



Tansley review no. 132

Evolutionary dynamics of grass genomes

Author for correspondence:

Brandon S. Gaut
Tel: +1 (949) 824 2564
Fax: +1 (949) 824 2181
Email: bgaut@uci.edu

Brandon S. Gaut

Department of Ecology and Evolutionary Biology, University of California, Irvine,
CA 92697-2525, USA

Received: 13 August 2001

Accepted: 6 December 2001

Contents

Summary	15	III. Comparative maps and sequencing	21
I. Introduction	16	1. Comparative maps of the grasses	21
II. Grass relationships, chromosome numbers and genome size	17	2. Limitations of map data for evolutionary analyses	22
1. Grass relationships	17	3. Synteny among grass genomes: a reanalysis	23
2. The timescale of grass evolution	17	4. Microsynteny: comparative grass sequences	24
3. The evolution of chromosome number	18	IV. Conclusions	25
4. The evolution of genome size	19	Acknowledgements	26
		References	26

Summary

Key words: genome size, phylogeny, comparative map, synteny, microsynteny.

The grass family (Poaceae) has been the subject of intense research over the past decade. Although other angiosperm families contain more species and more genera, the Poaceae exceed all other families in ecological dominance and economic importance. Research has focused on the evolutionary relationships among grasses as well as the structure of grass genomes. Here I examine the evolutionary dynamics of grass genomes in a phylogenetic framework. It is clear that grass genomes are evolutionarily labile for many characteristics, including genome size and chromosome number. Variation in genome size among grasses probably reflects fluctuations in the amount of repetitive DNA per genome, but the history and causes of chromosome number changes remain unclear. Despite substantial variation among genomes, comparative maps suggest that grass genomes retain extensive regions of colinearity. By reanalyzing some comparative map data and also by reviewing comparative sequence data, I argue that the current colinearity paradigm requires reassessment.

© *New Phytologist* (2002) **154**: 15–28

Abbreviations

Grass Phylogeny Working Group, GPWG; million years, my; Grass Genera of the World, GGW; Angiosperm C-value database, ACV

I. Introduction

The grass family (Poaceae) contains *c.* 10 000 species and 700 genera. Although other angiosperm families contain more species and more genera, the Poaceae exceeds all other families in one important trait: ecological dominance. Grasses are found throughout the globe and can dominate temperate and tropical habitats. Altogether, grasses cover > 20% of the earth's land surface (Shantz, 1954).

Given their ecological dominance, it is not surprising that grasses play a central role in the human endeavor. Grasses are a major food source for humans. Three grain crops – wheat (*Triticum aestivum*), rice (*Oryza sativa*) and maize (*Zea mays*) – are predominant food sources, but the grasses also include several additional and perhaps under-appreciated crops. For example, turfgrasses (*Lolium* and *Festuca* spp.) are a major crop group; in 1992, they generated \$600 million in seed sales in the United States, more that year than any other U.S. crop except corn (Ligon, 1993).

The economic incentive to work on the grasses is substantial, and their ecological dominance makes them intriguing from an evolutionary viewpoint. As a result, grasses have been the subject of intense phylogenetic, ecological, agronomic and molecular study. These studies have progressed particularly rapidly in the last decade, primarily due to the advent of high-throughput molecular biology. High-throughput methods have produced a wealth of genomic information encompassing molecular genetic maps, DNA sequences of large genomic regions, and large data sets for phylogenetic inference. From my perspective, the challenge of these data is to interpret them in a context that provides an accurate and useful picture of the evolutionary history of grass genomes.

To appraise our knowledge of grass genome evolution, this review is organized into two sections. The first section is centered on the phylogeny of the grass family and describes grass genome diversity in terms of genome size (DNA content) and chromosome number. The second section focuses on grass comparative maps, with the purpose of reassessing

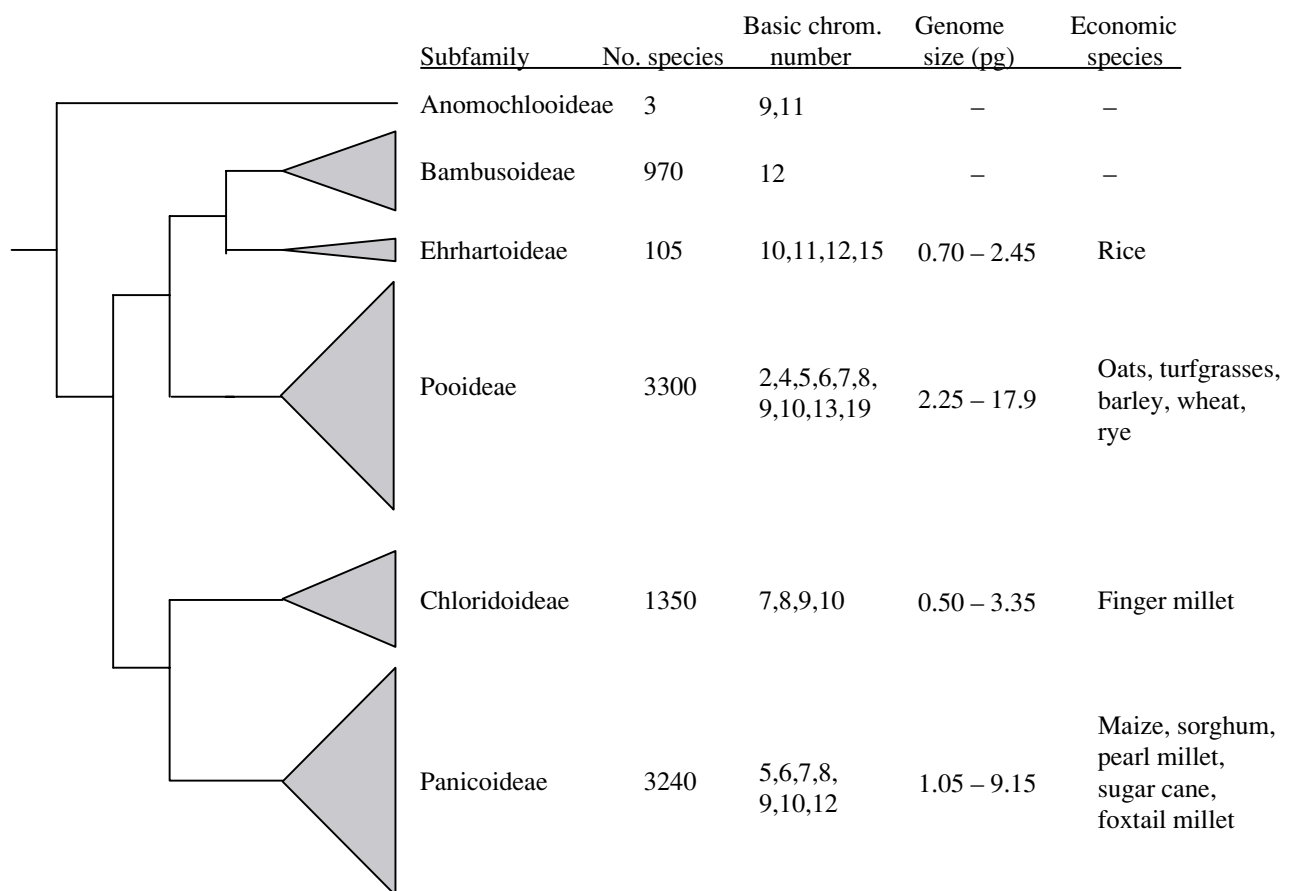


Fig. 1 A phylogeny of the grasses, featuring five major subfamilies and the earliest branching lineage. Each subfamily is represented by a triangle with height proportional to the number of species in the family. The number of species per subfamily is taken from Kellogg (2000). Basic chromosome numbers were collated from the Grass Genera of the World database (<http://biodiversity.uno.edu/delta/grass/index.htm>). Genome size is the range of haploid DNA content, in picograms, for diploid species. DNA content data were collated from the Angiosperm C-value database (<http://www.rbkew.org.uk/cval/>) at Kew Botanic Gardens.

the verity and limitations of conclusions based on map data. To do this, I briefly review the comparative map literature, discuss the interpretation and limitations of genetic maps, and reanalyze some comparative map data. The take-home points of this section are that the probability of randomly identifying an area of synteny between two well-diverged grass genomes can be quite low – on the order of 50% – and that rearrangement of syntenic regions occurs relatively regularly though time. Taken together, several disparate sources of information – for example, phylogeny, DNA content, chromosome number, comparative maps and comparative sequences – suggest that the grass genomes are evolutionarily labile, with perhaps less conservation than previously appreciated.

