Grasses as a Single Genetic System. Reassessment 2001

Michael Freeling*
Department of Plant and Microbial Biology, University of California, Berkeley, California 94720

Grasses are a large, diverse, and successful family of monocotyledonous flowering plants characterized most dramatically by their spikelet style of floral structure (Fig. 1a) and by the number and pattern of these spikelets in the inflorescence. There are about 10,000 species that are categorized as grasses by taxonomists who traditionally follow embryonic, architectural, and anatomical characters (e.g. Clayton and Renvoize, 1986). More recent comparative DNA sequence information confirms that all grasses examined to date did indeed diverge from one common ancestral population of "grass alleles" and are distinct from the nearest non-grass family, the Joinvilleae. In general, the taxonomic distinction within the grasses continuously divides these species into subfamilies, supertribes, tribes, subtribes, genera, species, and subspecies. As additional species have been added to the sequence databases, a phylogenetic trend has emerged. Where Clayton and Renvoize (1986) recognized six subfamilies, Kellogg (1998 and refs. therein) recognized 13, and the "in progress" data of the Grass Phylogenetics Working Group (http://www.ftg.fiu.edu/grass/gpwg) could be used to justify recognition of at least 16 subfamilies, often composed of one or a few species. At present the phylogenetic relationship—the exact branch relationships—among most of these 16 subfamilies is not known, as if they originated in a single grass adaptive radiation, estimated to be about 70 million years ago (Kellogg, 1998 and refs. therein). In fact, there is likely one specific order by which the several grass subfamilies are related, one to another, but sequence-based trees are necessarily quantitative and often cannot resolve deep branches. Perhaps phylogenetic trees based on chromosomal breakpoints, not being time-dependent, will fare better.

Until about 8 years ago, one grass species maps and mutant collections, however interesting, did not directly affect research on another grass species, even though grass genomes were known to be related. This state of being isolated by commodity changed in 1992–1993. Of particular notice was the research of Dr. Steven Tanksley and coworkers on gene order comparisons of maize and rice (Ahn and Tanksley, 1993) and Dr. Mike Gale and coworkers and their many international collaborators who moved from mapping wheat relatives to more distant grasses (summarized by Moore et al., 1995). The graphic summary of these mapping data, greatly simplified, has become popularly known as "The Circle Diagram" because of a method used to draw the expressed gene sequence (EST) maps of several different grass species on one radial axis. A recent Circle Diagram (Gale and Devos, 1998) includes crop grass species from four subfamilies: Pooids (wheat and oat), Panicoide (maize, sorghum, sugarcane, and foxtail millet), Oryzoids (rice), and the Chlorinoids (finger millet). In general, gene probes (ESTs) were chosen to span entire genomes at intervals of 10 or 20 map units. Comparative mapping in the grasses has been reviewed recently (Devos and Gale, 2000). The general conclusion is that all of the grasses in the four subfamilies examined have their genes in about the same order so that one can conclude with confidence that one ancestral genome remains recognizable in its descendents. The huge differences in DNA content/haploid genome and the differences in chromosome number seem to have little or nothing to do with gene number or order. Recent polyploids, like bread wheat, for example, certainly have multiples of the ancestral gene number, but even this simple expectation about polyploids proves to be false for descendents of ancient duplication events, as will be discussed.

DEFINITIONS OF MACROCOLINEARITY AND SYNTENY

Closely related species within a genus have the same chromosome numbers unless they are easily recognizable polyploids. It is expected and proved that the individual genomes in a polyploid series carry the same number of genes per genome, and that these genes are in the same order (as in the wheats; cited in Devos and Gale, 2000), allowing for the occasional chromosomal aberration (inversion, duplication, translocation, or deletion of previously duplicated chromosome) as well as the occasional "mutant." The result is called "macrocolinearity" because a typical mapping interval is 10 to 20 map units. In a particular rice heterozygotic region including the adh1 gene, 10 map units is 1.15 Mb and carries 147 potential genes (Tarchini et al., 2000). When gene orders, quantified in coarse mapping intervals, are compared species with species, estimates of "macrocolinearity" result. If five markers in a row covering 50 map units also occur in a row on another chromosome, these chromosome regions are macrocolinear. Often a breakpoint event disconnects this string of five markers, and colinearity is broken. Such chro-

