

# The Organization and Rate of Evolution of Wheat Genomes Are Correlated With Recombination Rates Along Chromosome Arms

Eduard D. Akhunov,<sup>1</sup> Andrew W. Goodyear,<sup>1</sup> Shu Geng,<sup>1</sup> Li-Li Qi,<sup>2</sup> Benjamin Echaliér,<sup>2</sup> Bikram S. Gill,<sup>2</sup> Miftahudin,<sup>3</sup> J. Perry Gustafson,<sup>4</sup> Gerard Lazo,<sup>5</sup> Shiaoman Chao,<sup>5</sup> Olin D. Anderson,<sup>5</sup> Anna M. Linkiewicz,<sup>1</sup> Jorge Dubcovsky,<sup>1</sup> Mauricio La Rota,<sup>6</sup> Mark E. Sorrells,<sup>6</sup> Deshui Zhang,<sup>7</sup> Henry T. Nguyen,<sup>3</sup> Venugopal Kalavacharla,<sup>8</sup> Khwaja Hossain,<sup>8</sup> Shahryar F. Kianian,<sup>8</sup> Junhua Peng,<sup>9</sup> Nora L.V. Lapitan,<sup>9</sup> Jose L. Gonzalez-Hernandez,<sup>10</sup> James A. Anderson,<sup>10</sup> Dong-Woog Choi,<sup>11</sup> Timothy J. Close,<sup>11</sup> Muharrem Dilbirligi,<sup>12</sup> Kulvinder S. Gill,<sup>12</sup> M. Kay Walker-Simmons,<sup>13</sup> Camille Steber,<sup>14</sup> Patrick E. McGuire,<sup>15</sup> Calvin O. Qualset,<sup>15</sup> and Jan Dvorak<sup>1,16</sup>

<sup>1</sup>Department of Agronomy and Range Science, University of California, Davis, California 95616, USA; <sup>2</sup>Department of Plant Pathology, Throckmorton Hall, Kansas State University, Manhattan, Kansas 66506, USA; <sup>3</sup>Department of Agronomy, University of Missouri, Columbia, Missouri 65211, USA; <sup>4</sup>USDA-ARS, University of Missouri, Columbia, Missouri 65211, USA; <sup>5</sup>USDA/ARS Western Regional Research Center, Albany, California 94710, USA; <sup>6</sup>Department of Plant Breeding, Cornell University, Ithaca, New York 14853, USA; <sup>7</sup>Department of Plant and Soil Science, Texas Tech University, Lubbock, Texas 79409, USA; <sup>8</sup>Department of Plant Sciences, Loftsgard Hall, North Dakota State University, Fargo, North Dakota 58105–5051, USA; <sup>9</sup>Department of Soil and Crop Sciences, Colorado State University, Fort Collins, Colorado 80524, USA; <sup>10</sup>Department of Agronomy and Plant Genetics, University of Minnesota, Twin Cities, St. Paul, Minnesota 55108–6026, USA; <sup>11</sup>Department of Botany and Plant Science, University of California, Riverside, California 92521, USA; <sup>12</sup>Department of Crop and Soil Sciences, Washington State University, Pullman, Washington 99164–6420, USA; <sup>13</sup>USDA-ARS National Program Staff, Beltsville, Maryland 20705–5139, USA; <sup>14</sup>USDA-ARS, Department of Crop and Soil Sciences, Washington State University, Pullman, Washington 99164–6420, USA; <sup>15</sup>Genetic Resources Conservation Program, University of California, Davis, California 95616, USA.

Genes detected by wheat expressed sequence tags (ESTs) were mapped into chromosome bins delineated by breakpoints of 159 overlapping deletions. These data were used to assess the organizational and evolutionary aspects of wheat genomes. Relative gene density and recombination rate increased with the relative distance of a bin from the centromere. Single-gene loci present once in the wheat genomes were found predominantly in the proximal, low-recombination regions, while multigene loci tended to be more frequent in distal, high-recombination regions. One-quarter of all gene motifs within wheat genomes were represented by two or more duplicated loci (paralogous sets). For 40 such sets, ancestral loci and loci derived from them by duplication were identified. Loci derived by duplication were most frequently located in distal, high-recombination chromosome regions whereas ancestral loci were most frequently located proximal to them. It is suggested that recombination has played a central role in the evolution of wheat genome structure and that gradients of recombination rates along chromosome arms promote more rapid rates of genome evolution in distal, high-recombination regions than in proximal, low-recombination regions.

The haploid chromosome complement of hexaploid wheat, *Triticum aestivum* L. ( $2n = 6x = 42$ ), consists of three genomes, designated A, B, and D, each including seven metacentric or submetacentric chromosomes. The genomes were contrib-

uted by three diploid species: *T. urartu* Thun. (the source of the A genome), *Aegilops speltoides* (Tausch) Gren. or a closely related species (the source of the B genome), and *Ae. tauschii* Coss. (the source of the D genome). Cytogenetic studies and comparative mapping indicated that, except for chromosomes 4A, 5A, and 7B, the *T. aestivum* homoeologous chromosomes are to a large extent homosequential and preserve the chromosome organization of wheat ancestors (Naranjo et

<sup>16</sup>Corresponding author.

E-MAIL [jdvorak@ucdavis.edu](mailto:jdvorak@ucdavis.edu); FAX (530) 752-4361.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.808603>. Article published online before print in April 2003.

al. 1987; Gale et al. 1993). Inferences on *T. aestivum* genome organization and evolution are therefore directly applicable to diploid relatives of wheat.

Comparisons of wheat linkage and deletion maps suggested that wheat chromosomes are subdivided into regions differing in relative gene density (Gill et al. 1993; Hohmann et al. 1994; Delaney et al. 1995a,b; Mickelson-Young et al. 1995; Gill et al. 1996a,b; Sandhu and Gill 2002). This is consistent with other lines of evidence suggesting that the chromosomes of most plants and animals are subdivided into gene-rich and gene-poor isochors (Sabeur et al. 1993; Bernardi 1996; Barakat et al. 1997; Bernardi et al. 1997; Carels et al. 1998; Bernardi 1999).

Recombination rates along wheat chromosomes are also not uniform. Recombination in the wheat genomes occurs preferentially in distal chromosome regions (Dvorak and Chen 1984; Kota et al. 1993; Lukaszewski and Curtis 1993; Hohmann et al. 1994; Delaney et al. 1995a,b; Mickelson-Young et al. 1995; Gill et al. 1996b; Dvorak et al. 1998).

These observations provoke the following questions: Is there some underlying order in the position of genes relative to recombination rates that has evolutionary significance for the organization of genomes? Conventional biological wisdom teaches that the function of biological organization often becomes apparent only if examined in evolutionary contexts. If there were an underlying order in the wheat genomes, what would be the causes and effects between evolutionary processes and this order? Would these evolutionary processes have homogeneous effects along chromosomes or would they cause chromosomes to differentiate (lose synteny) unevenly along their lengths?

To investigate these and other questions critical for the understanding of the organization and evolution of wheat genomes, a large population of wheat expressed sequence tags (ESTs) was developed, contigs were constructed, and genes detected by a representative cDNA clone from each contig were mapped. To map thousands of gene loci, a collaborative mapping strategy was developed that exploited the existence of numerous deletion lines in the nearly isogenic background of 'Chinese Spring' wheat (Endo and Gill 1996). A total of 159 homozygous terminal deletions were selected from this pool so that the breakpoints provided extensive coverage of the 21 wheat chromosomes. The positions of the breakpoints of the 159 deletions can be seen at <http://wheat.pw.usda.gov/west/binmaps>. The ten collaborating laboratories shared this set of deletions for the mapping of loci detected by Southern hybridization of cDNA clones utilizing digestion of genomic DNAs with the *EcoR* I restriction endonuclease.

By deletion mapping, loci were placed into bins delineated by the breakpoints of neighboring deletions (<http://wheat.pw.usda.gov/west/binmaps> and [http://wheat.pw.usda.gov/cgi-bin/westsql/map\\_locus.cgi](http://wheat.pw.usda.gov/cgi-bin/westsql/map_locus.cgi)). The only prerequisite for this technique to be fully successful was that restriction fragments hybridizing with a probe were of different lengths at each locus detected by a probe. Since this requirement was generally satisfied, it was often possible to map all or nearly all loci detected by a probe. This strategy therefore allowed for a routine mapping of all genes in a genome corresponding to a specific gene motif.

A database that was constructed is briefly described in Methods and will be fully described elsewhere. Data in the database were used to assess the level of gene redundancy, variation in relative gene density along chromosome arms, distribution of specific classes of genes along chromosome

arms, and differentiation between homoeologous chromosomes along chromosome arms. Additionally, recombination rates per bin were derived from published wheat linkage maps and bin sizes. Finally, these findings were used to investigate relationships between recombination rate and the various parameters of gene distribution along chromosome arms.

