Theor Appl Genet (2011) 123:455–464
DOI 10.1007/s00122-011-1598-4

ORIGINAL PAPER

Comparison of a high-density genetic linkage map to genome features in the model grass *Brachypodium distachyon*

Naxin Huo · David F. Garvin · Frank M. You · Stephanie McMahon · Ming-Cheng Luo · Yong Q. Gu · Gerard R. Lazo · John Philip Vogel

Received: 17 November 2010 / Accepted: 8 April 2011 / Published online: 20 May 2011

Abstract The small annual grass *Brachypodium distachyon* (Brachypodium) is rapidly emerging as a powerful model system to study questions unique to the grasses. Many Brachypodium resources have been developed including a whole genome sequence, highly efficient transformation and a large germplasm collection. We developed a genetic linkage map of Brachypodium using single nucleotide polymorphism (SNP) markers and an F_2 mapping population of 476 individuals. SNPs were identified by targeted resequencing of single copy genomic sequences. Using the Illumina GoldenGate Genotyping platform we placed 558 markers into five linkage groups corresponding to the five chromosomes of Brachypodium. The unusually long total genetic map length, 1,598 centiMorgans (cM), indicates that the Brachypodium mapping population has a high recombination rate. By comparing the genetic map to genome features we found that the recombination rate was positively correlated with gene density and negatively correlated with repetitive regions and sites of ancestral chromosome fusions that retained centromeric repeat sequences. A comparison of adjacent genome regions with high versus low recombination rates revealed a positive correlation between interspecific synteny and recombination rate.

Abbreviations

BAC	Bacterial artificial chromosome
BES	BAC end sequence
Bp	Base pair
cM	centiMorgans
kb	Kilobase
LTR	Long terminal repeat
SNP	Single nucleotide polymorphism
OPA	Oligo pool array
UTR	Untranslated region

Communicated by A. Schulman.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-011-1598-4) contains supplementary material, which is available to authorized users.

N. Huo · F. M. You · S. McMahon · Y. Q. Gu · G. R. Lazo · J. P. Vogel (✉)
USDA-ARS Western Regional Research Center, Albany, CA 94710, USA
e-mail: john.vogel@ars.usda.gov

N. Huo · F. M. You · M.-C. Luo
Department of Plant Sciences, University of California, Davis, CA 95616, USA

D. F. Garvin
USDA-ARS Plant Science Research Unit and Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108, USA

Introduction

The small annual grass *Brachypodium distachyon* (Brachypodium) possesses the physical and genomic attributes necessary to be a model system including: small size, rapid life cycle, and simple growth requirements (Draper et al. 2001; Garvin et al. 2008; Vogel and Bragg 2009). Recently, significant progress has been made in developing the genomic resources and experimental tools needed to expedite research in Brachypodium (Gu et al. 2009; Hasterok et al. 2006; Huo et al. 2006, 2008, 2009; Vain et al. 2008; Vogel et al. 2006; Vogel and Hill 2008). A draft genome sequence was recently produced and the 272 Mb sequence was nearly complete (only 0.4% predicted gaps) and of very high quality (International Brachypodium Initiative 2010). Lacking from the list of Brachypodium resources is a comprehensive genetic

linkage map, the foundation of any genetic system. Although an SSR-based Brachypodium genetic linkage map consisting of 20 linkage groups based on 139 markers was recently reported, this map failed to coalesce into five linkage groups representing the five Brachypodium chromosomes, indicating that more markers are needed (Garvin et al. 2010) not only to accomplish this goal, but also to make the map a broadly useful tool for structural and functional genomics research such as positional cloning. Single nucleotide polymorphisms (SNPs) are the most abundant type of genetic polymorphism and they are amenable to high-throughput analysis. Since SNPs are sequence tagged markers with codominant inheritance, they are suitable for comparative genomic studies (Luo et al. 2009) and for genome wide linkage disequilibrium and association studies (Chan et al. 2010; Cogan et al. 2006; Kim et al. 2007; Ponting et al. 2007). SNPs can also address fundamental questions relating to evolution and meiotic recombination (Groenen et al. 2009; Myers et al. 2005). Many technologies have been developed to genotype SNPs cost-effectively in an automated fashion (Butler and Ragoussis 2008).

It is known that recombination rate varies dramatically among genomic regions, and numerous studies have attempted to understand the factors responsible for this variation. An overall positive correlation between gene density and recombination rate has been observed in many species (Anderson et al. 2006; Dvorak et al. 2004; Tian et al. 2009). However, for other genome features correlations to recombination seem to be taxon-specific. For example, a positive correlation between recombination rate and GC content has been report in mammals (Groenen et al. 2009; Jensen-Seaman et al. 2004; Myers et al. 2005) whereas a negative correlation has been reported for Arabidopsis chromosome 4 (Drouaud et al. 2006). In rice, recombination hot spots were associated with high gene density (Wu et al. 2003) whereas this trend was not detected in Arabidopsis (Drouaud et al. 2006). A high density genetic map would allow us to examine the patterns of genome wide recombination in Brachypodium and compare those patterns to the sometimes conflicting patterns observed in other species as noted above.

Here, we describe the development of a Brachypodium SNP-based genetic linkage map, and compare the distribution of recombination events to that of genome features, with an emphasis on the unusually high recombination rate observed in Brachypodium.

Materials and methods

Mapping population and DNA extraction

Four hundred and seventy-six F_2 plants derived from the cross of two inbred diploid Brachypodium lines, Bd3-1 and Bd21 (Garvin et al. 2010) were used for linkage analysis. This population is a pool of F_2 individuals from two separate F_1 hybrids. Leaf tissues of the individual lines and parent lines were harvested from seedlings, frozen, lyophilized, and DNA was extracted using a standard phenol:chloroform extraction method (Sharp et al. 1988).

SNP discovery and SNP genotyping

Regions for SNP discovery were selected both from Brachypodium Bd21 BAC end sequences (BES) (Huo et al. 2008) and the Brachypodium 4× draft genome assembly. Unique sequences which were evenly spaced along physical contigs (Gu et al. 2009; Huo et al. 2008) and 4× scaffolds were selected for analysis. PCR primers were designed using BatchPrimer3 (You et al. 2008) and used to amplify the selected regions from Bd3-1. The resulting amplicons were sequenced on an ABI 3730 DNA Analyzer (Choi et al. 2007). The resulting sequences were trimmed to remove low quality sequence, and polymorphic sites were identified using an in-house SNP discovery pipeline. The pipeline includes four steps: (1) mapping Bd3-1 sequences to Bd21 4× draft genome assembly using BLAST (identity percentage $\geq$95%, and *E* value $\leq$1e−50); (2) finding SNPs on the basis of sequence alignments between those two lines; (3) assessment of quality scores of discovered SNPs on both Bd-21 and Bd3-1 (Phred score $\geq$20 on both lines at the same loci); and (4) generating sequence fragments (150 bp in length) with SNP variants in the middle for Illumina GoldenGate assays. The pipeline was written in Perl and is available upon request.

