A guided tour of large genome size in animals: what we know and where we are heading

France Dufresne · Nicholas Jeffery

Abstract The study of genome size diversity is an ever-expanding field that is highly relevant in today's world of rapid and efficient DNA sequencing. Animal genome sizes range from 0.02 to 132.83 pg but the majority of animal genomes are small, with the most of these genome sizes being less than 5 pg. Animals with large genomes (>10 pg) are scattered within some invertebrates, including the Platyhelminthes, crustaceans, and orthopterans, and also the vertebrates including the Actinopterygii, Chondrichthyes, and some amphibians. In this paper, we explore the connections between organismal phenotype, physiology, and ecology to genome size. We also discuss some of the molecular mechanisms of genome shrinkage and expansion obtained through comparative studies of species with full genome sequences and how this may apply to species with large genomes. As most animal species sequenced to date have been in the small range for genome size (especially invertebrates) due to sequencing costs and to difficulties associated with large genome assemblies, an understanding of the structural composition of large genomes is still lacking. Studies using next-generation sequencing are being attempted for the first time in animals with larger genomes. Such analyses using low genome coverage are providing a glimpse of the composition of repetitive elements in animals with more complex genomes. These future studies will allow a better understanding of factors leading to genomic obesity in animals.

Keywords Genome size · cell size · large genomes · transposable element · next-generation sequencing

Abbreviations
TE	Transposable elements
NGS	Next generation sequencing
LTR	Long terminal repeat retrotransposon
SINE	Short interspersed repetitive elements
LINE	Long interspersed repetitive elements
BES	BAC end sequences
pg	Picogram

F. Dufresne (✉)
Département de Biologie, Centre d'Études Nordiques,
Université du Québec à Rimouski,
300 allée des Ursulines,
Québec G5L 3A1, Canada
e-mail: france_dufresne@uqar.qc.ca

N. Jeffery
Department of Integrative Biology, University of Guelph,
Guelph, Ontario, N1G 2W1, Canada

Introduction to large genome sizes in animals

Genome size (the haploid DNA content per cell or C-value) estimates are now available for approximately 5000 animal species representing nearly 7,000-fold variation within the animals, which can be found online in the Animal Genome Size Database

(www.genomesize.com). However, the vast majority of these genome size estimates are not considered large based on the definition we provide here (e.g., <10 pg, where 1 pg is equal to 978 Mb) (Fig. 1) (Gregory 2011). In fact, the human genome, currently the largest sequenced animal genome, is only 3.5 pg in mass. Based on the frequency distribution of current data for animals, we classify large genomes as those larger than 10 pg, which is roughly three times the size of the human genome. Species with a genome size greater than 10 pg represent only 8% of all animal species currently in the Animal Genome Size Database (www.genomesize.com) (Fig. 1). While there are relatively few large genomes discovered in animals thus far, it is more than likely that many more large genomes will be revealed across a greater diversity of taxonomic groups as more genome sizes are estimated.

The goals of this review are to discuss the taxonomic distribution of large genomes in animals, the phenotypic and ecological parameters that correlate with increasing genome size, and the mechanisms that increase (or decrease) genome size. We also discuss the potential for sequencing large genomes with next-generation sequencing (NGS) techniques, with particular reference to completed whole-genome sequencing projects. There is little mention of large genomes in animals in the current literature, and so we aim to collate what data are available and provide suggestions for future research in the exciting field of genome size research. Note that in this review, we focus on animal species that have acquired large genomes but not as a result of polyploidy (though it is sometimes difficult to identify ancient polyploids that have diploidized).

Fig. 1 Histogram showing frequency distribution of haploid genome sizes (in picograms) of 4,972 animal species with genomes from 0.02 to 10 pg and from 10 to 132 pg. Compiled using the database of Gregory (2011)
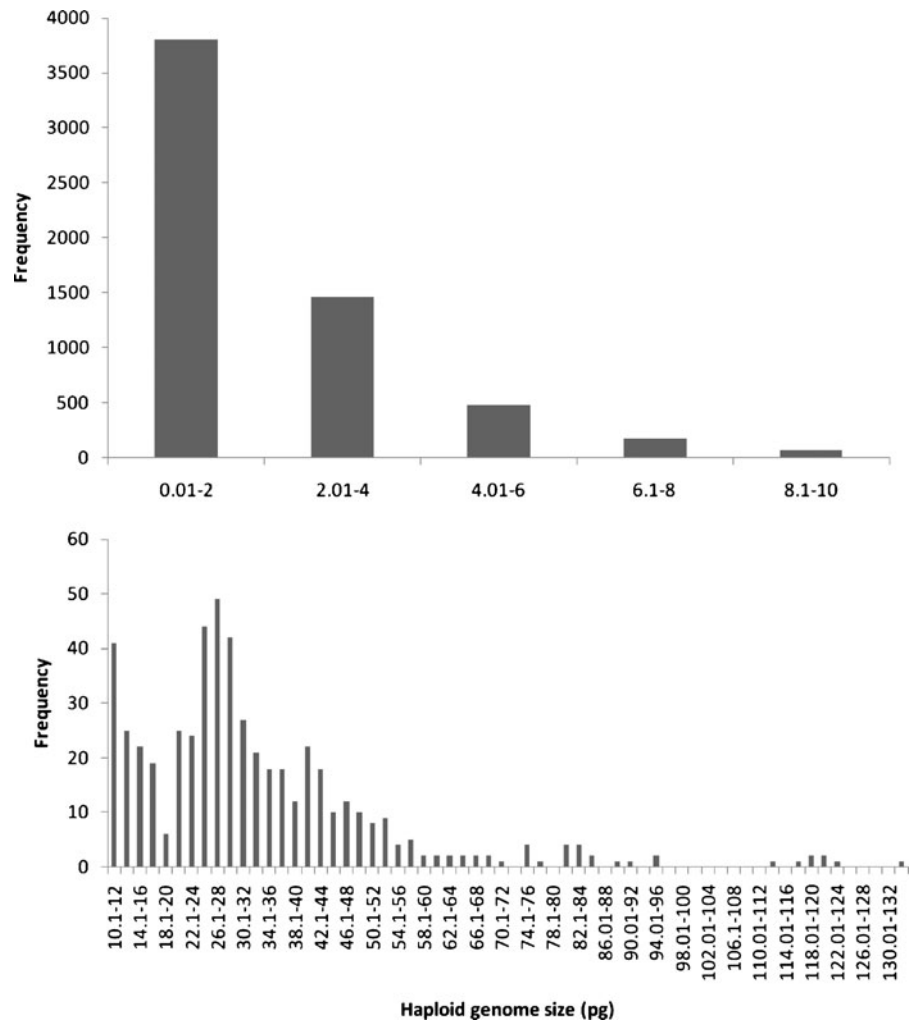
Taxonomic distribution of large genomes in animals

Lungfish

Lungfish are a group of freshwater fish that are able to breathe air and can be extremely long-lived. The largest animal genome resides within the marbled lungfish (*Protopterus aethiopicus*) with a mass of 132.83 pg, measured using Feulgen densitometry by Pedersen (1971). The South American lungfish *Lepidosiren paradoxa* has the second largest lungfish genome examined at roughly 80 pg (Vinogradov 2005). Previously, this species' genome size was estimated to be closer to 120 pg but has since been revised. This may be in part due to the methods involved in measuring these large genomes and the difficulty associated with finding appropriate standards for accurate estimates. There is considerable variation in the estimates for four of the species of lungfish on the Animal Genome Size Database, which again is likely due to the difficulty involved with quantifying the mass of large nuclei due to factors such as reduced stain uptake or finding a standard species with similar DNA compaction levels. However, thus far, the seven species and subspecies of lungfish that have been measured are all greater than 40 pg.

