Gene-Containing Regions of Wheat and the Other Grass Genomes[1]

Devinder Sandhu[2] and Kulvinder S. Gill*

Department of Agronomy and Horticulture, 362H Plant Science, P.O. Box 830915, University of Nebraska, Lincoln, Nebraska 68583–0915

Deletion line-based high-density physical maps revealed that the wheat (*Triticum aestivum*) genome is partitioned into gene-rich and -poor compartments. Available deletion lines have bracketed the gene-containing regions to about 10% of the genome. Emerging sequence data suggest that these may further be partitioned into "mini" gene-rich and gene-poor regions. An average of about 10% of each gene-rich region seem to contain genes. Sequence analyses in various species suggest that uneven distribution of genes may be a characteristic of all grasses and perhaps all higher organisms. Comparison of the physical maps with genetic linkage maps showed that recombination in wheat and barley (*Hordeum vulgare*) is confined to the gene-containing regions. Number of genes, gene density, and the extent of recombination vary greatly among the gene-rich regions. The gene order, relative region size, and recombination are highly conserved within the tribe Triticeae and moderately conserved within the family. Gene-poor regions are composed of retrotransposon-like non-transcribing repeats and pseudogenes. Direct comparisons of orthologous regions indicated that gene density in wheat is about one-half compared with rice (*Oryza sativa*). Genome size difference between wheat and rice is, therefore, mainly because of amplification of the gene-poor regions. Presence of species-, genera-, and family-specific repeats reveal a repeated invasion of the genomes by different retrotransposons over time. Preferential transposition to adjacent locations and presence of vital genes flanking a gene-rich region may have restricted retrotransposon amplification to gene-poor regions, resulting into tandem blocks of non-transcribing repeats. Insertional inactivation by adjoining retro-elements and selection seem to have played a major role in stabilizing genomes.

The grass family Poaceae includes major crop plants such as wheat (*Triticum aestivum*), barley (*Hordeum vulgare*), oat (*Avena sativa*), rice (*Oryza sativa*), and maize (*Zea mays*). One of its tribes, Triticeae, contains more than 15 genera and 300 species including wheat and barley. Among the important crop plants of the family, rice has the smallest genome (415 Mb), which is only about three times larger than the model plant Arabidopsis (Bennett and Smith, 1976; Arumuganathan and Earle, 1991; Table I). The barley genome is about 12 times larger than that of rice, and maize is about six times. Wheat and barley genomes are of about the same size except that the bread wheat is hexaploid and thus is about three times the size of barley.

The gene size and number in most of the higher plants, especially within the grass family, are expected to be the same. The gene-containing fraction of the Arabidopsis genome is 0.85 (Barakat et al., 1998). Thus, the corresponding fractions for rice, maize, barley, and wheat are expected to be 0.28, 0.05, 0.025, and <0.01, respectively. Therefore, it is imperative to localize and mark the gene-containing regions to understand and efficiently manipulate the crop plant genomes. Second, as shown by the comparisons of the genetic linkage maps, the gene order and synteny are conserved among various Poaceae species (Hart, 1987; Ahn and Tanksley, 1993; Ahn et al., 1993; Devos et al., 1994; Van Deynze et al., 1995a), despite the fact that wheat, rice, and maize diverged more than 50 million years ago (Bennetzen and Freeling, 1993). Based on the genetic linkage map comparisons, the rice genome was divided into conserved blocks, which were proposed to have assembled in different combinations in various Poaceae chromosomes (Moore et al., 1995). The above observations and other phylogenetic and molecular studies strongly suggested that cereal genomes originated from a common ancestor (Clayton and Renvoize, 1986; Kellogg, 1998; Watson and Dallwitz, 1992). Monophyletic origin of the grasses, resulting in as much as 35-fold genome size difference without altering gene number, synteny, and colinearity, is another enigma. In this article, we will focus on the current understanding of the structure, distribution, and evolution of the gene-containing regions of the grass genomes.

GENE DISTRIBUTION IN WHEAT

Studies on the genome organization of wheat suggested that more than 85% of the wheat genes are present in less than 10% of the chromosomal regions (Gill et al., 1996a, 1996b; Sandhu, 2000; Sandhu et al.,

[1] This paper is a contribution of the University of Nebraska Agricultural Research Division, journal series no. 13,450.

[2] Present address: Department of Agronomy, Iowa State University, Ames, IA 50011–1010.

* Corresponding author; e-mail kgill@unl.edu; fax 402–472–7904.

www.plantphysiol.org/cgi/doi/10.1104/pp.010745.

Table I. *DNA content and genome size of different members of family Poaceae*

Plant	Chromosome No.	Genome Size	DNA Content 2C
	n	*Mb*	*pg*
Rice	12	415	0.9[a]
Sorghum (*Sorghum bicolor*)	10	750	1.6[a]
Pennisetum americanum	7	2,410	4.8[b]
Maize	10	2,500	5.6[a]
Barley	7	5,000	10.1[a]
Aegilops tauschii	7	4,000	10.3[c]
Triticum monococcum	7	5,750	11.9[a]
Wheat	21	16,000	33.1[a]

[a] Arumuganathan and Earle (1991). [b] Bennett and Leitch (1995). [c] Furuta et al. (1986).

2001). A typical pattern of gene and recombination distribution on the wheat chromosomes is shown in Figure 1. The gene-containing regions of wheat were identified, marked, and bracketed by physical mapping of the genes and the markers on an array of single-break overlapping deletion lines (Endo and Gill, 1996; Gill et al., 1996a, 1996b; Faris et al., 2000; Sandhu et al., 2001). On each wheat chromosome, there are about six to eight gene-rich regions physically spanning about 10% of the region. The gene-rich regions are interspersed by gene-poor regions that are mainly comprised of retrotransposon-like repetitive sequences (SanMiguel et al., 1996; Barakat et al., 1997; Llaca and Messing, 1998; Feuillet and Keller, 1999). Physical location, structural organization, and gene densities of the gene-rich regions were similar among the three genomes of bread wheat (Gill et al., 1996a, 1996b; Sandhu, 2000; Sandhu et al., 2001). The gene-rich regions vary in the number of genes and the encompassed physical region. For example, most of the wheat homoeologous group 1 genes are present in eight distinct gene-rich regions. These gene-rich regions physically encompass 1% ("1S0.8 region") to 5% ("1L0.7 region") of the group 1 chromosomal region. Among the gene-rich regions, a minimum number of genes was present in the "1L1.0 region," which contained only 3% of the group 1 genes compared with maximum of 32% in the "1S0.8 region" (Fig. 1). In general, gene density in the distal gene-rich regions is higher as compared with the proximal.

