

Physical and Genetic Structure of the Maize Genome Reflects Its Complex Evolutionary History

Fusheng Wei^{1,2,3}, Ed Coe^{4,5}, William Nelson^{2,3,6}, Arvind K. Bharti⁷, Fred Engler^{2,3,6}, Ed Butler^{1,2,3}, HyeRan Kim^{1,2,3}, Jose Luis Goicoechea^{1,2,3}, Mingsheng Chen^{1,2,3}, Seunghee Lee^{1,2,3}, Galina Fuks⁷, Hector Sanchez-Villeda⁴, Steven Schroeder⁴, Zhiwei Fang⁴, Michael McMullen^{4,5}, Georgia Davis⁴, John E. Bowers⁸, Andrew H. Paterson⁸, Mary Schaeffer^{4,5}, Jack Gardiner⁴, Karen Cone⁹, Joachim Messing⁷, Carol Soderlund^{2,3,6*}, Rod A. Wing^{1,2,3*}

1 Arizona Genomics Institute, University of Arizona, Tucson, Arizona, United States of America, **2** Department of Plant Sciences, University of Arizona, Tucson, Arizona, United States of America, **3** BIO5 Institute, University of Arizona, Tucson, Arizona, United States of America, **4** Division of Plant Sciences, University of Missouri, Columbia, Missouri, United States of America, **5** Plant Genetics Research Unit, Agricultural Research Service, United States Department of Agriculture, Columbia, Missouri, United States of America, **6** Arizona Genomics Computational Laboratory, University of Arizona, Tucson, Arizona, United States of America, **7** Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers, The State University of New Jersey, Piscataway, New Jersey, United States of America, **8** Plant Genome Mapping Laboratory, Departments of Crop and Soil Science, Plant Biology, and Genetics, University of Georgia, Athens, Georgia, United States of America, **9** Division of Biological Sciences, University of Missouri, Columbia, Missouri, Arizona, United States of America

Maize (*Zea mays* L.) is one of the most important cereal crops and a model for the study of genetics, evolution, and domestication. To better understand maize genome organization and to build a framework for genome sequencing, we constructed a sequence-ready fingerprinted contig-based physical map that covers 93.5% of the genome, of which 86.1% is aligned to the genetic map. The fingerprinted contig map contains 25,908 genic markers that enabled us to align nearly 73% of the anchored maize genome to the rice genome. The distribution pattern of expressed sequence tags correlates to that of recombination. In collinear regions, 1 kb in rice corresponds to an average of 3.2 kb in maize, yet maize has a 6-fold genome size expansion. This can be explained by the fact that most rice regions correspond to two regions in maize as a result of its recent polyploid origin. Inversions account for the majority of chromosome structural variations during subsequent maize diploidization. We also find clear evidence of ancient genome duplication predating the divergence of the progenitors of maize and rice. Reconstructing the paleoethnobotany of the maize genome indicates that the progenitors of modern maize contained ten chromosomes.

Citation: Wei F, Coe E, Nelson W, Bharti AK, Engler F, et al. (2007) Physical and genetic structure of the maize genome reflects its complex evolutionary history. PLoS Genet 3(7): e123. doi:10.1371/journal.pgen.0030123

Introduction

Cereal crops, such as rice, maize, and wheat, are the major caloric source for humans and farm animals. The cereals shared a common ancestor some 50 million years ago (MYA) and even though their genome sizes vary considerably, their genetic map organization is highly conserved [1,2]. Rice (genome size = 389 Mb [3]) was the first cereal to have its genome completely sequenced and now serves as a reference sequence for comparative and functional genomics studies across the cereals. The 2,300-Mb [4] maize B73 genome is presently being sequenced by a clone-by-clone sequencing strategy. Although the maize genome behaves genetically as a simple diploid with ten pairs of chromosomes, its organization is quite complex. Early genetic analysis of duplicated genes suggested that maize had homoeologous regions [5–7]. This work was later supported by comparative restricted fragment length polymorphism (RFLP) mapping across the cereals, which showed that two maize chromosome sets aligned with one chromosome each of rice and sorghum, thereby demonstrating a whole genome duplication event [1,8]. Evolutionary analysis of duplicated genes indicated that maize may have arisen by allotetraploidy, suggesting that maize was formed by the hybridization of two slightly diverged progenitors rather than the duplication of a single

progenitor [9]. While cytogenetic studies suggested that the progenitors of maize were species consisting of five chromosomes [10], alignments of linkage maps of rice and maize hypothesized that the progenitors had eight chromosomes [11]. Furthermore, these maps also indicated that the ancient duplications in rice were also present in the progenitors of maize, suggesting that the duplications occurred before the split of the progenitors of rice and maize [12–15], and it is likely that they would be shared by most if not all cereals [16].

Editor: Joseph R. Ecker, The Salk Institute for Biological Studies, United States of America

Received February 22, 2007; **Accepted** June 11, 2007; **Published** July 20, 2007

A previous version of this article appeared as an Early Online Release on June 12, 2007 (doi:10.1371/journal.pgen.0030123.eor).

Copyright: © 2007 This is an open-access article distributed under the terms of the Creative Commons Public Domain declaration, which stipulates that, once placed in the public domain, this work may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose.

Abbreviations: BAC, bacterial artificial chromosome; BES, BAC end sequence; chr, chromosome; EST, expressed sequence tag; FPC, Fingerprinted contig; HICF, high information content fingerprint; IBM, intermated B73 × Mo17; MYA, million years ago; RFLP, restricted fragment length polymorphism; SSR, simple sequence repeat; SyMAP, synteny mapping and analysis program

* To whom correspondence should be addressed. E-mail: cari@agcol.arizona.edu (CS); rwing@ag.arizona.edu (RAW)

Author Summary

As a cash crop and a model biological system, maize is of great public interest. To facilitate maize molecular breeding and its basic biology research, we built a high-resolution physical map with two different fingerprinting methods on the same set of bacterial artificial chromosome clones. The physical map was integrated to a high-density genetic map and further serves as a framework for the maize genome-sequencing project. Comparative genomics showed that the euchromatic regions between rice and maize are very conserved. Physically we delimited these conserved regions and thus detected many genome rearrangements. We defined extensively the duplication blocks within the maize genome. These blocks allowed us to reconstruct the chromosomes of the maize progenitor. We detected that maize genome has experienced two rounds of genome duplications, an ancient one before maize–rice divergence and a recent one after tetraploidization.

Based on recent alignments of orthologous regions of rice and sorghum chromosomes with two homoeologous regions in maize, the progenitor of sorghum and the two progenitors of maize diverged at nearly the same time about 11.9 MYA. Furthermore, the two progenitors of maize may have hybridized as recently as 4.8 MYA, explaining the divergence of the duplicated regions in maize [17]. Alignments of chromosomal regions shared by common ancestry also showed that frequently one copy of the duplicated genes was lost [18,19].

Sequence analysis of random sheared fragments [20], BAC (bacterial artificial chromosome) end sequences (BESs) [21], and complete BAC sequences from 100 random regions [22] showed that about two-thirds of the maize genome consists of transposable elements, with 95% of those being retrotransposons. Because long terminal repeats are identical at the time of insertion, it was possible to determine the phase of expansion. It appears that retrotranspositions emerged after speciation of maize with the majority of elements occurring less than 1 MYA [23].

