The Plant Cell, Vol. 12, 1523–1539, September 2000, www.plantcell.org © 2000 American Society of Plant Physiologists

GENOMICS ARTICLE

Comparative Genomics of Plant Chromosomes

Andrew H. Paterson,[a,b,8] John E. Bowers,[a] Mark D. Burow,[a,b] Xavier Draye,[b,1] Christine G. Elsik,[c] Chun-Xiao Jiang,[b,2] Catherine Susan Katsar,[b,3] Tien-Hung Lan,[b,4] Yann-Rong Lin,[b,5] Reiguang Ming,[b,6] and Robert J. Wright[b,7]

[a] Applied Genetic Technology Center, Departments of Crop and Soil Science, Botany, and Genetics, University of Georgia, Athens, Georgia 30602

[b] Department of Soil and Crop Science, Texas A & M University, College Station, Texas 77843

[c] Department of Forestry, Texas A & M University, College Station, Texas 77843

FOUNDATIONS OF COMPARATIVE GENOMICS

Comparative genomics, the study of the similarities and differences in structure and function of hereditary information across taxa, uses molecular tools to investigate many notions that long preceded identification of DNA as the hereditary molecule. Vavilov's (1922) law of homologous series in variation was an early suggestion of the similarities in the genetic blueprints of many (plant) species. Genetic analysis based on morphological and isoenzyme markers hinted at parallel arrangements of genes along the chromosomes of various taxa. These hints were later borne out at the DNA level, in seminal investigations of nightshades (Tanksley et al., 1988; Bonierbale et al., 1988) and grasses (Hulbert et al., 1990).

Over the past two decades, multiple investigations of many additional taxa have delivered two broad messages: (1) In most plants, the evolution of the small but essential portion of the genome that actually encodes the organism's genes has proceeded relatively slowly; as a result, taxa that have been reproductively isolated for millions of years have retained recognizable intragenic DNA sequences as well as similar arrangements of genes along the chromosomes. (2) A wide range of factors, such as ancient chromosomal or segmental duplications, mobility of DNA sequences, gene deletion, and localized rearrangements, has been superimposed on the relatively slow tempo of chromosomal evolution and causes many deviations from colinearity.

[1] Current address: Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

[2] Current address: Garst Seeds, Slater, IA 50244.

[3] Current address: USDA-ARS, Manhattan, KS 66502.

[4] Current address: Novartis Agricultural Discovery Institute, La Jolla, CA 92037.

[5] Current address: Department of Biology, Fu-Jen Catholic University, Taipei, Taiwan, R.O.C.

[6] Current address: Hawaiian Agricultural Research Center, Aiea, HI 96701.

[7] Current address: Aventis Seeds, Gent, Belgium.

[8] To whom correspondence should be addressed. E-mail paterson @uga.edu; fax 706-583-0160.

COMPARATIVE MAPS IN CROP TAXA

The Brassicaceae

The Brassicaceae comprise 360 genera, organized into 13 tribes (Shultz, 1936; Al-Shehbaz, 1973). The most economically important species are in the genus Brassica (tribe Brassiceae), six species of which are cultivated worldwide. Three of these species are considered diploid (*B. rapa*, A genome, $2n = 2x = 20$; *B. nigra*, B genome, $2n = 2x = 16$; *B. oleracea*, C genome, $2n = 2x = 18$), and three are amphiploid derivatives of the diploids (*B. juncea*, AB, $2n = 4x = 36$; *B. napus*, AC, $2n = 4x = 38$; *B. carinata*, BC, $2n = 4x = 34$). At least 15 molecular genetic maps have been constructed for Brassica that include all of the major cultivated species (Slocum et al., 1990; Landry et al., 1991, 1992; Song et al., 1991; Chyi et al., 1992; Kianian and Quiros, 1992; Figdore et al., 1993; Ferreira et al., 1994; Teutonico and Osborn, 1994; Uzunova et al., 1995; Lagercrantz and Lydiate, 1996; Truco et al., 1996; Lan et al., 2000a, 2000b; T.-H. Lan and A.H. Paterson, manuscript submitted). These maps include at least 935 different publicly available probes, with many shared by multiple maps. Within a species, the maps show almost complete colinearity (see Lydiate et al., 1993), except for small inversions that differ among some *B. oleracea* morphotypes (Kianian and Quiros, 1992; Lan et al., 2000a). Comparison of *B. rapa* with *B. oleracea* and *B. napus* supports the close evolutionary relationship between the two diploids but indicates that deletions and insertions may have occurred after divergence of the two species (Hoenecke and Chyi, 1991). The genome of the synthetic *B. napus* is

essentially unrearranged with respect to its *B. oleracea* and *B. rapa* progenitors (Lydiate et al., 1993), although the evolution of wild *B. napus* (Cheung and Landry, 1996) has been accompanied by more complicated rearrangements. Extensive segmental duplications are found in both diploid and amphidiploid maps, supporting the hypothesis that diploid *Brassica* spp are derived from an ancestor with a lower original basic chromosome number; however, no duplications of whole linkage groups have been found. At least 22 chromosomal rearrangements differentiate *B. oleracea* homeologs from one another (Lan et al., 2000a) (see Figure 1).

The Poaceae

Most of the world's food and feed crops are members of the family Poaceae, including rice, wheat, maize, sorghum, sugarcane, barley, oat, rye, millet, and others. Conservation across large chromosomal tracts, spanning ~65 million years of evolution (Figure 2), is a recurring message from many investigations of diverse Poaceae (see Bennetzen and Freeling, 1993). Pioneering efforts in the comparative mapping of maize and sorghum (Hulbert et al., 1990) have been supported by more detailed studies (Whitkus et al., 1992; Berhan et al., 1993; Binelli et al., 1993; Chittenden et al., 1994; Pereira et al., 1994) and supplemented by the comparative organization of maize and rice (Ahn and Tanksley, 1993), wheat and rice (Kurata et al., 1994), and maize, wheat, and rice (Ahn et al., 1993). A host of investigations additionally encompasses many other cultivated Poaceae, with particular emphasis on the interrelationships among the homeologous chromosome sets of the Triticeae and their relatives (see Naranjo et al., 1987; Chao et al., 1989; Liu and Tsunewaki, 1991; Devos et al., 1992a, 1992b, 1993, 1995; Liu et al., 1992; Xie et al., 1993; Namuth et al., 1994; Hohmann et al., 1995; Marino et al., 1996; Mickelson-Young et al., 1995; Nelson et al., 1995a, 1995b, 1995c; Van Deynze et al., 1995). Curiously, even in the relatively "conservative" Poaceae, certain lineages appear to be rapidly evolving. The genomes of rye and wheat appear to differ by ~13 chromosomal rearrangements after only ~6 million years of divergence (Devos et al., 1992a, 1992b), a rate of reshuffling more than twice the average calculated for nine taxa (Paterson et al., 1996) and exceeded only by the Brassica–Arabidopsis lineage.

The Fabaceae

The Fabaceae are the third largest family of flowering plants, and like the Poaceae, they contribute substantially to the sustenance of humans. The Fabaceae include major legumes and oilseeds such as soybean (*Glycine max*, $2n = 4x = 40$), peanut (*Arachis hypogaea*, $2n = 4x = 40$), mung bean (*Vigna radiata*, $2n = 2x = 22$), chickpea (*Cicer arietinum*, $2n = 2x = 16$), and lentil (*Lens culinaris*, $2n = 2x = 14$), as well as vegetable crops such as common bean (*Phaseolus vulgaris*, $2n = 2x = 22$) and pea (*Pisum sativum*, $2n = 2x = 14$) and forages such as alfalfa (*Medicago sativa*, $2n = 4x = 32$). The Fabaceae are distinguished from other major crops in that ~90% of legume species interact symbiotically with Rhizobium to fix nitrogen and thus supply themselves with an otherwise costly fertilizer element. Detailed genetic maps have been assembled in at least eight genera of the Fabaceae (Arachis, Glycine, Lens, Medicago, Phaseolus, Pisum, Vicia, and Vigna). A pioneering report established common gene order across at least 40% of the lentil and pea genomes (Weeden et al., 1992) and further suggested, based on isoenzyme loci, conservation with many chromosomal regions in Vicia (Torres et al., 1993). Not surprisingly, mung bean and cowpea, both species of the genus Vicia, also exhibit a high degree of linkage conservation (Menacio-Hautea et al., 1993). Arachis adds a unique dimension to legume genomics in that genetic maps of both diploid (Halward et al., 1993) and tetraploid (M.D. Burow, J. Starr, C. Simpson, and A.H. Paterson, unpublished data) genomes are available. Further, the tetraploid map involves the use of a synthetic polyploid derived from three species that are thought to be quite distant from the cultigen, thereby introducing novel genetic variation that has proven successful for the first trait targeted (see Burow et al., 1996). Recent work on the small-genome taxon *Medicago trunculata* promises to provide a facile model and nodal point for comparative genomics of legumes (Cook, 1999).

The Solanaceae

The Solanaceae include several economically important plant species, such as tomato (*Solanum lycopersicum* [formerly *Lycopersicon esculentum*], $2n = 2x = 24$), pepper (*Capsicum annuum*, $2n = 2x = 24$), potato (*S. tuberosum*, $2n = 2x = 24$), eggplant (*S. melongena*, $2n = 2x = 24$), and tobacco (*Nicotiana tabacum*, $2n = 4x = 48$). The genomes of potato and tomato differ by only five chromosomal rearrangements (Bonierbale et al., 1988; Tanksley et al., 1992), all of which involve a single break at or near a centromere, resulting in paracentric inversions of the short arms of chromosomes 5, 9, 11, and 12 and of the long arm of chromosome 10. These findings further reinforce the relatively high propensity (or tolerance) of plants for intrachromosomal rearrangement. In contrast, the genomes of tomato and pepper are more extensively rearranged, with ~30 chromosome breaks since divergence from a common ancestor (Prince et al., 1993; Livingstone et al., 1999).

POLYPLOIDY: COMPARATIVE GENOMICS WITHIN A NUCLEUS

The joining of two divergent genomes into a common nucleus, thereby establishing polyploidy, is arguably the single

most important genetic mechanism in plant evolution and one that has affected the genomes of most angiosperms (Stebbins, 1966; Masterson, 1994). Polyploid formation is associated with extensive chromosome restructuring (Ahn and Tanksley, 1993; Reinisch et al., 1994; Song et al., 1995a, 1995b; Lagercrantz and Lydiate, 1996; Brubaker et al., 1999) and possibly with retrotransposon amplification (Matzke and Matzke, 1998; Zhao et al., 1998). Polyploidization may also allow further mechanisms of genome restructuring that result in additional gene duplications. For example, among 10 groups of maize *Myb* genes that are likely to have undergone duplication at the time of maize tetraploidization, five groups contain three or more members, indicating recent amplification above the number resulting directly from genome duplication (Rabinowicz et al., 1999).