All the SNPs were run through the Illumina assay design tool to predict their potential success as markers in the GoldenGate assay. A designability rank score was given to each SNP by Illumina, with scores ranging from 0 to 1.0, where a rank score of <0.4 has a low probability of success, 0.4 to <0.6 has a moderate probability of success, and >0.6 has a high probability of success. SNPs with designability rank score of 0.4 or higher were selected to be included in the oligo pool array (OPA).

The GoldenGate assay was performed by the UC Davis Genome Center per the manufacturer's protocol. Automatic allele calling for each locus was completed using BeadStudio software (Illumina, San Diego, CA). All genotype data were manually checked and re-scored if any errors in calling the homozygous or heterozygous clusters were evident.

Linkage map construction

Linkage analysis between markers, estimation of recombination frequencies, and determination of the linear order of loci was performed using JoinMap 4.0 software program with an initial logarithm of odds score of 10 (Van Ooijen

2006). Recombination rates were converted to genetic distances in centiMorgans (cM) using the Kosambi mapping function (Kosambi 1944).

Alignment to genomic sequence

The sequence flanking each SNP marker was compared to the assembled whole genome sequence using BLASTN to determine the position of each marker in the genome assemblies. Average recombination rates were obtained by dividing the total linkage distance (cM) by the total physical length (Mb) for each linkage group. These estimates were not adjusted for gaps between linkage groups or supercontigs or for differences in marker density.

Using the Brachypodium genome annotation v1.0 (International Brachypodium Initiative 2010), the percentages of gene sequence and LTR sequence of each marker interval were calculated and plotted against the recombination rate of each marker interval. Regression analysis was applied in logarithmic model.

Recombination rate and collinearity

Adjacent marker intervals with low recombination rates (<3.5 cM/Mb) and high recombination rates (>9 cM/Mb) were identified, and paired intervals with gene densities >8.5 gene/100 kb selected for further analysis. The sequences from selected intervals were extracted and GC content calculated using CoGe (Blas et al. 2009). Gene density, percentage of non-gene sequence and percentage of LTR sequences was calculated as described above. The rice orthologous regions to each such Brachypodium interval were determined using CoGe SynMap (Allerdings et al. 2006). Only gene pairs with BLAST values of $E \leq e^{-10}$ were used to determine collinearity, and small inversions were not counted.

Results

SNP discovery

The parents of the F_2 mapping population, Bd21 and Bd3-1, were chosen because they both exhibit a spring habit, and F_1 hybrids are fully fertile. They also appear to be highly divergent from each other (Vogel et al. 2009). To identify SNPs we designed 5,099 primer pairs to amplify unique sequences selected from Bd21 BES and preliminary 4× genome assembly sequences. We obtained useable sequence from 4,100 of these pairs to produce 3.8 Mb of sequence from Bd3-1. The remaining 999 primer pairs either failed to produce a PCR product or failed to produce a useable sequence. High quality sequences (Phred score >20) were compared to Bd21 sequence using an in-house SNP discovery pipeline to detect SNPs. Among the 4,100 sequences produced, 968 had at least one polymorphism and the remaining 3,132 loci were not polymorphic. In total, 3,387 SNPs (Phred >20) at 968 loci were detected, giving an average SNP frequency of ~1 SNP/kb of aligned sequence data from the two genotypes (Online Resource 1).

Table 1 Assignment of SNPs to genic or intergenic regions

	Genic region		Intergenic region (%)	Total
	Exon/CDS (%)	Intron/UTR (%)		
All markers	604 (17.8)	837 (24.7)	1,946 (57.5)	3,387
Mapped markers	123 (22.0)	177 (31.7)	258 (46.3)	558

We used the Brachypodium v1.0 genome annotation (International Brachypodium Initiative 2010) to determine whether SNPs were in genic or intergenic regions (Table 1). We define the genic region as anywhere between the 5′ and 3′ transcribed, untranslated regions (UTR), inclusive. Since not all genes had predicted UTRs, genic SNPs may be underestimated to a small degree. We showed that 58% (1,946) of the SNPs fell in intergenic regions and 43% (1,441) SNPs fell in genic regions. Among the SNPs located in genic regions, 604 were in exons and 837 were in introns or UTRs. The distribution of mapped markers was similar to the overall SNP distribution (Table 1).

The sequences flanking the SNPs were then analyzed using the Illumina assay design tool to predict potential success as markers in the GoldenGate assay. Of the 3,387 SNPs, 46 (1.4%) had a score <0.4, indicating they were not suitable for the GoldenGate platform, and were excluded from consideration. 259 (7.6%) had a score between 0.4 and 0.6 indicating a reasonable probability that they would work, and 3,082 (91%) had a score >0.6 indicating a high probability that they would work. Among the 3341 SNPs with design scores >0.4, 1976 occurred within 61 bp of another SNP, indicating they had a high probability of failing in the GoldenGate assay. The remaining 1,365 SNPs represented 872 loci. We considered SNPs originating from the same PCR amplicon as one locus. Of these 872 loci, 652 had at least one SNP with a Phred quality score >35. We chose to focus on SNPs meeting this higher quality threshold to avoid false-positives due to sequencing errors. SNPs from all 652 of these loci were selected for the OPA design. As a control, we included 46 SNPs that were <500 bp from another SNP in the OPA. Another 70 SNPs with Phred scores between 20 and 35 were included in the OPA because they had the potential to fill gaps along the sequence contigs.

