. BMC Evol Biol 5:28. doi:10.1186/1471-2148-5-28 1471-2148-5-28 [pii]

Ravi V, Venkatesh B (2008) Rapidly evolving fish genomes and teleost diversity. Curr Opin Genet Dev 18(6):544–550. doi:10.1016/j.gde.2008.11.001 S0959-437X(08)00151-2 [pii]

Richardson BE, Lehmann R (2010) Mechanisms guiding primordial germ cell migration: strategies from different organisms. Nat Rev Mol Cell Biol 11(1):37–49. doi:10.1038/nrm2815 nrm2815 [pii]

Risinger C, Larhammar D (1993) Multiple loci for synapse protein SNAP-25 in the tetraploid goldfish. Proc Nat Acad Sci USA 90(22):10598–10602

Robinson-Rechavi M, Marchand O, Escriva H, Bardet PL, Zelus D, Hughes S, Laudet V (2001) Euteleost fish genomes are characterized by expansion of gene families. Genome Res 11(5):781–788. doi:10.1101/gr.165601

Rohner N, Bercsenyi M, Orban L, Kolanczyk ME, Linke D, Brand M, Nusslein-Volhard C, Harris MP (2009) Duplication of *fgfr1* permits Fgf signaling to serve as a target for selection during domestication. Curr Biol 19(19):1642–1647. doi:10.1016/j.cub.2009.07.065 S0960-9822(09)01542-5 [pii]

Santini F, Harmon LJ, Carnevale G, Alfaro ME (2009) Did genome duplication drive the origin of teleosts? A comparative study of diversification in ray-finned fishes. BMC Evol Biol 9:194. doi:10.1186/1471-2148-9-194 1471-2148-9-194 [pii]

Sato Y, Nishida M (2007) Post-duplication charge evolution of phosphoglucose isomerases in teleost fishes through weak selection on many amino acid sites. BMC Evol Biol 7:204. doi:10.1186/1471-2148-7-204 Artn 204

Sato Y, Hashiguchi Y, Nishida M (2009) Temporal pattern of loss/persistence of duplicate genes involved in signal transduction and metabolic pathways after teleost-specific genome duplication. BMC Evol Biol 9:127. doi:10.1186/1471-2148-9-127 1471-2148-9-127 [pii]

Scannell DR, Wolfe KH (2008) A burst of protein sequence evolution and a prolonged period of asymmetric evolution follow gene duplication in yeast. Genome Res 18(1):137–147. doi:10.1101/gr.6341207 gr.6341207 [pii]

Schultz RJ (1980) Role of polyploidy in the evolution of fishes. In: Lewis WH (ed) Polyploidy—biological relevance. Basic life science, vol 13. Plenum Press, New York, pp 313–340

Semon M, Wolfe KH (2007a) Rearrangement rate following the whole-genome duplication in teleosts. Mol Biol Evol 24(3):860–867. doi:10.1093/molbev/msm003 msm003 [pii]

Semon M, Wolfe KH (2007b) Reciprocal gene loss between *Tetraodon* and zebrafish after whole genome duplication in their ancestor. Trends Genet 23(3):108–112. doi:10.1016/j.tig.2007.01.003 S0168-9525(07)00021-2 [pii]

Setiamarga DH, Miya M, Yamanoue Y, Azuma Y, Inoue JG, Ishiguro NB, Mabuchi K, Nishida M (2009) Divergence time of the two regional medaka populations in Japan as a new time scale for comparative genomics of vertebrates. Biol Lett 5(6):812–816. doi:10.1098/rsbl.2009.0419 rsbl.2009.0419 [pii]

Sha Z, Yu P, Takano T, Liu H, Terhune J, Liu Z (2008) The warm temperature acclimation protein Wap65 as an immune response gene: its duplicates are differentially regulated by temperature and bacterial infections. Mol Immunol 45(5):1458–1469. doi:10.1016/J.Molimm.2007.08.012

Shimada A, Yabusaki M, Niwa H, Yokoi H, Hatta K, Kobayashi D, Takeda H (2008) Maternal-zygotic medaka mutants for *fgfr1* reveal its essential role in the migration of the axial mesoderm but not the lateral mesoderm. Development 135(2):281–290. doi:10.1242/dev.011494 135/2/281 [pii]

Siegel N, Hoegg S, Salzburger W, Braasch I, Meyer A (2007) Comparative genomics of ParaHox clusters of teleost fishes: gene cluster breakup and the retention of gene sets following whole genome duplications. BMC Genomics 8:312. doi:10.1186/1471-2164-8-312 1471-2164-8-312 [pii]

Soltis DE, Albert VA, Leebens-Mack J et al (2009) Polyploidy and angiosperm diversification. Am J Bot 96(1):336–348

Song H, Yan YL, Titus T, He XJ, Postlethwait JH (2011) The role of *stat1b* in zebrafish hematopoiesis. Mech Develop 128(7–10):442–456. doi:10.1016/J.Mod.2011.08.004

Star B, Nederbragt AJ, Jentoft S, Grimholt U, Malmstrom M, Gregers TF, Rounge TB, Paulsen J, Solbakken MH, Sharma A, Wetten OF, Lanzen A, Winer R, Knight J, Vogel JH, Aken B, Andersen O, Lagesen K, Tooming-Klunderud A, Edvardsen RB, Tina KG, Espelund M, Nepal C, Previti C, Karlsen BO, Moum T, Skage M, Berg PR, Gjoen T, Kuhl H, Thorsen J, Malde K, Reinhardt R, Du L, Johansen SD, Searle S, Lien S, Nilsen F, Jonassen I, Omholt

SW, Stenseth NC, Jakobsen KS (2011) The genome sequence of Atlantic cod reveals a unique immune system. Nature 477(7363):207–210. doi:10.1038/nature10342 nature10342 [pii]
Steinke D, Salzburger W, Braasch I, Meyer A (2006) Many genes in fish have species-specific asymmetric rates of molecular evolution. BMC Genomics 7:20. doi:10.1186/1471-2164-7-20 1471-2164-7-20 [pii]
Stellwag EJ (1999) Hox gene duplication in fish. Semin Cell Dev Biol 10(5):531–540. doi:10.1006/scdb.1999.0334 S1084-9521(99)90334-8 [pii]
Stoltzfus A (1999) On the possibility of constructive neutral evolution. J Mol Evol 49(2):169–181. doi:JME1907 [pii]
Svärdson G (1945) Chromosome studies on Salmonidae. Rep Swed State Inst Fresh Fish Res 23:1–151
Taylor JS, Van de Peer Y, Braasch I, Meyer A (2001) Comparative genomics provides evidence for an ancient genome duplication event in fish. Philos Trans R Soc Lond B Biol Sci 356(1414):1661–1679. doi:10.1098/rstb.2001.0975
Taylor JS, Braasch I, Frickey T, Meyer A, Van de Peer Y (2003) Genome duplication, a trait shared by 22,000 species of ray-finned fish. Genome Res 13(3):382–390. doi:10.1101/gr.640303
Tumpel S, Cambronero F, Wiedemann LM, Krumlauf R (2006) Evolution of cis elements in the differential expression of two *Hoxa2* coparalogous genes in pufferfish (*Takifugu rubripes*). Proc Nat Acad Sci USA 103(14):5419–5424. doi:10.1073/pnas.0600993103 0600993103 [pii]
Udono T, Yasumoto K, Takeda K, Amae S, Watanabe K, Saito H, Fuse N, Tachibana M, Takahashi K, Tamai M, Shibahara S (2000) Structural organization of the human microphthalmia-associated transcription factor gene containing four alternative promoters. Biochim Biophys Acta 1491(1–3):205–219
Uyeno T, Smith GR (1972) Tetraploid origin of karyotype of catostomid fishes. Science 175(4022):644
Van de Peer Y, Taylor JS, Braasch I, Meyer A (2001) The ghost of selection past: rates of evolution and functional divergence of anciently duplicated genes. J Mol Evol 53(4–5):436–446. doi:10.1007/s002390010233
Van de Peer Y, Maere S, Meyer A (2009) The evolutionary significance of ancient genome duplications. Nat Rev Genet 10(10):725–732. doi:10.1038/nrg2600 nrg2600 [pii]
Vandepoele K, De Vos W, Taylor JS, Meyer A, Van de Peer Y (2004) Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates. Proc Nat Acad Sci USA 101(6):1638–1643. doi:10.1073/pnas.0307968100 0307968100[pii]
Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R, Birney E (2009) EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. Genome Res 19(2):327–335. doi:10.1101/Gr.073585.107
Volff JN (2005) Genome evolution and biodiversity in teleost fish. Heredity (Edinb) 94(3):280–294. doi:10.1038/sj.hdy.6800635 6800635 [pii]
Volff JN, Brunet F, Böhne A, Galiana-Arnoux D (2011) Evolution of fish genomes. In: Farrell AP, Cech JJ, Richards JG, Stevens ED (eds) Encyclopedia of fish physiology: from genome to environment. Elsevier Inc., San Diego
Werth CR, Windham MD (1991) A model for divergent, allopatric speciation of polyploid pteridophytes resulting from silencing of duplicate-gene expression. Am Nat 137(4):515–526
Winkler C, Schafer M, Duschl J, Schartl M, Volff JN (2003) Functional divergence of two zebrafish midkine growth factors following fish-specific gene duplication. Genome Res 13(6A):1067–1081. doi:10.1101/gr.1097503 GR-10975R [pii]
Wittbrodt J, Meyer A, Schartl M (1998) More genes in fish? BioEssays 20(6):511–515. doi:10.1002/(sici)1521-1878(199806)20:6<511:aid-bies10>3.0.co;2-3
Woltering JM, Durston AJ (2006) The zebrafish *hoxDb* cluster has been reduced a single microRNA. Nat Genet 38(6):601–602. doi:10.1038/Ng0606-601
Woods TD, Buth DG (1984) High-level of gene silencing in the tetraploid goldfish. Biochem Syst Ecol 12(4):415–421

Woods IG, Kelly PD, Chu F, Ngo-Hazelett P, Yan YL, Huang H, Postlethwait JH, Talbot WS (2000) A comparative map of the zebrafish genome. Genome Res 10(12):1903–1914

Woods IG, Wilson C, Friedlander B, Chang P, Reyes DK, Nix R, Kelly PD, Chu F, Postlethwait JH, Talbot WS (2005) The zebrafish gene map defines ancestral vertebrate chromosomes. Genome Res 15(9):1307–1314. doi:10.1101/gr.4134305 gr.4134305 [pii]

Woolfe A, Elgar G (2007) Comparative genomics using fugu reveals insights into regulatory subfunctionalization. Genome Biol 8(4):R53. doi:10.1186/gb-2007-8-4-r53 gb-2007-8-4-r53 [pii]

Yao K, Ge W (2010) Kit system in the zebrafish ovary: evidence for functional divergence of two isoforms of kit (kita and kitb) and kit ligand (kitlga and kitlgb) during folliculogenesis. Biol Reprod 82(6):1216–1226. doi:10.1095/biolreprod.109.082644 biolreprod.109.082644 [pii]

Yokoi H, Shimada A, Carl M, Takashima S, Kobayashi D, Narita T, Jindo T, Kimura T, Kitagawa T, Kage T, Sawada A, Naruse K, Asakawa S, Shimizu N, Mitani H, Shima A, Tsutsumi M, Hori H, Wittbrodt J, Saga Y, Ishikawa Y, Araki K, Takeda H (2007) Mutant analyses reveal different functions of *fgfr1* in medaka and zebrafish despite conserved ligand-receptor relationships. Dev Biol 304(1):326–337. doi:10.1016/j.ydbio.2006.12.043 S0012-1606(06)01502-8 [pii]