Geneticists have long debated whether the prevalence of polyploidy in plants simply reflects promiscuity or confers an inherent selective advantage. Polyploidy appears to contribute substantially to the productivity of many crop plants (e.g., see Simmonds, 1976), including cotton, wheat, oat, soybean, peanut, canola, tobacco, coffee, and banana. Remarkably, little is known about how divergent genomes coevolve within a common nucleus so as to function in concert. Both cotton, an allopolyploid, and sugarcane, an autopolyploid, exemplify successful polyploids, and some of their unique features are currently being revealed by comparative genomics**.**

Cotton: Permanent Hybridity by Allopolyploidy

In allopolyploidy, multiple genomes that previously had existed separately from one another coexist within a common nucleus, with strict one-to-one (disomic) pairing of chromosomes at meiosis. The predominant cultivated cottons, *Gossypium hirsutum* and *G. barbadense*, are allotetraploids, thought to have been formed ~1 to 2 million years ago by transoceanic migration of an Old World (A-genome) progenitor followed by hybridization with a New World (D-genome) progenitor (Wendel, 1989). The tetraploid chromosomes have been aligned with their diploid progenitors (Brubaker et al., 1999).

Both A-genome diploid and AD-tetraploid Gossypium taxa have been domesticated; however, intense directional selection by humans has consistently produced AD-tetraploid cottons with yield or quality (or both) superior to that of A-genome diploid cultivars.

Curiously, the superior genetic potential of tetraploid cotton over its A-genome progenitor is apparently attributable to contributions from the D subgenome, from the diploid ancestor that does not produce spinnable fiber. Most quantitative trait loci (QTLs) that account for genetic variation in fiber traits of modern *G. barbadense* and *G. hirsutum* map to the D subgenome (Jiang et al., 1998). A tendency toward relatively high polymorphism of the D subgenome has also been observed for other traits, including disease resistance (Wright et al., 1998), plant morphology (Jiang et al., 2000), and adaptations to arid conditions (Y. Saranga, M. Menz, C. Jiang, R. Wright, D. Yakir, and A.H. Paterson, unpublished data). An intriguing and perhaps significant observation is that many families of dispersed repetitive DNA elements, largely confined to A-genome diploid cottons, have spread to the D subgenome of tetraploid cottons (Hanson et al., 1998; Zhao et al., 1998). Together, these repetitive DNA families represent a potentially powerful mutagen, with an estimated 1.4 million copies present in the tetraploid genome (Zhao et al., 1998). Whether such repetitive DNA is responsible for generating the D-subgenome alleles that contribute to the transgressive variation of tetraploid cottons relative to their diploid progenitors remains to be determined.

Sugarcane: Genetic Buffering by Autopolyploidy

In autopolyploidy, multiple sets of chromosomes in a common nucleus retain the possibility of multipartner (polysomic) pairing of individual chromosomes at meiosis. Many evolutionary models consider autopolyploidy to be a transient phase, with selection gradually leading to reestablishment of disomic pairing relationships (allopolyploidy). The extent of diploidization of chromosomes, as reflected by bivalent inheritance, is considered one indicator of the antiquity of autopolyploid formation. Comparative genetic methods permit the extent of preferential chromosome pairing to be estimated directly (Al-Janabi et al., 1994).

The consequences of autopolyploid formation in plants are exemplified by comparing the genus Saccharum (sugarcane) with the closely related genus Sorghum. In as few as 5 million years since Saccharum and Sorghum diverged from a common ancestor (Al-Janabi et al., 1994; Sobral et al., 1994), *Saccharum* spp have reached gametic chromosome numbers ranging from 18 to 85 or more versus 10 for *Sorghum* spp. The high numbers of chromosome duplication and autogamous chromosome pairing in sugarcane preclude genetic mapping based on codominant "alleles." DNA markers showing simplex (single-dose) segregation (Wu et al., 1992) have been used to construct genetic linkage maps of at least five sugarcane populations (Al-Janabi et al., 1993; da Silva et al., 1993, 1995; Grivet et al., 1996; Mudge et al., 1996; Ming et al., 1998). The sugarcane maps range from 160 (Mudge et al., 1996) to 2460 loci (Ming et al., 1998) and collectively are estimated to cover ~70% of the genome (Ming et al., 1998).

The close relationship, high degree of colinearity, and cross-hybridization of DNA probes in *Saccharum* spp compel the use of the small genome of *Sorghum* to guide molecular mapping and positional cloning. Dufour et al. (1997) evaluated the colinearity of sugarcane and sorghum genomes based on 84 anchor probes distributed at intervals of ~20 centimorgans (cM) across the Sorghum genome.

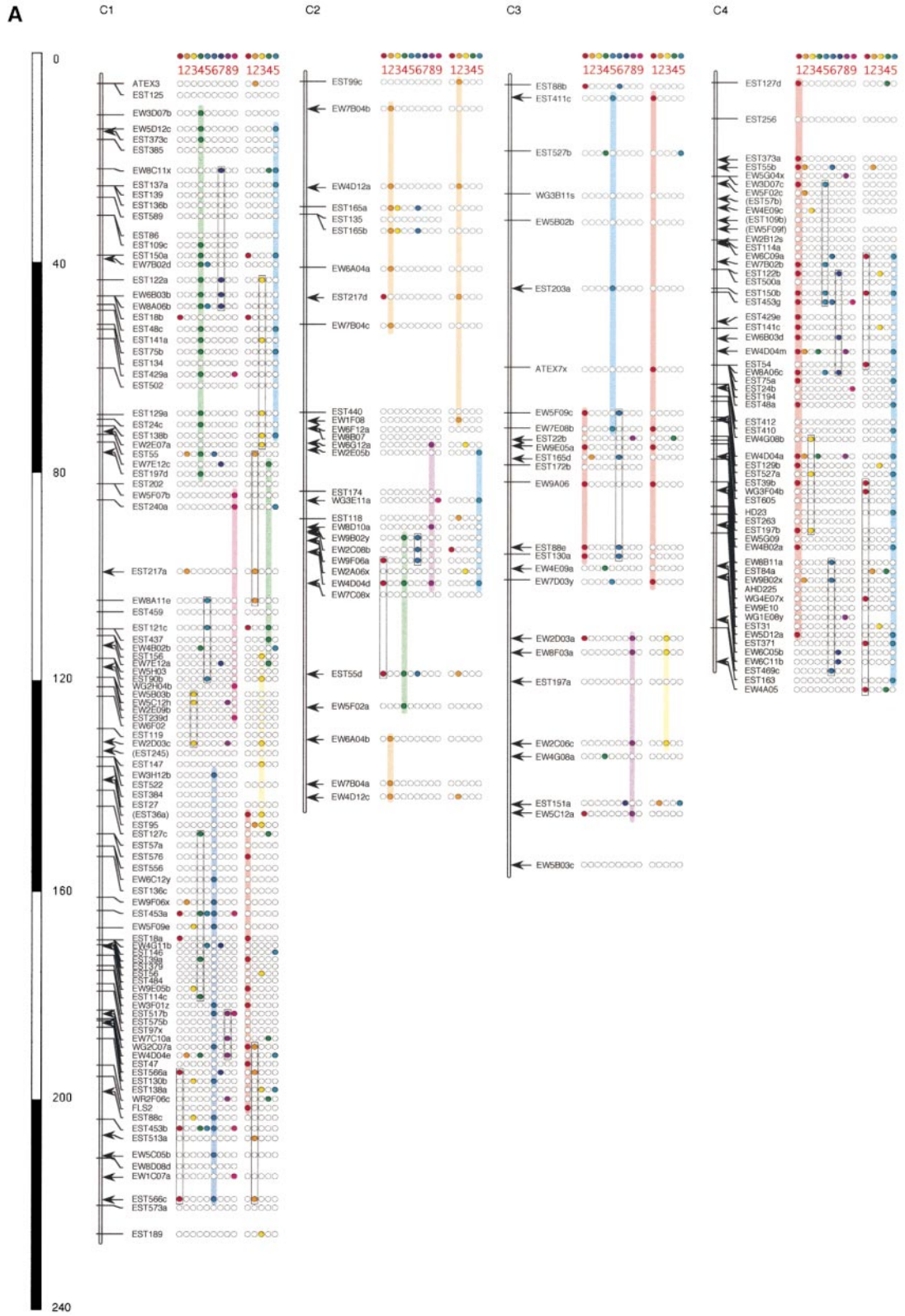

Figure 1. Composite Restriction Fragment Length Polymorphism Linkage Map of *Brassica oleracea*, and Its Alignment to the Map of *Arabidopsis thaliana*.

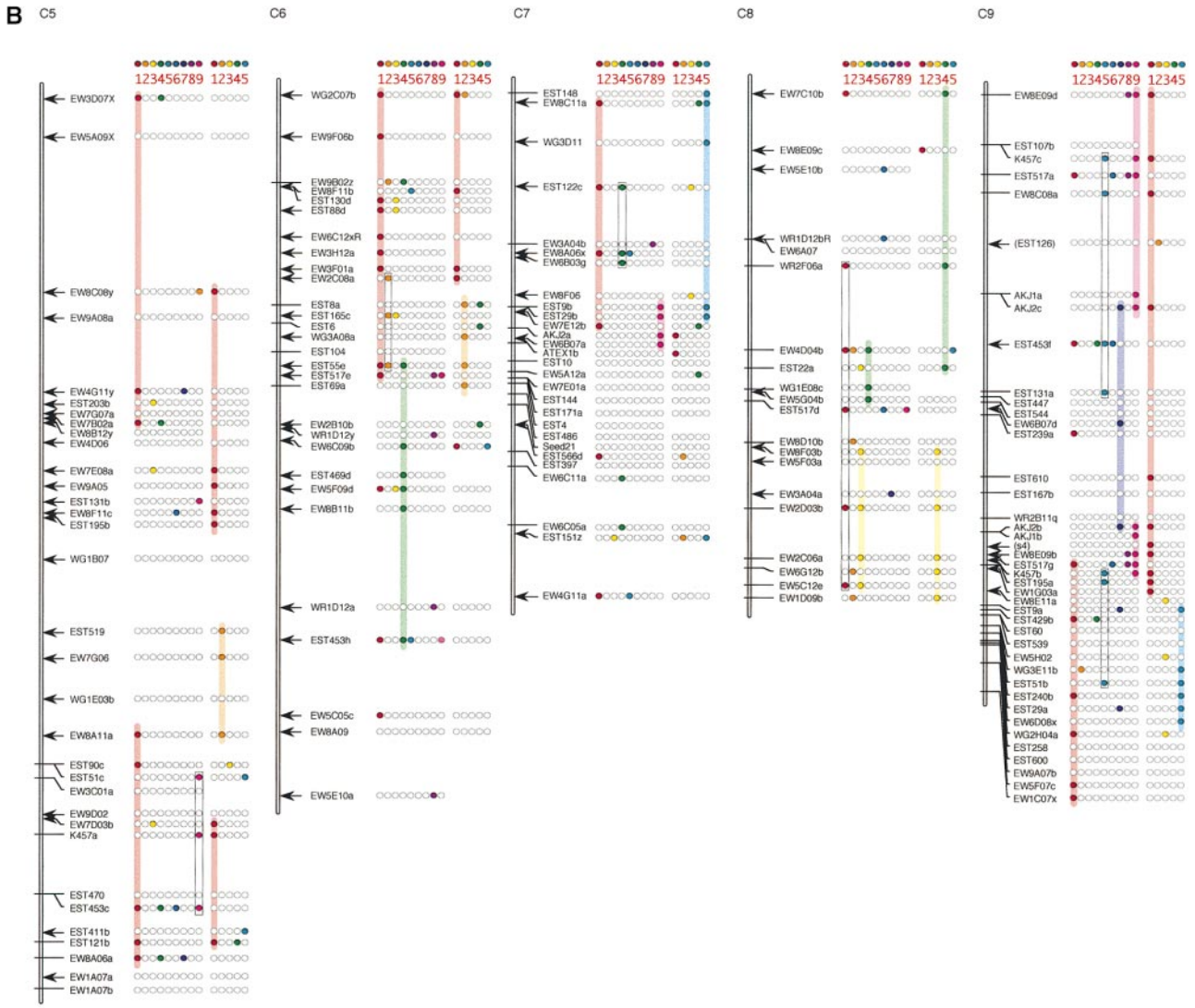

Figure 1. (continued).

