Structure and evolution of cereal genomes

Andrew H Paterson[*], John E Bowers[*], Daniel G Peterson[†], James C Estill[*] and Brad A Chapman[*]

The cereal species, of central importance to our diet, began to diverge 50–70 million years ago. For the past few thousand years, these species have undergone largely parallel selection regimes associated with domestication and improvement. The rice genome sequence provides a platform for organizing information about diverse cereals, and together with genetic maps and sequence samples from other cereals is yielding new insights into both the shared and the independent dimensions of cereal evolution. New data and population-based approaches are identifying genes that have been involved in cereal improvement. Reduced-representation sequencing promises to accelerate gene discovery in many large-genome cereals, and to better link the under-explored genomes of 'orphan' cereals with state-of-the-art knowledge.

Addresses
[*]Plant Genome Mapping Laboratory, University of Georgia, Athens, GA 30602, USA
[†]Mississippi Genome Exploration Laboratory, Mississippi State University, Mississippi, MO 39762, USA

Current Opinion in Genetics & Development 2003, **13**:644–650

This review comes from a themed issue on
Genomes and evolution
Edited by Evan Eichler and Nipam Patel

0959-437X/$ – see front matter

DOI 10.1016/j.gde.2003.10.002

Abbreviations

BAC	bacterial artificial chromosome
CBCS	Cot-based cloning and sequencing
EST	expressed sequence tag
SNP	single-nucleotide polymorphism
STS	sequence-tagged site

Introduction

The cereal crops, which provide about half of the calories in our diet, represent a relatively recent branch of the plant family tree. Although the angiosperm (flowering plant) lineage is thought to be ~200 million years old, cereals such as maize (*Zea*), rice (*Oryza*), sorghum (*Sorghum*), and wheat (*Triticum*) diverged from a common ancestor only ~50–70 million years ago [1]. Approximately 10,000 years ago, humans began to select cereals for traits including non-shattering, determinate growth, increased seed number, size, and carbohydrate content, and reduced dormancy [2]. Although these 'domestication' efforts were ostensibly independent and occurred on different continents — maize in America, sorghum in Africa, wheat in the Near East, and rice in both Africa and Asia — the possibility that mutations in some corresponding genes may have been selected (e.g. see [3,4]) is a general reflection of the many structural and functional parallels that appear to have persisted since the divergence of these lineages.

Most widely-grown cereals now enjoy detailed sequence-tagged site (STS) based genetic recombination maps [5–11,12•,13,14] that are suitable both for comparative biology and for crop improvement. Whereas these maps have been successfully applied for many uses using traditional restriction-fragment length polymorphism or simple sequence repeat based methods, genetically-mapped STSs can readily be used to discover single-nucleotide polymorphisms (SNPs) or small insertion/deletion polymorphisms [15•] that can then be genotyped by many new technologies. The ability to acquire such polymorphism information for corresponding loci in many genotypes increases the value of STS maps and reduces the costs associated with their wider utilization. STS-based maps also provide an excellent means by which physical maps based upon large-insert clones can be integrated with genetic recombination maps [16,17].

Gene repertoire and arrangement along the chromosomes of diverse cereals has evolved much more slowly than overall genome size and organization [18•]. For example, in the detailed genetic recombination maps of maize (>3400 loci; [9]; MaizeDB: http://www.agron.missouri.edu/maps.html), and sorghum (>2500 loci; [19]; Plant Genome Mapping Laboratory: http://www.plantgenome.uga.edu), ~55% of comparative loci show corresponding arrangement [19] although the underlying genomes differ fourfold in DNA content. The 35-fold divergence in genome size among major cereals, from ~0.5 pg (~490 Mb) per 1C (DNA content per haploid nucleus) for *Oryza sativa* (rice) to 17.33 pg (~16,979 Mb) per 1C for *Triticum aestivum* (bread wheat) [20], appears to be largely a result of a dynamic and lineage-specific balance between generation and elimination of mobile dispersed repetitive DNA elements [21]. Mechanisms that contribute to elimination of repetitive DNA may also contribute to the gene loss that follows polyploid formation in angiosperms [22••]. Differential gene loss after genome-wide duplication(s) may account for an appreciable fraction of genes that appear to be 'missing' in comparisons of micro-colinearity among some taxa.

Genomic archaeology: ancient events that shaped modern genomes

The large genomes of many cereal crops are not likely to be sequenced for many years yet, but the relatively close relationship among the major cereals suggests that their study and improvement can benefit considerably from the sequences of small-genome relatives. The emerging sequence of the rice chromosomes [23••,24••,25•], generated by the integration of 'genomic shotgun' data [26•,27•] with extensive genetic and physical mapping efforts [12•,28•,29], provides a foundation for organizing information about diverse cereals, and is yielding new insights into both the shared and the independent dimensions of cereal evolutionary history.

