

Transposable elements and the plant pan-genomes

Michele Morgante^{1,2}, Emanuele De Paoli¹ and Slobodanka Radovic¹

The comparative sequencing of several grass genomes has revealed that transposable elements are largely responsible for extensive variation in both intergenic and local genic content, not only between closely related species but also among individuals within a species. These observations indicate that a single genome sequence might not reflect the entire genomic complement of a species, and prompted us to introduce the concept of the plant pan-genome, which includes core genomic features that are common to all individuals and a dispensable genome composed of partially shared and/or non-shared DNA sequence elements. Uncovering the intriguing nature of the dispensable genome, namely its composition, origin and function, represents a step forward towards an understanding of the processes that generate genetic diversity and phenotypic variation. The developing view of transcriptional regulation as a complex and modular system, in which long-range interactions and the involvement of transposable elements are frequently observed, lends support to the possibility of an important functional role for the dispensable genome and could make it less dispensable than previously thought.

Addresses

¹ Dipartimento di Scienze Agrarie ed Ambientali, Università di Udine, Via delle Scienze 208, I-33100 Udine, Italy

² Istituto di Genomica Applicata, Parco Scientifico e Tecnologico di Udine, Via Linussio 51, 33100 Udine, Italy

Corresponding author: Morgante, Michele (michele.morgante@uniud.it)

Current Opinion in Plant Biology 2007, **10**:149–155

This review comes from a themed issue on
Genome studies and molecular genetics
Edited by Stefan Jansson and Edward S Buckler

Available online 14th February 2007

1369-5266/\$ – see front matter

© 2007 Elsevier Ltd. All rights reserved.

DOI 10.1016/j.pbi.2007.02.001

Introduction

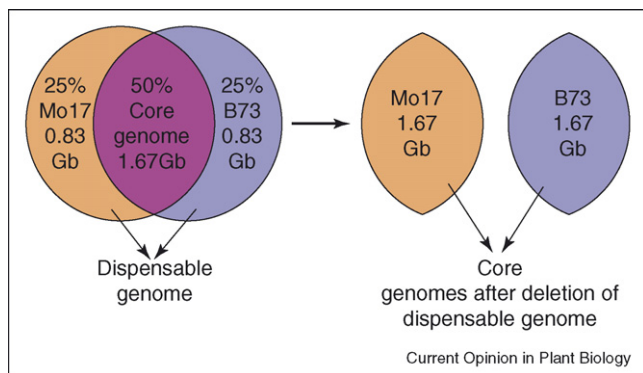
With the advent of high-throughput re-sequencing technologies that are based either on chip hybridization [1] or on sequencing by synthesis (SBS) [2,3], we have entered an exciting era in which we can finally learn what differences are found among individuals within a species at the DNA sequence level. Recent data obtained from different plant species have shown us how plastic, dynamic and variable plant genomes are. Transposable element movement is largely responsible for variation both in intergenic region sequence content and in local genic content. Both class I (long terminal repeat [LTR]-retrotransposons) and class II

(DNA transposons of different superfamilies) transposons contribute to sometimes dramatic differences in local sequence content among individuals belonging to the same species. A comparison of four randomly chosen genomic regions between the maize inbred lines B73 and Mo17 revealed that, on average, only 50% of the sequences are shared. Approximately 25% of the sequences were observed in a homologous location in one of the inbred lines but not in the other [4]. Similar, but less dramatic, differences have been observed in rice [5] and barley [6]. These observations have prompted us to borrow the concept of the pan-genome, which has been proposed for bacterial species [7^{*}], to describe the developing view of genomic variation within plant species. In this review, we use maize as an example to describe the contribution of transposable elements to the creation of the pan-genome, and discuss the implications of pan-genome structure for our understanding of the genetic bases of phenotypic variation.

