Chapter 6
Transposons in Cereals: Shaping Genomes and Driving Their Evolution

Jan P. Buchmann, Beat Keller and Thomas Wicker

6.1 Introduction

A molecular understanding of genome evolution depends on the availability of the complete DNA sequence of an organism, ideally in the form of a complete sequence or at least as good datasets of partial sequences. Thus, evolutionary genomics and specifically our understanding of the role of transposons in the evolution of (cereal) genomes have been intimately linked to the progress in whole genome analysis. The first completely sequenced plant genome was from the model plant *Arabidopsis thaliana* with a size of ~120 Mbp (AGI 2000). Two years later, the complete sequences of the first grass genomes were published: *Oryza sativa* L ssp. *japonica* and *Oryza sativa* L ssp. *indica* (Goff et al. 2002; Yu et al. 2002). These first genomes were sequenced by the "BAC-by-BAC" approach: constructing a BAC library of the genomes, fingerprinting the BAC clones and assembling them into a minimum tiling path which was then sequenced by shotgun-sequencing. This approach creates a high quality sequence which is ordered along the chromosomes, but is very laborious and expensive.

The rapid development in the field of DNA sequencing technology has resulted in faster and cheaper methods, allowing to sequence and assemble *de novo* entire genomes using a whole-genome shotgun (WGS) approach. This approach was used to sequence the genomes of the two grass species Sorghum and *Brachypodium*, as well as soybean (Paterson et al. 2009; IBI 2010; Schmutz et al. 2010). However, WGS sequencing has limits in cases where size and complexity of the analyzed genomes are large, e.g. the sequencing of the large and repetitive genomes from barley or wheat will critically depend on anchoring

J. P. Buchmann · B. Keller (✉) · T. Wicker
Institute of Plant Biology, University of Zurich, Zollikerstrasse 107, 8008 Zürich, Switzerland
e-mail: bkeller@botinst.uzh.ch

P. K. Gupta and R. K. Varshney (eds.), *Cereal Genomics II*,
DOI: 10.1007/978-94-007-6401-9_6,

shotgun sequences or individual BACs to genetic maps. The largest plant genome sequenced so far is the 2,300 Mbp genome of maize (Schnable et al. 2009).

The goal of each sequencing project is to retrieve the so called "pseudomolecules" which represent the chromosomes of the sequenced organism. However, the obtained sequence will not be one complete molecule per chromosome, whatever sequencing strategy and method is used. Problematic sequences which introduce gaps in genomic sequences are usually highly repetitive regions, e.g. centromeric regions or ribosomal DNA clusters. Therefore, genome sequences are gradually and continuously improved after the first release and even high quality genomes like those from rice and *Arabidopsis*, which are now in the sixth and ninth release, still contain gaps (Table 6.1). Not all plant genomes have yet reached such high quality standards. For example, the genomes of *Physcomitrella patens* (Rensing et al. 2008), a moss plant species, poplar (Tuskan et al. 2006) or grapevine (Jaillon et al. 2007) went through only one initial round of shotgun sequencing and the resulting assemblies consist of "supercontigs" or "scaffolds", which in many cases have not yet been assigned to specific chromosomes and contain thousands of sequence gaps (Table 6.1). Newer sequencing techniques, e.g. 454 and Illumina, create an enormous amount of sequence reads in every run and those assemblies are still a great challenge for existing software. In addition, these low-cost and faster sequencing techniques led to a rapid growth in the number of available genomes which show different levels of completeness. At the end of 2011, there were 25 genome projects listed on www.phytozome.org, of which five are from the grass family (Mayer et al. 2011; Berkman et al. 2011).

Chain et al. (2009) proposed a classification system, where the genomic sequences are classified into five categories depending upon technology used and the quality of the assembly. The categories range from the lowest *Standard Draft* (category 1), which represents a basic automated assembly of raw sequences to the highest *Finished* (category 5) with no gaps and less than 1 sequence error in 100 kb. Some microbial genomes have reached the *Finished* status, while plant genomes are found between the categories 1 through 4.

Table 6.1 Numbers of scaffolds and gaps in a selection of publicly available plant genomes

Organism	Size (Mbp)	Version	Scaffolds[a]	Gaps[b]	Gaps/Mbp	References
Arabidopsis	119	9	5	96	0.8	AGI (2000)
Brachypodium	271	1	5	1,625	5.9	IBI (2010)
Rice	372	6	12	203	0.5	IRGSP (2005)
Physcomitrella	462	1.6	506	14,910	32.2	Rensing et al. (2008)
Poplar	405	2	236	13,341	32.9	Tuskan et al. (2006)
Sorghum	659	1	10	6,907	10.4	Paterson et al. (2009)
Grapevine	342	1	32	165,717	25.7	Jaillon et al. (2007)
Maize	2,066	5b.60	11	125,338	60.8	Schnable et al. (2009)
Wheat 3B	14	1	10	282	19.8	Choulet et al. (2010)

[a]Scaffolds or supercontigs larger than 100 kb
[b]Total number of gaps longer than 5 bp

The availability of complete sequences from several grass genomes has allowed genome-wide studies on genome structure. Gene content can be compared and analyzed and the repetitive part of the genomes can be analyzed in great detail. This has allowed us to get a deep insight into the structure of grass genomes. In addition, this knowledge about genomes has allowed us to develop approaches for comparative and evolutionary genomics. This has resulted in new insight into the evolution of plant genomes, both concerning genes and well as repetitive elements. For example, the comparison of the *Brachypodium* and rice genomes revealed a model for chromosome fusion which could explain variation in chromosome numbers in the family of grasses (Paterson et al. 2009; IBI 2010). Thus, comparative and evolutionary analysis of genomes can result in new and exciting insights into genome structure and its evolution. In this chapter, we will focus on the role of transposons or repetitive elements on genome structure and evolution, a particularly active research field which is profiting from genome-wide analysis.

6.2 Comparative and Evolutionary Genomics in Grasses: The Early Studies

In the mid 1980s, restriction fragment length polymorphism (RFLP) markers were developed for applications in plant breeding and genetic research (Gale and Devos 1998). This resulted in the first genetic maps of cereal crop species. The potential of RFLP probes to hybridize to highly similar, but not perfectly identical sequences and lack of abundance of available markers at that time, stimulated the use of probes from one species for genetic studies in related species. Colinearity across genomes was first reported in the late 1980s between tomato and potato (Bonierbale et al. 1988) and between the three diploid genomes of hexaploid wheat (Chao et al. 1988). Soon after, RFLP-based genetic maps were developed for homoeologous chromosomes of group 7 of bread wheat (*Triticum aestivum*), revealing a high colinearity of marker order between them (Chao et al. 1989). This early work was followed by a number of studies using RFLP markers to establish complete maps of the wheat genome. The first consensus map in the grasses, known today as the 'crop circle', was published in 1995 by Moore et al. (1995a, b), providing the foundation of much of the later research, elaborating and refining the concept and establishing the grasses as a single genetic system. These early studies also revealed some rearrangements between similar genomes, starting the highly productive field of evolutionary genomics. In one such study, cross–hybridization of RFLP markers derived from bread wheat with rye (*Secale cereale*) and barley revealed evidence for a few translocations of chromosome arms in the rye genome if compared to the wheat genomes, while most probes showed that the order of the loci was conserved between those three species (Devos et al. 1993; Moore et al. 1995b).

Investigating the genomic relationships of wheat in maize and rice, Moore et al. (1995a) showed that, despite the divergence of those species ~60–70 million years

ago (MYA) and their massive differences in genome size, the gene order was still conserved along large stretches of the chromosomes. Assuming that the colinearity between rice and wheat is preserved, the genetic map of rice, the smallest grass genome known at that time, was divided into linkage groups and aligned against the genetic maps of wheat and maize. Indeed, it was possible to reconstruct the wheat and maize genome with the rice linkage groups (Moore et al. 1995a). This approach was extended to sugarcane and foxtail millet and led to the first version of the crop circle mentioned above, which has been updated and expanded later (Devos 2005; Salse and Feuillet 2011). The crop circle indicates that the grasses diverged from a common ancestor and that the gene order seems to be well conserved during evolution even after millions of years, despite chromosomal reorganization and remarkable changes in genome sizes.

However, due to the use of only relatively few DNA probes, the genetic resolution of the original crop circle was quite low and did not necessarily reflect the situation at the genomic level. The advances in sequence technology and the subsequent drop in costs created a vast amount of sequence information which offered a unique opportunity to investigate colinearity at the molecular level. In fact, already the first studies of genomic colinearity at the sequence level revealed various exceptions, demonstrating that genes were not always found at the expected position and, therefore, the hypothesis of gene movement was formulated (Gallego et al. 1998; Guyot et al. 2004).

The further comparative analysis of grass genomes revealed many surprising insights into genome evolution. For example, it was found the intergenic regions diverge completely within a few million years. Only in case of very recent evolutionary divergence, both genes and intergenic regions are still conserved. The finding that the intergenic space is changing at a faster pace than the genic space can easily be explained by the lower evolutionary pressure for conservation compared to the genes (Petrov 2001). Therefore, insertions by transposable elements (TE) or deletions caused by illegitimate recombination or unequal crossing over drive the fast turnover of intergenic sequences (Devos et al. 2002; Wicker et al. 2003). These first discoveries laid the foundation for much of the later work in genome-wide analysis described below.

6.3 Discovery of Transposons in Plants: Selfish DNA and Beyond

The research on transposable elements (TEs) in plant genomes has two different historical origins, each with independent lines of research. Their findings converged only relatively recently. First, in her pioneering work Barbara McClintock discovered the existence of jumping genes based on the careful analysis of several biological phenomena observed in maize genetic studies. B. McClintock suggested that genetic factors can move in the genome, thereby modifying gene expression and contributing to genome evolution (McClintock 1950 and summarized

in McClintock 1984). At that time, the molecular basis of the proposed mobile genetic elements remained unknown. It was later found that DNA transposons caused the observed effects (e.g. the Activator (*Ac*)/Dissociator (*Ds*) elements) (Fedoroff et al. 1983). Second, in an independent line of research, the analysis of complete genomes resulted in the discovery that some plant genomes have very high contents of repetitive DNA. Mostly, such studies were done in the 1970s and 1980s by DNA reassociation experiments (Cot analysis) which are based on DNA hybridization (Britten et al. 1974). The observation of a rapidly annealing fraction of genomic DNA suggested that many plant genomes are highly repetitive (Flavell et al. 1974). Only later it was found that this highly repetitive part of the genome consists mostly of transposon and retrotransposon DNA. Based on the history of this discovery, transposons are frequently also called repetitive elements.