Chondrichthyes

The Chondrichthyes (cartilaginous fishes) are the only other non-tetrapod vertebrates with large genomes. There are 15 species of sharks listed in the Animal Genome Size Database with genome sizes >10 pg, ranging from 10.13 pg in the Portuguese dogfish (Hardie and Hebert 2004) to 17.05 pg in the Angular roughshark (Stingo et al. 1989). The majority of the Chondrichthyes studied thus far, however, have genomes smaller than 10 pg, and so these 15 species represent just a small fraction of the relatively well-studied chondrichthyans.

Amphibians

Close behind the marbled lungfish, the second largest accepted animal genome sizes are found in the salamanders. Specifically, the species *Necturus lewisi* and *Necturus punctatus* have genome sizes of 120 and 119 pg, respectively (Sexsmith 1968; Olmo 1973). Other Urodeles with genome size estimates close to these values include *Necturus maculosus* (estimates range from 80.5 to 95 pg from Licht and Lowcock 1991 and Brown and Dawid 1968, respectively) and *Amphiuma means* with a similar range of estimates (e.g., Olmo 1974). While the majority of urodeles studied have genome sizes greater than 20 pg, only ten species of Anurans (frogs and toads) have a *C*-value greater than 10 pg, and four of these estimates are thought to be from polyploid species (Gregory 2005a, 2011). The third Amphibian order for which there are estimates, the *Gymnophiona* (also known as the caecilians), only has estimates for three species. One species, *Siphanops annulatus*, has a genome size of 13.95 pg and is the only caecilian with a large genome (Beçak et al. 1970). The fishes and the amphibians are the only vertebrates that contain species with large genomes.

Large genomes in invertebrates

Even though the vast majority of animal species that have genome size estimates are vertebrates (3,231 of approximately 4,900 estimates), there are a number of invertebrate phyla that have species with large genomes. Since invertebrates, which make up the majority of animal life on the planet, are vastly underrepresented in terms of genome size estimates, it is highly likely that more species with large genomes will be discovered with more intensive sampling.

Flatworms

The flatworms (Platyhelminthes) are a diverse group of invertebrates for which there are only 68 genome size estimates. Only two species have genomes larger than 10 pg, both of which are free-living species. The largest flatworm genome is found in *Otomesostoma auditivum* at 20.52 pg while the second largest genome is found in *Mesostoma ehrenbergii* at 14.80 pg according to Hebert and Beaton (1990) or 16.45 pg according to Gregory et al. (2000). The remaining species for which there are estimates in this group all have genome sizes below 6 pg.

Arthropods

The only other invertebrate phylum that (currently) contains species with large genomes is the

Arthropoda and, more specifically, the insects and crustaceans within the arthropods. The single insect order that contains species with large genomes is the Orthoptera, which includes the grasshoppers and crickets. There are 19 species of orthopterans with a *C*-value greater than 10 pg, the smallest of which is in the field grasshopper *Chorthippus brunneus* at 10.15 pg (Gosalvez et al. 1980) and the largest of which is in the mountain grasshopper *Podisma pedestris* at 16.93 pg (Westerman et al. 1987). The smallest orthopteran genome overall is 1.55 pg in the cave cricket *Hadenoecus subterraneus* (Rasch and Rasch 1981), which gives nearly 11-fold variation within the Orthoptera.

The Crustacea contains two classes that contain species with large genomes, namely the Maxillopoda and the Malacostraca. However, there are only estimates for species in four of the six classes of crustacean and so this number may expand. The species within the Maxillopoda that contain large genomes are all marine calanoid copepods which range from 9.30 pg in *Calanus glacialis* to 12.46 pg in *Calanus hyperboreus* (Escribano et al. 1992; McLaren et al. 1988). The smallest genome within the Calanoida is 0.63 pg in *Eurytemora affinis*, providing nearly 20-fold variation in this order. The class Malacostraca contains three orders with species with a large genome estimate. There are only two estimates for species in the order Stomatopoda, or mantis shrimps, one of which is considered large at 10.20 pg in *Pseudosquilla ciliata* (Bachmann and Rheinsmith 1973). There are 21 species in the order Decapoda (which includes crabs, shrimps, and lobsters) with large genomes ranging from 10 pg in the Atlantic vent shrimp *Rimicaris exoculata* to 40.89 pg in the polar shrimp *Sclerocrangon ferox*. Interestingly, nearly all decapods with large genomes are found in either the deep ocean (Bonnivard et al. 2009) or in cold, arctic waters (Rees et al. 2008). The third malacostracan order with large genomes is the Amphipoda, for which there are estimates for 27 species but only five of which have large genomes. The Amphipoda contains the two largest invertebrate genome size estimates at 64.62 pg in *Ampelisca macrocephala* and 50.91 pg in *Stegocephalus inflatus* (Rees et al. 2007). Similarly to the decapods, all five of the amphipods with genome sizes greater than 10 pg are found in Arctic species. A number of other species have recently been collected from Hudson Bay in the low

Arctic and have mostly also had genomes larger than 10 pg (e.g., *Atylus carinatus*, *Gammaracanthus loricatus*, *Onisimus litoralis*) (Jeffery, unpublished).

Phenotypic correlates of large genomes

The most universal effect of genome size increase is a concomitant increase in cell size (Gregory 2001). While early genome size biologists were surprised to discover that genome size did not correlate with apparent organismal complexity, positive correlations between erythrocyte size and genome size have long been noted in vertebrates (Mirsky and Ris 1951). Subsequent studies have confirmed this relationship between genome size and cell size in protists (Cavalier-Smith 1985), plants (Grime 1983; Gregory 2005a), invertebrates (Gregory 2005b), amphibians (Olmo and Morescalchi 1975; Horner and Macgregor 1983), reptiles (Olmo and Odierna 1982), fish (Pedersen 1971; Olmo 1983; Hardie and Hebert 2003), birds (Gregory 2002a), and mammals (Gregory 2000). Numerous theories have been invoked to account for the strong correlation between genome size and cell size (Gregory 2005a). Among these, the nucleoskeletal theory states that cell size is adjusted adaptively in response to selective pressures, and these changes enjoin correlated shifts in nucleus size (Cavalier-Smith 1985). In contrast, in the nucleotypic theory, DNA content is thought to exert a causative influence on cell parameters and is therefore seen as being subject to secondary selection via selection on cytological and organismal phenotypes (Bennett 1972). Lynch (2007) argues that the two above theories can be challenged by the fact that in prokaryotes, the strong relationship between cell size and genome size cannot be explained by cytoskeletal effects (as they lack nuclear membranes) nor can it be explained by the expansion of non-coding DNA. He advocates a population genetics approach to sort out cause and effect of genome size evolution and claims that heritable within-population variation in genome size covarying with cellular factors must be shown prior to asserting adaptive mechanisms. He further argues that the transitions from prokaryotes to unicellular eukaryotes to multicellular eukaryotes have been accompanied by reduced effective population size and these conditions likely led to the proliferation of genomic structures and hence to

increased genome size (Lynch and Conery 2003 but see Vinogradov 2004). This does not preclude adaptive phenotypic modifications as secondary events.