Genome size, reduplication, and high repetitive DNA content make maize an extremely challenging genome to sequence, even with the availability of the rice genome as a reference sequence. Sequencing projects using a clone-by-clone strategy, as well as many whole genome shotgun projects, require the availability of a high-resolution physical map. Here, we present the final assembly and analysis of the physical map of *Zea mays* ssp. *mays* cv. B73, which is comprised of many complementary resources including: (1) deep-coverage large insert BAC libraries [24,25]; (2) agarose and high information content fingerprint (HICF) datasets assembled with fingerprinted contig software (FPC) [26,27]; (3) a genetic marker dataset to anchor the physical map to the maize genetic map; and (4) sequence-based marker datasets (overgos [28] and BESs [21]) to infer contig position and orientation with respect to the rice reference sequence [3]. Alignment of the two genomes allowed us to shed new light on the organization of the chromosomes of the progenitors contributing to the maize polyploidization, unravel rearrangements associated with the subsequent diploidization of the maize genome, and infer the sizes and locations of homoeologous regions together with spatial patterns of genome size evolution. This also sets the stage for careful comparison of maize to sorghum and to begin to dissect the shared versus

lineage-specific consequences for these genomes of an ancient duplication common to all cereals.

Results/Discussion

Integrated Physical and Genetic Map of the Maize B73 Genome

A total of three deep-coverage large-insert BAC libraries covering ~30 genome equivalents were constructed using HindIII, EcoRI, and MboI digests of high molecular weight DNA isolated from maize inbred line B73 [24,25]. We used two different methods to fingerprint the same set of BAC clones, referred to as agarose [29] and HICF [30,31], to cross-confirm the assembled physical contigs of homoeologous regions and highly conserved sequence families. The agarose method resulted in 292,201 successful fingerprints that were automatically assembled into 4,518 contigs using FPC [26] at a Sulston score of $1e-12$ and tolerance of seven (Table S1). HICF generated 350,253 successful fingerprints that automatically assembled into 1,500 FPC contigs [27]. This was accomplished by assembling the initial contigs using a Sulston score of $1e-70$ and then automatically merging contigs with a Sulston score of $1e-21$.

A total of 25,908 markers were integrated into the FPC map that included 1,902 genetic markers and 24,006 overgo/BES markers. In addition to providing valuable genetic and biological information, these markers were also extremely useful in the validation of FPC contig assemblies. A total of 2,036 genetic markers including 1,307 SSRs (simple sequence repeat), 372 RFLPs, 189 SNPs, and 168 insertion/deletions were used to integrate the FPC map with the maize genetic map (Table S2). After integration, 1,902 genetic markers (93.4%; 1,240 SSRs, 345 RFLPs, 169 SNPs, and 148 insertion/deletions) could be placed confidently on the physical map (Table 1; Table S3 for the distribution of these markers). These included 822 well-ordered IBM (intermated B73 × Mo17) framework markers (94.6% success rate) and 1,080 binned IBM-neighbor placement markers (see Methods for marker definitions; 92.5% success rate) (Table S2).

In addition to the genetic markers, 24,006 additional sequence-based markers were integrated into the maize FPC map. Although these markers were not genetically mapped, they proved to be extremely powerful by providing sequence tag sites that could be used to order and orient contigs relative to the rice reference sequence [3]. Sequence-based markers included 3,438 privately contributed EST (expressed sequence tag)-derived unigene markers, 9,371 overgos derived from 70,716 maize ESTs [28], 2,068 highly conserved cereal sequences, and 9,129 gene-containing BESs [21]. As shown in Table 2, all marker types appear to be distributed across the maize chromosomes roughly in proportion to the sizes of the respective chromosomes.

In manual editing of the agarose physical map, four sources of evidence were assessed to indicate contig placement, orientation, or merging—the agarose FPC map, markers, the HICF FPC map, and synteny between maize and rice (see below). Only if evidence from three of the four criteria was met would two contigs be merged. The final FPC map resulted in 721 contigs covering 2,150 Mb (equal to 93.5% of the 2,300-Mb genome; see Methods for detailed calculations).

Of the final 721 contigs, 421 are anchored to the maize genome using 1,902 genetic markers. The anchored contigs

Table 1. Maize Genetic Markers

Chr	FPC Genetic Marker		IBM2	IBM2 Neighbor	IBM2 Neighbor Frame	
	Anchored					Total ^b
	Number ^a	Percent				
1	299	93.1	321	342	922	621
2	216	95.6	226	201	650	426
3	213	91.8	232	244	673	427
4	206	94.9	217	206	717	420
5	194	93.7	207	203	587	408
6	167	94.4	177	170	485	328
7	161	94.2	171	169	497	313
8	160	92.0	174	153	511	302
9	160	94.1	170	176	450	313
10	126	89.4	141	157	391	260
Total	1,902	93.4	2,036	2,021	5,883	3,818

^aNumber of markers anchored to physical contigs.

^bTotal number of markers in this study, including markers that failed to anchor contigs. doi:10.1371/journal.pgen.0030123.t001

cover 1,981 Mb, equal to 86.1% of the maize genome (Table 2). Of the remaining 300 unanchored contigs (7.4% of the genome), 189 contain fewer than ten BAC clones each. The average sizes of anchored and unanchored contigs were 4.7 Mb and 0.56 Mb, respectively. The longest anchored contig was 22.9 Mb on Chromosome 9 (chr9), while the longest unanchored contig was 6.7 Mb. The individual chromosome coverage varied from 94% on chr5 to 65% on chr9. The low coverage found on chr9 does not, perhaps, reflect the real situation due to the fact that 94.1% of the genetic markers could be placed on the FPC map for this chromosome. This discrepancy could possibly be explained by genome size disparity found in various maize cultivars, such as differences

in heterochromatic regions. The chromosome sizes used here, the only available maize data, were derived from sweet corn (Seneca 60) [32]. The fact that the largest anchored contig is on chr9 (22.9 Mb) may indicate that maize chr9 in B73 is highly euchromatic.

The average physical to genetic ratio in the maize genome is 182 kb per cM with great variation, ranging from >1.8 Mb/cM in centromeric regions to <10 kb in telomeric region. The rich marker information integrated into the physical map allowed us to plot marker distribution along the maize chromosomes. When compared with genetic marker distribution (Figure S1), we found that EST-derived markers are directly correlated with genetic markers, indicating a tight association of gene distribution and genetic recombination.

The maize physical map can be accessed at <http://www.genome.arizona.edu/fpc/maize>.

Maize–Rice Syntenic Block and Its Implications

Similarities in gene order between rice and maize have been extensively reported [1,11,15,33–36]. However, these studies used low-resolution genetic maps or incomplete rice sequences. Now, using the rich marker information embedded in our integrated maize physical map and the rice reference sequence [3], we were able to build a high-resolution comparative physical map between the maize and rice genomes. Using synteny mapping and analysis program (SyMAP) software [37] we generated a dotplot between the integrated maize physical map and the 12 rice pseudomolecules and then computed syntenic blocks (see Methods) with consideration of marker density and position within the integrated maize map (Figure 1). As shown in Table S4, 52 major maize–rice collinear blocks could be identified. Previous studies suggested 20 chromosomal rearrangements in maize compared to rice but could not provide accurate sizes of syntenic blocks because of the paucity of markers [11]. These 52 maize blocks varied in size from 760 kb to over 82

