NEWS AND COMMENTARY

Genome size of fossils
··

Genome sizes through the ages

IJ Leitch

···

The enormity of the c. 40 000-fold range in genome size (genome size or 1C nuclear DNA amount refers to the amount of DNA in the unreplicated gametic nucleus (e.g. pollen or sperm) and is usually measured in picograms (pg) or base pairs (bp); 1 pg ≈ 1 billion bp ≈ 1000 Mb) in eukaryotes and the lack of correlation with organismal complexity have intrigued scientists for over half a century. People have asked how and why genomes vary so extensively and whether it matters. The recent paper by Organ et al. (2007) has extended a paleogenomics dimension to these questions. By using the size of fossil dinosaur bone cells as proxies for genome size, they have attempted to trace the evolution of genome size in reptiles over 200 million years. Further, by analyzing currently available sequence data from a range of reptiles and birds, they have aimed to shed light on the genomic makeup of dinosaur genomes.

The first estimates of genome size were made in the late 1940s, but as data increased, it soon became clear that there was a huge disparity between organismal complexity and genome size. This led scientists such as Comings (1972) to question 'why the lowly liverwort has 18 times as much DNA as we have, and the slimy, dull salamander known as Amphiuma has 26 times our complement of DNA'. Since then, there has been much progress in understanding the molecular basis behind how genomes vary so extensively in size. It is now widely accepted that genome size diversity arises from differences in the amount of non-coding repetitive DNA (e.g. pseudogenes, retrotransposons, transposons satellite repeats, and so on.). The actual genome size of an organism is determined by the differential activity of mechanisms generating increases (e.g. retrotransposon amplification, polyploidy, segmental duplications) and decreases (e.g. illegitimate and unequal recombination, differences in double-strand break repair) in the DNA amount (reviewed in Gregory, 2005).

Nevertheless, despite genome size data for over 5000 plant and 4300 animal species readily available through the Plant DNA C-values database (Bennett and Leitch, 2005) and the Animal Genome Size database (Gregory, 2006), the reason(s) why genomes vary so extensively in size and the direction and nature of the evolutionary forces driving such changes are still not clearly understood.

Tracking the nature and direction of genome size evolution

Attempts to put a phylogenetic time frame on genome size evolution have been made using several approaches. High-resolution comparative genomics tools have been very informative in showing the timing, nature and mechanisms involved in genome size changes (Petrov, 2002; Hawkins et al., 2006; Vitte and Bennetzen, 2006).

A broader perspective on genome size evolution has been achieved by superimposing genome size data onto phylogenetic trees. When combined with estimates of divergence times between lineages, and with statistics (e.g. maximum parsimony, generalized least squares), such approaches can help pinpoint where and when changes might have taken place and make predictions as to the size of ancestral genomes. For example, Leitch et al. (2005) analyzed genome size data for 4538 land plant species. They revealed that genome size evolution was dynamic by showing that numerous independent increases and decreases had taken place during land plant evolution. They also suggested that the ancestral genome size of angiosperms was small (i.e. <1C = 1.4 pg). Increasingly, these approaches are being applied to both animal and plant systems at different taxonomic levels enabling species-specific patterns and unifying trends to be identified (Wendel et al., 2002; Honeycutt et al., 2003; Weiss-Schneeweiss et al., 2005).

Obtaining genome sizes from fossils

While extrapolating ancestral genome sizes using data from extant species is illuminating, verification of such approaches requires estimation of genome size from fossil material of known geological age. There are, of course, formidable obstacles for obtaining such data, but the well-documented correlation between genome size and cell size for certain cell types has been exploited by Organ et al. (2007) and a few other researchers to achieve this.

Masterson (1994) used leaf guard cell size as a proxy for genome size to track changes in some early angiosperm families such as magnolias (Magnoliaceae) over 100 million years ago (Mya). She showed that some fossil magnolias had genomes around 1 pg (∼1 billion bp) suggesting that a small genome may have been typical of early diverging angiosperms. It is reassuring that Masterson's (1994) data support the ancestral genome size predicted by Leitch et al. (2005) using genome size measurements of extant species (previously outlined).

In a geologically more extensive survey, Conway Morris and Harper (1988) analyzed genome size evolution in extinct conodonts over 270 million years using the size of epithelial cells as proxies for genome size, while Thomson (1972) used the size of bone osteocyte cells to track genome size changes in lungfish (Dipnoi) over an unbroken fossil record dating back to the Devonian (400 Mya). Thomson concluded that the truly enormous genomes of extant lungfish (the largest being for the marbled lungfish Protopterus aethiopicus 1 C = ∼133 pg or 133 billion bp) must be derived. In addition, he also found that during the period of greatest morphological diversification in the Devonian, cell size (and hence, by association, genome size) remained small. Increases in cell size were not observed until the Carboniferous, by which time the rate of morphological change in the group had slowed considerably. The data led Thomson (1972) to suggest that genome size was inversely correlated with evolutionary rate.