* E-mail freeling@nature.berkeley.edu; fax 510–642–4995.
mosomal aberrations are expected and are often readily explicable, so another more generous term, "synteny," is often used to refer to largely homologous chromosomal segments. Perfect synteny occurs as long as the ancestral gene order can be reconstructed, and accepts chromosomal breakpoints, polyploidy, and partial duplications. So, although closely related species in the same genus do not always display perfect macrocolinearity, they are syntenous. Comparing more diverged species before deciding whether macrocolinearity predicts microcolinearity. The first comparison between such regions in maize and sorghum discovered that maize carried massive amounts of retrotransposons between its genes, whereas sorghum did not, and that this difference roughly accounted for the 5-fold DNA content/genome difference between maize and sorghum (SanMiguel et al., 1996). Bennetzen (2000) measured microcolinearity in the adh1 region of maize and sorghum, species that are thought to have diverged about 20 million years ago. A 78-kb genomic sequence of sorghum carrying adh1 was compared with a potentially homologous adh1-containing maize sequence roughly five times larger because of retrotransposon filler. Considerable microcolinearity between maize and sorghum was found. A conservative estimate might be that eight obvious genes are present in perfect colinear order and transcriptional direction, and there are two genes in sorghum that may have no orthologs in the maize sequence. Tarchini and coworkers (2000) attempted to compare the adh1-adh2 region of rice with the Panoid (maize/sorghum) sequences, but encountered an apparent translocation just at Adh1; they did not find and sequence the missing segment. In the comparable region, considerable microcolinearity was found between rice and sorghum. Bennetzen (2000) reviews all of the other cases of intragrass sequence comparisons as well, and concludes that microcolinearity seems to be the rule so far, but there are many apparent exceptions. Bennetzen discusses mechanisms and selective pressures that could help explain these exceptions, including the interesting idea that there may be selection to break colinearity.

**MICROCOLINEARITY**

Ten map units of a rice chromosome in the adh1 region carry approximately 147 genes (Tarchini et al., 2000). Consider the question: is it certain that if a large chromosomal region displays macrocolinearity, then the genes between these markers should also be approximately colinear? The answer depends on the nature and frequency of small chromosomal aberrations, those involving one to a few genes, as compared with the traditional double-strand-break-type aberrations, and whether these small and large events share the same mechanism. If individual genes routinely excise and relocate at random in the genome, then synteny would be lost. Bennetzen (2000), in his recent review of microcolinearity in flowering plants, argues that it is best to compare all genes in macrocolinear regions of a number of species before deciding whether macrocolinearity predicts microcolinearity. The first comparison between such regions in maize and sorghum discovered that maize carried massive amounts of retrotransposons between its genes, whereas sorghum did not, and that this difference roughly accounted for the 5-fold DNA content/genome difference between maize and sorghum (SanMiguel et al., 1996). Bennetzen (2000) measured microcolinearity in the adh1 region of maize and sorghum, species that are thought to have diverged about 20 million years ago. A 78-kb genomic sequence of sorghum carrying adh1 was compared with a potentially homologous adh1-containing maize sequence roughly five times larger because of retrotransposon filler. Considerable microcolinearity between maize and sorghum was found. A conservative estimate might be that eight obvious genes are present in perfect colinear order and transcriptional direction, and there are two genes in sorghum that may have no orthologs in the maize sequence. Tarchini and coworkers (2000) attempted to compare the adh1-adh2 region of rice with the Panoid (maize/sorghum) sequences, but encountered an apparent translocation just at Adh1; they did not find and sequence the missing segment. In the comparable region, considerable microcolinearity was found between rice and sorghum.

Bennetzen (2000) reviews all of the other cases of intragrass sequence comparisons as well, and concludes that microcolinearity seems to be the rule so far, but there are many apparent exceptions. Bennetzen discusses mechanisms and selective pressures that could help explain these exceptions, including the interesting idea that there may be selection to break colinearity.
THE IMPORTANCE OF SYNTENY AMONG THE GRASSES

If all grass genomes were syntenous, then mapping any character to any grass gene in the progeny of any grass hybrid—in any of the 10,000 species—would lead to a map position via the rice sequence, and a deduced array of candidate genes. For example, this mapping might be between two newly discovered species of grass in a genus about which nothing is known. So, if any character or QTL (quantitative trait locus) can be mapped to one map unit, then there would be approximately 15 candidate genes in the region (using the rice estimate as already discussed). In short, any phene that can be mapped with precision can be reduced to nucleotide sequence. Synteny must hold if this pan-grass mapping potential is to be realized. Early data reviewed by Bennetzen (2000) are reassuring.