Analysis of the gene-containing regions and sequence data generated in different members of grass family suggest that the gene-containing regions are partitioned into the gene-rich and poor compartments (see later sections). The gene-rich regions appear to vary in the number and size of the "mini" gene-rich regions, and the size of the interspersed gene-poor compartments. For example, the "1S0.8 region" seems to have smaller number/proportion of gene-poor compartments as compared with the "1L0.7 region" because of the 12-times difference in gene density (Fig. 1). The gene density among the mini-gene-rich regions may further vary as much as 10-fold (see later sections).

The precision of localizing the gene-rich regions is a function of the number of deletion lines. The actual physical size of the gene-rich regions is probably much smaller because imprecise bracketing due to fewer number of deletion lines will result in overestimate of the spanning region. This is evident from the studies, where the individual maps of chromosomes 1A, 1B, and 1D localized the genes to 50% of the chromosomal region. With a 3-fold increase in the number of deletion line on the consensus map, genes for the same chromosome were localized to 10% of the chromosomal region (Gill et al., 1996a, 1996b; Sandhu, 2000). These results suggest that with availability of additional deletion breakpoints, the gene-rich regions may further be localized to much smaller physical regions than the proposed 10%.

The proposed gene distribution model of wheat is based on a fairly random sample of about 1% to 5% of the total genes (Gill et al., 1996a, 1996b; Sandhu et al., 2001). The genes/markers included in the gene distribution studies were: random clones from 26 different cDNA libraries of seven different Triticeae species, morphological markers, agronomically important genes, and *Pst*I genomic clones. A possible exception is the multicopy gene families, which are probably not well represented in the study because of the technical mapping difficulty. It would be interesting to study the distribution of multicopy gene families in comparison to the single/few copy genes.

GENE-POOR REGIONS

DNA reassociation kinetics studies showed that non-transcribing repeat (NTR)-DNA is an integral part of most plant genomes and its amount is proportional to the genome size (Flavell et al., 1974). Most plant genomes are large and complex and NTR-DNA is primarily composed of retrotransposons (Bennetzen et al., 1998; Shirasu et al., 2000; Wicker et al., 2001). The NTR-DNA of higher plants can be grouped into a few distinct classes based on the sequence comparisons (for review, see Bennetzen, 2000). Thus, the current composition of plant NTR-DNA seems to be a result of multiple invasions by different retrotransposons. Replication of retrotrans-

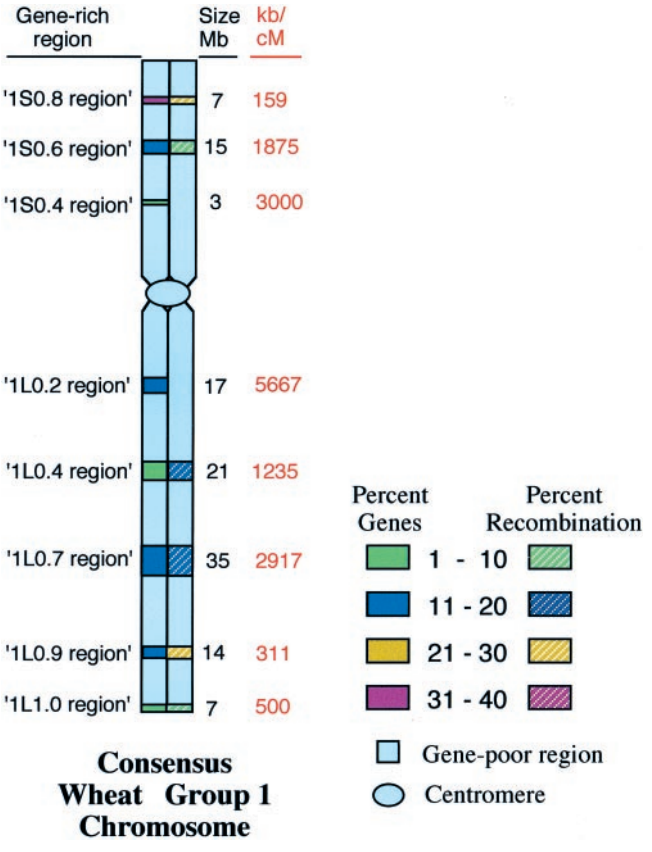

Figure 1. Distribution of genes and recombination on wheat homoeologous group 1 chromosomes. The consensus chromosome and the location and size of the gene-rich regions were drawn to scale based on the average size of the three homoeologous chromosomes 1A, 1B, and 1D. Names of the gene-rich regions are given on the left side of the consensus chromosome. In the nomenclature of the gene-rich regions (e.g. 1S0.8), the first digit represents wheat homoeologous group followed by the short arm (S) or long arm (L) letter. The last two numeral numbers represent fraction length of the gene-rich region location. Actual physical size (black) and the ratio of physical to genetic distance (red) for a region are given on the right-hand side of the consensus chromosome. Sizes (in Mb) of gene-rich regions were calculated based on the cytological measurements (measurements of the region bracketed by the flanking deletion line breakpoints in comparison with the total genome size in microns), and were drawn to scale. Percentage of genes in a gene-rich region with respect to the total group 1 genes was calculated from the overlapping deletion line mapping results for 147 cDNAs, and *Pst*I genome clones from 26 different libraries from seven different species of family Poaceae. Recombination in a chromosomal region was calculated by comparing a deletion line-based physical map with the consensus genetic linkage map of Triticeae (Van Deynze et al., 1995a; Sandhu, 2000).