II. Grass relationships, chromosome numbers and genome sizes

1. Grass relationships

Most taxonomic treatments of the grasses recognize six or seven major subfamilies, with several smaller subfamilies. Initial grass classifications were based on morphological structures like the spikelet, leaf blade (Ellis, 1986) and embryo (Reeder, 1957), but morphology alone failed to unambiguously resolve systematic relationships. As a result, molecular markers have been employed to construct grass phylogenies. Molecular studies initially focused on chloroplast markers, particularly the *rbcL* and *ndhF* genes (Clark *et al.*, 1995; Duvall & Morton, 1996), but more recently phylogenetic studies have been based on nuclear markers like Internal transcript space (ITS) (Hsiao *et al.*, 1998), *waxy* (Mason-Gamer *et al.*, 1998) and *phyB* (Mathews *et al.*, 2000). Some of these molecular and morphological studies have been combined by the Grass Phylogeny Working Group (GPWG) to yield a robust phylogeny of the family (Grass Phylogeny Working Group, 2001). An abbreviated version of the GPWG phylogeny is given in Fig. 1.

Phylogenetic approaches have provided unexpected information about the evolutionary history of the grasses. For example, before molecular phylogenetic analyses, the Anomochlooideae were considered members of subfamily Bambusoideae, and the bambusoids were considered early diverging grasses (Clark *et al.*, 1995). However, it is now known that the Anomochlooideae and Bambusoideae represent divergent grass lineages (Fig. 1), with the anomochlooids representing the basal, or most early diverged, grass lineage. By contrast, the bambusoids fall within a monophyletic group known as the 'BEP' clade because it contains subfamilies Bambusoideae, Ehrhartoideae and Pooideae (Kellogg, 2000, 2001). The latter two subfamilies include the economically important species rice (*Oryza sativa*), wheat (*Triticum aestivum*), barley (*Hordeum vulgare*) and oats (*Avena sativa*) (Fig. 1). The remaining major grass subfamilies fall into a second monophyletic clade, deemed the 'PACC' clade. The PACC clade contains subfamilies Panicoideae, Arundinoideae, Centothecoideae and

Chloridoideae; the phylogenetic placement of two of these subfamilies is given in Fig. 1, as are the names of some of the economically important species from these subfamilies.

The grass phylogeny forms the basis for functional, genomic and evolutionary studies. For example, Kellogg (2000) used the phylogeny to examine the evolution of C₄ photosynthesis in grasses. She found that all C₄ species fall within the PACC clade. Furthermore, the distribution of C₄ plants in the PACC clade suggests that C₄ photosynthesis originated at least four times. The functional implication of these findings is that regulation of C₄ photosynthesis may differ among species with independent origin of C₄ photosynthesis. The phylogeny also provides a conceptual framework for interpreting comparative genomic data. For example, given the inclusion of Ehrhartoideae in the BEP clade, it is clear that rice cannot be considered an ancestral grass genome despite its small size and relatively simple structure. Finally, the phylogeny provides a basis to generate expectations about genome relationships. Because rice and oats share a recent common ancestor, for example, the phylogeny suggests that the genomes of rice and oats should be more similar to one another than either is to maize and other members of the PACC clade.

2. The timescale of grass evolution

To better understand and discuss grass genome evolution, it is helpful to put the divergence of key grass taxa into a temporal framework. Figure 2 provides a phylogeny and divergence times among eight economically important grasses, along with a basal grass (*Anomochloa*) and an outgroup (*Joinvillea*). The tree topology is based on the GPWG phylogeny, but it is also the maximum parsimony topology for the *rbcL* and *ndhF* sequence data used in these analyses (Fig. 2). The divergence times on the nodes of the tree were estimated with the non-parametric rate smoothing method of Sanderson (Sanderson, 1997), assuming that maize and rice diverged 50 million years ago (Stebbins, 1981; Wolfe *et al.*, 1987). Sanderson's method does not assume a molecular clock, and for this reason the divergence time estimates in Fig. 2 may be improvements over some previously published estimates.

The estimates suggest the grass family originated roughly 77 million years (my) ago. The age of the family has previously been reported to be 55–70 my based on fossil evidence (Linder, 1987; Jacobs *et al.*, 1999; Kellogg, 2001). Because fossil data can only provide a minimum age of divergence, the higher estimate of 77 my seems reasonable. The divergence between Ehrhartoideae (rice) and the Pooideae (oats, barley and wheat) is estimated at 46 my, and this represents the time of origin of the BEP clade. Within the Pooideae, the Triticeae (barley and wheat) diverged from oats *c.* 25 my. Barley and wheat diverged *c.* 13 my ago, and this estimate is similar to a previous estimate of 10 my (Wolfe *et al.*, 1989).

Within the subfamily Panicoideae, the highest divergence estimate corresponds to the divergence between sorghum

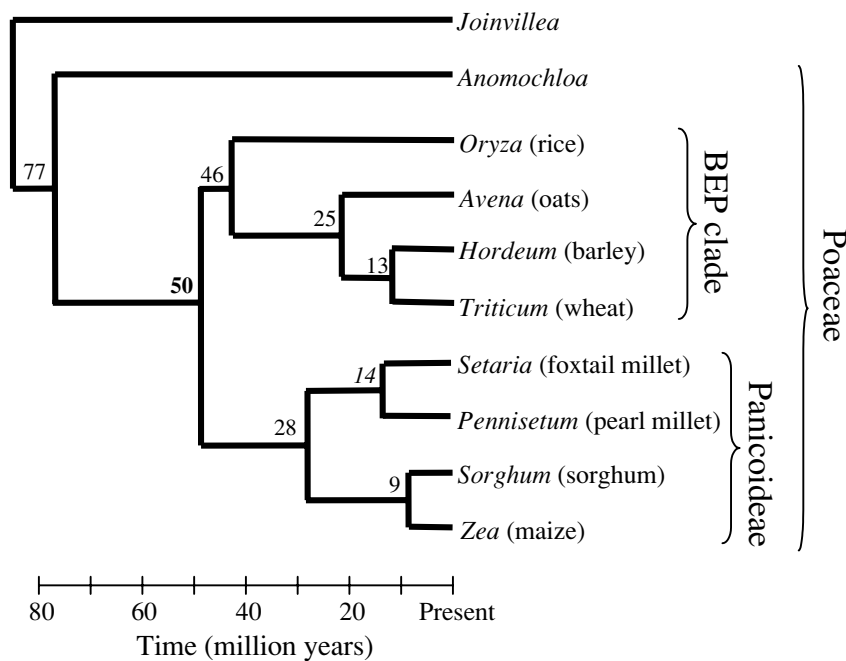


Fig. 2 The phylogeny of several well-studied grass species. The numbers on the node represent divergence times, as estimated by the method of Sanderson (1997). The number in bold is an assumed divergence time; nonitalic numbers on nodes are based on combined data from the chloroplast genes *ndhF* and *rbcl*; and numbers in italics are based on *rbcl* data alone. The genera *Sorghum* and *Zea* are members of the tribe Andropogoneae. All data are from GenBank.

and maize from pearl millet (*Pennisetum*) and foxtail millet (*Setaria*). The estimated divergence of 28 my closely matches a previous estimate of 30 my based on sequence data from maize and pearl millet nuclear genes (Gaut & Doebley, 1997). However, at least one estimate within the Panicoideae does not closely correspond to a previous estimate. Using data from two nuclear genes, Gaut & Doebley (1997) previously estimated the divergence between sorghum and maize to be 16.5 my, whereas chloroplast data provide an estimate of 9 my (Fig. 2). At present, it is not clear if the differences between the two estimates are due to different data sources (nuclear vs chloroplast genes) or statistical error. Because the sorghum-maize divergence time is important for understanding the history of the maize genome (Gaut & Doebley, 1997; Gaut *et al.*, 2000), this divergence time should probably be reexamined thoroughly.

It is important to recognize the limitations of divergence estimates. At least three factors contribute to uncertainty in these estimates. First, the assumed rice-maize divergence time of 50 my is based on unconvincing evidence (e.g. White & Doebley, 1999). Second, these estimates are based on sequences from two chloroplast genes; sequence data from other genes or genomes (i.e. nuclear or mitochondrial) may produce different estimates, as exemplified by the example of sorghum and maize. More thorough analyses with data from more genes and genomes will be enlightening but are beyond the scope of this review. Finally, the nonparametric method was designed to estimate rates in the absence of a molecular clock, but the method still assumes that evolutionary rates are auto-correlated across phylogenetic branches (Sanderson, 1997). Despite these caveats, the estimates in Fig. 2 provide a rough timeframe for the evolution of the grasses and hence are useful for further discussion.

3. The evolution of chromosome number

One interesting facet about the grass family is that chromosome number has fluctuated widely over 77 my of evolution. Variation in chromosome number is partly a consequence of polyploidy because extant grass polyploids comprise roughly 44% of the species in the family (DeWet, 1986). Nonetheless, variation in chromosome number cannot be attributed to polyploid events alone. Figure 1 provides basic chromosome numbers for genera within subfamilies. These basic chromosome numbers do not reflect recent increases in chromosome number due to extant polyploidy but can, in principle, reflect historical polyploid events (Stebbins, 1985). The basic chromosome numbers in Fig. 1 were collated from the Grass Genera of the World (GGW) database (<http://www.biodiversity.uno.edu/delta/grass/index.htm>), using taxa with unambiguous subfamily assignments. There was relatively little information about some subfamilies, however, so it is a near certainty that Fig. 1 underestimates basic chromosome number variation within subfamilies.

