Patterns in grass genome evolution

Jeffrey L Bennetzen

Increasingly comprehensive, species-rich, and large-scale comparisons of grass genome structure have uncovered an even higher level of genomic rearrangement than originally observed by recombinational mapping or orthologous clone sequence comparisons. Small rearrangements are exceedingly abundant, even in comparisons of closely related species. The mechanisms of these small rearrangements, mostly tiny deletions caused by illegitimate recombination, appear to be active in all of the plant species investigated, but their relative aggressiveness differs dramatically in different plant lineages. Transposable element amplification, including the acquisition and occasional fusion of gene fragments from multiple loci, is also common in all grasses studied, but has been a much more major contributor in some species than in others. The reasons for these quantitative differences are not known, but it is clear that they lead to species that have very different levels of genomic instability. Similarly, polyploidy and segmental duplication followed by gene loss are standard phenomena in the history of all flowering plants, including the grasses, but their frequency and final outcomes are very different in different lineages. Now that genomic instability has begun to be characterized in detail across an array of plant species, it is time for comprehensive studies to investigate the relationships between particular changes in genome structure and organismal function or fitness.

Addresses
Department of Genetics, University of Georgia, Athens, Georgia 30602-7223, USA

Corresponding author: Bennetzen, Jeffrey L (maize@uga.edu)

Current Opinion in Plant Biology 2007, **10**:176–181

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

Available online 8th February 2007

1369-5266/$ – see front matter

DOI 10.1016/j.pbi.2007.01.010

Introduction

The grasses have served as a model family for plant comparative genetics and genomics for more than a decade [1]. The importance of maize, rice, and wheat as the world's major food crops, and the extreme regional significance of sugarcane, sorghum, barley, oats, rye, various millets and forage grasses, has led to the creation of large research communities for these species, thereby guaranteeing enthusiastic proponents of comparative approaches. The more than 9000 current species of grasses are all derived from a common ancestor that lived about 50–80 million years ago (mya) [2–4]. Despite this relatively recent and monophyletic origin, grass genomes have diverged tremendously at the levels of chromosome number and genome size. Even among diploid species, grass genome size varies by more than 30-fold [5].

However, gene content has not been shown to be highly different between the grasses, or even between the grasses and more distant diploid relatives such as *Arabidopsis*. Although initial studies predicted that about half of the genes in rice had no strong homologs in *Arabidopsis* [6,7], and even that significant differences existed in local gene content between rice subspecies [8] or maize inbreds [9], subsequent studies determined that the great majority of these differences were due to inaccurate (i.e. excessive) gene predictions in the grass species studied [10–13,14••]. Hence, it appears that most genes, certainly greater than 90%, will be shared by any two compared grass species, although the copy numbers, expression patterns and precise uses of a few of these shared genes might have diverged to such a degree that they account for the unique properties of individual grass lineages. The search for these 'species-specific' or 'lineage-specific' genes or uniquely evolved gene roles will be a high-priority undertaking, and the grasses are well-suited for this pursuit. In the meantime, the general commonality in gene content will continue to serve as the foundation for comparative grass genomics.

Another vital tool in grass comparative genomics has been the colinearity of the genetic maps, as first evidenced by intraspecies recombinational maps that were based on shared DNA markers [15,16]. This colinearity was later confirmed and extended by DNA sequence comparisons of relatively small chromosomal segments from orthologous regions [17–19]. The observed colinearity of genes within compared segments of distant grass relatives, for instance rice and sorghum (which last shared a common relative ~50 mya), helped confirm the orthologous relationships of genes within the compared region. It also demonstrated that gene positions have been retained for most grass loci since their divergence from a common ancestor. Numerous small genic rearrangements were observed, however, often including one or two genes [20,21], although the mechanisms of these rearrangements were rarely clear. In a few cases, gene loss by the accumulation of small deletions [22] or gene inversion by unequal homologous recombination between flanking repeats [23] could be inferred, but the sequence comparisons of only a few regions from only a few species

made it unlikely that very recent rearrangements (with intact rearrangement hallmarks) would be discovered. This situation will soon be corrected by the recent dramatic expansion in the number of plant species targeted for extensive DNA sequence analyses. These projects include the ongoing whole-genome sequence analyses of sorghum, maize and Brachypodium, and the development of comprehensive projects for genomic investigations within specific plant lineages, such as the *Oryza* map alignment project (OMAP) for the genus *Oryza* [24] and similar studies in wheat and its close relatives [25•].

