Plant Breeding **130,** 533—539 (2011)
doi:10.1111/j.1439-0523.2011.01866.x

Recombination frequency variation in maize as revealed by genomewide single-nucleotide polymorphisms

Mohammad Farkhari[1,2,*], Yanli Lu[1,3,*], Trushar Shah[4], Shihuang Zhang[5], Mohammad R. Naghavi[2], Tingzhao Rong[3] and Yunbi Xu[1,6,7]

[1]International Maize and Wheat Improvement Center (CIMMYT), Km 45, Carretera, Mexico-Veracruz, El Batan, Texcoco, Mexico; [2]Plant Breeding Department, College of Agriculture, University of Tehran, Karaj, Iran; [3]Maize Research Institute, Sichuan Agricultural University, Chengdu 611130, Sichuan, China; [4]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru 502324, Andhra Pradesh, India; [5]Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, China; [6]Institute of Crop Science/International Maize and Wheat Improvement Center (CIMMYT), Chinese Academy of Agricultural Sciences, National Key Facilities for Crop Genetic Resources and Improvement, Beijing 100081, China; [7]Corresponding author, E-mail: y.xu@cgiar.org

With 2 tables

Received January 4, 2011/Accepted March 2, 2011
Communicated by T. Lübberstedt

Abstract

Recombination frequency greatly affects selection efficiency in plant breeding. A high-density single-nucleotide polymorphism (SNP) map integrated with physical map and other molecular maps is very useful for characterizing genetic recombination variation. In this study, recombination frequency in maize was investigated through SNP linkage maps constructed with three recombinant inbred line populations. The integrated map consisted of 1443 molecular markers, including 1155 SNPs, spanning 1346 cM. A 100-fold difference in recombination frequency was observed between different chromosomal regions, ranging from an average of 0.09 cM/Mb for pericentromeric regions to 7.08 cM/Mb for telomeric regions. Recombination suppression in non-centromeric regions identified nine recombination-suppressed regions, one of which likely contained condensed heterochromatin (knobs). Recombination variation along chromosomes was highly predictable for pericentromeric and telomeric regions, but population-specific with 4.5-fold difference for the same marker interval across the three populations or specific chromosome regions because of the presence of knobs. As recombination variation can be identified and well characterized as shown in this study, the related information will facilitate future genetic studies, gene cloning and marker-assisted plant breeding.

Key words: recombination frequency — integrated linkage map — selection — centromere — maize

An accurate and well-saturated genetic linkage map is fundamental for genetic research and molecular breeding because it facilitates the development of robust marker-assisted selection systems as well as fine mapping for the cloning of quantitative trait loci (QTL) controlling important agronomic traits (Chen et al. 2002, Hwang et al. 2009). There has been a long and successful history of genetic linkage mapping in maize from the first partial maps of the 1930s (McClintock 1931) to the recent innovations of intermated recombinant inbred line mapping (Cone and Coe 2009) and nested association mapping (NAM) for complex traits (Yu et al. 2008, McMullen et al. 2009). At the same time, the development of single-nucleotide polymorphism (SNP) markers and new high-throughput genotyping systems has revolutionized maize genetics (Zhu et al. 2003, Fan et al. 2006) with different genotyping technologies available (Gunderson et al. 2005, Appleby et al. 2009). Around 80% of the maize genome sequence consists of non-coding repetitive sequences (Blanc and Wolfe 2004), and the majority of sequence variation in these regions does not have any known relationship with agronomic phenotypes. Thus, SNP markers developed from candidate genes have a much higher probability of being associated with important genes than AFLP, RAPD and non-genic SSR markers. Saturated and high-resolution linkage maps constructed using this type of SNP are more suitable for positional cloning. As the construction of SNP linkage maps is relatively new and there are far more linkage maps based on SSR markers that are still widely used, an integrated genetic map that includes both SNP and SSR markers is particularly useful in genetics and breeding.

In maize, high-throughput chip-based SNP genotyping systems (such as the GoldenGate and Infinium assays) have been established (http://www.illumina.com). Four custom oligo pool assays (OPA) each containing 1536 SNPs have been developed in the public sector for the GoldenGate assay in maize. SNPs with even distribution in the genome were used to develop the first 1536 chip, of which 1106 SNPs have been mapped in an integrated linkage map (ILM) using the NAM population (McMullen et al. 2009). The second OPA was developed using the 1106 mapped SNPs from the first OPA plus 430 new SNPs selected from the PANZEA database (Yan et al. 2010), which is called random (RA) chip in this study. The third OPA was developed based on 582 loci, about half of which were from drought-related gene loci (Yan et al. 2010), and from each locus, one or more SNPs were developed. The fourth OPA was built from the 943 good-quality SNPs from the third OPA combined with 593 SNPs selected from the PANZEA database (http://www.panzea.org) of having a designability score higher than 0.5 (Wen et al. 2010), which is called drought tolerance (DT) chip in this study.

Variation in recombination frequencies across the genome has important implications for breeders, which results in

*Contributed equally to this study.

different probabilities of obtaining desirable recombinants across the genome and significantly influencing selection strategies in plant breeding. In addition, variation of recombination frequencies in the same genomic locations between populations confounds efforts to understand genetic basis of quantitative traits (McMullen et al. 2009). Variation in recombination frequencies is influenced by both environmental and genetic factors (Xu et al. 1997, McMullen et al. 2009). Identification of the loci that affect genetic recombination variation may greatly enhance the precision of molecular plant breeding.