An interesting feature about chromosome number is the pattern of variation both within and among subfamilies. For example, pooid and panicoid genera vary substantially in chromosome number, with 10 and 7 different basic chromosome numbers within each subfamily (Fig. 1). Basic chromosome number also varies substantially among subfamilies. No single basic chromosome number is shared by all the six grass subfamilies listed in Fig. 1. For example, the panicoids, chloridoideae and pooideae share four basic chromosome numbers in common ($x = 7, 8, 9, 10$), but these numbers are different than the only basic number shared by the bambusoideae and ehrhartoideae ($x = 12$). Because the GGW does not include data

for all grass genera, caution needs to be used for interpreting the numbers in Fig. 1. Indeed, basic chromosome numbers for subfamilies reported by de Wet (1986) vary from those in Fig. 1; some of the variation between this study and De Wet (1986) likely reflects differences in sampling as well as changes in classification over the past 15 yr.

Historically the basic chromosome number of grasses has received much attention. For example, Avdulov (1931) measured chromosome numbers for hundreds of grasses and speculated that the ancestral chromosome number of grasses was $x = 12$, with smaller basic chromosome numbers derived by aneuploid reduction. Flovik (1938) proposed an ancestral basic number of $x = 5$, whereas Sharma (1979) suggested the ancestral basic number was $x = 6$. Stebbins (1985) finally concluded that ancestral basic chromosome numbers of $x = 5, 6, 7$ were equally probable, with higher species' chromosome numbers derived either by polyploidy, by polyploidy followed by aneuploidy or by combinations (hybridization) of basic numbers. In short, the ancestral basic chromosome number of the grasses is uncertain, but many historical polyploid and/or aneuploid events are required to explain adequately the current distribution of basic chromosome numbers among grass taxa.

Uncertainty in basic chromosome number applies to more recently derived taxonomic groups as well. One example suffices to illustrate the point. The Andropogoneae is a tribe within subfamily Panicoideae that originated < 28 my ago (Fig. 2) and is thus much more recent than the grasses as a whole. Traditionally it has been assumed that the basic haploid chromosome number of the Andropogoneae was $n = 5$ (Celarier, 1956; Molina & Naranjo, 1987). More recently it has been suggested that the basic haploid chromosome of the tribe was $n = 10$ (Spangler *et al.*, 1999), based on phylogenetic arguments, and $n = 8$, based on comparative maps (Wilson *et al.*, 1999). It is not clear which number is correct, but all basic numbers require extensive chromosomal losses and gains within the tribe (Gaut *et al.*, 2000). As a result, the evolution of chromosome number is difficult to trace even in relatively recent grass groups.

How does basic chromosome number change? Polyploid events are common; they multiply the number of chromosomes in a taxon and lead to increased chromosome numbers over time. However, the mechanisms that lead to loss and gain of single chromosomes are less obvious, as are the mechanisms that lead to re-diploidization of polyploid genomes. It is known, however, that genomes can rearrange rapidly after polyploid events (Wendel, 2000), and this has been demonstrated experimentally. In one study, Song *et al.* (1995) created four synthetic *Brassica* allopolyploids, each of which was selfed from the F_2 to the F_5 generation. Each generation was subjected to Southern hybridization with a panel of 89 probes, and these probes revealed remarkable differences in fragment profiles from generation to generation. In one synthetic polyploid, 66% of the probes detected fragment loss, fragment gain or a change in fragment size through time, demonstrating

rapid genomic change after allopolyploid formation. Similar studies in *Triticum* and *Aegilops* suggest that allopolyploids lose non-coding sequences in a nonrandom fashion and that coding sequences can be extensively modified (Feldman *et al.*, 1997; Liu *et al.*, 1998a, 1998b).

It has not been demonstrated that rapid rearrangement in synthetic allopolyploids leads to chromosome loss or complete diploidization. However, it is clear that many extant diploid plants contain duplicated chromosomal regions that owe their origin to an ancient polyploid event. The list of ancient polyploid plants includes maize, soybean, *Brassica* species, cotton (which is an extant polyploid in addition to an ancient polyploid), and – perhaps most surprisingly – *Arabidopsis thaliana* (arabidopsis). Whole-genome sequence indicates that the arabidopsis genome is structurally complex, with *c.* 70% of the genome duplicated (Arabidopsis Genome Initiative, 2000). Furthermore, patterns of sequence divergence suggest that arabidopsis genome duplication was likely caused by five large-scale duplication events, each of which may have been a polyploid event. The five polyploid events are estimated to have occurred 50, 100, 140, 170, and 200 million years ago (Vision *et al.*, 2001), with the four most recent probably occurring after the divergence of monocots and dicots (Wolfe *et al.*, 1989; Laroche *et al.*, 1995). The emerging picture from arabidopsis and other plant taxa is that polyploidy, followed by chromosome rearrangement, is evolutionarily common.

Altogether, the grass chromosome numbers in Fig. 1, combined with examples of polyploid genome evolution, suggest that plant genomes are evolutionarily labile, with frequent chromosomal loss, chromosomal gain and perhaps commensurate genome rearrangement. These events must affect gene content and genome organization. Yet, comparative maps indicate that gene order has been conserved for many genomic segments throughout the 77 my history of grasses (Devos & Gale, 2000). Apparent inconsistencies between the rapid genomic change implied by chromosome numbers and the apparent conservation of genomes suggested by comparative maps will be discussed below.

4. The evolution of genome size

Grass taxa differ in chromosomal number and also exhibit extensive variation in genome size. Figure 1 also provides a range of genome sizes for grass subfamilies, with each range representing haploid genome contents of diploid species. All genome size estimates were taken from the Angiosperm *C*-value (ACV) database (<http://www.rbgekew.org.uk/cval/>), which summarizes decades of genome size measurements by Bennett and colleagues (Bennett & Smith, 1976, 1991; Bennett & Leitch, 1995). The range reported in Fig. 1 does not include data for species listed as polyploid or possibly polyploid, and it also does not contain data from genera with unconfirmed subfamilial classification. Classification was based on the GGW database.

Genome size estimates demonstrate two features of grass genome evolution. First, grass genomes vary considerably in size. For the subfamilies in Fig. 1, DNA content differs 36-fold between the smallest (*Oropetium thomaeum*) and largest diploid (*Psathyrostachys fragilis*) genomes in the database. Genome sizes also vary up to eightfold within subfamilies. Thus, rapid change in DNA content, as well as chromosome number, is a hallmark of grass genome evolution. Second, subfamilies differ in DNA content. For example, chloridoid species have low DNA contents, with the highest value at 3.35 pg DNA per 2C (haploid) nucleus, but pooids have relatively large genomes, with the smallest measured 2C nucleus at 2.25 pg. Comparison of DNA content between subfamilies should be made cautiously, because some subfamilies, like the chloridoids and the bambusoids, have not been sampled extensively. Nonetheless, the Pooideae and the Panicoideae have been sampled extensively (with 25 and 113 diploid taxa sampled, respectively), and there is strong statistical evidence that these two groups differ in DNA content (Mann–Whitney test; $U = 2438$; $P < 0.0001$). This difference could reflect sampling phenomena, and hence robust conclusions require additional taxon sampling. In any event, the available data suggest that DNA content has a phylogenetic component, with some grass clades containing higher DNA contents on average.

What mechanisms contribute to fluctuations in DNA content across grass species? Variation in genome size cannot be attributed solely to increases or decreases in chromosome number; in fact, for the 178 grass diploid species listed in ACV, there is a slight but significantly negative correlation between chromosome number and DNA content (Kendall coefficient = -0.37 ; $P < 0.001$). Instead of chromosome number *per se*, it is probable that the gain and loss of repeat sequences is the primary contributor to differences in DNA content between taxa. This point was first made by Flavell *et al.* (1974), who found that repetitive DNA (defined as DNA with more than 100 copies per genome) constitutes *c.* 80% of angiosperm genomes with a haploid DNA content greater than 5 picograms (pg). By contrast, they found that plant genomes with > 5 pg DNA content contain only 62% repetitive DNA, on average (Flavell *et al.*, 1974).

Grass studies support the view that genome size variation is largely a function of repetitive DNA. For example, barley and rice have similar complements of low-copy genes but a 12-fold difference in DNA content; most of this difference is attributable to amounts of repetitive DNA (Saghai-Maroo *et al.*, 1996). Similarly, comparative sequencing of the *adh1* region in sorghum and maize demonstrate that the two species vary threefold in length in this region (Tikhonov *et al.*, 1999). The length difference is primarily attributable to retrotransposons, which are absent from the sorghum *adh1* region but comprise 74% of the maize *adh1* region. Altogether, differences in the complement and number of retrotransposons explain much of the fourfold difference in DNA content between

maize and sorghum (SanMiguel *et al.*, 1998; Tikhonov *et al.*, 1999).