Yokoi H, Yan YL, Miller MR, BreMiller RA, Catchen JM, Johnson EA, Postlethwait JH (2009) Expression profiling of zebrafish *sox9* mutants reveals that *sox9* is required for retinal differentiation. Dev Biol 329(1):1–15. doi:10.1016/j.ydbio.2009.01.002 S0012-1606(09)00020-7 [pii]

Yu WP, Brenner S, Venkatesh B (2003) Duplication, degeneration and subfunctionalization of the nested synapsin-timp genes in fugu. Trends Genet 19(4):180–183 S0168952503000489[pii]

Yuan J, He Z, Yuan X, Jiang X, Sun X, Zou S (2010) Speciation of polyploid Cyprinidae fish of common carp, crucian carp, and silver crucian carp derived from duplicated Hox genes. J Exp Zool B Mol Dev Evol 314(6):445–456. doi:10.1002/jez.b.21350

Zapater C, Chauvigne F, Norberg B, Finn RN, Cerda J (2011) Dual neofunctionalization of a rapidly evolving aquaporin-1 paralog resulted in constrained and relaxed traits controlling channel function during meiosis resumption in teleosts. Mol Biol Evol 28(11):3151–3169. doi:10.1093/molbev/msr146 msr146 [pii]

Chapter 18
Polyploidization and Sex Chromosome Evolution in Amphibians

Ben J. Evans, R. Alexander Pyron and John J. Wiens

Abstract Genome duplication, including polyploid speciation and spontaneous polyploidy in diploid species, occurs more frequently in amphibians than mammals. One possible explanation is that some amphibians, unlike almost all mammals, have young sex chromosomes that carry a similar suite of genes (apart from the genetic trigger for sex determination). These species potentially can experience genome duplication without disrupting dosage stoichiometry between interacting proteins encoded by genes on the sex chromosomes and autosomal chromosomes. To explore this possibility, we performed a permutation aimed at testing whether amphibian species that experienced polyploid speciation or spontaneous polyploidy have younger sex chromosomes than other amphibians. While the most conservative permutation was not significant, the frog genera *Xenopus* and *Leiopelma* provide anecdotal support for a negative correlation between the age of sex chromosomes and a species' propensity to undergo genome duplication. This study also points to more frequent turnover of sex chromosomes than previously proposed, and suggests a lack of statistical support for male versus female heterogamy in the most recent common ancestors of frogs, salamanders, and amphibians in general. Future advances in genomics undoubtedly will further illuminate the relationship between amphibian sex chromosome degeneration and genome duplication.

B. J. Evans (✉)
Department of Biology, McMaster University, Life Sciences Building Room 328,
1280 Main Street West, Hamilton, ON L8S 4K1, Canada
e-mail: evansb@mcmaster.ca

R. Alexander Pyron
Department of Biological Sciences, The George Washington University,
2023 G St. NW, Washington, DC 20052, USA

J. J. Wiens
Department of Ecology and Evolution, Stony Brook University,
Stony Brook, NY 11794-5245, USA

P. S. Soltis and D. E. Soltis (eds.), *Polyploidy and Genome Evolution*,
DOI: 10.1007/978-3-642-31442-1_18,

18.1 Introduction

Why polyploidization is more common in plants than in animals is a central question in biology (Mable 2004; Muller 1925; Orr 1990), and multiple explanations have been put forward (reviewed in Gregory and Mable 2005; Mable 2004; Orr 1990; Otto and Whitton 2000). One possibility is that the propensity for a species to undergo polyploidization is related to the extent of sex chromosome degeneration. Sex chromosome degeneration is the evolution of differences in gene content that goes beyond the fundamental difference in the presence or absence of a genetic trigger for sex determination. Degenerate sex chromosomes could present problems during "diploidization" of a polyploid genome. Diploidization refers to the phenomenon by which a polyploid species transitions to a mode of chromosomal inheritance that is similar to a diploid species. A key feature of diploidization is the switch from polysomic inheritance, where multivalents form during cell division, to disomic inheritance, where only bivalents form (Wolfe 2001). This phenomenon is probably achieved via divergence between duplicated pairs of homologous chromosomes. Diploidization therefore could be instantaneous when polyploidization occurs via allopolyploidization (genome duplication associated with hybridization among diverged species) because duplicated homologous chromosome pairs already diverged from one another in the ancestral species. When a polyploid genome with duplicated sex chromosomes becomes diploidized, one pair of sex chromosomes presumably begins to segregate autosomally. With a degenerate Y-chromosome, for instance, the nascent autosomal pair that was previously a pair of sex chromosomes would initially have three possible genotypes: A_XA_X, A_X0, and 00 where A_X refers to an autosomal allele derived from an ancestral X-chromosome and 0 refers to a missing allele that was lost on the ancestral Y-chromosome. If the 00 genotype is deleterious or lethal, there would be reproductive incompatibilities in the early stages of diploidization until the degenerate chromosome is lost. This fitness cost could be mitigated if a functional paralogous allele were still present on the sex chromosomes, as would be expected in an autopolyploid (formed from genome duplication within a species). In allopolyploids, however, sex chromosomes could degenerate in unique ways in each ancestral species, giving rise to diverged gene content, so homozygous null genotypes would be a bigger problem. In both types of polyploids, dosage balance requirements—natural selection favoring a specific relative expression level of interacting genes (Papp et al. 2003; Qian and Zhang 2008)—could impose a fitness cost on a homozygous or heterozygous null autosomal genotype in a polyploid.

Polyploidization could also present challenges to species with degenerate sex chromosomes that have also evolved mechanisms for dosage compensation (Orr 1990). Dosage compensation equilibrates expression levels of genes that have one allele in one sex and two alleles in the other sex (for example, X-linked genes have one allele in XY males but two alleles in XX females). In this way the stoichiometry of expression of X-linked and autosomal genes is constant, or "balanced", in males

and females. Orr (1990) argued that a polyploid lineage would initially be established via backcrossing a new polyploid individual to diploid individuals, and that this would disrupt this balance.

The evolution of differences in gene content is a combined consequence of the migration of genes, especially genes with sex-specific function, to one or the other sex chromosome, and also the loss of genes from the region of suppressed recombination on the sex-specific sex chromosome (for example, the Y-chromosome). The disparity in gene content between the sex chromosomes is thought to increase over time as a consequence of natural selection (Bergero and Charlesworth 2009; Charlesworth et al. 2005). Substantial disparity in gene content between the sex chromosomes could be coupled with selective pressure favoring the evolution of mechanisms of dosage compensation. For this reason, the proposal that degenerate sex chromosomes deter polyploidization (including species that lack dosage compensation) is not independent of Orr's (1990) proposal that dosage compensation deters polyploidization. In either case, if sex chromosome degeneration acts as a barrier to polyploidization, this would predict that polyploid species or species with polymorphism in ploidy levels would have relatively "young", minimally degenerate sex chromosomes as compared with other species.

In this chapter we briefly review polyploidization in frogs and salamanders and general features of sex chromosome evolution. Using previously published information, we then use a maximum likelihood approach to analyze the evolution of new sex-determining mechanisms in frogs and salamanders in a phylogenetic context, where new mechanisms are inferred either from a change in which chromosomes are the sex chromosomes, the evolution of a new trigger for sex determination, or from observed polymorphism in sex-determining mechanisms. Following this, we explore whether polyploidization occurs more frequently soon after a new sex-determining mechanism evolves using a permutation test that accommodates uncertainty in ancestral reconstruction. We conclude that novel sex-determining mechanisms have evolved in amphibians even more frequently than previously proposed, and that amphibians with young sex chromosomes may be more likely to experience genome duplication, resulting either in polyploid speciation or in spontaneous polyploidy of individuals of an otherwise diploid species. A major caveat to the latter result is that information on the age of sex-determining mechanisms of most polyploid amphibians is lacking.

18.1.1 Sex Chromosome Evolution

Sex chromosomes originate from autosomes (Ohno 1967) but differ in carrying genetic information that (a) differs between the sexes and (b) triggers or represses sex-specific gonadal differentiation. The "heterogametic" sex produces two types of gametes, each type with a different sex chromosome, and the "homogametic" sex produces only one type of gamete with respect to the sex chromosomes. The sex chromosomes of species with male heterogamy are called "X" and "Y" (females

have two X chromosomes and males have an X and a Y), and the sex chromosomes of species with female heterogamy are called "Z" and "W" (males have two Z chromosomes, and females have a Z and a W). In species with genetic sex determination, gonadal differentiation—also known as primary sex determination—is achieved either using a sex chromosome-specific genetic trigger or by gene dosage, where the homogametic sex carries two doses of an activator of that sex, or a repressor of the heterogametic sex.

The age of sex chromosomes influences important aspects of their evolution and divergence, including divergence in gene content, and the origin of dosage compensation. For example, the sex chromosomes of therian (placental and marsupial) mammals are extremely old (>180 million years; Graves 2008), and the Y chromosome is much smaller than the X and carries fewer genes. This disparity in size and gene content arose due to Y chromosome "degeneration" as a consequence of suppressed recombination with the X chromosome (Charlesworth and Charlesworth 2000). Suppressed recombination ensures that male progeny inherit an intact copy of the genetic trigger for testis formation, which in therians is the *SRY* gene. But this also permits deleterious mutations to accumulate in Y-linked genes (Muller 1964), leading to loss of function and deletion. "Muller's ratchet", the stochastic loss of the least deleterious allele in a population (Felsenstein 1974), leads to a decline in fitness. This decline occurs more quickly in regions of the genome that do not recombine. Hill-Robertson effects, background selection, and hitchhiking of deleterious alleles also contribute to fitness declines of non-recombining portions of the genome (reviewed in Charlesworth and Charlesworth 2000). Degeneration of the therian Y-chromosome occurred in a stepwise fashion as the region of suppressed recombination expanded in large increments (Skaletsky et al. 2003). In therians the disparity in gene content increased over time after the origin of SRY and associated suppression of recombination between the X and Y chromosomes. Most angiosperm plants that have separate sexes (dioecy), in contrast, have comparatively young sex chromosomes that are not substantially differentiated, although exceptions exist (Bergero and Charlesworth 2011; Charlesworth 2002). Polyploid species are prevalent in angiosperms (Otto and Whitton 2000) but absent in therians (Svartman et al. 2005), and these observations thus provide anecdotal support for the contention that the extent of sex chromosome degeneration is negatively correlated with the incidence of polyploid speciation. Amphibians offer an interesting focal group with which to further evaluate this hypothesis because some features of amphibian genome evolution resemble plants more than other animal groups such as therian mammals. In particular, unlike therian mammals, sex chromosomes in many amphibians are relatively young, chromosome degeneration is modest or absent, and polyploidization is fairly common.