The cross-utilization of information from botanical models such as rice, in the study and improvement of major cereal crops requires a detailed understanding of the evolutionary history of cereal genomes. An especially important shared feature of cereal genome structure that will only be clear with a completed genome sequence is the pattern (or lack thereof) of ancient chromosomal or predominantly whole genome duplication, an evolutionary event with profound consequences for comparative biology [22••,30•]. Recent, extensive duplication in the genomes of some cereals such as maize [31,32•], and other grasses such as sugarcane [7], has long been recognized. Although a complete rice genome sequence is required for a definitive picture of the history of ancient duplications in the cereal lineage, an early glimpse on the basis of analysis of unordered sequence within series of physically-ordered BACs (bacterial artificial chromosomes; Figure 1) strongly supports suggestions [26•,33–35] that ancient genome-wide duplication has occurred here. 'Phylogenomic' analysis, merging phylogenetic inference with structural genomic data (in this case, regarding ancient duplication patterns) (Figure 1a), suggests that at least one duplication event predates the divergence of the major cereal lineages (Figure 1b). That such an event predates cereal divergence is consistent with the largely 'one-to-one' correspondence found between the chromosomes of ostensibly diploid cereals such as sorghum and rice.

Differential 'diploidization' (loss of duplicated genes [36]) in different lineages may contribute to deviations from synteny such as the 45% incongruency of anchor loci found between sorghum and maize [19]. For example, our comparison of the rice BACs (Figure 1) to the sorghum genetic map shows that 61.6% of loci (comprising 12 syntenic groups averaging 83.5 matching sequences each) fit with established primary syntenic relationships. Among those loci that deviate from the primary syntenic relationships, 25% (or 9.6% of all loci) fall at the single locations that are consistent with the proposed ancient duplications illustrated in Figure 1, with the remaining 75% distributed over the other 10 chromosomes. Together with other mechanisms of gene rearrangement, differential diploidization may also contribute to why some comparisons of more distantly related plant taxa have yielded complex mosaics of syntenic and non-syntenic loci [37–41].

Footprints of domestication and improvement

A particularly important application of structural genomic tools is the manipulation (in breeding) and even isolation of a growing number of genes important to agriculture, evolution, or development. Quantitative trait locus mapping approaches continue to be a powerful and widely-used means to identify genes that influence phenotypes for which information about related biochemical pathways or mechanisms is lacking. Progress in the use of structural genomic tools to dissect complex traits in rice [42] has been particularly good, although much progress has also been made in many other crops and botanical models too numerous to fully address herein.

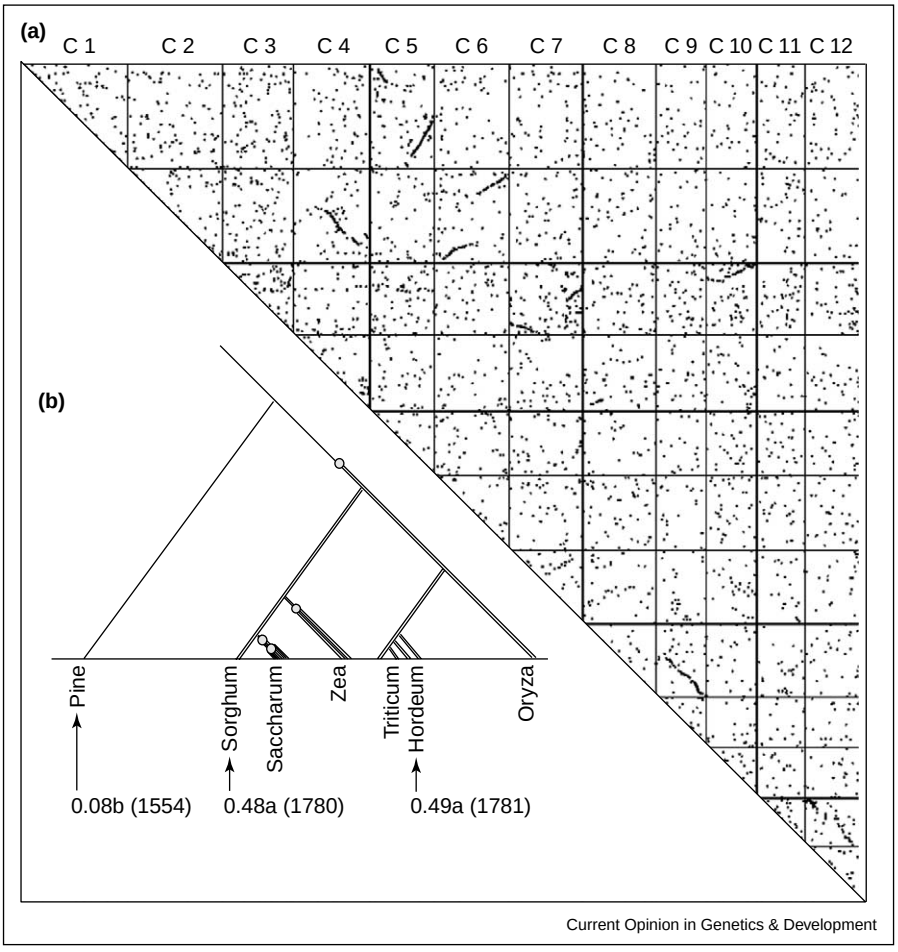
The rapidly-growing density of genetically-mapped STS markers, together with burgeoning DNA sequence data, provides new clues as to the locations and identities of phenotypically-important genes under selection [43,44•]. The past five years have seen an explosion in cereal genomics, with DNA sequence data for the cereal family tree expanding much more rapidly than for other taxa. In January of 1998, the total of 4 Mb of cereal sequence represented ~1% of the data held in GenBank. By 31 May of 2003, the total of 2,038 Mb of cereal sequence comprised ~6% that held in GenBank (Figure 2).