Genome size can change rapidly. The best illustration of rapid size change also comes from studies of the maize *adh1* region (Springer *et al.*, 1994; SanMiguel *et al.*, 1996, 1998; Tikhonov *et al.*, 1999). In these studies, Bennetzen and coworkers isolated a 280-kb YAC clone of the maize *adh1* region and characterized the composition of the repetitive intergenic DNA. The region contained 23 retrotransposons representing 10 distinct families, and these 10 families constitute *c.* 50% of the maize genome. By sequencing long-terminal repeats (LTRs) of retrotransposons and by applying molecular clock analyses, SanMiguel *et al.* (1998) were able to estimate the time of insertion of 17 of the 23 retrotransposons. Fifteen of 17 retrotransposons inserted into the *adh1* region within the last 3.0 my, and the oldest retrotransposon inserted *c.* 5.2 million years ago. If the *adh1* region is representative of genome-wide retrotransposon activity, the results imply that 50% of maize DNA content is attributable to retrotransposon proliferation during the last 5–6 my. In phylogenetic terms, 5–6 my is roughly the time of divergence between *Zea* and its sister genus *Tripsacum* (Hilton & Gaut, 1998; White & Doebley, 1999), and hence 5–6 my is a short time-scale relative to the 77 my age of the grass family. Although additional studies of this kind are lacking, it is likely that other grasses have experienced similarly rapid changes in genome size.

The proliferation of repetitive DNA is likely biased with respect to genomic region. For example, retrotransposons in the maize *adh1* region preferentially insert within the LTRs of other retrotransposons (SanMiguel *et al.*, 1996). Insertion within noncoding regions may be a successful evolutionary strategy for these 'selfish-genes', because it ensures that insertion does not interrupt genes of essential function, thereby killing the plant host. If insertion biases are general, it is easy to envision rapid physical expansion of repeat-rich regions without commensurate expansion of repeat-poor regions. As a result of repeat expansion, grass genomes are structurally heterogeneous, consisting of gene-rich and gene-poor regions (Barakat *et al.*, 1997). For example, a gene-rich region around the maize *bronze* gene contains 10 genes in 32 kilobases (kb), for an average density of one gene per 3.2 kb (Fu *et al.*, 2001). By contrast, the estimated average gene density for maize is one gene per 50 kb (Tikhonov *et al.*, 1999). Similar gene-rich regions have been isolated in barley (Panstruga *et al.*, 1998; Shirasu *et al.*, 2000) and wheat (Endo & Gill, 1996; Gill *et al.*, 1996).

Grass genomes can increase rapidly in size by gaining retrotransposons and other repetitive sequences, but can they also decrease rapidly in size by losing repetitive sequences? Unfortunately, this question has not been addressed with rigor in plants. One can, however, look to recent animal studies to begin to formulate expectations. Petrov and colleagues have examined rates and patterns of spontaneous deletion in animal pseudogene sequences, including retrotransposon

remnants (Petrov *et al.*, 1996, 2000). They found that the rate of DNA loss in *Drosophila* is 60 times higher than that of mammalian genomes and 40 times higher than that of Hawaiian crickets. The rate of spontaneous deletion correlates with genome size; *Drosophila* has a *c.* 20-fold smaller genome than humans and an 11-fold smaller genome size than the Hawaiian cricket. These studies demonstrate that some genomes are better able to combat 'genomic obesity' (Bennetzen & Kellogg, 1997) via a high rate of spontaneous deletion.

Molecular mechanisms for spontaneous deletion are not yet clear, but the sequence of a 66-kb barley fragment has led to one reasonable hypothesis (Shirasu *et al.*, 2000). The barley sequence contains numerous retrotransposons that lack one LTR. Shirasu *et al.* (2000) hypothesized that solo-LTRs are remnants of unequal crossing-over events that removed the matching LTR. If this hypothesis is correct, unequal crossing-over counteracts genome expansion. Rates of genome expansion and contraction are probably dependent upon myriad factors in addition to unequal crossing-over. These factors include: the types of elements that have invaded a genome (e.g. repetitive sequences, retrotransposons, DNA transposons, etc.), genome wide rates of mutation and spontaneous deletion, selective pressures for and against repeat proliferation, and stochastic events (Petrov, 2001). All of these factors need to be studied in much greater detail to facilitate an understanding of the forces underlying the evolution of grass genome size.

Although repetitive DNA is the primary contributor to genome size differences among grass taxa, it is important to note that differences in gene content (or gene copy number) probably also contribute to genome size. For example, sorghum and maize differ fourfold in DNA content, but retrotransposons apparently account for only a twofold difference in genome size (SanMiguel *et al.*, 1998). The additional twofold difference reflects an ancient polyploid event that duplicated genes as well as nongenic regions (Gaut *et al.*, 2000). Maize may not be a typical example because it has long been known to have a particularly complex genome (Helentjaris *et al.*, 1988). Nevertheless, fluxes in chromosome number and genome size among grasses must include fluctuations in gene content.

III. Comparative maps and sequencing

Despite extensive research on the ecology and phylogeny of a broad array of grasses, most work on grasses has focused on the few key crops listed in Fig. 1. Molecular genetic maps have been made for all of these crops, facilitating detailed comparison of genome structure and gene order. Map comparisons have revealed that grass genomes share large regions of synteny (in this context, shared molecular markers between chromosomes without regard to marker order) and colinearity (shared markers and shared order). These observations have contributed to the current paradigm of grass genome evolution, which asserts that grass genomes consist of *c.* 30 chromosomal 'building blocks' that have been shuffled through evolutionary

time (Devos & Gale, 2000). It is difficult to reconcile this paradigm with other aspects of grass genome evolution – that is, extensive variation in DNA content and chromosome number. The purpose of this section is to review comparative map and sequence data in order to reassess the current paradigm of grass genome evolution.

1. Comparative maps of the grasses

Genetic maps were first constructed with morphological and isozyme markers. These early maps indicated that linkage relationships among isozyme markers were often conserved among grass species, implying that gene order is also conserved (Hart, 1983). Studies of ribosomal DNA and 5S DNA corroborated isozyme studies, because the chromosomal position of both rDNA and 5S DNA are conserved across some grass species (Payne *et al.*, 1985; Appels *et al.*, 1986; Lawrence & Appels, 1986). Thus, early mapping studies suggested that gene order is conserved among grass taxa.

In the 1980s, RFLP markers became the method of choice for genetic map construction, and RFLP maps were eventually produced for the grass taxa listed in Figs 1 and 2. Many studies mapped the same RFLP markers to two or more species, and by the early 1990s there were sufficient data to compare genetic maps across grass genomes. Initial comparisons involved the A, B and D genomes of wheat (Chao *et al.*, 1988; Devos *et al.*, 1993b), as well as the genomes of wheat, barley and rye (Devos *et al.*, 1993a) and the genomes of maize and sorghum (Whitkus *et al.*, 1992). The basic conclusion of these studies was similar to that suggested by early isozyme data – that is, despite some rearrangement, most markers retain order across grass genomes. In 1995, the amount of data was sufficient to summarize grass genome relationships, and Moore produced the now-famous 'circle-format' grass map (Moore *et al.*, 1995). In one insightful swoop, Moore *et al.* (1995) provided a diagrammatic method to summarize grass genetic map information and also argued convincingly that grass genetic maps can be viewed as a reorganization of basic building blocks (or linkage groups).

The amount of genetic map data has multiplied substantially over the past few years. The data are too voluminous to summarize here, but mapping data have been reviewed several times recently (Gale & Devos, 1998; Devos & Gale, 2000; Paterson *et al.*, 2000). The basic conclusions of the reviews are that: gross chromosomal organization has remained largely conserved during *c.* 77 my of grass evolution; 30 rice linkage blocks adequately represent extant grass genomes, but these blocks are rearranged among grass taxa; and homologous blocks will prove useful for predicting the position of genes conferring key agronomic traits (Devos & Gale, 2000). This last point is especially important because it implies that knowledge gained about a trait in one grass species can be applied to other grasses, making the grasses a 'single genetic system' (Bennetzen & Freeling, 1993). That the grasses are a 'single

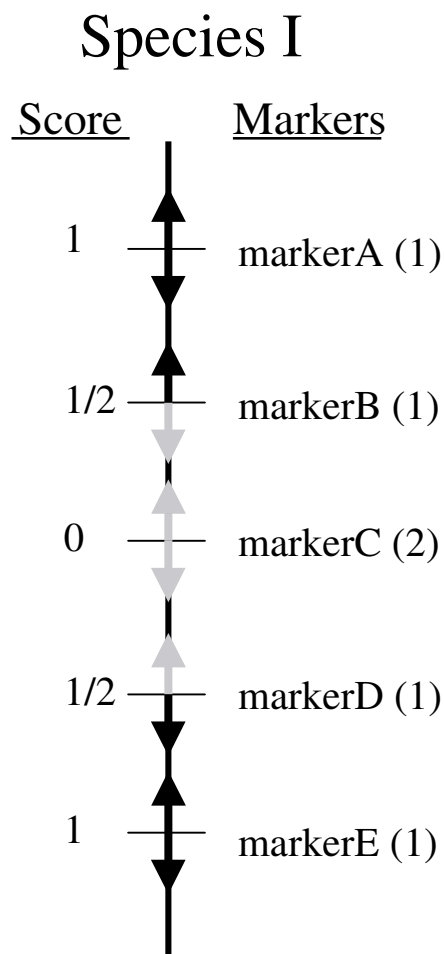
genetic system' has apparently been confirmed by several demonstrations that QTLs for important agronomic traits (like shattering; Paterson *et al.*, 1995) map to homologous regions in different grass genomes.