Associated with increases in cell size and genome size is a significant increase in the duration of cell division. Bennett (1977) reported that the duration of mitosis and meiosis were positively correlated with genome size, and therefore, large cells with high DNA contents develop more slowly. Since development at the organism level consists of division and growth at the cell level, this suggests that rate of development can be strongly influenced by variation in genome size. Evidence was provided using plant species with 3C nuclear DNA contents (three times the haploid genome size) ranging from 2.6 to 233.0 pg and in animals with 2C nuclear DNA contents ranging from 0.085 to 72.0 pg, suggesting that this pattern is widely applicable across the known range of genome size estimates in these groups (Bennett 1977). Developmental rate has also been found to be negatively correlated with genome size in some diploid plants and in some insects, crustaceans, salamanders, and mammals (Bennett 1987; White and McLaren 2000; Gregory 2002b, 2005a). Therefore, animals with large genomes can be expected to develop more slowly relative to related species with smaller genomes. This begs the question to be asked: Do slow-growing animals have large genomes because they tolerate more DNA, or do they have larger cells due to relaxed selection for fast growth? The fact that species with large genomes tend to inhabit environments with low biotic complexity, i.e., high latitudes (see below), would seem to favor the second hypothesis.

Species with large genomes often have larger body size. Positive correlations between genome size and body size have been found in aphids (Finston et al. 1995), flies (Gregory and Johnston 2008), mollusks (Hinegardner 1974), flatworms, and copepods (Gregory et al. 2000). This relationship is expected in species with a determinate growth pattern (i.e., when body size increases are mostly a result of cell size increases rather than increases in cell numbers). However, no correlation between genome size and body size was previously found in oligochaete annelids (Gregory and Hebert 2002), or beetles (Juan and Petitpierre 1991; Gregory et al. 2003). Mammals show a positive relationship between body size and genome size at the species and genus levels but not at higher levels (Gregory 2005b). Most of these examples are species that do not have large genomes under our definition as this data is still sparse. However, determining whether this pattern holds true in the groups discussed above by including species that have truly large genomes could be a rewarding line of research in this field.

A larger genome size can also have various effects on the physiology of an organism's cells. As cells become larger, the ratio of surface area to volume decreases, and this can have important effects on gas, ion, and protein exchange rates. Szarski (1970, 1983) noted that small cells (erythrocytes) and a small amount of DNA in the nucleus characterized groups with high metabolism. It is thought that physiological demands may have constrained the evolution of genome size in endothermic vertebrates by favoring smaller red blood cells to facilitate gas exchange (a constraint that mammals may have circumvented with enucleate red blood cells) (Cavalier-Smith 1985; Gregory 2000). Significant negative correlations between genome size and mass specific oxygen consumption have been found in birds and mammals (Vinogradov 1995, 1997; Gregory 2002a) but not in fishes (Hardie and Hebert 2004) nor in amphibians (Gregory 2001).

Ecological correlates with large genome sizes

An emerging trend from animal (and more specifically, crustacean) genome size studies is the positive relationship between genome size and latitude. This trend has been observed in many crop plants as genome size is generally larger in temperate species relative to tropical species (Bennett 1976 but see Bennett et al. 2008). These examples have been studied extensively in plants, but there is limited data on ecological correlations with genome size in animals, where more data are urgently needed to better understand these patterns. The positive correlation of genome size with latitude in crops has been influenced by the fact that humans have selected certain crops with high DNA amounts (as a by-product of other interesting characteristics) at higher latitudes (Bennett 1987). This generally holds true until a certain latitude, at which point average genome size for a particular group decreases with latitude as they extend into Arctic

ecosystems. This is thought to be due to the relatively short growing seasons in the Arctic where smaller genomes may be selected for so that these plants are able to develop and reproduce before they die (Bennett et al. 1982).

A similar trend may be expected for arctic terrestrial and freshwater organisms that are not able to survive the cold seasons. However, little genome size research has been conducted on arctic species with the exception of the genus *Daphnia*, where in fact the opposite trend is found. It is surprising to find *Daphnia* species with relatively large genomes (both diploid and polyploidy) within the Daphniidae (though still much smaller than 10 pg) and slow developmental rates in habitats with a short growing season. There is some evidence that polyploid *Daphnia* might have an advantage in subarctic zones as they show a tendency to mature earlier than diploid clones only at lower temperatures (Dufresne and Hebert 1998). Members of the *Daphnia pulex* complex have been found to be triploid in the Arctic, and some specimens of *Daphnia tenebrosa* have been found to be tetraploid which is likely due to hybrid origins (Vergilino et al. 2009). However, these genomes are still very small (less than 0.50 pg) and are not considered large in the sense of our definition, though these patterns may still apply to other species with truly large genomes such as some amphipods.

Interestingly, the only amphipods with large genomes examined thus far have been found in marine, arctic environments (Rees et al. 2007). Similarly, the majority of shrimps with large genomes are marine species found in either the arctic or near deep oceanic vents (Rees et al. 2008; Bonnivard et al. 2009). However, the only two shrimp with a genome larger than 10 pg that are not arctic or deep-sea species are the Ohio shrimp *Macrobrachium ohione* which is a freshwater species living in temperate North America with a genome size of 22.16 pg (Rheinsmith et al. 1974) and the grass shrimp *Palaemonetes* sp. with a genome size of 10.42 pg (Gregory, unpublished).

The relationship between habitat, latitude, and genome size in animals is as yet not well understood. A study by Hardie and Hebert (2003) on genome size diversity in fishes found that freshwater species tend to have larger genomes than marine species, possibly due to the freshwater fishes living in a harsher and fluctuating environment which may be linked to a higher DNA content. A similar suggestion was made by Vinogradov (2000) in which terrestrial mollusks had a higher DNA content than aquatic mollusks, possibly to promote genomic stability in an unstable terrestrial environment. However, the opposite trend seems to be true in many crustaceans in which the largest genomes are found in marine species. This may be due instead to living in a cold, stable environment that could allow for a longer development time (or eliminate the need for rapid development) rather than requiring more DNA to survive less stable environments. In the case of these arctic amphipods and shrimps, it is difficult to say whether their large genomes allows them to colonize a greater diversity of marine habitats across a wide range of latitudes, or whether these large genomes confine them to the colder waters of the poles or the abyss. If the latter is the case, climate change may greatly impact these species since they would not be able to survive the warming ocean. Climate change may also allow more temperate species with smaller genomes to colonize higher latitudes, further displacing the arctic species.

As large genomes in animals are still sparse, it is not yet known why certain lineages contain species with large genomes as others do not. Why, for example, do some crustaceans appear to have large genomes while other diverse phyla, such as the mollusks, do not? The mollusks are similar to the crustaceans in that they are diverse and can be found on land, in freshwater, and marine habitats at all latitudes, but no mollusk thus far has a genome size greater than 10 pg (Gregory 2011). Much more research needs to be conducted, particularly within the invertebrates, to determine other ecological correlates with genome size and to further explore the relationship between DNA content, habitat, and latitude. Future work should focus on estimating *C*-values for invertebrates from both poles and from a variety of habitats. Deep, cold lakes such as Lake Baikal in southern Siberia could be sampled to determine if benthic invertebrates have large genomes in a pattern similar to the marine benthic crustaceans examined thus far. This could provide interesting insight as to whether large genomes are more common to crustaceans living in all cold, stable environments such as the bottom of the ocean or deep lakes, or whether it is strictly marine benthic crustaceans that truly have the largest genomes of the invertebrates.

Phylogenetic trends of genome size evolution

The question of when and how often did large genomes originate requires an examination of genome size distribution in a phylogenetic context. However, the genome size database of animal species is too scattered to obtain as much insight on these questions as in plants.