Particularly important to relating such extensive sequence data to phenotype will be efficient re-sequencing approaches, using carefully-selected germplasm collections together with detailed knowledge of population structure and genetic relationships to implicate small subsets of sequences in the control of key traits [43]. Such 'association approaches' benefit from a good understanding of the extent of linkage disequilibrium in plant populations. In outcrossing species such as maize, linkage disequilibrium often decays to virtually undetectable levels even at opposite ends of a single gene, although rates of decay vary widely for different genes [45]. Linkage disequilibrium is moderately stronger in low-copy regions of predominantly-selfing taxa (M Hamblin, AH Paterson, S Kresovich, unpublished data) but still much less than expected if recombination were distributed evenly across genomic DNA, suggesting that the bulk of recombination in plants may occur within genes.

Toward a complete picture: broadening knowledge of diversity, and extending data from botanical models across (and beyond) the cereals

Enhanced knowledge of the adaptations that suit the cereals to agriculture, and specifically identifying the

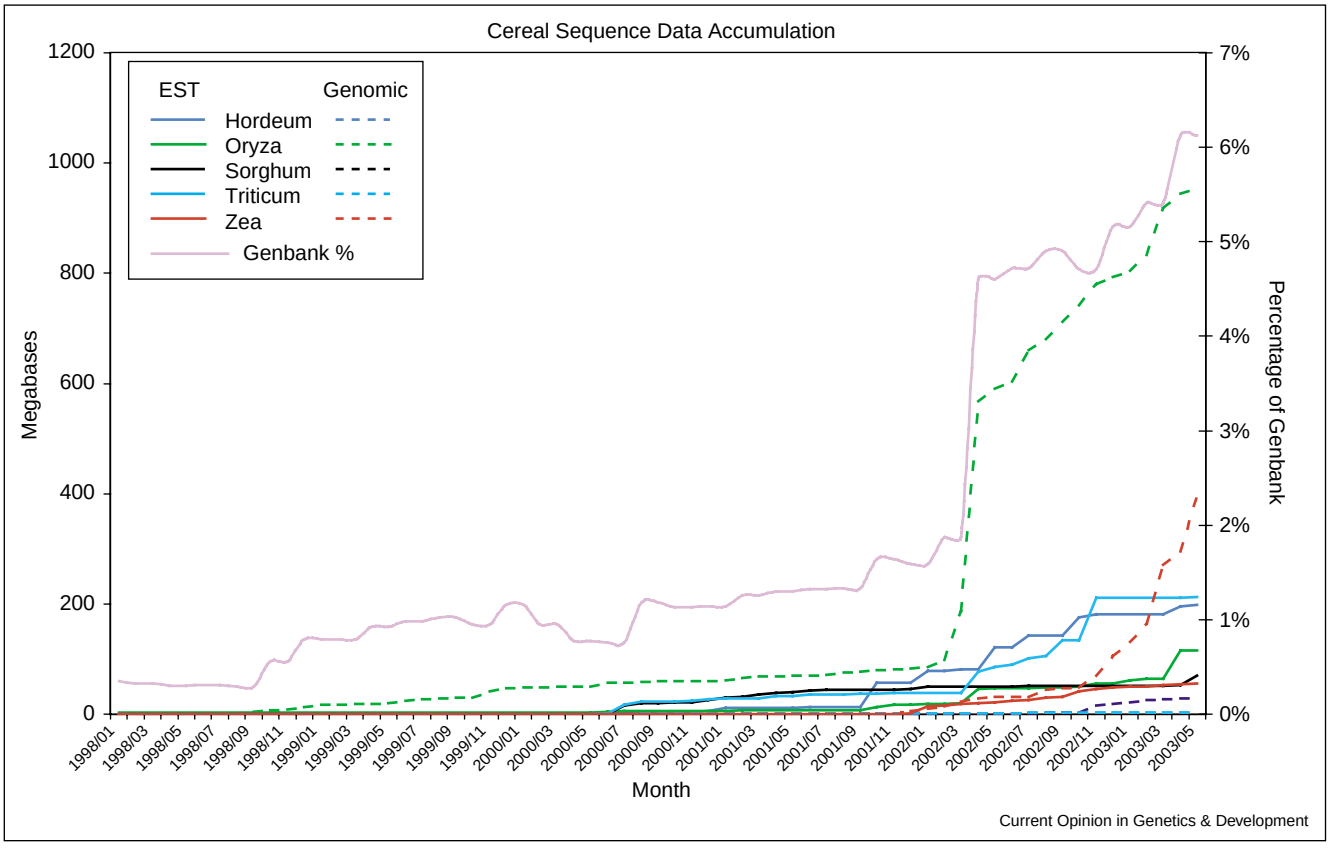
Figure 1



Genomic duplication in selected cereal genomes. **(a)** Patterns of ancient duplication in rice. For a first glimpse into the history of duplication early in the evolution of the cereal lineage, we used public data to construct a set of approximately physically-ordered BACs for rice, then identified genes that show high sequence similarity to un-ordered shotgun reads within the BACs, and studied the arrangements of closely-corresponding genes across the genome as described for *Arabidopsis* [59]. While a true assembled sequence will add important resolution and detail to this picture, it is clear already that there has been substantial, perhaps genome-wide, duplication of the rice genome. Previously reported duplication of chromosomes 1–5 [33] and 11–12 [34] is clear, together with nonoverlapping duplication of chromosome 2 with parts of chromosome 4 and 6; chromosome 3 with parts of chromosomes 7 and 10; and much of chromosomes 8 and 9. **(b)** Preliminary phylogenetic analysis for key cereal nodes and an outgroup (pine) were conducted as described [27•], including rooting using *Physcomitrella* sequences, with the frequencies of internal gene trees generated by