The pan-genome concept: a core and a dispensable genome

The comparison of the genomic sequences of eight strains of the bacterial species *Streptococcus agalactiae* [7^{*}] revealed that a bacterial species can be best described by its ‘pan-genome’ (from the Greek word *παν*, meaning whole). The pan-genome includes a core genome containing genes that are present in all strains and a dispensable genome composed of partially shared and strain-specific DNA sequence elements. Unique genes were detected in each of the eight sequenced genomes, and mathematical modelling indicates that new genes will still be found after sequencing many more strains. Thus, the genomes of multiple, independent isolates are required to understand the global complexity of bacterial species. We propose that the same concept of the pan-genome be used to describe the genome of plant species such as maize. In the two previously mentioned inbred lines, taking the estimates from the cited paper [4], the pan genome (Figure 1) would comprise a core genome representing the 50% of the genome that is shared between the two lines (corresponding to a size of 1.67 Gb, if we assume an approximate total genome size for each of the lines of 2.50 Gb) and a dispensable genome of the same total size that is equally distributed among the two lines. The core genome comprises both single-copy sequences (including most if not all genes) and transposable elements that are found among all individuals in a certain genomic location. The dispensable genome is made up mostly of transposable elements of different types that, although present in multiple copies in each individual, can be found in a specific location only in some of them. A gene-like fraction can also be found in the dispensable

Figure 1



A pan-genome view of the maize genome as defined by comparison of sequenced genomic regions in the B73 and Mo17 inbred lines. The relative and absolute sizes of the core and the dispensable genome are obtained from data provided in [4]. The results of an experiment in which the dispensable genome is selectively deleted are shown on the right. The different colours used for the genomes of the two lines indicate that while the structure of the core genome is the same for the two lines, allelic variations that are due to point mutations (i.e. single nucleotide polymorphisms [SNPs]) can still be observed.

maize genome. If the core genome is what we normally think of as a genome of a species, what remains to be determined is the role of the dispensable maize genome, namely its composition, origin and functional role.

The dispensable genome: origin and composition

The comparison described above of four orthologous genomic regions between two US maize inbred lines, representing a heterotic pattern used in agriculture, showed that approximately 50% of the sequences are not shared between the two lines [4]. Similarly, sequence diversity ranged from 25% to 84% in a horizontal comparison of the *bronze* (*bz*) genomic region among eight maize cultivars, with the physical size of the region varying between 52 and 159 kb [8^{••}]. In both studies, the differences in intergenic regions were accounted for by the transposition of several different families of retroelements. These retroelements are present in specific inbred lines, and have been inserted significantly more recently than the shared retroelements. Discrepancies in gene content resulted from insertions of complete or truncated genes that, unlike the shared genes, could not be found in the orthologous rice regions, suggesting their absence from the ancestral genome [4]. Similar observations have been made in barley in an approximately 300 kb region surrounding the *Rph7* leaf rust disease resistance locus, which was sequenced from two different cultivars [6]. Furthermore, two rice subspecies, *japonica* and *indica*, presented extensive differences in intergenic regions, with only 72% of the sequences being collinear [5]. These and other results

have paved the way to a new scenario of genome plasticity, in which genomic regions of common ancestry appear to have evolved very recently into a mosaic of syntenic blocks that have independently diverged between grass species [9[•]]. Different blocks have undergone specific processes of expansion and contraction in both genic and intergenic space, owing to differential insertions of repetitive elements and episodes of gene mobilization, which sometimes resulted in deep allelic differences even at the intraspecific level. Interestingly, the transposable element (TE) insertion polymorphisms observed by means of transposon display between two ecotypes of *Lotus japonicus* suggest that exceptions to microcolinearity also occur in dicotyledons [10].

What is evident is that the large amount of genomic variation in grasses and the occurrence of non-shared ('dispensable') genomic features can be ascribed to the very young age of their extant repetitive component. LTR-retrotransposons have undergone independent amplification in distinct lineages within single plant genera, i.e. at the species level, over shorter periods of time than previously supposed. In maize, the dating of retrotransposon insertions revealed that the majority of events have taken place within the past 400 000 years [4]. It is likely that there has not been enough time for many of the new insertions to be either eliminated from or fixed in the gene pool, either through genetic drift or through selection, so that they appear today as intraspecific polymorphisms. Episodes of recent (within the past million years) and lineage-specific amplification of LTR-retrotransposon have also occurred in the rice genus, leading to one case of genomic obesity in *Oryza australiensis* [11[•]]. In addition, the 130-kb intergenic region that is flanked by the leaf rust disease resistance gene *Lr10* and a second resistance gene analog (*RGA2*) diverged by more than 70% between diploid and tetraploid wheat, in part because of transposition activity that has occurred during the past two million years [12]. Latest data from cotton [13[•]] and the legume species *Vicia pannonica* [14], *Lotus japonicus* [10] and *Medicago truncatula* [15] show that this feature is widely recurrent outside of the grass family.