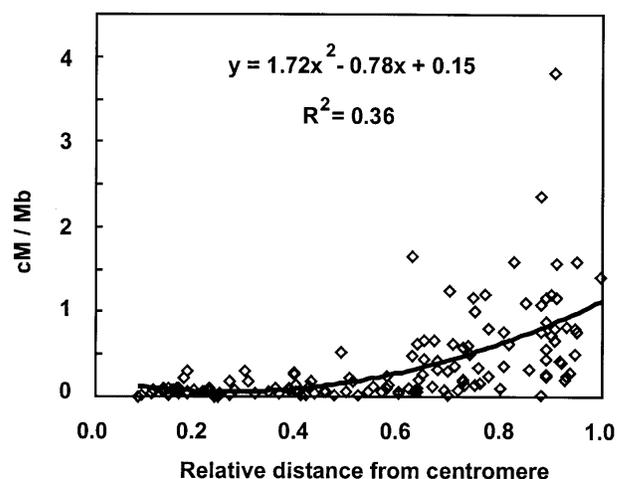
## RESULTS

### Recombination Rates

Recombination rates within bins were expressed as coefficients of exchange (CE), which is cM/Mb. Recombination rates increased as the relative distance of a bin from the centromere increased (Fig. 1; Table 1). Coefficient of exchange within a bin was positively correlated with the position of the center of that bin on the centromere-telomere axis (Spearman rank correlation coefficient  $r_s = 0.72$ ,  $P < 0.0001$ ). This relationship fit a squaring function ( $R^2 = 0.36$ ; Fig. 1), meaning recombination increased with approximately the square of the relative distance of a bin from the centromere. Most crossovers were physically located in the distal one-third of chromosome arms (Fig. 1; Table 1). The relationship was slightly stronger for the short arms ( $r_s = 0.78$ ,  $P < 0.0001$ ) than for the long arms ( $r_s = 0.71$ ,  $P < 0.0001$ ) but the difference between the two coefficients was not significant ( $P = 0.36$ ).

### Relative Gene Density

A total of 730 ESTs, each from a different contig, were used in this and several other analyses. Either one or several restriction fragments per gene may hybridize with an EST probe, depending on the length of the gene, length of the cDNA clone, and the distribution of the *EcoR* I restriction sites in the gene. It was assumed that variation due to these factors would be rendered homogeneous across a large number of genes and probes. Assuming that the population of 730 ESTs is representative, the distribution of restriction fragments detected by these clones estimates the overall density of genes along chromosome arms. This estimate will be called relative gene density. Relative gene density is heterogeneous along individual



**Figure 1** Estimates of recombination rates (cM/Mb) within individual bins in relation to the position of bin midpoints on the centromere-telomere axis of wheat chromosome arms. On the horizontal axis, 0.0 represents the centromere and 1.0 represents the telomere. The curve and its equation best fitting the data are shown.

**Table 1. Mean Recombination Rate**

Chromosome interval	Mean CE
0.000–0.169	0.06
0.170–0.299	0.10
0.330–0.499	0.16
0.500–0.669	0.29
0.670–0.829	0.53
0.830–1.000	0.87

Mean recombination rates along all 42 wheat chromosome arms expressed as mean coefficient of exchange (CE), which is cM/Mb, in six chromosome intervals of indicated position on the centromere-telomere axis. The centromere is at 0,000 and telomere is at 1,000.

chromosome arms in wheat (Sandhu and Gill 2002), and each arm may show its own unique pattern. The critical question for this study was whether or not there is an overall pattern in gene density along the centromere-telomere axis that is superimposed on any chromosome-specific patterns.

A total of 2743 restriction fragments hybridized with the 730 ESTs. Assuming that the number of restriction fragments approximates the number of genes, the estimate of relative gene density is approximately based on 2743 genes. The relative gene density for all 159 bins was plotted against bin position on the centromere-telomere axis (Fig. 2). Relative gene density increased along chromosome arms, being the lowest in the proximal bins and the highest in the distal bins. Relative gene density along chromosome arms was weakly correlated with the relative distance of a bin center from the centromere ( $r_s = 0.22$ ,  $P = 0.007$ ). An equally weak correlation was observed between relative gene density and recombination rate ( $r_s = 0.22$ ,  $P = 0.007$ ).

### Distribution of Single-Gene and Multigene Loci Along Chromosome Arms

Because a hybridization signal is affected not only by gene number but also by nucleotide sequence divergence between a probe and a gene, it is very difficult, if not impossible, to determine the number of genes at a locus with sufficient precision by quantification of Southern hybridization signals in hexaploid wheat. Rather than quantifying the absolute number of genes per locus, loci were allocated into the single-gene and multigene classes according to the following criteria.

#### Single-Gene Class

Hybridization profiles showing only three restriction fragments, one per each of the three orthologous bins, most likely originated by the hybridization of a cDNA clone with restriction fragments of orthologous single-gene loci in the A, B, and D genomes. An additional attribute of such loci was that they were present once in each of the three wheat genomes. To quantify the frequency of these loci in a bin, the number of restriction fragments originating from these loci in a bin (S) was divided by the total number of restriction fragments in the bin (T).

From a total of 2743 restriction fragments produced by hybridization of the 730 ESTs, 23% fell into this class. The remaining restriction fragments have probably originated from multigene loci, duplicated single-gene and multigene loci, or single-gene loci present once in a genome with more

than one *EcoR* I restriction fragment hybridizing with an EST probe.

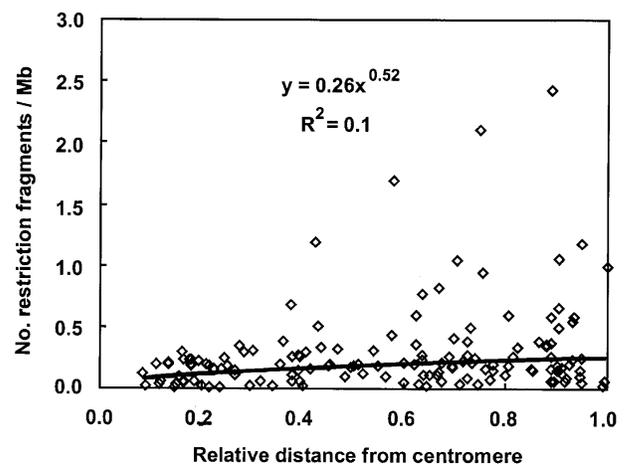
For most chromosomes, the S/T ratio declined ( $r_s = -0.48$ ,  $P < 0.0001$ ) with increasing relative distance of a bin from the centromere (Fig. 3). The S/T ratio also correlated negatively with recombination rate ( $r_s = -0.42$ ,  $P < 0.0001$ ). The S/T ratio decreased with approximately the square of bin relative distance from the centromere (Fig. 3). The relationship between the S/T ratio and the relative distance of a bin from the centromere was twice as strong for the short chromosome arms ( $r_s = -0.60$ ,  $p < 0.0001$ ,  $R^2 = 0.36$ ) than for the long chromosome arms ( $r_s = -0.36$ ,  $P < 0.0001$ ,  $R^2 = 0.13$ ). The difference between the two correlation coefficients was statistically borderline ( $P = 0.07$ ).

Correlation coefficients between the S/T ratio and the relative distance of a bin from the centromere ranged from  $r_s = -0.15$  for homoeologous group 5 to  $r_s = -0.66$  for homoeologous group 2. The coefficients were significantly different from zero for all homoeologous groups.

The extensively rearranged chromosome 4A in homoeologous group 4 was not included in any of these analyses. The structure of this chromosome was modified by a translocation and two inversions that happened during polyploid wheat evolution and one translocation that happened at the diploid level (Devos et al. 1995) and, therefore, the structure of wheat 4A does not reflect the structure of a single chromosome of any wheat diploid ancestor. Chromosome 4A did not show a gradient of the S/T ratio on the centromere-telomere axis in either arm.

#### Multigene Loci

Deletion mapping does not allow for an assessment of the frequency of multigene loci per bin. If an EST hybridized with several fragments that were mapped into a bin, a single multigene locus could be in the bin. However, the same pattern could also be due to the presence of two or more single-gene loci detected by the same probe in the bin, or because a single-gene locus was cleaved by *EcoR* I into several restriction fragments hybridizing with the probe. The sum of multiple restriction fragments hybridizing with ESTs per bin (M) there-



**Figure 2** Relative gene density along chromosome arms. On the horizontal axis, 0.0 represents the centromere and 1.0 represents the telomere. Relative gene density is estimated by the number of restriction fragments per Mb. The curve and its equation best fitting the data are shown.