By definition, molecular comparisons of related genomes unavoidably uncover the properties of genome evolution. Novel features must have evolved recently, whereas genes that have different degrees of sequence conservation indicate different degrees of selection for conserved or altered function. With common gene content and genomic colinearity as tools to investigate the divergent and shared properties of grass genomes, the basic outlines of grass genome evolution have been defined. Polyploidy, both ancient [3,26–28] and recent, has been a major factor in the evolution of all grass lineages by increasing gene numbers, by activating transposable elements [29], by altering the epigenetic landscape leading to new gene expression patterns [29], and by creating the potential for differential post-polyploid gene loss or divergence [27]. Different levels of transposable element amplification, especially amplification of the class I mobile DNAs called long terminal repeat (LTR) retrotransposons, have been the major factors responsible for variation in plant genome size [10,18,24]. Nonetheless, different rates of DNA removal by unequal homologous recombination or illegitimate recombination might also be a significant factor in some lineages [30,31].

In this review, I briefly present recent advances in understanding the general and lineage-specific properties of grass genome evolution. I discuss the degree and nature of conservation in chromosomal gene order and the major contributors to the overarching complexity of grass genomes. Possible differences in the properties of genome evolution across chromosomal regions, especially in the comparison of centromeres, pericentromeres and other heterochromatic regions to the gene-rich euchromatin, are also discussed.

Genomic colinearity: local, segmental and chromosomal

Recombinational maps, physical maps, and DNA sequences all have the potential to be utilized for comparative purposes when common DNA sequences are employed or identified. Initially, these comparisons required a very high level of investigator expertise and manual annotation, partly because of the sparse datasets (too few markers, too few species) and also because of related challenges in distinguishing orthologs from paralogs. Recently, however, software and database resources for comparative grass genomics have been expanded tremendously, allowing a much more efficient process for local and global genome comparisons [32–36].

Comparisons of recombinational maps in the grasses continue, particularly for enabling research on understudied cereal species like the millets [37]. These studies continue to confirm and enrich the basic story presented by Moore and colleagues [16], that most grass genomes can be described by various arrangements of a small number of conserved chromosomal segments that are often displayed as a series of concentrically comparable genetic maps. With the near-completion of the rice genome sequence [38], however, genetic or physical maps with sequenced markers for any other genome could be compared to rice in great detail. Not surprisingly, many more chromosomal rearrangements have been discovered by these probe-rich studies [33,39,40,41•] than were initially identified by comparative recombinational mapping of a small number of shared DNA markers [38]. In a wheat–rice comparison, for instance, several-fold more rearrangements were observed on detailed physical–sequence maps than on compared recombinational maps [16]. Virtually all of these newly discovered rearrangements were only a few cM in size at most [39], indicating that these smallish rearrangements are much more common than the large rearrangements presented in the comparative circle maps. The mechanisms of these predicted rearrangements, including a very high number of apparent non-reciprocal internal translocations, are not indicated by these analyses, nor is it clear how many rearrangements might be explained by a frequent inability to make incontrovertible determinations of orthology or paralogy.

Comparisons of DNA sequences of orthologous chromosome segments in the grasses has been expanded to include larger genomic regions of many Mb [42], and will reach its ultimate form when 'completed' sequences are available for two or more grass species. The discoveries from these studies have largely confirmed what had already been noted in comparisons of single large (e.g. bacterial artificial chromosome) inserts: the sequences between genes are not extensively conserved, primarily consisting of transposable elements and transposable element fragments, whereas genic rearrangements (i.e. deletions, inversions, duplications and so on) affect a substantial minority of the genes [20,21]. At a quantitative level, it appears that the frequency of chromosomal rearrangements is at least crudely inversely proportional to the size of the rearrangement, but it is not clear whether this is caused by biases in the mechanism(s) of the rearrangements, by stronger selection against larger rearrangements *per se*, or by a combination of these two factors. Numerous studies now make it clear that most

small rearrangements, primarily deletions of a single bp up to a few kb, are caused by illegitimate recombination [10,25•,43,44]. In particular, powerful analyses have come from a multi-species approach [22,25•], where enough comparisons were available to identify the mechanisms and approximate dates or lineages of the rearrangements.