Recent TE mobilization also accounts for the intraspecific differences in gene content reported in maize and barley. One mechanism is based on the insertional activity of *helitrons*, a class of DNA transposons, which appear to move by a copy-and-paste strategy that involves rolling-circle replication [16]. In maize, non-autonomous *helitrons* acquire and carry fragments of genes from different locations in the host genome through a duplicative mechanism, referred to as 'transduplication', which preserves the exon-intron structure of the gene fragments that are involved [17–20].

Transduplication capability is observed in other class II DNA transposon families. *Mutator*-like elements, namely

Pack-MULEs in rice [21,22], *Arabidopsis* [23**] and lotus [10] and the newly identified TA-flanked transposons (TAFTs) in maize [8**], appear to capture and transport host DNA fragments. The presence of a non-autonomous element of the CACTA superfamily that carries gene fragments in soybean [24] extends this ability to three different DNA transposon superfamilies, and new ones might be uncovered with further sequencing. Last, a different mechanism of gene mobility is mediated by retrotransposons. Here, a peculiar replication strategy, involving an RNA intermediate, is supposed to facilitate the fortuitous reverse-transcription of spliced host mRNAs and their insertion (retroposition) into new genomic positions to form intron-less retrogenes, some of which might be functional [25]. Retroposition is responsible for the generation of large numbers of retrogenes in rice [26] and has also been observed in *Arabidopsis*, in which 69 retrosequences have been identified [27]. Unexpectedly, intact genes, probably transferred by a not-yet-elucidated cut-and-paste mechanism, have also been found within LTR-retrotransposons in maize [28], confirming the unpredicted wealth of processes by which TE can mediate extensive sequence rearrangements and yield diversification of genomes.

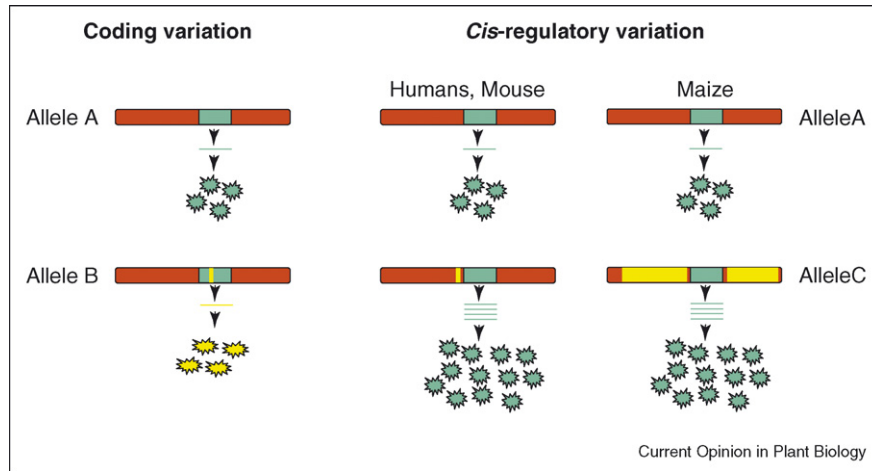
Is the dispensable genome really dispensable?