posons then occurred followed by their inactivation by transpositioning and/or heterochromatinization. The NTR-DNA is unevenly distributed in the plant genomes. Paucity of genes observed from physical maps (Sandhu et al., 2001) and an abundance of heterochromatin visualized as C-bands (Dvorak and Chen, 1984; Curtis and Lukaszewski, 1991; Gill et al., 1991; Jiang et al., 1996) allegorize that repetitive DNA is abundant around the centromeric regions (Copenhaver and Preuss, 1999; Puechberty et al., 1999). The same seems to be true in the smaller plant genomes, as shown by the sequencing information in Arabidopsis and some other organisms (Miller et al., 1998; Presting et al., 1998; Copenhaver et al., 1999).

In addition to the retrotransposons, pseudogenes seem to be an important part of the non-transcribed chromosomal regions (Watterson, 1983; Galili and Feldman, 1984; Zhu et al., 1994; Wendel, 2000). Resistance gene analogs are the best studied example. DNA fragments with structural similarities to the known disease and pest resistance genes have been isolated in many plant species such as soybeans (*Glycine max*), potato (*Solanum tuberosum*), rice, barley, wheat, beans (*Phaseolus vulgaris*), etc. (Kanazin et al., 1996; Leister et al., 1996, 1998; Yu et al., 1996; Feuillet et al., 1997; Zhou et al., 2001). Very few of the resistance gene analogs have been shown to be transcribing (Collins, 1999). Pseudogenes are expected to be particularly abundant in polyploids (for review, see Stephens, 1951; Wendel, 2000), such as bread wheat, where most genes have three structural copies. However, a significant proportion of the wheat genes follow a single-factor Mendalian inheritance (3:1 ratio in F_2), suggesting that only one of the three copies is functional. The other two copies are either nonfunctional or have acquired a different function. An example of such a case was observed in wheat, where the homoeologs of two cDNA clones were observed in the functional centromeric region (Sandhu et al., 2001). Most centromeres are highly heterochromatic and are expected to be inactive for gene expression. In the other two homoeologous chromosomes, these two cDNAs are not present in the centromeric region (Sandhu et al., 2001). Although not confirmed yet, expression of the centromeric copies of the genes is highly unlikely. These observations suggest that perhaps there are genes present in the highly heterochromatic, gene-poor blocks of the genome that are probably not expressing because of the surrounding chromatin structure. A similar observation, at a much finer scale, however, was made in yeast (*Schizosaccharomyces pombe*), where a stretch of approximately 11 kb of DNA was inactive for gene expression as well as recombination (Grewal et al., 1998). Histone deacetylase-mediated chromatin remodeling in the region, however, initiated both gene expression and recombination.

Centromeric regions are not the only places that are abundant in repeated DNA. Regions present between two gene-rich regions are also composed of NTR-DNA (Fig. 1). Some of the regions present even near to the tip of chromosome arms are deficient in genes.

GENE DISTRIBUTION IN TRITICEAE

Gene synteny and colinearity is conserved among Triticeae genomes. The indirect evidence came from Sears (1954), where he showed that loss of any Trit-

iceae chromosome can be at least partially compensated for by one of its homoeologous chromosomes. The term "homoeologous" was coined to represent loss-reparatory, non-homologous chromosomes. The first direct evidence came from the isozyme marker analysis in the early 1980s, where it was shown that the marker location and the relative order were conserved among homoeologs (Hart, 1987). Comparisons of the high-density genetic linkage maps have shown that the gene order and relative recombination are so conserved among Triticeae species that it is possible to construct an accurate consensus map (Van Deynze et al., 1995a). These observations strongly suggest that the structural and functional organization of Triticeae genomes is very similar. Confirmed in barley, the distribution of genes among other Triticeae species, therefore, is similar to that of wheat (Kunzel et al., 2000). Translocation breakpoint-based physical maps clearly partitioned the barley chromosomes into gene-rich and gene-poor compartments (Kunzel et al., 2000). The location and the relative size of gene-rich regions in barley were very similar to that of wheat.

DISTRIBUTION OF GENES IN POACEAE

Now, it is becoming evident that the gene order is conserved among the living organisms and its extent depends upon the evolutionary distance. Comparisons of different genetic maps exemplified this statement and showed that similarity is much greater among the Triticeae genomes as compared with other Poaceae species (Moore et al., 1995). The order of about 62% markers is conserved between rice and maize (Ahn and Tanksley, 1993) compared with 94% between wheat and oat (Van Deynze et al., 1995b). The similar estimates for the wheat and barley comparisons were even higher than 94%. Translocation breakpoints based barley physical maps showed that the gene order, and the extent and physical distribution of recombination were very similar to that of the three wheat genomes (Kunzel et al., 2000). Sequencing information near the *adh1-F* region in maize showed that intergenic regions were mainly composed of retrotransposons inserted within each other (SanMiguel et al., 1998). Wheat and maize diverged about 60 million years ago, but still conserved chromosomal segments are observed. The long arm of maize chromosome 9 and most of wheat chromosome 7 have originated from a common ancestral chromosome (Devos et al., 1994).

Various Poaceae genomes can be aligned by reshuffling the conserved linkage blocks of individual chromosomal segments (Moore et al., 1995). This alignment was primarily based on DNA and morphological markers that were used as anchors. Even in smaller genomes, like rice, gene-containing regions account for about 24% of the genome (Barakat et al., 1997). As observed in wheat, barley, maize, and rice, partitioning of genomes into gene-rich and gene-poor compartments; therefore, it most likely occurs in all Poaceae genomes. Gene distribution has been studied in animal systems and it seems that all animal genomes and chromosomes, to some degree, are divided into gene-rich and gene-poor compartments (Clay and Bernardi, 2001; for review, see Sumner et al., 1993).