Even these sequence comparisons lack some precision, however, because gene identification continues to be an imperfect process. The very high gene numbers initially proposed for rice and other cereal species were based on a common mis-annotation of both gene fragments and low copy number transposable elements as genes [10–13,14••]. This problem continues to persist, partly because gene fragments and low copy number transposable elements are so abundant in plant genomes, and because it is never certain whether any individual case of either of these classes of apparent pseudogene might actually have some genetic or epigenetic function.

The origins of genomic complexity

The five primary mechanisms of genomic instability in all flowering plants are polyploidy, transposon amplification, chromosome breakage, unequal homologous recombination and illegitimate recombination. These are not necessarily independent phenomena. For instance, polyploidy can induce transposable element activity [29], transposable elements can break and otherwise rearrange chromosomes, and one mechanism of illegitimate recombination is the repair of double-stranded DNA breaks [45]. Because all grasses are derived from an ancestor that had an ancient (~70 mya) genome duplication [3,26–28], many of the differences in apparent local gene content and colinearity could be due to lineage-specific (and often random) differences in which members of the pairs of homoeologous loci were subsequently deleted [22,46]. Some of these differences are likely to be associated with actual differences in functional selection, however, given that the genes retained in the duplicated state are biased towards those whose relative gene-product stoichiometry might be essential for appropriate cellular function [27,47,48].

It is clear that the major mechanisms of genome rearrangement are not equally active in all plant lineages. LTR retrotransposon amplification, the most important factor in plant genome expansion, is more active in some species than others for unknown reasons, and the most abundant elements that are responsible for 'genomic obesity' are often not from closely related element families in different species [31]. In some cases, a few LTR retrotransposon families that are active in a relatively short time frame seem to have given rise to a major genome expansion, as in the wild rice *Oryza australiensis* [49•], whereas in other cases (e.g. maize) a large number of different element families appear to have been continuously active over periods of many millions of years. Rates of DNA removal also appear to be quite different in various plant lineages [31,45]; for example, there is great variation in the relative efficiency of DNA removal by illegitimate recombination versus unequal homologous recombination [30,31].

Although the reasons for this amazing variation in the relative aggressiveness of basic DNA integrity processes such as DNA break repair, recombination and transposon suppression are not known, it is now clear that these processes can easily generate the great variation observed in angiosperm genome size and structure within the timeframes identified by phylogenetic analyses [5]. For instance, comparison of the genomes of the two domesticated Asian rice subspecies, *japonica* and *indica*, using the wild African rice *Oryza glaberrima* as an outgroup, indicated that these two ~400 Mb genomes had grown more than 2% in overall genome size over the past few hundred thousand years because of LTR retrotransposon amplification [10]. This, despite a rate of DNA removal by illegitimate and unequal homologous recombination that exceeds 40 Mb per million years [10].

Regional differences in genome evolution

The very different gene contents and recombinational properties of different chromosomal segments, for instance, euchromatin versus heterochromatin, led to the general expectation that genome evolution in these regions might proceed at different rates and/or by different mechanisms. Bowers *et al.* [41•] observed that local genic colinearity was greatest in high recombinational (i.e. euchromatic) regions in a sorghum–rice–maize comparison. The fact that gene fragments, LTR retrotransposons and other sequences that are possibly misannotated as genes tend to accumulate in repeat-rich blocks [50] might partly explain this result.

Akhunov and colleagues [51] observed a higher frequency of gene duplication in the highly recombinogenic distal regions of wheat chromosomes, suggesting that these regions would evolve more rapidly. In fact, genes that have a need for rapid evolution by unequal recombination, such as disease resistance loci of the gene-for-gene type, may preferentially accumulate in recombination-rich regions of the genome.