The filled circles next to the loci indicate homoeologous Brassica loci (chromosomes 1 to 9, near right) or homologous Arabidopsis loci (chromosomes 1 to 5, far right) detected by the same probe. When all the circles are open, no polymorphism was detected for homoeologous (Brassica) or homologous (Arabidopsis) loci. An R next to the probe name indicates that the probe hybridizes to a repetitive DNA sequence in Arabidopsis. Specific colors were assigned to each homoeologous (Brassica) and homologous (Arabidopsis) chromosome. Markers that appear to represent duplication of Brassica chromatin or orthology between Brassica and Arabidopsis (based on criteria described in Lan et al., 2000a) were connected by colored columns. Open columns indicate possible triplicated (Brassica) or duplicated (Arabidopsis) regions. Vertical axis indicates centimorgans. (Modified from Lan et al. [2000a], with permission.)

Guimaraes et al. (1997) compared *Saccharum* spp and Sorghum by using only 68 probes, and 63 (69%) of the Saccharum linkage groups they proposed contained two or fewer loci, making it impossible to test whether intrachromosomal rearrangements were present. Ming et al. (1998) found that ~84% of the loci mapped by 242 common probes were colinear between Saccharum and Sorghum. At least five homologous Saccharum groups correspond largely if not completely to single Sorghum chromosomes. Only one interchromosomal and two intrachromosomal rearrangements differentiated Saccharum and Sorghum, but several possible cases of chromosome polymorphism were found within Saccharum.

Evaluation of QTLs has provided a glimpse into the

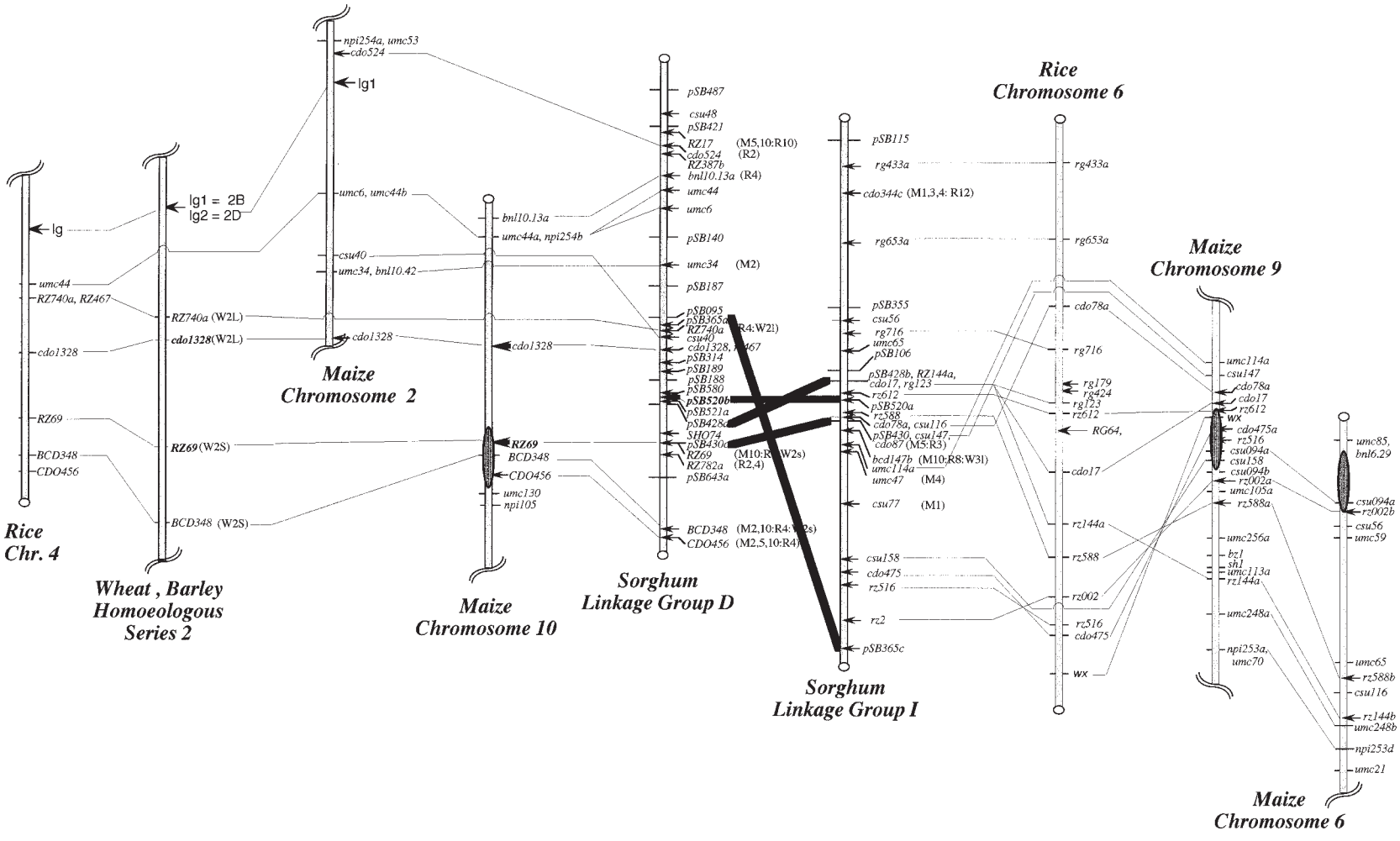

Figure 2. Comparative Mapping of Cereal Chromosomes.

molecular basis of phenotypic buffering that may contribute to the success of autopolyploid crops. The autopolyploidy of sugarcane is apparent in the high degree of observed duplication of QTLs. For example, 36 QTLs for sugar content, mapped in four different Saccharum genotypes, correspond to only seven distinct regions of the genome, suggesting that the 36 QTLs may be accounted for by many fewer genes (R. Ming, S.-C. Liu, J.E. Irvine, and A.H. Paterson, unpublished data). In several cases, two or more loci detected by the same DNA probe were each associated with variation in the same trait. Such associations may reflect either the multiplex segregation of individual QTLs or the coincidence of different QTL alleles at nearby loci; in either case, the consequences of multiple copies of a genomic region, each associated with common phenotypic effects, are apparent. In virtually all such cases, additional doses of QTLs are associated with diminishing returns, similar to the less-than-additive epistatic interactions of unlinked QTLs in diploid tomato (Eshed and Zamir, 1996). Epistasis in sugarcane is complicated by the possibility of nonlinear interactions between different alleles at homologous sites in addition to nonlinear interactions between unrelated loci (Eshed and Zamir, 1996).

Many Modern Diploids Appear to Be Ancient Polyploids

Contemporary polyploids provide further insight into instances of gene and genome duplication that have contributed to the evolution of genomes (Figures 2 and 3C) perceived today to be diploid. Indeed, most flowering plant genomes are thought to have gone through one or more cycles of polyploid formation (Stebbins, 1966; Masterson, 1994). A surprising discovery based on its near-complete sequencing is that much of the Arabidopsis genome appears to be duplicated (Figures 4A and 4B), supporting prior suggestions of segmental duplications (Kowalski et al., 1994; Paterson et al., 1996; Lan et al., 2000a) based on mapping restriction fragment length polymorphisms (RFLPs). Similar genetic mapping evidence of ancient duplication exists for sorghum (Chittenden et al., 1994; Pereira et al., 1994; Paterson et al., 1996) and rice (Kishimoto et al., 1994; Nagamura et al., 1995).

An important question concerns the fate of genes after duplication. For example, the data depicted in Figure 4 suggest that approximately two-thirds of the Arabidopsis genome has been duplicated; however, only ~19% of the total number of expressed sequence tags (ESTs) studied fell into the linked groups that suggest duplication. The vast majority of genes showed either no duplication or no discernible pattern in the arrangements of duplicated copies. If indeed there was a genome-wide duplication in an Arabidopsis progenitor, the vast majority of duplicated genes have been lost, moved to new locations, or diverged so greatly that they are no longer recognizable. Isoenzyme and molecular studies have provided evidence for gene silencing in polyploid angiosperms and ferns (Soltis and Soltis, 1993). Only a small proportion of the triplicated isoenzyme genes in hexaploid wheat have been lost (Hart, 1983); however, formation of hexaploid wheat is very recent. Duplicate isoenzyme loci have been reported in maize (Wendel et al., 1986, 1989) and sorghum (Morden et al., 1989), but many isoenzyme systems also occur in numbers typical of diploid plants (Wendel et al., 1986; Morden et al., 1989), which suggests gene loss. Molecular evidence suggests that both gene silencing and functional divergence have occurred in the leghemoglobin gene family in the polyploid angiosperm soybean (*G. max*; Lee and Verma, 1984). Gene loss, together with localized rearrangements, may contribute to deviations from synteny that are found both in genetic maps and in localized comparisons of bacterial artificial chromosome (BAC) sequences.

Consequences of Polyploidy for Physical Mapping

The combination of polyploidy and heterozygosity that is inherent to crops such as cotton and sugarcane but rare in model organisms such as Caenorhabditis, Arabidopsis, and

Figure 2. (continued).

The genetic maps of relevant regions of the genomes of rice, wheat, barley, and maize are aligned to the map of sorghum. DNA marker loci indicated by a line (at left of the gene name) were directly mapped in the cited populations, whereas those indicated by an arrow were mapped in other populations and the appropriate locations were inferred on the basis of map positions of flanking markers. Approximate locations of centromeres (ovals within linkage groups), telomeres (circles at ends of linkage groups), and breakpoints of chromosomal rearrangements (double squiggles) distinguishing taxa have been indicated. For DNA marker loci that conflict with the most-parsimonious interpretation of chromosomal correspondence between taxa, alternative map positions have been indicated in parentheses, adjacent to the appropriate locus in sorghum. For cases in which the exact determination of marker order in local regions could not be inferred, the order most parsimonious with that in other species was assumed. Because the populations used to map "anchor loci" were small (56 to 90 individuals), local reversals in order of closely linked markers (>3 cM) do not provide conclusive evidence of chromosomal rearrangement between taxa. Dark lines connecting selected probes on sorghum linkage groups D and I represent evidence of segmental duplication of this chromosomal region, as previously described (Chittenden et al., 1994). (Excerpted with permission from Paterson et al. [1995]. Copyright 1995, American Association for the Advancement of Science.)