A recent study using osteocyte lacunae size as a proxy for genome size in 14 extinct tetrapod genera from the Paleozoic and the early Mesozoic eras revealed that genome sizes were homogeneous among tetrapod lineages with values similar to extant mammals (Organ et al. 2011). These authors also report that the massive genome sizes of extant salamanders are likely derived, as genome sizes of extinct amphibians were found to be much smaller. Non-phylogenetically controlled studies of lung-fishes and amphibians also suggested an expansion of genome size in amphibians (Thomson and Muraszko 1978). Osteocyte measures further revealed that early reptiles had larger genome size than the average for extant non-avian reptiles, suggesting genome size contractions along this line (Organ et al. 2011).

Mode and tempo of genome size evolution

The skewed distribution of genome size has been attributed to the extinction of species with larger genomes due to deleterious effects (Knight et al. 2005), to higher fitness of smaller genomes, and to the inability of mobile elements to persist in large genomes due to the effects of sexual reproduction (Arkhipova and Meselson 2000). Oliver et al. (2007) examined the rate of genome size evolution across 20 eukaryotic clades and found that the tempo of genome size evolution was positively correlated to genome size. They suggested that a model of proportional evolution best explains this relationship. Under this model, the effects of DNA insertions and deletions depend on initial genome size; therefore, lineages with larger genomes should show more rapid genome size evolution. This stochastic proportionate draw of insertions and deletions is thus expected to produce far more small than large genomes.

Molecular mechanisms of genome size increases and decreases

It has been suggested that the net DNA content of an organism reflects the dynamic balance between the opposing forces of expansion and contraction (Oliver et al. 2007). In addition to these internal causal mechanisms, external selective factors are also expected to increase or decrease genome size via the correlated effects of nucleus/cell sizes (Gregory 2002b).

Expansion events include small-scale insertions of nucleotides or large-scale alterations such as gene duplications, transposon insertions, or more dramatic events such as polyploidy (Petrov 2001; Bennetzen 2002). By contrast, genomes contract following unequal intrastrand homologous recombination, double-strand break repair, illegitimate recombination, and retroelement extinction (Chen et al. 1998; Vicient et al. 1999; Shirasu et al. 2000; Orel and Puchta 2003). Somatic polyploidy is a mechanism that causes increases of DNA through endopolyploidy or polyteny. In endopolyploidy, the chromosome complement is progressively doubled within a nuclear membrane, while in polyteny the number of chromatids per chromosome increases. Somatic polyploidy is quite common, especially among arthropods (e.g., Bachmann and Rheinsmith 1973). It has been suggested that somatic polyploidy is inversely related to genome size (Nagl 1976) and therefore provides a compensation for these miniature genomes. For example, insects in the order Strepsiptera have tiny genomes and exhibit endoreduplication (Johnston et al. 2004). Likewise, *D. pulex* with a miniature genome of 0.24 pg has nearly half of its DNA in an endopolyploid state (Beaton and Hebert 1989). It would be tempting to speculate that species with large genomes lack polyteny but more genomes need to be examined to support this. Whole genome duplications have led to dramatic increases in genome size. These events have occurred less frequently in animals than in plants, though representative polyploid animals can be found in all major animal groups (to the exception of birds and mammals) (Otto and Whitton 2000). The recent advent of genome sequencing projects has revealed a surprising number of ancient polyploid species. For example, the yeast *Saccharomyces cerevisae* was found to be an ancient tetraploid (Wolfe and Shields 1997). It is now accepted that two rounds

of whole-genome duplications occurred during the early diversification of chordates and vertebrates, with strong evidence supporting a subsequent teleost fish-specific genome duplication (Dehal and Boore 2005; Hoegg and Meyer 2005). Ancient polyploids can be difficult to identify as extensive gene loss occurs following polyploidy and as rearrangements of chromosomal segments can further obscure the signature of polyploidy (Lynch 2007). Genome downsizing following polyploidy has been shown in a broad range of angiosperms, with species with larger genomes showing greatest genome size reduction, suggesting important biochemical costs of synthesizing and replicating extra DNA (Leitch and Bennett 2004).

Transposable elements (TEs) are major constituents of eukaryotic genomes and also have a great impact on genome structure and stability, in particular contributing to variations in genome size (Bennetzen 2005; Gregory 2005c; Oliver et al. 2007). Due to their ability to replicate at relatively rapid rates, TEs can increase genome size, often in response to specific stimuli such as stress and environmental changes (Grandbastien et al. 2005; Lesage and Todeschini 2005), and these changes can occur very rapidly. TEs will be discussed in greater detail in the section below.

Observations for mechanisms of genome size change from fully sequenced genomes

Each animal with a whole-genome sequence has had a genome size much smaller than 10 pg, and so these examples will be discussed and may be applicable to understanding the genome composition of species with truly large genomes. Results from fully sequenced genomes show a strong positive relationship between total genome size and the number of TE (Kidwell 2002; Lynch and Conery 2003). However, as information from full-genome sequence projects is still imperfect due to the low priority of sequencing the repetitive fraction of the genome (where most of these elements reside), these results are still subject to changes. Most TE can be assigned to two main classes: (1) the retrotransposons (short interspersed repetitive elements (SINEs), long interspersed repetitive elements (LINEs), long terminal repeat retrotransposon (LTR)) propagated by reverse transcription of an RNA intermediate and (2) DNA transposons that move through "cut and paste." DNA-based transposons,

LTR, and non-LTR retrotransposons (LINEs, SINEs) appear to show a threshold genome size below which mobile elements are unable to establish, an intermediate genome size where a fraction of species show them, and an upper threshold (100 Mb) where all species are infected (Lynch and Conery 2003). The nature of the relationship between TE and genome size suggests that factors other than TE may contribute more to genome size variation in the smaller (<0.5 pg) than in larger genomes (>0.5 pg) (Kidwell 2002). Preliminary data collected by Kidwell (2002) showed that the most striking difference between the larger genome (human) and the five smaller genomes (yeast, slime mold, *Drosophila*, nematodes) was the major increase in copy number per family in humans. The human genome harbors 45% TE (Fig. 2), SINE and LINE accounting for 11% and 17% of these. The proportion of the different classes, orders, and superfamilies of TEs in a given genome varies in different species; some harbor very abundant but few families of TEs such as mammals, where class I non-long terminal repeats (non-LTRs) predominate (Kidwell 2002; Wessler et al. 2006), while others contain many different yet less abundant families. Fish genomes usually contain several active clades of non-LTR retrotransposons, sometimes represented by a large diversity of families (Duvernell et al. 2004). This situation contrasts greatly with mammalian genomes that are dominated by a single clade, L1 in eutherians and marsupials (Furano et al. 2004; Gentles et al. 2007), and L2 in monotremes (Gilbert and Labuda 2000; Warren et al. 2008). The larger size of mammalian genomes relative to other vertebrates is thought to be directly related to the abundance of L1 elements that account for 17% of the human genome (Lander et al. 2001). More recently, the anole genome was found to contain an extraordinary diversity of non-LTR retrotransposons (Novick et al. 2009). This much greater diversity of non-LTR retrotransposons in the anole and in teleostean fish does not translate into larger genome sizes. In zebrafishes, mouse, and human, SINEs are the most abundant TEs (Sela et al. 2010). In the *Gallus gallus* genome, LINEs (belonging to the family of CR1 repeats) account for 79% of all TEs. In the *Danio rerio* genome, more than 75% of TEs are DNA transposons, whereas in *Drosophila melanogaster*, LTRs are the most abundant TEs, accounting for 44% of the elements observed. Finally, DNA transposons account for 95% of TEs in