The inevitable question of whether the dispensable genome contributes to phenotypic variation emerges when its significant size and its composition are taken into consideration. The best experimental approach to answering this question would be to remove the dispensable genome fraction from each of the two maize lines that we have previously used to describe the pan-genome concept, leaving them with just their core genomes (Figure 1; see [29] for a small-scale but elegant example of such an experiment in mouse). At this point, we could use the DNA-stripped-down lines to test if the dispensable genome is important in determining phenotype, that is, whether each of the two lines look like their genome-obese counterparts. We could also test if the dispensable genome is important in determining phenotypic variation among maize lines, that is, if the two lines look more similar or more different than before the DNA reduction. In reality, such a large-scale DNA deletion experiment is not feasible, and so we must resort to circumstantial evidence to try to address the role of the dispensable genome in determining phenotypic variation. For decades, repetitive elements have been referred to as 'selfish' or even 'junk' DNA that expands by cloning itself in the host genome, contributing to the phenotype merely by insertional mutagenesis. Just recently, these simplistic and self-centred concepts of the major component of eukaryotic genomes have been reconsidered. We have begun to recognize repetitive elements as natural molecular tools that have shaped the organization, structure, and function of genes and genomes throughout their evolution [30].

The latest findings in flowering plants have shown that transposable elements recurrently duplicate and move cellular gene sequences from one location to another through either transduplication or retroposition. One could hypothesize that the movement of genes into a different chromosomal context could lead to a novel regulation of an existing gene and/or provide redundancy of a crucial function. What is more, through 'exon shuffling' whereby fragments of genes are fused together, transduplication and retroposition provide a potentially rich mechanism for the evolution of new genes. Indeed, thousands of duplicated fragments appear to be a part of the transcriptome. This, however, does not necessarily indicate imminent functionality of these fragments. Closer inspection of the fragments carried by Pack-MULES and *helitrons* revealed that the vast majority represent pseudogenes. Still, there are rare but potentially important examples demonstrating that transduplicates can retain protein-coding capacity and evolve novel functions. For example, the *Arabidopsis* *KAONASHI* (*KI*) family of MULEs carry an apparently functional ubiquitin-like protein-specific protease (ULP)-like gene [23**] and a maize non-autonomous *helitron* contains an intact gene coding for a putative cytidine deaminase [20]. In the light of these discoveries, it might be more relevant to investigate whether transduplicates contribute to phenotypic variation by generating small interfering RNAs, which might participate in RNA-mediated silencing of the host genes from which the fragments were duplicated, and which therefore potentially represent *trans*-acting regulatory factors [31]. LTR-mediated retroposition seems to make a greater contribution than transduplication to the evolution of new protein functions. A survey of the rice genome identified 1235 primary retrogenes [26], which for the most part were not only functional but also have recruited nearby exons and regulatory sequences and have reincarnated into a novel chimeric genes.

Cellular genes that are nested in LTR-retrotransposons can show expression patterns that differ from those of their non-nested paralogs, implying that they have distinct roles and distinct regulation [28]. It is not known whether this difference in expression is based on the regulatory elements of the retrotransposon that flanks the gene. However, an increasing body of evidence, especially from mammals [32–35] and *Drosophila* [36*], suggests that TEs that are located in or near genes could affect their expression by providing transcriptional regulatory signals and through epigenetic silencing [37]. In view of these findings, it is interesting to note that maize, which harbours an extraordinary level of polymorphism in intergenic regions due to recent insertions of different LTR-retrotransposons, shows a widespread variation in allelic-expression owing to *cis*-regulatory effects ([38,39]; Figure 2). The intriguing connection between these two observations is strengthened by well-studied cases of phenotypic variation: variation at the *teosinte branched1*

Figure 2



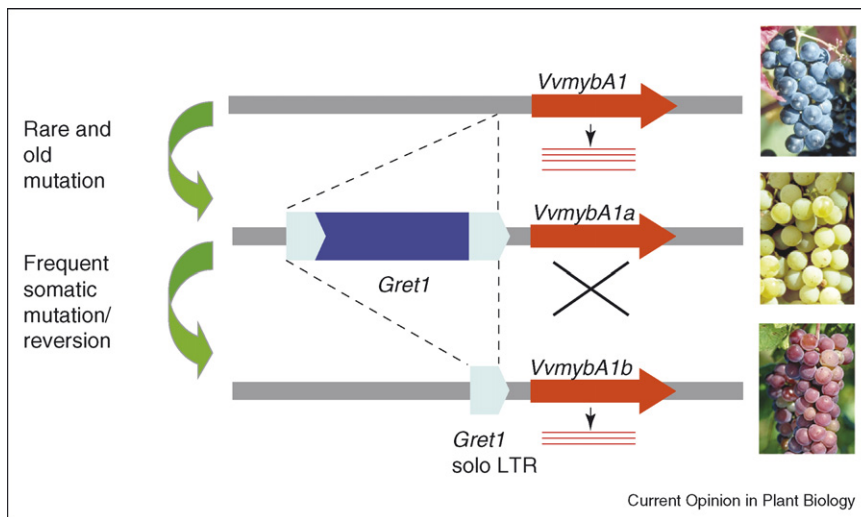
Coding variation versus *cis*-regulatory variation. The effects of different types of mutations (yellow lines/boxes) in a specific gene (green box) at the transcript and protein level. For *cis*-regulatory variation, we depict the very different types and extents of intergenic sequence variations that are commonly observed in species where *cis*-regulatory variation has been analysed (humans and mouse on one side, maize on the other).