Transposons were identified in many different organisms and in cases such as the *P* and *I* elements in Drosophila, they were found to have dramatic negative effects for the survival of the organism under certain conditions. In general, transposons did not have obvious adaptive value and were described as "selfish" DNA based on their ability to multiply (Doolittle and Sapienza 1980). Indeed, TEs are small genetic units, actual "minimal genomes", which contain exactly enough information to be able to replicate, move around in the genome, or both. They use the DNA replication and translation machinery of their "host" and thrive within the environment of the genome. In their paper, Doolittle and Sapienza (1980) made an argument for the hypothesis that the only function of transposable elements is the survival in the genome. However, they explicitly included the possibility that this "raw material" can have some adaptive value later in evolution. The concept of "selfish" DNA was rapidly adapted by the community. In addition, the term "junk DNA" was introduced, reflecting the idea that what is present in such enormous amounts in the genome without obvious consequences on the phenotype must be useless junk. Of course, this contrasts with the original findings of biological function of transposons and the hypothesis of B. McClintock that transposable element contribute to evolution by stimulating chromosomal and genomic rearrangements, resulting in new configurations of genes and changes in gene expression. Interestingly, and a main topic in this chapter below, it was recently found that transposons are major factors in moving around genes in genomes (Wicker et al. 2010). This is a highly relevant finding for understanding the evolution of plant genomes and fits perfectly with the earlier arguments of McClintock. Thus, the jury is out on the final decision on the role of transposable elements in evolution, i.e. selfish DNA verses adaptive value, and there are some arguments for both hypotheses.

6.4 Genome Size of Plants: Genes and Repetitive Elements

In the 1970s, it was found that eukaryotic genomes show an extreme variation in size (Bennett and Smith 1976). Some studies reported an over 200,000-fold variation in genome size, namely between the Amoeba *Amoeba dubia* which was found to have a genome size of 670,000 Mbp (Gregory 2001) and the 2.9 Mbp genome

of the microsporidium *Encephalitozoon cuniculi* (Biderre et al. 1995; Katinka et al. 2001). Plant genomes in particular show a vast variation in genome sizes, even between very closely related species. Most interestingly, there is almost no correlation between genome size and phylogenetic distance in plants (Fig. 6.1). Among the dicotyledonous plants, A*rabidopsis* has one of the smallest genomes known with only about 120 Mbp (AGI 2000). In contrast, the closely related *Brassica* species which diverged from *Arabidopsis* only 15–20 MYA(Yang et al. 1999) have 5–10 times larger genomes. In monocotyledonous plants, variation is even more extreme: The grasses *Brachypodium dystachion*, rice and sorghum have genome sizes of 273, 389 and 690 Mbp, respectively, considerably larger than the *Arabidopsis* genome but roughly an order of magnitude smaller than the genomes of some agriculturally important grass species such as diploid wheat or maize with haploid genome sizes of 5,700 and 2,500 Mbp, respectively. And even they are still dwarfed by the genomes of some lilies, among them *Fritillaria uva-vulpis* which has a genome size of more than 87,000 Mbp, over 700 times the size of the *Arabidopsis* genome (Leitch et al. 2007). Also among *Dicotyledons*, closely related species often differ dramatically in their genome sizes. Maize and sorghum, for example, diverged only about 12 MYA (Swigonová et al. 2004), but the maize genome is more than 4 times the size of the sorghum genome (Tables 6.1 and 6.2).

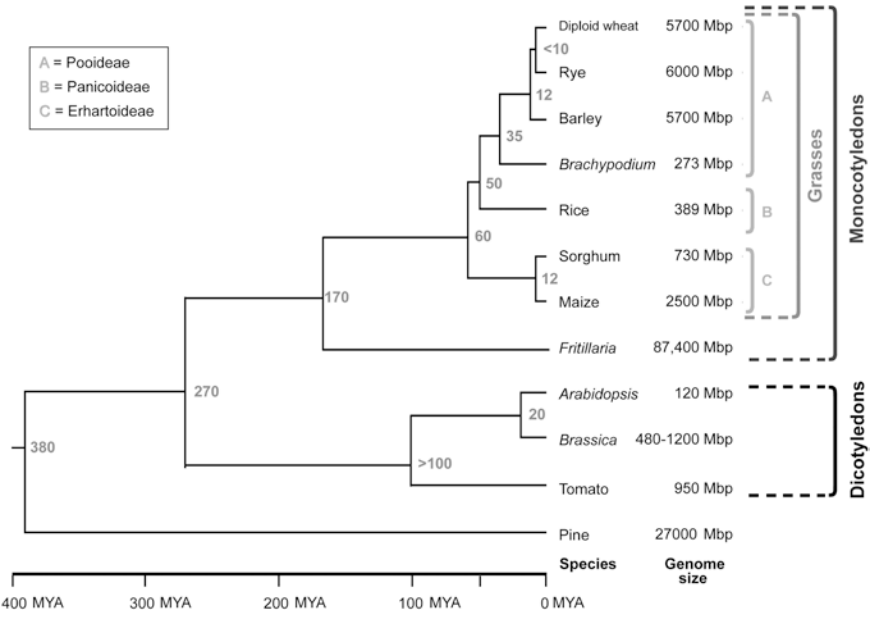

Fig. 6.1 Phylogenetic relationships and genome sizes in selected plant species. Divergence times of specific clades are indicated in *grey* numbers next to the corresponding branching. These numbers are averages of the published values provided in Table 6.1. The scale at the *bottom* indicates divergence times in million years ago (*MYA*). Major taxonomic groups that are discussed in the text are indicated at the *left*

Table 6.2 Plant genome sizes and gene numbers in a selection of publicly available genomes

Plant genomes	Size (Mbp)	Genes	Reference
Arabidopsis thaliana	120	26,200	AGI (2000)
Brachypodium distachyon	273	25,500	IBI (2010)
Fritillaria uva-vulpis	87,400	?	Leitch et al. (2007)
Hordeum vulgare	5,700	32,000	Mayer et al. (2011)
Oryza sativa	372	40,600	IRGSP (2005)
Physcomitrella patens	462	35,900	Rensing et al. (2008)
Populus trichocarpa	410	45,500	Tuskan et al. (2006)
Sorghum bicolor	659	34,500	Paterson et al. (2009)
Triticum aestivum	16,000	50,000	Choulet et al. (2010)
Vitis vinifera	342	30,400	Jaillon et al. (2007)
Zea mays	2,061	30,000	Schnable et al. (2009)

Abbreviations in references: *AGI Arabidopsis* Genome Initiative, *CSC C. elegans* Sequencing Consortium, *IBI* International *Brachypodium* Initiative, *ICGSC* International Chicken Genome Sequencing Consortium, *IHGSC* International Human Genome Sequencing Consortium, *IRGSP* International Rice Genome Sequencing Consortium, *MGSC* Mouse Genome Sequencing Consortium

Despite the vast differences in genomes sizes among plants, the number of genes is almost similar in all species investigated so far. In fact, in recent years, a consensus began to transpire that probably all angiosperm plants contain between 25,000 and 30,000 genes per haploid genome equivalent. This includes only protein-coding genes and excludes other components of gene space such as the highly repetitive ribosomal DNA clusters, tRNAs and small nucleolar and small interfering RNAs as well as conserved non-coding sequences (Freeling and Subramaniam 2009). However, the discussion about the actual gene number of plant genomes is far from over because of the technical difficulties of reliably predicting genes and the mere challenge of defining what a gene actually is.

6.5 Transposable Elements Determine Genome Size

The differences in genome sizes are caused by variation in the number and size of TEs. Especially in large genomes like barley, wheat or maize, TE contribute at least 80 % to the total genomic DNA (Schnable et al. 2009; Wicker et al. 2009b). Already early on, it became clear that there must be hundreds or even thousands of different TE families populating these large genomes (SanMiguel et al. 1998; Wicker et al. 2001). Thus, it has become an important research area to categorise and characterise at least the most abundant TE families in the different plant species. This is necessary for two practical reasons: first, TEs display such an enormous variety that some more exotic ones are often mistaken for genes and annotated as such and, second, transposable elements can cause problems during

sequencing, especially in large genomes. Good knowledge of TEs can therefore help order sequence fragments and close sequence gaps.

Although the necessity and practical value of databases of repetitive elements are apparent to researchers, in recent years, classification and characterisation of these repeats was done very much on a species-by-species level and no common guidelines and classification systems were ever consistently applied. In 2002, Jorge Dubcovsky (UC Davis, CA, USA), David Matthews (Cornell University, NY, USA) and Thomas Wicker (University of Zurich, Switzerland) initiated the first database for TE sequences from *Triticeae* (TREP, *Triticeae* repeat database). TREP originally included only sequences from wheat and barley (Wicker et al. 2002), but has been expanded to include other species since then. The 11th release of TREP contained over 1,500 DNA sequences of plant TEs plus 291 predicted TE protein sequences. Databases for TEs from *A. thaliana* (tigr.org) and rice (retroryza.org) have also become publicly available later.

In 2007, a group of TE experts met at the Plant and Animal Genome Conference in San Diego (California, USA) with the goal to define a broad consensus for the classification of all eukaryotic transposable elements. This included the definition of consistent criteria in the characterisation of the main superfamilies and families and a proposal for a naming system (Wicker et al. 2007a). The proposed system is a consensus of a previous TE classification system that groups all TEs into two major classes, 9 orders and 29 superfamilies (Fig. 6.2). Class 1 contains all TEs which replicate via an mRNA intermediate in a "copy-and-paste" process, while in Class 2 elements, the DNA itself is moved analogous to a "cut-and-paste" process. One novel aspect of the classification system is that the TE family name should be preceded by a three-letter code for class, order and superfamily (Fig. 6.2). This allows to immediately recognise the classification when seeing the name of TE. The proposed classification system is open to expansion as new types of TEs might still be identified in the future.