Species that determine sex exclusively using environmental triggers do not have genomic differences between the sexes and therefore have no sex chromosomes. In amphibians, sex determination is genetic so all species are expected to have sex chromosomes. In addition, temperature has been reported to influence offspring sex ratios of various species of the salamander genera *Pleurodeles* and

Hynobius and the frog genera *Bufo*, *Rana*, and laboratory-generated polyploids of the genus *Xenopus* (Hayes 1998; Kobel 1996; Schmid and Steinlein 2001). Sex chromosomes are cytologically distinct in some amphibian species (Schmid et al. 2010). Differences in gene content between the sex chromosomes, which is suggested by cytologically distinct sex chromosomes, led to the independent evolution of dosage compensation mechanisms in placental mammals, birds, and other groups such as *Drosophila* and *Caenorhabditis* (Arnold et al. 2008; Straub and Becker 2007). However, in amphibians evidence of dosage compensation has not been found (Hayes 1998; Schmid et al. 1986; Schmid and Steinlein 2001). One possible reason for this is that we do not yet know the identity of any amphibian genes that are hemizygous in the heterogametic sex, so a rigorous test for dosage compensation in amphibians is not yet possible. These genes would be restricted to the portion of the X or Z chromosome that does not recombine with the Y or W chromosome, respectively.

A recent study of European tree frogs identified one way that amphibians circumvent sex chromosome degeneration (Stöck et al. 2011). In three species, no recombination occurred between the sex chromosomes in males generated from intraspecific crosses, yet no intraspecific sex chromosome divergence was observed (Stöck et al. 2011). This suggests that Muller's ratchet is periodically reset in these species by infrequent recombination between the sex chromosomes. Sex chromosome degeneration can also be circumvented by genomic translocation of the sex-determining locus to another chromosomal pair, or by re-assignment of the sex-determining function to a gene located elsewhere in the genome. Both of these phenomena result in a change in which chromosomes are the sex chromosomes (hereafter "sex chromosome turnover"). Ancient examples of sex chromosome turnover are evinced in amniotes by homology between the sex chromosomes of platypuses and those of birds but not those of therian mammals (Graves 2008). In amphibians, sex chromosome turnover is common and is suggested by variation among and within species in male versus female heterogamy (Ezaz et al. 2006; Hillis and Green 1990). Using maximum parsimony, Hillis and Green (1990) analyzed variation in male and female heterogamy in amphibians in a phylogenetic context and concluded that sex chromosome turnover occurred at least seven times.

18.1.2 How Many Frog and Salamander Species are Polyploid?

Comprehensive reviews of polyploidization in amphibians are available in Bogart (1980), Kawamura (1984), Duellman and Trueb (1994), Beçak and Beçak (1998), Otto and Whitton (2000), Gregory and Mable (2005), Schmid et al. (2010), and Mable et al. (2011). The two most recent of these reviews have up-to-date lists of known polyploid species and associated citations that document polyploidy. Schmid et al. (2010) also summarize male and female heterogamy in frogs and salamanders, including information on species with cytologically detectable sex

chromosome divergence (their Table 8, pp 160–161). In their Supplementary Information, Mable et al. (2011) provide data on confirmed diploid species that are closely related to the polyploids. A key difference between these two reviews is that Mable et al. (2011) include only bisexually reproducing polyploids whereas Schmid et al. (2010) also include unisexual polyploids. We have attempted to compile this information as inclusively as possible in Table 18.1, including some minor corrections and a few additional species and associated citations. Thus, not all polyploids listed in this table are bisexual, and some are diploid species in which polyploid individuals occur spontaneously or by induction due to laboratory manipulation.

Fifty polyploid frog species have been described, including seven triploids, 30 tetraploids, 11 octoploids, and two dodecaploids derived from 15 families and 20 genera (Table 18.1). Three tetraploids and two dodecaploids have been reported from the genus *Xenopus* but not yet formally described as species (Evans 2007, 2008; Evans et al. 2004a, 2005a; Tymowska 1991). Stable triploids are known from three frog genera (*Bufo*, *Eupsophus*, and *Rana*), tetraploids from 16 (*Aphantophryne*, *Astylosternus*, *Bufo*, *Chiasmocleis*, *Dicroglossus*, *Eleuthrodactylus*, *Hyla*, *Neobatrachus*, *Odontophrynus*, *Phyllomedusa*, *Pleurodema*, *Pyxicephalus*, *Scaphiophryne Silurana*, *Tomopterna*, and *Xenopus*), octoploids from three (*Ceratophrys*, *Pleurodema*, and *Xenopus*), and dodecaploids only from *Xenopus*. Spontaneous or experimentally induced polyploidy has been reported in at least five frog species. Six polyploid species of salamander, including four triploids and two tetraploids, are known from only two genera (*Ambystoma* and *Siren*) from two families (Table 18.1). Spontaneous or experimentally induced triploidy or tetraploidy has been reported in eight salamander species.

The origin of polyploidy necessarily is preceded by the existence of one or more diploid ancestors. Interestingly, a number of polyploid frog species are inferred to have originated from ancestral diploid species that do not have known extant diploid descendants. In *Xenopus* and *Silurana*, for example, three currently unknown diploid species contributed their genomes to extant tetraploid species (reviewed in Evans 2008). There are also three currently unknown tetraploid species that contributed their genomes to extant octoploid and dodecaploid *Xenopus* species (reviewed in Evans 2008). The tetraploid *Hyla versicolor* is thought to be derived from multiple independent allopolyploidization events between three diploid species, two of which are currently unknown, and probably extinct given that the region in which they occur (temperate North America) is well studied (Holloway et al. 2006). In *Ceratophrys*, there are no known tetraploid species even though three species in this genus are octoploid (Table 18.1). Similarly, the tetraploid species *Bufo pewzowi* is thought to be derived from the diploid *B. turanensis* and another unidentified diploid (Stöck et al. 2009), and various tetraploid species of *Neobatrachus* are derived from diploid ancestors whose diploid descendants are currently unknown (Mable and Roberts 1997). It is tempting to speculate from these observations that polyploidization contributed to the long-term survival of these lineages, given that the diploid ancestors of extant polyploids seem to have gone extinct in many cases. However, we lack information on how frequently

Table 18.1 A list of known polyploid amphibians compiled primarily from Schmid et al. (2010) and Mabel et al. (2011). Additional citations are provided for examples not included in these references and for unnamed species

Family and species (Frogs)	Ploidy	Family and species (Frogs continued)	Ploidy
Alsodidae		Microhylidae	
Eupsophus vertebralis	Triploid	*Scaphiphryne gottlebei*	Tetraploid
		Aphantophryne (Cophixalus) pansa	Tetraploid
Arthroleptidae		*Chiasmocleis leucosticta*	Tetraploid
Astylosternus diadematus	Tetraploid	Odontophrynidae	
Bufonidae[1]		*Odontophrynus americanus*	Tetraploid
Bufo poweri	Triploid	Pipidae	
Bufo baturae	Triploid	*Silurana epitropicalis*	Tetraploid
Bufo pseudoraddei[1]	Triploid	"*Silurana* new tetraploid 1" ("*Silurana* sp. Nov VII", "*Silurana paratropicalis*")[2,3]	Tetraploid
Bufo zugmayeri[1]	Triploid	*"Silurana* new tetraploid 2"[2]	Tetraploid
Bufo viridis	Triploid	*Xenopus borealis*	Tetraploid
Bufo kerinyagae	Tetraploid	*Xenopus clivii*	Tetraploid
Bufo asmaerae/ Amietophrynus asmaerae	Tetraploid	*Xenopus fraseri*	Tetraploid
Bufo oblongus and subspecies (synonym of "*B. danatensis*")	Tetraploid	*Xenopus gilli*	Tetraploid
Bufo pewzowi and subspecies (also a synonym of *"B. danatensis"*)	Tetraploid	*Xenopus laevis*[4]	Tetraploid
		Xenopus muelleri	Tetraploid
Ceratophryidae		*Xenopus pygmaeus*	Tetraploid
Ceratophrys dorsata/ Ceratophrys aurita	Octoploid	*Xenopus largeni* ("*Xenopus* sp. Nov. III")[5]	Tetraploid
Ceratophrys ornata	Octoploid	"*Xenopus* new tetraploid 1" *("Xenopus* sp. Nov. VI"," *Xenopus* muelleri west")[5,6]	Tetraploid
Ceratophrys joazeirensis	Octoploid	*Xenopus amieti*	Octoploid
Craugastoridae		*Xenopus andrei*	Octoploid
Eleutherodactylus (Haddadus) binotatus	Tetraploid	*Xenopus boumbaensis*	Octoploid

(continued)

Table 18.1 (continued)

Family and species (Frogs)	Ploidy	Family and species (Frogs continued)	Ploidy
		Xenopus itombwensis[7]	Octoploid
Dicroglossidae		*Xenopus lenduensis*[8]	Octoploid
Dicroglossus (Hoplobotrachus) occipitalis	Tetraploid	*Xenopus vestitus*	Octoploid
		Xenopus wittei	Octoploid
Hylidae		*"Xenopus* sp. nov. X"[5,9]	Octoploid
Hyla versicolor	Tetraploid	*Xenopus longipes*	Dodecaploid
Phyllomedusa tetraploidea	Tetraploid	*Xenopus ruwenzoriensis*	Dodecaploid
		"X. cf. boumbaensis"[10]	Dodecaploid
Leptodactylidae		*"Xenopus* sp. Nov. VIII"[5,9]	Dodecaploid
Pleurodema bibroni	Tetraploid		
Pleurodema cordobae	Octoploid	Pyxicephalidae	
Pleurodema kriegi	Tetraploid	*Pyxicephalus (Tomopterna) delalandii*	Tetraploid
		Tomopterna tandyi	Tetraploid
Leiopelmatidae		Ranidae	
*Leiopelma hochstetteri***	Triploid	*Rana esculenta*	Triploid
		*Rana japonica**[11]	Triploid
Myobatrachidae		*Rana nigromaculata***	Triploid/Tetraploid
Neobatrachus aquilonius	Tetraploid	*Rana pipiens**[12]	Triploid
Neobatrachus centralis	Tetraploid	*Rana rugosa**[13]	Triploid
Neobatrachus kunapalari	Tetraploid		
Neobatrachus sudelli	Tetraploid		

Family and species (Salamanders)	Ploidy
Ambystomatidae	
Ambystoma jeffersonianum	Triploid
*Ambystoma mexicanum***	Triploid/Tetraploid
Ambystoma nothagenes	Triploid
Ambystoma platineum	Triploid
Ambystoma tremblayi	Triploid
Plethodontidae	
*Eurycea bislineata***	Triploid/Tetraploid
Salamandridae	
*Notophthalmus viridescens***[15]	Triploid
*Pleurodeles waltl**	Triploid
*lchthyosaura alpestris**	Triploid
*Cynops pyrrhogaster***[14]	Triploid
*Lissotriton vulgaris***[16]	Triploid
Sirenidae[17]	
Siren intermedia	Tetraploid
Siren lacertina	Tetraploid

(continued)

Table 18.1 (continued)

Family and species (Salamanders)	Ploidy

* Experimentally induced

** Spontaneously observed, in some cases also experimentally induced

[1] Following taxonomy of Stöck M, Ustinova J, Lamatsch DK, *et al.* (2009) A vertebrate reproductive system involving three ploidy levels: Hybrid origin of triploids in a contact zone of diploid and tetraploid paleartic green toads (*Bufo viridis* subgroup). *Evolution* **64**, 944–959. Further work is needed to confirm stable triploid ploidy of *B. pseudoraddei* and *B. zugmayeri* (Stöck et al. 2009)