comparison of each taxon shown to the rice duplicates shown below the taxa. Different letters represent statistically significant differences based on Tukey's test, and the number of trees that could be tested (according to criteria described in [27•]) is shown in parentheses. Open circles indicate possible chromosomal duplication events. The gene tree analyses support prior evidence of largely 'one-to one' correspondence of the rice chromosomes to those of other nominally diploid grasses such as sorghum [60] in suggesting that much of the duplication shown predates the divergence of the cereals from one another. Additional, more recent duplications within some cereal lineages such as maize [31,32•] and sugarcane [7] further complicate the comparative genomics of the cereals. Recent polyploidy formation within the Triticeae (not shown) similarly complicates comparisons at the tetraploid and hexaploid level. Although the drawing is congruent with contemporary phylogenetic inference, the lengths of the branches are illustrative and do not precisely represent any particular measure of taxonomic distance.

diversity that makes different cereals more productive in different environments or for different purposes, are important potential benefits that impel the acquisition of more and better information about the evolution of cereal genes and genomes. A major obstacle in gaining such information is repetitive DNA, which comprises the bulk of most cereal genomes (Figure 3) and therefore largely determines the cost of sequencing these genomes by shotgun approaches. As expressed sequence tag (EST) sequencing reaches a point of diminishing returns, a particularly promising approach to isolating comprehensive sets of cereal gene sequences is the use of Cot analysis [46] to fractionate repetitive genomes into components with similar degrees of sequence repetition, and then to sequence corresponding clone libraries to a depth sufficient to represent

Figure 2



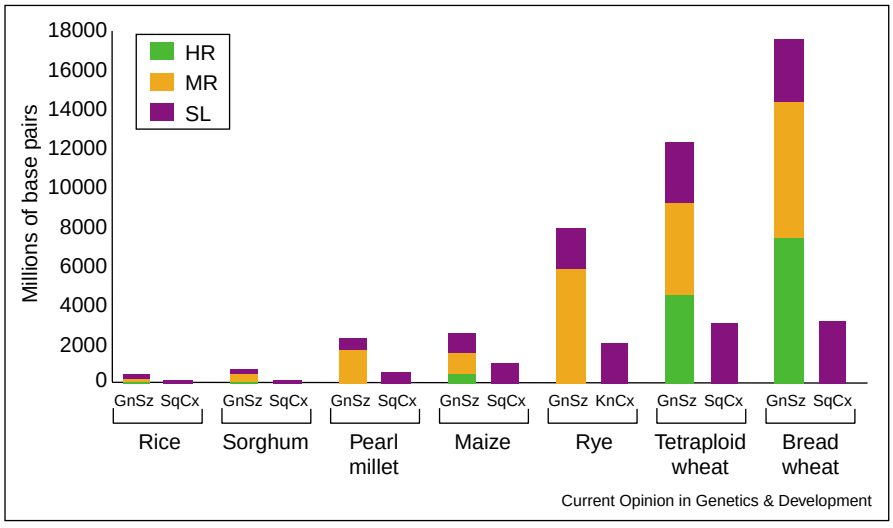
Public DNA sequence data for selected cereals. Color-coded curves illustrate the cumulative totals of EST and genomic survey sequence data deposited in GenBank for maize, sorghum, the Triticeae, and rice for the past 5 years in reference to the left axis, and the relative growth of cereal sequence data as a percentage of total GenBank entries in reference to the right axis.

the 'sequence complexity' of the respective components [47].