THE CASE OF THE GIBBERELLIC ACID-INSENSITIVE (GAI) DWARVES IN PLANTS: TOWARD MAPPING IN SILICO

GAI dwarves occur in several grasses and in Arabidopsis. The first sequence was from GAI (Arabidopsis), which was mapped onto maize, the three wheat genomes, and rice. Harberd and coworkers showed that a GAI ortholog existed in all of these grasses, and four of the positions were coincident with the dwarves that fuelled the Green Revolution (Peng et al., 1999). The homologous gibberellin-insensitive dwarf in barley was added subsequently to the list of GAI-dwarves (Ivandic et al., 1999). All GAs exist in syntenous positions in grasses. The data on which these studies were based involved wet-laboratory hybridization mapping. If the rice sequence had been available, GAs best homolog in rice would have anchored this gene to a syntenic region even though a dwarf is not known to map to this gene in rice. By mapping in rice, all of the syntenic (orthologous) map positions in maize, wheat, barley, and all other grasses would have become immediate candidates for dwarves. All of these dwarves would have been reduced to gene sequence simultaneously, each controlling the other to generate unequivocal data, and all would have been accomplished by working with information only (in silico).

It seems likely in the case of the GAI dwarves in grasses that the knockout phenotype in maize when it is learned, will also extrapolate to function throughout the grass family. (The Pioneer Hi-Bred International Company [Des Moines, IA] operates a maize gene knockout service that will search their proprietary F1 DNA samples for transposon Mu insertions between the PCR primer the collaborator sends them and a Mu transposon, and send out samples of F2 seed so that one might discover a phenotype caused by the mutant lesion [called TUSC; Bensen et al., 1995]. An NSF-funded reverse genetics service, the Maize Targeted Mutagenesis [MTM] service, http://mtm.cshl.org, began operation early in 2000.) Most genes are expected to function similarly in all grasses.

ON USING “FAMILY” RATHER THAN “SPECIES” AS A MODEL GENETIC SYSTEM

By considering individuals in many species related to one another in a known phylogeny, there is a logical possibility to deduce important facts about the origin of living designs where design represents evolution over millions of years by a process of natural selection of new mutants and new combinations of alleles, as well as manifold chance events, and perhaps the alleles or combinations of alleles that encode them. Figure 1 depicts grass flowers, and by no means covers the full breadth of species diversity. A plethora of flower branching designs, spikelet numbers, and spikelet arrangements are shown. Even so, each grass flower, and the spikelet itself, is made of the same modular segments—with the same organs and organ components—recognizable even if not elaborated. There is also much diversity for other characters, some of which are leaf anatomy, photosynthetic types, epidermal cell-type patterns, venation patterns, ligule shapes, apomictic style, weediness, and salt/drought/pest/hypoxia tolerance. Each of these variations was presumably evolved from a common grass ancestor (see Bennetzen and Freeling, 1998).

Figure 2 (Kellogg, 1998) shows a very minimized (four subfamilies) phylogenetic tree of the grasses, with a single character mapped upon this tree. This

![Figure 2](https://example.com/figure2.png)

**Figure 2.** A minimized phylogenetic tree of grass species on which is plotted DNA content per diploid (2C) prereplication nucleus, reproduced by permission from Kellogg (1998). C-value data and phylogenetic data are from others, cited in the original. Ancestral genome sizes were reconstructed according to squared-change parsimony.
character, called “C-value,” is the amount of DNA in the pre-replication, diploid, mitotic nucleus of the species identified. The subfamilies to which these species belong are marked on the right. Some common ancestors are identified by arrows in front of deduced 2C values; these notations constitute a possible model for the evolution of C-value in these grasses. The general conclusion from this study is that C-value can increase or decrease over evolutionary time. Increases in C-value are seen in the Pooids and the Zea/Tripsicum branches of the Panicoids. Decreases are particularly clear in the genus Corynephorus. Obvious mechanical hypotheses, probably involving retrotransposons (SanMiguel et al., 1996), can be formulated and tested on the basis of the phylogenetic data of Figure 2.