RELATIONSHIP BETWEEN DISTRIBUTION OF RECOMBINATION AND GENES

Comparisons of recombination among wheat and rye (*Secale cereale*) C-bands revealed that recombination is uneven along the Triticeae chromosomes (Dvorak and Chen, 1984; Curtis and Lukaszewski, 1991; Stein et al., 2000). Similar observations have been made in other plant and animal systems (Rick, 1971; Bollag et al., 1989; Ganal et al., 1989). Comparisons of deletion line-based physical maps with the genetic linkage maps elegantly showed that most of recombination occurs in the gene-rich regions of the wheat genome (Gill et al., 1996a, 1996b; Sandhu et al., 2001). Similar comparison in barley using translocation breakpoint-based physical maps confirmed its similarity in recombination distribution to that of wheat (Kunzel et al., 2000). In wheat, comparisons of the low-density wheat maps suggested that the recombination only occurs in the gene-containing regions (Werner et al., 1992; Gill et al., 1993; Kota et al., 1993; Delaney et al., 1995a, 1995b; Mickelson-Young et al., 1995). The high-density map comparisons confirmed that essentially no recombination occurs in the gene-poor regions (Gill et al., 1996a, 1996b; Weng et al., 2000; Sandhu et al., 2001). However, it became apparent that gene-rich regions differ in the extent of recombination (Sandhu, 2000). The gene-rich region "1S0.8" has 30-fold higher recombination as compared with the "1S0.4 region" (Fig. 1). Centromeres are known to suppress recombination in the vicinity in most of the higher eukaryotes (Puechberty et al., 1999). Essentially no recombination is observed in the proximal 30% of the wheat chromosomes, despite the presence of the gene-rich regions (Fig. 1).

Besides the centromere, other factors also seem to affect the extent of recombination. A proximal gene-rich region of wheat chromosome 6 ("6L0.4 region") has about 8-fold higher recombination than a distal gene-rich region ("6L0.7 region") of the same chromosome (D. Sandhu and K.S. Gill, unpublished data). The extent of recombination varies greatly even within the same gene-rich region. In barley, segments within a 1-Mb gene-rich region may vary as much as 10-fold for the extent of recombination (Wei et al., 1999).

STRUCTURE OF THE GENE-RICH REGIONS OF POACEAE

Despite the conservation of gene synteny and colinearity, Poaceae genome size may vary as much as 35 times among species. Given that all the Poaceae genomes seem to be partitioned into gene-rich and

-poor compartments, it is imperative to understand if the amplification is uniform within genomes or confined only to the gene-poor regions. The first line of evidence proposing a non-proportional amplification of the gene-poor regions of the larger genomes comes from the size estimates of the gene-rich regions. In the case of proportional amplification, the gene-containing fraction is expected to be the same in all Poaceae species. The best estimates for the gene-containing regions of wheat, barley, maize, and rice are 7%, 12%, 17%, and 24%, respectively (Carels et al., 1995; Barakat et al., 1997; Sandhu et al., 2001). Further, the gene density of the gene-containing regions should differ, proportional to the genome size. In case of uniform amplification, the gene density of wheat should be 35 times less than that of rice. However, recent studies suggest that the average gene density within the Triticeae gene-rich regions is 10 to 20 genes per 100 kb, compared with 15 to 25 in rice.

Gene density in gene-rich regions in grasses is comparable with the average gene density in Arabidopsis, which is one gene per 4 to 5 kb (Quigley et al., 1996). The higher gene density example is in *Lrk10* region of wheat, where an average distance between genes is 4.6 kb (Feuillet and Keller, 1999; Feuillet et al., 2001). The lower gene density example is for a gene-rich region around the *Mlo* locus of barley, where the same estimate is 20 kb (Panstruga et al., 1998). In this study, it is interesting to note that all three genes were present within 25 kb of the total 60-kb contig. In a 16-kb region containing the starch-branching enzyme I of *Ae. tauschii*, the gene density was 19 genes per 100 kb (Rahman et al., 1997). A 340-kb rice region around *Adh1-Adh2* contains 33 genes, with an average of one gene per 10.3 kb (Tarchini et al., 2000).

Direct comparisons of the similar regions between larger and smaller genome species suggest that the gene-containing regions are about two times larger irrespective of the difference in the genome size. Difference in genome size between maize and sorghum is 3.3 times compared with 35 times between wheat and rice. However, the difference in gene density is about the same. The distances among genes in the receptor-like kinase gene cluster in wheat were only 2 to 3 times more than the homoeologous region in rice (Feuillet and Keller, 1999). The sorghum *Adh* region of 78.2 kb corresponded to 225 kb in maize (Tikhonov et al., 1999). The distance between *Adh1* and *u22* was 2.4 times higher in maize as compared with sorghum (SanMiguel et al., 1996). These observations suggest that the amplification of a genome in gene-containing regions is much less than in gene-poor regions. Occasionally, localized amplification may occur in a region because of retrotransposon invasion irrespective of the species or the size difference. The genes *Sh2* and *A1* are 21 kb apart in rice, 22 kb in sorghum, and 140 kb in maize (Civardi et al., 1994; Bennetzen et al., 1998). For most other regions compared so far, average amplification in maize was only about 2 to 3 times that of sorghum and rice.

At a lower resolution, genes in cereal genomes appear to be clustered in small chromosomal regions separated by large blocks of repeat retrotransposon-like sequences (SanMiguel et al., 1996; Barakat et al., 1997; Llaca and Messing, 1998; Feuillet and Keller, 1999). At a higher resolution, gene-rich regions appear to be further consisted of mini-gene-rich regions interspersed by NTRs (Fig. 2). At sequence level, genes present in mini-gene-rich regions are further separated by intergenic NTR sequences consisting of retrotransposons (Fig. 2). Currently available data suggest that size and distribution of the mini-gene-rich region or the intergenic region do not follow any pattern.