6.6 TE-Driven Genome Expansion

The most abundant TE class in plant genomes are long terminal repeat (LTR) retrotransposons. They replicate via an mRNA intermediate which is reverse transcribed and integrated elsewhere in the genome. Thus, each replication cycle creates a new copy of the element. Most of the probably hundreds of LTR retrotransposon families in a genome are present in low or moderate copy numbers. However, especially the large plant genomes contain retrotransposon families which are extremely successful colonisers. For example, *BARE1* elements contribute more than 10 % to the barley genome (Vicient et al. 1999; Kalendar et al. 2000; Soleimani et al. 2006). A whole genome survey in barley showed that 50 % of its genome is made up by only 14 TE families and 12 of them are LTR retrotransposons (Wicker et al. 2009b). It is not known what makes certain LTR retrotransposon families particularly successful. Some retrotransposons have been shown to be activated by stress conditions

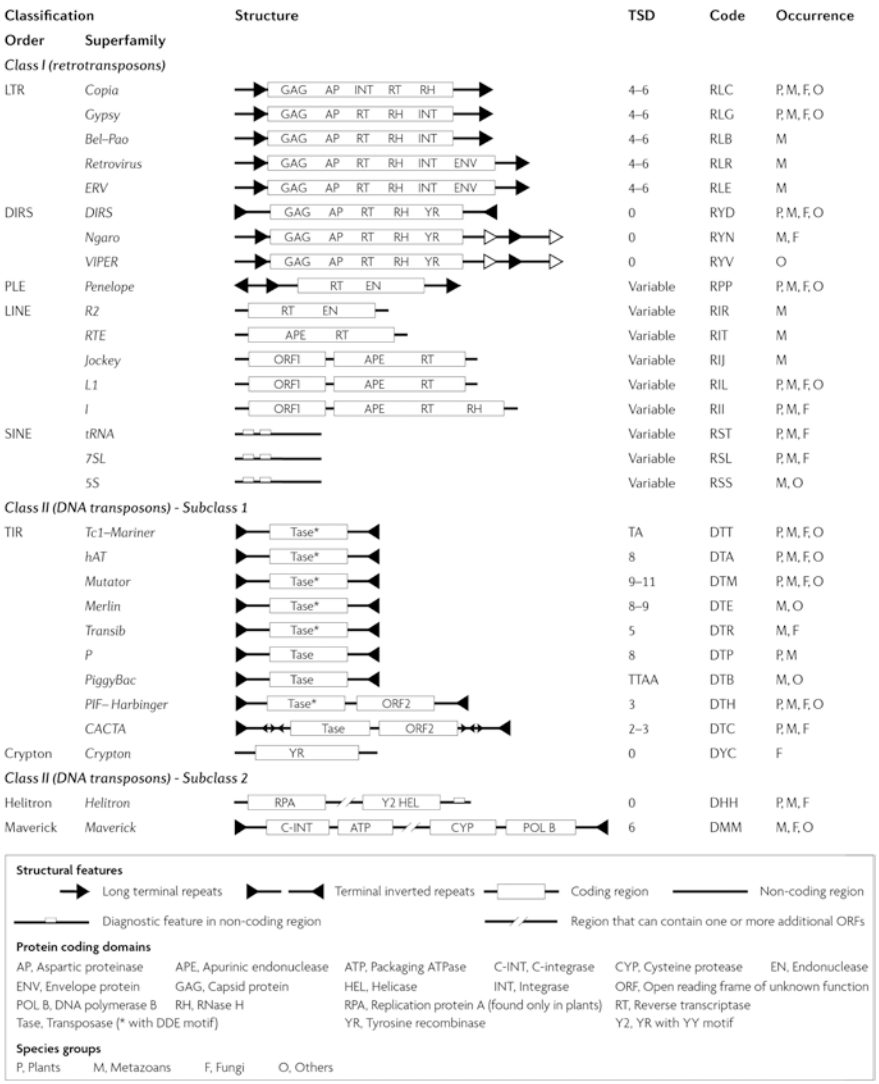

Fig. 6.2 Classification system for transposable elements (Wicker et al. 2007a). The classification is hierarchical and divides TEs into two main classes on the basis of the presence or absence of RNA as a transposition intermediate. They are further subdivided into subclasses, orders and superfamilies. The size of the target site duplication (*TSD*), which is characteristic for most superfamilies, can be used as a diagnostic feature. To facilitate identification, we propose a three-letter code that describes all major groups and that is added to the family name of each TE. *DIRS* Dictyostelium intermediate repeat sequence, *LINE* long interspersed nuclear element, *LTR* long terminal repeat, *PLE* Penelope-like elements, *SINE* short interspersed nuclear element, *TIR* terminal inverted repeat

such as drought (Kalendar et al. 2000). Additionally, analysis of *Copia* elements in rice and wheat showed that different families are active at different times in waves lasting for several hundreds of thousands of years (Wicker and Keller 2007).

In any case, the activity of LTR retrotransposons causes an increase in genome size. Indeed, it was shown that genome size in plants is largely determined by the amount of LTR retrotransposons, while all other TE superfamilies contribute only few percent to the total genomic DNA (Paterson et al. 2009; Schnable et al. 2009; Wicker et al. 2009b; IBI 2010). In large genomes, TEs often insert into one another, leading to complex nesting patterns with large regions that consist exclusively of TE sequences. This is illustrated in Fig. 6.3 which shows how the *rym4* locus in the barley variety *Morex* expanded to more than 65 kb by a series of TE insertions compared with the same locus in the variety *Haruna nijo*. The strong differences between the two varieties indicate that the two loci represent two ancient haplotypes which diverged approximately 930,000 years ago (Wicker et al. 2009a). These data illustrate that TE insertions can greatly expand intergenic regions within relatively short evolutionary time periods. Extensive regions consisting of nested TEs are a typical characteristic of large plant genomes, and they define the image of small gene islands being lost in an ocean of repetitive DNA.

6.7 Genome Contraction Through Deletion of Repetitive DNA

In the early 2000s, the discovery of genome expansion through TE replication led to the perception that plant genomes have "a one-way ticket to genomic obesity" (Bennetzen and Kellogg 1997). Indeed, the existence of plant genomes of several hundred times the size of the *Arabidopsis* genome suggested a one-way process. However, the model of ever-expanding genomes through TE activity could not explain why the genomes of some plant species would suddenly start to grow while others stayed small and compact. Neither could it explain the outright contradictions between taxonomy and genomes sizes (Fig. 6.1). For example, *Brachypodium* with its small genome lies in between two taxonomic groups with significantly larger genomes (*Triticeae* and *Panicoideae*, Fig. 6.1). The model of a one way-process could only explain this pattern if genome expansion had started in *Triticeae* and *Panicoideae* independently only after the three taxa had diverged.

Furthermore, comparative analysis of orthologous regions from barley and wheat revealed virtually no conservation of intergenic sequences (SanMiguel et al. 2002). Genes were found in the same linear order, while no TE was found to be conserved in both species in orthologous positions, i.e. TEs that have inserted in the common ancestor of wheat and barley (Fig. 6.4). Considering that wheat and barely have very similar genome sizes and diverged only about 12 MYA (Chalupska et al. 2008), this finding was surprising and could be best explained if there were mechanisms by which DNA could be removed from the genome.

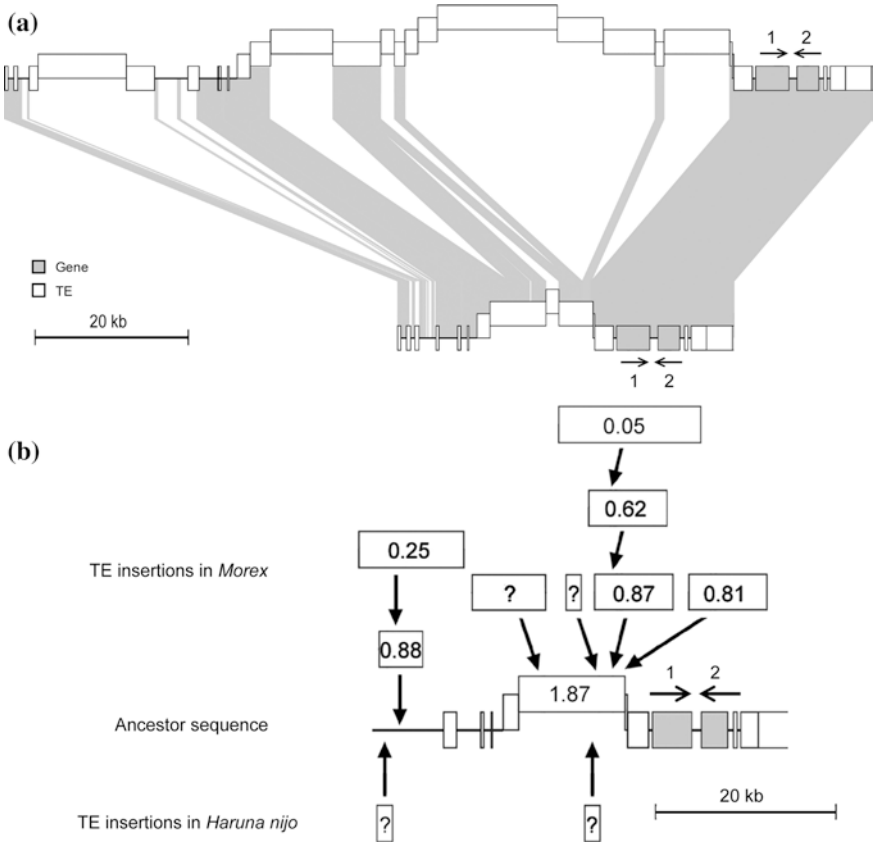

Fig. 6.3 Genome expansion through TE insertions. **a** Comparison of the *rym4* locus from the barley varieties *Morex* (*top*) and *Haruna nijo* (*bottom*). Two genes (#1 and #2) are conserved while intergenic regions differ strongly. The *Morex* locus is greatly expanded due to several TE insertions. Nested insertions of TEs are depicted as follows: TEs that have inserted into others are raised above the ones into which they have inserted. Regions that are conserved between the haplotypes of the two varieties are indicated with grey areas connecting the two maps. **b** Model for the evolution of the *rym4* locus in barley. The map depicts the sequence organization of the hypothetical ancestor sequence. Transposable elements that have subsequently inserted in *Morex* (*top*) and *Haruna nijo* (*bottom*) are indicated as colored boxes, with arrows pointing to their insertion sites. Estimated times of insertions in millions of years ago (*MYA*) are indicated inside the elements. Adapted from Wicker et al. (2009a)