As with genomic shotgun sequencing, the 'Cot-based cloning and sequencing' (CBCS) approach represents only one dimension of a multi-faceted strategy that will be needed to produce finished sequences of large cereal genomes. However, although the cost of standard genomic shotgun sequencing is proportional to total genome size, the cost of CBCS is proportional to genomic sequence complexity, yielding immense potential savings in characterizing the diversity of sequences in different genomes (Figure 3). First demonstrated in 2001 [48], early CBCS explorations show promise not only in characterization of the relatively small (~700 Mb) genome of sorghum [49••] but also in the larger and more complex genome of maize [50], and in non-cereals such as cotton (T Wicker *et al.*, unpublished data). Alternatives to CBCS based upon differential methylation of expressed versus non-expressed sequences [51] may also contribute significantly to 'skimming' of low-copy sequences from large-genome taxa, but unlike CBCS these techniques are subject to the variable relationship between methylation and gene expression across genes and taxa (see [47,49••] for an extensive discussion). CBCS has the further advantage that representative sequences of DNA families comprising the repetitive fraction(s) of a genome can also be efficiently obtained [47,48,49••]. This mitigates the risk that potentially valuable information is lost, in comparison to alternative methods [50,51] that disregard these fractions.

Prior to the production of sufficient data to cover the entire sequence complexity of a taxon, much may be learned from reduced-representation approaches that use, for example, restriction enzymes [52] or degenerate oligonucleotide primers [53•] to identify large numbers of SNPs in widely-distributed STSs. The STS-based information deriving from these methods is useful not only in detecting SNPs, but also in aligning the underlying genomes to those of well-studied models such as rice. Many 'orphan crops' [54] for which genomic information is presently lacking, but that are essential to sustain low-income human populations in harsh climates where

Figure 3



Contributions of highly repetitive (HR, green), moderately repetitive (MR, orange), and single/low-copy (SL, purple) components to genome size (GnSz) and sequence complexity (SqCx). Whereas HR and MR components collectively constitute the bulk of DNA in all cereal genomes, these components comprise only negligible portions of each genome's SqCx. CBCS allows elucidation of a genome's SqCx with minimal encumbrance by repetitive elements [49••]. Consequently, CBCS may accelerate exploration of large, highly repetitive genomes such as those of many cereals [47]. Cot data [49••,61–65] was normalized to reflect current genome size estimates [20] and differences in data formats (see www.msstate.edu/research/mgel/pdf_files/cogd_fig.pdf for details).

inputs such as fertilizer and water are prohibitively costly may be great beneficiaries from such approaches.

Conclusions and future directions

The completed rice sequence promises to shed much light on the early events that shaped the cereal lineage, and be invaluable as a framework for organizing comparative information [55] for both major and 'orphan' [54] cereals. The value of cross-utilizing rice genomic tools in other cereals is also considerable [56]. However, comparative phenotypic, genomic and sequence information from many additional taxa will be needed to elucidate the specific events responsible for the morphological and physiological diversity that adapts different cereals to different climates, production regimes, and human needs. Such information promises to grow at an accelerating rate by virtue of efficient new methods, and will help to reveal the relative roles of different genes, and different types of genomic changes, in the evolution of phenotypic diversity among and within cereal lineages. As the identities of growing numbers of key cereal genes become known in individual taxa (e.g. [57]), growing attention to comparative biology (e.g. [58]) promises to facilitate understanding of the relationship between DNA polymorphism and biological diversity.

Acknowledgements

We thank many members of the Paterson laboratory and our collaborators and colleagues for fruitful discussions, and the US National Science Foundation, US Department of Agriculture, International Consortium for Sugarcane Biotechnology, US Golf Association, and Georgia Agricultural Experiment Station for financial support.

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. Kellogg EA: **Relationships of cereal crops and other grasses**. *Proc Natl Acad Sci USA* 1998, **95**:2005-2010.

2. Harlan JR: *Crops and Man*. Madison, WI: Crop Science Society of America; 1975.

3. Paterson A, Lin Y-R, Li Z: **Convergent domestication of cereal crops by independent mutations at corresponding genetic loci**. *Science* 1995, **269**:1714-1718.

4. Hu FY, Tao DY, Sacks E, Fu BY, Xu P, Li J, Yang Y, McNally K, Khush GS, Paterson AH *et al.*: **Convergent evolution of perenniality in rice and sorghum**. *Proc Natl Acad Sci USA* 2003, **100**:4050-4054.

5. Qi XQ, Stam P, Lindhout P: **Comparison and integration of four barley genetic maps**. *Genome* 1996, **39**:379-394.

6. Roder MS, Korzun V, Wendehake K, Plaschke J, Tixier MH, Leroy P, Ganal MW: **A microsatellite map of wheat**. *Genetics* 1998, **149**:2007-2023.

