Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Genetics & Development

Genome size diversity in angiosperms and its influence on gene space

Steven Dodsworth[1], Andrew R Leitch[1] and Ilia J Leitch[2]

Genome size varies c. 2400-fold in angiosperms (flowering plants), although the range of genome size is skewed towards small genomes, with a mean genome size of 1C = 5.7 Gb. One of the most crucial factors governing genome size in angiosperms is the relative amount and activity of repetitive elements. Recently, there have been new insights into how these repeats, previously discarded as 'junk' DNA, can have a significant impact on gene space (i.e. the part of the genome comprising all the genes and gene-related DNA). Here we review these new findings and explore in what ways genome size itself plays a role in influencing how repeats impact genome dynamics and gene space, including gene expression.

Addresses
[1] School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK
[2] Department of Comparative Plant and Fungal Biology, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3DS, UK

Corresponding author: Leitch, Ilia J (i.leitch@kew.org)

Current Opinion in Genetics & Development 2015, **35**:73–78

This review comes from a themed issue on **Genomes and evolution**

Edited by **Antonis Rokas** and **Pamela S Soltis**

http://dx.doi.org/10.1016/j.gde.2015.10.006

0959-437/Published by Elsevier Ltd.

Introduction

Large-scale comparative analyses of plant genome sizes (GS) available in the Plant DNA C-values database (www.data.kew.org/cvalues) have shown that angiosperms (flowering plants) are remarkable in their GS diversity. Not only do they have the largest range for any comparable eukaryotic group, varying c. 2400-fold (1C = 0.063–148.8 Gb), but they also include the largest eukaryotic genome so far recorded[3] [i.e. *Paris japonica*, 1] which is c. 950× larger than the genome of *Arabidopsis thaliana* (1C = 0.157 Gb). Nevertheless, the distribution of GS is strongly skewed towards small genomes, with the modal and mean values being just 1C = 0.6 Gb and 5.7 Gb, respectively [Figure 1]. There are two major drivers of this astonishing GS diversity: (i) polyploidy, or whole-genome duplication [3,4], causing, at least initially, step-wise increases in GS, and (ii) deviation in repeat copy numbers, that can either result in subtle or more dramatic GS changes [2].

Repetitive DNA sequences account for the majority of the genomic DNA in most plant species, occurring in a few to millions of copies [5]. As GS increases, so does the proportion of repetitive DNA, up to a certain point, after which degraded repeats that are difficult to classify represent a significant portion of what has been termed the genomic 'dark matter' [6]. By comparison, the size of the gene space probably remains relatively constant. In angiosperms the repeat types (i.e. (retro)transposable elements, (micro)satellite DNA, and truncated derivatives — see Glossary) can be fast evolving in absolute copy numbers and sequence, such that in species from many plant families there are reports of repeat element half-lives of 3–4 million years [species in Poaceae, Brassicaceae, Fabaceae, and Vitaceae, 7], near complete repeat turnover in the genome over timeframes of 5–10 million years [Solanaceae, 8], and repeat copy numbers changing GS two or three fold over just a few million years [Poaceae and Malvaceae, 9].

Changes in the number, location, and diversity of repeat sequences have a significant impact on gene space evolution [9]. Here we focus on recent insights into this dynamic interplay. We propose that an understanding of gene evolution and gene regulation requires a deep understanding of the genomic context of gene space, that is, the repeat landscape and genome architecture within which a gene is embedded. In addition, we explore the extent to which these processes operate at the upper end of the scale in terms of GS.

Influence of repeats on gene expression and function

It has been widely documented that the mobility and amplification of repeats, both satellite and transposable elements (TEs), can influence gene expression and function [reviewed in, e.g. 9–13], and, if left unchecked, will lead to increasing GS with potentially detrimental consequences on the phenotype [14]. To reduce the frequency of these processes, eukaryote genomes have evolved a variety of mechanisms to epigenetically silence repeat activity, including RNA-directed DNA methylation (RdDM; involving small interfering RNAs, siRNAs),

[3] Although larger genome sizes have been reported in some unicellular eukaryotes, their estimates are considered unreliable (see http://www.genomesize.com/statistics.php) as they have never been confirmed using appropriate techniques.