EVOLUTION OF GENE-CONTAINING REGIONS IN POACEAE

It is now clear that retrotransposon invasion is the main factor contributing to the differential amplification of the Poaceae and other genomes. Retrotransposon insertions have doubled the size of the maize genome in the last 3 million years (SanMiguel et al., 1998). Presence of species- (Kumar et al., 1990; Pestova et al., 1998; Grutzner et al., 1999; Linares et al., 1999), genera- (SanMiguel et al., 1998; Schmidt et al., 1998; Staginnus et al., 1999; Pearce et al., 2000; Shi and Endo, 2000), and family- (Zhang et al., 1995) specific repeats strongly suggest a repeated invasion by different retrotransposons over time. Differences

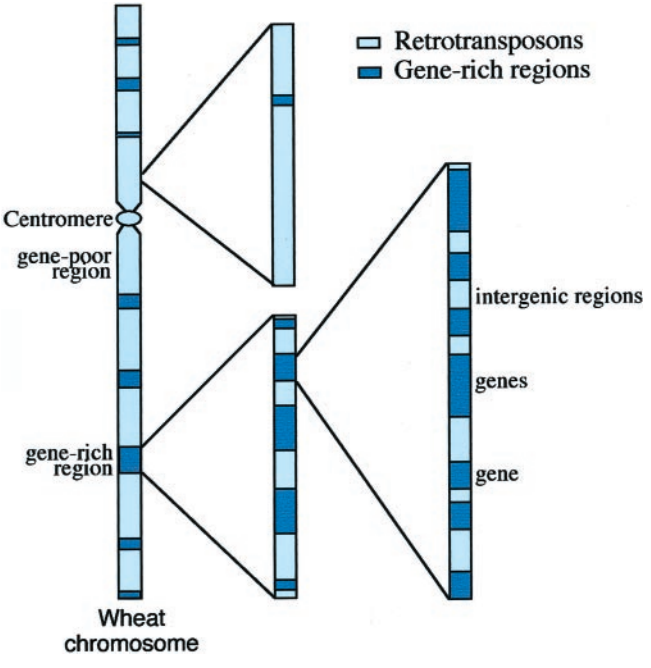

Figure 2. A wheat chromosome representing genome organization and gene distribution in cereal genomes. A gene-rich region and a gene-poor region are further amplified to show structural organization within these regions.

in the insertion dates for retrotransposons near the *Adh1* locus were observed between maize and sorghum (SanMiguel et al., 1998). Five repeat sequences were found common between maize and wheat, of which only one was common between oat and maize (Zhang et al., 1995). The repeats common between wheat and maize were probably in the progenitor genome before divergence. The same argument can be extended to tribe- and species-specific repeats (Fig. 3). In addition to the new invasions, the proportion of the existing repeats fluctuates and perhaps some repeats get eliminated. Absence of four repeats, common between maize and wheat, in oat supports this hypothesis (Zhang et al., 1995).

Upon invasion, two processes will restrict retrotransposon amplification to adjacent gene-poor regions, resulting into tandem blocks of NTRs. First, retrotransposons preferentially transpose to the nearby chromosomal location (Van Schaik and Brink, 1959; Greenblatt and Brink, 1962). Second, a gene-rich region containing genes flanked by genes important for fertility and viability will be resistant to retrotransposon invasion. However, transpositions in the intergenic regions of gene clusters will also be tolerated. Retrotransposon-like repeat sequences have been observed in almost all gene-containing regions sequenced, although their number and distribution vary (SanMiguel et al., 1996; Bennetzen et al., 1998; Llaca and Messing, 1998; Panstruga et al., 1998; Feuillet and Keller, 1999; Tarchini et al., 2000; Dubcovsky et al., 2001; Feuillet et al., 2001). The number of repeats in the intergenic regions should be a good indicator of the importance of the surrounding genes for the viability of the plant and its relative location within the gene-rich region. Sequence information around zein gene cluster in maize (Llaca and Messing, 1998) and *Mlo* locus of barley (Panstruga et al., 1998) suggested that very few retrotransposons were found in these regions. Because of the least selection pressure, amplification of the gene-poor regions will be most favored, resulting in the blocks of NTRs separating the gene-containing regions. Junctions of the gene-rich and -poor blocks are expected to have more intergenic repeats and thus lower gene density.

For the survival of any species, large-scale amplification of retro-elements needs to be stopped at some point. Insertional inactivation by adjoining retro-elements seems to have played a major role in restricting retro-element proliferation. The maize region containing *Adh1-F* and *u22* genes was mainly composed of nested clusters of retro-elements inserted within each other (SanMiguel et al., 1996). In the gene-poor regions, heterochromatinization because of long stretches of inactive DNA will silence any active retro-element or gene. Open reading frames have been observed in the centromeric regions of wheat, humans, and Arabidopsis, which are known to be transcriptionally inactive and highly hetrochromatic (Copenhaver et al., 1999; Puechberty et al., 1999; Sandhu et al., 2001).

The reason for the differential amplification of the genomes within the same family is not known; however, it seems to depend upon the type of the invading retro-element. For example, the long terminal repeat type of retrotransposons seems to transpose preferentially into the gene-poor region, whereas the miniature inverted-repeat transposable elements type prefer gene-rich regions (for review, see Bennetzen, 2000). Because of the selection pressure, the extent of amplification for the retro-elements preferring gene-poor regions would be higher than the elements preferring gene-rich regions. Majority of the NTRs in larger genomes are composed of elements preferring gene-poor regions for transposition.