(*tb1*) [40**] and *yellow1* (*y1*) [41] loci in maize has been attributed to intergenic polymorphisms that could be transposon indel polymorphisms rather than simple single nucleotide changes in specific regulatory elements. Retrotransposon methylation in rice [42] and *Arabidopsis* [43,44] has been found to modulate the activity of neighbouring genes, supporting another mechanism by which polymorphic retrotransposon insertions can affect *cis*-regulatory variation. Finally, a clear example of the phenotypic effects that can result from LTR-retrotransposon insertion, as well as from partial excision by intra-element

unequal recombination [45], is provided by mutational events in the *VvmybA1* gene in grape [46], which regulates the activity of the anthocyanin biosynthetic pathway that leads to berry pigmentation (Figure 3).

Until recently, it had been thought that plant TEs are dormant under normal conditions and become activated by genetic and environmental cues that potentially result in somaclonal variation. Discovery of effective transposition of the rice *hAT* superfamily transposon under natural growth conditions [47] tempts us to speculate that, in fact,

Figure 3



Model for the generation of variability in berry colour in grapevine (*Vitis vinifera*). The transcription factor gene *VvmybA1* regulates the anthocyanin biosynthetic genes, and its transcriptional activity is required to have anthocyanin production in the berry. We show here how the movement of an LTR-retrotransposon of the *Gret1* family (belonging to the *Gypsy* group) affects the transcription of *VvmybA1*, determining the creation of new phenotypes [46].

at least some of the genomic changes necessary for the uniqueness of individuals within a population might be driven by the activities of mobile elements.

Conclusions

Recent observations on DNA sequence variation in the maize genome imply that a complete description of the genomic structural variants and composition of the species can only be obtained by analysing more than one individual, and that we can apply the concept of the pan-genome, composed of a core and a dispensable genome. In maize, the dispensable genome is composed of sequences, i.e. TEs, that can have similar copies elsewhere in the genome but that are unique in their location and genomic context. How many different individuals will have to be sequenced before the pan-genome is completely described remains to be seen. Analysis of the *bz* region in eight maize lines [8**] reveals that although the core genome does not decrease dramatically in comparison to that defined by a pairwise comparison, for example, between the B73 and Mo17 inbred lines, the size of the dispensable genome increases as each one of the lines is added to the picture. Each line shows unique insertions of different sequence elements in their intergenic regions. The contribution of the dispensable genome to phenotypic variation within a species is still to be determined and is, of course, mainly dependent upon the possibility that different types of transposons contribute to the regulation of the neighbouring genes. In both animals and plants, the developing view of transcriptional regulation as a complex and modular system [48], in which long-range interactions are frequently observed [40**] and where transposable elements frequently provide regulatory elements [49,50], lends support to the possibility of an important functional role for the dispensable genome, and might make it less dispensable than previously thought. The creation of novel regulatory variants in plants by the movement of TEs is well demonstrated, for example, in determining berry colour in grape (Figure 3; [46]). The somatic mutations that give rise to red berries from white berry plants also provide an example of how TE movement can be responsible for the frequent creation of *de novo* mutations, which are then utilized in the breeding process, thus defining an important role for the pan-genome concept in applied science.