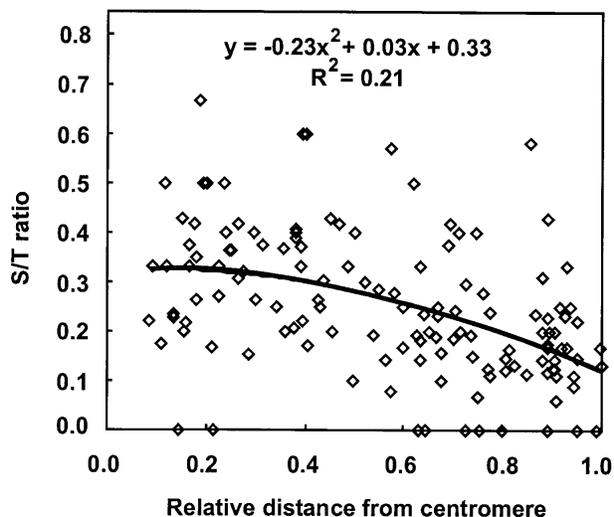
fore only approximates the number of genes located at multigene loci in a bin. The M/T ratio was generally higher in the distal bins than in the proximal bins ( $r_s = 0.29$ ,  $P = 0.0004$ ). Because the estimate of M was crude, no further characterization of the distribution of multigene loci was pursued.

### Duplicated Loci

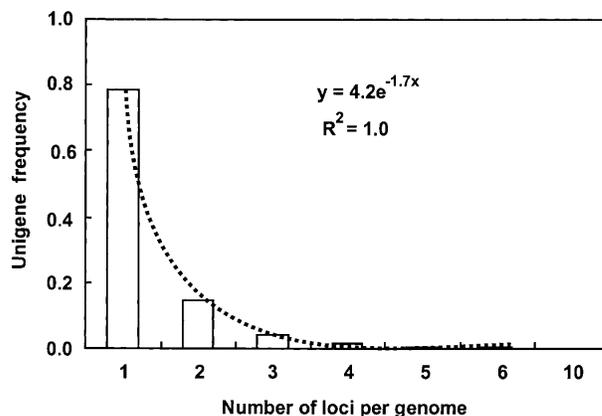
Data on deletion map positions of loci hybridizing with the 730 ESTs were used to assess the frequency of duplicated loci. A locus was assumed to be duplicated in a genome if an EST probe hybridized in a Southern blot with DNA fragments mapped to bins on different chromosomes within a genome. Linkage mapping showed that many intrachromosomally duplicated loci are located near each other. Since such loci often fall into a single bin, the number of intrachromosomally duplicated loci could not be estimated with any precision and therefore was not investigated. Loci detected by 79% of the ESTs mapped only to bins on homoeologous chromosomes, whereas those detected by the remaining 21% mapped from two to ten bins on different chromosomes within genomes (Fig. 4). The frequencies of ESTs hybridizing with 1, 2, ..., 10 paralogous loci closely fit an exponential function ( $R^2 = 1.0$ , Fig. 4).

### Chromosomal Distribution of Loci Within Paralogous Sets

The knowledge of homoeology between wheat and rice, *Oryza sativa* L., chromosomes (Gale and Devos 1998; M.E. Sorrells, C.M. La Rota, C.E. Bermudez-Kandianis, R.A. Greene, R. Kantety, J.D. Munkvold, Miftahudin, A. Mahmoud, J.P. Gustafson, L.L. Qi, B.E. Echaliier, B.S. Gill, D.E. Matthews, G.R. Lazo, S. Chao, O.D. Anderson, H. Edwards, A.M. Linkiewicz, J. Dubcovsky, E.D. Akhunov, J. Dvorak, D. Zhang, H.T. Nguyen, J. Peng, N.L.V. Lapitan, J.L. Gonzalez-Hernandez, J.A. Anderson, K.G. Hossain, V. Kalavacharla, S.F. Kianian, D.W. Choi, T.J. Close, M. Dilbirligi, K.S. Gill, C. Steber, M.K. Walker-



**Figure 3** Relationship between relative frequencies of single-gene loci (S/T ratio) per bin and the bin relative distance from the centromere. On the horizontal axis, 0.0 represents the centromere and 1.0 represents the telomere. The curve and its equation best fitting the data are shown.



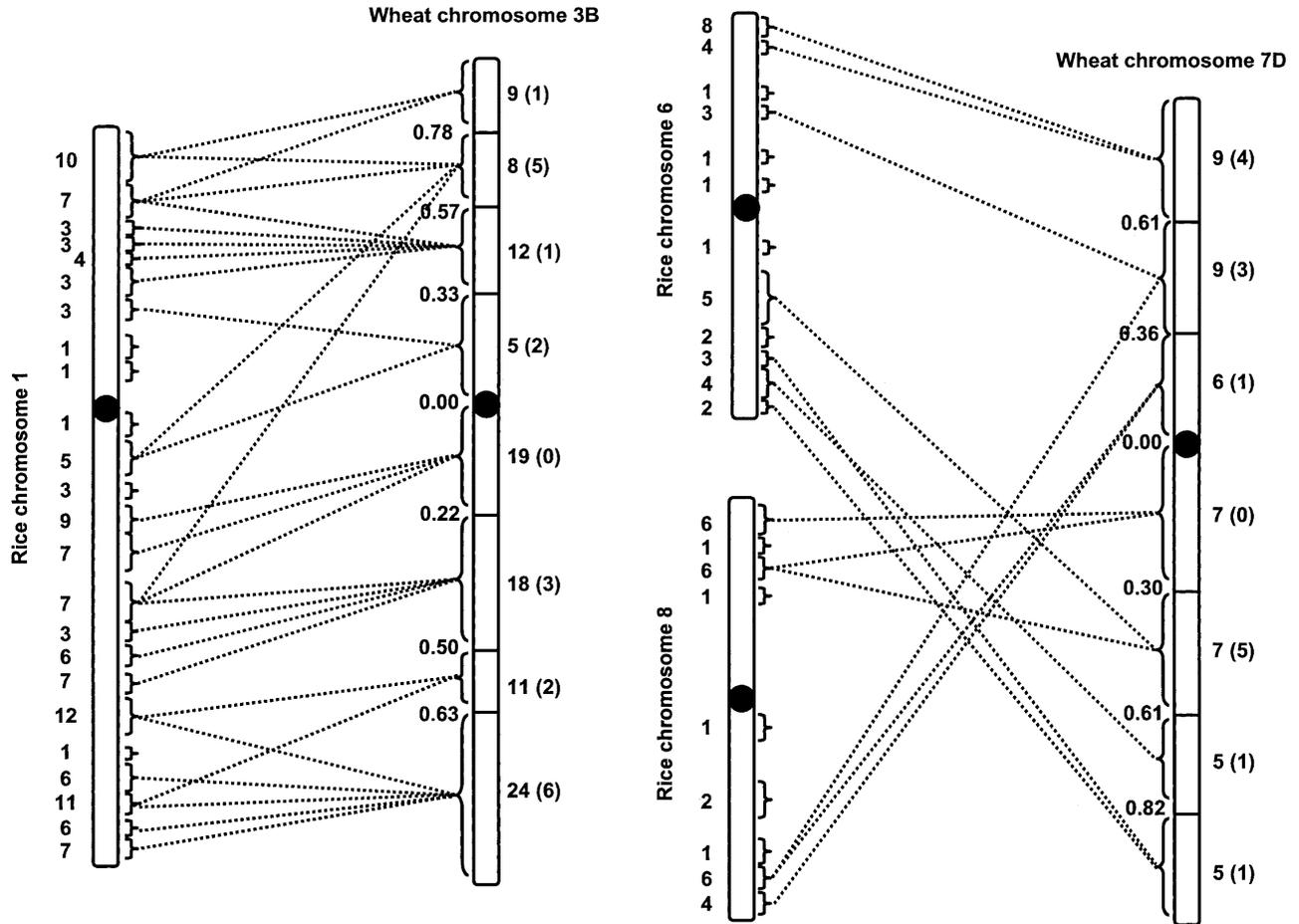
**Figure 4** Frequencies of EST unigenes detecting the indicated numbers of loci within a genome. The curve and the exponential equation best fitting the data are shown.