Maize linkage maps were divided into 100 bins each with an approximate length of 20 cM (Gardiner et al. 1993). Bins were distinguished from each other by one marker in each end of the bin, which was called core marker, and were designated with the chromosome number followed by a two-digit decimal (e.g. 1.00, 1.01, 1.02, etc). With this idea, rough comparisons among different linkage maps have been possible. A complete list for the core markers is available at MaizeGDB database (http://www.maizegdb.org) and from Davis et al. (1999).

In this study, linkage maps were constructed using three recombinant inbred line (RIL) populations genotyped with two 1536 SNP chips (RA and DT chips, as described earlier). The relationship between physical and genetic distances (variation in recombination frequencies) was then investigated at the whole genome level using three linkage maps. The SNP ILM was integrated with previously mapped molecular markers. The integrated map, along with the genomewide information on recombination frequency variation generated in this study, will facilitate future genetic studies, map-based cloning and marker-assisted plant breeding.

Materials and methods

Mapping populations: Three RIL populations were used in this study for construction of an ILM: 'XB' (X178 × B73) with 184 lines, developed by Institute of Crop Science, Chinese Academy of Agriculture Sciences (Hao et al. 2008), 'C5' (Ac7643 × Ac7729/TZSRW) (Ribaut et al. 1997) and 'C6' (CML444 × Malawi) (Messmer et al. 2009), both of which were developed at International Maize and Wheat Improvement Center (CIMMYT), Mexico, with 234 and 236 lines, respectively. These populations were originally developed for drought tolerance research. Drought-tolerant parental lines are X178, CML444 and Ac7643 (Ribaut et al. 1996, Hao et al. 2008, Trachsel 2009). 'XB' parents are temperate maize lines while 'C5' and 'C6' parents belong to subtropical and tropical groups of maize. Two subset (each with 31–39) lines, representing two extremes of anthesis-silking interval (a trait related to drought tolerance) and selected from each RIL population, were used for SNP linkage mapping with 71 lines from 'XB', 71 lines from 'C5' and 78 lines from 'C6'.

DNA extraction and SNP genotyping: Genomic DNA was extracted from seedling leaves using a locally optimized CTAB method (Laboratory Protocols, 2003). Illumina BeadStation 500 G (Illumina, Inc., San Diego, CA, USA) was employed for SNP genotyping at the Cornell University Life Sciences Core Laboratories Center following the protocols described by Fan et al. (2006) and Lu et al. (2009).

Construction of linkage maps: Mapmaker/exp v3.0 (Lander et al. 1987) was used to construct linkage maps. The map distance was calculated using the Kosambi mapping function, and Mapmaker error detection functions were employed to monitor potential errors at the probability level of 1% (Lincoln and Lander 1992). All pairs of linked SNP markers were first identified using the 'group' command with logarithm of the odds (LOD) score of ≥7.0 and recombination fraction = 0.3 (35 cM), a highly stringent significance threshold used to ensure that only tightly linked markers were selected for construction of framework maps. A primary core linkage map was then constructed for each linkage group (LG) by comparison with the physical map (maize-accessioned golden path, AGPv1: Release 4a.53; 38). Then, other markers were integrated using consecutive rounds of the Mapmaker 'build' command. Initially, a more restrictive criterion was used, including a 'multipoint exclusion threshold' with a LOD score of eight and a 'strict multipoint exclusion threshold' with a LOD score of nine. These criteria were then gradually decreased in subsequent steps until the 'multipoint exclusion threshold' and 'strict multipoint exclusion threshold' reached LOD scores of 2.5 and 2, respectively. In each step, after adding new markers to the map, the order of markers in the LG was verified and confirmed using the 'ripple' test. Markers were retained within the framework map only if they could pass the ripple test with LOD score ≥2. Flanking markers with no recombination in each round were defined as cosegregated markers, and only one of them was retained in the framework for the next round. Finally, using the 'build' command, the rest of markers that could not pass the ripple test were assigned to the most likely positions on a specific LG with a less stringent criterion and named 'try' markers in the final linkage map. With the relatively small population sizes used in this study, it can be expected that the resolution in mapping should be relatively low. However, using three independent mapping populations combined with the physical map should provide information of multiple sources to minimize the possible negative effect of small population size.

Construction of an ILM: An ILM was developed using genoplante® biomercator, Release V2.1 software (Arcade et al. 2004). The map from the 'XB' population was used as a reference map because it contained more markers with better quality than other populations. The order of markers and genetic distances in the reference map going through the integration steps were retained without undergoing any change. LGs belonging to the same chromosome (based on their physical locations) were then assembled together to develop a complete linkage map at LOD ≥ 3. Linkage maps constructed using 'C5' and 'C6' populations were then projected onto the reference map. Only framework and cosegregated markers were used to construct the ILM. Finally, the order of markers in the ILM was compared with the physical map, and some minor corrections were made for the markers from 'C5' or 'C6' linkage maps when the projection to the reference map had caused the order of markers to diverge from that of the physical map. New positions were calculated for these markers by integration from the physical map into the ILM using the standard procedure described for the genoplante® biomercator (Arcade et al. 2004). SNPs that could not pass ripple test ('try' markers) were assigned to their most probable marker intervals on the ILM without genetic distances, providing the marker order was consistent with the physical map.

