npg

NEWS AND COMMENTARY

The *Arabidopsis lyrata* genome sequence

Sizing up *Arabidopsis* genome evolution

SI Wright and JA Ågren

Plant genomes vary tremendously in size, even among close relatives (Bennett and Leitch, 2010). In addition, there is considerable variation in related features such as chromosome number and size, number of genes and transposable element (TE) content. There have been several attempts to shed light on both the relative importance of these contributors, and the relative role of mutation versus natural selection, in driving genome size evolution. However, our ability to understand the evolution of genome size to date has been limited, due in part to the lack of large-scale genomic information from closely related species. With the recent publication of the complete genome of the primarily self-incompatible plant *Arabidopsis lyrata* (Hu *et al.*, 2011), a close relative of the workhorse of plant genetics *Arabidopsis thaliana*, many of these questions can now be approached with increased strength under a comparative framework.

Several theoretical predictions have been made concerning genome size evolution. First, if genome expansion is governed by slightly deleterious mutations, species with larger effective population sizes (*Ne*) are expected to experience more efficient selection, and therefore maintain smaller genomes (Lynch and Conery, 2003). Second, asexual and highly selfing species are expected to experience reduced activity of TEs, which could contribute to genome shrinkage in these organisms (Bestor, 1999). This second prediction is somewhat at odds with the first, as the effective population size of selfing species should be reduced by at least half that of outcrossing species (Nordborg, 2000). Third, organisms with relatively quick development times tend to have smaller genomes, and there may be stronger selection for a compact genome in faster growing organisms (Pagel and Johnstone, 1992). Finally, species may differ in their relative rates of insertion and deletion, thereby driving genome expansion versus contraction under a purely neutral process (Petrov *et al.*, 2000).

It has long been known that the genome of *A. thaliana* genome was one of the smallest in angiosperms, and indeed this was a major reason for its adoption as a model system. Although the other species in the genus also have relatively small genomes, the *A. thaliana* genome is reduced by almost one half compared with the others, suggesting recent and rapid DNA loss. Moreover, compared with its closest relatives, the species also shows a number of relevant shifts in life history; it is highly selfing, in contrast with its predominantly self-incompatible and outcrossing relatives; it is annual rather than perennial, and patterns of nucleotide diversity suggest that it has a relatively small effectively population size (Clark *et al.*, 2007; Ross-Ibarra *et al.*, 2008), although diversity levels are still reasonably high. Thus, comparative genomics in this system provides an exciting framework for investigating the prime factors driving genome size evolution.

Despite the general similarity in gene order and a high level of sequence similarity in genes, Hu *et al.* (2011) report a difference of approximately 80 MB (over 200 MB compared with 125 MB) in genome size between *A. thaliana* and *A. lyrata*. What is perhaps most striking about the genome size shift is the consistency of genome loss; there are clear reductions in size due to chromosome rearrangements, TE copy number, small and large deletions, and even gene number. Furthermore, with the exception of single base pair deletions, the DNA size change is apparent for deletion/insertion events at all size ranges, although it is especially exaggerated at the larger size range.

What processes may be responsible for this genome loss? Clearly, the directionality is at odds with predictions based on differences in effective population size, as there is consistent evidence for a moderate loss of genetic variation in *A. thaliana*. In contrast, the global nature of genome size reduction suggests that directional shifts in muta-

tion pressure and/or stronger selection favoring deletions over insertions could have driven genome shrinkage in *A. thaliana*. Population genetic analysis can help distinguish these possibilities. If selection were acting against insertions but favoring deletions, we would expect to detect higher population frequencies of deletions compared with insertions. The authors analyzed polymorphism data and found support for this. By using *A. lyrata* to polarize the directionality of polymorphic size changes in *A. thaliana*, they show that insertions segregate at much lower population frequencies than deletions, suggesting ongoing selection pressures against insertions and in favor of deletions.

However, analysis of TEs suggests a somewhat different interpretation. Consistent with the general patterns, TE copy number is consistently higher in *A. lyrata* regardless of element type, ruling out a major expansion of any particular class of TE. In total, at least 30% of the *A. lyrata* genome is comprised of TEs, in contrast with 24% in *A. thaliana*. Analyses of sequence divergence of long-terminal repeats of retrotransposons and TE phylogenies suggest an enrichment of younger TE insertions in *A. lyrata* compared with *A. thaliana*. This suggests higher rates of TE activity in the outcrossing species, consistent with theoretical predictions about reduced genomic conflict following the evolution of selfing (Bestor, 1999). If a higher genome-wide rate of fixation of deletions is the only process governing genome loss in *A. thaliana*, we would expect the opposite pattern: fewer old insertions and a shift toward more young insertions. Furthermore, previous work suggests that TE insertions segregate at lower population frequencies in *A. lyrata* than in *A. thaliana* (Wright *et al.*, 2001; Lockton and Gaut, 2010), consistent with either stronger selection against TE insertions in *A. lyrata* or a reduced rate of transposition in *A. thaliana* (Wright *et al.*, 2001). Together with the genomic data and recent analysis of expression levels (Hollister *et al.*, 2011), the transposon patterns imply reduced activity in *A. thaliana* relative to *A. lyrata*. As one possible explanation, transposon silencing has been suggested to be more efficient in *A. thaliana*, although whether this is a cause or an effect of having more young insertions in *A. lyrata* remains unclear (Hollister *et al.*, 2011).