Recent studies by Ma and colleagues have made the surprising discovery that centromeric regions of rice are hotspots for rearrangement by unequal homologous recombination [52•], despite the general observation that centromeres and flanking pericentromeric heterochromatin are coldspots for homologous chromosomal exchange in meiosis. Earlier *in situ* hybridization studies in rice [53] and maize [54] had shown that the number and arrangement of centromeric-specific tandem repeats was surprisingly variable, even in haplotypic comparisons of the same centromere. Comprehensive sequence analysis

across a pair of rice centromeres, *Cen4* and *Cen8*, the first two centromeres completely sequenced in any higher eukaryote, uncovered very high complexity in core centromere structure, with blocks of tandem repeat arrays intermixed with 'centromere-specific' (actually enriched) LTR retrotransposons and other retrotransposons, transposons, and unattributed sequences [55–57]. Most surprising, active genes were found in these centromeric regions, and these genes exhibited fairly normal structural, expression and divergence properties [52•,55–57]. Many of the genomic rearrangements in these regions had structures suggestive of an origin by unequal recombination, especially within the centromere core, the site of kinetochore formation, where the high relative abundance of solo LTRs indicated an unusually high rate of unequal homologous recombination [52•].

Conclusions

Most, perhaps all, of the major mechanisms responsible for plant genome structural evolution have now been identified, and they appear to be active in all characterized plant genomes. However, plant genomes continue to reveal unexpected patterns of rapid structural change, including lineage-specific and chromosomal region-specific differences in the frequencies of gene and genome duplication and differences in rates and most-active types of DNA removal. The functional outcomes, if any, of most of these genomic changes remain to be identified. With the continuing proliferation of genome sequence analyses and genomic tool development into a wider array of grass species, it will be possible to identify very recent genomic rearrangements. These recent rearrangements will retain structural features that indicate the mechanisms of DNA rearrangement. Equally important, in genomes that differ by only a few rearrangements, such as those of the *O. sativa* subspecies *japonica* and *indica*, segregation and/or comprehensive expression studies might be able to associate particular rearrangements with physiological, morphological or developmental traits that differentiate these close relatives. Structure, after all, does determine function, and the plant science community now has the opportunity to begin to characterize structure–function relationships in rapidly evolving grass genomes.

Acknowledgements

Many of the experimental analyses discussed in this manuscript, and its preparation, were supported by NSF grant 030477-2.

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. Bennetzen JL, Freeling M: **Grasses as a single genetic system: genome composition, collinearity and compatibility**. *Trends Genet* 1993, **9**:259-261.
2. Crepet WL, Feldman GD: **The earliest remains of grasses in the fossil record**. *Am J Bot* 1991, **78**:1010-1014.
3. Paterson AH, Bowers JE, Chapman BA: **Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics**. *Proc Natl Acad Sci USA* 2004, **101**:9903-9908.
4. Prasad V, Strömberg CAE, Alimohammadian H, Sahni A: **Dinosaur coprolites and the early evolution of grasses and grazers**. *Science* 2005, **310**:1177-1180.
5. Kellogg EA, Bennetzen JL: **The evolution of nuclear genome structure in seed plants**. *Am J Bot* 2004, **91**:1709-1725.
6. Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X *et al.*: **A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*)**. *Science* 2002, **296**:79-91.
7. Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H *et al.*: **A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*)**. *Science* 2002, **296**:92-100.
8. Feng Q, Zhang Y, Hao P, Wang S, Fu G, Huang Y, Li Y, Zhu J, Liu Y, Hu X *et al.*: **Sequence and analysis of rice chromosome 4**. *Nature* 2002, **420**:316-320.
9. Fu H, Dooner HK: **Intraspecific violation of genetic colinearity and its implications in maize**. *Proc Natl Acad Sci USA* 2002, **99**:9573-9578.
10. Ma J, Bennetzen JL: **Rapid recent growth and divergence of rice nuclear genomes**. *Proc Natl Acad Sci USA* 2004, **101**:12404-12410.
11. Jabbari K, Cruveiller S, Clay O, Le Saux J, Bernardi G: **The new genes of rice: a closer look**. *Trends Plant Sci* 2004, **9**:281-285.
12. Bennetzen JL, Coleman C, Liu R, Jianxin Ma J, Ramakrishna W: **Consistent over-estimation of gene number in complex plant genomes**. *Curr Opin Plant Biol* 2004, **7**:732-736.
13. Lai J, Li Y, Messing J, Dooner HK: **Gene movement by Helitron transposons contributes to the haplotype variability in maize**. *Proc Natl Acad Sci USA* 2005, **102**:9068-9073.
14. •• 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.