Thomson and Muraszko (1978) went on to use osteocytes to trace genome size evolution in 26 genera of lobe-finned fish and amphibians. They concluded that ancestral fishes and tetrapods (i.e. early amphibians and reptiles) all had small cells and hence small genomes in the range 1C = 2.5–5 pg. Larger cells were observed only in some of the more recent fossils and extant species, representing the product of genome growth occurring independently in several lineages.

The paper by Organ *et al.* (2007) adds to these studies by looking at genome size evolution in fossil dinosaurs. In a similar approach to Thomson (1972), Organ *et al.* used osteocyte size as a proxy to estimate genome size in 31 extinct species of dinosaur selected from the major dinosaur lineages (ornithischians, sauropods and theropods) together with data for 26 extant species. Combining these data with current phylogenetic trees and statistical approaches, they estimated that the early evolving reptiles were characterized by genomes in the range of 2–3 pg. This is similar to the estimate by Thomson and Muraszko outlined above. However, they noted a marked reduction in genome size c. 230 Mya, right at the base of the branch leading to the fast moving carnivorous theropods such as *Tyrannosaurus rex* and long before the evolution of birds.

These studies have also helped to resolve the long standing debate concerning the origin of the very narrow range of genome sizes encountered in extant birds (1C = 1.0–2.2 pg). While some researchers argued that birds had evolved from reptiles with small genomes, others suggested that bird evolution had been accompanied by genome downsizing from larger reptilian genomes (Gregory, 2002). The paleogenomic approach of Organ *et al.* (2007) supports the former view. Nevertheless, given the prevalence of mechanisms capable of increasing genome size, the apparent stasis in DNA amount over 230 million years of evolution in the theropods and extant birds suggests that there have been strong selection pressures acting to keep genomes small.

It is clear that paleogenomic studies provide added perspective to understanding patterns of genome size evolution. It is heartening that the few studies to date reflect the patterns observed in the more widely applied genome size modelling approaches utilizing phylogenetic schemes and C-value data of extant species. Future studies need to be extended to determine other places where extensive changes in genome size have taken place during evolution. Furthermore, by combining such data with our increased understanding of paleoenvironments, there is the exciting potential of identifying possible ecological events that may have contributed to triggering major changes in genome size.

Dr IJ Leitch is at Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey, UK.

e-mail: i.leitch@kew.org

Bennett MD, Leitch IJ (2005). Plant DNA C-values database (release 4.0, Oct. 2005). http://www.kew.org/genomesize/homepage.

Comings DE (1972). The structure and function of chromatin. *Adv Hum Genet* **3**: 237–431.

Conway Morris S, Harper E (1988). Genome size in Conodonts (Chordata): inferred variation during 270 million years **241**: 1230–1232.

Gregory TR (2002). 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 (ed) (2005). *The Evolution of the Genome.* Elsevier: San Diego.

Gregory TR (2006). Animal genome size database. http://www.genomesize.com.

Hawkins JS, Kim H, Nason JD, Wing RA, Wendel JF (2006). Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. *Genome Res* **16**: 1252–1261.

Honeycutt RL, Rowe DL, Gallardo MH (2003). Molecular systematics of the South American caviomorph rodents: relationships among species and genera in the family Octodontidae. *Mol Phylogenet Evol* **26**: 476–489.

Leitch IJ, Soltis DE, Soltis PS, Bennett MD (2005). Evolution of DNA amounts across land plants (Embryophyta). *Ann Bot* **95**: 207–217.

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

Organ CL, Shedlock AM, Meade A, Pagel M, Edwards SV (2007). Origin of avian genome size and structure in non-avian dinosaurs. *Nature* **446**: 180–184.

Petrov DA (2002). Mutational equilibrium model of genome size evolution. *Theor Popul Biol* **61**: 531–544.

Thomson KS (1972). An attempt to reconstruct evolutionary changes in the cellular DNA content of lungfish. *J Exp Zool* **180**: 363–372.

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

Vitte C, Bennetzen JL (2006). Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. *Proc Natl Acad Sci USA* **103**: 17638–17643.

Weiss-Schneeweiss H, Greilhuber J, Schneeweiss GM (2005). Genome size evolution in holoparasitic *Orobanche* (Orobanchaceae) and related genera. *Am J Bot* **93**: 148–156.

Wendel JF, Cronn RC, Johnston JS, Price HJ (2002). Feast and famine in plant genomes. *Genetica* **115**: 37–47.

Editor's suggested reading

Castagnone-Sereno P (2006). Genetic variability and adaptive evolution in parthenogenetic root-knot nematodes. *Heredity* **96**: 282–289; Advance online publication 11 January 2006.

Kalisz S, Kramer EM (2007). Variation and constraint in plant evolution and development. *Heredity* (advance online publication, 31 January 2007; doi:10.1038/sj.hdy.6800939).

Leitch AR (2007). Conserved gene order belies rapid genome turnover: the dynamic interplay between genomic DNA and the outside world. *Heredity* **98**:61–62; Advance online publication 20 December 2006.

Sabot F, Schulman AH (2006). Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome. *Heredity* **97**: 381–388; advance online publication, 20 September 2006.