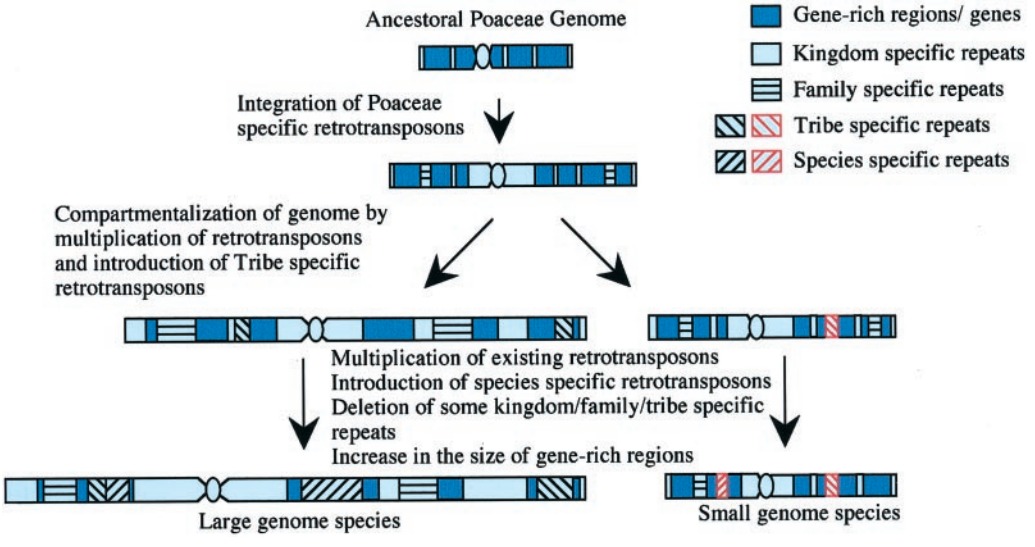

Figure 3. A hypothetical model for evolution of gene-containing regions in cereal genomes.

Received August 16, 2001; returned for revision October 4, 2001; accepted November 21, 2001.

LITERATURE CITED

Ahn SN, Anderson JA, Sorrells ME, Tanksley SD (1993) Homoeologous relationships of rice, wheat and maize. Mol Gen Genet **241:** 483–490

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

Arumuganathan K, Earle ED (1991) Estimation of nuclear DNA content of plants by flow cytometry. Plant Mol Biol Rep **9:** 229–233

Barakat A, Carels N, Bernardi G (1997) The distribution of genes in the genomes of Graminae. Proc Natl Acad Sci USA **94:** 6857–6861

Barakat A, Matassi G, Bernardi G (1998) Distribution of genes in the genome of *Arabidopsis thaliana* and its implications for the genome organization of plants. Proc Natl Acad Sci USA **95:** 10044–10049

Bennett MD, Leitch IJ (1995) Nuclear DNA amounts in angiosperms. Ann Bot **76:** 113–176

Bennett MD, Smith JB (1976) Nuclear DNA amounts in angiosperms. Philos Trans R Soc London Ser B **274:** 227–274

Bennetzen JL (2000) Transposable element contributions to plant gene and genome evolution. Plant Mol Biol **42:** 251–269

Bennetzen JL, Freeling M (1993) Grasses as a single genetic system. Trends Genet **9:** 259–261

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

Bollag RJ, Waldman AS, Liskay RM (1989) Homologous recombination in mammalian cells. Annu Rev Genet **23:** 199–225

Carels N, Barakat A, Bernardi G (1995) The gene distribution of the maize genome. Proc Natl Acad Sci USA **1995:** 11057–11060

Civardi L, Xia Y, Edwards KJ, Schnable PS, Nikolau BJ (1994) The relationship between genetic and physical distances in the cloned a1-sh2 interval of the *Zea mays* L. genome. Proc Natl Acad Sci USA **91:** 8268–8272

Clay O, Bernardi G (2001) The isochores in human chromosomes 21 and 22. Biochem Biophys Res Commun **285:** 855–856

Clayton WD, Renvoize SA (1986) Genera Graminum: Grasses of the World. Kew Bulletin Additional Series XIII, Royal Botanical Gardens, Kew, Her Majesty's Stationery Office, London

Collins FS (1999) Microarrays and macrosequences. Nat Genet **21:** 2

Copenhaver GP, Nickel K, Kuromori T, Benito MI, Kaul S, Lin X, Bevan M, Murphy G, Harris B, Parnell LD et al. (1999) Genetic definition and sequence analysis of Arabidopsis centromeres. Science **286:** 2468–2474

Copenhaver GP, Preuss D (1999) Centromeres in the genomic era: unraveling paradoxes. Curr Opin Plant Biol **2:** 104–108

Curtis CA, Lukaszewski AJ (1991) Genetic linkage between C-bands and storage protein genes in chromosome 1B of tetraploid wheat. Theor Appl Genet **81:** 245–252

Delaney D, Nasuda S, Endo TR, Gill BS, Hulbert SH (1995a) Cytologically based physical maps of the group-2 chromosomes of wheat. Theor Appl Genet **91:** 568–573

Delaney D, Nasuda S, Endo TR, Gill BS, Hulbert SH (1995b) Cytologically based physical maps of the group-3 chromosomes of wheat. Theor Appl Genet **91:** 780–782

Devos KM, Chao S, Li QY, Simonetti MC, Gale MD (1994) Relationship between chromosome 9 of maize and wheat homeologous group 7 chromosomes. Genetics **138:** 1287–1292

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

Dvorak J, Chen KC (1984) Phylogenetic relationships between chromosomes of wheat and chromosome 2E of *Elytrigia elongata*. Can J Genet Cytol **26:** 128–132

Endo TR, Gill BS (1996) The deletion stocks of common wheat. J Hered **87:** 295–307

Faris JD, Haen KM, Gill BS (2000) Saturation mapping of a gene-rich recombination hot spot region in wheat. Genetics **154:** 823–835

Feuillet C, Keller B (1999) High gene density is conserved at syntenic loci of small and large grass genomes. Proc Natl Acad Sci USA **96:** 8265–8270

Feuillet C, Penger A, Gellner K, Mast A, Keller B (2001) Molecular evolution of receptor-like kinase genes in hexaploid wheat: independent evolution of orthologs after polyploidization and mechanisms of local rearrangements at paralogous loci. Plant Physiol **125:** 1304–1313