We are therefore left with evidence for two processes explaining differences in

genome size, stronger selection for DNA removal and/or lower rates of TE activity. It is unclear whether the patterns could in fact be driven by a single mechanism. For example, higher TE activity via outcrossing in *A. lyrata* may generally lead to higher rates of gene duplication and a greater abundance of small pieces of neutral DNA. However, there are several future studies that could help distinguish whether either, or both, of these factors are contributing. If genome size evolution is mainly because of the differences in TE activity, we would expect polymorphism analysis in *A. lyrata* to reveal similar selection pressures on small insertions and deletions at homologous sequences, whereas basic transposition rates should differ between species. Alternatively, if selection favouring deletions in *A. thaliana* is the primary cause, we would expect similarly-aged TE insertions to show a higher rate of DNA loss in the selfing species. Finally, the sequencing of genome of the outgroup species *Capsella rubella* will be crucial for determining what factors are due to DNA loss in *A. thaliana* versus expansion in *A. lyrata*.

The contribution of *A. thaliana* genome sequencing to our understanding of plant genetics can hardly be exaggerated. The sequencing of the *A. lyrata* genome opens up many new avenues of research. In particular, the presence of the genome of two closely related species will create many possibilities to examine genome consequences of recent evolutionary changes, such as change alterations in mating system, life history and effective population size.

Conflict of interest

The authors declare no conflict of interest.

SI Wright and JA Ågren are at the Department of Ecology and Evolutionary Biology, University of Toronto, 25 Willcocks Street, Toronto, Ontario, Canada M5S 3B2.

e-mail: stephen.wright@utoronto.ca

Bennett MD, Leitch IJ (2010). Plant DNA C-value database (release 5.0, December 2010). Royal Botanic Gardens, Kew. Available at: http://data.kew.org/cvalues/.

Bestor TH (1999). Sex brings transposons and genomes into conflict. *Genetica* **107**: 289–295.

Clark RM, Schweikert G, Toomajian C, Ossowski S, Zeller G, Shinn P *et al.* (2007). Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science* **317**: 338–342.

Hollister JD, Smith LM, Guo Yl, Ott F, Weigel D, Gaut BS (2011). Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Proc Natl Acad Sci USA* **108**: 2322–2327.

Hu TT, Pattyn P, Bakker EG, Cao J, Cheng J, Clark RM *et al.* (2011). The Arabidopsis lyrata genome sequence and the basis of rapid genome size change. *Nat Genet* **43**: 476–481.

Lockton S, Gaut BS (2010). The evolution of transposable elements in natural populations of self-fertilizing Arabidopsis thaliana and its outcrossing relative Arabidopsis lyrata. *BMC Evol Biol* **10**: 10.

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

Nordborg M (2000). Linkage disequilibrium, gene trees and selfing: ancestral recombination graph with partial self-fertilization. *Genetics* **154**: 923–939.

Pagel M, Johnstone RA (1992). Variation across species in the size of the nuclear genome supports the junk-DNA explanation for the C-value paradox. *Proc Roy Soc Lond B* **249**: 119–124.

Petrov DA, Sangster TA, Johnston JS, Hartl DL, Shaw KL (2000). Evidence for DNA loss as a determinant of genome size. *Science* **287**: 1060–1062.

Ross-Ibarra J, Wright SI, Foxe JP, Kawabe A, DeRose-Wilson L, Gos G *et al.* (2008). Patterns of polymorphism and demographic history in natural populations of Arabidopsis lyrata. *PloS One* **3**: e2411.

Wright SI, Le QH, Schoen DJ, Bureau TE (2001). Population dynamics of an Ac-liketransposable element in self- and cross-pollinating *Arabidopsis*. *Genetics* **158**: 1279–1288.

Editor's suggested reading

Kejnovsky E, Hobza R, Cermak T, Kubat Z, Vyskot B (2009). The role of repetitive DNA in structure and evolution of sex chromosomes in plants. *Heredity* **102**: 533–541.

Kawakami T, Strakosh SC, Zhen Y, Ungerer MC (2010). Different scales of Ty1/copia-like retro-transposon proliferation in the genomes of three diploid hybrid sunflower species. *Heredity* **104**: 341–350.

Li XQ (2009). Developmental and environmental variation in genomes. *Heredity* **102**: 323–329.

Ellegren H (2011). Zebra finch genome sequenced: The singing genome. *Heredity* **106**: 533–534.