Glossary
Genome size: The amount of DNA in the nucleus. Usually this is given as a 1C-value that refers to the amount of DNA in the unreplicated gametic nucleus (units in Mb, Gb or pg; 1 pg = 978 Mb, thus 1 Gb ≈ 1 pg).
Polyploidy or whole-genome duplication: The presence of more than two genomes in the nucleus.
Gene space: The part of the genome comprising all the genes and gene-related DNA.
Repetitive DNA: Amongst the repeats, there are two major categories, tandem repeats (e.g. microsatellites, satellites, and ribosomal DNA) and dispersed repeats (comprising transposable elements (TE), including both DNA transposons and retroelements and their truncated and diverged derivatives).
Retroelements: These include (i) the LTR (long terminal repeat) retrotransposon families Ty3/*gypsy* and Ty1/*copia* which together usually account for the majority of angiosperm repetitive DNA [10,44] and (ii) non-LTR retroelements (LINEs and SINEs).
RdDM: RNA-directed DNA methylation: This is a mechanism involved in the silencing of repeats. It operates through RNA polymerase IV transcription of repeats, which generates small interfering RNAs (siRNAs). These are targeted back to the repeats where, through the activity of RNA polymerase V and other proteins, they trigger the methylation of cytosines and the recruitment of modified histones. Together this results in changes in chromatin conformation [15,16] and alters repeat activity.

maintenance methylation, and histone modifications [15,16]. Yet such silencing of repeats can have repercussions on adjacent gene domains as RdDM has been shown to 'seed' the spread of methylation into regions that were not originally targets of siRNAs.

Nevertheless, recent studies of different TE families across the whole genome in maize and *Arabidopsis* have shown that spreading of methylation is not a characteristic of all TEs [17,18••], hence not all TEs impact adjacent genes (within c. 1 kb) in this way. West *et al.* [18••] have also shown that there are differences in the amount of TE methylation spreading between species, with more spreading into flanking regions in maize than *Arabidopsis*. However, in maize, the boundaries between genes and TEs are marked by elevated cytosine methylation at CHH motifs (forming CHH methylation islands, triggered by the activities of RdDM), resulting in altered chromatin conformation. This may act to preferentially inhibit TE amplification whilst enabling gene expression [19].

Recently, it has become clear that as well as these *cis*-effects, siRNAs produced following activation of TEs can also regulate the expression of *Arabidopsis* genes in *trans* [20]. Given the relatively low number of genes that are targeted by siRNAs in *Arabidopsis* (30%) compared with rice (80%), which has a larger genome [21], this raises the question as to whether the impact of such *trans*-effects of siRNAs on gene expression may become increasingly complex as GS increases.

In addition to these repeat-silencing effects, specific structural features of LTR retrotransposons make them particularly likely to influence the expression of nearby genes. Promoter/regulatory sequences at both LTRs allow 3′ LTRs to drive bleed-through transcription, which can extend into neighbouring sequences. Indeed, it is likely that the average plant genome has hundreds or thousands of genes controlled by regulators originally derived from TEs [10]. Depending on both the position of insertion in gene regions, and the orientation of the LTR, this type of transcription can lead to multiple and antagonistic effects on gene expression [22]. Indeed, it is now becoming apparent that TEs enable fine-tuning of gene expression and can have an important regulatory role. This echoes original work on TEs in maize by Barbara McClintock, with her original name 'controlling elements' — 'The real point is control. The real secret of all of this is control. It is not transposition' [22,23].