2. Limitations of map data for evolutionary analyses

The pervading theme of comparative grass studies is that gene order is conserved across genomes, with rearrangements among linkage groups distinguishing taxa. This is a valuable contribution to our understanding of grass genome evolution, but this conclusion is incomplete because of the limitations of map data and their analyses. The data themselves are limited in at least three ways. First, most – if not all – genetic maps are based on low copy-number markers. Low copy-number markers are systematically biased against detecting homologies (or duplications) within a genome and therefore systematically underestimate genomic complexity. A revealing glimpse of the potentially misleading nature of this bias comes from arabidopsis. As previously noted, the arabidopsis genome sequence indicates that 70% of the genome is duplicated, and most of this duplication is located in chromosomal blocks. By contrast, only 10% (Kowalski *et al.*, 1994) and 17% (McGrath *et al.*, 1993) of RFLP markers included in arabidopsis genetic maps show evidence of genetic duplication. It is worth noting that unmapped markers showed a much higher incidence of duplication; 86% and 51% of RFLP markers, respectively, were not single copy. However, only a small percentage of multicopy RFLPs were mapped, owing either to limited polymorphism or a bias toward mapping single-copy markers. The important point is that arabidopsis genetic maps substantially underestimated the amount of the genome that is actually duplicated. Given this precedent, it is likely that grass comparative maps also grossly underestimate grass genome complexity.

Second, most genetic maps have low resolution, with average densities of < 1 marker per 10 centimorgans (cM) (Bennetzen, 2000). These densities ensure that most small (< 10 cM) rearrangements are not detected. If small rearrangements are common, their contribution to non-colinearity may be systematically underestimated by genetic maps. Finally, it is easier to locate markers in regions of high polymorphism, and hence maps overemphasize polymorphic genomic regions. As a result, physical regions of systematically low polymorphism, like centromeric regions (Dvorak *et al.*, 1998; Kraft *et al.*, 1998), are mapped sparsely. It is therefore unlikely that rearrangements are detected in some relatively large genomic regions, but it is unclear to what extent this phenomenon affects comparative map interpretation.

There are also analytical problems. As detailed by Bennetzen (2000), 'circular reasoning' biases the choice and interpretation of markers. This bias is introduced when an RFLP marker hybridizes to several loci but only one locus is polymorphic and mapped. If the mapped locus is in a colinear position in one species relative to another species, it is assumed to be an



$$\text{Synteny probability} = (1 + \frac{1}{2} + 0 + \frac{1}{2} + 1)/5 = 60\%$$

Fig. 3 A hypothetical example of five RFLP markers (markers A–E) that were mapped in two species (Species I and II). Their map positions in Species I are given by their order on the chromosome. Their map location for Species II is given in parentheses. For example, markerA maps to chromosome 1 of Species II, whereas markerC maps to chromosome 2 of Species II. In many cases the five markers would be interpreted as representing a region of homology between Species I and chromosome 1 of Species II. While this is probably a reasonable interpretation, conflicting information from markerC is often ignored. For the data in Table 1, marker data was interpreted in the following way. MarkerC receives a score of 0, because moving in either direction from markerC moves into a region of nonsynteny (gray arrows). For markerB, one direction leads to a nonsyntenic region while the other direction leads to a syntenic region (black arrows); hence moving from markerB has a 50% probability of moving into a syntenous region. In this example, the average probability of moving into a region of synteny is 60%.

ortholog. If it is not in a colinear position, it is often assumed that the locus is a paralog and interpreted as such. The net effect of this circularity is an over-emphasis on colinearity. Figure 3 provides an example of a case in which colinearity

is reasonably inferred, yet the inference ignores (or at least discounts) information from one marker. From a statistical and experimental standpoint, discounting information from any marker makes little sense, because each marker is mapped with the *a priori* expectation that it will provide useful information.

The second analytical problem is intimately associated with circularity; the problem is the lack of objective statistics for delineating regions of chromosomal homology. In some cases, authors rely on synteny to define chromosomal homology, and in other cases colinearity is used as evidence for homology. More importantly, the criterion for choosing a region of homology is rarely (if ever) stated. For example, when comparing maize and rice, are four colinear markers sufficient to declare regions of homology, or should more (or fewer) markers be required? What if one noncolinear marker interrupts several colinear markers (Fig. 3)? If synteny is the criterion, how many markers make a region homologous? Any answer to these questions is of necessity subjective, but by answering these questions at least some criterion is defined. Unfortunately, criteria have not been defined in most comparative mapping studies, leaving the reader uninformed about how homology relationships are identified and also discounting the value of conclusions.

Ideally, determination of homology should be answered in a statistical context. Recently Gaut (2001) introduced a first-step toward building such a context by introducing a simulation method to test whether colinear runs of markers are expected at random (i.e. are consistent with statistical noise) or provide evidence of underlying nonrandom pattern. Application of this method detected roughly 2.5-fold more homeologous regions within maize than previously noted. The method also facilitates estimation of the proportion of the genome that is duplicated. For maize, the current estimate is that roughly 80% of the genome is duplicated, and, as importantly, as much as one-third of the genome may be multi-copy. This method has not yet been applied to cross-species

comparisons, but this and similar methods will improve objectivity in map interpretation.

In the absence of physical maps and whole-genome sequence, marker-based mapping is still the most accessible way to gain a broad overview of whole-genome (or nearly whole-genome) structure and organization. Nonetheless, the limitations of genetic map data for comparative and evolutionary inference are substantial.

3. Synteny among grass genomes: a reanalysis

Ultimately the issue of synteny is important for functional applications; the impetus for finding synteny is to apply information from one species to a second species. Clearly there are large regions of synteny among grass genomes, but one must wonder whether synteny is useful for cross-species studies. One way to address this issue is to ask: what is the probability that any randomly chosen gene (or marker) is in a syntenous region? A simple attempt to answer this question is given in Table 1.

The data for Table 1 were gleaned from several sources, but the data for all species comparisons were collated in the same way. First, the total number of markers mapped between two species was counted. Next, each marker was given a score of either 1, $1/2$, or 0. The three scores correspond to markers for which both flanking markers are syntenous, for which one of two flanking markers is syntenous, and for which no flanking markers are syntenous, respectively (see Fig. 3). These scores represent the probability of moving randomly from the marker into a region of synteny, as defined by the marker of interest, its flanking marker and the position of the two markers in both species. The summation of scores, divided by the total number of markers, provides an average probability that a chromosome walk away from a marker will proceed into a region of synteny. Note that this treatment of the data weighs each marker equally – that is, all markers are assumed to provide information.

Table 1 Reanalysis of comparative map data

Taxa ¹	Total markers ²	Marker scores ³	Synteny probability	Divergence time (my) ⁴	Syntenic loss rate (% per 10 ⁶ yr) ⁵	Source
Rice-barley	103	52.5	51.0%	46	0.54	Saghai-Marooif <i>et al.</i> (1996)
Triticeae-rice	122	61.5	50.4%	46	0.54	Van Deynze <i>et al.</i> (1995b)
Oat-rice	83	42	50.6%	46	0.54	Van Deynze <i>et al.</i> (1995a)
Triticeae-oat	160	94	58.8%	25	0.83	Van Deynze <i>et al.</i> (1995b)
Foxtail millet-rice	143	62	43.4%	50	0.57	Devos <i>et al.</i> (1998)
Rice-maize	202	106	52.5%	50	0.48	Wilson <i>et al.</i> (1999)
Maize-sorghum	53	38.5	72.6%	9 (16.5)	1.52 (0.83)	Pereira <i>et al.</i> (1994)

¹Taxa are shown with the mapped species first. For example, the rice map included markers that contained information about the chromosomal location of the markers in barley. ²The total number of markers reported may differ from those reported in citations, because only markers with 'high support' (most studies used LOD values > 2.0 as evidence for high support) were used in analyses. Inclusion of markers with lower long odds ratio (LOD) values made little qualitative difference in results. ³Summation of the score for all markers, as per Fig. 3. ⁴Divergence times are based on chloroplast data analyses in Fig. 2, except the time in parentheses, which is taken from Gaut & Doebley (1997) and based on nuclear sequence data. ⁵Rate of syntenic loss is formulated as (100 – synteny probability) divided by 2 times the divergence time.