Table 2 Mapping summary by chromosome

	No. markers	Distance (cM)	kb/marker	cM/marker	Physical length (Mb)	Recombination rate (cM/Mb)
Group 1	152	449.1	492.1	3.0	74.8	6
Group 2	91	348.3	651.6	3.8	59.3	5.9
Group 3	137	350.9	437.2	2.6	59.9	5.9
Group 4	112	267	433.9	2.4	48.6	5.5
Group 5	66	182.7	430.3	2.8	28.4	6.4
Total	558	1598	485.7	2.9	271	5.9

Genotyping and linkage map construction

The 768 Illumina GoldenGate SNP markers were used to genotype 476 F_2 plants. Of these markers, 200 were not used for linkage analysis for the following three reasons: (1) 169 markers failed to detect a SNP in the mapping population (2) nine markers were apparently heterozygous in one of the parent lines or (3) 17 markers failed to produce good cluster separation on BeadStudio software platform. Another 10 markers were excluded because the segregation ratio exhibited significant segregation distortion ($P < 0.0001$, χ^2). Data from the remaining 558 markers was used for linkage analysis with the JoinMap4 program. All 558 markers were included in the resultant linkage map, including 34 pairs of control markers that were within 500 bp of one another. In every case, each partner in each control marker pair mapped to the same genetic locus as the other partner indicating that the Illumina genotyping and mapping were working properly. The five linkage groups (LOD value of 10) correspond to the five chromosomes of Brachypodium (Table 2; Fig. 1; Online Resources 2, 3). The map spanned a total length of 1,598 cM with an average of 2.9 cM between markers. There were two marker intervals greater than 30 cM and 21 intervals greater than 10 cM. The genetic length for each chromosome ranged from 182.7 cM (chromosome 5) to 449.1 cM (chromosome 1).

The total genetic length of the Brachypodium map was larger than expected when compared to other plants such as rice, sorghum, wheat, and Arabidopsis (Table 3) (Bowers et al. 2003; Dubcovsky et al. 1996; Harushima et al. 1998; Luo et al. 2009; Mace et al. 2009; Wright et al. 2003). Given its genome size, we expected the map length and recombination rate for Brachypodium to fall between those of Arabidopsis and rice. Instead, the Brachypodium map length was greater than rice, and Brachypodium had the highest recombination rate, even higher than Arabidopsis (Table 3). Given this surprising finding, we wanted to verify the map length obtained by comparing distances generated from another mapping project and by another method. To compare our map to the previously generated SSR map that was developed using 183 of the 476 F_2 plants used in this study, we combined the genotyping data from 69 Brachypodium SSR markers (Garvin et al. 2010) and 558 SNP markers derived from 177 F_2 lines (six lines were not included because too many SSR genotyping calls were missing) and created a combined SNP–SSR map with JoinMap (not shown). This map had a total map length of 1,464 cM, nearly the same as the SNP-only map. We compared this combined map to the five largest linkage groups (representing parts of Brachypodium chromosomes 1, 2, 3, and 4) from the SSR map (Garvin et al. 2010). In each case,

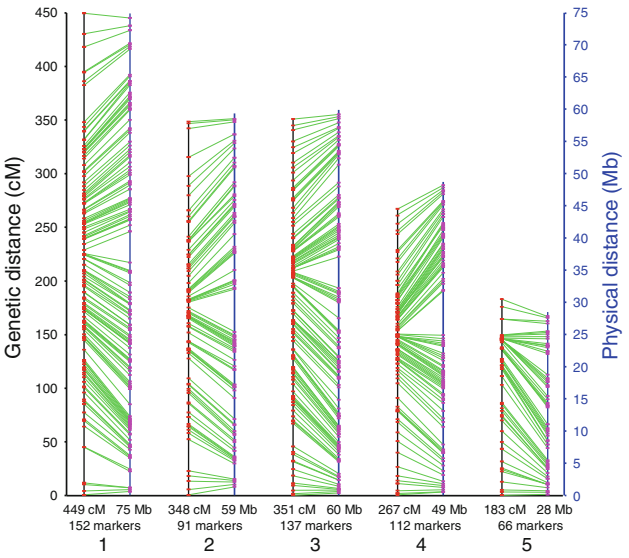

Fig. 1 Brachypodium linkage groups and their alignment to the Brachypodium genome sequence. Linkage group/pseudomolecule pairs are presented for each chromosome. In each pair, the *black bar* on the left represents the linkage group and the *blue bar* on the right represents the genomic sequence. They are oriented such that the start of the bp and map distances (0) is at the *top* of each chromosome. The location of each marker is indicated by a *red line* on the genetic map and a *pink line* on the genomic sequence. Corresponding locations for each marker is indicated by a *green line*. The scale for the linkage groups (cM) is on the left and the scale for the genomic sequence (Mb) is on the right. The sizes and number of markers are indicated below. Note that recombination rate is not uniform across the chromosome with higher recombination rates at the distal ends and lower in the central centromeric region. A figure showing the linkage groups with labels and distances can be found in Online Resource 2. A table listing all the markers and their positions can be found in Online Resource 3

Table 3 Comparison of recombination rate of plants with sequence genomes

	Genome size (Mb)[a]	Genetic map length (cM)[b]	Recombination rate (cM/Mb)[c]
Arabidopsis	119	597	5.0
Brachypodium	272	1,598	5.9
Rice	382	1,530	4.0
Sorghum	758	1,059	1.4

Note that the recombination rate for Brachypodium is greater than Arabidopsis despite having a genome over twice the size

[a] Assembled genome sizes (Arabidopsis Genome Initiative 2000; International Brachypodium Initiative 2010; International Rice Genome Sequencing Project 2005; Paterson et al. 2009)

[b] Map lengths from (Bowers et al. 2003; Harushima et al. 1998; Lister and Dean 1993)

[c] Calculated by dividing the map length by genome size

Table 4 Comparison of genetic distances between SSR and integrated map

Linkage group[a]	Chromosome	No. of Common SSR Markers	Distance in SSR map (cM)[a]	Distance in combined map (cM)
a	3	10	176.3	116.9
b	2	5	177.1	121.8
c	1	4	58.7	29
d	1	3	84.1	45.5
e	4	3	70.3	45.5

Note that for the common intervals, the SSR map sizes are slightly larger than the SNP map sizes. This is an independent verification of our map length

[a] From (Garvin et al. 2010)

the length of the common intervals was slightly longer in the SSR map (Table 4). In addition, the genetic length of linkage group 1 was calculated using another software package, QGA Station (http://ibi.zju.edu.cn/software/qga/index.htm), which resulted in a length estimate of 513 cM, longer than the length (449 cM) calculated by JoinMap4. Thus, we are confident that we have not overestimated the length of the Brachypodium map. Since many factors influence recombination frequency, it will be interesting to determine if other Brachypodium mapping populations have similarly high recombination rates.