This paper describes experiments indicating that a class of mobile DNAs called helitrons are responsible for much of the perceived variation in genic colinearity between maize inbreds and between maize and other grass species. More than 10 000 different gene fragments are predicted to have been acquired and amplified by helitrons in maize, often with fragments from several different genes in the same helitron. Some of these helitrons are expressed, occasionally yielding transcripts that include fused segments from different genes. These processes of fragment acquisition, amplification and exon shuffling are actively occurring in maize, thereby generating much of the perceived genetic diversity in this highly diverse species, and perhaps generating significant levels of functional novelty.

15. Hulbert SH, Richter TE, Axtell JD, Bennetzen JL: **Genetic mapping and characterization of sorghum and related crops by means of maize DNA probes**. *Proc Natl Acad Sci USA* 1990, **87**:4251-4255.
16. Moore G, Devos KM, Wang Z, Gale MD: **Grasses, line up and form a circle**. *Curr Biol* 1995, **5**:737-739.
17. Chen M, SanMiguel P, de Oliveira AC, Woo S-S, Zhang H, Wing RA, Bennetzen JL: **Microcolinearity in the *sh2*-homologous regions of the maize, rice and sorghum genomes**. *Proc Natl Acad Sci USA* 1997, **94**:3431-3435.
18. Tikhonov AP, SanMiguel PJ, Nakajima Y, Gorenstein NM, Bennetzen JL, Avramova Z: **Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum**. *Proc Natl Acad Sci USA* 1999, **96**:7409-7414.
19. Feuillet C, Keller B: **High gene density is conserved at syntenic loci of small and large grass genomes**. *Proc Natl Acad Sci USA* 1999, **96**:8665-8670.

20. Feuillet C, Keller B: **Comparative genomics in the grass family: molecular characterization of grass genome structure and function**. *Ann Bot* 2002, **89**:3-10.

21. Bennetzen JL, Ma J: **The genetic colinearity of rice and other cereals based on genomic sequence analysis**. *Curr Opin Plant Biol* 2003, **6**:128-133.

22. Ilic K, SanMiguel PJ, Bennetzen JL: **A complex history of rearrangement in an orthologous region of the maize, sorghum and rice genomes**. *Proc Natl Acad Sci USA* 2003, **100**:12265-12270.

23. Dubcovsky J, Ramakrishna W, SanMiguel PJ, Busso CS, Yan L, Shiloff BA, Bennetzen JL: **Comparative sequence analysis of colinear barley and rice bacterial artificial chromosomes**. *Plant Physiol* 2001, **125**:1342-1353.

24. Ammiraju JSS, Luo M, Goicoechea JL, Wang W, Kudrna D, Mueller C, Talag J, Kim HR, Sisneros NB, Blackmon B *et al.*: **The *Oryza* bacterial artificial chromosome library resource: construction and analysis of 12 deep-coverage large-insert BAC libraries that represent the 10 genome types of the genus *Oryza***. *Genome Res* 2006, **16**:140-147.

25. • Chantret N, Salse J, Sabot F, Rahman S, Bellect A, Laubin B, Dubois I, Dossat C, Sourdille P, Joudrier P *et al.*: **Molecular basis of evolutionary events that shaped the *Hardness* locus in diploid and polyploidy wheats**. *Plant Cell* 2005, **17**:1033-1045.