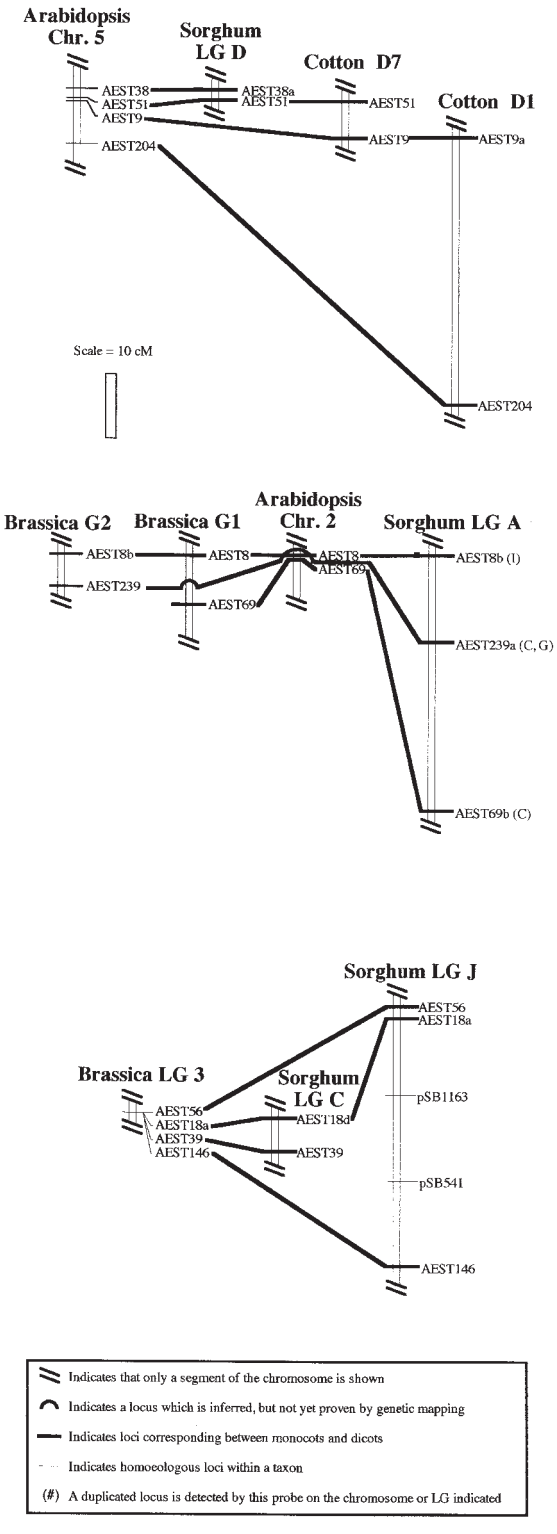

Figure 3. Examples of Colinearity of Monocot and Dicot Genes.

Arabidopsis cDNAs that show DNA sequence conservation (BLASTx > 150) with genes from monocots or more distant taxa were used to detect RFLPs and were then added to existing genetic maps of *Sorghum bicolor* × *S. propinquum*, *Arabidopsis thaliana*, *B. oleracea* (T.-H. Lan et al., 2000a), and *Gossypium trilobum* × *G. raimondii* (C. Brubaker, A.H. Paterson, and J.F. Wendel, unpublished data). Lowercase letters indicate that additional loci were detected by the same DNA probe. Chromosomal locations of duplicate loci are shown either directly or in parentheses. AEST, Arabidopsis cDNAs; pSB, sorghum PstI genomic DNA clones. LG, linkage group. (Modified from Paterson et al. [1996], with permission.)

mouse poses a new challenge to genomics. Traditional dot-blot hybridization or polymerase chain reaction–based assays for identifying BAC clones corresponding to a mapped DNA landmark usually do not provide sufficient information to distinguish between allelic and nonallelic loci. Use of DNA fingerprint data to assemble locus-specific contigs for duplicated chromosomal regions requires a means for distinguishing not only between the genomic DNA of the underlying clones but also between distinct loci that have arisen subsequent to the establishment of polyploidy. To assemble a contig that truly represents a continuous sequence within a polyploid genome, it is necessary to ascertain that the component BAC (or other) clones are derived from the same genetic locus. Individual BACs can be assigned to their source loci within the polyploid genome by detecting sequence variation among homologous and homeologous DNA sequences (i.e., alloalleles; Reinisch et al., 1994); such detection is not possible by traditional dot-blot or polymerase chain reaction–based screening methods. In the BAC-RF method (Lin et al., 2000), DNA probes hybridize to both restriction digestions of BAC DNA and dot blots. Restriction digestion of pools of BAC DNA indicates the size of the hybridizing fragment or fragments, which can be compared with the sizes of alleles that have been genetically mapped to their chromosomal locations. Dot blots yield qualitative data about which BACs hybridize to which probes. BAC pools confer advantages similar to those realized by use of radiation hybrid cell lines (Goss and Harris, 1975; Cox et al., 1990), obviating the need for DNA polymorphism to map loci to their chromosomal locations at fine-scale resolution. BAC-RF will facilitate contig assembly for many polyploid crops and similarly provides a facile means to identify and isolate individual members of multigene or repetitive DNA families.

ANGIOSPERM CHROMOSOME EVOLUTION

As exemplified above, closely related genera often retain similar gene order along large segments of their genetic

maps, punctuated by structural mutations such as inversions and translocations. Several examples point to the possibility that conservation of gene order might extrapolate to more distantly related taxa. The two major subclasses of flowering plants, monocots and dicots, are thought to have diverged from a common ancestor ~130 to 240 million years ago (Wolfe et al., 1989; Crane et al., 1995). The possibility that discernible parallels in gene order may have persisted over this time is supported in a general sense by the observation that nine pairs of taxa for which both comparative genetic maps and plausible estimates of divergence time are available show an average of 0.14 (±0.06) structural differences per chromosome per million years of divergence (Paterson et al., 1996); in monocots and dicots, this rate of divergence would suggest that 43 to 58% of the chromosomal tracts that span ≤3 cM would remain uninterrupted by this type of major structural rearrangement. The extent of parallel arrangement of common DNA probes mapped in sorghum, cotton, Arabidopsis, and Brassica is not explicable by chance (Paterson et al., 1996; see Figure 3 for examples). However, conservation of order among syntenic but widely dispersed markers may not represent colinearity of all genes between the markers (see Paterson et al., 1996), and determination of the shortest conserved evolutionary unit sequence (SCEUS) depends on the density of comparative markers (O'Brien et al., 1993). Indeed, recent studies suggest that many localized rearrangements can be accounted for by smaller scale events superimposed on chromosome-level events.

Local Genomic Colinearity

Comparisons of recombination-based genetic maps are limited in sensitivity, partly by the number of orthologous probes available and partly by population sizes relative to the frequency of recombination. Most RFLP maps contain only a few hundred comparable markers at most, confining analysis of colinearity to 5- to 10-cM segments. Furthermore, many mapped populations are based on <100 individuals, which means that unambiguous resolution of the orders of genes is limited to those that are at least 1 cM apart. This low sensitivity could miss many small deletions, duplications, and inversions. In contrast, comparative restriction mapping or sequence analysis of orthologous genomic segments of 100 kb or so cloned into BAC or yeast artificial chromosome vectors can provide highly detailed comparisons, albeit for only a small portion of any genome.

Comparative studies of the restriction maps of yeast artificial chromosome or BAC clones (Dunford et al., 1995; Kilian et al., 1995) and genomic sequence comparisons (Chen et al., 1998; Tikhonov et al., 1999) continue to support the notion of colinear gene orders; intergenic sequences, however, are highly variable. In particular, the large maize genome contains substantial amounts of intergenic DNA, most of which consists of retrotransposons inserted within each other (SanMiguel et al., 1996; Chen et al., 1997, 1998; Tikhonov et al., 1999). These intergenic retrotransposons show a preference for insertion between genes and have been amplified within the last 2 to 6 million years to make up >50% of the total maize nuclear genome (SanMiguel et al., 1996, 1998; SanMiguel and Bennetzen, 1998). Hence, intergenic retrotransposons largely account for the greater physical distance between genes in large-genome grasses in comparison to small-genome grasses (Chen et al., 1997; Tikhonov et al., 1999). Clearly, these mobile DNAs are usually able to insert into genomic regions without affecting local gene content or order.

Some exceptions to colinearity have also been observed with genes in otherwise orthologous regions. For instance, orthologous regions of the maize, sorghum, and rice genomes all exhibit tight genetic and physical linkage of *sh2* and *a1* homologs, but these two loci are not linked in Arabidopsis (Bennetzen et al., 1998). Even in maize and sorghum, two closely related species that descended from a common ancestor ~15 to 20 million years ago (Gaut and Doebley, 1997), some exceptions to colinearity were observed in *adh1* homologous regions (Tikhonov et al., 1999).

Sequence-Based Comparisons

Comparisons of DNA sequences from taxa whose genomes have been only partially sequenced provide mixed results in supporting or refuting the value of gene order in one taxon for predicting gene order in other taxa (Gale and Devos, 1998). Occasional artifacts may result from "best available matches" between genes that are in fact not orthologous but merely similar, perhaps reflecting the suggestion that the "universe of exons" may be as few as 5000, arranged in various combinations and permutations into a much larger universe of genes (Dorit et al., 1990). A modal SCEUS length of 3 cM between monocots and dicots has been estimated (Paterson et al., 1996), but variable rates of chromosome structural divergence are evident in different lineages, contributing to a coefficient of variation of ~40% (0.06/0.14). Given these considerations, along with well-established variation in the relationship between genetic (cM) and physical (kb) distances, the observation that sampling of different genomic regions from different taxa would show different degrees of structural conservation is not startling.

A recent comparison of rice ESTs with the Arabidopsis genomic sequence suggests that identification of orthologs by using BLASTN may not always be reliable when dealing with divergent, incompletely sequenced genomes (Devos et al., 1999). That BLASTN encounters difficulties in identifying orthologs is not surprising, given that the BLASTN algorithm imposes a tradeoff between speed and accuracy (Altschul et al., 1990) and does not recognize moderately distant nucleotide homologies efficiently. Indeed, Karlin and Altschul (1990) have shown, using default scoring matrices, that BLASTN targets matches of >98% sequence similarity.