Table 2. Summary of the Maize Physical Map

Overall	Chr	Contigs	Contig Range	Length (kb)	Real (kb)	Percent	Overgoes					
							CL	PCO	si	dd	SOG	Total
	1	67	1–67	290,381	337,000	86.2	1,342	2,064	584	1,488	3,017	8,495
	2	43	68–110	234,774	278,000	84.5	1,202	1,611	488	1,206	2,368	6,875
	3	43	111–153	223,965	259,000	86.5	1,021	1,455	470	1,083	2,168	6,197
	4	50	154–203	254,675	271,000	94.0	1,045	1,507	469	1,101	2,288	6,410
	5	52	204–255	232,573	250,000	93.0	1,089	1,630	458	1,089	2,348	6,614
	6	36	256–291	162,307	208,000	78.0	778	1,122	328	854	1,666	4,748
	7	34	292–325	152,162	199,000	76.5	682	990	318	834	1,421	4,245
	8	41	326–366	166,918	202,000	82.6	837	1,190	354	795	1,696	4,872
	9	25	367–391	123,588	191,000	64.7	599	864	257	705	1,214	3,639
	10	30	392–421	139,447	170,000	82.0	653	913	244	693	1,352	3,855
Total		421	NA	1,980,790	2,365,000	86.1 ^a	9,248	13,346	3,970	9,848	19,538	55,950
Actual overgo number							3,075	4,819	1,425	3,438	2,068	14,825
Number of copy in the genome							3.0	2.8	2.8	2.9	9.4	3.8

^aBased on 2,300-Mb genome of B73 [4], instead of the 2,365-Mb genome of Seneca 60 (sweet corn [32]), in which chromosome size is the only available maize data. The numbers under Overgoes are the number of probes in each overgo type. The number difference between the map and the actual number is due to multiple copies of each sequence in the genome. SOG probes were designed for sequence conservation and can hit gene families resulting in higher gene copies.

CL, clusters assembled from public EST sequences; dd, anonymous clusters assembled by DuPont (sequences are not available); NA, not applicable; PCO, public sequences combined with DuPont sequences; si, public singletons that did not cluster with DuPont sequences; SOG, overgos from the Paterson lab derived from probes that have been mapped in sorghum and other grasses.

doi:10.1371/journal.pgen.0030123.t002

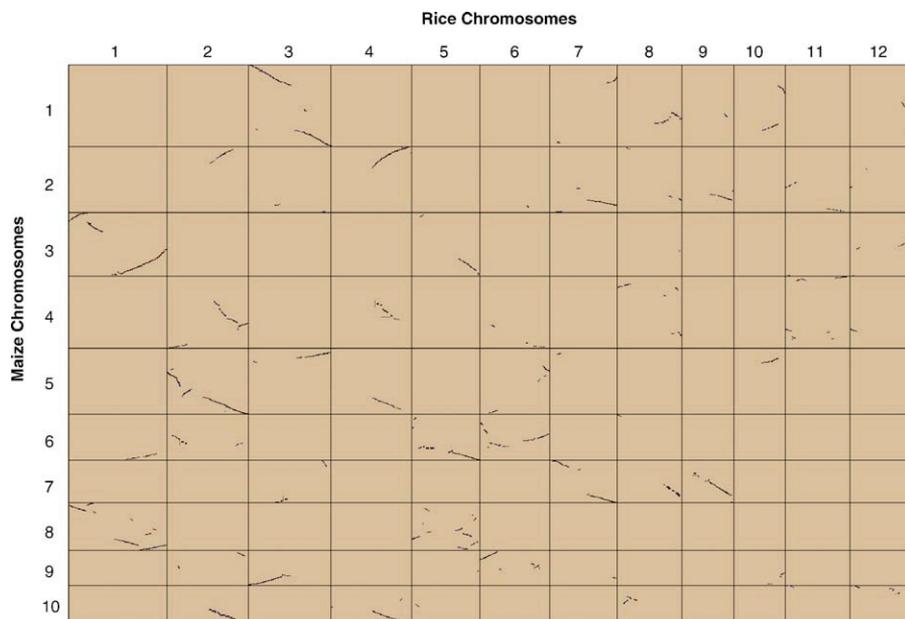


Figure 1. Dotplot Analysis of the Integrated Maize Map against Rice Pseudomolecules

Syntenic blocks were detected, and background noise was filtered with SyMAP [37]. The interactive dotplot can be viewed at <http://www.agcol.arizona.edu/symap>. When clicking the related syntenic block, the detailed window with contig number will pop up. The viewer can select the preferred area and double click the selection, and then a graphic alignment is displayed. doi:10.1371/journal.pgen.0030123.g001

Mb, and their corresponding rice regions fluctuated from 360 kb to over 22 Mb.

Although the maize genome is about 6-fold larger than rice, in collinear regions on average, every kilobase of rice sequence expands to only about 3.2 kb of sequence in maize, with a range of 1.1–9.5 kb in different regions (Table S4). This uneven expansion of syntenic blocks appears to be due to differential loss of genic regions and insertion of retrotransposable elements as recently shown by sequence comparison of a pair of homoeologous regions of maize chr1S and chr9L with rice chr3S [18]. Since on average one region in rice corresponds to two maize regions, one would expect the expansion factor for each maize region to be half of the total genome expansion. Therefore, the derived factor of 3.2 is consistent with the whole genome duplication event described in Figure 2. The maize syntenic regions (1,474.5 Mb in total) cover 74.5% of the anchored map (64% of the genome). The regions in Figure S2A that do not show maize–rice correspondence appear to be heterochromatic, also as reported for the sorghum–rice comparison [38].

An interactive display of the rice–maize synteny is available at <http://www.agcol.arizona.edu/symap>.

Genome Duplication and Genome Rearrangement during Maize Evolution

From the dotplot in Figure 1, we observed that in addition to the primary syntenic blocks in rice (Table S3) each maize syntenic block corresponds to another syntenic region in rice, consistent with previous reports [12–15] that used translated protein sequences to show ancient duplications in rice, which were phylogenetically dated before the divergence of most if not all cereals [16]. Whereas they used rice-to-rice comparisons, we used the maize alignment to rice to detect syntenic blocks that elucidated the rice duplication.

The advantage of comparing rice to maize instead of rice to rice is that all ancient duplications that predate the ancestor of rice and maize should also be present in the progenitors of maize. Therefore, if maize arose from two progenitors by allotetraploidization, segmental duplications should be present in four copies. Indeed, our high-resolution integrated physical map shows examples of such segmental quadruplications and allows us to more precisely determine their coordinates. Using the rice genome as a reference to investigate the maize genome, we found that each syntenic region in rice has three to four corresponding regions in maize (Figures 1 and S2; Table 3). The orthologous maize–rice regions (primary blocks) have higher density syntenic alignments than the “paleologous” regions resulting from ancient duplications (secondary blocks). Figure 2A shows the primary synteny of the long arm of rice chr2 with maize chr4 and chr5, and Figure 2B shows secondary synteny of the same rice chr2 region with maize chr2 and chr10 (see also Figure S2B). Figure 2C and 2D shows that the long arm of rice chr4 is primarily syntenic with maize chr2 and chr10 and secondarily syntenic to maize chr4 and chr5 (also see Table 3). Figures 1 and S2A show that maize chr2, chr4, chr5, and chr10 all align to the same region of rice chr2 and chr4. These results are consistent with previous reports that showed the long arms of rice chr2 and chr4 to be syntenic to one another [12,13,15]. Table 3 summarizes all the primary and secondary duplication blocks in maize on the basis of our analysis. In addition, the detection of an average of around three copies for each overgo (Table 2) further supports the idea that the present maize genome emerged from tetraploidization of several ancient duplicated regions. On the other hand, formation of duplications in the rice genome is not limited to the ancestor of rice and maize. The most recent segmental duplication in rice of 3 Mb on the tips of rice chr11 and chr12 is estimated