Integration of genetic linkage map and genome sequence

To integrate the genetic linkage map and the Brachypodium genome sequence, we assigned each marker to a unique position in the 8× preliminary genome assembly by BLAST searches. Since only unique sequences were used to construct the markers, we were able to unambiguously assign each marker to one location in the genome (Fig. 1; Online Resource 3). The genetic map was then used to assemble the preliminary 8× sequence scaffolds into five pseudomolecules (International Brachypodium Initiative 2010). Where genetic mapping data was unable to determine marker order due to a lack of recombination between markers, we placed the markers into the same order as they are found in the genomic sequence The marker density of each chromosome ranged from 430 to 652 kb/marker, with an average of 486 kb/marker. There were 55 (10%) marker intervals larger than 1 Mb and 17 (3%) marker intervals larger than 2 Mb.

Recombination frequency in a genomic context

Genome-wide patterns of recombination have been described for Arabidopsis and rice. In both cases, recombination rates in the repeat-rich centromeric and pericentromeric regions are much lower than rates in gene-rich distal regions (Drouaud et al. 2006; Jensen-Seaman et al. 2004; Wright et al. 2003; Wu et al. 2003). By comparing genetic and physical distances between neighboring markers, we were able to examine the relative changes of recombination rate along each chromosome (Table 2; Fig. 2). The average recombination rate for each chromosome ranged from 5.5 to 6.4 cM/Mb with a genome-wide average of 5.9 cM/Mb. In contrast, recombination rates for smaller intervals varied dramatically, from 0 to 43.4 cM/Mb. As has been demonstrated in numerous other systems, we observed low recombination rates in centromeric and pericentromeric regions of each chromosome and higher recombination rates in regions with high gene density (Fig. 2).

To quantitatively examine the correlation between recombination rate and gene or long terminal repeat (LTR) density, we plotted the gene and LTR content of each map interval against the corresponding recombination rate (Fig. 3). Regression analysis showed a significant positive correlation between recombination rate and the percentage of gene sequence in an interval (Fig. 3). There was a significant negative correlation between recombination rate and the percentage of LTR sequence in an interval (Fig. 3).

Another approach to compare recombination rate with genomic feature composition is to rank all intervals from high to low recombination rate and then calculate a cumulative sum of each feature. In fact, such an analysis answers the question of how much of a given feature is contained in various fractions of high and low recombination regions. Figure 4 shows a deviation from a straight line which would be expected if there was no correlation between recombination rate and physical distance, coding sequence, or LTR sequence. The trend is most dramatic for LTR sequence where only 6.7% of the genetic distance accounts for 50% of the LTR sequence. It is less dramatic for

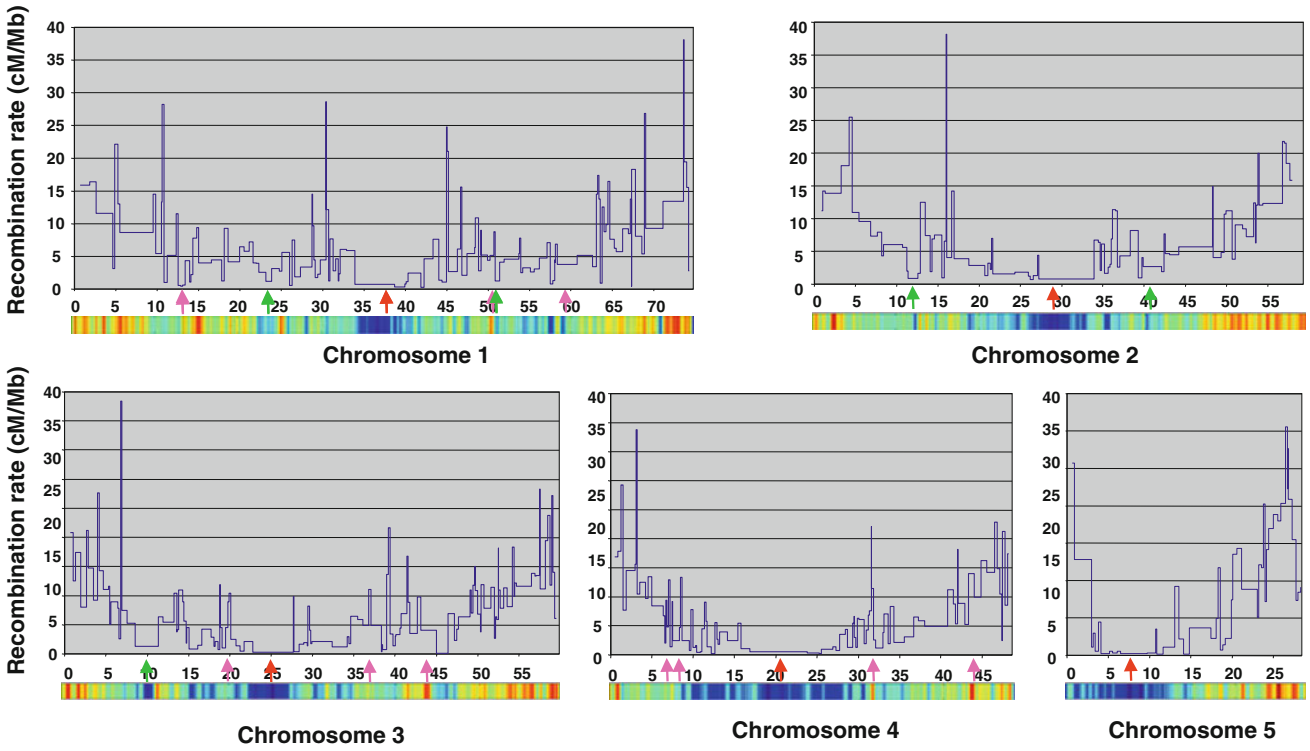

Fig. 2 Recombination rates of Brachypodium chromosomes. The *x* axis represents the physical length of each chromosome (Mb). The *y* axis represents the recombination rate (cM/Mb) calculated for the genomic regions between adjacent markers. *Red arrows* indicate the centromeres; *green arrows* indicate syntenic breakpoints with centromere repeats; pink arrows indicate the syntenic breakpoints without centromere repeats. The heat map at the *bottom* of each graph was adapted from (International Brachypodium Initiative 2010) and represents the percentage of coding sequence (exons) with the scale from *blue* to *red* representing a range of 0–22.3%

physical distance and gene sequence, where 50% of the physical distance or coding sequence is contained in 15.2 or 23.4% of the genetic distance, respectively. While there are obvious deviations from an even distribution, the curves are much less skewed than they would be expected to be for large genome species like wheat, maize or even sorghum.