[2] Evans BJ, Kelley DB, Tinsley RC, Melnick DJ, Cannatella DC (2004) A mitochondrial DNA phylogeny of clawed frogs: phylogeography on sub-Saharan Africa and implications for polyploid evolution. Molecular Phylogenetics and Evolution 33, 197–213

[3] "*Silurana paratropicalis*" is a nomen nudem; see Blackburn DC (2011)

[4] Here we consider as *X. laevis* all diverged populations within this clade as identified by Evans et al. (2004). This includes *Xenopus sp. Nov. IX (a.k.a.* X. congo 3) from Tymowska (1991), *X. petersi*, and *X. victorianus*

[5] Tymowska J (1991) Polyploidy and cytogenetic variation in frogs of the genus Xenopus. In: Amphibian cytogenetics and evolution (eds. Green DS, Sessions SK), pp. 259–297. Academic Press., San Diego

[6] Kobel HR, Loumont C, Tinsley RC (1996) The extant species. In: The Biology of *Xenopus* (eds. Tinsley RC, Kobel HR), pp. 9–33. Clarendon Press, Oxford

[7] Table 13 of Schmid et al. (2010) incorrectly lists *Xenopus itombwensis* as a dodecaploid

[8] Evans BJ, Greenbaum E, Kusamba C, et al. (2011) Description of a new octoploid frog species (Anura: Pipidae: Xenopus) from the Democratic Republic of the Congo, with a discussion of the biogeography of African clawed frogs in the Albertine Rift. Journal of Zoology, London 283, 276–290

[9] Species status requires further investigation; for example, *Xenopus* sp Nov. VIII may be the same as *Xenopus cf. boumbaensis*

[10] Evans BJ (2007) Ancestry influences the fate of duplicated genes millions of years after duplication in allopolyploid clawed frogs (Xenopus). Genetics 176, 1119–1130

[11] Kawamura T, Tokunaga C (1952) The sex of triploid frogs, *Rana japonica* Günther. Journal of Science of the Hiroshima University, Series B, Division 1 (Zoology) 13

[12] Briggs R (1947) The experimental production and development of triploid frog embryos. Journal of Experimental Zoology 106, 237–266

[13] Kashiwagi, K. (1993) Production of triploids and their reproductive capacity in *Rana rugosa*. Sci. Rep. Lab. Amphibian Biol. Hiroshima Univ. 12: 23–36

[14] Fankhauser, G, Crotta, R., Perrot, M. (1942) Spontaneous and cold-induced triploidy in the japanese newt, *Triturus pyrrhogaster*. Journal of Experimental Zoology 89(1) 167-181

[15] Fankhauser, G. (1941) The frequency of polyploidy and other spontaneous aberrations of chromosome number among larvae of the newt, *Triturus viridescens*. PNAS 27(11): 507-512. Fankhauser, G. and Watson, R. C. (1942) Heat-indiced triploidy in the Newt, *Triturus viridescens*. Proceedings of the National Academy of Sciences 28(10): 436–440

[16] Litvinchuk, S. N., Rosanov, J. M., Borkin, L. J. 1998. A case of natural triploidy in a smooth newt *Triturus vulgaris* (Linneaus, 1958), from Russia (Caudata: Salamandridae). Herpetozoa 11: 93-95

[17] Morescalchi A, Olmo E (1974) Sirenids: a family of polyploid Urodeles? Experientia 30, 491–492 found *Pseudobranchus striatus* to be polyploid but this result was not supported by the analysis of Moler PE, Kezer J (1993) Karyology and systematics of the salamander genus Pseudobranchus (Sirenidae). Copeia 1993, 39–47

polyploidization occurs and how frequently diploids outcompete polyploids, so it is difficult to test this. It is also plausible, for example, that variation in ploidy level is a neutral phenomenon influenced by stochastic survival and extinction of polyploids and diploids, or by variation among lineages, including polyploids (Mayrose et al. 2011), in their ability to speciate by polyploidization.

18.1.3 Examples of polyploidy in species with demonstrably young sex chromosomes

About one third of the described polyploid frog species belong to the genus *Xenopus*. At least six independent instances of genome duplication gave rise to the ploidy levels seen among extant species in this group, including multiple episodes that generated the highest ploidy level of any vertebrate—dodecaploidy (reviewed in Evans 2008). Tetraploid *Xenopus* evolved at least once, octoploid *Xenopus* evolved independently at least three times, and dodecaploid *Xenopus* evolved independently at least two times (and possibly more depending on the species status of *Xenopus* cf. *boumbaensis* and of *Xenopus* sp. nov. VIII; Table 1, Evans 2007, 2008; Evans et al. 2008a, 2011, 2005a; Tymowska 1991). Tetraploidy also occurred independently in *Silurana* (Evans 2007; Evans et al. 2005a).

With respect to genome duplication, something is clearly special about *Xenopus*—but what? One possible clue emerges from the recent discovery of the first known genetic trigger of sex determination in amphibians by Yoshimoto et al. (2008). These researchers identified a female-specific gene called *DMW* in the tetraploid species *Xenopus laevis*. *DMW* is a W-chromosome linked gene that evolved via gene duplication from another important regulator of sexual differentiation called *DMRT1* (Yoshimoto et al. 2008). *DMW* may function by blocking *DMRT1* induction of testis differentiation (Yoshimoto et al. 2010, 2008). Potentially relevant to the high incidence of polyploidization in *Xenopus* is the discovery that *DMW* originated extremely recently in amphibian evolution—after divergence of *Silurana* and *Xenopus*, but before diversification of most or all extant species of *Xenopus* (Bewick et al. 2011). Not surprisingly, the sex chromosomes of *Xenopus* are not cytologically distinct (Tymowska 1991). Gene contents of the W and Z chromosomes of *Xenopus* are therefore probably very similar, and *Xenopus* species presumably lack mechanisms of dosage compensation operating over most sex-linked genes because both sexes have two alleles at most loci on the sex chromosomes. The preponderance of polyploids in *Xenopus* is therefore consistent with the proposal that polyploidization is more likely to occur in lineages with young, minimally degenerate sex chromosomes.

Another possible link between sex chromosome evolution and polyploidization is provided by *Leiopelma hochstetteri*. This species has intraspecific variation in the presence of a recently evolved univalent W chromosome that governs sex determination in females (Green 1988). *Leiopelma hochstetteri* is diploid but also has spontaneous triploidy (that is, polyploidy without speciation;

Green et al. 1984). It is not clear whether novel mechanisms for sex determination are more likely to evolve and persist in species that have nondegenerate sex chromosomes, but this seems plausible under the same reasoning discussed above with respect to the propensity for lineages to experience polyploidization. More specifically, if a new pair of sex chromosomes appears in a population then the old ones would segregate as a newly established autosomal pair. For this reason, ancestral sex chromosomes with similar gene content would lack or have few null alleles when they segregate autosomally. While the observation of spontaneous triploidy suggests a tolerance of polyploidy, *L. hochstetteri* is not a polyploid species, so a direct link between the age of the sex chromosomes and polyploid speciation (as opposed to the toleration of polyploidy) is not established by this species.

18.2 Evolution of Sex Determination Systems in Amphibians

18.2.1 Methods

Changes in the heterogametic sex, evolution of new triggers for sex determination, and polymorphism in sex chromosomes mark the origin of novel features in genetic pathways for sex determination. In order to quantify how many times this has happened in amphibians, we began with the large amphibian phylogeny reported by Pyron and Wiens (2011). We trimmed from this tree all species except those for which we had information on either heterogamy or polyploidy, or both, and retained the original maximum likelihood branch lengths among the retained species. For illustrative purposes, we also retained diploid species (confirmed or presumed) from the phylogeny of Pyron and Wiens (2011) that are closely related to polyploid species. In many cases diploidy has been confirmed in these species or other closely related species (see Supplementary Information of Mable et al. 2011). We had heterogamy information for *Physalaemus (Engystomops) petersi*, but this species was not included in the phylogeny. Therefore, we used a closely related species (*Physalaemus cuvieri*) that is included in the phylogeny to represent *Physalaemus petersi*. *Physalaemus petersi* was the only species from this genus that was analyzed, so this substitution should be uncontroversial (note that placing some *Physalaemus* in *Engystomops* makes no difference as *Engystomops* and *Phsalaemus* are sister taxa). To better illustrate the phylogenetic distribution of polyploid species in Fig. 18.1, we also substituted *Chiasmocleis hudsoni*, which was present in the phylogeny of Pyron and Wiens (2011), with the tetraploid species *C. leucosticta*, which was not present in the phylogeny of Pyron and Wiens (2011). However, *C. leucosticta* was not included in any of the analyses described below because we lack data on heterogamy for this species.

A total of 143 species (97 frogs, 45 salamanders, and one caecilian as an outgroup) were included. We then converted this tree to a chronogram (a time-

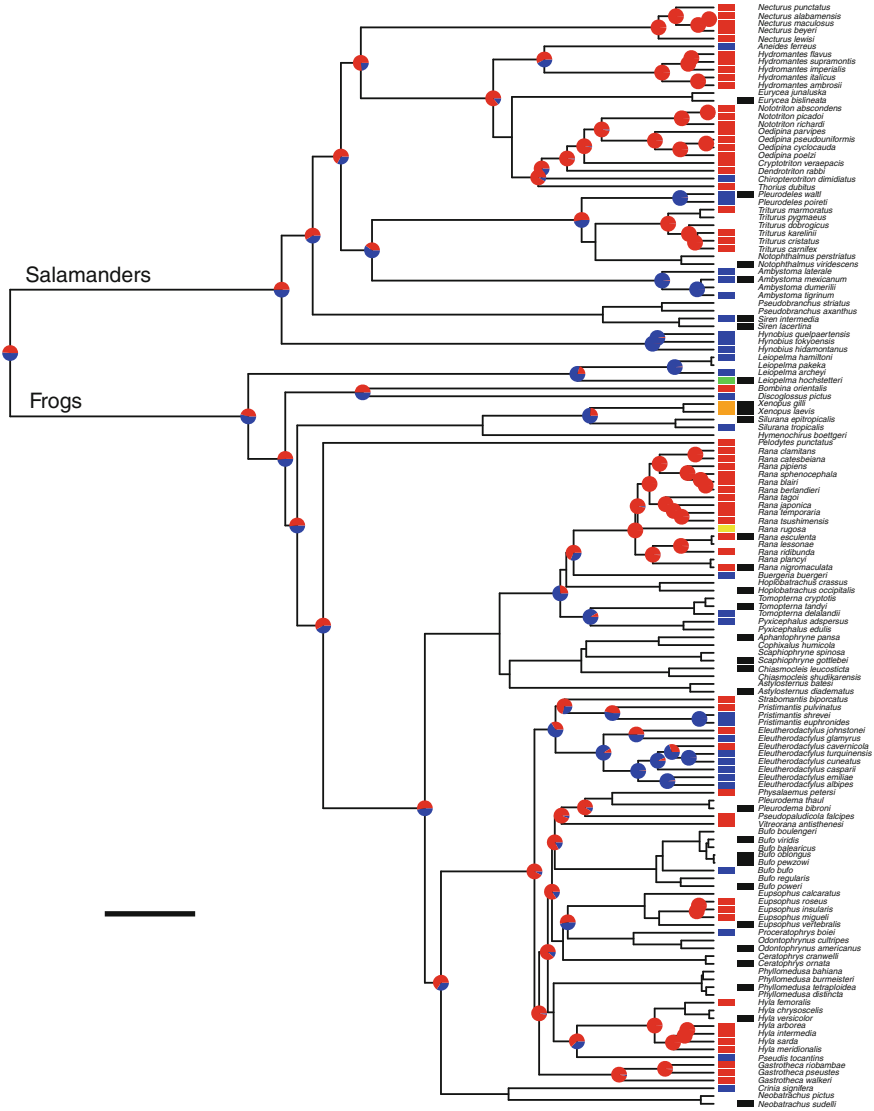

Fig. 18.1 Heterogamy and polyploidy in salamanders and frogs. The phylogeny is from Pyron and Wiens (2011) with branch lengths proportional to time based on a relaxed molecular clock and calibration points described in the text. A *black* scale bar indicates 40 million years of evolution. *Red* and *blue rectangles* on tips indicate male or female heterogamy respectively; *green*, *orange*, and *yellow rectangles* indicate de novo sex determining systems; missing data on heterogamy have no *rectangles* in this column. In the *right* column, *black rectangles* indicate polyploid species; other species are either confirmed or assumed diploid. Some polyploid species listed in Table 18.1 are not included in this figure due to a lack of phylogenetic information. Pie charts on nodes indicate ancestral reconstructions of heterogametic state