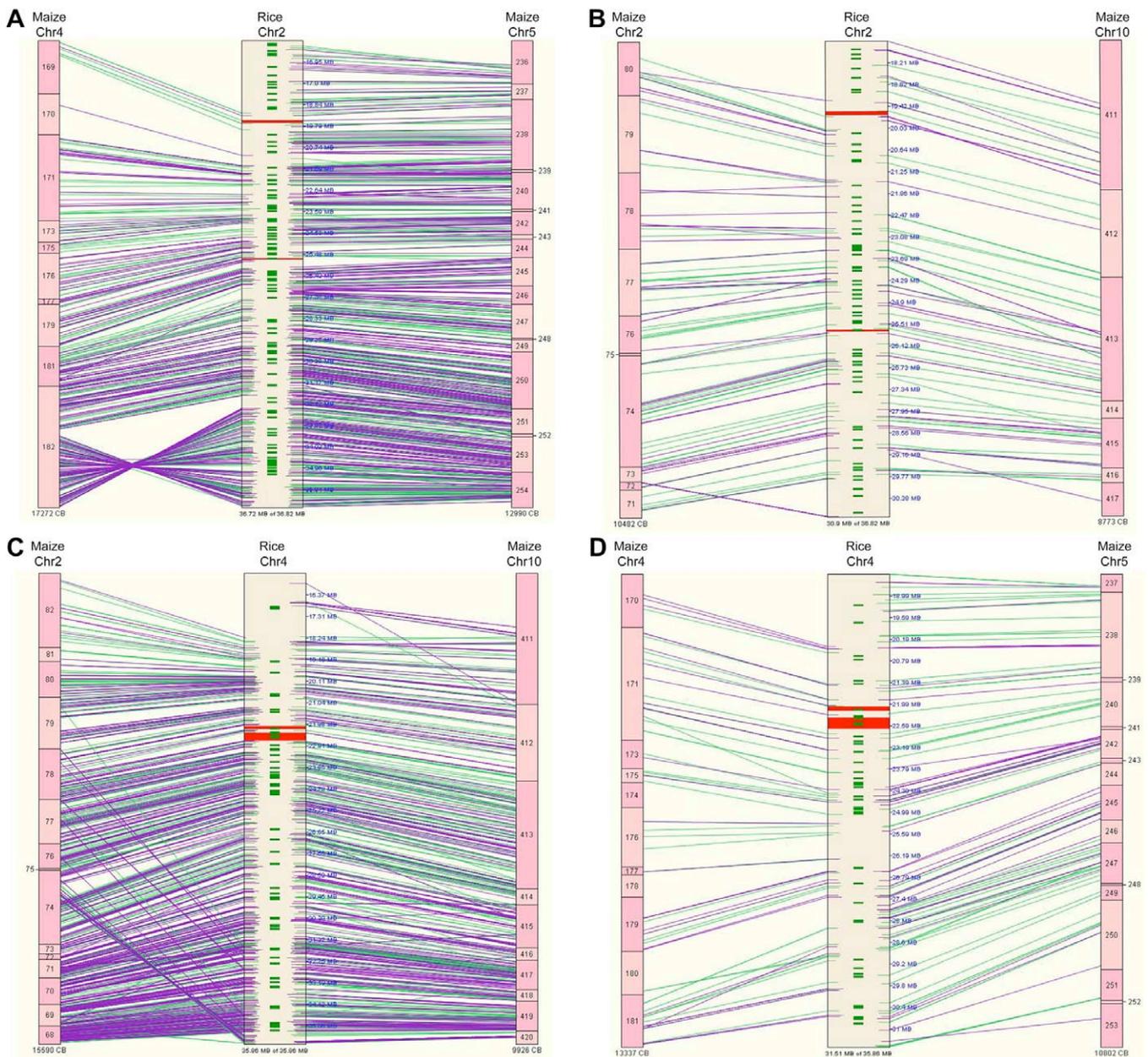


Figure 2. Graphic Display of Two Rounds of Maize Genome Duplication

The picture was captured from the comparative block display of the integrated maize–rice synteny map generated by SyMAP [37]. Green alignment line showed result from overgo markers, and violet line showed alignment from low-copy BES.

(A) Recent duplication resulted from ditetraploidization of maize chr4L and 5L with reference to rice 2L.

(B) Ancient duplication before maize and rice divergence of maize chr2S and chr10L with reference to rice 2L is presented.

(C and D) show that with reference to rice chr4L, recent and ancient duplications have occurred of maize chr2S and chr10L, and chr4L and chr5L, respectively.

doi:10.1371/journal.pgen.0030123.g002

to be about 7.7 MYA [39], well after rice–maize divergence. This illustrates a general model that small segmental duplications (such as the 7.7 MYA one) continue to arise over long periods of time unlike a whole genome duplication event and therefore contribute to the uniqueness of each genome [14,39]. Multiple duplication events may also contribute to the maize-complex duplication pattern reported previously [11,40], in which the recent maize duplication is much easier to be identified than the ancient genome duplication in a cereal ancestor.

Our synteny analysis also shows that several genome rearrangements have occurred since the divergence of maize and rice. As shown in Table S5, we identified 62 rearrangements (>125 kb in corresponding rice sequence) across all chromosomes, which account for 281.2 Mb (19%) of the maize 1,474.5 Mb syntenic sequence with a block size variation from 440 kb to over 32 Mb. These rearrangements include 39 inversions (207.2 Mb), 14 inversion and translocation events (56.7 Mb), and eight translocations (16.3 Mb). Most of these rearrangements were detected near telomeres.

Table 3. Two Genome Duplication Events in Maize Genome Evolution

Block Pairs	Recent Duplication				Rice Syntenic Region	Ancient Duplication				
	Region 1		Region 2			Region 1		Region 2		Rice Ancient Duplication Region
	Physical Block ^a	Chr	Physical Block ^a	Chr		Physical Block ^a	Chr	Physical Block ^a	Chr	
1S	111–120	3	326–334	8	1S					
1L	125–151	3	354–366	8	1L	285–290	6	336–351	8	5L
2S	196–203	4	223–236	5	2S	271–277	6	376	9	6L
2La	169–181	4	234–250	5	2L	71–80	2	411–417	10	4L
2Lb	181–184	4	251–254	5	2L	280–281	6	367–373	9	6L
3S	1–19	1	383–391	9	3S	104–105	2	323–325	7	7L
3L	50–67	1	202–210	5	3L	109	2	293–299	7	7S
4S	86	2	409–411	10	4S					
4L	68–82	2	411–420	10	4L	170–181	4	237–253	5	2L
5Sa	260–268	6	354	8	5S	405–409	10			Unknown
5Sb	281–283	6	338–340,348–349	8	5S	329–340	8			1S
5L	283–291	6	345–353,331–336	8	5L	132–151	3	355–366	8	1L
6S	368–374	9	265–270, 279–281	6	6S	182–184	4	251–254	5	2L
6La	270–281	6	374–380	9	6L	195,198–199	4	223–229	5	2S
6Lb	218–223	5	270–271	6	6L	199–201	4	223	5	2S
7S	109–110	2	292–301	7	7S	57–65	1	202	5	3L
7L	100–106	2	320–325	7	7L	11–19	1	385–387	9	3S
8Sa	164	4	256–260	6	8S					
8Sb	162–164	4	399–405	10	8S	69–70	2	417–418	10	4L
8L	37–44	1	164–193	4	8L	91–103	2	307–320	7	9L
9L	91–103	2	304–320	7	9L	38–44	1			8L
10La	45–49	1	212–219	5	10L	4–5	1			3S
10Lb	20–27	1	382–385	9	10L	1–4	1			3S
11S	89–91	2	183–187,193–195,157–159	4	11S	392–393	10	151	3	12S
11L	108–109	2	160,188,194,154–157	4	11L					
12S	392–395	10	121,125–126,151–152	3	12S	184–187	4	80–91	2	11S
12La	124	3	394–398	10	12L					
12Lb	30–37	1	121–124	3	12L	51–55	1	207–210	5	3L

^aFPC contigs have been numbered sequentially along each chromosome. SyMAP defined “blocks” on chromosomes that have contiguous contigs synteneous with rice. doi:10.1371/journal.pgen.0030123.t003

A total of 12 of the 20 maize telomeric ends or 15 of 24 rice telomeric ends have undergone rearrangements in one way or the other; most inversions may be due to high recombination frequency in the regions. It is rare to find interchromosomal rearrangement, supporting the role of recombination in these macrorearrangements. Although the majority of rearrangements can be confirmed using our integrated map, it is possible that a few may have resulted from fingerprint or sequence misassembly in maize or rice, respectively. The maize genome-sequencing project as well as rice and maize optical maps will likely resolve such discrepancies in the near future.