Caenorhabditis elegans (Sela et al. 2010). TEs have contributed significantly to the ~5-fold size difference in the *Aedes aegypti* and the *Anapheles gambiae* genomes with the former species' genome being comprised of 47% TEs (Nene et al. 2007), a very large value for an invertebrate genome (Fig. 2). Smaller genomes may maintain a higher diversity and activity of transposable elements than larger genomes as a result of the strong intragenomic selective pressure to maintain a small genome (Gregory 2005c). To control TE proliferation, natural selection acting against the mutagenic effects of insertions, the metabolic costs of TE processing, and chromosomal damages have been credited as the dominant force limiting TE explosion (Charlesworth and Langley 1989). More recently, host-mediated RNA-based genome defense mechanisms have been found in eukaryotes (e.g., Malone and Hannon 2009). Small interfering RNA (siRNA) and piwi-interacting RNA repress TEs through post-transcriptional gene silencing and transcriptional silencing by DNA methylation and heterochromatin formation (in Blumenstiel 2011). Therefore, after the initial phase of TE invasion within a population, there will be proliferation of TE and the production of siRNA in response to the increase in TE insertion alleles in the genome. As explained by Blumenstiel (2011), transposition will continue at a rate that depends on the effects of silencing alleles. As this is a population genetics process, drift will lower the probabilities of fixation

by natural selection of the repressor alleles, and linkage between the repressor alleles and TE insertion sites will increase the effects of repressor alleles. The outcomes of these evolutionary games between TE invasion and the host immune system adaptation by these small RNA will likely influence genome size, with hosts winning at the game having smaller genomes and hosts losing at the game having larger genomes.

Intron size and genome size are known to be positively correlated between species of *Drosophila* (Moriyama et al. 1998), within the class of mammals (Ogata et al. 1996), and across eukaryotes in general (Vinogradov 1999). A comparison of 199 introns in 22 orthologous genes revealed that introns in the pufferfish *Fugu* were on average eight times smaller than those in humans, consistent with their ratio of genome sizes, though *Fugu* appears to have a uniquely small genome (McLysaght et al. 2000). By contrast, intron sizes in plants appear remarkably static even when confronted with mechanisms that massively expand or contract genome size (Wendel et al. 2002).

DNA-sequence studies have suggested that deletion biases might be responsible for important DNA loss leading to genome contractions in *Drosophila* (Petrov 2002). A strong negative correlation between rate of DNA loss and genome size has been observed upon comparing four species with different genome sizes. The grasshopper, *Podisma*, with a genome size of 17 pg lost DNA more slowly than humans (3.5 pg), which in turn had slower DNA losses than the Hawaiian cricket, *Laupala* (1.93 pg), while *Drosophila*, with a miniature genome of 0.16 pg, had the fastest DNA deletion pattern of all four species (Graur et al. 1989; Ophir and Graur 1997; Petrov and Hartl 1998; Petrov et al. 2001). Some pufferfish species in the families Tetraodontidae and Diodontidae also exhibited a similar bias toward DNA loss, generated by larger and more frequent deletions than insertions (Neafsey and Palumbi 2003) but did not differ in deletion rates despite a 2-fold difference in genome size. Petrov and Hartl (1998) also failed to detect deletion rate differences in two *Drosophila* species with a 2-fold difference in genome size. Gregory (2004) has questioned the importance of small indel bias as a determinant of genome size based on the small dataset and some problems associated with existing data and suggested that the specific mechanism relating to small deletion bias was too weak to play a
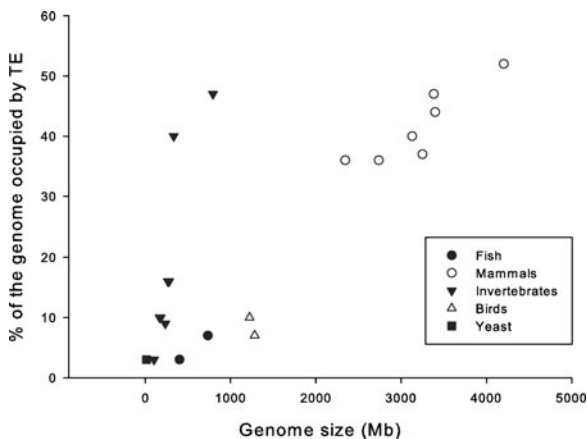
Fig. 2 The relationship between haploid genome size (one million base pairs) and the percent of the genome occupied by transposable elements in different animal taxa. The data have been obtained from species subject to large-scale sequencing analyses

determinant role in genome size variation. In addition, as reported in Lynch (2007), long-term accumulation studies in yeast have revealed a 15:4 insertion/deletion ratio (Denver et al. 2004).

Comparative genomic analysis within taxonomic groups is a powerful tool to uncover the mechanisms of genome size variation. It has been suggested that the reduced genome size of birds as compared to other tetrapods is an adaptation to the high rate of oxidative metabolism caused by the high energetic demands of flight (e.g., Tiersch and Wachtel 1991; Gregory 2002a). Birds have provided a powerful model to study mechanisms of genome shrinkage. Past studies have found that chickens have, on average, shorter introns than humans (Hughes and Hughes 1995), but this is not true of all mammalian species (Vinogradov 1999). A comparative analysis of 18,516 gene families in the genomes of five mammals and chicken revealed that the latter had a lower number of gene families than mammals (Hughes and Friedman 2008), congruent with their smaller genome sizes. The amount of simple sequence repeat array length in chicken is also smaller than in humans (Hughes and Piontkivska 2005). Through these comparative studies, we are starting to understand which genomic elements are associated with genome contraction and expansion, but we are still far from understanding why some species attain such large genome sizes. Are they less efficient than small genome species at DNA removal by recombination, or are the mechanisms of genome growth less prevalent in their genomes?

Methods for exploring large genome composition: next-generation sequencing

Complete genome sequencing using traditional Sanger approach has provided considerable insight into genome size evolution but has been largely restricted to animal species with small genomes due to costs and assembly problems associated with repetitive sequences found in large genomes. For example, whereas most vertebrate species sequenced to date have typical genome size values for animals, most invertebrate species sequenced have very small genomes (e.g., 0.10 pg for *C. elegans*, 0.16 pg for *D. melanogaster*, 0.23 pg for *D. pulex*). Our current understanding of structure, organization, and

composition of genomes is thus highly biased toward species with small genomes. With NGS techniques, there will be an explosion of genomics data and information on non-model species and on species with more complex genomes. Knowledge of large-scale genomic structure of species with large genomes will be an asset to better comprehend genome size evolution. NGS allows parallel sequencing of hundreds of thousands of individual templates immobilized on microbeads, thus producing megabases of sequence data in a single run (Mardis 2008).

The repetitive nature of TE sequences poses fundamental challenges to genome sequencing, assembly, annotation, and alignment (Devine et al. 1997; Myers et al. 2000; Bray et al. 2003). A hybrid strategy of Sanger and pyrosequencing was thought to be the best approach as the short read lengths produced by pyrosequencing make it difficult to span repetitive genomic elements (Goldberg et al. 2006) and rendering the assembly challenging. Therefore, for organisms with a large genome size, many sequencing gaps, and hard stops, an initial sequencing of 5.3× Sanger data followed by the addition of two Roche-454 paired-end runs is a cost-effective approach (Goldberg et al. 2006). Recently, the draft of the Argentine ant, *Linepithema humile*, was obtained based on a combination of Roche-454 and Illumina sequencing at 23× coverage (Smith et al. 2011). This species has a small genome of 0.26 pg and was found to have 31% of its genome made of repetitive elements (Smith et al. 2011). For comparison, in other Hymenopteran species, *Apis mellifera* with a genome size of 0.17 pg has a genome composed of 10% of repetitive elements, and *Nasonia vitripennis* with a genome size of 0.34 pg has 40% repetitive elements in its genome, thus showing the positive relationship between genome size and abundance of repetitive elements.