Chaloub and colleagues sequence an orthologous region from domesticated bread wheat, *Triticum aestivum* (AABBDD), and three close relatives, *Triticum turgidum* (AABB), *Aegilops tauschii* (DD) and *Triticum monococcum* (A^mA^m). With this array of species, it was possible to determine specific rearrangement events, their timing in the descent of these species, and their apparent mechanisms of origin. Illegitimate recombination was found to be a particularly major contributor to the observed rearrangements.

26. Vandepoele K, Simillion C, Van de Peer Y: **Evidence that rice and other cereals are ancient aneuploids**. *Plant Cell* 2003, **9**:2192-2202.

27. Tian CG, Xiong YQ, Liu TY, Sun SH, Chen LB, Chen MS: **Evidence for an ancient whole genome duplication event in rice and other cereals**. *Yi Chuan Xue Bao* 2005, **32**:519-527.

28. 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.

29. Kashkush K, Feldman M, Levy AA: **Gene loss, silencing and activation in a newly synthesized wheat allotetraploid**. *Genetics* 2002, **160**:1651-1659.

30. Bennetzen JL, Ma J, Devos KM: **Mechanisms of recent genome size variation in flowering plants**. *Ann Bot* 2005, **95**:127-132.

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

32. Dong Q, Schlueter SD, Brendel B: **PlantGDB, plant genome database and analysis tools**. *Nucleic Acids Res* 2004, **32**:D354-D359.

33. Odland W, Baumgarter A, Phillips R: **Ancestral rice blocks define multiple related regions in the maize genome**. *Crop Sci* 2006, **46**:S41-S48.

34. Soderlund C, Nelson W, Shoemaker A, Paterson A: **SyMAP: a system for discovering and viewing syntenic regions of FPC maps**. *Genome Res* 2006, **16**:1159-1168.

35. Jaiswal P, Ni J, Yap I, Ware D, Spooner W, Youens-Clark K, Ren L, Liang C, Zhao W, Ratnapu K *et al.*: **Gramene: a bird's eye view of cereal genomes**. *Nucleic Acids Res* 2006, **34**:D717-D723.

36. Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, Childs K, Thibaud-Nissen F, Malek RL, Lee Y, Zheng L *et al.*: **The TIGR rice genome annotation resource: improvements and new features**. *Nucleic Acids Res* 2007, **35**:D883-D887.

37. Devos KM, Pittaway TS, Reynolds A, Gale MD: **Comparative mapping reveals a complex relationship between the pearl millet genome and those of foxtail millet and rice**. *Theor Appl Genet* 2000, **100**:190-198.

38. International Rice Genome Sequencing Project: **The map-based sequence of the rice genome**. *Nature* 2005, **436**:793-800.

39. Sorrells ME, La Rota M, Bermudez-Kandianis CE, Greene RA, Kantety R, Munkvold JD, Miftahudin, Mahmoud A, Ma X, Gustafson PJ *et al.*: **Comparative DNA sequence analysis of wheat and rice genomes**. *Genome Res* 2003, **13**:1818-1827.

40. The Rice Chromosome 3 Sequencing Consortium: **Sequence, annotation, and analysis of synteny between rice chromosome 3 and diverged grass species**. *Genome Res* 2005, **15**:1284-1291.

41. • Bowers JE, Arias MA, Asher R, Avise JA, Ball RT, Brewer GA, Buss RW, Chen AH, Edwards TM, Estill JC *et al.*: **Comparative physical mapping links conservation of microsynteny to chromosome structure and recombination in grasses**. *Proc Natl Acad Sci USA* 2005, **102**:13206-13211.

These studies employed DNA sequences that are shared between two sorghum physical maps and the sequenced rice genome to characterize the distribution of genic rearrangements. Greater microcolinearity was observed in euchromatic (i.e. gene-rich and recombination-rich) regions. The authors propose that rearrangements may often be deleterious, and therefore could only accumulate in regions where their removal by recombination is relatively slow.

42. 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.

43. Devos KM, Brown JK, Bennetzen JL: **Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis***. *Genome Res* 2002, **12**:1075-1079.