Simmons, P.E. McGuire, and C.O. Qualset) was exploited to infer for 40 sets of wheat paralogous loci which locus was ancestral and which was/were derived from it by duplication. Loci mapped on chromosomes of a specific wheat homoeologous group were recorded, and each locus was scrutinized for the presence of a homologous gene on a rice homoeologous chromosome. If a homologous gene was present on a rice homoeologous chromosome, the wheat and rice genes were assumed to be orthologous. If such a gene was a member of a wheat paralogous set, it was assumed that the locus was the ancestral locus of the wheat paralogous set. The locus (or loci) of the set present on a nonhomoeologous wheat chromosome was assumed to have originated by duplication of this ancestral locus. The prerequisite for these analyses was the existence of extensive homoeology between wheat and rice chromosomes and the availability of rice genomic nucleotide sequences. These requirements limited these analyses to only the chromosomes of wheat homoeologous groups 3 and 7.

Wheat chromosomes 3A, 3B, and 3D (wheat homoeologous group 3, henceforth W3) are homoeologous with rice chromosome 1 (henceforth R1) and wheat chromosomes 7A, 7B, and 7D (wheat homoeologous group 7, henceforth W7) are homoeologous with rice chromosomes 6 and 8 (henceforth R6 and R8). From 2100 ESTs, each selected from a unique EST contig, 373 detected loci on W3 and 347 detected loci on W7. Of the 373 loci, 146 had an ortholog on R1. These 146 loci encompassed the entire lengths of W3 and R1 (Fig. 5). Of these, 25 ESTs detected loci derived by duplication on wheat chromosomes belonging to five of the six remaining homoeologous groups. These 25 W3 loci were assumed to be the ancestral loci of the 25 paralogous sets.

Since the patterns of chromosomes 3A, 3B, and 3D were similar, only chromosome 3B, which is divided by deletion breakpoints into the highest number of bins among the three chromosomes in W3, is used as an example in Figure 5. In some cases, a locus was mapped on only two of the three wheat homoeologous chromosomes. Hence, for any of the three homoeologous chromosomes, the numbers of loci in each class were lower than the numbers for all three chromosomes. Only 20 of the above 25 loci of W3 were mapped on chromosome 3B (Fig. 5).

Of the 347 loci mapped on W7, 72 had an ortholog on R6 or R8 (wheat chromosome 7D is used as an example of W7 in



**Figure 5** Synteny between wheat chromosome 3B and rice chromosome 1 and wheat chromosome 7D and rice chromosomes 6 and 8. The boundaries of the bins delineated by deletion breakpoints on chromosomes 3B and 7D are indicated by horizontal lines across chromosomes. The distances of the breakpoints from the centromere, expressed as the fraction of arm length, are shown to the left of the breakpoints. The numbers to the right of the wheat chromosomes indicate the numbers of loci detected by ESTs per bin that had orthologous genes on rice homoeologous chromosomes. Numbers of wheat/ice orthologous loci that are members of wheat paralogous sets are indicated in parentheses. Wheat bins and rice contigs (indicated by brackets) containing two or more orthologous loci are connected by dotted lines. Rice contigs containing only a single putative ortholog are unconnected. The numbers of wheat/ice orthologs within each rice contig are indicated to the left of rice chromosomes. The black circles indicate centromeres.

Fig. 5). Of these, 15 detected a locus (or loci) derived by duplication on chromosomes belonging to all six remaining homoeologous groups. These 15 W7 loci were assumed to be the ancestral loci of 15 paralogous sets.

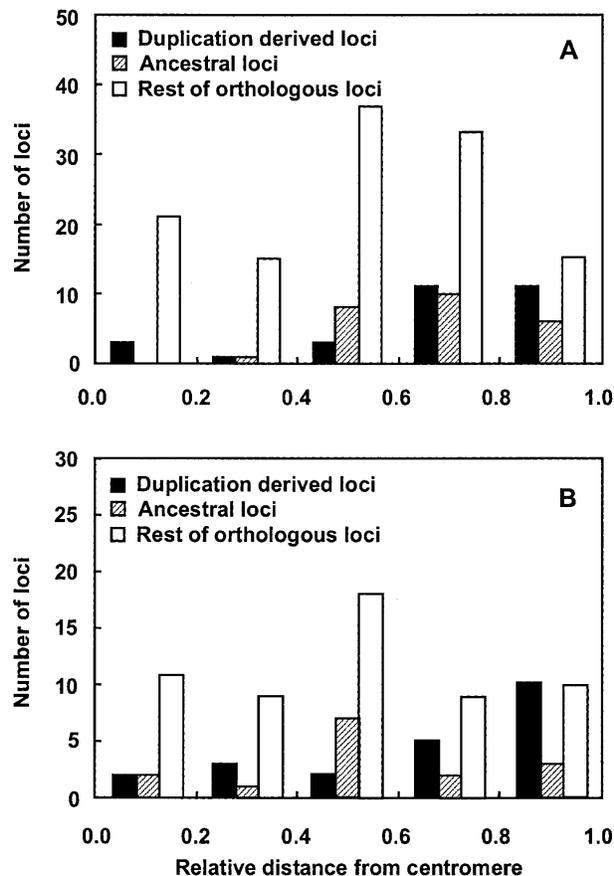
There was an extensive synteny between W3 and R1 for all W3 bins (Fig. 5). The same was true in the reverse direction; genes in contigs spanning the entire R1 showed synteny with W3 (Fig. 5). Synteny of W7 with rice chromosomes R6 and R8 was not as extensive, given the current state of the rice genome sequence and BAC ordering information available. All 7D bins had loci in common with R6 or R8 (Fig. 5). In addition to the extensive synteny between the investigated wheat and rice homoeologous chromosomes, the putative orthologs delineated large colinear segments (Fig. 5). This colinearity provided additional evidence for orthology of these loci.

The identification of the ancestral loci and loci derived by duplication for 40 paralogous sets made it possible to compare the distribution of the two classes of loci along chromosome arms. A majority of the derived loci were located in the distal regions of chromosome arms, but a majority of the an-

cestral loci were located proximal to them (Fig. 6). The distribution of the ancestral loci along chromosome arms was similar to the distribution of the remaining wheat/ice orthologs that were not subjected to locus duplication (Fig. 6).

The W3 and W7 data were combined to increase the sample size for statistical tests. The distributions of the ancestral and derived loci differed significantly from each other ( $P = 0.02$ ,  $\chi^2$ -test). The distribution of the ancestral loci did not significantly differ from the distribution of the remaining orthologs that were not subjected to locus duplication ( $P = 0.1$ ,  $\chi^2$ -test). The similarity of the distribution of the ancestral loci and the distribution along chromosome arms of the remaining orthologs that were not subjected to locus duplication suggested that there was no regional preference for a locus to be a source of a locus duplication.

To assess the correlation between the distribution of ancestral and derived loci along chromosome arms and recombination rate, the midpoints of bins were grouped according to the relative physical distance from a bin midpoint to the centromere into six equal intervals (rather than five shown in



**Figure 6** The distribution of loci derived by duplication of ancestral loci on W3, the ancestral loci, and the rest of W3 loci with orthologs on rice chromosome 1 along wheat chromosome arms (A) and the distribution of loci derived by duplication of ancestral loci on W7, ancestral loci, and the rest of W7 loci with orthologs on rice chromosome 6 plus 8 along wheat chromosome arms (B). The horizontal axes, on which 0.0 represents the centromere and 1.0 represents the telomere, show grouping of bin midpoints in which loci reside on the centromere-telomere axis into five equal intervals.

Fig. 6) to make the correlation statistically meaningful (data not shown). The CEs in each of the six intervals were averaged, and correlation coefficients between mean CE and the number of duplicated loci in the intervals were computed. The distribution of loci derived by duplication along wheat chromosomes was positively correlated with recombination rates along wheat chromosome arms (Pearson  $r = 0.97$ ,  $P = 0.002$ ). The distribution of the ancestral loci in the 40 paralogous sets did not significantly correlate with recombination rates along chromosome arms.

### Differentiation of R1 From Wheat Chromosome 3B

To quantitatively assess differentiation (perturbation of synteny) along chromosome arms, a comparison was made between wheat chromosome 3B (used as an example of W3) and R1. Whether or not there was a homologous locus on R1 was determined for 365 ESTs that detected loci on chromosome 3B in a database of 3977 ESTs. ESTs that detected loci on chromosome 3B in the previous section were a subset of these ESTs. Homologues on R1 were found for 158 of the 365 ESTs.

Differentiation increased (di) from the proximal bin to the distal bin in both 3B arms (Fig. 7).

## DISCUSSION

### Relative Gene Density

Distal regions of wheat chromosomes have been previously equated with gene-rich regions (Gill et al. 1993). Present findings showed that relative gene density per bin generally increase with the increase in relative distance of a bin from the centromere, and the highest gene densities were indeed found in several distal bins. However, the relationship between relative gene density and the bin position on the centromere telomere axis was overall weak.