Chromosome Reconstruction of Maize and Its Ancestors

The 52 major maize–rice collinear blocks (Table S4) and 28 maize duplication block pairs (Table 3) allow us for the first time to demonstrate the extent of chromosome breakage and fusion that occurred in the maize genome after chromosome doubling from its two progenitors. Furthermore, with the assistance of the rice reference sequence [3], we can propose a model for the chromosome structure of the progenitors of maize and their ancestors.

When the maize homoeologous regions are reassembled into homoeologous chromosome pairs and are then aligned with homologous rice chromosomes (Figure 3A), a picture

emerges that the two progenitors of maize each consisted of ten chromosomes resulting in ten chromosomes after hybridization. The most parsimonious explanation is that the progenitor chromosomes underwent breakage and fusion and eventually reassembled into ten new mosaic chromosomes that are genetically diploid as opposed to being allotetraploid. Therefore, it is most likely that the syntenic blocks discovered here reflect the new junctions that were formed in this process. One possible explanation for the formation of a $2N = 20$ diploid, composed of reshuffled chromosomes rather than a $2N = 40$ allotetraploid after hybridization, is that maize lacks the equivalent of the *Ph1* product/structure that prevents pairing of homoeologous chromosomes in wheat [41]. There are examples where allopolyploidy does not result in reshuffling of chromosomes as in the case of polyploid wheat. It is interesting to note that the *Ph1* locus in wheat prevents the pairing of nonhomologous chromosomes. One can envision that the lack of such a function in maize has resulted in alignments of nonhomologous chromosomes after allotetraploidy and contributed at least in part to their rearrangements. In addition, a burst of transpositions of transposable elements or ectopic recombination or conversion might have triggered chromosome breakages and fusions on a larger scale. However, these are just possible scenarios, and perhaps sequences of multiple

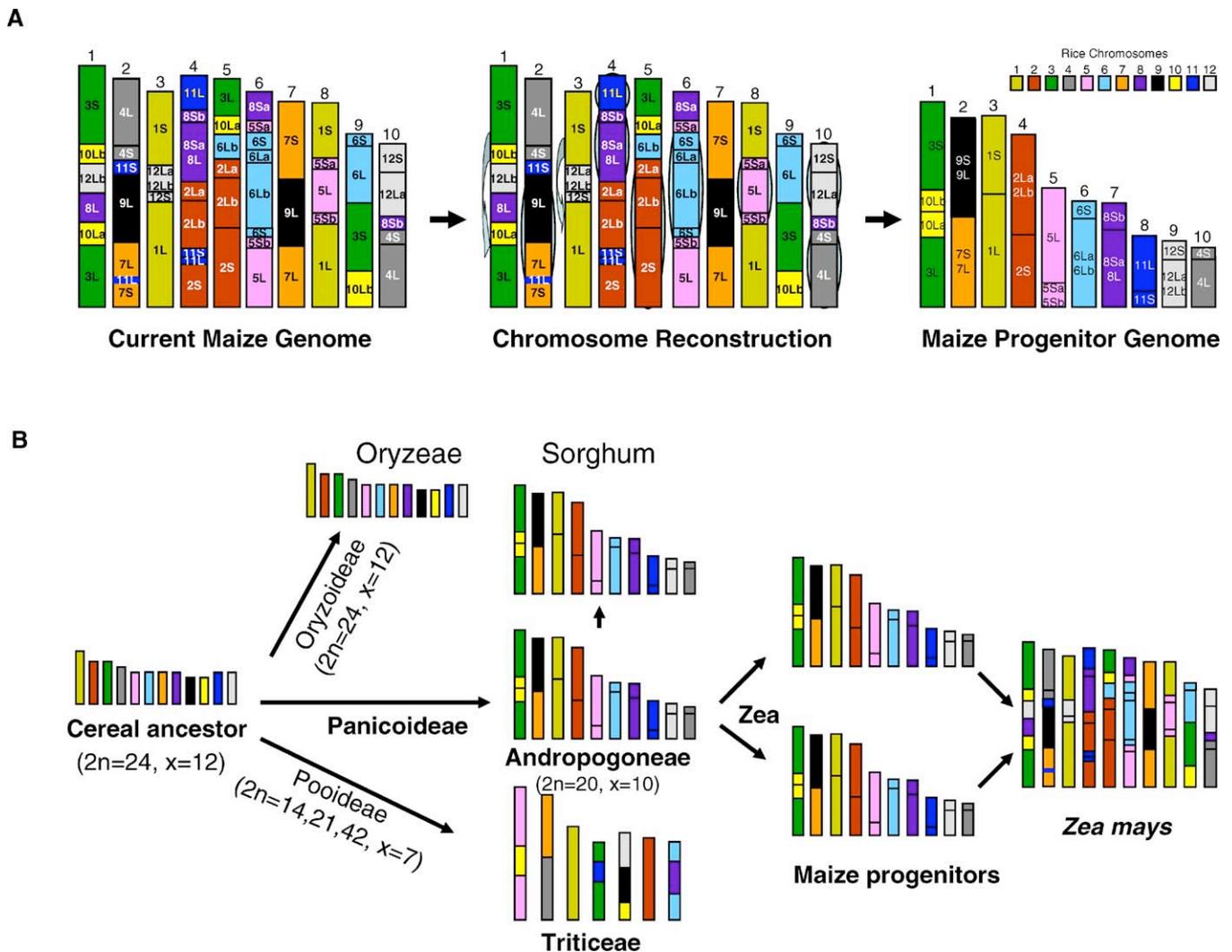


Figure 3. Change of Chromosome Numbers during Speciation of Cereals

Based on the dotplot comparison of rice, maize, sorghum, and wheat chromosomes, synteny blocks have been used to assemble progenitor chromosomes of these species. Rice synteny blocks have been color coded.

(A) Using the rice color-code syntenic block from Table 3 and Table S4, the chromosomes of the progenitors of maize have been reconstructed. The block names in the figure are the same in Table 3. No change in chromosome number occurred, but an increase of maize chromosome sizes did.

(B) Comparison of the relationship of the maize progenitors with sorghum and wheat has been used to reconstruct the changes and conservation of chromosomes during speciation.

doi:10.1371/journal.pgen.0030123.g003

inbred lines could be more informative about the possible mechanisms.

When compared to the rice–sorghum synteny, the ten maize progenitor chromosomes appear to be the same as the ten sorghum linkage groups (Figure 3B). After divergence from rice, the common ancestor of maize and sorghum combined rice chr3 and chr10 to form one chromosome and rice chr7 and chr9 to form another chromosome. This resulted in ten chromosomes in the maize and sorghum ancestors (Figure 3A). Using a more limited dataset, Wilson et al. [11] proposed that in addition to the above chromosome fusions, two additional fusions occurred—one between rice chr1 and chr5 and the other between rice chr4 and chr12—hypothesizing an eight-chromosome model for the progenitors of maize. Our data did not support these two fusions. By comparison of this rice–Andropogoneae synteny to synteny

between rice and wheat [42], we found there are no common chromosome combinations with respect to the rice genome after divergence from the common ancestor. These data suggest that the rice genome may represent the ancient ancestral form of cereal genomes (Figure 3B). Clearly, the high-density gene map of maize has raised the level of our understanding of the role of speciation in the chromosome evolution in plants.