One mechanism how TE sequences can be deleted from the genome is through unequal homologous recombination between the LTRs of retrotransposons. This leads to the generation of a solo-LTR while the internal domain and one equivalent of an LTR is eliminated from the genome (Fig. 6.5a). This phenomenon was long known in animals (who have only relatively few LTR retrotransposons) and was first described as a possible mechanism of genome size reduction in plants (Shirasu et al. 2000). This mechanism also provides an elegant explanation how large parts of retrotransposons

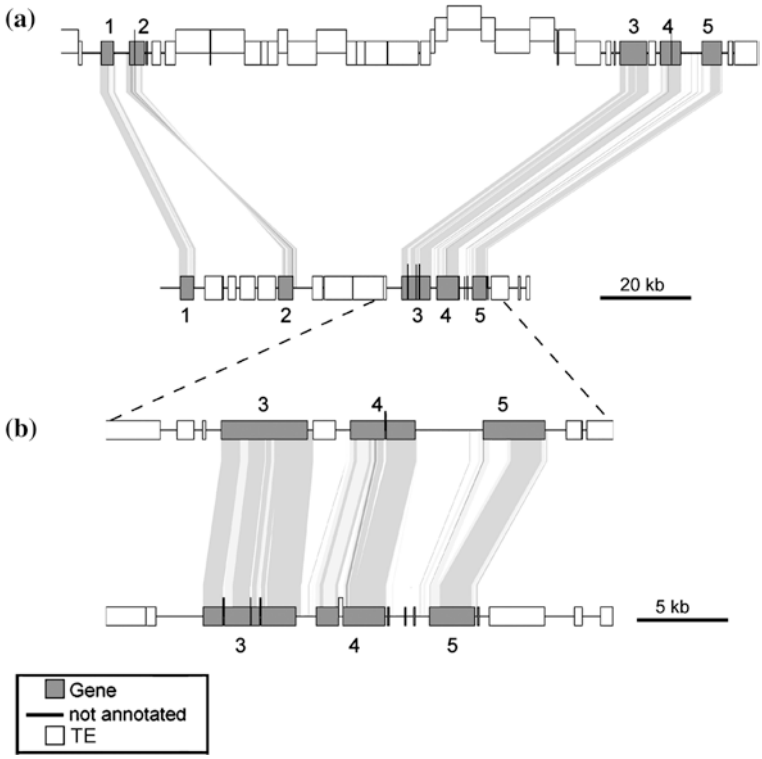

Fig. 6.4 Comparison of orthologous loci from diploid wheat *Triticum monococcum* and barley. **a** Comparison of complete BAC sequences with *T. monococcum* at the *top* and barley at the *bottom*. Orthologous regions conserved in both species are connected by shaded areas. Note that genes are conserved while intergenic regions are completely different due to genomic turnover caused by TE insertions and deletions of DNA. The only difference in the genes is an inversion of the gene #2. Nested insertions of TEs are depicted as described in Fig. 6.3. **b** Detail view of the gene island containing genes 3, 4 and 5. Almost exclusively coding sequences of genes are conserved while promoters and downstream regions have diverged to a degree that they can hardly be aligned. Based on SanMiguel et al. (2002)

can be eliminated from the genome. However, the resulting solo-LTRs still mean a net increase in genome size. The formation of solo-LTRs can therefore not explain the complete absence of colinearity in intergenic regions.

The discovery of apparently "random deletions" in the large intergenic sequences suggested a new mechanism by which repetitive DNA is eliminated independent of its sequence (Wicker et al. 2001). The mechanism that causes this kind of deletions was later described as "illegitimate recombination" (Devos et al. 2002). The term illegitimate recombination includes multiple molecular mechanisms such as replication slippage (reviewed by Lovett (2004)) or double strand break (DSB) repair through non-homologous end-joining (reviewed by van Rijk and Bloemendal (2003)) or single-strand annealing (SSA, reviewed by Hartlerode

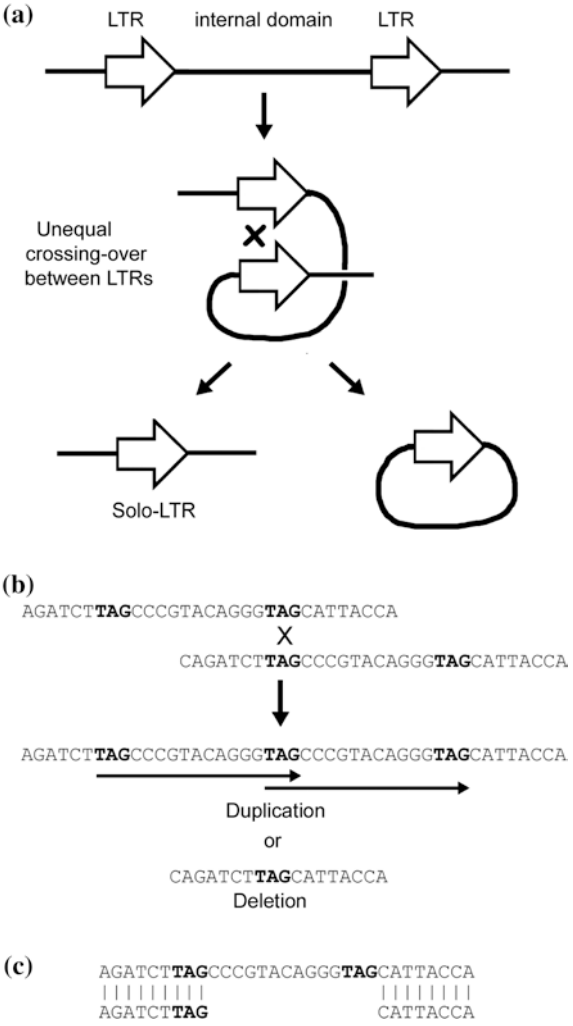

Fig. 6.5 Mechanisms for reduction of genome size. **a** Unequal homologous recombination can occur between the two LTRs of a retrotransposon. It results in a solo-LTR and a circular molecule which is then degraded. **b** Schematic depiction of illegitimate recombination. A short motif of only 3 bp that occurs twice by chance in a short stretch serves as a template for the recombination event. The two products of the recombination are a duplication and a deletion. In the case of the duplication, both units are flanked by the 3 bp motif that served as template. In that case, the 3 bp motif is referred to as the illegitimate recombination signature. **c** A deletion can only be detected if a sequence from an orthologous or paralogous locus is available. The typical illegitimate recombination signature (i.e. the sequences that served as templates for the recombination event) is printed in bold

and Scully (2009)). Whatever the precise molecular mechanisms are, they all result in recombination between very short stretches of homology (e.g. a single or a few bp), thus leading to the apparent random nature and distribution of illegitimate recombination events (Fig. 6.5b).

Studies in *Arabidopsis* (Devos et al. 2002), wheat (Wicker et al. 2003) and rice (Bennetzen et al. 2004) showed that illegitimate recombination is a major mechanism for genome contraction and might have a larger effect on genome size than the generation of solo-LTRs. As shown in Fig. 6.5b, illegitimate recombination leads either to a deletion or a duplication. However, our recent data indicate that deletions caused by DSB repair through SSA probably strongly outnumber duplications (Buchmann et al. 2012). Nevertheless, such duplications can sometimes

act as a creative force, for example in the generation of sequence variability of NBS-LRR resistance gene analogs (Wicker et al. 2007b). There, they can trigger the expansion of the leucine-rich repeat (LRR) domain which is responsible for the specific recognition of pathogens.

The processes of TE amplification and DNA removal drive a "genomic turnover" which is characterized by a balance between TE-driven duplication of DNA and deletions. This results in a permanent reshuffling of all intergenic sequences. Obviously, any alterations in the sequences essential for the immediate survival of the organism will be selected against. For example, if an important gene is disrupted by a TE insertion or partially deleted by illegitimate recombination, the offspring of that cell is not viable. However, parts of the genome which are not under selection pressure, namely TE sequences, can accumulate such rearrangements without negative effects on the fitness of the organism. Apparently, in plants this process is quite rapid and dynamic. As described above, between barley and wheat, intergenic sequences are completely reshuffled within a few million years. In fact, even between different *Triticum* species, only very limited conservation of intergenic sequences was found, although these species have diverged less than three million years ago (Wicker et al. 2003).

A first step toward unravelling the rate at which genomic turnover occurs was the introduction of a method for estimating the age of retrotransposons based on the divergence of their LTRs (SanMiguel et al. 1998). Because of the mechanism of reverse transcription, both LTRs are identical at the time of insertion (Lewin 1997). Since retrotransposons are largely free from selection pressure (Petrov 2001), they accumulate mutations at a background rate which was estimated to be 1.3×10^{-8} substitutions per site per year (Ma and Bennetzen 2004). Thus, the number of differences between the two LTR sequences is proportional to their age.

Numerous surveys have since studied age distributions of LTR retrotransposons in several species, including *Arabidopsis* (Pereira 2004), rice (Gao et al. 2004; Ma et al. 2004; Piegu et al. 2006; Wicker and Keller 2007), wheat (SanMiguel et al. 2002; Wicker and Keller 2007), maize (Du et al. 2006; Wolfgruber et al. 2009) and sorghum (Du et al. 2006). The finding common to all these studies was that hardly any retrotransposons older than 6–7 million years were found, indicating that the removal of repetitive sequences in plant genomes is rather efficient.