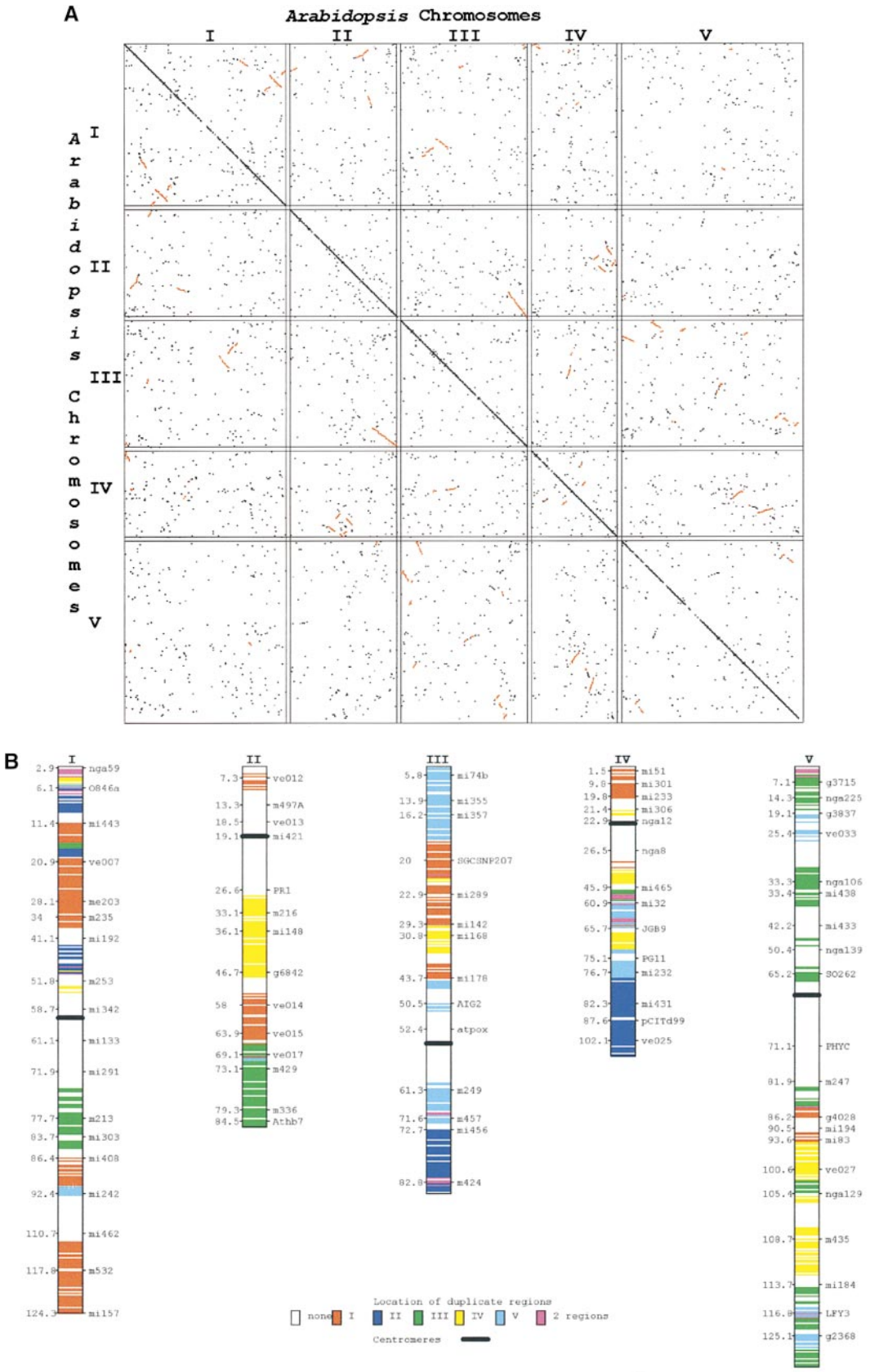

Figure 4. Evidence for Nearly Genome-wide Duplication of Arabidopsis Chromatin.

Nucleotide sequence similarities between divergent species may thus be eroded to the extent that sequence identity calculated for orthologs is similar to that among other members of the gene family. Alternatively, amino acid sequence similarity has been shown to identify homologs that diverged as much as 2.5 billion years ago (Pearson, 1997). A comparison at the amino acid level would thus be more effective than use of BLASTN to establish orthologous relationships between rice and Arabidopsis.

In addition, nucleotide or amino acid sequence identity is inadequate for identifying orthologs from highly divergent species (Tatusov et al., 1997). Gene loss after gene duplication complicates identification of orthologs when highest sequence identity is the sole criterion. If two genomes lose different paralogs of an ancestral gene that was duplicated before species divergence, the remaining paralogs will have the highest sequence identity (Huynen and Bork, 1998). Orthologs within large multigene families, which are common in angiosperms, may be especially difficult to identify by using sequence identity.

In summary, comparative information about chromosome organization is of high value in closely related taxa, in which deviations from colinearity are infrequent. Across more distantly related taxa (such as different taxonomic families), chromosomal rearrangements, together with gene duplication, divergence, and localized rearrangement, gradually reduce the value of comparative maps for predicting gene content and order. Although patterns can still be discerned that reflect gene arrangements as ancient as those tracing back to the common ancestor of monocots and dicots, or a lower ploidy ancestor of modern *Arabidopsis thaliana*, such patterns are considerably obscured by extensive gene deletion or divergence. Ancient chromosomal duplication, now suspected to have occurred even in the streamlined genome of Arabidopsis, may also explain some of the apparent incongruities in comparative maps. However, even modest degrees of conservation across large evolutionary distances are of interest in providing a framework for studying the course of plant chromosome evolution.

Comparative QTL Mapping

The molecular dissection of complex traits through breeding approaches (see Paterson et al., 1990), coupled with the development of comprehensive EST databases and their use for high-throughput parallel analysis of gene expression (see DeRisi et al., 1997), promises to add much to our understanding of the relationship of genes to phenotypes. At present, however, the vast majority of phenotypes cannot yet be assigned to specific genes and are known only to lie within some chromosomal region perhaps 10 to 30 cM long, with a specified level of confidence. Statistical assessments continue to be the best available means of evaluating the extent of similarity in the genetic determinants of such complex traits across different taxa. A sampling of the following three statistical studies illustrates the messages that are emerging from comparative QTL mapping.

QTLs Associated with Domestication of Cereal Crops

An early study (Paterson et al., 1995) described the comparative molecular analysis of the sets of genes associated with domestication of three major cereal crops (Sorghum, Oryza, and Zea), each on a different continent. A framework of common DNA markers provided a basis for evaluating how well the locations of genes that confer common phenotypes corresponded. Correspondence was evaluated among

Figure 4. (continued).

(A) Dot plot showing locations of duplicated genes in Arabidopsis. The *x* and *y* axes each represent 1452 completely sequenced Arabidopsis BACs, arranged in their chromosomal order, according to the reference genetic markers in **(B)**. Points on the graph represent the best two BLAST hits on the Arabidopsis BACs, for 15,199 Arabidopsis ESTs showing high similarity ($<1 \times 10^{-14}$ likelihood of occurring by chance) to two nonoverlapping BACs, selected from >30,000 Arabidopsis ESTs available in GenBank (accession numbers AV530001 to AV560000). The 15,199 ESTs that were duplicated hit 5084 different BAC pairs, shown as points in the figure. Regions in which five or more BAC pairs showed hits in a square 6×6 region of 36 BACs are shown in red, indicating regions of duplication. Because of the resolution of this figure, it is difficult visually to resolve the number of points in a region; a tight clustering of points may appear less significant than a more dispersed grouping. A total of 42 segments appear to be duplicated. The 880 points that are shown in red represent 5780 (19.2%) of the total number of ESTs and 38% of the ESTs that were duplicated.

(B) Representation of duplications on Arabidopsis chromosomes. Duplicated regions from **(A)** are shown along the Arabidopsis genetic map, including commonly used reference markers. Colored blocks represent the chromosomal location of the duplicated copy for the indicated chromosomal region. Of the 1452 completely sequenced Arabidopsis BACs on which this comparison was based, 841 (58%) appear to have been duplicated. Because these BACs cover only ~90% of the genome, this suggests that at least two-thirds of the Arabidopsis genome has been duplicated. Only 49 (3%) appear to be represented in more than one duplicated region, suggesting that relatively little of the genome is involved in higher order levels of duplication (such as triplication). Curiously, the regions surrounding the centromeres show little evidence of duplication, although the physical maps and sequences of these regions tend to be less complete. Centimorgans are shown to left of chromosomes.

genes/QTLs associated with temperate (day-neutral) flowering, restricted seed dispersal, and increased seed size in crosses between divergent Sorghum, Oryza, and Zea taxa. Genes/QTLs were found to correspond far more often than would be expected to occur by chance, suggesting that corresponding genes may be involved in evolution of these phenotypes. Convergent domestication of sorghum, rice, and maize, as reflected by mutations at corresponding genetic loci, suggests that a few genes with large effects may mostly determine the phenotypes studied.

Correspondence in location of QTLs in different taxa does not prove identity between the underlying genes, but it does suggest the identity of some of them. Map-based cloning of QTLs, a heretofore refractory objective, appears more tractable in view of these results. Specifically, "parallel walks" to corresponding genes in different species may obviate some obstacles often associated with positional cloning.

In all three taxa used in these studies, the crosses studied were highly divergent. In both Sorghum and Zea, a cultivated type was crossed to a wild relative; in Oryza, divergent subspecies were crossed to one another. Hence, the genes that were mapped are likely to reflect fundamental changes in the early stages of domestication, and the results may not therefore be predictive of the levels and patterns of conservation of subtle differences among elite cultivars.

Comparative Analysis of Pest Resistance Genes

The greenbug, *Schizaphis graminum* (Rondani), survives on a wide range of grass species. The parthenogenetic nature of this aphid's reproduction leads to rapid propagation of virulent biotypes: in only 30 years, *S. graminum* has given rise to eight new biotypes that have overcome both host plant resistance and susceptibility to organophosphate insecticides. Sorghum, wheat, and barley each contain greenbug resistance genes at comparable chromosomal locations—specifically, among sorghum linkage groups D, F, I, and J, and Triticeae homologous groups 1, 6, and 7 (C.S. Katsar, G. Peterson, Y. Lin, G.L. Teetes, and A.H. Paterson, unpublished data). Crop rotation has long been a key component of integrated pest management, with the tacit assumption that different crops offer different obstacles to the pest. The sorghum–wheat rotation commonly used in the Southern Plains of the United States, however, may be inadequate to disrupt the life cycle of this pest, and correspondence among resistance loci across various taxa suggests the need for more effective "gene rotation" strategies. Thus, positional and functional information about resistance genes in different crops is vital to ensure that crop rotation truly accomplishes the goal of challenging pests with a series of diverse resistance genes/mechanisms. Further, many of sorghum's greenbug resistance genes seem to fall within duplicated chromosomal regions, heightening the possibility of further redundancy of resistance mechanisms.