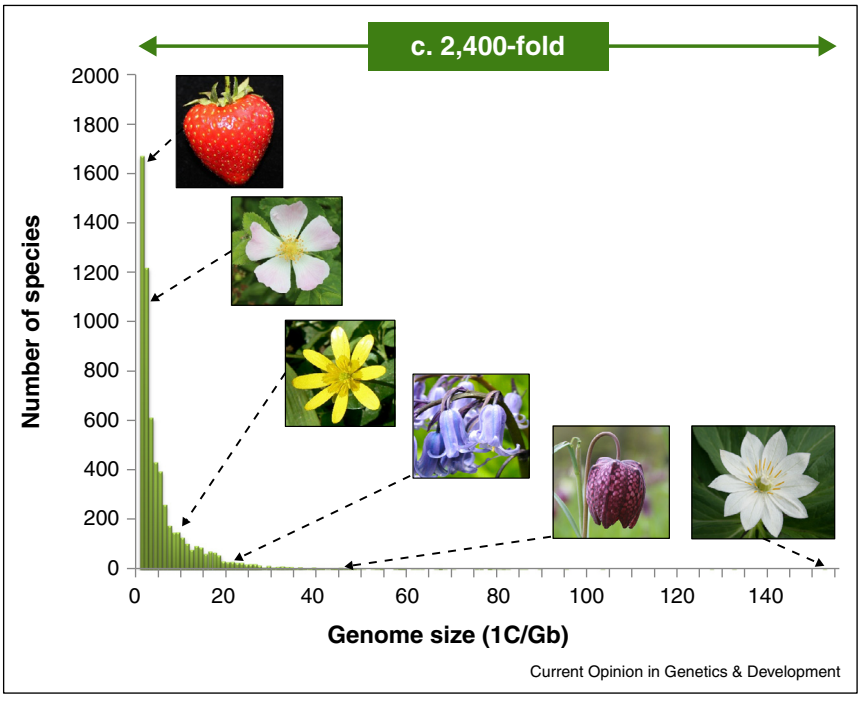
It is known that exaptation of TEs is important in enabling phenotypic plasticity in plants, producing variability in some important agricultural traits, and in responses to stress. The classic examples of red and white grape phenotypes and the blood-orange phenotype are each associated with insertion of TEs into upstream promoters [11]. More recently, a genome-wide association study (GWAS) of 368 maize inbred lines showed that a *CACTA*-like TE within the promoter of the gene *ZmCCT* was associated with reduced photoperiod sensitivity in maize and reduced flowering time, and hence likely to be significant in its post-domestication expansion to temperate regions [24]. Evidence is also growing that shows TEs have been exapted in plant defence responses. For example, knock out of a Ty1/*copia* retrotransposon, *ATCOPIA4*, adjacent to *RPP5* genes was shown to result in increased susceptibility to downy mildew in *Arabidopsis* [25]. Furthermore, in *Arabidopsis* the study of cytosine methylation mutants with increased susceptibility to *Fusarium oxysporum* showed that a significant fraction of genes that are differentially expressed are also associated with reduced CHH methylation in upstream promoters carrying TEs [26••]. Such results suggest that the TEs and their regulation are involved in disease resistance. There is also likely to be a wider role of TEs in regulating the nature of the transcriptome response to abiotic stress. For example, an analysis of the maize inbred line B73 under temperature and UV stress has shown that genes which are up-regulated are significantly more likely to be closely associated with 20 TE families, whilst genes that are down-regulated are frequently associated with a further three TE families [27••].

Together these studies highlight the increasingly diverse ways in which repeats can impact how genes are regulated and expressed in response to the environment.

Influence of repeats on the evolution of gene space

Most flowering plant lineages have undergone multiple rounds of polyploidy in their ancestry, a process that is

Figure 1



Histogram showing the distribution of genome sizes (GS) in 7542 angiosperms using data taken from the Plant DNA C-values database (www.data.kew.org/cvalues). Note the strong skew towards small GS, with a mean 1C-value of 5.7 Gb and a modal 1C-value of 0.6 Gb. Images of representative plants from left to right are: *Fragaria* × *ananassa* 1C/0.60 Gb; *Rosa canina* 1C/1.39 Gb; *Ranunculus ficaria* 1C/9.12 Gb; *Hyacinthoides non-scripta* 1C/20.73 Gb; *Fritillaria meleagris* 1C/46.26 Gb; *Paris japonica* 1C/148.8 Gb (image of *P. japonica* from Alpsdake/Wikimedia).

ongoing in many lineages [28]. Consequently, supposedly diploid species are in fact palaeopolyploids. For example, the 'diploid' *A. thaliana* may be as high as 48-ploid, and 'tetraploid' cytotypes (96-ploid) exist in nature [28]. Recurring polyploidy has resulted in the evolution of large gene families, with genes frequently occurring in multiple syntenic blocks arranged in a co-linear order, the distribution of these blocks reflecting the lineage's history of chromosomal rearrangements. However, the fate of the duplicated gene copies themselves following polyploidy can differ. Some genes are resistant to losses post-polyploidy [i.e. gene balance hypothesis, 29], whilst others are free to drift back to lower 'diploid' copy numbers.