Genome-wide surveys of LTR retrotransposon age distributions showed that most retrotransposons are young and the older they get, the rarer they are. This finding suggested that intergenic sequences might be removed at a more or less constant rate from genomes. In fact, in rice and *Arabidopsis*, age distribution of *Copia* retrotransposons approximately follows a hyperbolic distribution, allowing to postulate a "half-life" value that describes the time it takes until half of the retrotransposons are at least partially removed from the genome (i.e. at least one LTR is deleted so that the time of insertion of the element can not be estimated anymore). Interestingly, this value was estimated to be 470,000 years in *Arabidopsis* (Pereira 2004) and 790,000 years in rice (Wicker and Keller 2007). This is consistent with the fact that rice has a significantly larger genome than *Arabidopsis*.

6.8 Dynamic Equilibrium of Genome Size

The findings on genomic turnover led to the emergence of the "increase/decrease" model for genome size evolution (Vitte and Panaud 2005) which describes genome size as a function of the rate of DNA increase through TE amplification and DNA decrease through TE removal. The balance of these two rates determines the current genome size. A change in one or both rates can therefore lead to an increase or decrease of genome size.

The rapid turnover of intergenic sequences makes them a perfect chronicle to study the background processes that shape genomes over time. Especially in large genomes like that of wheat or barley, the study of intergenic sequences allows detailed reconstruction of past events and gives an insight of the mechanisms at work. In the case of the *rym4* locus, evolutionary events could be traced back for approximately 7 million years (Wicker et al. 2005) and revealed a turbulent mixture of insertions, deletions and duplications. However, reconstruction of evolutionary events has its limitations due to high rate of genomic turnover. As mentioned above, within a few million years, intergenic sequences are completely reshuffled. Thus, detailed reconstruction of evolutionary events is not possible past that time frame.

6.9 The Molecular Basis for Gene Movement in Grass Genomes

As described above, the genomic turnover removes almost all sequence similarities outside of protein coding regions within a few million years. Therefore, among more distantly related species, sequence conservation is limited to regions which are under selection. We also discussed above that among grasses, many genetic markers are found in the same order in different species, reflecting the conserved chromosome structure of a common ancestor (Gale and Devos 1998). At the DNA sequence level, one finds that the majority of genes are in the same linear order across species, a finding that is commonly referred to as "synteny" or "colinearity". Comparison of the complete genomes of *Brachypodium*, rice and sorghum showed that at least 60–70 % of all genes are found in the same order in the three species (IBI 2010). This allows us to identify corresponding chromosomal regions with relative ease by comparing the positions of orthologous genes.

The more distantly related two species are, the fewer genes are found in colinear positions. This decrease of the number of collinear genes between species is generally attributed to "gene movement". The molecular mechanism of gene movement and the erosion of synteny that goes with it has been an unsolved riddle since the advent of comparative genomics. Several studies have shown that genes or gene fragments are sometimes captured by TEs and moved or copied to a different location (Wicker et al. 2003; Jiang et al. 2004; Lai et al. 2005; Morgante et al.

2005; Paterson et al. 2009). However, most of these captured pieces of DNA are very small and usually only contain fragments of genes. Thus, none of these studies could so far provide a robust explanation for the movement of large fragments which sometimes contain several genes.

A recent study investigated the molecular basis of the movement of large gene-containing fragments (Wicker et al. 2010). Three-way comparison of the genomes of *Brachypodium*, rice and sorghum identified genes which are specifically non-colinear in only one species (i.e. one could identify in which species the movement had occurred). This approach revealed evidence that gene movement is

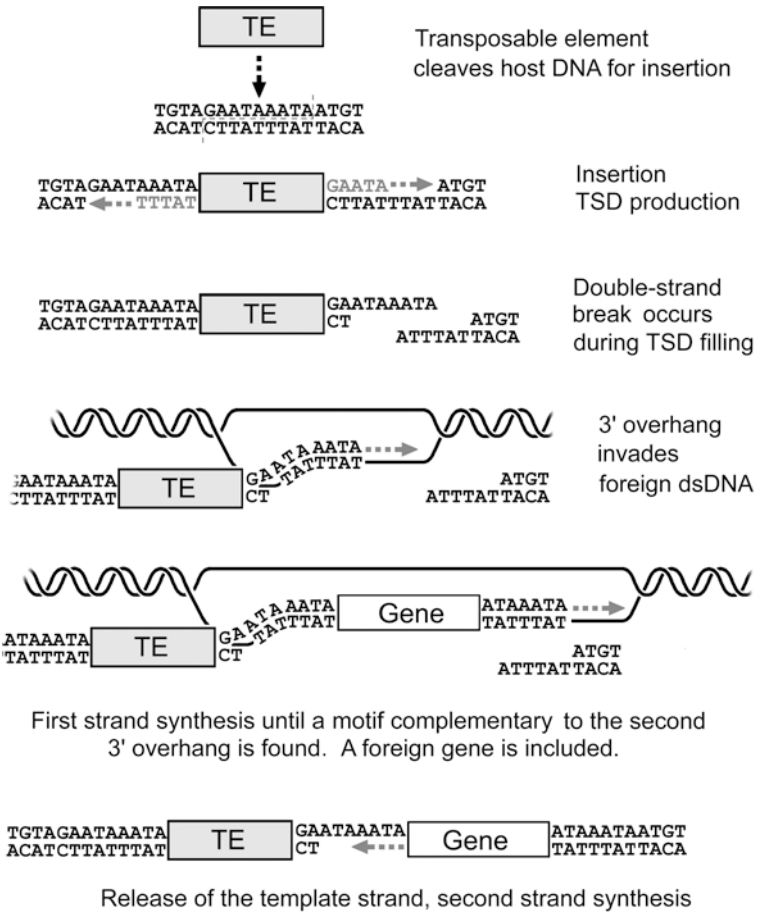

Fig. 6.6 Model for molecular events that lead to a duplication of a foreign gene. A DSB is introduced after the insertion of e *Mutator* element in the genome. A sequence fragment from elsewhere in the genome containing foreign gene is used as filler DNA to repair the DSB (adapted from Wicker et al. (2010))

mostly the result of double-strand breaks (DSBs) that are repaired by the "synthesis-dependent strand annealing" mechanism where a copy of the foreign fragment is used as "filler DNA" to repair the DSB (reviewed by Hartlerode and Scully 2009). Thus, gene movement is largely a copy-and-paste process. The duplicated fragments ranged from a few hundred bp to more than 50 kb and sometimes contained multiple genes. Most DSBs were apparently caused by TE insertions because we often found a TE immediately adjacent to the duplicated fragment (Wicker et al. 2010).

In several cases, highly diagnostic sequence motifs such as target site duplications of TEs on both sides of the duplicated fragments were found, strongly supporting the hypothesis that TE elements cause gene movement. A detailed example for the molecular events is provided in Fig. 6.6: In the first step, a *Mutator* element is inserted into the genome. The transposase cutting the host DNA creates the typical 9 bp overhangs bordering the termini of the element. Usually these gaps would be filled by cellular DNA repair enzymes, resulting in the characteristic target site duplication. We assume that during this process, a DSB can occur either precisely at the insertion point or a few bp away from it. The 3′ overhang produced by the transposase invades a complementary motif elsewhere in the genome. A filler strand is synthesized until a second matching motif is reached. The result is that the filler DNA is immediately adjacent to the TE insertion. Apparently, matching motifs of only a few bp in size are sufficient for strand invasion and priming of synthesis (Puchta 2005).

These findings suggest that gene movement is in fact the result of a rather routine process, namely the "patching up" of gaps in the genome. Besides TE insertions, there are several other mechanisms that can induce DSBs in genomes such as template slippage or unequal crossing-over (Wicker et al. 2010). In fact, recent analyses strongly suggest that the excision of TEs might also be a frequent source of DSBs (Buchmann et al. 2012).

The above observations indicate that more or less random fragments are used as filler DNA in "patching up" of gaps. If gene-containing segments are used to patch the gaps, most of these duplicated genes will probably degenerate as they are not under selection. In a few cases, the duplicated genes will gain a new function or be retained through genetic drift and therefore become established in a population.

6.10 Other Contributions of Transposable Elements to Evolution

Recent discoveries have supported the importance of TEs as a major evolutionary force (Biémont and Vieira 2006). There are a number of cases which clearly demonstrate a role of TEs in evolution. Very importantly, they create diversity in gene expression, either as a consequence of direct changes of the genome

sequence or by epigenetic mechanisms. A very good example (although not from a Triticeae species) of a direct change is the determination of grape color in grapevine: there, the insertion of a retrotransposon changes the grape color from blue to white by insertion into the promoter region of a *Myb* transcription factor gene which is involved in the control of anthocyanin production (Kobayashi et al. 2004; Morgante et al. 2007). The excision of this retrotransposon by unequal intrachromosomal recombination between the long terminal repeat sequences (LTR), resulted in one remaining LTR sequence in the *myb* promoter region. This (partial) excision formed a promoter which is only partially active, but the gene expression at a low level restores MYB activity sufficiently to allow the synthesis of some anthocyanins. This results in the production of red grapes, a nice example how retrotransposon activity is related to an economically highly relevant agronomical trait. As mentioned above, transposons can also change gene expression by epigenetic mechanisms. A classical example of such an epigenetic gene regulation based on a transposable element is the coat color of mice controlled by the *agouti* gene (Morgan et al. 1999). Finally, there is the suprising finding that six lineages of the *copia* retrotransposon show a surprising degree of conservation across phylogenetically different species such as rice and Arabidopsis, indicating some type of selection (Wicker and Keller 2007).

In an intriguing recent discovery, it was found that LINE (long interspersed nuclear element) retrotransposon activity is elevated in brain tissue vs. other somatic tissue in humans. The differential transposon activity in cells of the brain results in brain-specific genetic mosaicism. This brain-specific activity of LINE retrotransposons possibly has consequences on gene expression and neuronal function (Muotri et al. 2010; Singer et al. 2010). Whether this individual-specific diversity results in biologically significant traits remains to be determined.