Determination of physical positions for SNP, SSR and core markers: To determine the physical positions of the SNP markers, the original SNP sequences were used to perform a blastn (Altschul et al. 1990) against AGPv1 (Release 4a.53) downloaded from the Arizona Genomics Institute (http://www2.genome.arizona.edu/genomes/maize).

The SSR markers to be used in ILM were selected, one SSR every 5–10 cM, from IBM2 2008 ILM map (http://www.maizegdb.org). The respective amplicon sequence (amplicon plus flanking sequence) was then used to perform a blastn search against the AGPv1 to determine the physical position. A total of 901 SSR markers with an e-value score $\leq e^{-15}$ were selected, and then, an exact physical position was calculated for each marker with $(b - a)/2 + a$, where a and b were the start and finish positions of the related SSR sequence. The amplicon sequence of core markers also was retrieved from MaizeGDB (http://www.maizegdb.org). The physical positions for the core markers were determined using the same procedure as used for the SSR markers. The selected SSRs and the core markers that defined the bin borders were first

intercalated onto our physical SNP map based on their physical positions and then incorporated into the ILM using GENOPLANTE® BIOMERCATOR v2.1 (Arcade et al. 2004). The intercalation was performed based on homothetic projection. The distance ratio between shared markers for two different linkage maps (ILM and physical map) was calculated. Then, SSR and core markers from physical map were introduced to their corresponding position in the target map (ILM) based on their related distance ratios.

Detection of genetic recombination variation: Physical and genetic linkage maps were compared based on the cM/Mbp ratio for each population, which was used to reveal the variation in genetic recombination. The ratio was calculated for each bin that was defined by core markers. The first and last SNP markers with physical position in each bin were used to calculate cM/Mbp ratio. The markers with distance larger than 10 cM were excluded from the analysis to avoid the bigger standard deviation for the ratio estimate caused by the loosely linked markers. Centromeric regions were determined based on the cM/Mbp ratio. The bin with the lowest rate of recombination frequencies in each chromosome was considered as centromeric bin. As 'XB' population map had more markers with better quality and one of the parents was B73, variation in genetic recombination as revealed by cM/Mbp ratios along each chromosome was depicted by plotting cM against Mbp using data from the 'XB' population.

Results

SNP marker information

For the total of 3072 SNPs genotyped using two SNP chips, 1049 and 1090 SNPs were scorable with high quality for RA and DT chips, respectively. For these high-quality SNPs, polymorphic SNP markers in three RIL populations 'XB', 'C5' and 'C6' were 46%, 34% and 32%, respectively. The number of polymorphic markers was almost equal (330–370) with DT and RA chips for populations 'C5' and 'C6'. For the 'XB' population, RA chip revealed much higher polymorphism (Table S1). Both parental lines of the 'XB' population are temperate inbreds while the parental lines of the 'C5' and 'C6' populations are tropical/subtropical inbreds (Lu et al. 2009). Both our result and Lu et al. (2009) indicated that the RA chip could reveal a higher level of genetic polymorphism within temperate lines than within tropical/subtropical lines. A total of 1765 SNP markers could be assigned physical positions by using the original SNP sequences to perform a BlastN (Altschul et al. 1990) search against AGPv1 (Release 4a.53). Over 98% of the SNPs mapped physically have been placed on the same chromosomes as IBM2 2008 ILM. For the 2011 SSR markers in IBM2 2008 ILM (http://www.maizegdb.org), 1087 markers have a sequence available including the DNA region that could be amplified by SSR primers plus some other flanking nucleotides. When the related sequence at MaizeGDB was used to perform a BlastN search against the maize AGPv1 genome sequence database, a total of 903 SSR markers had a blast match with an e-value score $\leq e^{-15}$, of which 60 markers (6.6%) had a different chromosome assignment in the physical map from IBM2 2008 ILM and were thus excluded from further consideration in this study.

Linkage map and ILM

After grouping polymorphic SNP markers at LOD = 7, more than 10 (16–28) LGs were formed for each RIL population. Total chromosome (genome) lengths estimated from the 'XB', 'C5' and 'C6' populations were 1346.0, 899.5, 1426.2 cM, respectively (Table S2). This indicates that the 'C5' map was not complete as also suggested by the relatively small number of framework SNPs being mapped with this population (185 vs. 391 in 'XB'). With the stringent critical threshold (LOD ≥ 7) used in this study for construction of the framework maps, the average distance between LGs within chromosomes was 4.25, 20.0 and 22.0 cM for 'XB', 'C5' and 'C6' populations, respectively. Considering all criteria, the 'XB' map had the best resolution and thus used as a reference map for integrated linkage mapping.