It is clear that for many genes, copy numbers can diverge quickly, with a long-term tendency to reduce copy number for all gene duplicates that do not have selection pressures maintaining them. One mechanistic driver of that reduction in copy number is likely to be the proximity of LTRs and the frequency of unequal recombination, which leads to the deletion of sequences between adjacent LTRs. Recombination between TEs and indeed any adjacent repeats can have multiple effects. First, recombination-based removal from the genome limits the impact of repeats on gene expression [e.g. methylated LTR retrotransposons in rice are preferentially removed from regions surrounding genes, 13,30]. Second, recombination between adjacent repeats can also involve the deletion or duplication of intervening genes, giving rise to copy number and presence/absence variants.

These structural variants (SVs) generate genomic complexity that differentiates species and populations/lines within species [31]. SVs can occur at an astonishingly high frequency. In maize it has been shown that SVs influence thousands of genes [e.g. ~83% of 8681 transcripts were only expressed in subsets of 503 diverse inbred lines, 32]. Indeed, the maize reference genome, B73 [33], carries only c. 70% of all the low-copy sequences identified in 27 diverse maize accessions [34].

Thus, removal or amplification of repeats and genes generates considerable structural variation upon which selection can act. Over time, in *Arabidopsis*, Gaut *et al.* [35] suggest that the dynamics of gene duplication via ancestral polyploidy, and losses and gains of genes through recombination has resulted in a genome whereby surviving duplicates derived from each mechanism occur in similar numbers. Furthermore, differential selection pressures on duplicates lead to genome fractionation,

whereby regions of the genome become enriched for genes resistant to post-duplication losses [36].

Influence of repeats on gene space across the range of plant GSs

Comparative mapping in a range of grass species differing in GS has revealed that the evolution of gene space in terms of organization, duplication rates, and number of genes reflects the GS of the species. An accelerated rate of evolution was found in the larger genomes of species in the Triticeae tribe of the grass family compared with the smaller genomes of *Oryza sativa*, *Brachypodium distachyon*, and *Sorghum bicolor* [37–39].

Given these observations, and extrapolating these data to the biggest angiosperm genomes which are several-fold larger than the grasses studied above, one might expect that gene regulation and expression networks in species with giant genomes would be in utter chaos. Clearly this is not the case. One reason could be that repeats do not accumulate randomly in the genome and/or their removal is not random. Consequently, with increasing genome size, repeats can accumulate in ever increasing blocks, pushing genes into islands in an ever more partitioned genome [40]. It is also possible that genomic and epigenetic processes, influencing chromatin conformation, gene expression, and recombination, are not operating in the same way across the range of GSs encountered in angiosperms.

The GS of an individual represents the balance between processes that amplify and delete sequences, for example, polyploidy, (retro)transposition, illegitimate and unequal recombination, and non-homologous end joining in DNA repair. The epigenetic silencing of repeats described above, whilst it may indeed reduce the frequency of, for example, retrotransposition, will also influence recombination and DNA repair pathways because the chromatin may be heterochromatinised and hence rendered less accessible. In particular it will influence the balance between homologous and non-homologous DNA repair and the frequency of DNA deletion through unequal recombination and illegitimate recombination — both of which have been shown to contribute to genome downsizing. Indeed Fedoroff [41] stated, 'I contend that it was precisely the evolution of prokaryotic mechanisms to regulate homologous recombination within the eukaryotic genome that made it possible for genomes to grow'. Thus, it can be argued that large, repeat-rich genomes become locked down by epigenetic silencing, reducing the frequency of repeat removal [42••].

In support of this hypothesis, Kelly *et al.* [42••] showed that in *Fritillaria*, the genus with the largest known GSs amongst diploid plants, the repeat profile is not dominated by a few, rather homogeneous repeats that make up a substantial proportion of the genome, as is typical of species with small genomes. This phenomenon is also seen in some species of amphibians and lung fish which also have very large genomes [43]. Instead, in *Fritillaria*, at least, there is a plethora of diverse repeats, many in large copy number, but each accounting for only a small proportion of the whole genome. The data also indicate that the repeats are ancient, suggesting that they are not being deleted and turned over, but rather are slowly accumulating over time.