Several pairs of grass species are compared in Table 1, and two points are clear. First, the average probability of moving from a marker into a syntenous region is not exceptionally high for any of the species pairs examined. For example, the average probability of randomly moving from a marker into a syntenic region between maize and sorghum is only 73%, despite the fact that these species have diverged relatively recently (Table 1). The 73% probability can be interpreted in the following manner: if a researcher knows that a gene (or QTL) of interest is near a marker in sorghum, there is only an *a priori* 73% probability that the same gene is near that marker in maize. Because 73% is an average, it is obvious that some genomic regions have higher synteny probabilities, and other regions have lower probabilities. However, the average probability is < 50% for more diverged species like foxtail millet and rice (Table 1). It is important to note that the probabilities do not imply that there is no synteny between genomes. (In fact, genomes with no synteny should have probabilities that approach but do not reach 0.0%; the limiting probability is a function of the number of markers under comparison.) Nonetheless, these observations do raise the issue as to whether synteny is extensive enough to justify the study of small genome grasses (e.g. rice or sorghum) as a proxy for more complex genomes (e.g. wheat or maize). I should note that probabilities for strict colinearity, as opposed to synteny, will be substantially smaller than the probabilities given in Table 1.

The second point from Table 1 is that the rate of loss of synteny is reasonably steady. In the comparison between rice and barley, for example, the rate of loss of synteny is 0.54% per my. The rate of synteny loss is very similar for 5 of the 7 comparisons in Table 1, all of which include rice. However, there is some variation in rate, with two comparisons, maize-sorghum and Triticeae-oat, suggesting a rate at least 1.5-fold higher. It is not clear if this higher rate reflects bouts of rapid genome rearrangement in these lineages or rather reflects statistical oddities in the data. Previous studies have indicated that the rate of genome rearrangement has not been constant through grass evolution (Gale & Devos, 1998). The 1.5-fold difference in rates reported here is consistent with the previous finding, but additional studies of rearrangement rates are merited.

I would like to reemphasize that the average probability of synteny is not extremely high among the species pairs in Table 1 and also that low probabilities do not imply there is no colinearity among taxa, because there certainly are highly conserved regions among grass taxa. Nonetheless, the relatively low probabilities in Table 1 can be interpreted as an indication that genome rearrangement in grasses is extensive, resulting in many exceptions to colinearity. As a result, colinearity in the grasses is reduced to the famous scenario of the 'half-filled' glass. Is the glass half-empty or is it half-full? In other words, do grass genomes contain extensive colinearity or are they substantially rearranged? Given limitations of map data, evolu-

tionary lability of chromosome number and vast variation in genome size, my (admittedly subjective) belief is that there is insufficient data to argue that grass genomes are sufficiently similar to consider them either 'well-conserved' or a 'single genetic system'.

4. Microsynteny: comparative grass sequences

With some grass physical maps nearing completion and the whole-genome of rice sequenced but not yet freely available, new tools will soon be available for investigating grass genome evolution. Thus far, few studies have addressed conservation of grass genomes using comparative sequence data. Although these studies have limitations in their own right, they provide a second means of evaluating colinearity and genome conservation in grasses.

These studies have been collectively called 'microsynteny' studies because they examine synteny at the DNA sequence level. Microsynteny studies have been reviewed recently (Bennetzen, 2000). For our purposes it is sufficient to ask whether microsynteny studies are consistent with the paradigm of extensive gene-order and genome conservation in the grasses or instead consistent with the dynamic picture of grass evolution provided by data like synteny probabilities, genome size and chromosome numbers. Before discussing microsynteny data in detail, however, it is important to recognize two limitations of the basic approach. First, unlike genetic map data, microsynteny data fail to provide a 'whole-genome' view. Conclusions are necessarily limited to the regions under study. Second, the sequences under study are subject to ascertainment biases. Because conserved probes are used to isolate the region from multiple species, isolation necessarily targets regions that may be relatively well conserved.

Despite these limitations, microsynteny studies have been insightful. Several aspects of these studies have already been discussed – for example, solo-LTRs in barley, retrotransposon proliferation in maize, etc. – and hence a full summary is not necessary here. Instead, I would like to comment on two studies that provide contrasting views of grass genome evolution. The first study, a comparison of the *sh2-a1* region of sorghum, maize and rice (Chen *et al.*, 1998), found that the four genes in the *sh2-a1* region were conserved and collinear among taxa, substantiating that gene order can be well-conserved among grasses. The striking aspect of this study was that *sh2* and *a1* were physically separated by 140 kb in maize but only *c.* 19 kb in rice and sorghum (Chen *et al.*, 1997), the distances among genes reflecting differences in the amount of intergenic repetitive DNA. Another surprising feature was that one putative gene had lost its zinc finger domain in sorghum relative to rice, suggesting functional divergence of this gene. Overall, studies of the *sh2-a1* region indicate retention of colinearity despite putative functional divergence.

By contrast, the *adh1* region has undergone substantial rearrangement in sorghum, maize and rice. Nine genes were

shared in colinear order between maize and sorghum, but three genes were missing from this region in maize relative to sorghum (Tikhonov *et al.*, 1999). By contrast, the rice *adh1* region exhibited little colinearity between sorghum and maize; the only apparent commonality among species was the *adh1* gene itself (Tarchini *et al.*, 2000). One sobering aspect of the rice *adh1* study was that 8 of 13 putative rice genes did not cross-hybridize to maize, suggesting either gene deletion in maize or high sequence divergence between rice and maize homologs. Whatever the cause, a lack of cross-hybridization severely limits the value of cross-species comparisons. Unfortunately, the proportion of genes that evolve rapidly and thus fail to cross-hybridize among grasses is not yet known.

Without substantially more DNA sequence data, it is challenging to draw general conclusions from microsynteny studies. Perhaps the most basic conclusion is that there are 'many exceptions' (Bennetzen, 2000) to microsynteny. It is not clear, however, whether small rearrangements identified by microsynteny studies occur more or less frequently than the larger chromosomal rearrangements identified by comparative mapping. Bennetzen (2000) posits that small rearrangements are an order of magnitude more frequent than large chromosomal events. However, in some respects the relative rates of these events are not particularly important, because both small and large rearrangements affect colinearity, thereby potentially complicating cross-species studies. In addition, both small and large rearrangements contribute to our understanding of grasses as entities of substantial genomic change.

IV. Conclusions

It is clear that grass genomes evolve with frequent loss and gain of chromosomes and DNA content through time. The increasingly robust grass phylogeny provides an evolutionary framework to examine the pattern of loss and gain. In this framework, analyses suggest that genome content varies among grass phylogenetic groups. Changes in DNA content may primarily reflect proliferation and removal of repetitive DNA, but it also seems likely that gene content (or copy number) has changed often among grass genomes, especially given the frequent occurrence of polyploidy throughout the family. By contrast, the phylogeny does not yet provide extensive insights into the evolution of basic chromosome number, and as a result the evolutionary mechanisms contributing to chromosome loss and gain are unclear (Moore *et al.*, 1997).

DNA content and chromosome numbers suggest that grass genomes are dynamic, rapidly evolving entities. Nonetheless, most comparative mapping literature concludes that the major hallmark of grass genome evolution is the retention of extensive colinear regions. It is time to re-evaluate this conclusion, based on several observations. First, microsynteny studies

suggest that small-scale rearrangement can be frequent. Second, synteny probabilities, which are based on genetic maps, are not exceptionally high (Table 1). Third, mapping data are limited, both because of the nature of data and because of methods of interpretation. This is not to imply that there are no syntenic regions among grass genomes, but the more pressing question is whether extensive genome conservation is the hallmark of grass evolution. Unfortunately, the current data are too limited, in my view, to make strong conclusions about genome conservation or potential mechanisms of genome conservation. These issues can be addressed further with improved methods of map interpretation and additional sequence and physical map data.

Despite rapid progress in the last decade, our understanding of grass genomes (and plant genomes as a whole) is rudimentary. However, the spectacular advances of the last decade have spawned an abundance of additional questions about genome evolution. For example, are there any constraints on genome size and content? The vast size of some grass genomes suggests that some evolutionary lineages have relatively few constraints on genome size, but we do not yet know about constraints (or lack thereof) on gene content and copy number.

Is there selection for or against gene order? The lack of colinearity between grasses and arabidopsis has been interpreted as evidence that there has been natural selection against colinearity in these evolutionary distant taxa (Bennetzen, 2000). It is entirely possible, however, that the apparent lack of colinearity between arabidopsis and grasses is the realization of an approximately steady-state process of synteny disruption. For example, one can calculate the expected percentage loss of synteny in rice vs arabidopsis using the rates in Table 1. Assuming a rate of loss of 0.54% synteny per my and a monocot-dicot divergence of 200 my, the expected loss of synteny between arabidopsis and rice is > 100%. This calculation predicts that colinearity between arabidopsis and rice will not exceed what is expected by random chance in the absence of stabilizing forces. By contrast, there have also been suggestions that there is selection for gene order, as first hypothesized by Stebbins (1971). For the time being, it is difficult to assess whether selection is playing a role in maintaining linkage groups. With few exceptions (Rieseberg *et al.*, 1996), we have virtually no knowledge of the evolutionary forces that shape linkage relationships in plant genomes.

Many more questions need to be answered – that is, why are gene families prevalent in grass genomes? What is the evolutionary fate of duplicated genes? What proportion of genes are evolving so rapidly that cross-hybridization between highly diverged grasses is unlikely to be successful? What mechanisms play a role in generating large (chromosomal) and small (microsyntenic) rearrangements, and how do those mechanisms differ? The address of these questions has just begun, but the near future promises to yield fascinating glimpses into grass genome evolution.