Loci in distal, high-recombination chromosome regions are more polymorphic than those in the proximal, low-recombination regions in Triticeae (Dvorak et al. 1998; 1999). It is therefore easier to find RFLP for clones that originated from distal, high-recombination chromosome regions than for those that originated from proximal, low-recombination regions. Triticeae RFLP markers are therefore enriched for those detected by clones that originated from distal chromosome regions. Inclusion of RFLP markers into a set of markers used to study relative gene density may have therefore somewhat inflated the previous estimates of relative gene densities in the distal chromosome regions relative to estimates obtained with unselected cDNA clones used in the present study.

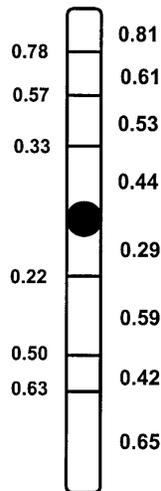
### Recombination Rates

Recombination rates have been previously assessed in the B genome on the basis of recombination of polymorphic C-bands in tetraploid wheat (Lukaszewski and Curtis 1993). Although a different method was used to estimate recombination rates in the present study, the results are similar. In both studies, recombination rate was a squaring function of the position of a locus on the centromere-telomere axis.

Lukaszewski and Curtis (1993) reported that short chromosome arms had steeper gradients than long chromosome arms. In the present study, recombination rates were studied without making distinction among the A, B, and D genomes of *T. aestivum*. The average gradient of recombination rates in the short arms was steeper than that in the long arms, but the difference was minor, possibly due to combining data for all three genomes. No distinction was therefore made between the short and long arms in most correlation analyses performed.

Mean recombination rate expressed as mean CE along wheat chromosome arms ranged from 0.06 (which is equal to 16.7 Mb/cM) in the most proximal interval to 0.87 (which is equal to 1.1 Mb/cM) in the most distal interval. Several estimates of local recombination rates computed from recombination of markers flanking sequenced genomic regions in wheat and related species have been reported. The estimates of local recombination rates in regions located in the most distal interval ranged from as low as 0.02 Mb/cM, measured in a short interval on the short arm of *Ae. tauschii* chromosome 1D (Spielmeier et al. 2000), to 2.3 Mb/cM in a 300-kb interval in the distal region of the short arm of *T. monococcum* L. chromosome 1A<sup>m</sup> (Stein et al. 2000). Recombination in the distal region of the short arm of 1A<sup>m</sup> is suppressed by the terminally located nucleolus organizing locus *Nor9* (Luo et al. 1998). This suppression of recombination may be responsible for the slightly lower local recombination rate than the aver-

Wheat chromosome 3B



**Figure 7** Differentiation rice chromosome 1 (R1) from between wheat chromosome 3B. The boundaries of bins (breakpoints) are indicated by horizontal lines across the chromosome. Their distances from the centromere, expressed as the fraction of arm length, are shown to the left of the breakpoints. The estimates of differentiation,  $d_i$ , of R1 from individual bins of 3B are to the right of the chromosome.

age rate of 1.1 Mb/cM. An estimated local recombination rate of 0.64 Mb/cM in a 90-kb region in the distal interval of *T. monococcum* 5A<sup>m</sup> (Tranquilli et al. 1999) is close to the average rate of 1.1 Mb/cM. Thus, only the 1D estimate of local recombination rate differs greatly from the average rate in the distal interval. Only one estimate of recombination rate in an interval that was not near the end of a chromosome has been reported. The midpoint of a bin in which this interval resided was at 0.73 of the 5AL chromosome arm length (San Miguel et al. 2002). A recombination rate of 2.7 Mb/cM was observed in a 215 kb interval (San Miguel et al. 2002). Substituting 0.73 in the equation in Figure 1 yields a CE of 0.5, which is equal to 2.0 Mb/cM. This is close to the observed local rate of 2.7 Mb/cM. Although the estimates of local recombination rates approximate average rates along chromosome arms reported here, they also reveal that local recombination rates may in specific cases differ from the average rates reported here.

The vast majority of homologous recombination in plants occurs within genes and in their immediate vicinity (Schnable et al. 1998). It was therefore expected that recombination would strongly correlate with gene density. However, this was not found. The relationship between recombination rate and relative bin position on the centromere-telomere axis was stronger ( $P < 0.001$ ) than the relationship between relative gene density and relative bin position on the centromere-telomere axis. This suggested that the relative position of a chromosome region on the centromere-telomere axis was more important for recombination rate than the local gene density.

A comparative study of recombination in structurally modified chromosomes and chiasmate pairing of deletion chromosomes in wheat suggested that the levels of recombination along wheat chromosomes are principally determined by the physical distances of chromosome regions from the telomere and positive interference distance (Lukaszewski and

Curtis 1993; Devos et al. 1995; Jones et al. 2002). These are probably the principal factors affecting the steepness of recombination rate gradients along wheat chromosome arms. In some species, such as rice, the gradients are minor (Wu et al. 2002). In species with the proximal localization of chiasmata (Darlington 1958) the gradients are probably oriented in the direction opposite to that found in wheat.

### Gene Duplication

A total of 21% of the ESTs in the present study detected duplicated loci on different chromosomes within wheat genomes. Anderson et al. (1992) used the nullisomic-tetrasomic synteny mapping strategy and reported 19.5% of clones detecting duplicated loci on different wheat chromosomes. Linkage mapping of the *T. monococcum* genome revealed higher levels of locus duplication; 27.7% cDNA clones and 34.4% *Pst*I genomic clones detected duplicated loci (Dubcovsky et al. 1996). Estimates of locus duplication in barley (*Hordeum vulgare* L.) obtained by linkage mapping ranged from 20% (Graner et al. 1991) to 30% (Kleinbols et al. 1993). Of all of the duplicated loci detected by linkage mapping in *T. monococcum*, 79% were interchromosomal and 21% intrachromosomal. Assuming that the ratio is similar in *T. aestivum*, the total frequency of inter- and intrachromosomally duplicated loci is expected to be about 25% of the wheat gene motifs. This estimate still represents a minimum since longer autoradiography or more strongly labeled probes may reveal additional fragments from paralogous loci that were not detected in some of the original autoradiograms.

If these duplicated loci were generated by polyploidy, the frequencies of ESTs hybridizing with even numbers of duplicated loci would have to be elevated relative to those hybridizing with odd numbers of loci, since only even sporophytic ploidy levels exist in nature in sexually reproducing grasses. The frequencies of ESTs hybridizing with even numbers of duplicated loci were not elevated relative to those hybridizing with odd numbers of duplicated loci (Fig. 4). The frequencies of ESTs hybridizing with 1, 2, ..., 10 paralogous loci closely fit an exponential function ( $R^2 = 1.0$ ; Fig. 4). The multiplicative nature of the function implies an approximately equal probability of duplication events. Furthermore, if loci derived by interchromosomal duplication of ancestral loci located on W3 and W7 originated by ancient polyploidy, they should be located on only one or few of the remaining chromosomes within wheat genomes and should largely be in blocks. The fact that loci derived by duplication appeared on either all or almost all chromosomes within a genome provided an additional argument that ancient polyploidy was not responsible for the origin of most of the paralogous sets of loci within wheat genomes. Although we do not know what exact mechanism was responsible for their origin, DNA transposition is the most likely possibility.

### Organization and Evolution of the Wheat Genomes

The wheat genomes can be characterized by two contrasting relationships. The first contrasting relationship was in the locations of single-gene loci present once in the wheat genomes and in the locations of multigene loci. The proximal, low-recombination regions of chromosome arms were enriched for the former, while the distal, high-recombination regions appeared to be enriched for the latter.

The second contrasting relationship was in the distribution of ancestral loci and duplicated loci derived from them

within paralogous sets of loci. While the majority of the ancestral loci of the investigated paralogous sets were located proximally, the majority of loci derived from them by duplication were located distally.

These two contrasting relationships are probably interrelated. Loci derived by duplication should represent more than one-quarter of all gene loci present in the wheat genomes since 25% or more of all wheat gene motifs are represented by paralogous sets, often containing more than one duplicated locus. For the 40 paralogous sets investigated, 61% of the loci derived by duplication were physically located in the distal one-third of chromosome arms. Therefore, a large proportion of all loci in the distal one-third of wheat chromosome arms were probably derived by gene duplication. This accumulation of duplicated genes in the distal chromosome regions must have influenced the distribution of genes along chromosome arms during the evolution of wheat genomes by increasing the absolute gene density in the high-recombination regions but reducing the relative proportion of genes that were originally located in the high-recombination regions. Therefore, the density of single-gene loci present once in the wheat genomes, i.e., those that were not subjected to duplication, should be the highest in the low-recombination regions, which was found. Single-gene loci present once in the genome should represent the ancient component of the genome. In chromosome 3B, used as an example of wheat homoeologous group 3, 56% and 71% of genes in the most proximal 3B bins were shared with R1. In contrast, only 35% and 19% of genes in the most distal 3B bins were shared with R1. Although differentiation of homoeologous chromosomes along their lengths has been investigated in only one of the seven wheat homoeologous groups, the finding is consistent with the prediction that the proximal regions of wheat chromosome arms may retain the largest proportion of ancient genes.