Furthermore, the dynamic means that they are free to diverge through accumulation of mutations, becoming low-abundance unique and inactive DNA that represents substantial proportions of the genome [perhaps up to 40–50% in very large genomes, 42••]. A consequence of that erosion of repeats is that as GS increases, the genes may be found in an accumulating sea of non-repetitive DNA, despite the overall huge GS. Thus, paradoxically gene space may be less vulnerable to the effects of repeats than in species with small genomes. However, that stability may itself come at an evolutionary cost, as it is the variation generated by repeat dynamics that makes up a significant amount of genetic variation upon which selection can act.

Acknowledgements

We thank NERC for funding (NE/G01724/1) and a PhD studentship to SD.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Pellicer J, Fay MF, Leitch IJ: **The largest eukaryotic genome of them all?** *Bot J Linnean Soc* 2010, **164**:10-15.

2. Leitch IJ, Leitch AR: **Genome size diversity and evolution in land plants**. In *Plant Genome Diversity, Vol 2, Physical Structure, Behaviour and Evolution of Plant Genomes.* Edited by Leitch IJ, Greilhuber J, Doležel J, Wendel JF. Springer-Verlag; 2013:307-322.

3. Leitch AR, Leitch IJ: **Ecological and genetic factors linked to contrasting genome dynamics in seed plants**. *New Phytol* 2012, **194**:629-646.

4. Soltis PS, Soltis DE: **The role of hybridization in plant speciation**. *Annu Rev Plant Biol* 2009, **60**:561-588.

5. Schmidt T, Heslop-Harrison JS: **Genomes, genes and junk: the large-scale organization of plant chromosomes**. *Trends Plant Sci* 1998, **3**:195-199.

6. Maumus F, Quesneville H: **Deep investigation of *Arabidopsis thaliana* junk DNA reveals a continuum between repetitive elements and genomic dark matter**. *PLoS ONE* 2014, **9**:e94101.

7. El Baidouri M, Panaud O: **Comparative genomic paleontology across plant kingdom reveals the dynamics of TE-driven genome evolution**. *Genome Biol Evol* 2013, **5**:954-965.

8. Lim KY, Kovarik A, Matyasek R, Chase MW, Clarkson JJ, Grandbastien MA, Leitch AR: **Sequence of events leading to near-complete genome turnover in allopolyploid *Nicotiana* within five million years**. *New Phytol* 2007, **175**:756-763.

9. Slotkin RK, Nuthikattu S, Jiang N: **The impact of transposable elements on gene and genome evolution**. In *Plant Genome Diversity, Vol 1, Plant Genomes, Their Residents, and Their*

Evolutionary Dynamics. Edited by Wendel JF, Greilhuber J, Doležel J, Leitch IJ. Springer-Verlag; 2012:35-58.

10. Bennetzen JL, Wang H: **The contributions of transposable elements to the structure, function, and evolution of plant genomes**. *Annu Rev Plant Biol* 2014, **65**:505-530.

11. Lisch D: **How important are transposons for plant evolution?** *Nat Rev Genetics* 2013, **14**:49-61.

12. Pezer Z, Brajkovic J, Feliciello I, Ugarkovic D: **Satellite DNA-mediated effects on genome regulation**. In *Repetitive DNA*, 7. Edited by Garrido-Ramos MA. Karger, Basel; 2012:153-169.

13. Mirouze M, Vitte C: **Transposable elements, a treasure trove to decipher epigenetic variation: insights from Arabidopsis and crop epigenomes**. *J Exp Bot* 2014, **65**:2801-2812.

14. Greilhuber J, Leitch IJ: **Genome size and the phenotype**. In *Plant Genome Diversity, Vol 2, Physical Structure, Behaviour and Evolution of Plant Genomes.* Edited by Leitch IJ, Greilhuber J, Doležel J, Wendel JF. Springer-Verlag; 2013:323-344.

15. Matzke MA, Mosher RA: **RNA-directed DNA methylation: an epigenetic pathway of increasing complexity**. *Nat Rev Genetics* 2014, **15**:394-408.