An adaptive value of transposable elements was suggested by the findings in wild barley populations from Israel. There, two populations from very different, but geographically close, microclimates were analyzed for TE insertion patterns and copy number (Kalendar et al. 2000). The two populations were located in the so called evolution canyon in Northern Israel. This ecological site has two slopes which differ sharply in a number of ecologically important aspects. Specifically in wild barley plants harvested from the drier slope of the canyon, the genome was enriched with *BARE* 1 retrotransposons. Based on these data, the authors speculated about a possible adaptive selection for increased genome size caused by retrotransposon activity. There are some other data which can be viewed as being supportive for such a hypothesis, e.g. it was observed that individual families of the *copia* retrotransposon in rice and wheat were active at different time windows during evolution (Wicker and Keller 2007). These spikes of activity are possibly caused by the evolution of aggressively multiplying element followed by the evolution of efficient silencing mechanisms. However, it cannot be excluded that certain transposon families react to specific environmental conditions. Thus, one can speculate that transposon activity might leave a footprint of past environmental conditions in the genome, an intriguing and fascinating aspect of whole genome analysis in plants.

6.11 The Use of Transposons as Tools for Functional Studies

The ability of transposable elements to move to a new location in the genome has made them important tools for the analysis of gene function, particularly in plants. Insertion of a transposable element into a gene will mostly result in inactivation of this particular gene and the obtained insertion mutant can be used for further experimental studies (for a review on insertion mutagenesis in plants see Ramachandran and Sundaresan 2001). Both DNA transposons as well as retrotransposons have been used for insertional mutagenesis in plants in general, and cereals in particular. If the transposon moves quite frequently in a genome, it can create large sets of insertion mutants, ideally allowing identification of a mutant in any desired gene. Transposon insertion mutagenesis has been particularly important in maize and rice, but there is increasing interest in barley also.

As described above, the crop plant maize is at the origin of important transposable elements. The *Ac* transposon and derived *Ds* deletion variants were first isolated at the molecular level by Fedoroff et al. (1983). Based on the *Ac*/*Ds* elements as well as the *MuDr*/*Mu* transposons, saturation mutagenesis was established in maize (Walbot 2000; Fernandes et al. 2004). A large-scale study comparing the insertion patterns of *Ac*/*Ds* and *MuDR*/*Mu* revealed distinct and complementary target site preferences of the two systems (a review on the different systems available in maize was published by Weil and Monde (2007). The available collections of transposon insertion mutants in maize were recently summarized by Balyan et al. (2008).

One of the most studied and used insertional mutagenesis system applied in heterologous plant species lacking efficient endogenous transposons is based on the above mentioned *Ac*/*Ds* elements derived from maize. For instance, very efficient systems for transposon mutagenesis have been developed for rice using modified versions of these *Ac*/*Ds* elements (Qu et al. 2008). In this crop, starting with only 26 primary transformants, a total of 638 stable *Ds* insertions were identified, with a very wide distribution of the inserted sequences over the whole rice genome. Similarly, Kim et al. (2004) have shown the feasibility of using *Ds* elements for the generation of a large number of *Ds* insertion mutants in rice. A very large genetic resource based on *Ds* in rice was recently described in japonica rice cultivar Dongjin. In this study, 115,000 *Ds* insertion lines were produced, making it an excellent source for the identification of mutants (Park et al. 2009). A summary on the rice genetic resources with transposon insertions is found in a recent review (Balyan et al. 2008).

Interestingly, and very much in contrast to many other plant species, no highly active transposons have yet been identified in the economically important Triticeae species which include wheat, rye and barley. Thus, this very relevant tool for functional studies is not available in this important group of crop plants. Therefore, there are considerable efforts to establish a transgenic system based on the *Ac*/*Ds* system mostly in barley, with some first work done also in wheat. Barley can be relatively easily transformed by Agrobacterium transformation and is a diploid species, so that insertion mutants in this crop would be highly informative for

functional analysis for all Triticeae crops. Recently, several groups have described significant progress in establishing a transposon-tagging system in barley. In one study, more than 100 independent *Ds* insertions were identified and mapped. They were well distributed across the whole genome and integrated preferentially into gene-containing regions. These insertions can now be used as launch pads for further saturation of the genome with insertions (Zhao et al. 2006). Similarly, in an independent study a large number of single copy *Ds* insertions were generated and flanking sequences were determined (Singh et al. 2006). High frequencies of secondary and tertiary transpositions were observed, possibly allowing the development of large populations with independent insertions. More recently, Randhawa et al. (2009) have located single-copy *Ds* insertion events in barley by using wheat cytogenetic stocks. They concluded that it might be possible to target all genes by transposon tagging even with low transposition frequency in gene poor regions. The *Ac*/*Ds* system was also used for additional, more specific applications in barley. A gene trap approach was successfully implemented which will allow gene identification by expression studies as well as by forward and reverse genetics (Lazarow and Lütticke 2009). Along a similar line, an activation tagging system was developed in barley, based on a modified *Ds* element fused to the maize ubiquitin promoter (Ayliffe et al. 2007; Ayliffe and Pryor 2009). This system should allow identification of dominant over-expression phenotypes as done in several other plant species.

Large-scale collections of transposon insertion mutants were developed using DNA transposons, mostly the *Ac*/*Ds* and *Mu*/*MuDR* systems described above. However, in rice, a highly efficient approach was used for insertion mutagenesis based on a retrotransposon called *Tos17*. This element has a very low copy number in the rice genome, particularly if compared to other retrotransposon families. For instance, the well-studied cultivar Nipponbare with a completely sequenced genome contains only two copies. *Tos17* is activated specifically by tissue culture which is used to induce new insertions. In contrast to the *Ac*/*Ds* elements which preferentially insert into closely linked DNA, retrotransposons are mostly transposed to unlinked sequences. The molecular basis of this system and its applications for insertion-based mutagenesis in rice have been reviewed (Hirochika 2001; Kumar and Hirochika 2001). A recent summary of the research field and the complete overview on the available resources based on Tos17 insertions are also available in a recent review (Hirochika 2010).

Although *Tos17* is the only retrotransposon that has been used for large scale mutagenesis in plants, there are other retrotransposons in different plant species which are active and cause mutations. For instance, the spontaneous iron-inefficient mutant *fer* in tomato was recently shown to be caused by an insertion of the *Rider* retrotransposon into the first exon of the gene (Cheng et al. 2009). As this mutant was not derived from tissue culture, it must be assumed that it originates from spontaneous transposition in the plant. Interestingly, there is evidence that retrotransposons can also be used in a transgenic form in other species. The *Tnt1* retrotransposon was originally identified in tobacco and was subsequently used in a transgenic form in the heterologous system of lettuce. There, *Tnt1* gets frequently inserted into genes and the insertions were stably inherited (Mazier et al. 2007). As the lettuce genome

is large, it is tempting to speculate that *Tnt1* might also be useful as an active retrotransposon insertion system in the large genomes of barley and wheat. However, to our knowledge this has not yet been tested so far.

6.12 Summary and Outlook

Studying genome evolution is a complex and mostly theoretical field of research, because most theories and models can not be proved experimentally but have to be inferred between sequence and comparative analysis. Nevertheless, as described in this chapter, our understanding of the molecular mechanisms that shape genomes has greatly improved. The current knowledge opens up many new possible areas of research, some of which are outlined here. We have seen that DNA repair is a major driving force for genomic rearrangements, but we are only beginning to understand what is its actual impact on genome evolution. It will be fascinating to further explore the causes of DSBs, the role of TEs in causing DSBs and the various ways in which they are repaired. Quantitative analyses will be necessary to determine the average size ranges of filler DNA and deletions that are introduced during DSB repair. This will allow conclusions on the magnitude of the impact of DSB repair on genome evolution. Of particular interest is the question why gene colinearity erodes much more rapidly in plants than in animals. Are animal genes less likely to be moved because of their much larger size or is their large size an adaptation that prevents their movement? An increasing number of available eucaryotic genome sequences that are becoming publicly available will allow targeted comparative analyses to address these questions.

One of the central themes of this chapter is the role of TEs in genome evolution. It is essential to study further several fundamental aspects of TEs and their interaction with their host genome. Besides being a frequent source of DSBs, different types of TEs seem to be confined to specific “genomic compartments”. For example, miniature inverted-repeat transposable elements (MITEs) are almost exclusively found in or near genes and their sequence composition is very similar to that of con-coding regions of genes (i.e. promoters, introns and downstream regions). It is therefore perceivable that many if not most gene promoters and regulatory sequences are actually derived from such TE sequences. This would provide an elegant explanation for the observation that non-coding parts of genes are almost completely divergent even between closely related plant species.

In contrast to MITEs, some retrotransposon families are specifically and exclusively found in centromeres of grasses. It is suspected that specific protein domains encoded by the retrotransposon are responsible for guiding the insertion of the DNA copy to specific locations in the genome. This suggests that some of these centromeric elements indeed play a vital role in centromere function.

TEs have proven extremely useful as agents for mutagenesis as well as in gene tagging systems. The more we expand our knowledge of TEs, the more we will discover useful properties that can expand our set of molecular tools to investigate

and manipulate target organisms. Most likely, future studies will help to differentiate our general perception that TEs are purely selfish genetic elements but have, at least to some degree, been recruited by the host to fulfil specific functions.