calibrated phylogeny) using the penalized likelihood approach (Sanderson 2002), implemented in r8s version 1.71 (Sanderson 2003). We used the calibration points detailed in Wiens (2011) but some had to be excluded given the more limited taxon sampling used here, and some were added or modified given the differences in taxon sampling (e.g., we added two calibration points within the genus *Hyla*, given our more extensive sampling of species in that genus relative to Wiens (2011). We used the following 17 calibration points. The first 16 were treated as constraints on the minimum age of each clade, and the final calibration point was a fixed age for the root of the tree.

(1) Most recent common ancestor (MRCA) of extant salamanders, at least 150.8 Mya (Millions of years ago), based on the fossil *Iridotriton hechti* of the Kimmeridgian/Early Tithonian (Late Jurassic), which is considered to be a crown-group caudate (Evans et al. 2005b).
(2) MRCA of Salamandroidea (all salamanders exclusive of cryptobranchids, hynobiids, and sirenids), at least 125.0 Mya (early Barremian, Cretaceous), based on *Galverpeton* and *Valdotriton* (Evans and Milner 1996).
(3) MRCA of plethodontids and proteids, at least 65.5 Mya. The oldest known amphiumid fossil (*Proamphiuma cretacea*) is late Maastrichtian or early Paleocene, and thus from approximately 65.5 Mya (Gardner 2003). The split between Plethodontidae and Amphiumidae must be at least this old. We do not have amphiumids included here, but the clade of plethodontids and proteids must be at least this old given the well-supported clade consisting of proteids, rhyacotritonids, amphiumids, and plethodontids; see Pyron, Wiens (2011) and earlier studies.
(4) MRCA of *Aneides* and *Hydromantes*, at least 19 Mya. Given the presence of an *Aneides* vertebra in the Arikareean period (Tihen and Wake 1981), the MRCA of the clade containing modern *Aneides* must be at least 19 Myo (Millions of years old).
(5) MRCA of *Triturus* and *Notophthalmus* at least 33.9 Mya, based on fossils of *Triturus* from the Eocene of Europe (33.9–55.8 Mya; Milner 2000)
(6) MRCA of Ambystomatidae and Salamandridae, at least 56.8 Mya, based on a fossil dicamptodontid (Paleocene; Tiffanian; 60.2–56.8 Mya; Naylor and Fox 1993), and given that the ambystomatidae is the sister group to the Dicamptodontidae (so the sister group to Ambystomatidae + Dicamptodontidae must be at least this old).
(7) MRCA of frogs and salamanders, at least 245 Mya, based on a fossil anuran (*Triadobatrachus*) from the Early Triassic (251–245 Mya) of Madagascar (Carroll 1988; Rage and Rocek 1989)
(8) MRCA of pipoids and all other frogs, at least 145.5 Mya, given *Rhadinosteus parvus*, ostensibly a rhinophrynid and clearly a pipoid, from the Late Jurassic (Tithonian, 145.5–150.8 Mya; Rocek 2000).
(9) MRCA of *Hymenochirus* and *Xenopus*, at least 83.5 Mya, given the pipid *Pachybatrachus taqueti* from the Upper Cretaceous (Coniacian-Santonian, 83.5–89.3 Mya), which is thought to be closely related to *Hymenochirus* (Rocek 2000).

(10) MRCA of Myobatrachidae (represented here by the limnodynastine *Neobatrachus* and the myobatrachine *Crinia*) at least 54.6 Mya, given fossils assigned to the limnodynastine genus *Lechriodus* (Evans et al., 2008b; Sanchiz, 1998).
(11) MRCA of Bufonidae + Leptodactylidae + Centrolenidae, at least 55.8 Mya, given putative fossil *Bufo* from the late Paleocene (55.8–58.7 Mya; Baéz 2000).
(12) MRCA of Ranidae (sensu Wiens et al. 2009) at least 33.9 Mya, given fossil *Rana* from the Late Eocene (37.2–33.9 Mya; Rocek and Rage 2000).
(13) Crown group of Terrarana (the clade including the families Brachycephalidae, Ceuthomantidae, Craugastoridae, Eleutherodactylidae, and Strabomantidae, or more simply, the clade including *Eleutherodactylus* and related genera) at least 35 Mya, based on an *Eleutherodactylus* fossil in amber from the La Toca formation (Dominican Republic) estimated to be ~35 Myo (Poinar and Cannatella 1987).
(14) Stem group of Ceratophryidae (the clade including the genera *Ceratophrys*, *Chacophrys*, and *Lepidobatrachus*) at least 65.5 Mya based on the late Cretaceous fossil genera *Beelzebufo* and *Baurubatrachus* (Evans et al. 2008b). Evans et al. (2008b) considered the Madagascan taxon *Beelzebufo* to be a ceratophryine. This taxon is of Maastrichtian (Late Cretaceous) age (65.5–70.6 Mya). The South American genus *Baurubatrachus* is also considered to be a ceratophryine (Evans et al. 2008b; Rocek 2000). Although the exact relationships of these taxa are somewhat uncertain, the presence of seemingly ceratophryine fossils in South America suggests that the stem group age of Ceratophryidae is at least 65.5 Mya. The relationships of ceratophryids are uncertain, but in this molecular analysis, they appear as the sister group to a clade including Odontophrynidae and Alsodidae (*Eupsophus)*.
(15) Crown-group age of North American and European *Hyla* clade, at least 16 Myo; given fossil *Hyla* similar to extant *H. arborea* and *H. meridionalis* in the Lower Miocene of Austria (~16 Myo; Sanchiz 1998). We assume that these *Hyla* are closely related to *Hyla* presently extant in Europe. However, we cannot assume that these fossils are younger than the crown-group age of the extant European species. We assume instead that the crown group of the clade of *Hyla* is at least 16 Myo based on these European fossils.
(16) MRCA of *H. avivoca-H. chrysocelis-H. versicolor* clade; *H. miocenica* is thought to be closely related to *H. chrysocelis* and *H. versicolor* and occurs in the Barstovian of the Middle Miocene (14–16 Myo; Holman 2003). In our phylogeny, *H. avivoca, H. chrysocelis*, and *H. versicolor* form a clade. We assume that the stem group age of these three species is at least 14 Myo.
(17) We fixed the root age of the tree using the estimated age from Wiens (2011) for the MRCA of lissamphibians (frogs, salamanders, caecilians) of 368.3 Mya, using penalized likelihood. Although the use of a fixed calibration point (rather than a minimum constraint) may seem controversial, it should be noted that at least one node must be given a fixed age. Furthermore, our focus here is not on re-estimating these ages, but providing relative assessments of clade ages (see below).

Construction of a chronogram using r8s requires a cross-validation step that identifies a best-fitting value for the "smoothing parameter", which specifies the cost of differing rates of evolution between neighboring branches (Sanderson 2002). Cross-validation considered smoothing parameter values from 10^0 –$10^{5.5}$ in exponential increments of 0.5. These cross-validation analyses failed until a species with a zero-length branch (*Bufo pewzoi*) was removed. After removing this species, the cross-validation analyses showed that a value of 10^1 gave the lowest Chi-squared error. We then generated a chronogram for the 142 remaining species using this smoothing parameter. We then performed a second analysis using this same smoothing parameter but including all 143 species. This second analysis gave identical divergence-date estimates throughout the tree as the first analysis with 142 species. The resulting chronogram (Fig. 18.1) shows species for which we have information on phylogenetic relationships and either heterogamy and/or ploidy.

Using the R package "ape" (Paradis et al. 2004), which is a software package for phylogenetic analysis, we tested the fit of alternative models for the evolution of new sex determination systems for a total of 90 species (55 frogs and 35 salamanders) for which we had data on heterogamy or sex chromosome polymorphism. This was done to select an appropriate model for ancestral reconstruction and for use in a permutation test described below. We coded all species as having either male heterogamy (0), female heterogamy (1), or, for each of three species (*Xenopus laevis*, *Leiopelma hochstetteri*, and *Rana rugosa*), a unique "de novo" sex-determining mechanism (2, 3, and 4), in which categories we include species with polymorphisms in mechanisms for sex determination. *Xenopus laevis* was assigned a unique heterogametic state in order to account for the finding that this species evolved its W-linked sex-determining gene after the split from *S. tropicalis* (Bewick et al. 2011), which also has female heterogamy. *Leiopelma hochstetteri* was assigned a unique heterogametic state in order to account for the finding that this species recently evolved a derived W0 sex-determining system that is unique to this lineage (Green et al. 1993; Sharbel et al. 1998). *Rana rugosa* was assigned a unique heterogamy state in order to accommodate evidence for recent and possibly repeated instances of sex chromosome turnover (Ogata et al. 2008). The three de novo states were coded as separate character states in order to ensure that known instances of novel sex determination mechanisms were included in the analysis even though they did not necessarily involve a change in heterogamy. We note that the newly evolved sex chromosomes of *L. hochstetteri* and *R. rugosa* are polymorphisms, and it is not clear whether these new polymorphisms will eventually fix in each species, and thus actually constitute a sex chromosome turnover. However, in both of these examples, at least one of the polymorphic systems for sex determination is species-specific and therefore new.

Using the "ace" function of the "ape" package, we then evaluated the following models for evolution of the five heterogamy states:

(1) All rates equal (one rate for all possible transitions between states, one parameter).
(2) One reversible rate between XY and ZW, and one reversible rate to and from any of the de novo states (two rates in total, two parameters). By reversible, we

mean, for example, that the rate of change from XY to ZW is equal to the rate of change from ZW to XY.
(3) One rate for XY to ZW, another rate for ZW to XY, and one reversible rate to and from any of the de novo states (three rates in total, three parameters).
(4) Rates between each of the five heterogamy states are reversible and unique (ten rates in total, ten parameters).
(5) All rates unique (twenty rates in total, twenty parameters).