16. Matzke MA, Kanno T, Matzke AJM: **RNA-directed DNA methylation: the evolution of a complex epigenetic pathway in flowering plants**. *Annu Rev Plant Biol* 2015, **66**:243-267.

17. Eichten SR, Ellis NA, Makarevitch I, Yeh C-T, Gent JI, Guo L, Mcginnis KM, Zhang X, Schnable PS, Vaughn MW *et al.*: **Spreading of heterochromatin is limited to specific families of maize retrotransposons**. *PLOS Genetics* 2012, **8**:e1003127.

18. •• West PT, Li Q, Ji L, Eichten SR, Song J, Vaughn MW, Schmitz RJ, Springer NM: **Genomic distribution of h3k9me2 and DNA methylation in a maize genome**. *PLoS ONE* 2014, **9**:e105267.
Patterns of genome-wide DNA methylation and H3K9me2 were investigated in maize B73 and compared with *Arabidopsis*. Flanking regions of genes had elevated methylation marks in maize, that is, methylation spreading. This is probably due to a more dynamic chromosomal context in maize with transposons being more interspersed with genes.

19. Gent JI, Ellis NA, Guo L, Harkess AE, Yao Y, Zhang X, Dawe RK: **CHH islands: de novo DNA methylation in near-gene chromatin regulation in maize**. *Genome Res* 2013, **23**:628-637.

20. McCue AD, Nuthikattu S, Slotkin RK: **Genome-wide identification of genes regulated in trans by transposable element small interfering RNAs**. *RNA Biol* 2013, **10**:1379-1395.

21. Beló A, Nobuta K, Venu RC, Janardhanan P, Wang G-l, Meyers B: **Transposable element regulation in rice and Arabidopsis: diverse patterns of active expression and siRNA-mediated silencing**. *Trop Plant Biol* 2008, **1**:72-84.

22. Grandbastien M-A: **LTR retrotransposons, handy hitchhikers of plant regulation and stress response**. *Biochim Biophys Acta* 2015, **1849**:403-416.

23. Comfort NC: **''The real point is control'': the reception of Barbara McClintock's controlling elements**. *J History Biol* 1999, **32**:133-162.

24. Yang Q, Li Z, Li W, Ku L, Wang C, Ye J, Li K, Yang N, Li Y, Zhong T *et al.*: **CACTA-like transposable element in ZmCCT attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize**. *Proc Natl Acad Sci* 2013, **110**:16969-16974.

25. Wang Y-H, Warren JT Jr: **Mutations in retrotransposon *AtCOPIA4* compromises resistance to *Hyaloperonospora parasitica* in *Arabidopsis thaliana***. *Genetics Mol Biol* 2010, **33**:135-140.

26. •• Le T-N, Schumann U, Smith N, Tiwari S, Au P, Zhu Q-H, Taylor J, Kazan K, Llewellyn D, Zhang R *et al.*: **DNA demethylases target promoter transposable elements to positively regulate stress responsive genes in Arabidopsis**. *Genome Biol* 2014, **15**:458.
Using a triple DNA demethylase mutant, *rdd*, of *Arabidopsis*, the authors show that a significant proportion of the 348 differentially expressed genes examined in the mutant are stress response genes that are not only down-regulated but also enriched for short TEs in their promoters. These TEs and their surrounding sequences show a general reduction in CHH methylation, suggesting interplay between RdDM and DNA demethylases in regulation of stress response genes via TEs in promoter regions.

27. •• Makarevitch I, Waters AJ, West PT, Stitzer M, Hirsch CN, Ross-Ibarra J, Springer NM: **Transposable elements contribute to activation of maize genes in response to abiotic stress**. *PLOS Genet* 2015, **11**:e1004915.
This paper investigated TE associations with nearby genes in a genome-wide manner, showing a significant association of 20 TE families with up-regulated genes during abiotic stress in maize, thereby providing stress-responsive enhancer-like functions. They also provide evidence that TE insertions create allelic diversity influencing stress responses.

28. Renny-Byfield S, Wendel JF: **Doubling down on genomes: polyploidy and crop plants**. *Am J Bot* 2014, **101**:1711-1725.