A draft of a larger and more complex genome, the one of the giant panda *Ailuropoda melanoleura* genome, was recently successfully assembled using Illumina Genome Analyzer sequencing technology alone at 56× coverage (Li et al. 2010). TEs comprised approximately 36.2% of the 2.40-pg panda genome, which is similar to that of the dog genome (36.1%), and lower than the human genome (46.1%) (Li et al. 2010). The turkey genome sequence represents the first eukaryotic genome completely sequenced and assembled de novo from data produced by a combination of two NGS platforms, Roche-454 and

Illumina Genome Analyzer II (Dalloul et al. 2010). Approximately 6.94% of the turkey genome was found to consist of interspersed repeats, most of which belonged to three groups of TEs: the *CR1*-type non-LTR retrotransposons, the LTR retrotransposons, and the *mariner*-type DNA transposons, a situation very similar to the chicken genome (Dalloul et al. 2010).

A number of recent studies using low genome coverage have proved useful to estimate the fraction of repetitive elements in the genomes of non-model species. For example, random pyrosequencing of the scuttle fly, *Megaselia scalaris*, with a genome size of 0.46 pg (C. J. Picard, personal communication) at 0.1× genome coverage has provided a glimpse of the genomic composition of repetitive elements (45% LTR, 38% non-LTR LINES, and 4% DNA tranpo-sons) (Rasmussen and Noor 2009). BAC end sequen-ces (BES) analyses at 3.34× clonal coverage in the common carp genome, *Cyprinus carpio*, a tetraploid species with a genome size of 1.80 pg, have revealed that repetitive DNA comprised approximately 28% of the genome (2.5% representation in the BES). The most abundant type of repetitive elements were DNA transposons (6.67%), mostly hobo-activator (2.25%), and followed by retroelements (4.52%) including LINEs (2.33%), LTR elements (1.98%), and SINEs (0.2%) (Xu et al. 2011). BES analyses studies on Atlantic salmon, with a genome size of 3.27 pg, revealed that repetitive elements accounted for 30–35% of the genome (Davidson et al. 2010). Fourteen families of DNA transposons, of which 11 were novel, comprised 6–10% of the genome. Sequencing this fish genome will be particularly challenging due to its autotetraploid nature, its long and frequent repeats, and the lack of a close genome reference (de Boer et al. 2007).

Conclusions and future research

The field of genome size research is becoming increasingly important as whole-genome sequencing becomes more efficient and as new trends are revealed that relate to differences in genome size. However, our knowledge of the distribution of and reasons for why certain lineages contain large genomes while others do not is poor at best. Thus far, no studies have attempted to obtain a full-genome sequence of an animal species with a genome size greater than 10 pg. Analyses using NGS at low coverage could help provide information on the genomic composition of these genomic giants. These studies will be highly important to understand which types of repetitive elements account for these large genomes and in what proportions. For example, recently, a strong correlation between genome size and the amount of satellite DNA underreplication was found in 12 *Drosophila* species (Bosco et al. 2007). It would be interesting to test if satellite DNA is also correlated with genome size in species with larger genomes. It is likely that there remain many types of TEs awaiting discovery within these large genomes. Assembly and annotation of the repeat-rich hetero-chromatin regions of sequenced genomes will likely refine our understanding of true composition of transposable elements in eukaryotic genomes. As many species with very large genomes are ancient polyploids in the process of diploidization, the role of TEs in these genomes will be particularly revealing.

The taxonomic distribution of large genomes in animals is currently quite scattered, being confined to the lungfishes, some amphibians, flatworms, orthop-terans, and some crustaceans, though it is likely the number of species with large genomes will increase as more estimates are provided. There are several phenotypic and ecological parameters that may correlate with increasing genome size, including developmental rate, body size, and latitude, though more research is needed to fill in the gaps and to determine how very large genomes correlate with these parameters specifi-cally. Much also remains to be learned about the respective roles of neutral processes versus natural selection in genome size alterations. The advent of NGS on animal species with large genomes should also prove useful to understand forces that allow genome expansion and also to understand why so few species have attained a very large genome size.

References

Arkhipova I, Meselson M (2000) Tranposable elements in sexual and ancient asexual taxa. Proc Natl Acad Sci 97:14473–14477

Bachmann K, Rheinsmith EL (1973) Nuclear DNA amounts in pacific Crustacea. Chromosoma 43:225–236

Beaton MJ, Hebert PDN (1989) Miniature genomes and endopolyploidy in cladoceran crustaceans. Genome 32:1048–1053

Beçak W, Beçak ML, Schreiber G, Lavalle D, Amorim FO (1970) Interspecific variability of DNA content in Amphibia. Experientia 26:204–206

Bennett MD (1972) Nuclear DNA content and minimum generation time in herbeacous plants. Proc Roy Soc Lond B 181:109–135

Bennett MD (1976) DNA amount, latitude, and crop plant distribution. Env Exp Biol 16:93–108

Bennett MD (1977) The time and duration of meiosis. Phil Trans R Soc Lond B 277:201–277

Bennett MD (1987) Variation in genomic form in plants and its ecological implications. New Phyt 106:177–200

Bennett MD, Smith JB, Lewis Smith RI (1982) DNA amounts of angiosperms from the Antarctic and South Georgia. Env Exp Biol 22:307–318

Bennett MD, Price HJ, Johnston JS (2008) Anthocyanin inhibits propidium iodide DNA fluorescence in *Euphorbia pulcherrima*: implications for genome size variation and flow cytometry. Ann Bot 101:777–790

Bennetzen JL (2002) Mechanisms and rates of genome expansion and contraction in flowering plants. Genetica 115:29–36

Bennetzen JL (2005) TE, gene creation and genome rearrangement in flowering plants. Curr Opin Genet Dev 15:621–627

Blumenstiel JP (2011) Evolutionary dynamics of transposable elements in a small RNA world. Trends Genet 27:23–31

Bonnivard E, Catrice O, Ravaux J (2009) Survey of genome size in 28 hydrothermal vent species covering 10 families. Genome 52:524–536

Bosco G, Campbell P, Leiva-Neto JT, Markow TA (2007) Analysis of *Drosophila* species genome size and satellite DNA content reveals significant differences among strains as well as between species. Genetics 177:1277–1290

Bray N, Dubchak I, Pachter L (2003) AVID: a global alignment program. Gen Res 13:97–102

Brown DD, Dawid IB (1968) Specific gene amplification in oocytes. Science 160:272–280

Cavalier-Smith T (1985) Cell volume and the evolution of eukaryotic genome size. In: Cavalier-Smith T (ed) The evolution of genome size. Wiley, Chichester, pp 104–184

Charlesworth B, Langley CH (1989) The population genetics of *Drosophila* transposable elements. Ann Rev Genet 23:251–287

Chen MS, SanMiguel P, Bennetzen JL (1998) Sequence organization and conservation in sh2/a1-homologous regions of sorghum and rice. Genetics 148:435–443