A comparison of recombination-suppressed heterochromatic regions in rice and sorghum physical maps revealed that these regions were less syntenic than regions with higher recombination rates (Bowers et al. 2005). We examined previously described syntenic breakpoints and paired regions to see if synteny correlates with recombination rate in Brachypodium. Previously, 14 major syntenic disruptions between Brachypodium and rice/sorghum were identified (International Brachypodium Initiative 2010). These breaks in synteny are at the sites of ancient chromosome fusion events and describe a general role for the insertion of whole chromosomes into centromeric regions during grass genome evolution (International Brachypodium Initiative 2010). Five of these regions are also characterized by the presence of low copy number centromeric repeats (International Brachypodium Initiative 2010). All five of these regions had suppressed recombination (<3.1 cM/Mb) (Fig. 2). The recombination rate at the nine syntenic breakpoints without centromeric repeats varied from 0.6 to 13.9 cM/Mb (Fig 2). In these cases, recombination rate correlated with overall percentage of gene sequences indicating that a historical break in synteny does not predict continued low recombination rates.

To further examine the correlation between recombination rate and genomic features, including synteny to rice, we compared regions of high (>8.5 cM/Mb) and low (<3.5 cM/Mb) recombination rates. Since chromosomal position and gene density have large effects on recombination rate (Fig. 2), we focused our analysis on a paired comparison of low recombination intervals and high recombination intervals that were immediately adjacent to or within 211 kb and that had similar gene densities. First, we identified 155 regions with recombination rates less than 3.5 cM/Mb. As expected, most (107) of these intervals had lower than average gene densities (<8.5 genes/100 kb, the genome-wide gene density is 9.4 gene/100 kb) and were excluded from the analysis. Another 15 marker intervals were smaller than 100 kb or larger than 500 kb and were excluded from the analysis because small regions provided too few genes to accurately gauge synteny, and large regions are less likely to be uniform and may have sub regions that vary in recombination rate. We examined the remaining 33 intervals to determine if they bordered, or were very close to, intervals with a high recombination rate

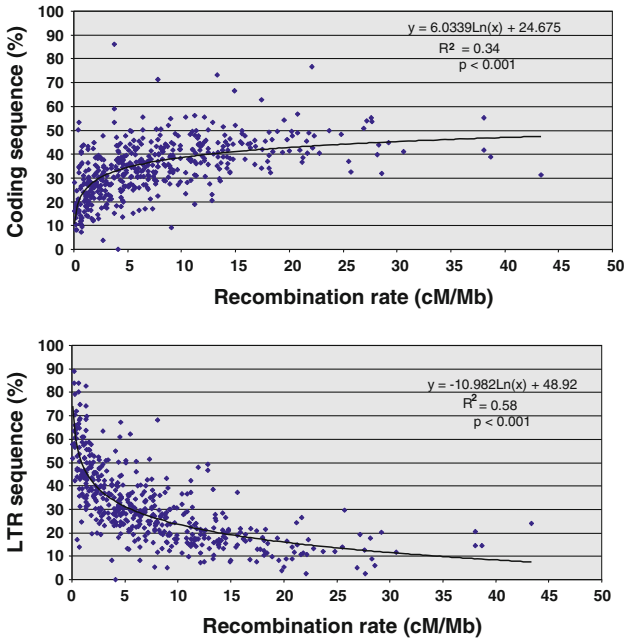

Fig. 3 Correlation of genome features with recombination rates. The **a** percentage of coding sequence and **b** percentage of LTR sequence contained in each interval was calculated and plotted against the recombination rate (cM/Mb). A logarithmic model curve was fitted to each dataset and the correlation coefficient calculated. Note that recombination rate is positively correlated with the coding sequence and negatively correlated with LTR sequence

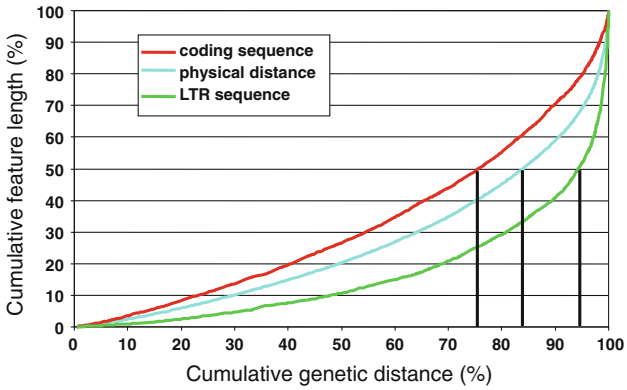

Fig. 4 Cumulative genetic distance plotted against cumulative physical distance gene and LTR sequence. Map intervals were ranked from highest to lowest recombination rate. Cumulative percentage of coding sequence, physical distance and LTR sequence were calculated by summing consecutive percentages of each feature in each map interval with the next interval. The resulting lines deviate from the *straight line* expected if there was no relationship between recombination rate and these features

(Online Resource 4). Fourteen intervals fit these criteria, 13 were adjacent to and one was within 211 kb of a high recombination rate interval. We then determined the gene density, the percentage of gene sequence, the percentage of non-gene sequence, the percentage of LTR sequence and GC content for each interval. We also determined the percentage of genes that were collinear when compared to the rice orthologous region. A paired Student's *t* test revealed that the degree of collinearity was significantly different ($P < 0.001$) between the high and low recombination rate intervals. In fact, in 13 out of 14 interval pairs the interval with the low recombination rate had a lower degree of synteny than the interval with the high recombination rate. To see if there were other distinguishing features of these intervals, we also use paired *t* tests to examine the gene density, the percentage of non-gene sequence and GC content. In all cases, the regions were not significantly different between the two groups (Table 5). The only other significant difference ($P < 0.01$) we noted was in the percentage of LTR sequences. The high recombination rate regions had 50% fewer LTR sequence than the low recombination rate regions. Thus, a major difference between these paired high and low recombination rate regions is in the degree of synteny to another species. While it is obvious that rearrangements between the parents of a mapping cross will suppress recombination in the affected region, it is surprising that synteny between two species that diverged over 40 million years ago would still track so significantly with recombination rate. Whether this is due to LTR sequences that continue to cause rearrangements or some other mechanism and how widely applicable this phenomenon is will require more detailed studies. As a practical application, if it is shown to be widespread, the correlation between synteny and recombination rate may be useful in predicting the utility of particular regions as a structural model for genomic regions in related grasses such as wheat and barley.