Feuillet C, Schachermayr G, Keller B (1997) Molecular cloning of a new receptor-like kinase gene encoded at the Lr10 disease resistance locus of wheat. Plant J **11:** 45–52

Flavell RB, Bennett MD, Smith JB, Smith DB (1974) Genome size and the proportion of repeated nucleotide sequence DNA in plants. Biochem Genet **12:** 257–269

Furuta Y, Nishikawa K, Yamaguchi S (1986) Nuclear DNA content of in diploid wheat and its relatives in relation to the phylogeny of tetraploid wheat. Japan J Genet **61:** 97–105

Galili G, Feldman M (1984) Intergenomic suppression of endosperm protein genes in common wheat. Can J Genet Cytol **26:** 651–656

Ganal MW, Young ND, Tanksley SD (1989) Pulsed field gel electrophoresis and physical mapping of large DNA fragments in the *Tm-2a* region of chromosome 9 in tomato. Mol Gen Genet **215:** 395–400

Gill BS, Friebe B, Endo TR (1991) Standard karyotype and nomenclature system for description of chromosome bands and structural aberrations in wheat (*Triticum aestivum*). Genome **34:** 830–839

Gill KS, Gill BS, Endo TR (1993) A chromosome region-specific mapping strategy reveals gene rich telomeric ends in wheat. Chromosoma **102:** 374–381

Gill KS, Gill BS, Endo TR, Boyko EV (1996a) Identification and high-density mapping of gene-rich regions in chromosome group *5* of wheat. Genetics **143:** 1001–1012

Gill KS, Gill BS, Endo TR, Taylor T (1996b) Identification and high-density mapping of gene-rich regions in chromosome group *1* of wheat. Genetics **144:** 1883–1891

Greenblatt IM, Brink RA (1962) Twin mutations in medium variegated pericarp maize. Genetics **47:** 489–501

Grewal SI, Bonaduce MJ, Klar AJ (1998) Histone deacetylase homologs regulate epigenetic inheritance of transcriptional silencing and chromosome segregation in fission yeast. Genetics **150:** 563–576

Grutzner F, Lutjens G, Rovira C, Barnes DW, Ropers HH, Haaf T (1999) Classical and molecular cytogenetics of the pufferfish *Tetraodon nigroviridis*. Chromosome Res **7:** 655–662

Hart GE (1987) Genetic and biochemical studies of enzymes. *In* EG Heyne, ed, Wheat and Wheat Improvement, Ed 2, Vol 13. American Society of Agronomy, Madison, WI, pp 199–214

Jiang J, Nasuda S, Dong F, Scherrer CW, Woo SS, Wing RA, Gill BS, Ward DC (1996) A conserved repetitive DNA element located in the centromeres of cereal chromosomes. Proc Natl Acad Sci USA **93:** 14210–14213

Kanazin V, Marek LF, Shoemaker RC (1996) Resistance gene analogs are conserved and clustered in soybean. Proc Natl Acad Sci USA **93:** 11746–11750

Kellogg EA (1998) Relationships of cereal crops and other grasses. Proc Natl Acad Sci USA **97:** 9121–9126

Kota RS, Gill KS, Gill BS, Endo TR (1993) A cytogenetically based physical map of chromosome 1B in common wheat. Genome **36:** 548–554

Kumar LS, Gupta VS, Ranjekar PK (1990) Identification and partial characterization of two species-specific repeat families in the great millet (*Sorghum vulgare*, Poaceae) genome. Plant Syst Evol **171:** 249–257

Kunzel G, Korzum L, Meister A (2000) Cytologically integrated physical restriction fragement length polymorphism maps for the barley genome based on translocation breakpoints. Genetics **154:** 397–412

Leister D, Ballvora A, Salamini F, Gebhardt C (1996) A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. Nat Genet **14:** 421–429

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

Linares C, Serna A, Fominaya A (1999) Chromosomal organization of a sequence related to LTR-like elements of Ty1-copia retrotransposons in *Avena* species. Genome **42:** 706–713

Llaca V, Messing J (1998) Amplicons of maize zein genes are conserved within genic but expanded and constricted in intergenic regions. Plant J **15:** 211–220

Mickelson-Young L, Endo TR, Gill BS (1995) A cytogenetic ladder map of the wheat homoeologous group-4 chromosomes. Theor Appl Genet **90:** 1007–1011

Miller JT, Dong F, Jackson SA, Song J, Jiang J (1998) Retrotransposon-related DNA sequences in the centromeres of grass chromosomes. Genetics **150:** 1615–1623

Moore G, Devos KM, Wang Z, Gale MD (1995) Cereal genome evolution: grasses, line up and form a circle. Curr Biol **5:** 737–739

Panstruga R, Buschges R, Piffanelli P, Schulze-Lefert P (1998) A contiguous 60 kb genomic stretch from barley reveals molecular evidence for gene islands in a monocot genome. Nucleic Acids Res **26:** 1056–1062

Pearce SR, Knox M, Ellis TH, Flavell AJ, Kumar A (2000) Pea Ty1-copia group retrotransposons: transpositional activity and use as markers to study genetic diversity in *Pisum*. Mol Gen Genet **263:** 898–907

Pestova EG, Goncharov NP, Salina EA (1998) Elimination of a tandem repeat of telomeric heterochromatin during the evolution of wheat. Theor Appl Genet **97:** 1380–1386

Presting GG, Malysheva L, Fuchs J, Schubert I (1998) A Ty3/gypsy retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. Plant J **16:** 721–728

Puechberty J, Laurent AM, Gimenez S, Billault A, Laurent MEB, Calenda A, Marcais B, Prades C, Loannou P, Yurov Y et al. (1999) Genetic and physical analyses of the centromeric and pericentromeric regions of human chromosome 5: recombination across 5cen. Genomics **56:** 274–287

Quigley F, Dao P, Cottet A, Mache R (1996) Sequence analysis of an 81 kb contig from *Arabidopsis thaliana* chromosome III. Nucleic Acids Res **24:** 4313–4318