18.2.2 Results and Discussion

The likelihoods of each of these models were compared using the Akaike Information Criterion (Akaike 1974) calculated as $2k–2\ln(L)$ where k is the number of parameters in the model and L is the maximum value of the likelihood function of the model. A P value was generated with a hierarchical likelihood ratio test (hLRT) with degrees of freedom equal to the difference in free parameters of the models under the assumption of a Chi-squared distribution. For the hLRT we evaluated whether adding complexity to the models resulted in a significant increase in model fit. The likelihoods of Models 1, 2, 3, 4 and 5 were -70.74572 (AIC = 143.5), -53.82688 (AIC = 111.7), -53.81533 (AIC = 113.6), -52.31287 (AIC = 124.6), and -49.97701 (AIC = 140.0) respectively. Model 2 thus was favored by the Akaike Information Criterion. According to the hLRT, Model 2 was also preferred over Model 1 ($P<0.0001$), but Model 3 was not preferred over Model 2 ($P = 0.879$). Model 4 was not preferred over Model 2 ($P = 0.932$) or Model 3 ($P = 0.885$), and Model 5 was not preferred over Model 2 ($P = 0.982$), or Model 3 ($P = 0.999$), or Model 4 ($P = 0.912$). These results suggests that the transition rate from ZW to XY is not significantly higher than the transition rate from XY to ZW. Model 2 was therefore used to reconstruct ancestral heterogamy states and also used for simulations in our permutation test described below. The model used in the ancestral reconstructions differs slightly from the model used in the permutation test in that the rate of reversal from the de novo heterogamy states to other heterogamy states was set to zero for the ancestral reconstructions. This was not possible with the permutation test, which requires a reversible model. We present results from this slightly different version for the ancestral reconstructions for illustrative purposes because with this model there is zero likelihood for all of the de novo states in all of the ancestral reconstructions. Other inferences, such as the likelihood of male and female heterogamy in the most recent common ancestor of frogs, of salamanders, and of frogs and salamanders discussed below, are identical with both of these models.

The ancestral state reconstructions estimated from the analysis with 90 species but plotted on the chronogram with 143 species (Fig. 18.1) suggest that there is not strong statistical support to distinguish whether the ancestral hetrogamy state was female or male heterogamy (that is, ZW females and ZZ males) in frogs or salamanders. The marginal likelihood of female heterogamy (ZW) for the most recent common ancestor of frogs is 0.544, for the most recent common ancestor of

salamanders is 0.499, and for the most recent common ancestor of frogs and salamanders is 0.513. Thus, the support for ZW versus XY heterogamy as the ancestral state in each group is effectively equivocal.

18.3 Is Polyploidy Tolerated to a Greater Degree in Species with Young Sex Chromosomes?

18.3.1 Methods

In our analysis, we consider the sex-determining system to have changed every time that a change in heterogamy occurred from XY to ZW, from ZW to XY, or from ZW or XY to one of the three de novo sex-determining systems. To quantify the number of times that the system for sex determination changed in amphibians, we used the stochastic character mapping approach proposed by Nielsen (2002) as implemented by the R package "phytools" (Revell 2011). This approach simulates character evolution on a phylogeny, conditioning on the observed character states of the terminals. In this way, one can estimate the number of character-state transitions that occurred and also evaluate where in the phylogeny changes are likely to have occurred. The stochastic mapping simulations were performed using a reversible version of Model 2 described above.

As discussed earlier, recent discoveries implicate sex chromosome turnover in facilitating polyploid speciation or the tolerance of polyploidization, at least in *Xenopus* and *Leiopelma*. These observations raise the question: is this a general phenomenon in amphibians? Since a change in heterogamy necessarily involves a change in the sex-determining system, we predicted that the time since a change in heterogamy (XY to ZW or ZW to XY) would be lower in species that are polyploid or that tolerate polyploidy than expected by chance, if the same number of polyploid species (or polyploid-tolerant species) were to evolve randomly on the phylogeny. We note that this hypothesis does not involve correlation between polyploidy and a particular heterogamy state, so standard approaches to test for phylogenetic correlation among traits cannot be used to test this prediction. Instead, we developed a novel permutation test that accommodates uncertainty in when and in which lineages sex chromosome turnover occurred during amphibian evolution.

From the set of 90 species for which heterogamy information was available, we identified five phylogenetically independent instances of stable or spontaneous polyploidy. We emphasize that this is an underestimate of the number of independent polyplodization events, and we were able to use only five instances of polyploidization because we lack heterogamy and/or phylogenetic information for the other examples listed in Table 18.1. Diploidy has either been confirmed for the other species for which heterogamy information was available (Mable et al. 2011) or was assumed. The five examples of independent polyploidization or tolerance of polyploidization are:

(1) *Siren intermedia*, a tetraploid, which may have descended from a tetraploid ancestor that also gave rise to *S. lacertina* (Morescalchi and Olmo 1974). We note that the polyploid status of the family Sirenidae has not been confirmed by additional studies (Mable et al. 2011).
(2) *Ambystoma mexicanum*, a species with spontaneous triploidy (Humphrey 1963) that is closely related to the unisexual triploids *A. jeffersonianum, A. platineum,* and *A. tremblayi* (see notes on species status of unisexuals in Table 18.1).
(3) *Xenopus laevis*, a tetraploid species (Tymowska 1991). We excluded *X. gilli* from this analysis even though we have heterogamy data for this species because it shares a polyploid ancestor with *X. laevis*.
(4) *Rana esculenta,* a naturally occurring diploid/triploid hybridogenic species formed from hybridization of *R. lessonae* and *R. ridibunda* (Uzzell et al. 1975).
(5) *Leiopelma hochstetteri*, a diploid species with spontaneous triploidy (Green et al. 1984).

For this permutation test, we used as a test statistic the mean time since the origin of the current heterogamy state for the five polyploid lineages. This mean was calculated from 1,000 simulations that are conditioned on the observed heterogamy states, using the stochastic character mapping approach described by Nielsen (2002) and implemented by the R package "phytools" (Revell 2011). Each stochastic mapping simulation provides one possible evolutionary scenario that is consistent with the data. It was necessary to perform many (1,000) stochastic mapping simulations for the observed data in order to accommodate uncertainty in these evolutionary scenarios (that is, to accommodate uncertainty in the ancestral reconstruction of the evolution of heterogamy).

If polyploidy tends to occur soon after a change in heterogamy, then the observed test statistic should be smaller than the distribution of statistics calculated after repeatedly randomly selecting five species across the tree to be polyploid, and performing 1,000 stochastic character mapping simulations for each of the randomizations. The observed test statistic was therefore compared to a distribution of statistics generated from 100 randomizations where, in each randomization, five species are selected to be polyploid, with 1,000 stochastic character mapping simulations that were conditioned on the observed heterogamy states performed for each of the randomizations. The difference between the test statistic and the randomizations therefore is that the mean path length (that is, for each polyploid, the path length between the most recent change in heterogamy and the branch tip) was calculated respectively either from real polyploid species (for the test statistic), or from five species selected at random from the 90 species for which we have heterogamy data (for each of the randomizations).

This test is conservative in the sense that it does not consider subsequent, phylogenetically independent instances of polyploidization that occurred in *Xenopus* (three additional independent instances of octoploidization and at least two additional independent instances of dodecaploidization). More specifically, because *Xenopus* has a relatively young sex-determining system, a test statistic generated by counting each of the

independent polyploidization events in *Xenopus* is even lower than the test statistic that counts polyploidization of *Xenopus* only once (see below).

One limitation of this analysis is that the frequency of transitions in amphibian sex-determining systems is undoubtedly underestimated. It is possible, for example, that changes in heterogamy occurred in other species that were excluded in the analysis because differences between the sex chromosomes were not cytologically detectable. It is also possible that some species experienced a change in the sex-determining system that did not involve a change in heterogamy. *Gastrotheca pseustes*, for instance, is known to have polymorphism in the morphology of the Y chromosome (Schmid et al. 1990), but this species was coded as XY because these size variants may involve homologous chromosomes and no change in the sex-determining system. Intraspecific polymorphism in sex-determining mechanisms has also been observed in *Rana narina*, *Eleutherodactylus maussi*, *Rana japonica* and *R. narina* (Eggert 2005), but phylogenetic information was lacking from these species in the phylogeny of Pyron and Wiens (2011). Furthermore, adding more taxa might influence the inferred timing of the transitions to heterogamy, even if the phylogeny remains the same (e.g., added taxa could subdivide long branches and help clarify where on a given branch the heterogamy transition occurred).

Another limitation of this analysis is that we use the total time that the polyploid species have been in the observed heterogamy states for our test statistic, rather than the difference between these times and the age of each polyploid species. The latter difference would be a better metric for this test because it focuses on events prior to polyploidization. However, it is difficult to estimate the age of each polyploid (or of a randomly selected species in the permutation) because in some cases the diploid ancestor of the polyploid is unknown or extinct (see above), or other cases because we lack phylogenetic information from the sister taxon. We note that results are contingent on the phylogeny and evolutionary model, and that this analysis does not accommodate uncertainty in divergence times and phylogenetic relationships.

18.3.2 Results and Discussion

Stochastic mapping of heterogamy state, including the independent evolution of three de novo sex chromosomes provides an average estimated number of times that the sex chromosomes turned over of 32 (95% confidence interval: 25–41) based on 1,000 simulations that were conditioned on the observed heterogamy states. This is much higher than the maximum parsimony inference of only seven changes by Hillis and Green (1990). In the example simulation depicted in Fig. 18.2a, for instance, there are 28 changes in heterogamy.

These simulations also suggest that more changes occurred from male heterogamy to female heterogamy than the reverse, even though our model comparison suggested that the rate of change in each direction was not significantly different. Out of 1,000 stochastic mapping simulations, the mean number of changes from

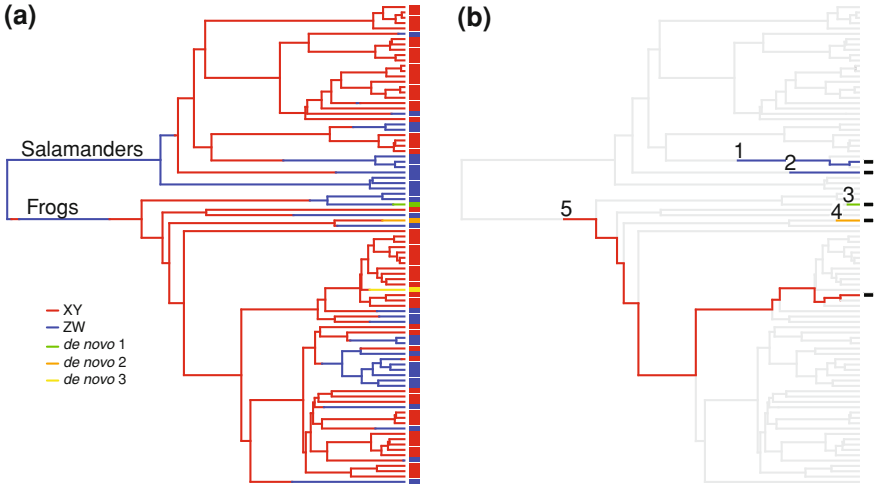

Fig. 18.2 Simulations that stochastically map evolution of heterogamy form the basis of the permutation test. Shown here is (**a**) an example of a stochastic mapping simulation for the evolution of the observed heterogamy states and (**b**) highlighted simulated paths to each of the five observed polyploid species. Species with missing heterogamy data from Fig. 18.1 have been removed for these analyses and species names are omitted for clarity. In (**b**) simulated ages of the observed heterogamy states for *Ambystoma mexicanum*, *Siren intermedia*, *Leiopelma hochstetteri*, *Xenopus laevis*, and *Rana esculentia* are labeled 1, 2, 3, 4, and 5 respectively. The test statistic is the average age of the observed heterogamy state for these five polyploid species, averaged over 1,000 stochastic mapping simulations. This test statistic is compared to analogous calculations from 100 permutations where five species were randomly selected to be polyploid and 1,000 stochastic mapping simulations were performed for each randomization

male to female heterogamy (XY to ZW) was 22 (95 % confidence interval: 13–28) and the mean number of changes from female to male heterogamy (ZW to XY) was 7 (95 % confidence interval: 2–14). The example simulation depicted in Fig. 18.2a is typical of the other simulations in the sense that there are 19 changes from male to female heterogamy but only 6 changes from female to male heterogamy.