How likely we are to identify a model of genome variation similar to that observed in maize, i.e. a pan-genome model, in other plant species will depend upon different factors. Species that have larger genomes have a higher density of TEs and are therefore more likely to show variation that is due to their movement. The recent transpositional activity of the elements is also a very important factor: the observation that, in most angiosperms analysed to date, TEs have moved in very recent evolutionary times or are still moving encourages us to think that they might contribute to extant intraspecific sequence variation. Finally, the

mating system of the species could have an influence: outcrossing species usually have larger effective population sizes than selfing species, and therefore new TE insertions will have greater persistence in a polymorphic state, i.e. longer periods before they are either lost or fixed due to genetic drift, in large outcrossing populations than in selfing populations.

Acknowledgements

The preparation of this article was supported by an Italian Ministry of University and Research (MIUR), PRIN 2005 projects grant.

References and recommended reading

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

- of special interest
- of outstanding interest

1. Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR, Kautzer CR, Lee DH, Marjoribanks C, McDonough DP *et al.*: **Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21.** *Science* 2001, **294**:1719-1723.
2. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z *et al.*: **Genome sequencing in microfabricated high-density picolitre reactors.** *Nature* 2005, **437**:376-380.
3. Bentley DR: **Whole-genome re-sequencing.** *Curr Opin Genet Dev* 2006, **16**:545-552.
4. Brunner S, Fengler K, Morgante M, Tingey S, Rafalski A: **Evolution of DNA sequence nonhomologies among maize inbreds.** *Plant Cell* 2005, **17**:343-360.
5. Yu J, Wang J, Lin W, Li S, Li H, Zhou J, Ni P, Dong W, Hu S, Zeng C *et al.*: **The genomes of *Oryza sativa*: a history of duplications.** *PLoS Biol* 2005, **3**:e38.
6. Scherrer B, Isidore E, Klein P, Kim JS, Bellec A, Chalhoub B, Keller B, Feuillet C: **Large intraspecific haplotype variability at the *Rph7* locus results from rapid and recent divergence in the barley genome.** *Plant Cell* 2005, **17**:361-374.
7. Tettelin H, Massignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS *et al.*: **Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial 'pan-genome'.** *Proc Natl Acad Sci USA* 2005, **102**:13950-13955.
This paper introduces the concept of a pan-genome that includes a shared component common to all strains within the species and a component that includes all the sequence fragments that are found in a subset of strains only.
8. Wang Q, Dooner HK: **Remarkable variation in maize genome structure inferred from haplotype diversity at the *bz* locus.** *Proc Natl Acad Sci USA* 2006, **103**:17644-17649.
By describing the unprecedented variation in the *bz* genomic region observed in eight maize inbred lines, the authors show that a complete description of the maize pan-genome might require a very deep sequencing of the maize gene pool, especially if all of the structural variants of the intergenic regions have to be described.
9. Bruggmann R, Bharti AK, Gundlach H, Lai J, Young S, Pontaroli AC, Wei F, Haberer G, Fuks G, Du C *et al.*: **Uneven chromosome contraction and expansion in the maize genome.** *Genome Res* 2006, **16**:1241-1251.
This paper provides a picture of grass genome evolution in which different chromosomal regions have evolved into a mosaic of syntenic blocks. The differential expansion of these blocks is caused by the contraction of genic and intergenic space in combination with the addition of different combinations of repeat elements.
10. Holligan D, Zhang X, Jiang N, Pritham EJ, Wessler S: **The transposable element landscape of the model legume *Lotus japonicus*.** *Genetics* 2006, **174**:2215-2228.

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

This is the first study to provide direct evidence at the genome scale that the activity of LTR-retrotransposons can increase genome size to a very large extent over a short evolutionary period.

12. Isidore E, Scherrer B, Chalhoub B, Feuillet C, Keller B: **Ancient haplotypes resulting from extensive molecular rearrangements in the wheat A genome have been maintained in species of three different ploidy levels.** *Genome Res* 2005, **15**:526-536.
13. Hawkins JS, Kim H, Nason JD, Wing RA, Wendel JF: **Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*.** *Genome Res* 2006, **16**:1252-1261.

A key conclusion of this study is that variation in genome size within a single genus of plants reflects not only the differential amplification of diverse types of repetitive sequences but also that specific families within a repetitive sequence type proliferate differentially.