7. Ming R, Liu SC, Lin YR, da Silva J, Wilson W, Braga D, van Deynze A, Wenslaff TF, Wu KK, Moore PH *et al.*: **Detailed alignment of Saccharum and Sorghum chromosomes: comparative organization of closely related diploid and polyploid genomes**. *Genetics* 1998, **150**:1663-1682.

8. 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.

9. Lee M, Sharopova N, Beavis WD, Grant D, Katt M, Blair D, Hallauer A: **Expanding the genetic map of maize with the intermated B73 × Mo17 (IBM) population**. *Plant Mol Biol* 2002, **48**:453-461.

10. McCouch SR, Teytelman L, Xu YB, Lobos KB, Clare K, Walton M, Fu BY, Maghirang R, Li ZK, Xing YZ *et al.*: **Development and**

mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). *DNA Res* 2002, **9**:199-207.

11. Sharopova N, McMullen MD, Schultz L, Schroeder S, Sanchez-Villeda H, Gardiner J, Bergstrom D, Houchins K, Melia-Hancock S, Musket T *et al.*: **Development and mapping of SSR markers for maize**. *Plant Mol Biol* 2002, **48**:463-481.

12. • Wu JZ, Maehara T, Shimokawa T, Yamamoto S, Harada C, Takazaki Y, Ono N, Mukai Y, Koike K, Yazaki J *et al.*: **A comprehensive rice transcript map containing 6591 expressed sequence tag sites**. *Plant Cell* 2002, **14**:525-535.
Gene mapping, based on the hybridization of ESTs to large-insert DNA clones, provides many of the benefits of a completed sequence to many cereal genomes at much less time and cost. Although the rice genome will soon be fully sequenced, the methods described are applicable to many additional genomes that may not be sequenced for quite some time yet.

13. Wight CP, Tinker NA, Kianian SF, Sorrells ME, O'Donoughue LS, Hoffman DL, Groh S, Scoles GJ, Li CD, Webster FH *et al.*: **A molecular marker map in 'Kanota' × 'Ogle' hexaploid oat (Avena spp.) enhanced by additional markers and a robust framework**. *Genome* 2003, **46**:28-47.

14. 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.

15. • Nasu S, Suzuki J, Ohta R, Hasegawa K, Yui R, Kitazawa N, Monna L, Minobe Y: **Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa*, *Oryza rufipogon*) and establishment of SNP markers**. *DNA Res* 2002, **9**:163-171.
Among the first illustrations of a genome-wide SNP map in plants, including early quantification of rates and types of SNPs within *Oryza*, and demonstration of an uneven genomic landscape of SNP distribution.

16. Draye X, Lin Y-R, Bowers JE, Burow GB, Morrell PL, Peterson DG, Presting GG, Ren SX, Wing RA, Paterson AH: **Toward integration of comparative genetic, physical, diversity, and cytomolecular maps for grasses and grains, using the *Sorghum* genome as a foundation**. *Plant Physiol* 2001, **125**:1325-1341.

17. Coe E, Cone K, McMullen M, Chen SS, Davis G, Gardiner J, Liscum E, Polacco M, Paterson A, Sanchez-Villeda H *et al.*: **Access to the maize genome: an integrated physical and genetic map**. *Plant Physiol* 2002, **128**:9-12.

18. • Feuillet C, Keller B: **Comparative genomics in the grass family: molecular characterization of grass genome structure and evolution**. *Ann Bot (Lond)* 2002, **89**:3-10.
A succinct review of micro-colinearity and its exceptions in the grasses.

19. Bowers JE, Abbey C, Anderson S, Chang C, Draye X, Hoppe AH, Jessup R, Lemke C, Lennington J, Li Z *et al.*: **A high-density genetic recombination map of sequence-tagged sites for sorghum, as a framework for comparative structural and evolutionary genomics of tropical grains and grasses**. *Genetics* 2003, in press.

20. Bennett MD, Leitch IJ: **Angiosperm DNA C-values database (release 4.0, Jan. 2003)**. http://www.rbgkew.org.uk/cval/homepage.html

21. Bennetzen JL: **Mechanisms and rates of genome expansion and contraction in flowering plants**. *Genetica* 2002, **115**:29-36.

22. •• Bowers JE, Chapman BA, Rong JK, Paterson AH: **Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events**. *Nature* 2003, **422**:433-438.
An early application of 'phylogenomics', a merger of phylogenetic inference with structural genomic information that is expected to be of growing importance in the analysis and interpretation of genomic data.

23. •• Sasaki T, Matsumoto T, Yamamoto K, Sakata K, Baba T, Katayose Y, Wu JZ, Niimura Y, Cheng ZK, Nagamura Y *et al.*: **The genome sequence and structure of rice chromosome 1**. *Nature* 2002, **420**:312-316.
See annotation [24••].

24. •• Feng Q, Zhang YJ, Hao P, Wang SY, Fu G, Huang YC, Li Y, Zhu JJ, Liu YL, Hu X *et al.*: **Sequence and analysis of rice chromosome 4**. *Nature* 2002, **420**:316-320.
These papers [23••,24••], published simultaneously, represent the first essentially completed rice chromosomes. Reference [23••] is distinguished in also being the largest rice chromosome.