A total of 1155 SNP markers were mapped onto the ILM, which spanned 1346 cM (Figure S1; Table 1), including 570 framework SNPs that exceeded the stringent linkage threshold of LOD ≥ 7. Of 1155 SNP markers, 123 were added to the map using 'try' command, which did not have exact positions on the ILM but were placed coarsely around framework SNP markers. A total of 462 SNP markers were cosegregated. By the end, the complete map shown in Figure S1 consisted of three types of SNP markers: framework (570), cosegregated (462) and try markers (123). For the ILM, the largest genetic distance between any two neighbouring SNPs was 18 cM (chromosome 2) while other SNPs formed several SNP clusters. The average marker density ranged from 1.13 SNP/cM for chromosome 5 to 0.63 SNP/cM for chromosome 7,

Table 1: Basic information of integrated linkage maps

Chromosome	# Framework SNPs	# Co-segregated SNPs	# Try SNPs	# Total SNPs	# SSRs	# Core markers	# Shared SNPs	# Total markers	# Disorder SNPs	Length (cM)	SNP/cM
Ch1	101	77	14	192	34	12	44	238	7	221.9	0.87
Ch2	70	54	13	137	25	9	39	171	4	164.0	0.84
Ch3	57	44	11	112	24	8	16	144	6	146.5	0.76
Ch4	59	29	14	102	22	7	11	131	4	120.0	0.85
Ch5	68	78	16	162	23	7	40	192	1	144.0	1.13
Ch6	47	28	7	82	18	7	19	107	5	112.2	0.73
Ch7	41	22	8	71	18	5	13	94	2	112.5	0.63
Ch8	50	70	15	135	21	6	24	162	6	128.0	1.05
Ch9	34	26	22	82	15	8	15	105	6	108.5	0.76
Ch10	43	34	3	80	14	5	28	99	2	88.4	0.9
Total and average	570	462	123	1155	214	74	249	1443	43	1346.0	0.86

Try SNPs, SNP markers that were added to the integrated map using 'try' command of Mapmaker to determine the linkage status without exact positions provided on the linkage maps.
Shared SNPs are those that have been mapped with at least two populations.
SNP, single-nucleotide polymorphism.

with an average of 0.86 SNP/cM. The number of SNPs on each chromosome ranged from 71 to 192. A total of 249 mapped markers (framework and cosegregated SNPs) were positioned in the map based on at least two mapping populations. For 43 SNPs (1–7 per chromosome), the marker orders on the ILM were not in agreement with those of the physical map (Table 1).

Some SSR and core markers were also added to the ILM. A total of 214 SSRs were projected to the ILM. Seventy-four of the 90 core markers had a sequence in the MaizeGDB database (http://www.maizegdb.org) and an e-value smaller than e^{-15} in blast results, which were also added to the ILM. For chromosomes 1 and 9, it was possible to add all the core markers to the ILM. Specific details about the ILM are shown in Table 1.

Genetic recombination and the ratio of genetic and physical distances

The average ratios of genetic and physical distances for 'XB', 'C5' and 'C6' populations were 0.63, 0.74 and 0.59 cM/Mbp, respectively. Variation in genetic recombination among the three populations was higher in suppressed recombination frequency regions than other regions, as indicated by the coefficient of variation (C.V.) that increased from telomere toward centromeric regions (Table 2).

Variation in genetic recombination across different bins within a chromosome was analysed (Table S3). Most severe recombination suppression occurred in centromere bins, and genetic recombination increased gradually from centromere toward telomeric regions. Recombination frequencies ranged from 0.06 cM/Mbp in centromeric regions to 8.47 cM/Mbp in telomeric regions, the latter being over 100 times higher than the former. For all chromosomes except chromosomes 1 and 3, the centromere bins defined by Davis et al. (1999) showed the most severe genetic recombination suppression in all of the three populations studied here. For chromosomes 1 and 3, bins 1.05 and 3.04 should be the centromere bins as more severe recombination suppression was detected across all the three populations in our study than the bins that were defined as centromeres by Davis et al. (1999). Among four bins (8.00–8.03) on chromosome 8 that were detected as the likely centromere location by Davis et al. (1999), bin 8.03 exhibited the most severe recombination suppression in this study.

Several long non-centromeric regions showed high recombination suppression in the 'XB' population. These long recombination suppression regions resided beside chromosome centromeres, except one of them which was located at long arm of chromosome 4 (Figure S2). Each of these regions was more than 10 cM long with a genetic recombination frequency of less than 1.5 cM/Mbp.

Table 2: Relationship between physical and genetic linkage maps for different chromosome regions as revealed by cM/Mbp ratios

Regions	'XB'	'C5'	'C6'	Average	C.V.
Telomere	3.29	4.46	3.51	3.81	0.11
Telomere neighbours	1.07	1.40	1.19	1.22	0.14
Centromere neighbours	0.49	0.71	0.46	0.53	0.25
Centromere	0.17	0.30	0.21	0.21	0.28
Average	0.63	0.74	0.59	0.63	0.12

Centromere neighbour regions include the two bins flanking the centromeric bin. Telomere regions consist of two terminal bins while the remaining bins form telomere neighbour regions.
C.V., coefficient of variation.

Discussion

ILM and its use in genetics and breeding

The physical map (AGPv1, release 4a.53) available currently and used in this study is almost complete, and only 6% (127.8 Mbp) of which has not been genetically ordered and oriented (Wei et al. 2009). Our results also confirmed that SSR and core marker positions were well placed, and reliable and physical orders of projected SSR and core markers were in agreement with their genetic orders as shown in the IBM2 2008 ILM and our ILM.