The observed average path length to a change in the sex-determining system for the five polyploids averaged over 1,000 stochastic mapping simulations, was 89.0 million years. This is not to suggest that no sex chromosome degeneration occurred within this period. In fact, if sex chromosome degeneration in amphibians occurred at a similar rate as it did in therian mammals (4.6 genes per million years; Graves 2004), the ancestors of these polyploids probably did have somewhat degenerate sex chromosomes. Rather, this result suggests that the amount of sex chromosome degeneration that typically occurs within this period of time was not of sufficient magnitude to prevent polyploidization or the tolerance of spontaneous polyploidization. The permutation test indicates that this test statistic is not significantly lower than the distribution of statistics calculated when polyploids

evolved five times on random branches in this phylogeny ($P = 0.059$; average path length in permutations was 141.2 million years and the standard deviation was 32.7 million years), although the P value is close to 0.05. If we include six independent polyploidizations in *Xenopus* (one tetraploidization, three octoploidizations, and at least two dodecaploidizations; reviewed in Evans 2008) in the observed test statistic, in addition to the four other examples of polyploidization itemized above, the average observed path length to a change in the sex-determining system for the ten independent polyploid lineages is 68.1 million years. This test statistic is significantly smaller than statistics calculated from 100 permutations where one of the five randomly selected polyploids is also assumed to undergo six independent polyploidizations ($P = 0.020$; average path length in permutations was 169.8 million years and the standard deviation was 55.3 million years). Although there are at least six independent polyploidizations in *Xenopus*, this test suffers from pseudoreplication in that these polyploid lineages may share the same system for sex determination (i.e., *DMW*). Additional data on whether other polyploid species of amphibians have male or female heterogamy would clearly help illuminate the question of whether species with young sex chromosomes are more tolerant of polyploidization. It is surprising how little is known about heterogamy of polyploid amphibians given that karyotypes of essentially all of these species were inspected in order to identify polyploidy in the first place. One possible reason for this dearth of information on heterogamy of polyploid species is that many of these species may lack morphologically distinct sex chromosomes. This proposal, if accurate, would be consistent with the contention that polyploidization is better tolerated by species with minimally degenerate sex chromosomes.

Additional insights into the influence of sex chromosome evolution and polyploidization may be gained from studies of laboratory-generated polyploid *Xenopus*. Laboratory allopolyploidization in *Xenopus* duplicates autosomal chromosomes but generates female polyploid individuals with 3 Z chromosomes and 1 W chromosome and male polyploid individuals with 4 Z chromosomes (Kobel and Du Pasquier 1986). Thus, the W chromosome of one of the ancestral diploids is not inherited by *Xenopus* allopolyploids, and therefore never gets a chance to segregate as an autosome. In this way, the mechanism of *Xenopus* allopolyploidization in nature may circumvent autosomal segregation of a W chromosome (whether degenerate or not) (reviewed in Evans 2008), and this could account for the unusually high incidence of polyploidization in this genus.

Analyses presented here show that changes in the system for sex chromosome turnover were much more common in amphibians than previously proposed [$\sim$32 versus 7 as proposed by Hillis and Green (1990)], and that there is not strong support for female versus male heterogamy in the ancestor of salamanders, frogs, or the most recent common ancestor of salamanders and frogs. These changes need not involve the evolution of completely novel systems for sex determination, and some of these inferred changes may be reversals to an ancestral system. We also found that changes from male to female heterogamy occurred more frequently than changes from female to male heterogamy, although the transition rates

between each state were not significantly different. This result also contradicts the conclusions of Hillis and Green (1990), who stated that there was a bias in evolution from female heterogamy to male heterogamy. One reason for these differences is that our analysis included new data on heterogamy (reviewed in Schmid et al. 2010), more species, and a more comprehensive phylogeny (Pyron and Wiens 2011). Another reason for our higher estimate in the number of changes is that we considered three de novo changes in sex determination as a new heterogamy state, thereby forcing a sex chromosome turnover in these lineages (but this is only 3 out of $\sim$32 changes). Differences in the resolution, relationships, and branch lengths of the phylogenies used in each study are also likely to have played a role in these differing conclusions. Finally, and importantly, the different analytical approaches may have influenced the results (that is, the use of maximum parsimony by Hillis and Green (1990) and maximum likelihood here).

18.4 Conclusions

Polyploidization generates new species and duplicates genes; the resulting genetic redundancy has the capacity to degrade or to undergo innovation. The question of why some lineages frequently undergo polyploidization whereas others do not thus has important implications for evolution and adaptation. Eventually we will have a much more comprehensive understanding of genetic variation in (a) the triggers of sex determination in amphibians, (b) the extent of suppressed recombination that surrounds these genetic triggers, (c) the extent of sex chromosome degeneration that exists in amphibians, and (d) whether or not other lineages of polyploid amphibians have minimally degenerate sex chromosomes. Future discoveries in these areas can undoubtedly be leveraged to provide exciting new insights into the role of sex chromosome degeneration in the propensity of species to tolerate polyploidization.

Acknowledgements We are particularly grateful to Barbara Mable who provided a comprehensive critical assessment of an earlier version of this chapter. We also thank Liam Revell for advice and assistance with the R package "phytools" and Jim Bogart and Ben Bolker for comments.

References

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Autom Contr 19:716–723

Arnold AP, Itoh Y, Melamed E (2008) A bird's-eye view of sex chromosome dosage compensation. Annu Rev Genomics Hum Genet 9:109–127

Baéz AM (2000) Tertiary anurans from South America. In: Heatwole H, Carroll RL (eds) Amphibian biology. Surrey Beatty, Chipping Norton, Australia, pp 1388–1401

Beçak ML, Beçak W (1998) Evolution by polyploidy in amphibia: new insights. Cytogenet Cell Genet 80:28–33

Bergero R, Charlesworth D (2009) The evolution of restricted recombination in sex chromosomes. Trends Ecol Evol 24:94–102
Bergero R, Charlesworth D (2011) Preservation of the Y transcriptome in a 10 million-year-old plant sex chromosome system. Curr Biol 21:1470–1474
Bewick AJ, Anderson DW, Evans BJ (2011) Evolution of the closely related, sex-related genes DM-W and DMRT1 in African clawed frogs (*Xenopus*). Evolution 65:698–712
Blackburn DC and Beier M (2011) *"Xenopus paratropicalis"* is not a valid name. Zootaxa 3035:57–58
Briggs R (1947) The experimental production and development of triploid frog embryos. J Exp Zool 106:237–266
Bogart JP (1980) Evolutionary significance of polyploidy in amphibians and reptiles. In: Lewis WH (ed) Polyploidy, biological relavance. Basic life sciences, New York, pp 341–378
Carroll RL (1988) Vertebrate paleontology and evolution W. H. Freeman and Company, New York
Charlesworth B, Charlesworth D (2000) The degeneration of Y chromosomes. Philos Trans R Soc London B 355:1563–1572
Charlesworth D (2002) Plant sex determination and sex chromosomes. Heredity 88:94–101
Charlesworth D, Charlesworth B, Mariais G (2005) Steps in the evolution of heteromorphic sex chromosomes. Heredity 95:118–128
Duellman WE, Trueb L (1994) Biology of amphibians. The Johns Hopkins University Press, Baltimore
Eggert C (2005) Sex determination: the amphibian models. Reprod Nutr Dev 44:539–549
Evans BJ (2007) Ancestry influences the fate of duplicated genes millions of years after duplication in allopolyploid clawed frogs (*Xenopus*). Genetics 176:1119–1130
Evans BJ (2008) Genome evolution and speciation genetics of allopolyploid clawed frogs (*Xenopus* and *Silurana*). Front Biosci 13:4687–4706
Evans BJ, Cannatella DC, Melnick DJ (2004a) Understanding the origins of areas of endemism in phylogeographic analyses: a reply to Bridle et al. Evolution 58:1397–1400
Evans BJ, Kelley DB, Tinsley RC, Melnick DJ, Cannatella DC (2004b) A mitochondrial DNA phylogeny of clawed frogs: phylogeography on sub-Saharan Africa and implications for polyploid evolution. Mol Phylogenet Evol 33:197–213
Evans BJ, Carter TF, Hanner R et al (2008a) A new species of clawed frog (genus *Xenopus*), from the Itombwe Plateau, Democratic Republic of the Congo: implications for DNA barcodes and biodiversity conservation. Zootaxa 1780:55–68
Evans BJ, Greenbaum E, Kusamba C et al (2011) Description of a new octoploid frog species (Anura: Pipidae: *Xenopus*) from the Democratic Republic of the Congo, with a discussion of the biogeography of African clawed frogs in the Albertine Rift. J Zool London 283:276–290
Evans BJ, Kelley DB, Melnick DJ, Cannatella DC (2005a) Evolution of RAG-1 in polyploid clawed frogs. Mol Biol Evol 22:1193–1207
Evans SE, Jones MEH, Krause DW (2008b) A giant frog with South American affinities from the Late Cretaceous of Madagascar. Proc Nat Acad Sci 105:2951–2956
Evans SE, Lally C, Chure DC, Elder A, Maisano JA (2005b) A new fully metamorphosed salamander from the Late Jurassic of North America. Zool J Linn Soc 143
Evans SE, Milner AR (1996) A metamorphosed salamander from the early Cretaceous of Las Hoyas, Spain. Philos Trans R Soc London B 351:627–646
Ezaz T, Stiglec R, Veyrunes F, Graves JAM (2006) Relationships between vertebrate ZW and XY sex chromosome systems. Curr Biol 16:R736–R743
Felsenstein J (1974) The evolutionary advantage of recombination. Genetics 78:737–756
Fankhauser G, Crotta R, Perrot M (1942) Spontaneous and cold-induced triploidy in the Japanese newt *Triturus pyrrhogaster*. J Exp Zool 89:167–181
Fankhauser G (1941) The frequency of polyploidy and other spontaneous aberrations of chromosome number among larvae of the newt *Triturus viridescens*. Proc Nat Acad Sci 27:507–512
Fankhauser G, Watson RC (1942) Heat-indiced triploidy in the newt, *Triturus viridescens*. Proc Nat Acad Sci 28:436–440
Gardner JD (2003) The fossil salamander *Proamphiuma cretacea* Estes (Caudata: Amphiumidae) and relationships within the Amphiumidae. J Vertebr Paleontol 23:769–782