Methods

DNA fingerprinting and band analysis. The three BAC libraries (30× total coverage) used in this study were HindIII (136-kb average insert size), EcoRI (163-kb), and MboI (167-kb) libraries [24,25], with 14.2×, 7.6×, and 7.8× coverage respectively. The use of three BAC libraries helped to ensure the maize genome was fully represented. Maize cultivar B73 was selected because of its widespread use in

breeding and because it is one of the parents of a public high-resolution genetic reference mapping population (IBM [43]).

BAC clones starting with a “b” are from the B73 HindIII library (ZMMBBb). BAC clones starting with a “c” are from the B73 EcoRI and MboI libraries known as CHORI-201 (Clemson University Genomics Institute/Arizona Genomics Institute [CUGI/AGI] name ZMMBBc). In the c library, the first half (288 384-well plates) was from the EcoRI digest, and the second half was from the MboI digest. Clones ending in “sd1” are sequenced BACs downloaded from GenBank (<http://www.ncbi.nlm.nih.gov/Genbank>) digested with HindIII in silico [44]. If the sequences were longer than 150 kb, they were artificially fragmented into multiple overlapping clones, i.e., sd1, sd2, etc. and then digested with HindIII in silico. For the agarose method [29,45], all BAC DNAs were digested with the restriction enzyme HindIII, run on high-resolution 1.0% agarose gels, stained with SYBR Green (FMC BioProducts, <http://www.fmc.com>), and the Image software [46] was used for interactive band calling.

For the HICF method [27,30,31], BAC DNAs were digested with type IIS restriction enzyme *EcoRI*, in which the resulting ends were tagged with fluorescently labeled ddNTPs and TaqI, which is to reduce the fragment sizes of the *EcoRI* digestion. The first base in a two-base overhang from TaqI is G, therefore ddCTP was not used in the end-labeling reaction because the majority of fragments would contain at least one TaqI site. Thus we only used C, T, and A overhangs for end labeling by using the fluorescent dyes ddGTP (blue), ddATP (green), and ddTTP (yellow), respectively. Red dye (GeneScan-500 ROX, Applied Biosystems, <http://www.appliedbiosystems.com>) was employed as an internal size standard, which was run in each capillary. Reaction products were resolved on ABI3700 automated DNA sequencers. ABI GeneScan version 3.7.1 software (Applied Biosystems) was used for band detection and extraction. Only fragments ranging in size from 75 to 500 bp were used for HICF assembly. Because FPC can only input one set of band values for each clone, we used the technique of Ding et al. [31] to convert the size/color pairs generated by HICF. Each band was multiplied by 20, and the fractional part discarded. An offset was then added to each band as follows: 0 to blue, 10,000 to green, and 20,000 to yellow bands. Since we used fragments in the range 75–500 bp, the result of this conversion was bands occupying the ranges 1,500–10,000, 11,500–20,000, and 21,500–30,000.

Contig assembly and manual editing. FPC software [26] was used for fingerprint assembly. To automatically build the agarose map, two clones assemble together if they share at least N markers and have a Sulston score of less than M , using the following (N, M) pairs: (0, 1e–12), (1, 1e–11), (2, 1e–10), or (3, 1e–09). False-positive overlaps generally result in a stack of clones within the contig whose bands do not align well with the underlying consensus band map; these clones are called Q (questionable) clones [47]. The FPC DQer function reassembles all contigs that have >5 Q clones using a more stringent cutoff, such as 1e–13, 1e–14, or 1e–15. If the more stringent cutoff breaks the contig into multiple contigs, the function tries to join the subcontigs by their end clones; if this is not possible, it creates new contigs from the subcontigs. If it cannot reduce the number of Qs by a 1e–15, it leaves the contig alone. Generation of the maize FPC map was performed in stages including manual editing that utilized the agarose assembly of 4,518 contigs and all genetic and overgo markers. The addition of these markers allowed us to reduce the Sulston score for contig merging. Singletons were added in to merge contigs when they overlapped with the ends of two contigs. The total contig number decreased from initial build of 4,518 to 2,085, and the assembly was publicly released in February 2004 (Table S1).

To automatically build the HICF map, possible contaminated clones were screened for same-plate overlaps at 1e–45, followed by removal of all remaining clones over 175 bands. The initial contigs were built at 1e–70 at the tolerance of 4, then the DQer reassembled all contigs that had $>15\%$ Q clones. The end-merging function was run at 1e–61, 1e–52, 1e–45, 1e–40, and 1e–21, where the end-merging function checks the clones at the ends of contigs, and if at least two pairs of clones overlap based on the given cutoff, the contigs are merged. Singletons were placed at 1e–43 to their best location, before the final end-merge.

The agarose map was edited to split contigs with false-positive joins and to merge contigs that had overlapping end clones that were too small to be detected by FPC. First, manual end joining was performed if the end clones from two contigs had a Sulston score of N and at least M shared markers, where the valid (N, M) pairs were (0, 1e–10), (1, 1e–09), (2, 1e–08), or (3, 1e–07). We also extensively used the HICF map, marker information, and maize–rice synteny for contig merging. A new marker type called HICF was created, which was used to link each pair of clones that are bridged by an HICF clone

that did not occur in the agarose map. We know that the pair of clones lie close to each other (within 150 kb) and probably overlap, but it cannot be confirmed. The agarose method produced an average of 30 bands for each maize clone (band size range from 1 to 23 kb), while the HICF method generated an average of 107 bands for each clone (size range 75 bp–500 bp [27]).

The metric of a FPC contig is the consensus band (CB) unit. If the length of a contig is N CB units, then its approximate length in base pairs is $N \times 4,900$. The 4,900 was calculated as the average size restriction fragment as follows: A total of 20 completely sequenced maize BACs that had corresponding fingerprints were downloaded from GenBank. The average size of all sequenced BACs was divided by the average number of fingerprinted bands, which resulted in an average band size of 4,900 bp. Since the number of bands was taken from the real fingerprint and not from the simulated fingerprint, this average size took into account missing bands. This estimation is similar to that in the rice physical map [45].

Genetic mapping. As an essential component to establish an integrated map for the maize genome, a saturated and evenly distributed genetic map was developed to anchor and fully orient physical contigs along the chromosome. We built a dense genetic framework using the IBM mapping population [43]. The inbred parents of IBM are highly polymorphic and the progeny experienced four rounds of intermating, which makes IBM nearly four times expanded in size and 18-fold higher in resolution than previous public standard populations [48]. The order of loci on the IBM map has proven to match closely that developed by fingerprinting and BAC assembly.

The number of genetically mapped markers was increased by intercalating locations of additional markers into the IBM, constructing “neighbors” maps (Table 1), i.e., by importing markers from maps of other mapping populations and interpolating for genetic coordinates [49]. Successive iterations of neighbors’ maps have incorporated new map data from other projects and have refined orders according to FPC results. Maps have been updated regularly in MaizeGDB [50].