Development of high-density genetic linkage maps is highly dependent on the level of genetic polymorphism in mapping populations as well as the inherent level of detectable polymorphism of the molecular markers used. As the number of used markers increases, the number of distinguishable polymorphic alleles that exist between two parents will also increase. Because each type of DNA marker needs different equipment for genotyping, a specific genetic mapping project is usually based on one or two types of DNA markers. The latest estimate suggested that the maize genome is 2300 Mbp (Schnable et al. 2009) with one SNP every 44 bp (Gore et al. 2009). Thus, potentially, about 52 000 000 SNPs will be available for the whole maize genome, whereas until now, the total number of SSRs in the Panzea database (http://www.panzea.org) is only around 216 000. As all these SNPs can be genotyped using a high-throughput genotyping platform similar to that used in this study, the best type of DNA marker for construction of a high-density linkage map will be SNPs.

The level of genetic polymorphism in mapping populations can be increased by choosing diverse inbred parental lines as parents. A much higher level of genetic polymorphism was detected in this study with RA chip and 'XB' population, which can be partially attributed to the fact that the chip was developed by maximizing genetic polymorphism for the NAM population where a temperate maize line B73 was used to cross with 25 maize lines, including nine temperate lines (Lu et al. 2009). As mapping populations are usually developed for mapping of one or few traits, there often has to be a compromise regarding the extent of genetic diversity that can be captured for one trait while still providing a good level of variation for the second trait. The ILM approach allows a focus on trait-specific populations followed by using a genetic map with integration of different types of markers mapped previously (Falque et al. 2005).

The ILM developed in this study can be exploited in different ways. In molecular plant breeding, an ILM can have an axial role, particularly as the number of QTL mapping publications dramatically increases. This will allow breeders to utilize information from integrated maps to dissect the genomic regions most likely to influence a target agronomic trait and then dissect them using specific populations. Recent advances in genome sequencing platforms now offer a cost-effective option for genotyping large numbers of points on the genome across large population sizes common in plant breeding programmes and then empirically develop indices for genomewide selection (Heffner et al. 2009) and molecular breeding (Xu 2010).

Variation in genetic recombination along chromosomes

The frequency of recombination varies substantially between species, individuals and across different regions of the genome

(Nachman 2002). The genetic complexity of the maize genome creates a serious obstacle to exploring recombination frequency variation and the relationship between physical and genetic maps. Retrotransposon clusters are one of the factors that account for most of the repetitive DNA in maize and can suppress or enhance the rate of recombination in small genomic regions (Dooner and He 2008), and they vary markedly in composition and location relative to genes (Wang and Dooner 2006). In addition, recombinant modifier genes (Ji et al. 1999) and heterochromatic knobs may also influence genetic recombination. Dooner et al. (1985) combined molecular and genetic analyses of the *bronze* (*bz*) locus in maize and found that the level of meiotic recombination per unit of physical length is > 100-fold higher than the genome's average. Centromeric regions show dramatic recombination suppression (Round et al. 1997, Chen et al. 2002). In this study, the recombination suppression in these regions reached to more than 100 times in comparison with recombination frequencies at telomeric regions. Our results also indicate recombination frequencies gradually increased with distance from the centromere (Table 2, Table S3 and Figure S2). The current report provides detailed information about the variation in genetic recombination across the whole maize genome.

The centromeric bins defined by Davis et al. (1999) generally had the lowest recombination frequency compared to bins across the rest of each chromosome throughout the genomes of the three populations studied here. However, in two cases, the immediate neighbouring bin had the lowest values in our populations. Thus, our results can be used to infer most likely centromere locations for chromosomes 1 and 3, which should be in bins 1.05 and 3.04, respectively, instead of the bins 1.06 and 3.05 as defined by Davis et al. (1999). In both cases, there was a threefold difference in the recombination frequency, which was much larger than what would be expected from the difference in location of these core markers (Table S3). Other studies have also reported the same differences in predicted centromere locations for chromosomes 1 and 3 (Okagaki et al. 2008), including a recent precision mapping of centromeres (Wolfgruber et al. 2009). Another case is the centromeric bin for chromosome 8 where Davis et al. (1999) had defined a broad genomic region of bins 8.00–8.03 as the likely position. We observed severe suppression of recombination in bin 8.03, and thus, we conclude that this is the most likely bin for the centromere (Table S3), which is also supported by the markers defined for the centromere location by Wolfgruber et al. (2009).

When variation in recombination frequency was depicted along the chromosomes using 'XB' population (Figure S2), like as Gore et al. (2009), all chromosomes had a pericentromeric region which extended up to three bins beside the centromeres. Gore et al. (2009) investigated recombination variation along the maize chromosomes by genetic distances using the NAM population.