Discussion

This study provides answers to several important questions that will allow Brachypodium to serve as a model system. First, we showed that there is sufficient polymorphism (~1 SNP/kb of single-copy sequence) between Brachypodium genotypes to enable studies that require genetic diversity (e.g. map-based cloning, QTL analysis, association mapping, etc.). This supports earlier work based on a survey of SSR markers (Garvin et al. 2010), but since the current study focused on large numbers of single copy sequences it provides a more refined quantitative estimate of molecular diversity. The level of polymorphism we observed is similar to that observed in many other species including between the low copy regions of the indica and japonica subspecies of rice (Feltus et al. 2004). Also, our exclusion of multicopy sequences and repetitive DNA has likely reduced our SNP rate relative to other studies (Feng et al. 2002; Nasu et al. 2002) because genes and low-copy DNA

Table 5 Genomic composition of adjacent low and high recombination rate intervals

	Mean low recombination rate interval[b]	Mean high recombination rate interval[b]	Absolute mean difference	SD of mean difference	T^c	P value[c]
Genes collinear with rice orthologs (%)[a]	47.7 ± 19.3	65.7 ± 21.3	18	15.5	4.4	0.001
Gene density (genes/kb)	11.5 ± 1.7	12 ± 1.9	0.44	1.9	0.9	0.387
Gene sequence (%)	36.5 ± 6.23	38.7 ± 6.88	2.2	9.53	0.864	0.403
Non-gene sequence (%)	66.9 ± 7.3	64 ± 7.7	2.9	10.8	1	0.333
LTR sequence (%)	33.1 ± 8.9	22 ± 6.1	11	12.9	3.2	0.007
GC content (%)	45.4 ± 1.2	45.8 ± 0.8	0.38	1.3	1.1	0.287

The genomic composition of 14 adjacent intervals with similar gene densities that had high (≥9 cM/Mb) or low (≤3.5 cM/Mb) recombination rates were compared

[a] Genes were designated as collinear if they have BLAST hit at $E \leq e^{-10}$ to rice orthologous region, small inversions were not counted

[b] The mean ± the standard deviation of each feature is presented

[c] t statistic and P value were calculated using a paired Student's t test. Note that only the degree of collinearity and percentage of LTR sequence are significantly different between adjacent high and low recombination intervals

A table containing details for individual paired intervals can be found in Online Resource 4

generally evolve more slowly than multicopy sequences. We also showed similar rates of polymorphism in both genic and non-genic sequences. Second, we demonstrated that there are no particular difficulties in mapping markers in Brachypodium. This is an important finding because the first attempt to make a genetic map for Brachypodium using SSR markers did not produce five linkage groups, which raised the possibility of inherent problems associated with the cross, or with Brachypodium biology (e.g. large regions that lack polymorphism or unusual recombination behavior). Third, this genetic linkage map provided a template with which to assemble sequence scaffolds into five pseudomolecules representing the five Brachypodium chromosomes (International Brachypodium Initiative 2010). All of the mapped markers were collinear with the sequence assembly, giving us confidence that both the genetic map and sequence scaffolds are largely correct. Together these findings pave the way for future genetic studies in Brachypodium.

This study also identifies some potentially interesting aspects of recombination rate that may be worth further study. First, since Brachypodium is almost entirely inbreeding (Vogel et al. 2009), the unusually high recombination rate observed may be an adaptation to maximize generation of novel genetic variation when rare outcrossing events do occur. Since recombination rate is affected by many factors, including genetic interference, much more work needs to be done to determine if this finding is a broad feature of Brachypodium. Second, by examining the genomic composition of the recombination intervals, we showed that, as reported elsewhere, recombination rate correlates positively with gene density and negatively with LTR density. This is in agreement with the conclusions from studies in rice, maize, and wheat (Anderson et al. 2006; Dvorak et al. 2004; Tian et al. 2009). We also showed that recombination was suppressed in regions containing centromere repeat sequences present at ancestral chromosome insertion events, but found that regions of ancestral chromosome insertions without residual centromere repeat sequences were not inherently prone to low rates of recombination.

A more surprising finding was that recombination rate was positively correlated with the degree of synteny between rice and Brachypodium. This became apparent when we examined 14 intervals in which a high recombination region bordered a low recombination region. In all except one case, the region that exhibited higher synteny with rice had a higher recombination rate. Importantly, these regions were remarkably similar not only in terms of gene density, but also in percent of non-coding sequence and GC content. We did note a significantly higher LTR content in the low recombination intervals. It will be interesting to compare the genome sequence of both parents of the mapping cross to explore possible reasons for the differences in recombination rate.

Acknowledgments We would like to thank Haiming Xu (Zhejiang University) for QGA station analysis, Alan Schulman for LTR locations, and Edward Buckler for suggestions on presenting cumulative features (Fig. 4). This work was funded by an NSF 06-555 Plant Genome Research program grant and by USDA CRIS projects 5325-21000-013-00 and 3640-21000-021-00D.

References

Allerdings E, Ralph J, Steinhart H, Bunzel M (2006) Isolation and structural identification of complex feruloylated heteroxylan side-chains from maize bran. Phytochemistry 67(12):1276–1286

Anderson LK, Lai A, Stack SM, Rizzon C, Gaut BS (2006) Uneven distribution of expressed sequence tag loci on maize pachytene chromosomes. Genome Res 16(1):115

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

Blas AL, Yu Q, Chen C, Veatch O, Moore PH, Paull RE, Ming R (2009) Enrichment of a papaya high-density genetic map with AFLP markers. Genome 52(8):716

Bowers JE, Abbey C, Anderson S, Chang C, Draye X, Hoppe AH, Jessup R, Lemke C, Lennington J, Li Z, Lin YR, Liu SC, Luo L, Marler BS, Ming R, Mitchell SE, Qiang D, Reischmann K, Schulze SR, Skinner DN, Wang YW, Kresovich S, Schertz KF, Paterson AH (2003) A high-density genetic recombination map of sequence-tagged sites for Sorghum, as a framework for comparative structural and evolutionary genomics of tropical grains and grasses. Genetics 165(1):367