References

AGI (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408:796–815

Ayliffe MA, Pallotta M, Langridge P, Pryor AJ (2007) A barley activation tagging system. Plant Mol Biol 64:329–347

Ayliffe M, Pryor A (2009) Transposon-based activation tagging in cereals. Funct Plant Biol 36:915–921

Balyan H, Sreenivasulu N, Lizarazu OR, Azhaguvel P, Kianian S (2008) Mutagenesis and high throughput functional genomics. In: Sparks DL (ed) Cereal crops: current status. Academic Press, New York, pp 357–414

Bennett MD, Smith JB (1976) Nuclear DNA amounts in angiosperms. Philos Trans R Soc Lond B Biol Sci 274:227–274

Bennetzen JL, Coleman C, Liu R, Ma J, Ramakrishna W (2004) Consistent over-estimation of gene number in complex plant genomes. Curr Opin Plant Biol 7:732–736

Bennetzen JL, Kellogg EA (1997) Do plants have a one-way ticket to genomic obesity? Plant Cell 9:1509–1514

Berkman PJ, Skarshewski A, Lorenc MT, Lai K, Duran C, Ling EYS, Stiller J, Smits L, Imelfort M, Manoli S, McKenzie M, Kubaláková M, Šimková H, Batley J, Fleury D, Doležel J, Edwards D (2011) Sequencing and assembly of low copy and genic regions of isolated Triticum aestivum chromosome arm 7DS. Plant Biotech J 9:768–775

Biderre C, Pagès M, Méténier G, Canning EU, Vivarès CP (1995) Evidence for the smallest nuclear genome (2.9 Mb) in the microsporidium *Encephalitozoon cuniculi*. Mol Biochem Parasitol 74:229–231

Biémont C, Vieira C (2006) Genetics: junk DNA as an evolutionary force. Nature 443:521–524

Bonierbale MW, Plaisted RL, Tanksley SD (1988) RFLP maps based on a common set of clones reveal modes of chromosomal evolution in Potato and Tomato. Genetics 120:1095–1103

Britten RJ, Graham DE, Neufeld BR (1974) Analysis of repeating DNA sequences by reassociation. Method Enzymol 29:363–418

Buchmann JP, Matsumoto T, Stein N, Keller B, Wicker T (2012) Interspecies sequence comparison in Brachypodium reveals how transposon activity corrodes colinearity. Plant J 71(4):550–563

Chain PSG, Grafham DV, Fulton RS, Fitzgerald MG, Hostetler J, Muzny D, Ali J, Birren B, Bruce DC, Buhay C, Cole JR, Ding Y, Dugan S, Field D, Garrity GM, Gibbs R, Graves T, Han CS, Harrison SH, Highlander S, Hugenholtz P, Khouri HM, Kodira CD, Kolker E, Kyrpides NC, Lang D, Lapidus A, Malfatti SA, Markowitz V, Metha T, Nelson KE, Parkhill J, Pitluck S, Qin X, Read TD, Schmutz J, Sozhamannan S, Sterk P, Strausberg RL, Sutton G, Thomson NR, Tiedje JM, Weinstock G, Wollam A, Consortium GS, Detter JC (2009) Genomics. Genome project standards in a new era of sequencing. Science 326:236–237

Chalupska D, Lee HY, Faris JD, Evrard A, Chalhoub B, Haselkorn R, Gornicki P (2008) *Acc* homoeoloci and the evolution of wheat genomes. Proc Natl Acad Sci USA 105:9691–9696

Chao S, Sharp PJ, Gale MD (1988) In: Miller TE, KoebnerK RMD (ed) Proceedings of the 7th international wheat genetics symposium. IPSR, Cambridge Laboratory, Cambridge

Chao S, Sharp PJ, Worland AJ, Warham EJ, Koebner RMD, Gale MD (1989) RFLP-based genetic maps of wheat homoeologous group 7 chromosomes. Theor Appl Genet 78:495–504

Cheng X, Zhang D, Cheng Z, Keller B, Ling H-Q (2009) A new family of *Ty1*-copia-like retrotransposons originated in the tomato genome by a recent horizontal transfer event. Genetics 181:1183–1193

Choulet F, Wicker T, Rustenholz C, Paux E, Salse J, Leroy P, Schlub S, Paslier M-CL, Magdelenat G, Gonthier C, Couloux A, Budak H, Breen J, Pumphrey M, Liu S, Kong X, Jia

J, Gut M, Brunel D, Anderson JA, Gill BS, Appels R, Keller B, Feuillet C (2010) Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. Plant Cell 22:1686–1701

Devos KM (2005) Updating the 'crop circle'. Curr Opin Plant Biol 8:155–162

Devos KM, Atkinson MD, Chinoy CN, Francis HA, Harcourt RL, Koebner RMD, Liu CJ, Masojć P, Xie DX, Gale MD (1993) Chromosomal rearrangements in the rye genome relative to that of wheat. Theor Appl Genet 85:673–680

Devos KM, Brown JKM, Bennetzen JL (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. Genome Res 12:1075–1079

Doolittle WF, Sapienza C (1980) Selfish genes, the phenotype paradigm and genome evolution. Nature 284:601–603

Du C, Swigonová Z, Messing J (2006) Retrotranspositions in orthologous regions of closely related grass species. BMC Evol Biol 6:62

Fedoroff N, Wessler S, Shure M (1983) Isolation of the transposable maize controlling elements *Ac* and *Ds*. Cell 35:235–242

Fernandes J, Dong Q, Schneider B, Morrow DJ, Nan G-L, Brendel V, Walbot V (2004) Genome-wide mutagenesis of Zea mays L. using RescueMu transposons. Genome Biol 5:R82

Flavell RB, Bennett MD, Smith JB, Smith DB (1974) Genome size and the proportion of repeated nucleotide sequence DNA in plants. Biochem Genet 12:257–269

Freeling M, Subramaniam S (2009) Conserved noncoding sequences (CNSs) in higher plants. Curr Opin Plant Biol 12:126–132

Gale MD, Devos KM (1998) Comparative genetics in the grasses. Proc Natl Acad Sci USA 95:1971–1974

Gallego F, Feuillet C, Messmer M, Penger A, Graner A, Yano M, Sasaki T, Keller B (1998) Comparative mapping of the two wheat leaf rust resistance loci *Lr*1 and *Lr*10 in rice and barley. Genome 41:328–336

Gao L, McCarthy EM, Ganko EW, McDonald JF (2004) Evolutionary history of *Oryza sativa* LTR retrotransposons: a preliminary survey of the rice genome sequences. BMC Genomics 5:18

Goff SA, Ricke D, Lan T-H, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S, Colbert M, lin Sun W, Chen L, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu Y, 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

Gregory TR (2001) Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. Biol Rev Camb Philos Soc 76:65–101

Guyot R, Yahiaoui N, Feuillet C, Keller B (2004) In silico comparative analysis reveals a mosaic conservation of genes within a novel colinear region in wheat chromosome 1AS and rice chromosome 5S. Funct Integr Genomics 4:47–58

Hartlerode AJ, Scully R (2009) Mechanisms of double-strand break repair in somatic mammalian cells. Biochem J 423:157–168

Hirochika H (2001) Contribution of the *Tos*17 retrotransposon to rice functional genomics. Curr Opin Plant Biol 4:118–122

Hirochika H (2010) Insertional mutagenesis with *Tos*17 for functional analysis of rice genes. Breeding Sci 60:486–492

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

IRGSP (2005) The map-based sequence of the rice genome. Nature 436:793–800

Jaillon O, Aury J-M, 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, Bruyère C, Billault A, Segurens B, Gouyvenoux M, Ugarte E, Cattonaro F, Anthouard V, Vico V, Fabbro CD, Alaux M, Gaspero GD, Dumas V, Felice N, Paillard S,

Juman I, Moroldo M, Scalabrin S, Canaguier A, Clainche IL, Malacrida G, Durand E, Pesole G, Laucou V, Chatelet P, Merdinoglu D, Delledonne M, Pezzotti M, Lecharny A, Scarpelli C, Artiguenave F, Pè ME, Valle G, Morgante M, Caboche M, Adam-Blondon A-F, Weissenbach J, Quétier F, Wincker P (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463–467

Jiang N, Feschotte C, Zhang X, Wessler SR (2004) Using rice to understand the origin and amplification of miniature inverted repeat transposable elements (MITEs). Curr Opin Plant Biol 7:115–119

Kalendar R, Tanskanen J, Immonen S, Nevo E, Schulman AH (2000) Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. Proc Natl Acad Sci USA 97:6603–6607

Katinka MD, Duprat S, Cornillot E, Méténier G, Thomarat F, Prensier G, Barbe V, Peyretaillade E, Brottier P, Wincker P, Delbac F, Alaoui HE, Peyret P, Saurin W, Gouy M, Weissenbach J, Vivarès CP (2001) Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. Nature 414:450–453

Kim CM, Piao HL, Park SJ, Chon NS, Je BI, Sun B, Park SH, Park JY, Lee EJ, Kim MJ, Chung WS, Lee KH, Lee YS, Lee JJ, Won YJ, Yi G, Nam MH, Cha YS, Yun DW, Eun MY, deok Han C (2004) Rapid, large-scale generation of *Ds* transposant lines and analysis of the *Ds* insertion sites in rice. Plant J 39:252–263

Kobayashi S, Goto-Yamamoto N, Hirochika H (2004) Retrotransposon-induced mutations in grape skin color. Science 304:982

Kumar A, Hirochika H (2001) Applications of retrotransposons as genetic tools in plant biology. Trends Plant Sci 6:127–134

Lai J, Li Y, Messing J, Dooner HK (2005) Gene movement by *Helitron* transposons contributes to the haplotype variability of maize. Proc Natl Acad Sci USA 102:9068–9073

Lazarow K, Lütticke S (2009) An *Ac*/*Ds*-mediated gene trap system for functional genomics in barley. BMC Genomics 10:55

Leitch IJ, Beaulieu JM, Cheung K, Hanson L, Lysak MA, Fay MF (2007) Punctuated genome size evolution in *Liliaceae*. J Evol Biol 20:2296–2308

Lewin B (1997) Genes VI. Oxford University Press, New York

Lovett ST (2004) Encoded errors: mutations and rearrangements mediated by misalignment at repetitive DNA sequences. Mol Microbiol 52:1243–1253

Ma J, Bennetzen JL (2004) Rapid recent growth and divergence of rice nuclear genomes. Proc Natl Acad Sci USA 101:12404–12410

Ma J, Devos KM, Bennetzen JL (2004) Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. Genome Res 14:860–869

Mayer KFX, Martis M, Hedley PE, Simková H, Liu H, Morris JA, Steuernagel B, Taudien S, Roessner S, Gundlach H, Kubaláková M, Suchánková 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. Plant Cell 23:1249–1263

Mazier M, Botton E, Flamain F, Bouchet J-P, Courtial B, Chupeau M-C, Chupeau Y, Maisonneuve B, Lucas H (2007) Successful gene tagging in lettuce using the *Tnt*1 retrotransposon from tobacco. Plant Physiol 144:18–31