One non-pericentromeric region on bin 4.08–4.09 showed severe recombination suppression. The suppression of recombination in this large non-pericentromeric region was most likely caused by large blocks of constitutively condensed heterochromatin (called knobs) that were found throughout the maize genome (Peacock et al. 1981). Because knobs and centromeres share similarities in some aspects, it is possible that knobs also cause recombination suppression like centromeres. For example, knobs can have centromeric behaviour (Dawe 2009). Like centromeres, knobs also consist of heterochromatic blocks, which is composed of DNA repeats (180 bp repeat and TR1) (Peacock et al. 1981, Ananiev et al. 1998) that are arrayed in tandem for many kilobases such as CentC (major tandem repeat in centromere with 156 bp) (Adawy et al. 2004). On the other hand, we cannot attribute severe genetic recombination suppression in a long region (more than 10 cM) of chromosome 4 to any other known genetic structures. Therefore, it is concluded that depressed recombination regions observed on the long arm of chromosome 4 is a knob. This knob is one of the four knobs identified in inbred line B73 (Chughtai and Steffensen 1987).

We have also observed substantial variation in recombination frequency in the same bin between different populations (Table S3). For example, a 4.5-fold difference was found for bin 5.05 across the three populations. The variation might have been created by genetic properties associated with each population. Gore et al. (2009) observed up to 30-fold differences in recombination frequencies for the same locations in different populations.

A precise understanding of the relationship between physical and genetic distance along chromosomes will have significant impacts on genetics and breeding research. The reliability of genetic maps and success of genetic mapping and map-based cloning depend on the accurate estimation of recombination frequencies across the genome and at specific genomic regions, which in turn determines the selection strategies that can be used in plant breeding. With accurately estimated genetic distance between genes for specific populations under specific environmental conditions, it would be possible to calculate the minimum population size required to obtain desirable recombinants at a given chromosomal region. In backcross breeding programmes, for example, the population size and the number of generations required to recover the recurrent parent genome around the donor gene depend on recombination frequencies. The present study provides genomewide recombination frequencies that in most cases, especially at telomeric regions (Table S3), showed great consistency across the three populations tested. The overall difference in genetic recombination observed between centromeric and telomeric regions provides a general guidance for developing strategies for selecting desirable recombinants for the genes located in these regions and for map-based cloning of the genes even in the regions with low recombination rates such as pericentromeric regions where it would be very difficult without the complete genome sequence available. Significant differences identified for a specific genomic region across different populations, as shown in this study, could be also exploited for plant breeding. The populations with significantly increased recombination frequencies can be used to improve the probability of obtaining desirable recombinants around the regions, while the populations with significantly decreased recombination frequencies can be used to suppress the undesirable crossovers. Genes and genetic factors controlling crossovers may be cloned and used to engineer the genomic regions to obtain desirable recombination frequencies. Along with the recombination frequency information available genomewide, other technologies may be developed to either enhance or suppress recombination rates at the target genomic regions.

Acknowledgements

The SNP chips used in this study were developed through a collaborative project between Institute for Genomic Diversity, Cornell University (Ed. Buckler) and CIMMYT (Marilyn Warburton and

Jianbing Yan). Two RIL populations, 'C5' (Ac7643 × Ac7729/TZSRW) and 'C6' (CML444 × Malawi), were developed at CIMMYT for maize DT project (Jean-Marcel Ribaut). Dr Jonathan Crouch provided a comprehensive review on the draft manuscript. We thank Eva Huerta Miranda, Carlos Martinez Flores, Martha Hernandez Rodríguez, Alberto Vergara Alva, Maria Asunción Moreno Ortega and Jose Simon Pastrana Marias for laboratory and field assistance. This work was funded by the Rockefeller Foundation, the Bill and Melinda Gates Foundation, and the European Community, and through other attributed or unrestricted funds provided by donors of the Consultative Group on International Agricultural Research (CGIAR) especially the national governments of USA, Japan and UK. Mohammad Farkhari and Yanli Lu as joint PhD students were partially supported by Ministry of Science, Research and Technology of Iran and China Scholarship Council, respectively, for their research at CIMMYT, Mexico.

References

Adawy, S. S. M., R. M. Stupar, and J. Jiang, 2004: Fluorescence in situ hybridization analysis reveals multiple loci of knob-associated DNA elements in one-knob and knobless maize lines. J. Histochem. Cytochem. **52,** 1113—1116.

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

Ananiev, E. V., R. L. Phillips, and H. W. Rines, 1998: Complex structure of knob DNA on maize chromosome 9: retrotransposon invasion into heterochromatin. Genetics **149,** 2025—2037.

Appleby, N., D. Edwards, and J. Batley, 2009: New technologies for ultra-high throughput genotyping in plants. In: D. Somers, P. Langridge, and J. Gustafson (eds), Plant Genomics Methods and Protocols, 19—39. Humana Press, New York.

Arcade, A., A. Labourdette, M. Falque, B. Mangin, F. Chardon, A. Charcosset, and J. Joets, 2004: BioMercator: integrating genetic maps and QTL towards discovery of candidate genes. Bioinformatics **20,** 2324—2326.

Blanc, G., and K. H. Wolfe, 2004: Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell **16,** 1667—1678.

Chen, M., G. Presting, W. B. Barbazuk, J. L. Goicoechea, B. Blackmon, G. Fang, H. Kim, D. Frisch, Y. Yu, and S. Sun, 2002: An integrated physical and genetic map of the rice genome. Plant Cell **14,** 537—545.