Bowers JE, Arias MA, Asher R, Avise JA, Ball RT, Brewer GA, Buss RW, Chen AH, Edwards TM, Estill JC, Exum HE, Goff VH, Herrick KL, Steele CLJ, Karunakaran S, Lafayette GK, Lemke C, Marler BS, Masters SL, McMillan JM, Nelson LK, Newsome GA, Nwakanma CC, Odeh RN, Phelps CA, Rarick EA, Rogers CJ, Ryan SP, Slaughter KA, Soderlund CA, Tang H, Wing RA, Paterson AH (2005) Comparative physical mapping links conservation of microsynteny to chromosome structure and recombination in grasses. Proc Natl Acad Sci USA 102(37):13206

Butler H, Ragoussis J (2008) BeadArray-Based genotyping. Methods Mol Biol 439:53–74

Chan EKF, Rowe HC, Kliebenstein DJ (2010) Understanding the evolution of defense metabolites in *Arabidopsis thaliana* using genome-wide association mapping. Genetics 185(3):991–1007. doi:10.1534/genetics.109.108522

Choi IY, Hyten DL, Matukumalli LK, Song Q, Chaky JM, Quigley CV, Chase K, Lark KG, Reiter RS, Yoon MS, Hwang EY, Yi SI, Young ND, Shoemaker RC, Van Tassell CP, Specht JE, Cregan PB (2007) A soybean transcript map: Gene distribution, haplotype and single-nucleotide polymorphism analysis. Genetics 176(1):685

Cogan NOI, Ponting RC, Vecchies AC, Drayton MC, George J, Dracatos PM, Dobrowolski MP, Sawbridge TI, Smith KF, Spangenberg GC, Forster JW (2006) Gene-associated single nucleotide polymorphism discovery in perennial ryegrass (*Lolium perenne* L.). Mol Genet Genomics 276(2):101

Draper J, Mur LAJ, Jenkins G, Ghosh-Biswas GC, Bablak P, Hasterok R, Routledge APM (2001) *Brachypodium distachyon*. A new model system for functional genomics in grasses. Plant Physiol 127(4):1539–1555

Drouaud J, Camilleri C, Bourguignon PY, Canaguier A, Berard A, Vezon D, Giancola S, Brunel D, Colot V, Prum B, Quesneville H, Mezard C (2006) Variation in crossing-over rates across chromosome 4 of *Arabidopsis thaliana* reveals the presence of meiotic recombination "hot spots". Genome Res 16(1):106

Dubcovsky J, Luo MC, Zhong GY, Bransteitter R, Desai A, Kilian A, Kleinhofs A, DvoÅ™Ã¡k J (1996) Genetic map of diploid wheat, *Triticum monococcum* L., and its comparison with maps of *Hordeum vulgare* L. Genetics 143(2):983

Dvorak J, Yang ZL, You FM, Luo MC (2004) Deletion polymorphism in wheat chromosome regions with contrasting recombination rates. Genetics 168(3):1665

Feltus FA, Wan J, Schulze SR, Estill JC, Jiang N, Paterson AH (2004) An SNP resource for rice genetics and breeding based on subspecies indica and japonica genome alignments. Genome Res 14(9):1812–1819

Feng Q, Zhang Y, Hao P, Wang S, Fu G, Huang Y, Li Y, Zhu J, Liu Y, Hu X, Jia P, Zhang Y, Zhao Q, Ying K, Yu S, Tang Y, Weng Q, Zhang L, Lu Y, Mu J, Lu Y, Zhang LS, Yu Z, Fan D, Liu X, Lu T, Li C, Wu Y, Sun T, Lei H, Li T, Hu H, Guan J, Wu M, Zhang R, Zhou B, Chen Z, Chen L, Jin Z, Wang R, Yin H, Cai Z, Ren S, Lv G, Gu W, Zhu G, Tu Y, Jia J, Zhang Y, Chen J, Kang H, Chen X, Shao C, Sun Y, Hu Q, Zhang X, Zhang W, Wang L, Ding C, Sheng H, Gu J, Chen S, Ni L, Zhu F, Chen W, Lan L, Lai Y, Cheng Z, Gu M, Jiang J, Li J, Hong G, Xue Y, Han B (2002) Sequence and analysis of rice chromosome 4. Nature 420(6913):316

Garvin DF, Gu YQ, Hasterok R, Hazen SP, Jenkins G, Mockler TC, Mur LAJ, Vogel JP (2008) Development of genetic and genomic research resources for *Brachypodium distachyon*, a new model system for grass crop research. Crop Sci 48:S-69–S-84

Garvin DF, McKenzie N, Vogel JP, Mockler TC, Blankenheim ZJ, Wright J, Cheema JJS, Dicks J, Huo N, Hayden DM, Gu Y, Tobias C, Chang JH, Chu A, Trick M, Michael TP, Bevan MW, Snape JW (2010) An SSR-based genetic linkage map of the model grass *Brachypodium distachyon*. Genome 53(1):1–13

Groenen MAM, Wahlberg P, Foglio M, Cheng HH, Megens HJ, Crooijmans RPMA, Besnier F, Lathrop M, Muir WM, Wong GKS, Gut I, Andersson L (2009) A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate. Genome Res 19(3):510

Gu YQ, Ma Y, Huo N, Vogel JP, You FM, Lazo GR, Nelson WM, Soderlund C, Dvorak J, Anderson OD, Luo MC (2009) A BAC-based physical map of *Brachypodium distachyon* and its comparative analysis with rice and wheat. BMC Genomics 10:496

Harushima Y, Yano M, Shomura A, Sato M, Shimano T, Kuboki Y, Yamamoto T, Lin SY, Antonio BA, Parco A, Kajiya H, Huang N, Yamamoto K, Nagamura Y, Kurata N, Khush GS, Sasaki T (1998) A high-density rice genetic linkage map with 2275 markers using a single F_2 population. Genetics 148(1):479

Hasterok R, Marasek A, Donnison IS, Armstead I, Thomas A, King IP, Wolny E, Idziak D, Draper J, Jenkins G (2006) Alignment of the genomes of *Brachypodium distachyon* and temperate cereals and grasses using bacterial artificial chromosome landing with fluorescence in situ hybridization. Genetics 173(1):349–362