CONVERGENCE IN A PHYLOGENETIC TREE MARKS AN EXPERIMENTAL OPPORTUNITY

Because DNA content is a continuous character, no branch in the tree in Figure 2 demonstrates an origin of novelty. The informative lineages of Figure 2 are those demonstrating convergence. For example, that rice and a Corynephorous have low DNA/genome content is not because of lineage (which would make the characters homologous), but convergence. The mechanism of lowering genome size rapidly within an otherwise high DNA content lineage is of obvious interest specifically because it is able to adapt or change. Even more useful experimentally are cases of convergence that happen between more closely related species, implying rapid evolution. The most useful cases of convergence occur between subspecies, or species that can still cross pollinate to make fertile hybrids, or in the rapid process of domestication. Such convergences must have occurred quickly, and involve one or a few alleles, and the wide hybrid would permit mapping of these alleles as QTLs. Every character that is polymorphic in the grasses—apomixis, Kranz anatomy, and the others listed previously—display convergences. Geographies where unfilled niches appeared quickly, as with islands, or where the environment is marginal to grass life, as with sand dunes or thermal pools, are especially good places to search for convergences. Each convergence constitutes a rapid evolutionary adaptation involving one or a few alleles, and each is a case study in the “genetic engineering” of useful phenotype. Often a particular character or adaptation such as short generation time, perennial habit, C4 photosynthesis, and aposporic apomixis occurs multiple times in a lineage. To what extent such “repeating” characters are convergent (truly of genetically independent origin) or divergent (at least partially sharing the genetic capacity to evolve a character) can make matters academically confusing. Speaking practically, if the character can be mapped, there is a good chance the character can be reduced to DNA sequences.

Using “grasses as a single genetic system” logic, Paterson and coworkers (1995) were able to explain the convergence of domestication traits, those selected by indigenous breeders, in the cereal grasses by showing that mutant alleles of a small set of grass genes were involved.

THE GRASS HYBRIDS DATABASE:
HTTP://128.32.88.35/GRAWWEB/

Why does hybridization sometimes occur between different species or between species in different genera? The domesticated races of wild species are often classed as different species by taxonomists, although domesticated and wild species can effectively cross. The explanation is obvious: The taxonomists were fooled by convergences caused by human selection. Some genera such as the Festuca-Lolium complex of species are famous for wide crosses that yield fertile hybrids; other subtribes exhibit no such propensity. There are approximately 4,000 hybrids, mostly infertile, identified and referenced in the grass hybrids database (http://128.32.88.35/grassweb), and there are certainly many more hybrids known, but not yet reported.

The point of the grass hybrids database is to facilitate discovery of individual fertile hybrids that are heterozygous for alleles specifying profound character differences. Such profound character differences result from convergence or perhaps, hopefully, the evolution of a truly novel phene. Given this hybrid, QTL mapping of alleles to pan-grass markers should automatically generate candidate genes from the fully sequenced anchor genome (this genome is rice for the grass family). The best support for this approach (Lan and Paterson, 2000) is the recent success at mapping QTLs important for “sculpting the curd,” the unique flower arrangement of broccoli, for example, in the Brassica family of dicots, using Arabidopsis as the anchor genome from which candidate genes are drawn.

REALY WIDE CROSSES BETWEEN GRASSES: CHROMOSOME ADDITION LINES

There have been attempts to cross grass species in different subfamilies, species diverged to about as far as one can get within the grasses. Crosses between maize and wheat, and many others, always with the most polyploid as female, have been tried and the resulting embryos, when such occur, have been rescued by tissue culture. The results are maternally haploid progeny and also include cell lines that have additional chromosomes that are eliminated during mitoses and plant regeneration; hybrids created by protoplast fusion fared similarly (Laurie et al., 1990). There is one glaring, very wide cross success that was between grasses that are maximally diverged: pollen from maize (Panicoid; ancient tetraploid) onto eggs...
of oats (Pooid; recent hexaploid). A recent publication (Riera-Lizarazu et al., 2000) refers to success at adding each chromosome from maize into oat and having this chromosome predictably transmit to the next generation. This paper highlights one of these monosomic lines containing maize chromosome 9 as a source for construction of radiation hybrid mapping populations. Oat seeds with one chromosome 9 were irradiated. The often deleted or minimized chromosome 9 was recovered in oats, identified because these oats contained maize specific mid-repetitive sequences, and then mapped using 39 maize markers spaced along the 151 map units, and 191 Mb, of chromosome 9. Results indicate that 100 informative radiation hybrid lines should permit mapping to 0.5 to 1.0 Mb, which should be very approximately 10 to 20 genes as extrapolated from data (Bennetzen, 2000) on the maize adh1 region. Even higher resolution is possible with more lines. Using maize-specific probes to examine addition line karyotypes by fluorescent in situ hybridization shows exactly where these fragments of maize DNA end up in oat chromosomes, providing an unparalleled tool for examining aspects of chromatin-level gene regulation. As referenced by Riera-Lizarazu et al. (2000, citing Muehlbauer et al., 2000), there is at least one case where a gene on the added maize chromosome expresses and alters phenotype.

By applying “grass as a single genetic system,” it seems possible that mapping in outlying grasses might be to PCR-generated gene fragments, and then these gene fragments might be mapped in maize using radiation hybrids. Gene mapping in any grass should identify the candidate genes in the rice sequence.