Diversity in QTLs Suggests That Brassica Is a Rapidly Evolving Genus

Comparing three morphologically diverse *B. oleracea* genotypes, each crossed to a common rapid-cycling genotype, we studied the extent to which apparently similar phenotypes were under common genetic control in the Brassica genus. Although this experiment was designed to investigate the comparative evolution of the inflorescence, other aspects of plant morphology were also studied to provide a more detailed understanding of the suite of changes needed to transform a small, ephemeral plant into, for example, cauliflower. For example, we inferred that the 47 QTLs found to affect measures of plant size and shape in the three different crosses reflected variation in ~35 different genetic loci, illustrating that most QTLs differed between populations (T.-H. Lan and A.H. Paterson, manuscript submitted). This picture tends to contradict the pattern found in the grasses, which showed highly similar genetic controls of a trait in different genotypes. However, diversity of QTLs is reinforced by QTL data for other traits in Brassica (Lan and Paterson, 2000b); previously published data (Kowalski et al., 1994) also suggest that the Brassica genome may be relatively rapidly evolving at the structural level.

SYNTHESIS

Approximately a decade ago, the suggestion that diverse plant species might resemble each other in terms of gene content (Tanksley et al., 1988) and order (Bonierbale et al., 1988; Hulbert et al., 1990) came as something of a surprise, especially in view of the tremendous developmental, biochemical, and morphological diversity that has arisen over the 130- to 240-million-year history of angiosperm evolution. Today, comparative genetic maps allow model species to guide inquiry and provide insights into the genomes and evolution of multiple experimental species. Widespread use of heterologous DNA markers has accelerated genome analysis in many crops.

Nevertheless, present-day comparative genomic maps must be interpreted with caution. Most comparative genetic maps continue to be based on a relative paucity of DNA markers, and virtually all are dependent on the availability of DNA polymorphism, which means that only a subset of loci can be mapped. In addition, regardless of how much genomic colinearity may exist between two species, the very fact of speciation means that a certain degree of noncolinearity is likely (see, e.g., Dubcovsky and Dvorak, 1995; Leister et al., 1998).

Much current activity is centered on aligning the physical maps and genomic sequences from models such as Arabidopsis with genetic maps of major crops, potentially to aid in the cloning of agriculturally important genes or QTLs. Thousands of genetically mapped mutants of Arabidopsis,

maize, rice, and other taxa might similarly be united into a central tool for the comparative study of plant development. However, although chromosomal parallels appear to have persisted among distantly related plants, they are considerably obscured by extensive gene deletion or divergence. Ancient chromosomal duplication, now evident even in the streamlined genome of Arabidopsis, may also cause apparent incongruities that can be resolved only when comparative maps become very detailed.

Colinear RFLP maps and a close phylogenetic relationship are good indicators of local colinearity, but only detailed analysis of the targeted local region can reveal microcolinearity. Next-generation "gene maps" (Deloukas et al., 1998), and ultimately additional genomic sequences, will empower comparative studies at a level of resolution that far exceeds present capabilities. Fine-scale scans of genomes for features that implicate particular regions in the genetic control of key phenotypes are on the horizon and will reveal heretofore unappreciated evolutionary trends (G. Wang et al., 1995; Aaltonen et al., 1997; Peltonen and Uusitalo, 1997; Cohen, 1999; R. Wang et al., 1999). Predicted SCEUS lengths will prove useful in identifying unusual features of particular genomes. "Gene blocks" that are conserved because of advantages of particular structural features or gene arrangements ("position effects"; see Lewis, 1950) that impact expression and function will be of special interest. Methods for parallel "expression profiling" of thousands of genes simultaneously (DeRisi et al., 1997) will be indispensable to analyses of chromosome function on a global scale. Taxa such as Brassica and Secale, which show particularly high rates of chromosome structural rearrangement (see Paterson et al., 1996), might be fertile systems for investigating the structural mutation process. New methods also promise to reveal more detail about the arrangements of repetitive DNA elements on a genome-wide scale (Lin et al., 2000; Zhang et al., 2000). An especially intriguing issue is how and why parallel gene order is preserved along the chromosomes of taxa that have been reproductively isolated for millions of years despite rapidly changing arrangements of intergenic DNA.

Plant biology is entering a golden age in which high-throughput genomic enterprises, such as those providing the complete sequences for Oryza and Arabidopsis, will provide powerful new approaches to solving puzzles that have perplexed botanists for centuries. The extent of parallelism in gene orders is only one example of the kinds of learning opportunities that will reward public investment in comprehensive plant genomics programs. It is particularly exciting that such programs are beginning to expand beyond selected models to include the food, feed, and fiber crops that sustain humanity. Crops are domesticated because they exhibit one or more extraordinary features, such as the large carbohydrate-rich seeds of the cultivated cereals, the remarkably long and strong single-celled fibers of cotton, the curd-like semi-sterile inflorescence of cauliflower, and the bulbous berry of tomato. Each crop is intrinsically an elegant "model" that offers unique opportunities to make new advances in (comparative) plant biology. As such advances are made, crop improvement will be equipped with an unprecedented set of tools for further adaptation of plants to better meet human needs.

Received March 7, 2000; accepted August 14, 2000.

REFERENCES

Aaltonen, J., Horelli-Kuitunen, N., Fan, J., Björses, P., Perheentupa, J., Myers, R., Palotie, A., and Peltonen, L. (1997). High-resolution physical and transcriptional mapping of the autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy locus on chromosome 21q22.3 by FISH. Genome Res. **7,** 820–829.

Ahn, S., and Tanksley, S.D. (1993). Comparative linkage maps of rice and maize genomes. Proc. Natl. Acad. Sci. USA **90,** 7980–7984.

Ahn, S., Anderson, J.A., Sorrells, M.E., and Tanksley, S.D. (1993). Homoeologous relationships of rice, wheat, and maize chromosomes. Mol. Gen. Genet. **241,** 483–490.

Al-Janabi, S.M., Honeycutt, R.J., McClelland, M., and Sobral, B.W.S. (1993). A genetic linkage map of *Saccharum spontaneum* L. 'SES 208.' Genetics **134,** 1249–1260.

Al-Janabi, S.M., Honeycutt, R.J., Peterson, C., and Sobral, B.W.S. (1994). Phylogenetic analysis of organellar DNA sequences in the Andropogoneae: *Saccharum*. Theor. Appl. Genet. **88,** 933–944.

Al-Shehbaz, I.A. (1973). The biosystematics of the genus *Thelypodium* (Cruciferae). Contrib. Gray Herb. Harv. Univ. **204,** 3–148.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. **215,** 403–410.

Bennetzen, J.L., and Freeling, M. (1993). Grasses as a single genetic system: Genome composition, collinearity and compatibility. Trends Genet. **9,** 259–261.

Bennetzen, J.L., SanMiguel, P., Chen, M., Tikhonov, A., Francki, M., and Avramova, Z. (1998). Grass genomes. Proc. Natl. Acad. Sci. USA **95,** 1975–1978.

Berhan, A.M., Hulbert, S.H., Butler, L.G., and Bennetzen, J.L. (1993). Structure and evolution of the genomes of *Sorghum bicolor* and *Zea mays*. Theor. Appl. Genet. **86,** 598–604.

Binelli, G., Gianfrancesci, L., Pè, M.E., Taramino, G., Busso, C., Stenhouse, J., and Ottaviano, E. (1993). Similarity of maize and sorghum genomes as revealed by maize RFLP probes. Theor. Appl. Genet. **84,** 10–16.

Bonierbale, M.D., Plaisted, R.L., and Tanksley, S.D. (1988). RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics **120,** 1095–1103.

Brubaker, C.L., Paterson, A.H., and Wendel, J.F. (1999). Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. Genome **42,** 184–203.

Burow, M.D., Simpson, C.E., Paterson, A.H., and Starr, J.L.

(1996). Tagging of a gene for resistance to *Meloidogyne arenaria* in peanut. Mol. Breeding **2,** 369–379.

Chao, S., Sharp, P.J., Worland, A.J., Warham, E.J., Koebner, R.M.D., and Gale, M.D. (1989). RFLP-based genetic maps of wheat homoeologous group 7 chromosomes. Theor. Appl. Genet. **78,** 495–504.

Chen, M., SanMiguel, P., de Oliveira, A.C., Woo, S.-S., Zhang, H., Wing, R.A., and Bennetzen, J.L. (1997). Microcolinearity in the *sh2*-homologous regions of the maize, rice and sorghum genomes. Proc. Natl. Acad. Sci. USA **94,** 3431–3435.

Chen, M., SanMiguel, P., and Bennetzen, J.L. (1998). Sequence organization and conservation in *sh2/a1*-homologous regions of sorghum and rice. Genetics **148,** 435–443.

Cheung, W.Y., and Landry, B.S. (1996). Current status of genome mapping in the Cruciferae. In Genome Mapping in Plants, A.H. Paterson, ed (Landes Bioscience Press, Austin TX), pp. 193–210.

Chittenden, L.M., Schertz, K.F., Lin, Y.-R., Wing, R.A., and Paterson, A.H. (1994). A detailed RFLP map of *Sorghum bicolor* × *S. propinquum* suitable for high-density mapping suggests ancestral duplication of chromosomes or chromosomal segments. Theor. Appl. Genet. **87,** 925–933.

Chyi, Y.-S., Hoenecke, M.E., and Sernyk, J.L. (1992). A genetic linkage map of restriction fragment length polymorphism loci for *Brassica rapa* (syn. campestris). Genome **35,** 746–757.

Cohen, D. (1999). Plenary speech, International Plant and Animal Genome Conference VII, San Diego, CA, 18 January 1999.

Cook, D.R. (1999). *Medicago truncatula*—A model in the making! Curr. Opin. Plant Biol. **2,** 301–304.

Cox, D., Burmeister, M., Price, E., Kim, S., and Myers, R. (1990). Radiation hybrid mapping: A somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. Science **250,** 245–250.

Crane, P.R., Friis, E.M., and Raunsgaard-Pedersen, K. (1995). The origin and early diversification of angiosperms. Nature **374,** 27–33.

da Silva, J., Sorrells, M.E., Burnquist, W., and Tanksley, S.D. (1993). RFLP linkage map and genome analysis of *Saccharum spontaneum*. Genome **36,** 782–791.

da Silva, J., Honeycutt, R.J., Burnquist, W., Al-Janabi, S.M., Sorrells, M.E., Tanksley, S.D., and Sobral, B.W.S. (1995). *Saccharum spontaneum* L. 'SES 208' genetic linkage map combining RFLP- and PCR-based markers. Mol. Breeding **1,** 165–179.

Deloukas, P. (1998). A physical map of 30,000 human genes. Science **282,** 744–746.

DeRisi, J.L., Iyer, V.R., and Brown, P.O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. Science **278,** 680–686.

Devos, K., Atkinson, M.D., Chinoy, C.N., Liu, C.J., and Gale, M.D. (1992a). Chromosomal rearrangements in rye genome relative to that of wheat. Theor. Appl. Genet. **85,** 673–680.

Devos, K.M., Atkinson, M.D., Chinoy, C.N., Liu, C.J., and Gale, M.D. (1992b). RFLP-based genetic map of the homoeologous group 3 chromosomes of wheat and rye. Theor. Appl. Genet. **83,** 931–939.

Devos, K.M., Atkinson, M.D., Chinoy, C.N., Liu, C.J., and Gale, M.D. (1993). RFLP-based genetic map of the homoeologous group 2 chromosomes of wheat, rye, and barley. Theor. Appl. Genet. **85,** 784–792.