25. • The Rice Chromosome 10 Sequencing Consortium: **In-depth view of structure, activity, and evolution of rice chromosome 10**. *Science* 2003, **300**:1566-1569.
Near-completed sequencing and detailed analysis of the smallest and most heterochromatic of the rice chromosomes.

26. • Yu J, Hu SN, Wang J, Wong GKS, Li SG, Liu B, Deng YJ, Dai L, Zhou Y, Zhang XQ *et al.*: **A draft sequence of the rice genome (Oryza sativa L. ssp indica)**. *Science* 2002, **296**:79-92.
See annotation [27•].

27. • Goff SA, Ricke D, Lan TH, Presting G, Wang RL, 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.
Papers [26•,27•], published simultaneously, comprise the first genomic shotgun sequence of a cereal, and shed light on the genomic diversity of cereals from dicot plants such as *Arabidopsis*, as well as non-plant taxa.

28. • Chen MS, Presting G, Barbazuk WB, Goicoechea JL, Blackmon B, Fang FC, Kim H, Frisch D, Yu YS, Sun SH *et al.*: **An integrated physical and genetic map of the rice genome**. *Plant Cell* 2002, **14**:537-545.
Together with [26•,27•], this work sets the stage for a finished sequence of the first cereal genome, emerging now as illustrated in [23••,24••,25•].

29. Zhao Q, Zhang Y, Cheng ZK, Chen MS, Wang SY, Feng Q, Huang YC, Li Y, Tang YS, Zhou B *et al.*: **A fine physical map of the rice chromosome 4**. *Genome Res* 2002, **12**:817-823.

30. • Kellogg EA: **It's all relative**. *Nature* 2003, **422**:383-384.
A succinct and illustrative critique of new methods described in [27•] and elsewhere for application of 'phylogenomics' to the analysis of plant chromosome evolution.

31. Gaut BS: **Patterns of chromosomal duplication in maize and their implications for comparative maps of the grasses**. *Genome Res* 2001, **11**:55-66.

32. • Gaut BS: **Evolutionary dynamics of grass genomes**. *New Phytol* 2002, **154**:15-28.
A 'Tansley review' (by invitation only), providing an excellent overview of existing knowledge, challenges, and opportunities associated with comparative genomics in the grasses.

33. Kishimoto N, Higo H, Abe K, Arai S, Saito A, Higo K: **Identification of the duplicated segments in rice chromosomes 1 and 5 by linkage analysis of cDNA markers of known functions**. *Theor Appl Genet* 1994, **88**:722-726.

34. Nagamura Y, Inoue T, Antonio B, Shimano T, Kajiya H, Shomura A, Lin S, Kuboki Y, Harushima Y, Kurata N *et al.*: **Conservation of duplicated segments between rice chromosomes 11 and 12**. *Breed Sci* 1995, **45**:373-376.

35. Wang SP, Liu KD, Zhang QF: **Segmental duplications are common in rice genome**. *Acta Bot Sin* 2000, **42**:1150-1155.

36. Eckhardt N: **A sense of self: the role of DNA sequence elimination in allopolyploidization**. *Plant Cell* 2001, **13**:1699-1704.

37. Devos KM, Beales J, Nagamura Y, Sasaki T: ***Arabidopsis* - rice: will colinearity allow gene prediction across the eudicot-monocot divide?** *Genome Res* 1999, **9**:825-829.

38. Liu H, Sachidanandam R, Stein L: **Comparative genomics between rice and *Arabidopsis* shows scant collinearity in gene order**. *Genome Res* 2001, **11**:2020-2026.

39. Vandepoele K, Saeys Y, Simillion C, Raes J, Van de Peer Y: **The automatic detection of homologous regions (ADHoRe) and its application to microcolinearity between *Arabidopsis* and rice**. *Genome Res* 2002, **12**:1792-1801.

40. Salse J, Piegu B, Cooke R, Delseny M: **Synteny between *Arabidopsis thaliana* and rice at the genome level: a tool to identify conservation in the ongoing rice genome sequencing project**. *Nucleic Acids Res* 2002, **30**:2316-2328.

41. Song R, Llaca V, Messing J: **Mosaic organization of orthologous sequences in grass genomes**. *Genome Res* 2002, **12**:1549-1555.

42. Yano M: **Genetic and molecular dissection of naturally occurring variation**. *Curr Opin Plant Biol* 2001, **4**:130-135.

43. Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES: **Dwarf8 polymorphisms associate with variation in flowering time**. *Nat Genet* 2001, **28**:286-289.

44. • Tenaillon MI, Sawkins MC, Anderson LK, Stack SM, Doebley J, Gaut BS: **Patterns of diversity and recombination along chromosome 1 of maize (*Zea mays* ssp mays L.)**. *Genetics* 2002, **162**:1401-1413.
An elegant integration of cytological, structural and population genomic data, yielding new insights into how allelic diversity is distributed across genomes.

45. Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doeblay J, Kresovich S, Goodman MM, Buckler ES: **Structure of linkage disequilibrium and phenotypic associations in the maize genome**. *Proc Natl Acad Sci USA* 2001, **98**:11479-11484.

46. Goldberg RB: **From cot curves to genomics. How gene cloning established new concepts in plant biology**. *Plant Physiol* 2001, **125**:4-8.

47. Peterson DG, Wessler SR, Paterson AH: **Efficient capture of unique sequences from eukaryotic genomes**. *Trends Genet* 2002, **18**:547-550.

48. Peterson DG, Schulze SR, Sciara EB, Lee SA, Bowers JE, Nagel A, Tibbitts DC, Wessler SR, Paterson AH: **Integration of Cot analysis, DNA cloning, and high-throughput sequencing facilitates genome characterization and gene discovery**. http://www.ncbi.nlm.nih.gov/entrez (2001)

49. •• Peterson DG, Schulze SR, Sciara EB, Lee SA, Bowers JE, Nagel A, Jiang N, Tibbitts DC, Wessler SR, Paterson AH: **Integration of Cot analysis, DNA cloning, and high-throughput sequencing facilitates genome characterization and gene discovery**. *Genome Res* 2002, **12**:795-807.
Merger of a classic approach with modern methods to 'fractionate' genomes based on the degree of repetition of DNA element families. Together with [47], this illustrates an alternative to 'genomic shotgun sequencing' that may permit us to economically access the sequence complexity (set of unique sequences) in large, highly-repetitive genomes.

50. Yuan YN, SanMiguel PJ, Bennetzen JL: **High-Cot sequence analysis of the maize genome**. *Plant J* 2003, **34**:249-255.

51. Rabinowicz PD, McCombie WR, Martienssen RA: **Gene enrichment in plant genomic shotgun libraries**. *Curr Opin Plant Biol* 2003, **6**:150-156.

52. Altshuler D, Pollara VJ, Cowles CR, Van Etten WJ, Baldwin J, Linton L, Lander ES: **An SNP map of the human genome generated by reduced representation shotgun sequencing**. *Nature* 2000, **407**:513-516.

53. • Jordan B, Charest A, Dowd JF, Blumenstiel JP, Yeh RF, Osman A, Housman DE, Landers JE: **Genome complexity reduction for SNP genotyping analysis**. *Proc Natl Acad Sci USA* 2002, **99**:2942-2947.
A promising reduced-representation method to rapidly identify sufficient populations of SNPs for genetic mapping of genomes for which a minimum of *a priori* information exists.

54. Goodman RM, Naylor R, Tefera H, Nelson R, Falcon W: **The rice genome and the minor grains**. *Science* 2002, **296**:1801.

55. Yuan QP, Quackenbush J, Sultana R, Pertea M, Salzberg SL, Buell CR: **Rice bioinformatics. Analysis of rice sequence data and leveraging the data to other plant species**. *Plant Physiol* 2001, **125**:1166-1174.

56. Yamada T, Kishida T: **Genetic analysis of forage grasses based on heterologous RFLP markers detected by rice cDNAs**. *Plant Breed* 2003, **122**:57-60.

57. Spielmeyer W, Ellis MH, Chandler PM: **Semidwarf (sd-1), ''green revolution'' rice, contains a defective gibberellin 20-oxidase gene**. *Proc Natl Acad Sci USA* 2002, **99**:9043-9048.

58. Lukens L, Doebley J: **Molecular evolution of the teosinte branched gene among maize and related grasses**. *Mol Biol Evol* 2001, **18**:627-638.

59. Paterson AH, Bowers JE, Burow MD, Draye X, Elsik CG, Jiang CX, Katsar CS, Lan TH, Lin YR, Ming RG *et al.*: **Comparative genomics of plant chromosomes**. *Plant Cell* 2000, **12**:1523-1539.

60. Paterson AH, Lin YR, Li ZK, Schertz KF, Doebley JF, Pinson SRM, Liu SC, Stansel JW, Irvine JE: **Convergent domestication of cereal crops by independent mutations at corresponding genetic-loci**. *Science* 1995, **269**:1714-1718.

61. Mitra R, Bhatia CR: **Repeated and non-repeated nucleotide sequences in polyploid wheat species**. *Heredity* 1973, **31**:251-262.

62. Smith DB, Flavell RB: **Nucleotide sequence organisation in the rye genome**. *Biochim Biophys Acta* 1977, **474**:82-97.

63. Wimpee CF, Rawson JRY: **Characterization of the nuclear genome of pearl-millet**. *Biochim Biophys Acta* 1979, **562**:192-206.

64. Hake S, Walbot V: **The genome of Zea-Mays, its organization and homology to related grasses**. *Chromosoma* 1980, **79**:251-270.

65. Gupta VS, Gadre SR, Ranjekar PK: **Novel DNA-sequence organization in rice genome**. *Biochim Biophys Acta* 1981, **656**:147-154.