McClintock B (1950) The origin and behavior of mutable loci in maize. Proc Natl Acad Sci USA 36:344–355

McClintock B (1984) The significance of responses of the genome to challenge. Science 226:792–801

Moore G, Devos KM, Wang Z, Gale MD (1995a) Cereal genome evolution. Grasses, line up and form a circle. Curr Biol 5:737–739b

Moore G, Foote T, Helentjaris T, Devos K, Kurata N, Gale M (1995b) Was there a single ancestral cereal chromosome? Trends Genet 11:81–82a

Morgan HD, Sutherland HG, Martin DI, Whitelaw E (1999) Epigenetic inheritance at the *agouti* locus in the mouse. Nat Genet 23:314–318

Morgante M, Brunner S, Pea G, Fengler K, Zuccolo A, Rafalski A (2005) Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. Nat Genet 37:997–1002

Morgante M, Paoli ED, Radovic S (2007) Transposable elements and the plant pan-genomes. Curr Opin Plant Biol 10:149–155

Muotri AR, Marchetto MCN, Coufal NG, Oefner R, Yeo G, Nakashima K, Gage FH (2010) L1 retrotransposition in neurons is modulated by *MeCP*2. Nature 468:443–446

Park D-S, Park S-K, Han S-I, Wang H-J, Jun N-S, Manigbas N, Woo Y-M, Ahn B-O, Yun D-W, Yoon U-H, Kim Y-H, Lee M-C, Kim D-H, Nam M-H, Han C-D, Kang H-W, Yi G (2009) Genetic variation through Dissociation (*Ds)* insertional mutagenesis system for rice in Korea: progress and current status. Mol Breeding 24:1–15

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, ur Rahman M, Ware D, Westhoff P, Mayer KFX, Messing J, Rokhsar DS (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556

Pereira V (2004) Insertion bias and purifying selection of retrotransposons in the *Arabidopsis thaliana* genome. Genome Biol 5:R79

Petrov DA (2001) Evolution of genome size: new approaches to an old problem. Trends Genet 17:23–28

Piegu B, Guyot R, Picault N, Roulin A, Saniyal A, Kim H, Collura K, Brar DS, Jackson S, Wing RA, Panaud O (2006) Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. Genome Res 16:1262–1269

Puchta H (2005) The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. J Exp Bot 56:1–14

Qu S, Desai A, Wing R, Sundaresan V (2008) A versatile transposon-based activation tag vector system for functional genomics in cereals and other monocot plants. Plant Physiol 146:189–199

Ramachandran S, Sundaresan V (2001) Transposons as tools for functional genomics. Plant Physiol Biochem 39:243–252

Randhawa HS, Singh J, Lemaux PG, Gill KS (2009) Mapping barley *Ds* insertions using wheat deletion lines reveals high insertion frequencies in gene-rich regions with high to moderate recombination rates. Genome 52:566–575

Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud P-F, Lindquist EA, Kamisugi Y, Tanahashi T, Sakakibara K, Fujita T, Oishi K, Shin-I T, Kuroki Y, Toyoda A, Suzuki Y, Hashimoto S-I, Yamaguchi K, Sugano S, Kohara Y, Fujiyama A, Anterola A, Aoki S, Ashton N, Barbazuk WB, Barker E, Bennetzen JL, Blankenship R, Cho SH, Dutcher SK, Estelle M, Fawcett JA, Gundlach H, Hanada K, Heyl A, Hicks KA, Hughes J, Lohr M, Mayer K, Melkozernov A, Murata T, Nelson DR, Pils B, Prigge M, Reiss B, Renner T, Rombauts S, Rushton PJ, Sanderfoot A, Schween G, Shiu S-H, Stueber K, Theodoulou FL, Tu H, de Peer YV, Verrier PJ, Waters E, Wood A, Yang L, Cove D, Cuming AC, Hasebe M, Lucas S, Mishler BD, Reski R, Grigoriev IV, Quatrano RS, Boore JL (2008) The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. Science 319:64–69

van Rijk A, Bloemendal H (2003) Molecular mechanisms of exon shuffling: illegitimate recombination. Genetica 118:245–249

Salse J, Feuillet C (2011) Palaeogenomics in cereals: modeling of ancestors for modern species improvement. C R Biol 334:205–211

SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL (1998) The paleontology of intergene retrotransposons of maize. Nat Genet 20:43–45

SanMiguel PJ, Ramakrishna W, Bennetzen JL, Busso CS, Dubcovsky J (2002) Transposable elements, genes and recombination in a 215-kb contig from wheat chromosome $5A^m$. Funct Integr Genomics 2:70–80

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, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang X-C, 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 palaeopolyploid soybean. Nature 463:178–183

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, 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 F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen W, 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 R, 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 J, 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 L, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Buren PV, Vaughn MW, Ying K, Yeh C-T, Emrich SJ, Jia Y, Kalyanaraman A, Hsia A-P, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia J-M, Deragon J-M, Estill JC, Fu Y, Jeddeloh JA, Han Y, Lee H, Li P, Lisch DR, Liu S, Liu Z, 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 L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen JL, Dawe RK, Jiang J, 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

Shirasu K, Schulman AH, Lahaye T, Schulze-Lefert P (2000) A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. Genome Res 10:908–915

Singer T, McConnell MJ, Marchetto MCN, Coufal NG, Gage FH (2010) LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes? Trends Neurosci 33:345–354

Singh J, Zhang S, Chen C, Cooper L, Bregitzer P, Sturbaum A, Hayes PM, Lemaux PG (2006) High-frequency Ds remobilization over multiple generations in barley facilitates gene tagging in large genome cereals. Plant Mol Biol 62:937–950

Soleimani VD, Baum BR, Johnson DA (2006) Quantification of the retrotransposon BARE-1 reveals the dynamic nature of the barley genome. Genome 49:389–396

Swigonová Z, Lai J, Ma J, Ramakrishna W, Llaca V, Bennetzen JL, Messing J (2004) Close split of sorghum and maize genome progenitors. Genome Res 14:1916–1923

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen G-L, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Déjardin A, Depamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjärvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leplé J-C, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouzé P, Ryaboy D, Schmutz J,

Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai C-J, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, de Peer YV, Rokhsar D (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313:1596–1604

Vicient CM, Kalendar R, Anamthawat-Jónsson K, Schulman AH (1999) Structure, functionality, and evolution of the BARE-1 retrotransposon of barley. Genetica 107:53–63

Vitte C, Panaud O (2005) LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. Cytogenet Genome Res 110:91–107

Walbot V (2000) Saturation mutagenesis using maize transposons. Curr Opin Plant Biol 3:103–107

Weil CF, Monde R-A (2007) Induced mutations in maize Israel. J Plant Sci 55:183–190

Wicker T, Buchmann JP, Keller B (2010) Patching gaps in plant genomes results in gene movement and erosion of colinearity. Genome Res 20:1229–1237

Wicker T, Keller B (2007) Genome-wide comparative analysis of copia retrotransposons in Triticeae, rice, and Arabidopsis reveals conserved ancient evolutionary lineages and distinct dynamics of individual copia families. Genome Res 17:1072–1081

Wicker T, Krattinger SG, Lagudah ES, Komatsuda T, Pourkheirandish M, Matsumoto T, Cloutier S, Reiser L, Kanamori H, Sato K, Perovic D, Stein N, Keller B (2009a) Analysis of intraspecies diversity in wheat and barley genomes identifies breakpoints of ancient haplotypes and provides insight into the structure of diploid and hexaploid triticeae gene pools. Plant Physiol 149:258–270a

Wicker T, Matthews D, and Keller B (2002) TREP: a database for Triticeae repetitive elements. Trends Plant Sci 7:561–562

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH (2007a) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8:973–982a

Wicker T, Stein N, Albar L, Feuillet C, Schlagenhauf E, Keller B (2001) Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. Plant J 26:307–316

Wicker T, Taudien S, Houben A, Keller B, Graner A, Platzer M, Stein N (2009b) A whole-genome snapshot of 454 sequences exposes the composition of the barley genome and provides evidence for parallel evolution of genome size in wheat and barley. Plant J 59:712–722b

Wicker T, Yahiaoui N, Guyot R, Schlagenhauf E, Liu Z-D, Dubcovsky J, Keller B (2003) Rapid genome divergence at orthologous low molecular weight glutenin loci of the A and A^m genomes of wheat. Plant Cell 15:1186–1197

Wicker T, Yahiaoui N, Keller B (2007b) Illegitimate recombination is a major evolutionary mechanism for initiating size variation in plant resistance genes. Plant J 51:631–641b

Wicker T, Zimmermann W, Perovic D, Paterson AH, Ganal M, Graner A, Stein N (2005) A detailed look at 7 million years of genome evolution in a 439 kb contiguous sequence at the barley *Hv-eIF4E* locus: recombination, rearrangements and repeats. Plant J 41:184–194

Wolfgruber TK, Sharma A, Schneider KL, Albert PS, Koo D-H, Shi J, Gao Z, Han F, Lee H, Xu R, Allison J, Birchler JA, Jiang J, Dawe RK, Presting GG (2009) Maize centromere structure and evolution: sequence analysis of centromeres 2 and 5 reveals dynamic Loci shaped primarily by retrotransposons. PLoS Genet 5:e1000743

Yang YW, Lai KN, Tai PY, Li WH (1999) Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between Brassica and other angiosperm lineages. J Mol Evol 48:597–604

Yu J, Hu S, Wang J, Wong GK-S, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Liu Z, Li L, Liu J, Qi Q, Liu J, Li L, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Zhang J, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Ren X, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Wang J, Zhao

W, Li P, Chen W, Wang X, Zhang Y, Hu J, Wang J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Li G, Liu S, Tao M, Wang J, Zhu L, Yuan L, Yang H (2002) A draft sequence of the rice genome *Oryza sativa*. Science 296:79–92

Zhao T, Palotta M, Langridge P, Prasad M, Graner A, Schulze-Lefert P, Koprek T (2006) Mapped *Ds*/T-DNA launch pads for functional genomics in barley. Plant J 47:811–826