### Recombination Rates Drive the Rates of Evolution of the Wheat Genomes

Duplicated loci that are selectively neutral are analogous to other neutral DNA insertions responsible for RFLP in wheat and related diploid species. The levels of RFLP along chromosome arms are positively correlated with recombination rates along chromosome arms in wheat and its diploid relatives and other plants (Dvorak et al. 1998; 1999; Kraft et al. 1998; Stephan and Langley 1998). This relationship has been attributed to indirect positive selection, i.e., hitchhiking of neutral polymorphism with genes selected for by natural selection (Maynard Smith and Haigh 1974) or indirect negative selection, i.e., background selection against mildly deleterious genes (Charlesworth 1994). In both scenarios, there is a greater chance for neutral polymorphism and, by extension, polymorphism for loci derived by duplication, to be eliminated in low-recombination regions than in high-recombination regions. This differential chance of survival of polymorphism for duplicated loci in low- and high-recombination regions may account for the accumulation of loci derived by duplication in high-recombination regions of chromosome arms.

The same pattern would also emerge if the mechanism(s) responsible for transposition of genes were to target preferentially chromosome regions with high levels of homologous recombination. We are not aware of such a mechanism. Mobile genetic elements are known to target preferentially gene

neighborhoods (Cresse et al. 1995). In that case, however, the distribution of loci derived by duplication would have to be similar to that of other genes, which it is not.

Present findings suggest that the higher incidence of genes derived by duplication in the distal chromosome regions, diluting the relative proportion of ancestral loci and single-gene loci present only once in the wheat genomes proximally, is causally related to high recombination rates in distal chromosome regions in the wheat genomes. Assuming that this causality is true, the distribution of these various classes of genes along the chromosome arms would depend on the steepness of recombination rate gradients along chromosome arms and would likely be in a dynamic state reflecting changes in recombination rate gradients along chromosome arms during genome evolution.

Gene duplication is the principal path of new gene evolution (Ohno 1970; Ohta 1990; Hughes 1994; Walsh 1995; Force et al. 1999; Lynch and Force 2000). Many duplicated genes are inactive pseudogenes (Hollis et al. 1982; Rogers 1985). Dispersed duplicated genes that retain activity have a greater chance to evolve new function than tandem duplicated genes because they have a reduced chance for gene conversion and are not subjected to homogenization by unequal crossover, as are duplicated genes within a locus. Additionally, retroposition and other mechanisms of transposition of genes and gene fragments may occasionally generate novel, chimeric genes (Long and Langley 1993; Gilbert et al. 1997). The accumulation of dispersed copies of genes and gene fragments in the high-recombination regions may therefore increase the probability of the evolution of novel genes in those regions and faster differentiation of genomes in the high-recombination regions than in the low-recombination regions of chromosomes.

Correlations described earlier suggest that homoeologous chromosome differentiate (lose synteny) from each other faster in high-recombination regions than in low-recombination regions. Differentiation between homoeologous chromosomes 3B and R1 is consistent with this prediction: synteny of wheat chromosome 3B with its rice homoeolog declined from the proximal, low-recombination region to the distal high-recombination region in both arms. Unfortunately, the current state of rice genome sequencing and extensive structural differentiation between most of the wheat and rice homoeologous chromosomes allowed testing this prediction only in a single of the seven wheat homoeologous groups. In another analysis of the wheat EST database, differentiation along chromosome arms in relation to the recombination rate gradient was investigated between wheat homoeologous chromosomes of the A, B, and D genomes (E.D. Akhunov and J. Dvorak, unpubl.). The patterns were identical to that described here between chromosome 3B and R1; wheat homoeologous chromosomes were more differentiated from each other in distal, high-recombination regions than in proximal, low-recombination regions.

In *Caenorhabditis elegans* chromosomes I and V, genes with lethal mutant phenotype tend to be located predominantly in the low-recombination regions (Johnsen et al. 2000). Assuming that essential genes represent the conservative component of the genome, these findings would be consistent with the prediction that low-recombination regions tend to conserve ancient genes more than the high-recombination regions in genomes with steep recombination gradients along chromosomes.

Selecting wheat for the investigation of questions posed

in the introductory paragraphs was a fortunate choice. Wheat has steep gradients of recombination rates along chromosome arms and its large genomes have been subjected to extensive gene duplication and dispersion of duplicated genes. We were consequently able to observe an evolutionary process that has not been observed in studies of model organisms. Experimental evidence was obtained for the first time indicating that there may be a causal relationship between heterogeneity in recombination rates along chromosome arms and the underlying order in the position of genes in the chromosomes. The gradient of recombination rates on the centromere-telomere axis appears to be the causal factor in these relationships. Furthermore, evidence was obtained indicating that the rates with which genomes differentiate may not be constant along chromosome arms. Variation in recombination rates along chromosomes is again seen as the causal factor.

## METHODS

### ESTs

From 1000 to 11,000 clones were randomly selected from each of 41 cDNA libraries comprising most organs and developmental phases of the wheat plant. An average of 460 bp of 80,000 cDNA clones have been sequenced and deposited in GenBank ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). ESTs, their characteristics, and mapping status can be viewed at [http://wheat.pw.usda.gov/NSF/progress\\_est.html](http://wheat.pw.usda.gov/NSF/progress_est.html).

### Database

By September, 2002, a total of 3977 ESTs selected from different EST contigs had been hybridized with DNAs of deletion lines and other mapping stocks. Mapping data and images of autoradiograms were deposited in a database ([http://wheat.pw.usda.gov/cgi-bin/westsq1/map\\_locus.cgi](http://wheat.pw.usda.gov/cgi-bin/westsq1/map_locus.cgi)). The numbers of ESTs in the database were steadily increasing during the studies reported here, and, therefore, the various analyses reported here had to utilize different numbers of ESTs. For analyses, the database was queried to obtain the number of restriction fragments with which an EST probe hybridized and chromosome and bin location for each mapped fragment. Links from each EST in the database to GenBank were used to obtain nucleotide sequences when needed.

### Deletion Mapping

A total of 101 *T. aestivum* "Chinese Spring" deletion stocks, comprising 159 deletions, and 24 ditelosomic stocks and 21 nullisomic-tetrasomic stocks were used in EST mapping. The breakpoint of each deletion was expressed as a fraction of the chromosome arm length. Overlapping deletions delimited a physical region on a chromosome arm called a bin. The most proximal bin in each arm was delimited by the most proximal breakpoint and by the centromeric breakpoint in the relevant Chinese Spring ditelosome. A total of 159 bins were delimited by the breakpoints in the 21 wheat chromosomes. DNA was isolated from plants of the deletion stocks, nulli-tetrasomic stocks, and ditelosomic stocks, digested with *EcoRI* and fractionated in 1% agarose gels. DNA was depurinated with 0.25 M HCl for 25 min, and rinsed twice in distilled water followed by neutralization in 0.4 M NaOH for 20 min. Denatured DNA was transferred to Hybond N+ nylon membranes (Amersham) by capillary transfer in 0.4 M NaOH overnight. The membranes were then rinsed for 5 min in 2 × SSC and immediately prehybridized in either 6 × SSPE (1 × SSPE is 0.15 M NaCl, 0.01 M NaH<sub>2</sub>PO<sub>4</sub>, and 0.001 EDTA), 5 × Denhart's solution, 0.5% sodium dodecyl sulfate (SDS), 50 µg/mL sheared, single-stranded salmon sperm DNA for at least 6 h, or 2.5 × SSPE, 1%

SDS, 0.01% sodium pyrophosphate, and 0.1% polyanethanol-sulphonic acid for at least 30 min. cDNA inserts were polymerase chain reaction (PCR)-amplified using M13 Reverse and Universal primers. The products were sequenced and compared with the nucleotide sequence of each EST in the GenBank database to verify the authenticity of the PCR products. The products were labeled with <sup>32</sup>P using the random primer method. Hybridization was performed at 65°C overnight in either 6 × SSPE, 5 × Denhart's solution, 0.5% SDS or 7% SDS, 0.29M Na<sub>2</sub>HPO<sub>4</sub>, 0.22M NaH<sub>2</sub>PO<sub>4</sub>, and 5mM EDTA. The membranes were washed in 2 × SSC or SSPE and 0.5% SDS for 30 min at 65°C, 1 × SSPE or SSC and 0.5% SDS for 30 min at 65°C, and 0.5 × SSPE or SSC and 0.5% SDS for 10 to 30 min. at 65°C. The profiles of ESTs mapped into a bin were systematically compared with each other and duplicates were eliminated.