14. Neumann P, Koblikova A, Navratilova A, Macas J: **Significant expansion of *Vicia pannonica* genome size mediated by amplification of a single type of giant retroelement.** *Genetics* 2006, **173**:1047-1056.
15. Vitte C, Bennetzen JL: **Eukaryotic transposable elements and genome evolution special feature: analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution.** *Proc Natl Acad Sci USA* 2006, **103**:17638-17643.
16. Kapitonov VV, Jurka J: **Rolling-circle transposons in eukaryotes.** *Proc Natl Acad Sci USA* 2001, **98**:8714-8719.
17. Lai J, Li Y, Messing J, Dooner HK: **Gene movement by *Helitron* transposons contributes to the haplotype variability of maize.** *Proc Natl Acad Sci USA* 2005, **102**:9068-9073.
18. Morgante M, Brunner S, Pea G, Fengler K, Zuccolo A, Rafalski A: **Gene duplication and exon shuffling by *helitron*-like transposons generate intraspecies diversity in maize.** *Nat Genet* 2005, **37**:997-1002.
19. Brunner S, Pea G, Rafalski A: **Origins, genetic organization and transcription of a family of non-autonomous *helitron* elements in maize.** *Plant J* 2005, **43**:799-810.
20. Xu JH, Messing J: **Maize haplotype with a *helitron*-amplified cytidine deaminase gene copy.** *BMC Genet* 2006, **7**:52.
21. Jiang N, Bao Z, Zhang X, Eddy SR, Wessler SR: **Pack-MULE transposable elements mediate gene evolution in plants.** *Nature* 2004, **431**:569-573.
22. Ohtsu K, Hirano HY, Tsutsumi N, Hirai A, Nakazono M: ***Anaconda*, a new class of transposon belonging to the *Mu* superfamily, has diversified by acquiring host genes during rice evolution.** *Mol Genet Genomics* 2005, **274**:606-615.
23. Hoen DR, Park KC, Elrouby N, Yu Z, Mohabir N, Cowan RK, Bureau TE: **Transposon-mediated expansion and diversification of a family of ULP-like genes.** *Mol Biol Evol* 2006, **23**:1254-1268.

Genome-wide survey of MULE-mediated transduplication in *Arabidopsis* identified the *KI* family, which contains at least 97 potentially functional genes and pseudogenes that have strong similarity to peptidase C48. The authors examine the characteristics, phylogeny, conservation, age and expression patterns of *KI* transduplicates and their cellular paralogs.

24. Zabala G, Vodkin LO: **The *wp* mutation of *Glycine max* carries a gene-fragment-rich transposon of the CACTA superfamily.** *Plant Cell* 2005, **17**:2619-2632.
25. Weiner AM, Deininger PL, Efstratiadis A: **Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information.** *Annu Rev Biochem* 1986, **55**:631-661.
26. Wang W, Zheng H, Fan C, Li J, Shi J, Cai Z, Zhang G, Liu D, Zhang J, Yang S *et al.*: **High rate of chimeric gene origination by retroposition in plant genomes.** *Plant Cell* 2006, **18**:1791-1802.

27. Zhang Y, Wu Y, Liu Y, Han B: **Computational identification of 69 retroposons in *Arabidopsis*.** *Plant Physiol* 2005, **138**:935-948.

28. Du C, Swigonova Z, Messing J: **Retrotranspositions in orthologous regions of closely related grass species.** *BMC Evol Biol* 2006, **6**:62.
29. Nobrega MA, Zhu Y, Plajzer-Frick I, Afzal V, Rubin EM: **Megabase deletions of gene deserts result in viable mice.** *Nature* 2004, **431**:988-993.
30. Miller WJ, Capy P: **Applying mobile genetic elements for genome analysis and evolution.** *Mol Biotechnol* 2006, **33**:161-174.
31. Meister G, Tuschl T: **Mechanisms of gene silencing by double-stranded RNA.** *Nature* 2004, **431**:343-349.
32. Bejerano G, Lowe CB, Ahituv N, King B, Siepel A, Salama SR, Rubin EM, Kent WJ, Haussler D: **A distal enhancer and an ultraconserved exon are derived from a novel retroposon.** *Nature* 2006, **441**:87-90.
33. Thornburg BG, Gotea V, Makalowski W: **Transposable elements as a significant source of transcription regulating signals.** *Gene* 2006, **365**:104-110.
34. Polak P, Domany E: ***Alu* elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes.** *BMC Genomics* 2006, **7**:133.
35. Johnson R, Gamblin RJ, Ooi L, Bruce AW, Donaldson IJ, Westhead DR, Wood IC, Jackson RM, Buckley NJ: **Identification of the REST regulon reveals extensive transposable element-mediated binding site duplication.** *Nucleic Acids Res* 2006, **34**:3862-3877.