Graves JAM (2004) The degenerate Y chromosome—can conversion save it? Reprod Fertil Dev 16:527–534
Graves JAM (2008) Weird animal genomes and the evolution of vertebrate sex and sex chromosomes. Annu Rev Genet 42:565–586
Green DM (1988) Cytogenetics of the endemic New Zealand frog, *Leiopelma hochstetteri*: extraordinary supernumerary chromosome variation and a unique sex-chromosome system. Chromosoma 97:55–70
Green DM, Kezer J, Nussbaum RA (1984) Triploidy in Hochstetter's frog, *Leiopelma hochstetteri*, from New Zealand. New Zealand J Zool 11:457–460
Green DM, Zeyl CW, Sharbel TF (1993) The evolution of hypervariable sex and supernumerary (B) chromosomes in the relict New Zealand frog, *Leiopelma hochstetteri*. J Evol Biol 6:417–441
Gregory TR, Mable BK (2005) Polyploidy in animals. In: Gregory TR (ed) The Evolution of the Genome. Elsevier Academic Press, Burlington, pp 428–517
Hayes TB (1998) Sex determination and primary sex differentiation in amphibians: genetic and developmental mechanisms. J Exp Zool 281:373–399
Hillis DM, Green DM (1990) Evolutionary changes of heterogametic sex in the phylogenetic history of amphibians. J Evol Biol 3:49–64
Holloway AK, Cannatella DC, Gerhardt HC, Hillis DM (2006) Polyploids with different origins and ancestors form a single sexual polyploid species. Am Nat 167:E88–E101
Holman JA (2003) Fossil frogs and toads of North America Indiana University Press. Bloomington and Indianapolis, IN
Kashiwagi K (1993) Production of triploids and their reproductive capacity in *Rana rugosa*. Sci Rep Lab Amphibian Biol Hiroshima Univ 12:23–36
Kawamura T, Tokunaga C (1952) The sex of triploid frogs, *Rana japonica* Günther. J Sci Hiroshima Univ, Ser B, Div 1 (Zoology) 13
Humphrey RR (1963) Polyploidy in the Mexican axolotl (*Ambystoma mexicanum*) resulting from multinucleate ova. Proc Nat Acad Sci 50:1122–1127
Kobel HR, Loumont C, Tinsley RC (1996) The extant species. In: Tinsley RC, Kobel HR (eds) The Biology of *Xenopus*. Clarendon Press, Oxford, pp 9–33
Kawamura T (1984) Polyploidy in amphibians. Zool Sci 1:1–15
Kobel HR (1996) Allopolyploid speciation. In: Tinsley RC, Kobel HR (eds) The Biology of *Xenopus*. Clarendon Press, Oxford, pp 391–401
Kobel HR, Du Pasquier L (1986) Genetics of polyploid *Xenopus*. Trends Genet 2:310–315
Litvinchuk SN, Rosanov JM, Borkin LJ (1998) A case of natural triploidy in a smooth newt *Triturus vulgaris* (Linneaus, 1958), from Russia (Caudata: Salamandridae). Herpetozoa 11:93–95
Mable BK (2004) 'Why polyploidy is rarer in animals than in plants': myths and mechanisms. Biol J Linn Soc 82:453–466
Mable BK, Alexandrou MA, Taylor MI (2011) Genome duplication in amphibians and fish: an extended synthesis. J Zool 284:151–182
Mable BK, Roberts JD (1997) Mitochondrial DNA evolution in the genus *Neobatrachus* (Anura: Myobatrachidae). Copeia 1997:680–689
Mayrose I, Zhan SH, Rothfels CJ et al (2011) Recently formed polyploid plants diversify at lower rates. Science 333:1257
Milner AR (2000) Mesozoic and Tertiary Caudata and Albanerpetontidae. In: Heatwole H, Carrol RL (eds) Amphibian Biology. Surrey Beatty, Chipping Norton, Australia, pp 31–108
Morescalchi A, Olmo E (1974) Sirenids: a family of polyploid urodeles? Experientia 30:491–492
Moler PE, Kezer J (1993) Karyology and systematics of the salamander genus *Pseudobranchus* (Sirenidae). Copeia 1993:39–47
Muller HJ (1925) Why polyploidy is rarer in animals in plants. Am Nat 59:346–353
Muller HJ (1964) The relation of recombination to mutational advance. Mutat Res 106:2–9
Naylor BG, Fox RC (1993) A new ambystomatid salamander *Dicamptodon antiquus* n. sp. from the Paleocene of Alberta. Can J Earth Sci 30:814–818
Nielsen R (2002) Mapping mutations on phylogenies. Syst Biol 51:729–739

Ogata M, Hasegawa Y, Ohtani H, Mineyama M, Miura I (2008) The ZZ/ZW sex-determining mechanism originated twice and independently during evolution of the frog, *Rana rugosa*. Heredity 100:92–99

Ohno S (1967) Sex chromosomes and sex-linked genes. Springer, Berlin

Orr HA (1990) 'Why polyploidy is rarer in animals than in plants' revisited. Am Nat 136:759–770

Otto SP, Whitton J (2000) Polyploid incidence and evolution. Annu Rev Genet 34:401–437

Papp B, Pál C, Hurst LD (2003) Dosage sensitivity and the evolution of gene families in yeast. Nature 424:194–197

Paradis E, Claude J, Strimmer K (2004) APE: analysis of phylogenetics and evolution in R language. Bioinf 20:289–290

Poinar GO, Cannatella DC (1987) An upper Eocene frog from the Dominican Republic and its implication for Caribbean biogeography. Science 237:1215–1216

Pyron RA, Wiens JJ (2011) A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. Mol Phylogenet Evol 61:543–583

Qian W, Zhang J (2008) Gene dosage and gene duplicability. Genetics 179:2319–2324

Rage JC, Rocek Z (1989) Redescription of *Triadobatrachus massinoti* (Piveteau, 1936) an anuran amphibian from the early Triassic. Palaeontographica Paleontologica 206:1–16

Revell JJ (2011) Phytools: phylogenetic tools for comparative biology (and other things). (R Package)

Rocek Z (2000) Mesozoic anurans. In: Heatwole H, Carrol RL (eds) Amphibian biology. Surrey Beatty, Chipping Norton, Australia, pp 1295–1331

Rocek Z, Rage J-C (2000) Tertiary Anura of Europe, Africa, Asia, North America, and Australia. In: Heatwole H, Carrol RL (eds) Amphibian Biology. Surrey Beatty, Chipping Norton, Australia, pp 1332–1387

Sanchiz FB (1998) Salienta. In: Wellnhofer P (ed) Encyclopedia of paleoherpetology, Part 4, Salienta. Verlag, Pfeil, Munich, pp 1–276

Sanderson MJ (2002) Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol Biol Evol 19:1218–1231

Sanderson MJ (2003) r8s: inferring absolute rates of evolution and divergence times in the absence of a molecular clock. Bioinformatics 19:301–302

Schmid M, Sims SH, Haaf T, Macgregor HC (1986) Chromosome banding in amphibia X. 18S and 28S ribosomal RNA genes, nucleolus organizers and nucleoli in *Gastrotheca riobambae*. Chromosoma 94:139–145

Schmid M, Steinlein C (2001) Sex chromosomes, sex-linked genes, and sex determination in the vertebrate class Amphibia. In: Scherer G, Schmid M (eds) Genes and mechanisms in vertebrate sex determination. Verlag, Basel, pp 143–176

Schmid M, Steinlein C, Bogart JP et al (2010) The chromosomes of terraranan frogs. Cytogenetic Genome Res 130–131:1–568

Schmid M, Steinlein C, Friedl R et al (1990) Chromosome banding in Amphibia. XV. Two types of Y chromosomes and heterochromatin hypervariability in *Gastrotheca pseustes* (Anura, Hylidae). Chromosoma 99:413–423

Sharbel TF, Green DM, Houben A (1998) B-chromosome origin in the endemic New Zealand frog *Leiopelma hochstetteri* through sex chromosome devolution. Genome 41:14–22

Skaletsky H, Kuroda-Kawaguchi T, Minx PJ et al (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 432:823–837

Stöck M, Horn A, Grossen C et al (2011) Ever-young sex chromosomes in European tree frogs. PLoS Biol 9:e1001062

Stöck M, Ustinova J, Lamatsch DK et al (2009) A vertebrate reproductive system involving three ploidy levels: Hybrid origin of triploids in a contact zone of diploid and tetraploid Paleartic green toads (*Bufo viridis* subgroup). Evolution 64:944–959

Straub T, Becker PB (2007) Dosage compensation: the beginning and end of a generalization. Nat Rev Genet 8:47–57

Svartman M, Stone G, Stanyon R (2005) Molecular cytogenetics discards polyploidy in mammals. Genomics 85:425–430

Tihen JA, Wake DB (1981) Vertebrae of plethodontid salamanders from the Lower Miocene of Montana. J Herpetology 15:35–40

Tymowska J (1991) Polyploidy and cytogenetic variation in frogs of the genus *Xenopus*. In: Green DS, Sessions SK (eds) Amphibian cytogenetics and evolution. Academic Press, San Diego, pp 259–297

Uzzell T, Berger L, Günther R (1975) Diploid and triploid progeny from a diploid female of *Rana esculenta* (Amphibia Salientia). Proc Acad Nat Sci Philadelphia 127:81–91

Wiens JJ (2011) Re-evolution of lost mandibular teeth in frogs after more than 200 million years, and re-evaluating Dollo's Law. Evolution 65:1283–1296

Wiens JJ, Sukumaran J, Pyron RA, Brown RM (2009) Evolutionary and biogeographic origins of high tropical diversity in Old World frogs (Ranidae). Evolution 63:1217–1231

Wolfe KH (2001) Yesterdays's polyploids and the mystery of diploidization. Nat Rev Genet 2:333–341

Yoshimoto S, Ikeda K, Izutsu Y et al (2010) Opposite roles of *DMRT1* and its W-linked paralog, *DM-W*, in sexual dimorphism of *Xenopus laevis*: implications of a ZZ/ZW-type sex-determining system. Development 137:2519–2526

Yoshimoto S, Okada E, Umemoto H et al (2008) A W-linked DM-domain gene, DM-W, participates in primary ovary development in *Xenopus laevis*. Proc Nat Acad Sci 105:2469–2474

Chapter 19
Erratum to—Polyploid Evolution in *Spartina*: Dealing with Highly Redundant Hybrid Genomes

M. Ainouche, H. Chelaifa, J. Ferreira, S. Bellot, A. Ainouche and A. Salmon

Erratum to: Polyploid Evolution in *Spartina*: Dealing with Highly Redundant Hybrid Genomes, DOI: 10.1007/978-3-642-31442-1_12

Page	Item or line	Corrections
225	Line 9, Abstract	"the TETRAPLOID maritima (a European native)..." should be modified as "...the HEXAPLOID maritima (a European native)... "
232		"The Californian S. DENSIFLORA x foliosa hybrids ..." should be "The Californian S. ALTERNIFLORA x foliosa hybrids ..."

The online version of the original chapter can be found at 10.1007/978-3-642-31442-1_12.

M. Ainouche (✉) · H. Chelaifa · J. Ferreira · S. Bellot · A. Ainouche · A. Salmon
University of Rennes 1, UMR CNRS 6553 Ecobio, Bât. 14A, Campus Scientifique de Beaulieu, 35042, Rennes Cedex, France
e-mail: malika.ainouche@univ-rennes1.fr

P. S. Soltis and D. E. Soltis (eds.), *Polyploidy and Genome Evolution*,
DOI: 10.1007/978-3-642-31442-1_19,

Index

P. S. Soltis and D. E. Soltis (eds.), *Polyploidy and Genome Evolution*,
DOI: 10.1007/978-3-642-31442-1, © Springer-Verlag Berlin Heidelberg 2012