Acknowledgements

P. Tiffin, M. Tenaillon, A. Barakat, L. Zhang, J. Wendel and one anonymous reviewer provided valuable comments, and L. Zhang assisted in compiling GWP data. The work was supported by NSF (DBI-9872631 and DEB-9815855) and USDA (98-35301-6153).

References

- Appels R, Scoles G, Chapman CGD. 1986. The nature of change in nuclear DNA in the evolution of the grasses. In: Soderstrom T, Hilu KW, Campbell CS, Barkworth ME, eds. *Grass systematics and evolution*. Washington DC, USA: Smithsonian Institution Press, 73–87.
- Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796–815.
- Avdulov NP. 1931. Karyosystematic studies in the grass family. *The Bulletin of Applied Botany Genetics and Plant Breeding*, S44: 1–325.
- Barakat A, Carels N, Bernardi G. 1997. The distribution of genes in the genomes of Gramineae. *Proceedings of the National Academy of Sciences, USA* 94: 6857–6861.
- Bennett MD, Leitch IJ. 1995. Nuclear DNA amounts in angiosperms. *Annals of Botany* 76: 113–176.
- Bennett MD, Smith JB. 1976. Nuclear DNA amounts in angiosperms. *Philosophical Transactions of the Royal Society of London* 274: 227–274.
- Bennett MD, Smith JB. 1991. Nuclear DNA amounts in angiosperms. *Philosophical Transactions of the Royal Society of London B* 334: 309–345.
- Bennetzen JL. 2000. Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions. *Plant Cell* 12: 1021–1029.
- Bennetzen JL, Freeling M. 1993. Grasses as a single genetic system – genome composition, colinearity and compatibility. *Trends in Genetics* 9: 259–261.
- Bennetzen JL, Kellogg EA. 1997. Do plants have a one-way ticket to genomic obesity? *Plant Cell* 9: 1509–1514.
- Celarier RP. 1956. Additional evidence of five as the basic chromosome number of the Andropogoneae. *Rhodora* 58: 135–143.
- Chao S, Sharp PJ, Gale MD. 1988. A linkage map of wheat homoeologous group of chromosomes using RFLP markers. In: Miller TE, Koebner RMD, eds. *Proceedings of the 7th International Wheat Genet Symposium IPSR*. Cambridge, UK: Institute of Plant Science Research, Cambridge, 000–000.
- Chen MS, SanMiguel P, Bennetzen JL. 1998. Sequence organization and conservation in sh2/a1-homologous regions of sorghum and rice. *Genetics* 148: 435–443.
- Chen M, SanMiguel P, de Oliveira AC, Woo S-S, Zhang H, Wing RA, Bennetzen JL. 1997. Microcolinearity in sh2-homologous regions of the maize, rice, and sorghum genomes. *Proceedings of the National Academy of Sciences, USA* 94: 3431–3435.
- Clark LG, Zhang W, Wendel JF. 1995. A phylogeny of the Grass family (Poaceae) based on *ndhF* sequence data. *Systemic Botany* 20: 436–360.
- Devos KM, Atkinson MD, Chinoy CN, Harcourt RL, Koebner RMD, Liu CJ, Masojc P, Xie DX, Gale MD. 1993a. Chromosomal rearrangements in the rye genome relative to that of wheat. *Theoretical Applied Genetics* 85: 784–792.
- Devos KM, Gale MD. 2000. Genome relationships: The grass model in current research. *Plant Cell* 12: 637–646.
- Devos KM, Millan T, Gale MD. 1993b. The genetic maps of wheat and their potential in plant breeding. *Theoretical Applied Genetics* 85: 673–680.
- Devos K, Wang ZM, Beales J, Sasaki Y, Gale MD. 1998. Comparative genetic maps of foxtail millet (*Setaria italica*) and rice (*Oryza sativa*). *Theoretical Applied Genetics* 96: 63–68.
- DeWet JMJ. 1986. Hybridization and polyploidy in the Poaceae. In: Soderstrom T, Hilu KW, Campbell CS, Barkworth ME, eds. *Grass systematics and evolution*. Washington DC, USA: Smithsonian Institution Press, 188–194.
- Duvall MR, Morton BR. 1996. Molecular phylogenetics of Poaceae: an expanded analysis of *rbcL* sequence data. *Molecular Phylogenetics and Evolution* 5: 352–358.
- Dvorak J, Luo M-C, Yang Z-L. 1998. Restriction fragment length polymorphism and divergence in the genomic regions of high and low recombination in self-fertilizing and cross-fertilizing *Aegilops* species. *Genetics* 148: 423–434.
- Ellis RP. 1986. A review of comparative leaf blade anatomy in the systematics of the Poaceae: The past twenty-five years. In: Soderstrom TR, Hilu KH, Campbell CS, Barkworth ME, eds. *Grass systematics and evolution*. Washington DC, USA: Smithsonian Institution Press, 3–11.
- Endo TR, Gill BS. 1996. The deletion stocks of common wheat. *Journal of Heredity* 87: 295–307.
- Feldman M, Liu B, Segal G, Abbo S, Levy AA, Vega JM. 1997. Rapid elimination of low-copy DNA sequences in polyploid wheat: a possible mechanism for differentiation of homoeologous chromosomes. *Genetics* 147: 1381–1387.
- Flavell RB, Bennett MD, Smith JB, Smith DB. 1974. Genome size and the proportion of repeated nucleotide sequence DNA in plants. *Biochemistry and Genetics* 12: 257–269.
- Flovik K. 1938. Cytological studies of arctic grasses. *Hereditas* 24: 265–376.
- Fu H, Park W, Yan X, Zheng Z, Shen B, Dooner HK. 2001. The highly recombinogenic bz locus lies in an unusually gene-rich region of the maize genome. *Proceedings of the National Academy of Sciences, USA* 98: 8903–8908.
- Gale MD, Devos KM. 1998. Comparative genetics in the grasses. *Proceedings of the National Academy of Sciences, USA* 95: 1971–1974.
- Gaut BS. 2001. Patterns of chromosomal duplication in maize and their implications for comparative maps of the grasses. *Genome Research* 11: 55–66.
- Gaut BS, Doebley JF. 1997. DNA sequence evidence for the segmental allotetraploid origin of maize. *Proceedings of the National Academy of Sciences, USA* 94: 6809–6814.
- Gaut BS, Le Thierry d'Ennequin M, Peek AS, Sawkins MC. 2000. Maize as a Model for the Evolution of Plant Nuclear Genomes. *Proceedings of the National Academy of Sciences, USA* 97: 7008–7015.
- Gill KS, Gill BS, Endo TR, Boyko EV. 1996. Identification and high density mapping of gene-rich regions in chromosome group 5 of wheat. *Genetics* 143: 1001–10012.
- Grass Phylogeny Working Group. 2001. Phylogeny and subfamilial classification of the grasses (Poaceae). *Annals of the Missouri Botanical Garden* 88: 373–457.
- Hart GE. 1983. Hexaploid wheat (*Triticum aestivum* L. em Thell). In: Tanksley SD, Orton TJ, eds. *Isozymes in Plant Genetics and Breeding*. New York, USA: Elsevier, 35–36.
- Helentjaris T, Weber D, Wright S. 1988. Identification of the genomic locations of duplicate nucleotide sequences in maize by analysis of restriction fragment length polymorphism. *Genetics* 118: 353–363.
- Hilton H, Gaut BS. 1998. Speciation and Domestication in maize and its wild relatives: evidence from the *Globulin-1* gene. *Genetics* 150: 863–872.
- Hsiao C, Jacobs SWL, Barker NP, Chatterton JJ. 1998. A molecular phylogeny of the subfamily Arundoideae (Poaceae) based on sequences of rDNA. *Australian Systematic Botany* 11: 41–52.
- Jacobs BF, Kingston JD, Jacobs LL. 1999. The origin of grass-dominated ecosystems. *Annual Review of Ecology and Systematics* 86: 590–643.