36. Savitskaya E, Melnikova L, Kostuchenko M, Kravchenko E, Pomerantseva E, Boikova T, Chetverina D, Parshikov A, Zobacheva P, Gracheva E *et al.*: **Study of long-distance functional interactions between *Su(Hw)* insulators that can regulate enhancer-promoter communication in *Drosophila melanogaster*.** *Mol Cell Biol* 2006, **26**:754-761.

The paper reports a functional analysis of the interplay of the two and three *Su(Hw)*, the most potent enhancer blocking insulator found in *gypsy* retrotransposons, and offer some inferences on the possible modes of insulator action and the ensuing regulatory schemes.

37. Reiss D, Mager DL: **Stochastic epigenetic silencing of retrotransposons: does stability come with age?** *Gene* 2006, in press.
38. Guo M, Rupe MA, Zinselmeier C, Habben J, Bowen BA, Smith OS: **Allelic variation of gene expression in maize hybrids.** *Plant Cell* 2004, **16**:1707-1716.
39. Stupar RM, Springer NM: ***Cis*-transcriptional variation in maize inbred lines B73 and Mo17 leads to additive expression patterns in the F₁ hybrid.** *Genetics* 2006, **173**:2199-2210.
40. Clark RM, Wagler TN, Quijada P, Doebley J: **A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture.** *Nat Genet* 2006, **38**:594-597.

An excellent example of long-range effects on gene regulation that points to the contribution of intergenic sequences that are rich in junk DNA to regulatory and quantitative variation.

41. Palaisa KA, Morgante M, Williams M, Rafalski A: **Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci.** *Plant Cell* 2003, **15**:1795-1806.
42. Cheng C, Daigen M, Hirochika H: **Epigenetic regulation of the rice retrotransposon *Tos17*.** *Mol Genet Genomics* 2006, **276**:378-390.
43. Lippman Z, Gendrel AV, Black M, Vaughn MW, Dedhia N, McCombie WR, Lavine K, Mittal V, May B, Kasschau KD *et al.*: **Role of transposable elements in heterochromatin and epigenetic control.** *Nature* 2004, **430**:471-476.
44. Huettel B, Kanno T, Daxinger L, Aufsatz W, Matzke AJ, Matzke M: **Endogenous targets of RNA-directed DNA methylation and *Pol IV* in *Arabidopsis*.** *EMBO J* 2006, **25**:2828-2836.

45. Devos KM, Brown JK, Bennetzen JL: **Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis***. *Genome Res* 2002, **12**:1075-1079.
46. Kobayashi S, Goto-Yamamoto N, Hirochika H: **Retrotransposon-induced mutations in grape skin color**. *Science* 2004, **304**:982.
47. Tsugane K, Maekawa M, Takagi K, Takahara H, Qian Q, Eun CH, Iida S: **An active DNA transposon *nDart* causing leaf variegation and mutable dwarfism and its related elements in rice**. *Plant J* 2006, **45**:46-57.
48. Levine M, Tjian R: **Transcription regulation and animal diversity**. *Nature* 2003, **424**:147-151.
49. Jordan IK, Rogozin IB, Glazko GV, Koonin EV: **Origin of a substantial fraction of human regulatory sequences from transposable elements**. *Trends Genet* 2003, **19**:68-72.
50. van de Lagemaat LN, Landry JR, Mager DL, Medstrand P: **Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions**. *Trends Genet* 2003, **19**:530-536.