Integration of genetic markers with the FPC map. We used probe hybridization on BAC filters, primer amplification methods on 6-D BAC pools, and overgo probe hybridizations to integrate genetically mapped markers with the related BAC clones in the maize physical map. Since genetic markers can hybridize to multiple contigs because of paralogous sequences, it was necessary to distinguish between marked contigs that could be genetically placed and all other copies. To this end, anchoring markers retained their original name, and any unanchored copies were assigned a suffix (.A). Markers included RFLPs, SSRs, ESTs, SNPs, and InDels. Filter hybridization has low throughput, so we used only 90 RFLP probes on filters [24] and carried out further integrating with two high-throughput methods.

Hybridization of overgos was conducted on four filters of a 6 × 6 grid with a total of 165,888 BACs, half from each of the HindIII and EcoRI libraries. The 40-bp overgo probes were designed from unigenes of the DuPont/MMP/Incyte Genomics partnership and were screened with a maize repeat database to ensure that each probe was low copy. The overgos were classified into five marker types where the prefix identifies the marker type: (1) the prefix CL identifies clusters assembled from public EST sequences; (2) PCO represents public sequences combined with DuPont sequences (Unigene Consensus deposited in GenBank); (3) dd signifies anonymous clusters assembled by DuPont (sequences are not available); (4) si represents public singletons that did not cluster with DuPont sequences; and (5) SOG markers are overgos from the Paterson lab, derived from probes that have been mapped in sorghum and other grasses. Except for the SOG markers, which were designed for sequence conservation, all other overgos were designed for their diversity. We employed a 2-D 24 × 24 pool strategy for overgo hybridization and scored dual grid spots as positive clones [28].

PCR methods were used to screen BACs with SSR or genomic sequence markers. A 6-D pool strategy [51] was used for library screening to construct 288 pools from a 48 × 48 × 48 block for a total of 82,944 BACs in 216 384-well plates from the HindIII library. Primers for selected sequences were applied to pools, and the data were deconvoluted using criteria that assured disambiguation by multiple confirmation.

Maize–rice synteny analysis. We used the SyMAP [37] to compute and view the syntenic blocks. The algorithm first computed anchors from the alignment of the maize markers and BES sequences to the rice genomic sequence and then calculated chains of anchors using dynamic programming and merged chains into blocks. The graphical results from a versatile CGI/HTML and Java Display allow users to observe synteny from different views, including dotplot, genome

blocks, chromosome blocks, close-up chromosomes, close-up contigs, and table view. The results can be interactively browsed at <http://www.agcol.arizona.edu/symap>.

Supporting Information

Figure S1. Correlation between Gene Distribution and Genetic Recombination

The distribution of EST-derived overgo markers and genetic markers was plotted against maize chr1. The overgos, which hybridized to at least two BAC clones in each contig and hit fewer than ten contigs, were counted for the plot. The overgos and genetic markers were counted as a total in each Mb region.

Found at doi:10.1371/journal.pgen.0030123.sg001 (44 KB PPT).

Figure S2. Overall Block View of Maize–Rice Synteny

(A) General view of maize–rice synteny alignment is presented. This is the block view generated from dotplot analysis in Figure 1. Rice chromosomes are in the middle of each synteny block as gray vertical bars, and rice centromeres are shown in red. Maize chromosomes are colored as shown in the horizontal color key.

(B) Physical synteny map of rice chr2 with maize chromosomal regions is presented. Each maize physical block is indicated by chromosome number and FPC contig numbers. When clicked in our interactive website (<http://www.agcol.arizona.edu/symap>), an individual block view with rice alignment or a comparative block view like Figure 2 will pop up.

Found at doi:10.1371/journal.pgen.0030123.sg002 (72 KB PPT).

Table S1. Major Release of the Maize Physical Map

Found at doi:10.1371/journal.pgen.0030123.st001 (16 KB XLS).

Table S2. Summary of the Maize Genetic Map

Found at doi:10.1371/journal.pgen.0030123.st002 (26 KB XLS).

Table S3. Distribution of Different Marker Types among Chromosomes

Found at doi:10.1371/journal.pgen.0030123.st003 (16 KB XLS).

References

- Gale M, Devos K (1998) Comparative genetics in the grasses. *Proc Natl Acad Sci U S A* 95: 1971–1974.
- Kellogg E (1998) Relationships of cereal crops and other grasses. *Proc Natl Acad Sci U S A* 95: 2005–2010.
- International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature* 436: 793–800.
- Rayburn A, Biradar D, Bullock D, McMurphy L (1993) Nuclear DNA content in F1 hybrids of maize. *Heredity* 70: 294–300.
- Rhoades M (1951) Duplicate genes in maize. *Am Nat* 85: 105–110.
- Goodman MM, Stuber CW, Newton K, Weissinger HH (1980) Linkage relationships of 19 enzyme loci in maize. *Genetics* 96: 697–710.
- Wendel J, Stuber C, Edwards M, Goodman M (1986) Duplicated chromosome segments in *Zea mays* L.: Further evidence from hexokinase enzymes. *Theor Appl Gen* 72: 178–185.
- Moore G, Devos KM, Wang Z, Gale MD (1995) Cereal genome evolution. Grasses, line up and form a circle. *Curr Biol* 5: 737–739.
- Gaut B, Doebley J (1997) DNA sequence evidence for the segmental allotetraploid origin of maize. *Proc Natl Acad Sci U S A* 94: 6809–6814.
- Molina M, Naranjo C (1987) Cytogenetic studies in the genus *Zea*: 1. Evidence for five as the basic chromosome number. *Theor Appl Gen* 73: 542–550.
- Wilson WA, Harrington SE, Woodman WL, Lee M, Sorrells ME, et al. (1999) Inferences on the genome structure of progenitor maize through comparative analysis of rice, maize and the domesticated panicoids. *Genetics* 153: 453–473.
- Paterson AH, Bowers J, Peterson D, Estill J, Chapman B (2003) Structure and evolution of cereal genomes. *Curr Opin Genet Dev* 13: 644–650.
- Vandepoele K, Simillion C, Van de Peer Y (2003) Evidence that rice and other cereals are ancient aneuploids. *Plant Cell* 15: 2192–2202.
- Wang X, Shi X, Hao B, Ge S, Luo J (2005) Duplication and DNA segmental loss in the rice genome: Implications for diploidization. *New Phytol* 165: 937–946.
- Yu J, Wang J, Lin W, Li S, Li H, et al. (2005) The genomes of *Oryza sativa*: A history of duplications. *PLoS Biol* 3: e38. doi:10.1371/journal.pbio.0030038
- Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc Natl Acad Sci U S A* 101: 9903–9908.
- Swigonova Z, Lai J, Ma J, Ramakrishna W, Llaca V, et al. (2004) Close split of sorghum and maize genome progenitors. *Genome Res* 14: 1916–1923.

Table S4. Primary Synteny of Maize and Rice

Found at doi:10.1371/journal.pgen.0030123.st004 (30 KB XLS).

Table S5. Chromosomal Fragment Rearrangement after Maize–Rice Divergence

Found at doi:10.1371/journal.pgen.0030123.st005 (31 KB XLS).

Acknowledgments

We thank the CUGI/AGI fingerprinting and sequencing production teams, especially John and Dao Phimphilai, for hard work and dedication to the project. In addition, we thank Scott Tingey at Dupont Agriculture and Nutrition-Molecular Genetics for sharing some overgo markers with this project.

Author contributions. FW performed manual editing of the physical map, maize–rice synteny study, maize genome duplication and rearrangement analysis, maize chromosomal reconstruction, and wrote the paper. EC provided the genetic markers and overgo markers with their BAC clone association and aided the manual editing. WN performed the HICF analysis and helped in the manual editing and synteny analysis. FE was responsible for loading markers into the agarose-based map and analyzing their quality. AKB, EB, HRK, MC, GF, JM, and RAW produced HICF or agarose fingerprint and BES data. AKB, JLG, and SL participated in the manual editing. HSV, SS, and ZF derived the LIMS, mapping analysis, and pipeline utilities for genetic mapping and data transfers. MM, GD, JEB, AHP, MS, JG, and KC produced genetic and overgo markers, their association with BAC clones, and genetic maps. CS was involved with HICF, agarose, and synteny analysis. EC, AHP, JM, CS, and RAW participated in data generation, data analysis, and manuscript preparation.