44. Wicker T, Yahiaoui N, Guyot R, Schlagenhauf E, Liu Z-D, Dubcovsky J, Keller B: **Rapid genome divergence at orthologous low molecular weight glutenin loci of the A and A^m genomes of wheat**. *Plant Cell* 2003, **15**:1186-1197.

45. Kirik A, Salomon S, Puchta H: **Species-specific double-strand break repair and genome evolution in plants**. *EMBO J* 2000, **19**:5562-5566.

46. Lai J, Ma J, Swigonova Z, Ramakrishna W, Linton E, Llaca V, Tanyolac B, Park YJ, Jeong OY, Bennetzen JL, Messing J: **Gene loss and movement in the maize genome**. *Genome Res* 2004, **14**:1924-1931.

47. Thomas BC, Pedersen B, Freeling M: **Following tetraploidy in an *Arabidopsis* ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes**. *Genome Res* 2006, **16**:934-946.

48. Chapman BA, Bowers JE, Feltus FA, Paterson AH: **Buffering of crucial functions by paleologous duplicated genes may contribute cyclicality to angiosperm genome duplication**. *Proc Natl Acad Sci USA* 2006, **103**:2730-2735.

49. • 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 retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice**. *Genome Res* 2006, **16**:1262-1269.

This manuscript demonstrates that the recent amplification of three LTR retrotransposons is responsible for the relatively large size of the *O. australiensis* genome. Members of these LTR retrotransposon families are all present in the *O. sativa* genome, but with intact element copy numbers of less than 20. The species-specific amplifications of these elements to copy numbers of more than 10 000 each has occurred in the past 2.5 million years, adding ~600 Mb to the ~965 Mb genome of *O. australiensis*.

50. Ma J, SanMiguel P, Lai J, Messing J, Bennetzen JL: **DNA rearrangement in orthologous *orp* regions of the maize, rice and sorghum genomes**. *Genetics* 2005, **170**:1209-1220.

51. Akhunov ED, Goodyear AW, Geng S, Qi L-L, Echalier B, Gill BS, Miftahudin, Gustafson JP, Lazo G, Chao S *et al.*: **The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms**. *Genome Res* 2003, **13**:753-763.

52. Ma J, Bennetzen JL: **Recombination, rearrangement, reshuffling, and divergence in a centromeric region of rice**. *Proc Natl Acad Sci USA* 2006, **103**:383-388.
This study investigated the divergence and rearrangement of genes, tandem repeats, and LTR retrotransposons in the centromeric region of rice chromosome 8. Unequal homologous recombination was found to be a major contributor to the instability of this region, and this process was particularly active in the core of the centromere that directs spindle attachment. Because meiotic homolog exchange has been observed to be suppressed in this and all other studied eukaryotic centromeric regions, the authors propose that recombination in this area is quite active, but is channeled primarily into non-crossover outcomes and/or sister chromatid exchange.

53. Cheng Z, Dong F, Langdon T, Ouyang S, Buell CR, Gu M, Blattner FR, Jiang J: **Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon**. *Plant Cell* 2002, **14**:1691-1704.

54. Kato A, Lamb JC, Birchler JA: **Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize**. *Proc Natl Acad Sci USA* 2004, **101**:13554-13559.

55. Wu J, Yamagata H, Hayashi-Tsugane M, Hijishita S, Fujisawa M, Shibata M, Ito Y, Nakamura M, Sakguchi M, Yoshihara R *et al.*: **Composition and structure of the centromeric region of rice chromosome 8**. *Plant Cell* 2004, **16**:967-976.

56. Zhang Y, Huang Y, Zhang L, Li Y, Lu T, Lu Y, Feng Q, Zhao Q, Cheng Z, Xue Y *et al.*: **Structural features of the rice chromosome 4 centromere**. *Nucleic Acids Res* 2004, **32**:2023-2030.

57. Nagaki K, Cheng Z, Ouyang S, Talbert PB, Kim M, Jones KM, Henikoff S, Buell CR, Jiang J: **Sequencing of a rice centromere uncovers active genes**. *Nat Genet* 2004, **36**:138-145.