- Kellogg EA. 2000. The grasses: a case study of macroevolution. *Annual Review of Ecology and Systematics* 31: 217–238.
- Kellogg EA. 2001. Evolutionary History of the Grasses. *Plant Physiology* 125: 1198–1205.
- Kowalski SP, Lan TH, Feldmann KA, Paterson AH. 1994. Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea* chromosomes reveals islands of conserved organization. *Genetics* 138: 499–510.
- Kraft T, Sall T, Magnusson-Rading I, Nilsson NO, Hallden C. 1998. Positive correlation between recombination rates and levels of genetic variation in natural populations of sea beet (*Beta vulgaris* subsp. *maritima*). *Genetics* 150: 1239–1244.
- Laroche J, Li P, Bousquet J. 1995. Mitochondrial DNA and monocot-dicot divergence time. *Molecular Biological Evolution* 12: 1151–1156.
- Lawrence GL, Appels R. 1986. Mapping the nucleolus organizer region, seed protein loci, and isozyme loci on chromosome 1R in rye. *Applied Genetics* 71: 742–749.
- Ligon PC. 1993. Seeds of change. *Dealer Progress* Nov–Dec: 29–30.
- Linder HP. 1987. The evolutionary history of the Poales/Restionales: a hypothesis. *Kew Bulletin* 42: 297–318.
- Liu BV, Vega JM, Feldman M. 1998b. Rapid genomic changes in newly synthesized amphiploids of *Triticum* and *Aegilops*. II. Changes in low-copy coding DNA sequences. *Genome* 41: 535–542.
- Liu B, Vega JM, Segal G, Abbo S, Rodova H, Feldman M. 1998a. Rapid genomic changes in newly synthesized amphiploids of *Triticum* and *Aegilops*. I. Changes in low-copy noncoding DNA sequences. *Genome* 41: 272–277.
- Mason-Gamer RJ, Weil CF, Kellogg EA. 1998. Granule-bound starch synthase: structure, function and phylogenetic utility. *Molecular Biological Evolution* 15: 1658–1673.
- Mathews S, Tsai RC, Kellogg EA. 2000. Phylogenetic structure in the grass family (Poaceae): Evidence from the nuclear gene phytochrome B. *American Journal of Botany* 87: 96–107.
- McGrath JM, Jansco MM, Pichersky E. 1993. Duplicated sequences with a similarity to expressed genes in the genome of *Arabidopsis thaliana*. *Theoretical Applied Genetics* 86: 880–888.
- Molina MD, Naranjo CA. 1987. Cytogenetic studies in the genus *Zea*: 1. Evidence for five as the basic chromosome number. *Theoretical Applied Genetics* 73: 542–550.
- Moore G, Aragon-Alcaide L, Roberts M, Reader S, Miller T, Foote T. 1997. Are rice chromosomes components of a holocentric chromosome ancestor? *Plant Molecular Biology* 35: 17–23.
- Moore G, Devos KM, Wang Z, Gale MD. 1995. Grasses, line up and form a circle. *Current Biology* 5: 737–739.
- Panstruga R, Buschges R, Piffanelli P, Schulze-Lefert P. 1998. A contiguous 60 kb genomic stretch from barley reveals molecular evidence for gene islands in a monocot genome. *Nucleic Acids Research* 26: 1056–1062.
- Paterson AH, Bowers JE, Burow MD, Draye X, Elsik CG, Jiang C-X, Katsar CS, Lan T-H, Lin Y-R, Ming R, Wright RJ. 2000. Comparative Genomics of Plant Chromosomes. *Plant Cell* 12: 1523–1539.
- Paterson AH, Lin Y-R, Li Z, Schertz KF, Doebley JF, Pinson SRM, Liu S-C, Stansel JW, Irvine JE. 1995. Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. *Science* 269: 1714–1718.
- Payne PI, Holt LM, Hutchinson J, Bennett MD. 1985. Development and characterization of a line of bread wheat, *Triticum aestivum*, which lacks the short-arm satellite of chromosome 1B and the *Gli-B1* locus. *Theoretical Applied Genetics* 68: 327–334.
- Pereira MG, Lee M, Bramel-Cox P, Woodman W, Doebley J, Whitkus R. 1994. Construction of an RFLP map in sorghum and comparative mapping in maize. *Genome* 37: 236–243.
- Petrov DA. 2001. Evolution of genome size: new approaches to an old problem. *Trends in Genetics* 17: 23–28.
- Petrov DA, Lozovskaya ER, Hartl DL. 1996. High intrinsic rate of DNA loss in *Drosophila*. *Nature* 384: 346–349.
- Petrov DA, Sangster TA, Johnston JS, Hartl DL, Shaw KL. 2000. Evidence for DNA loss as a determinant of genome size. *Science* 287: 1060–1062.
- Reeder JR. 1957. The embryo in grass systematics. *American Journal of Botany* 44: 756–769.
- Rieseberg LH, Sinervo B, Linder CR, Ungerer MC, Arias DM. 1996. Role of gene interactions in hybrid speciation: evidence from ancient and experimental hybrids. *Science* 272: 741–745.
- Saghai-Marooof MA, Yang GP, Biyashev RM, Maughan PJ, Zhang Q. 1996. Analysis of the barley and rice genomes by comparative RFLP linkage mapping. *Theoretical Applied Genetics* 92: 541–551.
- Sanderson MJ. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Molecular Biological Evolution* 14: 1218–1231.
- SanMiguel PJ, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL. 1998. The paleontology of intergene retrotransposons of maize: dating the strata. *Nature Genetics* 20: 43–45.
- SanMiguel P, Tikhonov A, Jin Y-K, Melake-Berhan A, Springer PS, Edwards KJ, Avramova Z, Bennetzen JL. 1996. Nested retrotransposons in the intergenic regions of the maize genome. *Science* 274: 765–768.
- Shantz HL. 1954. The place of grasslands in the earth's cover of vegetation. *Ecology* 35: 143–145.
- Sharma ML. 1979. Some considerations on the phylogeny and chromosomal evolution in grasses. *Cytologia* 44: 679–685.
- Shirasu K, Schulman AH, Lahaye T, Schulze-Lefert P. 2000. A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. *Genome Research* 10: 908–915.
- Song K, Lu P, Tang K, Osborn TC. 1995. Rapid genome change in synthetic polyploids of *Brassica* and its implication for polyploid evolution. *Proceedings of the National Academy of Sciences, USA* 92: 7719–7723.
- Spangler R, Zaitchik B, Russo E, Kellogg EA. 1999. Andropogoneae evolution and generic limits in *Sorghum* (Poaceae) using *ndhF* sequences. *Systematic Botany* 24: 267–281.
- Springer PS, Edwards KJ, Bennetzen JL. 1994. DNA class organization on maize *adh1* yeast artificial chromosomes. *Proceedings of the National Academy of Sciences, USA* 91: 863–867.
- Stebbins GL. 1971. *Chromosomal evolution in higher plants*. London, UK: Arnold.
- Stebbins GL. 1981. *Annals of the Missouri Botanical Garden* 68: 75–86.
- Stebbins GL. 1985. Polyploidy, hybridization and the invasion of new habitats. *Annals of the Missouri Botanical Garden* 72: 824–832.
- Tarchini R, Biddle R, Wineland R, Tingey S, Rafalski A. 2000. The complete sequence of 340 kb of DNA around the rice *adh1-adh2* region reveals interrupted colinearity with maize chromosome 4. *Plant Cell* 12: 381–391.
- Tikhonov AP, SanMiguel PJ, Nakajima Y, Gorenstein NM, Bennetzen JL, Avramova Z. 1999. Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum. *Proceedings of the National Academy of Sciences, USA* 96: 7409–7414.
- Van Deynze AE, Nelson JC, O'Donoghue LS, Ahn SN, Siripoonwivat W, Harrington SE, Yglesias ES, Braga DP, McCouch SR, Sorrells SR. 1995a. Comparative mapping in grasses. Oat relationships. *Theoretical Applied Genetics* 249: 349–356.
- Van Deynze AE, Nelson JC, Yglesias ES, Harrington SE, Braga DP, McCouch SR, Sorrells SR. 1995b. Comparative mapping grasses. Wheat relationships. *Molecular and General Genetics* 248: 744–754.
- Vision TJ, Brown DG, Tanksley SD. 2001. The origins of genomic duplication in the *Arabidopsis* genome. *Science* 290: 2114–2117.
- Wendel JF. 2000. Genome evolution in polyploids. *Plant Molecular Biology* 42: 225–249.
- White SE, Doebley JF. 1999. The molecular evolution of *terminal ear 1*, a regulatory gene in the genus *Zea*. *Genetics* 153: 1455–1462.

Whitkus R, Doebley J, Lee M. 1992. Comparative genome mapping of sorghum and maize. *Genetics* 132: 1119–1130.

Wilson WA, Harrington SE, Woodman WL, Lee M, Sorrells ME, McCouch SR. 1999. Inferences on the genome structure of progenitor maize through comparative analysis of rice, maize and the domesticated panicoids. *Genetics* 153: 453–473.

Wolfe KH, Gouy M, Yang Y-W, Sharp PM, Li W-H. 1989. Date of

the monocot-dicot divergence estimated from chloroplast DNA sequence data. *Proceedings of the National Academy of Sciences, USA* 86: 6201–6205.

Wolfe KH, Li W-H, Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast and nuclear DNAs. *Proceedings of the National Academy of Sciences, USA* 84: 9054–9058.



About *New Phytologist*

- *New Phytologist* is owned by a non-profit-making **charitable trust** dedicated to the promotion of plant science. Regular papers, Letters, Research reviews, Rapid reports and Methods papers are encouraged. Complete information is available at www.newphytologist.com
- All the following are **free** – essential colour costs, 100 offprints for each article, online summaries and ToC alerts (go to the website and click on Synergy)
- You can take out a **personal subscription** to the journal for a fraction of the institutional price. Rates start at £83 in Europe/\$133 in the USA & Canada for the online edition (go to the website and click on Subscriptions)
- If you have any questions, do get in touch with Central Office (newphytol@lancaster.ac.uk; tel +44 1524 594691) or, for a local contact in North America, the USA Office (newphytol@ornl.gov; tel 865 576 5251)