Devos, K.M., Dubcovsky., J., Dvorak, J., Chinoy, C.N., and Gale, M.D. (1995). Structural evolution of wheat chromosomes 4A, 5A, and 7B and its impact on recombination. Theor. Appl. Genet. **91,** 282–288.

Devos, K.M., Beales, J., Nagamura, Y., and Sasaki, T. (1999). *Arabidopsis*—Rice: Will colinearity allow gene prediction across the eudicot–monocot divide? Genome Res. **9,** 825–829.

Dorit, R.L., Schoenbach, L., and Gilbert, W. (1990). How big is the universe of exons? Science **250,** 1377–1382.

Dubcovsky, J., and Dvorak, J. (1995). Ribosomal RNA multigene loci: Nomads of the Triticeae genomes. Genetics **140,** 1367–1377.

Dufour, P., Deu, M., Grivet, L., D'Hont, A., Paulet, F., and Glaszmann, J.C. (1997). Construction of a composite sorghum genome map and comparison with sugarcane, a related complex polyploid. Theor. Appl. Genet. **94,** 409–418.

Dunford, R.P., Kurata, N., Laurie, D.A., Money, T.A., Minobe, Y., and Moore, G. (1995). Conservation of fine scale DNA marker order in the genomes of rice and the Triticeae. Nucleic Acids Res. **23,** 2724–2728.

Eshed, Y., and Zamir, D. (1996). Less-than-additive epistatic interactions of quantitative trait loci in tomato. Genetics **143,** 1807–1817.

Ferreira, M.E., Satagopan, J., Yandell, B.S., and Osborn, T.C. (1994). RFLP mapping of *Brassica napus* using doubled haploid lines. Theor. Appl. Genet. **89,** 615–621.

Figdore, S.S., Ferreira, M.E., Slocum, M.K., and Osborn, T.C. (1993). Association of RFLP markers with trait loci affecting clubroot resistance and morphological characters in *Brassica oleracea*. Euphytica **69,** 33–44.

Gale, M.D., and Devos, K.M. (1998). Plant comparative genetics after 10 years. Science **282,** 656–659.

Gaut, B.S., and Doebley, J.F. (1997). DNA sequence evidence for the segmental allotetraploid origin of maize. Proc. Natl. Acad. Sci. USA **94,** 6809–6814.

Goss, S., and Harris, H. (1975). New method for mapping genes in human chromosomes. Nature **255,** 680–684.

Grivet, L., D'Hont, A., Roques, D., Feldmann, P., Lanaud, C., and Glaszmann, J.C. (1996). RFLP mapping in cultivated sugarcane (*Saccharum* spp.): Genome organization in a highly polyploid and aneuploid interspecific hybrid. Genetics **142,** 987–1000.

Guimaraes, C.T., Sills, G.R., and Sobral, B.W.S. (1997). Comparative mapping of Andropogoneae: *Saccharum* L. (sugarcane) and its relation to sorghum and maize. Proc. Natl. Acad. Sci. USA **94,** 14261–14266.

Halward, T., Stalker, H.T., and Kochert, G. (1993). Development of an RFLP linkage map in diploid peanut species. Theor. Appl. Genet. **87,** 379–384.

Hanson, R.E., Zhao, X., Paterson, A.H., Islam-Faridi, M.N., Zwick, M.S., Crane, C.F., McKnight, T.D., Stelly, D.M., and Price, H.J. (1998). Concerted evolution of 20 interspersed repetitive elements in a polyploid. Am. J. Bot. **85,** 1364–1368.

Hart, G.E. (1983). Genetics and evolution of multilocus isozymes in hexaploid wheat. Isozymes Curr. Top. Biol. Med. Res. **10,** 365–380.

Hoenecke, M., and Chyi, Y.-S. (1991). Comparison of *Brassica napus* and *B. rapa* genomes based on restriction fragment length polymorphism mapping. In Proceedings of the Eighth International Rapeseed Congress GCIRC, Saskatoon, D.I. McGregor, ed pp. 1102–1107.

Hohmann, U., Graner, A., and Endo, T.R. (1995). Comparison of wheat physical maps with barley linkage maps for group 7 chromosomes. Theor. Appl. Genet. **91,** 618–626.

Hulbert, S.H., Richter, T.E., Axtell, J.D., and Bennetzen, J.L. (1990). Genetic mapping and characterization of sorghum and related crops by means of maize DNA probes. Proc. Natl. Acad. Sci. USA **87,** 4251–4255.

Huynen, M.A., and Bork, P. (1998). Measuring genome evolution. Proc. Natl. Acad. Sci. USA **95,** 5849–5856.

Jiang, C., Wright, R., El-Zik, K., and Paterson, A.H. (1998). Polyploid formation created unique avenues for response to selection in Gossypium (cotton). Proc. Natl. Acad. Sci. USA **95,** 4419–4424.

Jiang, C., Wright, R., Woo, S.-S., Delmonte, T.A., and Paterson, A.H. (2000). QTL analysis of leaf morphology in tetraploid *Gossypium* (cotton). Theor. Appl. Genet. **100,** 409–418.

Karlin, S., and Altschul, S.F. (1990). Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. Proc. Natl. Acad. Sci. USA **87,** 2264–2268.

Kianian, S.F., and Quiros, C.F. (1992). Generation of a *Brassica oleracea* composite RFLP map: Linkage arrangements among various populations and evolutionary implications. Theor. Appl. Genet. **84,** 544–554.

Kilian, A., Kudrna, D.A., Kleinhofs, A., Yano, M., Kurata, N., Steffenson, B., and Sasaki, T. (1995). Rice–barley synteny and its application to saturation mapping of the barley *Rpg1* region. Nucleic Acids Res. **23,** 2729–2733.

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

Kowalski, S.P., Lan, T.-H., Feldmann, K.A., and Paterson, A.H. (1994). Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea* chromosomes reveals islands of conserved organization. Genetics **138,** 499–510.

Kurata, N., Moore, G., Nagamura, Y., Foote, T., Yano, M., Minobe, Y., and Gale, M. (1994). Conversation of genome structure between rice and wheat. Bio/Technology **12,** 276–278.

Lagercrantz, U., and Lydiate, D. (1996). Comparative genome mapping in Brassica. Genetics **144,** 1903–1910.

Lan, T.H., Delmonte, T.A., Reischmann, K.P., Hyman, J., Kowalski, S., McFerson, J., Kresovich, S., and Paterson, A.H. (2000a). EST-enriched comparative map of *Brassica oleracea* and *Arabidopsis thaliana*. Genome Res. **10,** 776–788.

Lan, T.H., Kowalski, S., Delmonte, T., and Paterson, A.H. (2000b). Comparative evolution of QTLs sculpting the curd of *Brassica oleracea*. Genetics, in press.

Landry, B.S., Hubert, N., Etoh, T., Harada, J., and Lincoln, S.E. (1991). A genetic map for *Brassica napus* based on restriction fragment length polymorphisms detected with expressed DNA sequences. Genome **34,** 543–552.

Landry, B.S., Hubert, N., Crete, R., Chang, M.S., Lincoln, S.E., and Etoh, T. (1992). A genetic map of *Brassica oleracea* based on RFLP markers detected with expressed DNA sequences and mapping resistance genes to race 2 of *Plasmodiophora brassicae* (Woronin). Genome **35,** 409–419.

Lee, J.S., and Verma, D.P. (1984). Structure and chromosomal arrangement of leghemoglobin genes in kidney bean suggest divergence in soybean leghemoglobin gene loci following tetraploidization. EMBO J. **3,** 2745–2752.

Leister, D., Kurth, J., Laurie, D.A., Yano, M., Sasaki, T., Devos, K., Graner, A., and Schulze-Leifert, P. (1998). Rapid reorganization of resistance gene homologues in cereal chromosomes. Proc. Natl. Acad. Sci. USA **95,** 370–375.

Lewis, E.B. (1950). The phenomenon of position effect. Adv. Genet. **3,** 73–115.

Lin, Y., Draye, X., Qian, X., Ren, S., Zhu, L., and Paterson, A.H. (2000). Fine-scale mapping and sequence-ready contig assem-bly in highly-duplicated genomes, using the BAC-RF method. Nucleic Acids Res. **28,** e23. http://www3.oup.co.uk/nar/methods/Volume_28/Issue_07/gnd023_gml.abs.html.

Liu, C.J., Atkinson, M.D., Chinoy, C.N., and Gale, M.D. (1992). Nonhomoeologous translocations between group 4, 5, and 7 chromosomes within wheat and rye. Theor. Appl. Genet. **83,** 305–312.

Liu, Y.G., and Tsunewaki, K. (1991). Restriction fragment length polymorphism (RFLP) analysis in wheat. II. Linkage maps of the RFLP sites in common wheat. Jpn. J. Genet. **66,** 617–633.

Livingstone, K.D., Lackney, V.K., Blauth, J.R., van Wijk, R., and Jahn, M.K. (1999). Genome mapping in Capsicum and the evolution of genome structure in the Solanaceae. Genetics **152,** 1183–1202.

Lydiate, D., Sharpe, A., Lagercrantz, U., and Parkin, I. (1993). Mapping the *Brassica* genome. Outlook Agric. **2,** 85–92.

Marino, C.L., Nelson, J.C., Lu, Y.H., Sorrells, M.E., Leroy, P., Tuleen, N.A., Lopes, C.R., and Hart, G.E. (1996). RFLP-based linkage maps of the homoeologous group 6 chromosomes of hexaploid wheat (*Triticum aestivum* L. em. Thell). Genome **39,** 359–366.

Masterson, J. (1994). Stomatal size in fossil plants: Evidence for polyploidy in majority of angiosperms. Science **264,** 421–424.

Matzke, M.A., and Matzke, A.J.M. (1998). Polyploidy and transposons. Trends Ecol. Evol. **13,** 241.

Menacio-Hautea, D., Fatokum, C.A., Kumar, L., Danesh, D., and Young, N.D. (1993). Comparative genome analysis of mungbean (*Vigna radiata* L. Wilczek) and cowpea (*V. unguiculata*) using RFLP analysis. Theor. Appl. Genet. **86,** 797–810.

Mickelson-Young, L., Endo, T.R., and Gill, B.S. (1995). A cytogenetic ladder-map of the wheat homoeologous group-4 chromosomes. Theor. Appl. Genet. **90,** 1007–1011.

Ming, R., et al. (1998). Alignment of the *Sorghum* and *Saccharum* chromosomes: Comparative genome organization and evolution of a polysomic polyploid genus and its diploid cousin. Genetics **150,** 1663–1682.

Morden, C.W., Doebley, J.F., and Schertz, K.F. (1989). Allozyme

variation in old world races of *Sorghum bicolor* (Poaceae). Am. J. Bot. **76,** 247–255.

Mudge, J., Anderson, W.R., Kehrer, R.L., and Fairbanks, D.J. (1996). A RAPD genetic map of *Saccharum officinarum*. Crop Sci. **36,** 1362–1366.