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

30. Vonholdt BM, Takuno S, Gaut BS: **Recent retrotransposon insertions are methylated and phylogenetically clustered in Japonica rice (*Oryza sativa* spp. *japonica*)**. *Mol Biol Evol* 2012, **29**:3193-3203.

31. Marroni F, Pinosio S, Morgante M: **Structural variation and genome complexity: is dispensable really dispensable?** *Curr Opin Plant Biol* 2014, **18**:31-36.

32. •• Hirsch CN, Foerster JM, Johnson JM, Sekhon RS, Muttoni G, Vaillancourt B, Penagaricano F, Lindquist E, Pedraza MA, Barry K *et al.*: **Insights into the maize pan-genome and pan-transcriptome**. *Plant Cell* 2014, **26**:121-135.
The authors investigate 503 inbred maize lines to characterise the maize pan-genome, showing that 16.4% of 8681 representative transcript assemblies are expressed in all lines, and that 82.7% are only expressed in subsets of the lines.

33. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA *et al.*: **The B73 maize genome: complexity, diversity, and dynamics**. *Science* 2009, **326**:1112-1115.

34. Gore MA, Chia J-M, Elshire RJ, Sun Q, Ersoz ES, Hurwitz BL, Peiffer JA, McMullen MD, Grills GS, Ross-Ibarra J *et al.*: **A first-generation haplotype map of maize**. *Science* 2009, **326**:1115-1117.

35. Gaut BS, Wright SI, Rizzon C, Dvorak J, Anderson LK: **Opinion — recombination: an underappreciated factor in the evolution of plant genomes**. *Nat Rev Genetics* 2007, **8**:77-84.

36. Freeling M: **Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition**. *Annu Rev Plant Biol* 2009, **60**:433-453.

37. Massa AN, Wanjugi H, Deal KR, O'Brien K, You FM, Maiti R, Chan AP, Gu YQ, Luo MC, Anderson OD *et al.*: **Gene space dynamics during the evolution of *Aegilops tauschii*, *Brachypodium distachyon*, *Oryza sativa*, and *Sorghum bicolor* genomes**. *Mol Biol Evol* 2011, **28**:2537-2547.

38. Luo MC, Deal KR, Akhunov ED, Akhunova AR, Anderson OD, Anderson JA, Blake N, Clegg MT, Coleman-Derr D, Conley EJ *et al.*: **Genome comparisons reveal a dominant mechanism of chromosome number reduction in grasses and accelerated genome evolution in Triticeae**. *Proc Natl Acad Sci* 2009, **106**:15780-15785.

39. Choulet F, Wicker T, Rustenholz C, Paux E, Salse J, Leroy P, Schlub S, Le Paslier M-C, Magdelenat G, Gonthier C *et al.*: **Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces**. *Plant Cell* 2010, **22**:1686-1701.

40. Gottlieb A, Müller H-G, Massa AN, Wanjugi H, Deal KR, You FM, Xu X, Gu YQ, Luo M-C, Anderson OD *et al.*: **Insular organization of gene space in grass genomes**. *PlOS ONE* 2013, **8**:e54101.

41. Fedoroff NV: **Transposable elements, epigenetics, and genome evolution**. *Science* 2012, **338**:758-767.

42. •• Kelly LJ, Renny-Byfield S, Pellicer J, Macas J, Novák P, Neumann P, Lysak MA, Day PD, Berger M, Fay MF *et al.*: **Analysis of the giant genomes of *Fritillaria* (Liliaceae) indicates that a**

lack of DNA removal characterizes extreme expansions in genome size. *New Phytol* 2015, **208**:596-607.
These authors investigate the nature of repeats in the angiosperm genus *Fritillaria*, which includes diploids ranging in GS by 30 Gb. They show that the genome is composed of many diverse repeats, none of which constitutes a substantial proportion of the genome. The data are consistent with a hypothesis that failure to remove DNA rather than runaway repeat expansion causes GS enlargement.

43. Metcalfe CJ, Casane D: **Accommodating the load — the transposable element content of very large genomes**. *Mobile Genetic Elements* 2013, **3**:e24775.

44. Kejnovsky E, Hawkins J, Feschotte C: **Plant transposable elements: biology and evolution**. In *Plant Genome Diversity 1.* Edited by Greilhuber J, Dolezel J, Leitch IJ. Springer-Velag; 2012.