Huo N, Gu Y, Lazo G, Vogel J, Coleman-Derr D, Luo M, Thilmony R, Garvin D, Anderson O (2006) Construction and characterization of two BAC libraries from *Brachypodium distachyon*, a new model for grass genomics. Genome 49:1099–1108

Huo N, Lazo GR, Vogel JP, You FM, Ma Y, Hayden DM, Coleman-Derr D, Hill TA, Dvorak J, Anderson OD, Luo M-C, Gu YQ (2008) The nuclear genome of *Brachypodium distachyon*: analysis of BAC end sequences. Funct Integr Genomics 8(2):135–147

Huo N, Vogel JP, Lazo GR, You FM, Ma Y, McMahon S, Dvorak J, Anderson OD, Luo MC, Gu YQ (2009) Structural characterization of Brachypodium genome and its syntenic relationship with rice and wheat. Plant Mol Biol 70(1–2):47

International Brachypodium Initiative (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463:763–768

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436:793–800

Jensen-Seaman MI, Furey TS, Payseur BA, Lu Y, Roskin KM, Chen CF, Thomas MA, Haussler D, Jacob HJ (2004) Comparative recombination rates in the rat, mouse, and human genomes. Genome Res 14(4):528

Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S, Ecker JR, Weigel D, Nordborg M (2007) Recombination and linkage disequilibrium in *Arabidopsis thaliana*. Nat Genet 39(9):1151–1155

Kosambi D (1944) The estimation of map distances from recombination values. Ann Eugen 12:172–175

Lister C, Dean C (1993) Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*. Plant J 4:745–750

Luo MC, Deal KR, Akhunov ED, Akhunova AR, Anderson OD, Anderson JA, Blake N, Clegg MT, Coleman-Derr D, Conley EJ, Crossman CC, Dubcovsky J, Gill BS, Gu YQ, Hadam J, Heo HY, Huo N, Lazo G, Ma Y, Matthews DE, McGuire PE, Morrell PL,

Qualset CO, Renfro J, Tabanao D, Talbert LE, Tian C, Toleno DM, Warburton ML, You FM, Zhang W, Dvorak J (2009) Genome comparisons reveal a dominant mechanism of chromosome number reduction in grasses and accelerated genome evolution in Triticeae. Proc Natl Acad Sci USA 106(37):15780

Mace ES, Rami JF, Bouchet S, Klein PE, Klein RR, Kilian A, Wenzl P, Xia L, Halloran K, Jordan DR (2009) A consensus genetic map of sorghum that integrates multiple component maps and high-throughput Diversity Array Technology (DArT) markers. BMC Plant Biol 9

Myers S, Bottolo L, Freeman C, McVean G, Donnelly P (2005) Genetics: a fine-scale map of recombination rates and hotspots across the human genome. Science 310(5746):321–324

Nasu S, Suzuki J, Ohta R, Hasegawa K, Yui R, Kitazawa N, Monna L, Minobe Y (2002) Search for and analysis of single nucleotide polymorphisms (SNPS) in rice (Oryza sativa, Oryza rufipogon) and establishment of SNP markers. DNA Res 9(5):163–171

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob Ur R, Ware D, Westhoff P, Mayer KFX, Messing J, Rokhsar DS (2009) The Sorghum bicolor genome and the diversification of grasses. Nature 457:551–556

Ponting RC, Drayton MC, Cogan NOI, Dobrowolski MP, Spangenberg GC, Smith KF, Forster JW (2007) SNP discovery, validation, haplotype structure and linkage disequilibrium in full-length herbage nutritive quality genes of perennial ryegrass (*Lolium perenne* L.). Mol Genet Genomics 278(5):585–597

Sharp PJ, Kreis M, Shewry PR, Gale MD (1988) Location of β-amylase sequence in wheat and its relatives. Theor Appl Genet 75:286–290

Tian Z, Rizzon C, Du J, Zhu L, Bennetzen JL, Jackson SA, Gaut BS, Ma J (2009) Do genetic recombination and gene density shape the pattern of DNA elimination in rice long terminal repeat retrotransposons? Genome Res 19(12):2221–2230

Vain P, Worland B, Thole V, McKenzie N, Alves SC, Opanowicz M, Fish LJ, Bevan MW, Snape JW (2008) *Agrobacterium*-mediated transformation of the temperate grass *Brachypodium distachyon* (genotype Bd21) for T-DNA insertional mutagenesis. Plant Biotechnol J 6:236–245

Van Ooijen JW (2006) JoinMap®4, software for the calculation of genetic linkage maps in experimental population. Kyazma B.V., Wageningen, the Netherlands

Vogel JP, Bragg JN (2009) *Brachypodium distachyon*, a new model for the Triticeae. In: Feuillet C, Muehlbauer G (eds) Genetics and genomics of the Triticeae, vol 7. Plant genetics and genomics: crops and models. Springer, New York, pp 427–449

Vogel J, Hill T (2008) High-efficiency Agrobacterium-mediated transformation of *Brachypodium distachyon* inbred line Bd21-3. Plant Cell Rep 27(3):471–478

Vogel J, Gu Y, Twigg P, Lazo G, Laudencia-Chingcuanco D, Hayden D, Donze T, Vivian L, Stamova B, Coleman-Derr D (2006) EST sequencing and phylogenetic analysis of the model grass *Brachypodium distachyon*. Theor Appl Genet 113:186–195

Vogel JP, Tuna M, Budak H, Huo N, Gu YQ, Steinwand MA (2009) Development of SSR markers and analysis of diversity in Turkish populations of Brachypodium distachyon. BMC Plant Biology 9:Art. No.: 88

Wright SI, Agrawal N, Bureau TE (2003) Effects of recombination rate and gene density on transposable element distributions in *Arabidopsis thaliana*. Genome Res 13(8):1897–1903

Wu J, Mizuno H, Hayashi-Tsugane M, Ito Y, Chiden Y, Fujisawa M, Katagiri S, Saji S, Yoshiki S, Karasawa W, Yoshihara R, Hayashi A, Kobayashi H, Ito K, Hamada M, Okamoto M, Ikeno M, Ichikawa Y, Katayose Y, Yano M, Matsumoto T, Sasaki T (2003) Physical maps and recombination frequency of six rice chromosomes. Plant J 36(5):720–730

You FM, Huo N, Gu YQ, Luo MC, Ma Y, Hane D, Lazo GR, Dvorak J, Anderson OD (2008) BatchPrimer3: a high throughput web application for PCR and sequencing primer design. BMC Bioinform 9:253