Chughtai, S., and D. Steffensen, 1987: Heterochromatic knob composition of commercial inbreds of maize. Maydica **32,** 171—187.

Cone, K., and E. Coe, 2009: Genetic mapping and maps. In: J. Bennetzen, and S. Hake (eds) Handbook of Maize: Its Biology, 507—522. Springer, New York.

Davis, G. L., M. D. McMullen, C. Baysdorfer, T. Musket, D. Grant, M. Staebell, G. Xu, M. Polacco, L. Koster, and S. Melia-Hancock, 1999: A maize map standard with sequenced core markers, grass genome reference points and 932 expressed sequence tagged sites (ESTs) in a 1736-locus map. Genetics **152,** 1137—1172.

Dawe, R., 2009: Maize centromeres and knobs (neocentromeres). In: J. Bennetzen, and S. Hake (eds), Handbook of Maize: Genetic and Genomics, 239—251. Springer, New York.

Dooner, H. K., and L. He, 2008: Maize genome structure variation: interplay between retrotransposon polymorphisms and genic recombination. Plant Cell **20,** 249—257.

Dooner, H. K., E. Weck, S. Adams, E. Ralston, M. Favreau, and J. English, 1985: A molecular genetic analysis of insertions in the bronze locus in maize. Mol. Gen. Genet. **200,** 240—246.

Falque, M., L. Decousset, D. Dervins, A. M. Jacob, J. Joets, J. P. Martinant, X. Raffoux, N. Ribiere, C. Ridel, and D. Samson, 2005: Linkage mapping of 1454 new maize candidate gene loci. Genetics **170,** 1957—1966.

Fan, J. B., K. L. Gunderson, M. Bibikova, J. M. Yeakley, J. Chen, E. Wickham Garcia, L. L. Lebruska, M. Laurent, R. Shen, and D. Barker, 2006: [3] Illumina Universal Bead Arrays. Methods Enzymol. **410,** 57—73.

Gardiner, J. M., E. H. Coe, S. Melia-Hancock, D. A. Hoisington, and S. Chao, 1993: Development of a core RFLP map in maize using an immortalized F2 Population. Genetics **134,** 917—930.

Gore, M. A., J. M. Chia, R. J. Elshire, Q. Sun, E. S. Ersoz, B. L. Hurwitz, J. A. Peiffer, M. D. McMullen, G. S. Grills, and J. Ross-Ibarra, 2009: A first-generation haplotype map of maize. Science **326,** 1115—1117.

Gunderson, K. L., F. J. Steemers, G. Lee, L. G. Mendoza, and M. S. Chee, 2005: A genome-wide scalable SNP genotyping assay using microarray technology. Nat. Genet. **37,** 549—554.

Hao, Z. F., X. H. Li, C. X. Xie, M. S. Li, D. G. Zhang, L. Bai, and S. H. Zhang, 2008: Two consensus quantitative trait loci clusters controlling anthesis–silking interval, ear setting and grain yield might be related with drought tolerance in maize. Ann. Appl. Biol. **153,** 73—83.

Heffner, E. L., M. E. Sorrells, and J. L. Jannink, 2009: Genomic selection for crop improvement. Crop Sci. **49,** 1—12.

Hwang, T. Y., T. Sayama, M. Takahashi, Y. Takada, Y. Nakamoto, H. Funatsuki, H. Hisano, S. Sasamoto, S. Sato, and S. Tabata, 2009: High-density integrated linkage map based on SSR markers in soybean. DNA Res. **16,** 213—225.

Ji, Y., D. M. Stelly, M. De Donato, M. M. Goodman, and C. G. Williams, 1999: A candidate recombination modifier gene for *Zea mays* L. Genetics **151,** 821—830.

Laboratory Protocols, 2003: CIMMYT Applied Molecular Genetics Laboratory, 3rd edn, 7—11. CIMMYT, Mexico DF.

Lander, E. S., P. Green, J. Abrahamson, A. Barlow, M. J. Daly, S. E. Lincoln, and L. Newburg, 1987: MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics **1,** 174—181.

Lincoln, S. E., and E. S. Lander, 1992: Systematic detection of errors in genetic linkage data. Genomics **14,** 604—610.

Lu, Y., J. Yan, C. T. Guimarães, S. Taba, Z. Hao, S. Gao, S. Chen, J. Li, S. Zhang, B. S. Vivek, C. Magorokosho, S. Mugo, D. Makumbi, S. N. Parentoni, T. Shah, T. Rong, J. H. Crouch and Y. Xu, 2009: Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. Theor. Appl. Genet. **120,** 93—115.

McClintock, B., 1931: The order of the genes C, Sh and Wx in *Zea Mays* with reference to a cytologically known point in the chromosome. Proc. Natl Acad. Sci. USA **17,** 485—491.

McMullen, M. D., S. Kresovich, H. S. Villeda, P. Bradbury, H. Li, Q. Sun, S. Flint-Garcia, J. Thornsberry, C. Acharya, and C. Bottoms, 2009: Genetic properties of the maize nested association mapping population. Science **325,** 737—740.