Dalloul RA, Long JA, Zimin AV (2010) Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): genome assembly and analysis. PLoS Biol 8 (9):e1000475. doi:10.1371/journal.pbio.1000475

Davidson WS, Koop BF, Jones SJM, Iturra P, Vidal R, Maass A et al (2010) Sequencing the genome of the Atlantic salmon *Salmo salar*. Genome Biol 11:403

de Boer JG, Yazawa R, Davidson WS, Koop BF (2007) Bursts and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids. BMC Genomics 8:422

Dehal P, Boore JL (2005) Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol 3:e314

Denver DR, Morris K, Lynch M, Thomas WK (2004) High direct estimates of the mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome. Nature 430:679–682

Devine SE, Chissoe SL, Eby Y, Wilson RK, Boeke JD (1997) A transposon-based strategy for sequencing repetitive DNA in eukaryotic genomes. Genet Res 7:551–563

Dufresne F, Hebert PDN (1998) Temperature-related differences in life-history characteristics between diploid and polyploid clones of the Daphnia pulex complex. Ecoscience 5:433–437

Duvernell DD, Pryor SR, Adams SM (2004) Teleost fish genomes contain a diverse array of L1 retrotransposon lineages that exhibit a low copy number and high rate of turnover. J Mol Evol 59:298–308

Escribano R, McLaren IA, Klein Breteler WCM (1992) Innate and acquired variation of nuclear DNA contents of marine copepods. Genome 35:602–610

Finston TL, Herbert PD, Foottit RB (1995) Genome size variation in aphids. Insect Biochem Mol Biol 25:189–196

Furano AV, Duvernell DD, Boissinot S (2004) L1 (line-1) retrotransposon diversity differs dramatically between mammals and fish. Trends Genet 20:9–14

Gentles AJ, Wakefield MJ, Kohany O, Gu W, Batzer MA, Pollock DD et al (2007) Evolutionary dynamics of transposable elements in the short-tailed opossum *Monodelphis domestica*. Genome Res 17:992–1004

Gilbert N, Labuda D (2000) Evolutionary inventions and continuity of CORE-SINEs in mammals. J Mol Biol 298:365–377

Goldberg SMD, Johnson J, Busam D, Feldblyum T, Ferriera S, Friedman R et al (2006) A Sanger pyrosequencing hybrid approach for the generation of high-quality draft assemblies of marine microbial genomes. Proc Natl Acad Sci 103:11240–11245

Gosalvez J, López-Fernandez C, Esponda P (1980) Variability of the DNA content in five orthopteran species. Caryologia 33:275–281

Grandbastien M, Audeon C, Bonnivard E (2005) Stress activation and genomic impact of Tnt1 retrotransposons in Solanaceae. Cyt Genome Res 110:229–241

Graur D, Shuali Y, Li W-H (1989) Deletions in processed pseudogenes accumulate faster in rodents than in humans. J Mol Evol 28:279–285

Gregory TR (2000) Nucleotypic effects without nuclei: Genome size and erythrocyte size in mammals. Genome 43:895–901

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

Gregory TR (2002a) A bird's eye view of the C-value enigma: genome size, cell size, and metabolic rate in the class Aves. Evolution 56:121–130

Gregory TR (2002b) Genome size and developmental complexity. Genetica 115:131–146

Gregory TR (2004) Insertion-deletion biases and the evolution of genome size. Gene 324:15–34

Gregory TR (2005a) The C-value enigma in plants and animals: a review of parallels and an appeal for partnership. Ann Bot 95:133–146

Gregory TR (2005b) The evolution of genome size in animals. In: TR Gregory (ed) The evolution of the genome. pp. 4–71

Gregory TR (2005c) Synergy between sequence and size in Large-scale genomics. Nature Rev Genetics 6:699–708

Gregory TR (2011). Animal Genome Size Database. http://www.genomesize.com

Gregory TR, Hebert PDN (2002) Genome size estimates for some oligochaete annelids. Can J Zool 80:1485–1489

Gregory TR, Johnston JS (2008) Genome size diversity in the family Drosophilidae. Heredity 101:228–238

Gregory TR, Hebert PDN, Kolasa J (2000) Evolutionary implications of the relationship between genome size and body size in flatworms and copepods. Heredity 84:201–208

Gregory TR, Nedved O, Adamowicz SJ (2003) C-value estimates for 31 species of ladybird beetles (Coleoptera: Coccinellidae). Hereditas 139:121–127

Grime JP (1983) Prediction of weed and crop response to climate based upon measurements of nuclear DNA content. Aspect Appl Biol 4:87–98

Hardie DC, Hebert PDN (2003) The nucleotypic effects of cellular DNA content in cartilaginous and ray-finned fishes. Genome 46:683–706

Hardie DC, Hebert PDN (2004) Genome size evolution in fishes. Can J Fish Aquat Sci 61:1636–1646

Hebert PDN, Beaton MJ (1990) Breeding system and genome size of the rhabdocoel turbellarian Mesostoma ehrenbergii. Genome 33:719–724

Hinegardner R (1974) Cellular DNA content of the Mollusca. Comp Biochem Physiol 47A:447–460

Hoegg S, Meyer A (2005) Hox clusters as models for vertebrate genome evolution. Trends Genet 21:421–424

Horner HA, Macgregor HC (1983) C value and cell volume: their significance in the evolution and development of amphibians. J Cell Sci 63:135–146

Hughes AL, Friedman R (2008) Genome size reduction in the chicken has involved massive loss of ancestral protein-coding genes. Mol Biol Evol 25:2681–2688

Hughes AL, Hughes MK (1995) Small genomes for better flyers. Nature 377:391

Hughes A, Piontkivska H (2005) DNA repeat arrays in chicken and human genomes and the adaptive evolution of avian genome size. BMC Evol Biol 5:12

Johnston JS, Ross LD, Hughes DP, Kathirithamby J (2004) Tiny genomes and endoreduplication in Strepsiptera. Insect Mol Biol 13:581–585

Juan C, Petitpierre E (1991) Evolution of genome size in darkling beetles (Tenebrionidae, Coleoptera). Genome 34:169–173

Kidwell MG (2002) TE and the evolution of genome size in eukaryotes. Genetica 115:49–63

Knight C, Molinari NA, Petrov DA (2005) The large genome constraint hypothesis: evolution, ecology, and phenotype. Ann Bot 95:177–190

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J et al (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921

Leitch IJ, Bennett MD (2004) Genome downsizing in polyploid plants. Biol J Linn Soc Lond 82:651–663

Lesage P, Todeschini AL (2005) Happy together: the life and times of Ty retrotransposons and their hosts. Cyt Genome Res 110:70–90

Li R, Fan W, Tian G, Zhu H, He L, Cai J et al (2010) The sequence and de novo assembly of the giant panda genome. Nature 463:311–317

Licht LA, Lowcock LE (1991) Genome size and metabolic-rate in salamanders. Comp Biochem Physiol 100:83–92

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

Lynch M, Conery JS (2003) The origins of genome complexity. Science 302:1401–1404

Malone CD, Hannon GJ (2009) Small RNA as guardians of the genome. Cell 136:656–668

Mardis ER (2008) The impact of next-generation sequencing technology on genetics. Trends Genet 24:133–141

McLaren IA, Sevigny JM, Corkett CJ (1988) Body sizes, development rates, and genome sizes among Calanus species. Hydrobiologia 167–168:275–284

McLysaght A, Enright AJ, Skrabanek L, Wolfe KH (2000) Estimation of synteny conservation and genome compaction between pufferfish (Fugu) and human. Yeast 17:22–36