Rahman S, Abrahams S, Abbott D, Mukai Y, Samuel M, Morell M, Appels R (1997) A complex arrangement of genes at a starch branching enzyme I locus in the D-genome donor of wheat. Genome **40:** 465–474

Rick CM (1971) Some cytological features of the genome in diploid plant species. Stadler Genet Symp **1,2**: 153–174

Sandhu D (2000) Molecular characterization of a major gene-rich region in wheat (*Triticum aestivum* L.). PhD thesis. University of Nebraska, Lincoln

Sandhu D, Champoux JA, Bondareva SN, Gill KS (2001) Identification and physical localization of useful genes and markers to a major gene-rich region on wheat group *1S* chromosomes. Genetics **157:** 1735–1747

SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL (1998) The paleontology of intergene retrotransposons of maize. Nat Genet **20:** 43–45

SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z et al. (1996) Nested retrotransposons in the intergenic regions of the maize genome. Science **274:** 765–768

Schmidt T, Kubis S, Katsiotis A, Jung C, Heslop Harrison JS (1998) Molecular and chromosomal organization of two repetitive DNA sequences with intercalary locations in sugar beet and other *Beta* species. Theor Appl Genet **97:** 696–704

Sears ER (1954) The aneuploids of common wheat. Research Bulletin 572. University of Missouri Agricultural Experiment Station, Columbia, pp 1–58

Shi F, Endo TR (2000) Genetic induction of chromosomal rearrangements in barley chromosome 7H added to common wheat. Chromosoma **109:** 358–363

Shirasu K, Schulman AH, Lahaye T, Schulze-Lefert P (2000) A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. Genome Res **10:** 908–915

Staginnus C, Winter P, Desel C, Schmidt T, Kahl G (1999) Molecular structure and chromosomal localization of major repetitive DNA families in the chickpea (*Cicer arietinum* L.) genome. Plant Mol Biol **39:** 1037–1050

Stein N, Feuillet C, Wicker T, Schlagenhauf E, Keller B (2000) Subgenome chromosome walking in wheat: A 450-kb physical contig in *Triticum monococcum* L. spans the *Lr10* resistance locus in hexaploid wheat (*Triticum aestivum* L.). Proc Natl Acad Sci USA **97:** 13436–14441

Stephens SG (1951) Possible significance of duplication in evolution. Adv Genet **4:** 247–265

Sumner AT, Torre JDL, Stuppia L (1993) The distribution of genes on chromosomes: a cytological approach. J Mol Evol **37:** 117–122

Tarchini R, Biddle P, Wineland R, Tingey S, Rafalski A (2000) The complete sequence of 340 kb of DNA around the rice Adh1-Adh2 region reveals interrupted colinearity with maize chromosome 4. Plant Cell **12:** 381–391

Tikhonov AP, SanMiguel PJ, Nakajima Y, Gorenstein NM, Bennetzen JL, Avramova Z (1999) Colinearity and its exceptions in orthologous adh regions of maize and sorghum. Proc Natl Acad Sci USA **96:** 7409–7414

Van Deynze AE, Dubcovsky J, Gill KS, Nelson JC, Sorrells ME, Dvorak J, Gill BS, Lagudah ES, McCouch SR, Appels R (1995a) Molecular-genetic maps for group 1 chromosomes of Triticeae species and their relation to chromosomes in rice and oat. Genome **38:** 45–59

Van Deynze AE, Nelson JC, Yglesias ES, Harrington SE, Braga DP, McCouch SR, Sorrells ME (1995b) Comparative mapping in grasses. Wheat relationships. Mol Gen Genet **248:** 744–754

Van Schaik NW, Brink RA (1959) Transpositions of modulator, a component of the variegated pericarp allele in maize. Genetics **44:** 725–738

Watson L, Dallwitz MJ (1992) The grass genera of the world. C.A.B. International, Wallingford, Oxon, UK

Watterson GA (1983) On the time for gene silencing at duplicate loci. Genetics **105:** 745–766

Wei F, Gobelman-Werner K, Morroll SM, Kurth J, Mao L, Wing RA, Leister D, Schulze-Lefert P, Wise RP (1999) The Mla (powdery mildew) resistance cluster is associated with three NBS-LRR gene families and suppressed recombination within a 240-kb DNA interval on chromosome 5S (1HS) of barley. Genetics **153:** 1929–1948

Wendel JF (2000) Genome evolution in polyploids. Plant Mol Biol **42:** 225–249

Weng Y, Tuleen NA, Hart GE (2000) Extended physical maps and a consensus physical map of the homoeologous group-*6* chromosomes of wheat (*Triticum aestivum* L. em Thell.). Theor Appl Genet **100:** 519–527

Werner JE, Endo TR, Gill BS (1992) Toward a cytogenetically based physical map of the wheat genome. Proc Natl Acad Sci USA **89:** 11307–11311

Wicker T, Stein N, Albar L, Feuillet C, Schlagenhauf E, Keller B (2001) Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. Plant J **26:** 307–316

Yu YG, Buss GR, Maroof MA (1996) Isolation of a superfamily of candidate disease-resistance genes in soybean based on a conserved nucleotide-binding site. Proc Natl Acad Sci USA **93:** 11751–11756

Zhang J, Friebe B, Gill BS (1995) Detection of maize DNA sequences amplified in wheat. Genome **38:** 946–950

Zhou F, Kurth J, Wei F, Elliott C, Vale G, Yahiaoui N, Keller B, Somerville S, Wise R, Schulze-Lefert P (2001) Cell-autonomous expression of barley *Mla1* confers race-specific resistance to the powdery mildew fungus *via* a *Rar1*-independent signaling pathway. Plant Cell **13:** 337–350

Zhu T, Schupp JM, Oliphant A, Keim P (1994) Hypomethylated sequences: characterization of the duplicate soybean genome. Mol Gen Genet **244:** 638–645