Funding. This work was funded by National Science Foundation (NSF) Plant Genome grants DBI-9872655 (EC, KC, MM, JG, GD, MS, AHP, CS, and RAW), DBI-0211851 (JM, CS, and RAW), and DBI-0115903 (AP, CS, and RAW).

Competing interests. The authors have declared that no competing interests exist.

- Bruggmann R, Bharti AK, Gundlach H, Lai J, Young S, et al. (2006) Uneven chromosome contraction and expansion in the maize genome. *Genome Res* 16: 1241–1251.
- Lai J, Ma J, Swigonova Z, Ramakrishna W, Linton E, et al. (2004) Gene loss and movement in the maize genome. *Genome Res* 14: 1924–1931.
- Meyers B, Tingey S, Morgante M (2001) Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. *Genome Res* 11: 1660–1676.
- Messing J, Bharti AK, Karlowski WM, Gundlach H, Kim HR, et al. (2004) Sequence composition and genome organization of maize. *Proc Natl Acad Sci U S A* 101: 14349–14354.
- Haberer G, Young S, Bharti AK, Gundlach H, Raymond C, et al. (2005) Structure and architecture of the maize genome. *Plant Physiol* 139: 1612–1624.
- Du C, Swigonova Z, Messing J (2006) Retrotranspositions in orthologous regions of closely related grass species. *BMC Evol Biol* 6: e62.
- Yim YS, Davis GL, Duru NA, Musket TA, Linton EW, et al. (2002) Characterization of three maize bacterial artificial chromosome libraries toward anchoring of the physical map to the genetic map using high-density bacterial artificial chromosome filter hybridization. *Plant Physiol* 130: 1686–1696.
- Tomkins JP, Davis G, Main D, Yim Y, Duru N, et al. (2002) Construction and characterization of a deep-coverage bacterial artificial chromosome library for maize. *Crop Sci* 42: 928–933.
- Soderlund C, Longden I, Mott R (1997) FPC: A system for building contigs from restriction fingerprinted clones. *Comput Appl Biosci* 13: 523–535.
- Nelson WM, Bharti AK, Butler E, Wei F, Fuks G, et al. (2005) Whole-genome validation of high-information-content fingerprinting. *Plant Physiol* 139: 27–38.
- Gardiner J, Schroeder S, Polacco ML, Sanchez-Villeda H, Fang Z, et al. (2004) Anchoring 9,371 maize expressed sequence tagged unigenes to the bacterial artificial chromosome contig map by two-dimensional overgo hybridization. *Plant Physiol* 134: 1317–1326.
- Marra M, Kucaba T, Dietrich N, Green E, Brownstein B, et al. (1997) High throughput fingerprint analysis of large-insert clones. *Genome Res* 7: 1072–1084.
- Luo MC, Thomas C, You FM, Hsiao J, Ouyang S, et al. (2003) High-throughput fingerprinting of bacterial artificial chromosomes using the snapshot labeling kit and sizing of restriction fragments by capillary electrophoresis. *Genomics* 82: 378–389.

31. Ding Y, Johnson MD, Chen WQ, Wong D, Chen YJ, et al. (2001) Five-color-based high-information-content fingerprinting of bacterial artificial chromosome clones using type IIS restriction endonucleases. *Genomics* 74: 142–154.
32. Bennett M, Laurie D (1995) Chromosome size in maize and sorghum using EM serial section reconstructed nuclei. *Maydica* 40: 199–204.
33. Ahn S, Tanksley S (1993) Comparative linkage maps of the rice and maize genomes. *Proc Natl Acad Sci U S A* 90: 7980–7984.
34. Paterson AH, Lin Y-R, Li Z, Schertz KF, Doebley JF, et al. (1995) Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. *Science* 269: 1714–1718.
35. Goff SA, Ricke D, Lan TH, Presting G, Wang R, et al. (2002) A draft sequence of the rice genome (*Oryza sativa L. ssp. japonica*). *Science* 296: 92–100.
36. Salse J, Piegu B, Cooke R, Delseny M (2004) New in silico insight into the synteny between rice (*Oryza sativa L.*) and maize (*Zea mays L.*) highlights reshuffling and identifies new duplications in the rice genome. *Plant J* 38: 396–409.
37. Soderlund C, Nelson W, Shoemaker A, Paterson A (2006) SyMAP: A system for discovering and viewing syntenic regions of FPC maps. *Genome Res* 16: 1159–1168.
38. Bowers JE, Arias M, Asher R, Avise J, Ball R, et al. (2005) Comparative physical mapping links conservation of microsynteny to chromosome structure and recombination in grasses. *Proc Natl Acad Sci U S A* 102: 13206–13211.
39. Rice Chromosomes 11 and 12 Sequencing Consortia (2005) The sequence of rice Chromosomes 11 and 12, rich in disease resistance genes and recent gene duplications. *BMC Biol* 3: 20.
40. Gaut B (2001) Patterns of chromosomal duplication in maize and their implications for comparative maps of the grasses. *Genome Res* 11: 55–66.
41. Griffiths S, Sharp R, Foote T, Bertin I, Wanous M, et al. (2006) Molecular characterization of Ph1 as a major chromosome pairing locus in polyploid wheat. *Nature* 439: 749–752.
42. Sorrells ME, La Rota M, Bermudez-Kandianis CE, Greene RA, Kantety R, et al. (2003) Comparative DNA sequence analysis of wheat and rice genomes. *Genome Res* 13: 1818–1827.
43. Cone KC, McMullen MD, Bi IV, Davis GL, Yim YS, et al. (2002) Genetic, physical, and informatics resources for maize. On the road to an integrated map. *Plant Physiol* 130: 1598–1605.
44. Engler F, Hatfield J, Nelson W, Soderlund C (2003) Locating sequence on FPC maps and selecting a minimal tiling path. *Genome Res* 13: 2152–2163.
45. Chen M, Presting G, Barbazuk WB, Goicoechea JL, Blackmon B, et al. (2002) An integrated physical and genetic map of the rice genome. *Plant Cell* 14: 537–545.
46. Sulston J, Mallett F, Durbin R, Horsnell T (1989) Image analysis of restriction enzyme fingerprint autoradiograms. *Comput Appl Biosci* 5: 101–106.
47. Soderlund C, Humphray S, Dunham A, French L (2000) Contigs built with fingerprints, markers, and FPC V4.7. *Genome Res* 10: 1772–1787.
48. Davis G, McMullen M, Baysdorfer C, Musket T, Grant D, et al. (1999) A maize map standard with sequenced core markers, grass genome reference points and 932 expressed sequence tagged sites (ESTs) in a 1,736-locus map. *Genetics* 152: 1137–1172.
49. Coe E, Schaeffer M (2005) Genetic, physical, maps, and database resources for maize. *Maydica* 50: 285–303.
50. Lawrence C, Seigfried T, Brendel V (2005) The maize genetics and genomics database: The community resource for access to diverse maize data. *Plant Physiol* 138: 55–58.
51. Yim Y, Moak P, Sanchez-Villeda H, Musket T, Close P, et al. (2007) A BAC pooling strategy combined with PCR-based screenings in a large, highly repetitive genome enables integration of the maize genetic and physical maps. *BMC Genomics* 8: 47.