**THE CONSEQUENCES OF ROUNDS AND ROUNDS OF DUPLICATIONS AND BREAKPOINTS: THE IMPORTANCE OF CONSOLIDATION OF CHROMOSOMES BEFORE COLINEARITY COMPARISONS**

Internal sequence comparisons of yeast have discovered much duplication and a tetraploid ancestry is suspected (Wolfe and Shields, 1997). Arabidopsis is largely duplicated as well and a comparison of tomato and Arabidopsis, two widely diverged dicots, leads to the general conclusion that Arabidopsis has undergone an ancient tetraploidization event and subsequent large-scale duplication events as well (Blanc et al., 2000; Ku et al., 2000). Duplications of the genome, in whole and in part, may be the rule rather than the exception.

Maize is thought to be descended from a segmental allotetraploidy event (Gaut and Doebley, 1997) that happened approximately 20 million years ago (for discussion, see Bennetzen, 2000). The case of allotetraploid maize illustrates an important point I call “consolidation.” There are numerous cases of duplications that map to the suspected homeologs, but there also numerous cases of sequences that seem to hybridize as single genes and there are certainly many, many recessive mutant alleles identified in maize. Thus, it is difficult to know what to expect of a once-duplicated genome after 20 million years of evolution. One study (Sentoku et al., 1999 and refs. therein) generated data on which some useful predictions can be made. After exhaustive hybridization searches, there turns out to be seven class I homeobox genes (knock genes) in rice, a presumed true diploid, and nine in maize, not 14 as would be initially expected on the basis of polyploidy. Only one
of the extra genes in maize maps to the syntenic region and is diverged to the extent expected of the genomic duplicate; the other extra knox gene is an obviously recent, linked duplication. This computes to 14% (1/7) retained tetraploidy, 14% recent duplication, 86% winnowing, where one of the once-duplicated genes has been deleted, and 100% synteny, where every rice gene maps to a macrocolinear position in maize. Figure 3 is a totally hypothetical model where an ancestral stretch of chromosome marked in blocks 1 through 10 is duplicated and in the process, partially duplicated again and inserted within itself in inverted order. One possible evolved consequence of this duplication-triplication inversion event is shown, using the large winnowing frequency found by Sentoku and coworkers. The point of this exercise is to emphasize that duplications establish macrocolinear arrangements that are expected to alter over time. (Ku and coworkers [2000], working with dicots, used the term “network of synteny” when describing the consequences of evolution from duplicated genomes; their discussions are conceptually similar to my own.) It is premature to call deviations from colinearity within single chromosomes “exceptions” until all of the duplicated regions are sequenced and consolidated into one virtual chromosome. The inset of Figure 3 shows a comparison of the original ancestral sequence with the consolidated sequence; the result is 100% synteny and some added complexity.

THE MECHANICAL BASIS FOR DARWINIAN EVOLUTION

Using “grasses as a single genetic system,” there is a logical avenue to discover the sorts of genes that when mutant, generate macromutations important for speciation and adaptation. The mechanical basis for Darwinian evolution is, of course, grain-like in importance (Goldschmidt, 1952) and cannot be addressed in any single species genetic system.

CONCLUSIONS

Most data support and none refute the general conclusion that grasses share “alleles” of the ancestral grass genome rearranged in blocks that remain long enough to reconstruct, by the process of consolidation, useful syntenic relationships. The vast paradigm called “grasses as a single genetic system,” (Bennetzen and Freeling, 1993) has now been redefined in these few paragraphs. This paradigm provides a way to extract from natural variation those gene sequences, alleles, that caused profound evolutionary changes, and not to be sidetracked by all of those alleles “along for the ride.” In studying these sequences, the mechanisms by which they are expressed, and the biological context within which they acquire meaning, it is possible that we may yet be able to reduce profound diversity—the sort of diversity represented by the morphological differences in Figure 1—to principles of design.

There are many components necessary to proceed within “grasses as a single genetic system.” Progress could be impeded by inadequate talent, funding, or political will. Thinking by commodity is certainly not helpful. Excessive self-interest among institutions, for-profits, and governments, or barriers to export of germplasm could also slow progress. There is one component that is fundamental and irreversibly limiting: the availability of wild germplasm. It would be ironic to say the least if our most specialized and vulnerable wild species become extinct before we are able to understand and value their alleles.

ACKNOWLEDGMENTS

I would like to thank the members of the Freeling Laboratory at the University of California, Berkeley for their useful discussions.

Received November 7, 2000; accepted December 6, 2000.

LITERATURE CITED


Goldschmidt RB (1952) Evolution, as viewed by one geneticist. Am Sci 40: 84–98