Mirsky AE, Ris H (1951) The deoxyribonucleic acid content of animal cells and its evolutionary significance. J Gen Physiol 34:451–462

Moriyama EN, Petrov DA, Hartl DL (1998) Genome size and intron size in Drosophila. Mol Biol Evol 15:770–773

Myers EW, Sutton GG, Delcher AL (2000) A whole-genome assembly of Drosophila. Science 287:2196–2204

Nagl W (1976) DNA endoreduplication and polyteny understood as evolutionary strategies. Nature 261:614–615

Neafsey DE, Palumbi SR (2003) Genome size evolution in pufferfish: a comparative analysis of diodontid and tetraodontid pufferfish genomes. Genome Res 13:821–830

Nene V, Wortman JR, Lawson D et al (2007) Genome sequence of Aedes aegypti, a major arbovirus vector. Science 316:1718–1723

Novick PA, Basta H, Floumanhaft M (2009) The evolutionary dynamics of autonomous non-LTR retrotransposons in the lizard Anolis carolinensis shows more similarity to fish than mammals. Mol Biol Evol 26:1811–1822

Ogata H, Fujbuchi W, Kanehisa M (1996) The size differences among mammalian introns are due to the accumulation of small deletions. Febs Letters 390:99–103

Oliver MJ, Petrov D, Ackerly D, Falkowski P, Schofield OM (2007) The mode and tempo of genome size evolution in eukaryotes. Genome Res 17:594–601

Olmo E (1973) Quantitative variations in nuclear DNA and phylogenesis of Amphibia. Karyologia 26:43–68

Olmo E (1974) Further data on the genome size in the urodeles. B Zool 41:29–33

Olmo E (1983) Nucleotype and cell size in vertebrates: a review. Basic Appl Histochem 27:227–256

Olmo E, Morescalchi A (1975) Evolution of the genome and cell sizes in salamanders. Experientia 31:804–806

Olmo E, Odierna E (1982) Relationships between DNA content and cell morphometric parameters in reptiles. Basic Appl Histochem 26:27–34

Ophir R, Graur D (1997) Patterns and rates of indel evolution in processed pseudogenes from humans and murids. Gene 205:191–202

Orel N, Puchta H (2003) Differences in the processing of DNA ends in *Arabidopsis thaliana* and tobacco: possible implications for genome evolution. Plant Mol Biol 51:523–531

Organ CL, Canoville A, Reisz RR (2011) Paleogenomic data suggest mammal-like genome size in the ancestral amniote and derived large genome size in amphibians. J Evol Biol 24:372–380

Otto SP, Whitton J (2000) Polyploidy: incidence and evolution. Annu Rev Genet 34:401–437

Pedersen RA (1971) DNA content, ribosomal gene multiplicity, and cell size in fish. J Exp Zool 177:65–79

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

Petrov DA (2002) DNA loss and evolution of genome size in *Drosophila*. Genetica 115:81–91

Petrov DA, Hartl DL (1998) High rate of DNA loss in the *Drosophila melanogaster* and *Drosophila virilis* species groups. Mol Biol Evol 15:293–302

Petrov DA, Sangster TA, Johnston JS, Hartl DL, Shaw KL (2001) Evidence for DNA loss as a determinant of genome size. Science 287:1060–1062

Rasch EM, Rasch RW (1981) Cytophotometric determination of genome size for two species of cave crickets (Orthoptera, Rhaphidophoridae). J Histochem Cytochem 29:885

Rasmussen DA, Noor MAF (2009) What can you do with 0.1× genome coverage? A case study based on a genome survey of the scuttle fly *Megaselia scalaris* (Phoridae). BMC Genomics 10:382. doi:10.1186/1471-2164-10-382

Rees DJ, Dufresne F, Glémet H (2007) Amphipod genome sizes: first estimates for Arctic species reveal genomic giants. Genome 50:151–158

Rees DJ, Belzile C, Glémet H, Dufresne F (2008) Large genomes among caridean shrimp. Genome 51:159–163

Rheinsmith EL, Hinegardner R, Bachmann K (1974) Nuclear DNA amounts in Crustacea. Comp Biochem Physiol 48B:343–348

Sela N, Kim E, Ast G (2010) The role of transposable elements in the evolution of non-mammalian vertebrates and invertebrates. Genome Biol 11:R59

Sexsmith LE (1968) DNA values and karyotypes of amphibians. Ph.D. thesis, Department of Botany, University of Toronto

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

Smith CD, Ziminb A, Holtc A, Abouheif E, Benton R, Cash E et al (2011) Draft genome of the globally widespread and invasive Argentine ant (*Linepithema humile*). Proc Natl Acad Sci 108:5673–5678

Stingo V, Capriglione T, Rocco L, Improta R, Morescalchi A (1989) Genome size and A-T rich DNA in selachians. Genetica 79:197–205

Szarski H (1970) Changes in the amount of DNA in cell nuclei during vertebrate evolution. Nature 226:651–652

Szarski H (1983) Cell size and the concept of wasteful and frugal evolutionary strategies. J Theor Biol 105:201–209

Thomson KS, Muraszko K (1978) Estimation of cell size and DNA content in fossil fishes and amphibians. J Exp Zool 205:315–320

Tiersch TR, Wachtel SS (1991) On the evolution of genome size of birds. J Hered 82:363–368

Vergilino R, Belzile C, Dufresne F (2009) Genome size evolution and polyploidy in the *Daphnia pulex* complex (Cladocera: Daphniidae). Biol J Linn Soc 97:68–79

Vicient CM, Suoniemi A, Anamthawat-Jonsson K (1999) Retrotransposon BARE-1 and its role in genome evolution in the genus Hordeum. Plant Cell 11:1769–1784

Vinogradov AE (1995) Nucleotypic effect in homeotherms: body-mass-corrected basal metabolic rates of mammals is related to genome size. Evolution 49:1249–1259

Vinogradov AE (1997) Nucleotypic effect in homeotherms: Body-mass independent resting metabolic rate of passerine birds is related to genome size. Evolution 51:220–225

Vinogradov AE (1999) Intron–genome size relationship on a large evolutionary scale. J Mol Evol 49:376–384

Vinogradov AE (2000) Larger genomes for molluskan land pioneers. Genome 43:211–212

Vinogradov AE (2004) Testing genome complexity. Science 304:389–390

Vinogradov AE (2005) Genome size and chromatin condensation in vertebrates. Chromosoma 113:362–369

Warren WC, Hillier LW, Marshall Graves JA, Birney E, Ponting CP, Grützner F et al (2008) Genome analysis of the platypus reveals unique signatures of evolution. Nature 453:175–183

Wendel JF, Cronn RC, Alvarez I (2002) Intron size and genome size in plants. Mol Biol Evol 19:2346–2352

Wessler SR (2006) Transposable elements and the evolution of eukaryotic genomes. Proc Natl Acad Sci 103:17600–17601

Westerman M, Barton NH, Hewitt GM (1987) Differences in DNA content between two chromosomal races of the grasshopper *Podisma pedestris*. Heredity 58:221–228

White MM, McLaren IA (2000) Copepod development rates in relation to genome size and 18S rDNA copy number. Genome 43:750–755

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

Xu P, Li J, Li Y, Cui R, Wang J, Wang J et al (2011) Genomic insight into the common carp (*Cyprinus carpio*) genome by sequencing analysis of BAC-end sequences. BMC Genomics 12:188. doi:10.1186/1471-2164-12-188