### Recombination Rate

Recombination rates were expressed as CEs (Lindsley and Sandler 1977). To compute CE for each bin, the genetic distances in cM between loci and their locations in a bin were determined from published wheat linkage and deletion maps (Devos et al. 1992; 1993; Devos and Gale 1993; Gill et al. 1993; 1996a,b; Hohmann et al. 1994; Delaney et al. 1995a,b; Mickelson-Young et al. 1995; Nelson et al. 1995a,b; Van Deynze et al. 1995; Dubcovsky et al. 1996; Jia et al. 1996; Marino et al. 1996). The physical distance in terms of the fraction of chromosome arm length between the centers of bins flanking a specific bin (<http://wheat.pw.usda.gov/west/binmaps>) was converted to megabase pairs as described earlier (Dvorak et al. 1998). For the most proximal bins, the physical distance was from the centromere to the center of the flanking bin. For the most distal bin, the physical distance was from the telomere to the center of the flanking bin. The estimate of genetic distance in cM was divided by the estimate of the length in Mb. The resulting fraction, cM/Mb, is defined as the CE within a bin.

### Identification of Wheat/Rice Orthologous Genes

The locations of the ancestral loci and those derived by duplication were determined for 40 paralogous sets. For all sets, the ancestral locus was on wheat homoeologous group 3 (W3) or homoeologous group 7 (W7). For W7, the distal portion of the short arm of chromosome 7B was not considered since that region was involved in a translocation with chromosome 4A, and does not represent the chromosome structure of wheat diploid ancestors. The following strategy was used to identify paralogous sets of loci with ancestral loci on W3 or W7. The hybridization patterns of ESTs mapped on W3 and W7 were examined. ESTs producing from one to three fragments for each of the orthologous bins of W3 or W7 were selected. The list of ordered BAC/PAC clones representing the full genomic sequence of rice chromosomes 1, 6, and 8 was downloaded from TIGR (<http://www.tigr.org>). The nucleotide sequences of those clones were retrieved from the GenBank database (<http://www.ncbi.nlm.nih.gov>) and used for the BLASTN (Altschul et al. 1990) search against wheat EST nucleotide sequences. The nucleotide sequences of the ESTs were used for BLASTN search with a threshold probability of  $e^{-10}$  against the genomic sequence of rice homoeologs. The presence of a homologous sequence on the rice homoeolog suggested the existence of putative orthologous genes on homoeologs.

### Correlation Between the Distribution of Ancestral Loci, Derived Loci, and Recombination Rates

Bins were grouped according to the distances of their midpoints from the centromere into six equal intervals along chromosome arms. Because means of the intervals were used

as variables in the correlation analysis, six rather than five intervals were used to make the correlation analysis statistically valid. The CEs in each of the six intervals were averaged and Pearson correlation coefficients were computed for the mean CE and the number of duplicated loci in the interval.

### Differentiation of R1 From Wheat Chromosome 3B

From a population of 3977 ESTs present in the database by August, 2002, 365 ESTs detected loci on chromosome 3B. Nucleotide sequences of these 365 ESTs were used for a BLASTN search against the genomic sequence of rice chromosome 1 using NCBI BLASTN 2.0. Sequences having e-values lower than  $e^{-10}$  and alignment length greater than 100 bp were considered homologous. The differentiation of R1 from 3B in the  $i$ th 3B bin ( $d_i$ ) was computed as follows:

$$d_i = 1 - \frac{n_i}{N_i}$$

where  $n_i$  is the number of EST clones that hybridized with loci in the  $i$ th 3B bin and had a homologs on R1 and  $N_i$  is the total number of EST clones that hybridized with loci in the  $i$ th 3B bin.

### Curve Fitting

The following types of functions were used to fit curves to data: linear, logarithmic, polynomial, exponential, and power. The fit showing the largest  $R^2$  was chosen as the best fit for each set of data.

### ACKNOWLEDGMENTS

This publication is based upon work supported by the National Science Foundation under Cooperative Agreement No. DBI-9975989.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

### REFERENCES

- 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.
- Anderson, J.A., Ogihara, Y., Sorrells, M.E., and Tanksley, S.D. 1992. Development of a chromosomal arm map for wheat based on RFLP markers. *Theor. Appl. Genet.* **83**: 1035–1043.
- Barakat, A., Carels, N., and Bernardi, G. 1997. The distribution of genes in the genomes of Gramineae. *Proc. Natl. Acad. Sci.* **94**: 6857–6861.
- Bernardi, G. 1996. The organization of the human genome. *Chemische Listy* **90**: 549–553.
- Bernardi, G., Hughes, S., and Mouchiroud, D. 1997. The major compositional transitions in the vertebrate genome. *J. Mol. Evol.* **44**: S44–S51.
- Bernardi, G. 1999. The genome organization in plants. In *Plant and Animal Genome VII* (ed. S.R. Heller), p. 131. Scherago Intl. Inc. San Diego, CA. abstract, www.intl-pag.org
- Carels, N., Hatey, P., Jabbari, K., and Bernardi, G. 1998. Compositional properties of homologous coding sequences from plants. *J. Mol. Evol.* **46**: 45–53.
- Charlesworth, B. 1994. The effect of background selection against deleterious mutations on weakly selected, linked variants. *Genet. Res. Cambridge* **63**: 213–227.
- Cresse, A.D., Hulbert, S.H., Brown, W.E., Lucas, J.R., and Bennetzen, J.L. 1995.  $\mu$ 1-related transposable elements of maize preferentially inserted into low copy number DNA. *Genetics* **140**: 315–324.
- Darlington, C.D. 1958. *Evolution of genetic systems*. Oliver and Boyd, Edinburgh, Scotland.
- Delaney, D., Nasuda, E.S., Endo, T.R., Gill, B.S., and Hulbert, S.H. 1995a. Cytologically-based physical maps of the group-2 chromosomes of wheat. *Theor. Appl. Genet.* **91**: 568–573.
- Delaney, D., Nasuda, E.S., Endo, T.R., Gill, B.S., and Hulbert, S.H. 1995b. Cytologically-based physical maps of the group 3 chromosomes of wheat. *Theor. Appl. Genet.* **91**: 780–782.
- Devos, K.M., Atkinson, M.D., Chinoy, C.N., Liu, C., and Gale, M.D. 1992. RFLP based genetic maps of the homoeologous group 3 chromosomes of wheat and rye. *Theor. Appl. Genet.* **83**: 931–939.
- Devos, K.M. and Gale, M.D. 1993. Extended genetic maps of the homoeologous group 3 chromosomes of wheat, rye, and barley. *Theor. Appl. Genet.* **85**: 649–652.
- Devos, K.M., Milan, T., and Gale, M.D. 1993. Comparative RFLP maps 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.
- Dubcovsky, J., Luo, M.C., Zhong, G.Y., Bransteitter, R., Desai, A., Kilian, A., Kleinhofs, A., and Dvorak, J. 1996. Genetic map of diploid wheat, *Triticum monococcum* L., and its comparison with maps of *Hordeum vulgare* L. *Genetics* **143**: 983–999.
- Dvorak, J. and Chen, K.-C. 1984. Distribution of nonstructural variation between wheat cultivars along chromosome arm 6Bp: Evidence from the linkage map and physical map of the arm. *Genetics* **106**: 325–333.
- Dvorak, J., Luo, M.-C., and Yang, Z.-L. 1998. Restriction fragment length polymorphism and divergence in the genomic regions of high and low recombination in self-fertilizing and cross-fertilizing Aegilops species. *Genetics* **148**: 423–434.
- Dvorak, J., Yang, Z.-L., and Luo, M.-C. 1999. Genetic variation, recombination and divergence along chromosomes. In *Plant and Animal Genome VII* (ed. S.R. Heller) p. 126. Scherago Intl. Inc. San Diego, CA. abstract, www.intl-pag.org
- Endo, T.R. and Gill, B.S. 1996. The deletion stocks of common wheat. *J. Hered.* **87**: 295–307.
- Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., and Postlethwait, J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**: 1531–1545.
- Gale, M.D., Atkinson, M.D., Chinoy, C.N., Harcourt, R.L., Jia, J., Li, Q.Y., and Devos, K.M. 1993. Genetic maps of hexaploid wheat. In *8th International Genetic Symposium* (eds. Z.S. Li and Z.Y. Xin), pp. 29–40. China Agricultural Sciencetech Press, Beijing, China.
- Gale, M.D. and Devos, K.M. 1998. Comparative genetics in the grasses. *Proc. Natl. Acad. Sci.* **95**: 1971–1974.
- Gilbert, W., deSouza, S.J., and Long, M.Y. 1997. Origin of genes. *Proc. Natl. Acad. Sci.* **94**: 7698–7703.
- Gill, K.S., Gill, B.S., and Endo, T.R. 1993. A chromosome region-specific mapping strategy reveals gene-rich telomeric ends in wheat. *Chromosoma* **102**: 374–381.
- Gill, K.S., Gill, B.S., Endo, T.R., and Boyko, E.V. 1996a. Identification and high-density mapping of gene-rich regions in chromosome group 5 of wheat. *Genetics* **143**: 1001–1012.
- Gill, K.S., Gill, B.S., Endo, T.R., and Taylor, T. 1996b. Identification and high-density mapping of gene rich regions in chromosome group 1 of wheat. *Genetics* **144**: 1883–1891.
- Granan, A., Jahoor, A., Schondelmeier, J., Siedler, H., Pillen, K., Fischbeck, G., Wenzel, G., and Herrman, R.G. 1991. Construction of an RFLP map of barley. *Theor. Appl. Genet.* **83**: 250–256.
- Hohmann, U., Endo, T.R., Gill, K.S., and Gill, B.S. 1994. Comparison of genetic and physical maps of group 7 chromosomes from *Triticum aestivum* L. *Mol. Gen. Genet.* **245**: 644–653.
- Hollis, G.F., Hieter, P.A., McBride, O.W., Swan, D., and Leder, P. 1982. Processed genes: A dispersed human immunoglobulin gene bearing evidence of RNA-type processing. *Nature* **296**: 321–325.
- Hughes, A.L. 1994. The evolution of functionally novel proteins after gene duplication. *Proc. Royal Soc. Lond. Ser. B. Biol. Sci.* **256**: 119–124.
- Jia, J., Devos, K.M., Chao, S., Miller, T.E., Reader, S.M., and Gale, M.D. 1996. RFLP-based maps of the homoeologous group-6 chromosomes of wheat and their application in the tagging of *Pm12*, a powdery mildew resistance gene transferred from *Aegilops speltoides* to wheat. *Theor. Appl. Genet.* **92**: 559–565.
- Johnsen, R.C., Jones, S.J.M., and Rose, A.M. 2000. Mutational accessibility of essential genes on chromosome I (left) in *Caenorhabditis elegans*. *Mol. Gen. Genet.* **263**: 239–252.
- Jones, L.E., Rybka, K., and Lukaszewski, A.J. 2002. The effect of a deficiency and a deletion on recombination in chromosome 1BL in wheat. *Theor. Appl. Genet.* **104**: 1204–1208.
- Kleinhofs, A., Kilian, A., Maroof, M.A.S., Biyashev, R.M., Hayes, P., Chen, F.Q., Lapitan, N., Fenwick, A., Blake, T.K., Kanazin, V., et