Messmer, R., Y. Fracheboud, M. Bänziger, M. Vargas, P. Stamp, and J. M. Ribaut, 2009: Drought stress and tropical maize: QTL-by-environment interactions and stability of QTLs across environments for yield components and secondary traits. Theor. Appl. Genet. **119,** 913—930.

Nachman, M. W., 2002: Variation in recombination rate across the genome: evidence and implications. Curr. Opin. Genet. Dev. **12,** 657—663.

Okagaki, R. J., M. S. Jacobs, A. O. Stec, R. G. Kynast, E. Buescher, H. W. Rines, M. I. Vales, O. Riera-Lizarazu, M. Schneerman, and G. Doyle, 2008: Maize centromere mapping: a comparison of physical and genetic strategies. J. Hered. **99,** 85—93.

Peacock, W. J., E. S. Dennis, M. M. Rhoades, and A. J. Pryor, 1981: Highly repeated DNA sequence limited to knob heterochromatin in maize. Proc. Natl Acad. Sci. USA **78,** 4490—4494.

Ribaut, J. M., D. A. Hoisington, J. A. Deutsch, C. Jiang, and D. Gonzalez-de-Leon, 1996: Identification of quantitative trait loci under drought conditions in tropical maize. 1. Flowering parameters and the anthesis-silking interval. Theor. Appl. Genet. **92,** 905—914.

Ribaut, J. M., C. Jiang, D. Gonzalez-de-Leon, G. O. Edmeades, and D. A. Hoisington, 1997: Identification of quantitative trait loci under drought conditions in tropical maize. 2. Yield components

and marker-assisted selection strategies. Theor. Appl. Genet. **94,** 887—896.

Round, E. K., S. K. Flowers, and E. J. Richards, 1997: *Arabidopsis thaliana* centromere regions: genetic map positions and repetitive DNA structure. Genome Res. **7,** 1045—1053.

Schnable, P. S., D. Ware, R. S. Fulton, J. C. Stein, F. Wei, S. Pasternak, C. Liang, J. Zhang, L. Fulton, and T. A. Graves, 2009: The B73 maize genome: complexity, diversity, and dynamics. Science **326,** 1112—1115.

Trachsel, S., 2009: Genetic Analysis of Root Morphology and Growth of Tropical Maize and Their Role in Tolerance to Desiccation, Aluminum Toxicity and High Temperature. Doctoral and Habilitation Theses. ETH Zurich, Zürich.

Wang, Q., and H. K. Dooner, 2006: Remarkable variation in maize genome structure inferred from haplotype diversity at the *bz* locus. Proc. Nat. Acad. Sci. **103,** 17644—17649.

Wei, F., J. Zhang, S. Zhou, R. He, M. Schaeffer, K. Collura, D. Kudrna, B. P. Faga, M. Wissotski, W. Golser, S. M. Rock, T. A. Graves, R. S. Fulton, E. Coe, P. S. Schnable, D. C. Schwartz, D. Ware, S. W. Clifton, R. K. Wilson, and R. A. Wing, 2009: The physical and genetic framework of the maize B73 genome. PLoS Genet. **5,** e1000715.

Wen, W., S. Taba, T. Shah, V. H. Chavez Tovar, and J. Yan, 2010: Detection of genetic integrity of conserved maize (*Zea mays* L.) germplasm in genebanks using SNP markers. Genet. Resour. Crop Evol. **58,** 189—207.

Wolfgruber, T. K., A. Sharma, K. L. Schneider, P. S. Albert, D. H. Koo, J. Shi, Z. Gao, F. Han, H. Lee, and R. Xu, 2009: Maize centromere structure and evolution: sequence analysis of centromeres 2 and 5 reveals dynamic loci shaped primarily by retrotransposons. PLoS Genet. **5,** e1000743.

Xu, Y., 2010: Molecular Plant Breeding, 734. CABI Publishing, Wallington, UK.

Xu, Y., L. Zhu, J. Xiao, N. Huang, and S. R. McCouch, 1997: Chromosomal regions associated with segregation distortion of molecular markers in F2, backcross, doubled haploid, and recombinant inbred populations in rice (*Oryza sativa* L.). Mol. Gen. Genet. **253,** 535—545.

Yan, J., X. Yang, T. Shah, H. Sánchez-Villeda, J. Li, M. Warburton, Y. Zhou, J. H. Crouch, and Y. Xu, 2010: High-throughput SNP genotyping with the GoldenGate assay in maize. Mol. Breeding **25,** 441—451.

Yu, J., J. B. Holland, M. D. McMullen, and E. S. Buckler, 2008: Genetic design and statistical power of nested association mapping in maize. Genetics **178,** 539—551.

Zhu, Y. L., Q. J. Song, D. L. Hyten, C. P. Van Tassell, L. K. Matukumalli, D. R. Grimm, S. M. Hyatt, E. W. Fickus, N. D. Young, and P. B. Cregan, 2003: Single nucleotide polymorphisms in soybean. Genetics **163,** 1123—1134.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Integrated maize linkage map constructed using three RIL populations and physical map.

Figure S2. Relationship between physical and genetic distance in the map derived from 'XB' population.

Table S1. Information for three SNP linkage maps.

Table S2. Information of genetic linkage maps constructed using three RIL populations.

Table S3. Relationship between physical and genetic distances for maps derived from three RIL populations.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.