Nagamura, Y., et al. (1995). Conservation of duplicated segments between rice chromosomes 11 and 12. Breeding Sci. **45,** 373–376.

Namuth, D.M., Lapitan, N.L.V., Gill, K.S., and Gill, B.S. (1994). Comparative RFLP mapping of *Hordeum vulgare* and *Triticum tauschii*. Theor. Appl. Genet. **89,** 865–872.

Naranjo, T., Roca, A., Goicoechea, P.G., and Giraldez, R. (1987). Arm homoeology of wheat and rye chromosomes. Genome **29,** 873–882.

Nelson, J.C., Van Deynze, A.E., Autrique, E., Sorrells, M.E., Lu, Y.H., Merlino, M., Atkinson, M., and Leroy, P. (1995a). Molecular mapping of wheat. Homoeologous group 2. Genome **38,** 116–124.

Nelson, J.C., Van Deynze, A.E., Autrique, E., Sorrells, M.E., Lu, Y.H., Negre, S., Bernard, M., and Leroy, P. (1995b). Molecular mapping of wheat. Homoeologous group 3. Genome **38,** 125–133.

Nelson, J.C., Sorrells, M.E., Van Deynze, A.E., Lu, Y.H., Atkinson, M., Bernard, M., Leroy, P., Faris, J.D., and Anderson, J.A. (1995c). Molecular mapping of wheat. Major genes and rearrangements in homoeologous groups 4, 5, and 7. Genetics **141,** 721–731.

O'Brien, S.J., Womack, J.E., Lyons, L.A., Moore, K.J., Jenkins, N.A., and Copeland, N.G. (1993). Anchored reference loci for comparative genome mapping in mammals. Nat. Genet. **3,** 103–112.

Paterson, A.H., DeVerna, J., Lanini, B., and Tanksley, S.D. (1990). Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes, from an interspecies cross of tomato. Genetics **124,** 735–742.

Paterson, A.H., Lin, Y.R., Li, Z., Schertz, K.F., Doebley, J.F., Pinson, S.R.M., Liu, S.C., Stansel, J.W., and Irvine, J.E. (1995). Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science **269,** 1714–1718.

Paterson, A.H., et al. (1996). Toward a unified map of higher plant chromosomes, transcending the monocot–dicot divergence. Nat. Genet. **14,** 380–382.

Pearson, W.R. (1997). Identifying distantly related protein sequences. Comput. Appl. Biol. Sci. **13,** 325–332.

Peltonen, L., and Uusitalo, A. (1997). Rare disease genes: Lessons and challenges. Genome Res. **7,** 765–767.

Pereira, M.G., Lee, M., Bramel-Cox, P., Woodman, W., Doebley, J., and Whitkus, R. (1994). Construction of an RFLP map in sorghum and comparative mapping in maize. Genome **37,** 236–243.

Prince, J., Pochard, P.E., and Tanksley, S.D. (1993). Construction of a molecular linkage map of pepper, and a comparison of synteny with tomato. Genome **36,** 404–417.

Rabinowicz, P.D., Braun, E.L., Wolfe, A.D., Bowen, B., and Grotewald, E. (1999). Maize R2R3 Myb genes. Sequence analysis reveals amplification in the higher plants. Genetics **153,** 427–444.

Reinisch, A.R., Dong, J.-M., Brubaker, C., Stelly, D., Wendel, J., and Paterson, A.H. (1994). An RFLP map of cotton (*Gossypium hirsutum* × *G. barbadense*): Chromosome organization and evolution in a disomic polyploid genome. Genetics **138,** 829–847.

SanMiguel, P., and Bennetzen, J.L. (1998). Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. Ann. Bot. **82,** 37–44.

SanMiguel, P., Tikhonov, A., Jin, Y.-K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S., Edwards, K.J., Lee, M., Avramova, Z., and Bennetzen, J.L. (1996). Nested retrotransposons in the intergenic regions of the maize genome. Science **274,** 765–768.

SanMiguel, P., Gaut, B., and Bennetzen, J.L. (1998). The paleontology of intergene retrotransposons in maize. Nat. Genet. **20,** 43–45.

Shultz, O.E. (1936). In Die Natürlichen Pflanzenfamilien, A. Engler and H. Harms, eds Vol. 17b, pp. 227–658.

Simmonds, N.W. (1976). Evolution of Crop Plants. Essex: Longman Scientific and Technical.

Slocum, M.K., Figdore, S.S., Kennard, W.C., Suzuki, J.Y., and Osborn, T.C. (1990). Linkage arrangement of restriction fragment length polymorphism loci in *Brassica oleracea*. Theor. Appl. Genet. **80,** 57–64.

Sobral, B.W.S., Braga, D.P.V., LaHood, E.S., and Keim, P. (1994). Phylogenetic analysis of chloroplast restriction enzyme site mutations in the *Saccharinae* Griseb. Subtribe of the *Andropogoneae* Dumort. Tribe. Theor. Appl. Genet. **87,** 843–853.

Soltis, D.E., and Soltis, P.S. (1993). Molecular data and the dynamic nature of polyploidy. Crit. Rev. Plant Sci. **12,** 243–273.

Song, K.M., Suzuki, J.Y., Slocum, M.K., Williams, P.H., and Osborn, T.C. (1991). A linkage map of *Brassica rapa* (syn. *campestris*) based on restriction fragment length polymorphism loci. Theor. Appl. Genet. **82,** 296–304.

Song, K., Lu, P., Tang, K., and Osborn, T.C. (1995a). Rapid genome change in synthetic polyploids of Brassica and its implications for polyploid evolution. Proc. Natl. Acad. Sci. USA **92,** 7719–7723.

Song, K., Slocum, M.K., and Osborn, T.C. (1995b). Molecular marker analysis of genes encoding morphological variation in *Brassica rapa* (syn. *campestris*). Theor. Appl. Genet. **90,** 1–10.

Stebbins, G.L. (1966). Chromosomal variation and evolution; polyploidy and chromosome size and number shed light on evolutionary processes in higher plants. Science **152,** 1463–1469.

Tanksley, S.D., Bernatzky, R., Lapitan, N.L., and Prince, J.P. (1988). Conservation of gene repertoire but not gene order in pepper and tomato. PNAS-USA **85,** 6419–6423.

Tanksley, S.D., et al. (1992). High density molecular linkage maps of the tomato and potato genomes. Genetics **132,** 1141–1160.

Tatusov, R.L., Koonin, E.V., and Lipman, D.J. (1997). A genomic perspective on protein families. Science **278,** 631–637.

Teutonico, R.A., and Osborn, T.C. (1994). Mapping of RFLP and qualitative trait loci in *Brassica rapa*, and comparison to linkage maps of *B. napus, B. oleracea,* and *Arabidopsis thaliana*. Theor. Appl. Genet. **89,** 885–894.

Tikhonov, A.P., SanMiguel, P.J., Nakajima, Y., Gorenstein, N.D., Bennetzen, J.L., and Avramova, Z. (1999). Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum. Proc. Natl. Acad. Sci. USA **96,** 7409–7414.

Torres, A.M., Weeden, N.F., and Martin, A. (1993). Linkage among isozyme, RFLP, and RAPD markers in *Vicia faba*. Theor. Appl. Genet. **85,** 937–945.

Truco, M.J., Hu, J., Sadowski, J., and Quiros, C.F. (1996). Inter- and intra-genomic homology of the Brassica genomes: Implica-

tions for their origin and evolution. Theor. Appl. Genet. **93,** 1225–1233.

Uzunova, M., Ecke, W., Weissleder, K., and Robbelen, G. (1995). Mapping of the genome of rapeseed (*Brassica napus* L.). I. Construction of an RFLP linkage map and localization of QTLs for seed glucosinolate content. Theor. Appl. Genet. **90,** 194–204.

Van Deynze, A.E., Dubcovsky, J., Gill, K.S., Nelson, J.C. Sorrells, M.E., Dvorak, J., Gill, B.S., Lagudah, E.S., McCouch, S.R., and Appels, R. (1995). Molecular-genetic maps for chromosome 1 in *Triticeae* species and their relation to chromosomes in rice and oats. Genome **38,** 47–59.

Vavilov, N.I. (1922). The law of homologous series in variation. J. Genet. **12**, 1.

Wang, G., Dong, J., and Paterson, A.H. (1995). The distribution of *Gossypium hirsutum* chromatin in *G. barbadense* germplasm: Molecular analysis of introgressive plant breeding. Theor. Appl. Genet. **91,** 1153–1161.

Wang, R.-L., Stec, A., Hey, J., Lukens, L., and Doebley, J. (1999). The limits of selection during maize domestication. Nature **398,** 236–239.

Weeden, N.L., Muehlbauer, F.J., and Ladizinsky, G. (1992). Extensive conservation of linkage relationships between pea and lentil genetic maps. J. Hered. **83,** 123–129.

Wendel, J.F. (1989). New World cottons contain Old World cytoplasm. Proc. Natl. Acad. Sci. USA **86,** 4132–4136.

Wendel, J.F., Stuber, C.W., Edwards, M.D., and Goodman, M.M. (1986). Duplicated chromosome segments in maize (*Zea mays* L.): Further evidence from hexokinase isozymes. Theor. Appl. Genet. **72,** 178–185.

Wendel, J.F., Stuber, C.W., Goodman, M.M., and Beckett, J.B. (1989). Duplicated plastic and triplicated cytosolic isozymes of triosephosphate isomerase in maize (*Zea mays* L.). J. Hered. **80,** 218–228.

Whitkus, R., Doebley, J., and Lee, M. (1992). Comparative genome mapping of sorghum and maize. Genetics **132,** 119–130.

Wolfe, K.H., Gouy, M., Yang, Y.W., Sharp, P.M., and Li, W.H. (1989). Date of the monocot–dicot divergence estimated from chloroplast DNA sequence data. Proc. Natl. Acad. Sci. USA **86,** 6201–6205.

Wright, R., Thaxton, P., Paterson, A.H., and El-Zik, K. (1998). Polyploid formation in *Gossypium* has created novel avenues for response to selection for disease resistance. Genetics **149,** 1987–1996.

Wu, K.K., Burnquist, W., Sorrells, M.E., Tew, T.L., Moore, P.H., and Tanksley, S.D. (1992). The detection and estimation of linkage in polyploids using single-dose restriction fragments. Theor. Appl. Genet. **83,** 294–300.

Xie, D.X., Devos, K.M., Moore, G., and Gale, M.D. (1993). RFLP-based genetic maps of the homoeologous group 5 chromosomes of bread wheat (*Triticum aestivum* L.). Theor. Appl. Genet. **87,** 70–74.

Zhao, X., Si, Y., Hanson, R., Price, H.J., Stelly, D., Wendel, J., and Paterson, A.H. (1998). Dispersed repetitive DNA has spread to new genomes since polyploid formation in cotton. Genome Res. **8,** 479–492.

Zhang, Q., Arbuckle, J., and Wessler, S.R. (2000). Recent, extensive and preferential insertion of members of the MITE family *Heartbreaker (Hbr)* into genic regions of maize. Proc. Natl. Acad. Sci. USA, in press.