- al. 1993. A molecular, isozyme and morphological map of the barley (*Hordeum vulgare*) genome. *Theor. Appl. Genet.* **86**: 705–712.
- Kota, R.S., Gill, K.S., Gill, B.S., and Endo, T.R. 1993. A cytogenetically based physical map of chromosome-1B in common wheat. *Genome* **36**: 548–554.
- Kraft, T., Sall, T., Magnusson Rading, I., Nilsson, N.O., and Hallden, C. 1998. Positive correlation between recombination rates and levels of genetic variation in natural populations of sea beet (*Beta vulgaris* subsp. *maritima*). *Genetics* **150**: 1239–1244.
- Lindsley, D.L. and Sandler, L. 1977. The genetic analysis of meiosis in female *Drosophila melanogaster*. *Phil. Transact. Royal Soc. Lond. B* **277**: 295–312.
- Long, M.Y. and Langley, C.H. 1993. Natural selection and the origin of *Jingwei*, a chimeric processed functional gene in *Drosophila*. *Science* **260**: 91–95.
- Lukaszewski, A.J. and Curtis, C.A. 1993. Physical distribution of recombination in B-genome chromosomes of tetraploid wheat. *Theor. Appl. Genet.* **84**: 121–127.
- Luo, M.C., Yang, Z.L., and Dvorak, J. 1998. Position effects of ribosomal RNA multigene loci on meiotic recombination in wheat. *Genetics* **149**: 1105–1113.
- Lynch, M. and Force, A. 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**: 459–473.
- Marino, C.L., Nelson, Y.H., Lu, Y.H., Sorrells, M.E., Leroy, P., Tuleen, N.A., Lopes, C.R., and Hart, G.E. 1996. Molecular genetic maps of the group 6 chromosomes of hexaploid wheat (*Triticum aestivum* L. mend. Thell.). *Genome* **39**: 359–366.
- Maynard Smith, J. and Haigh, J. 1974. The hitchhiking effect of a favorable gene. *Genetic Res. Cambridge* **23**: 23–35.
- 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.
- Naranjo, T., Roca, A., Goicochea, P.G., and Giraldez, R. 1987. Arm homoeology of wheat and rye chromosomes. *Genome* **29**: 873–882.
- Nelson, J.C., VanDeynze, 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**: 525–533.
- Nelson, J.C., Vandeynze, 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**: 525–533.
- Ohno, S. 1970. *Evolution by gene duplication*. Springer, Berlin, Germany.
- Ohta, T. 1990. How gene families evolve. *Theor. Pop. Biol.* **37**: 213–219.
- Rogers, J.H. 1985. The origin and evolution retroposons. *Intl. Rev. Cyto.* **93**: 187–279.
- Sabeur, G., Macaya, G., Kadi, F., and Bernardi, G. 1993. The isochore patterns of mammalian genomes and their phylogenetic implications. *J. Mol. Evol.* **37**: 93–108.
- Sandhu, D. and Gill, K.S. 2002. Gene-containing regions of wheat and the other grass genomes. *Plant Physiol.* **128**: 803–811.
- San Miguel, P.J., Ramakrishna, W., Bennetzen, J.L., Busso, C.S., and Dubcovsky, J. 2002. Transposable elements, genes and recombination in a 215-kb contig from wheat chromosome 5Am. *Funct. Integr. Genomics* **2**: 70–80.
- Schnable, P.S., Hsia, A.P., and Nikolau, B.J. 1998. Genetic recombination in plants. *Curr. Opin. Plant Biol.* **1**: 123–129.
- Spielmeier, W., Moullet, O., Laroche, A., and Lagudah, E.S. 2000. Highly recombinogenic regions at seed storage protein loci on chromosome 1DS of *Aegilops tauschii*, the D-genome donor of wheat. *Genetics* **155**: 361–367.
- Stein, N., Feuillet, C., Wicker, T., Schlagenhauf, E., and 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.* **97**: 13436–13441.
- Stephan, W. and Langley, C.H. 1998. DNA polymorphism in *Lycopersicon* and crossing-over per physical length. *Genetics* **150**: 1585–1593.
- Tranquilli, G., Lijavetzky, D., Muzzi, G., and Dubcovsky, J. 1999. Genetic and physical characterization of grain texture-related loci in diploid wheat. *Mol. Gen. Genet.* **262**: 846–850.
- 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 group 1 chromosomes of Triticeae species and their relation to chromosomes in rice and oat. *Genome* **38**: 45–59.
- Walsh, J.B. 1995. How often do duplicated genes evolve new functions? *Genetics* **139**: 421–428.
- Wu, J.Z., Maehara, T., Shimokawa, T., Yamamoto, S., Harada, C., Takazaki, Y., Ono, N., Mukai, Y., Koike, K., Yazaki, J., et al. 2002. A comprehensive rice transcript map containing 6591 expressed sequence tag sites. *Plant Cell* **14**: 525–535.

## WEB SITE REFERENCES

- <http://www.tigr.org>; The Institute for Genomic Research (TIGR).
- <http://www.ncbi.nlm.nih.gov>; GenBank database at National Center for Biotechnology Information.
- <http://wheat.pw.usda.gov/wEST/>; wEST database (wheat EST) at USDA-ARS Western Regional Research Center server.
- <http://wheat.pw.usda.gov/west/binmaps/>; A database of bins delineated by deletion breakpoints at USDA-ARS Western Regional Research Center server.
- [http://wheat.pw.usda.gov/cgi-bin/westsq1/map\\_locus.cgi](http://wheat.pw.usda.gov/cgi-bin/westsq1/map_locus.cgi); A database of mapped loci in the bins at USDA-ARS Western Regional Research Center server.

Received September 16, 2002; accepted in revised form February 18, 2003.