ORIGINAL PAPER

Comparative analyses reveal high levels of conserved colinearity between the finger millet and rice genomes

**Srinivasachary · Mathews M. Dida ·
Mike D. Gale · Katrien M. Devos**

Abstract Finger millet is an allotetraploid ($2n = 4x = 36$) grass that belongs to the Chloridoideae subfamily. A comparative analysis has been carried out to determine the relationship of the finger millet genome with that of rice. Six of the nine finger millet homoeologous groups corresponded to a single rice chromosome each. Each of the remaining three finger millet groups were orthologous to two rice chromosomes, and in all the three cases one rice chromosome was inserted into the centromeric region of a second rice chromosome to give the finger millet chromosomal configuration. All observed rearrangements were, among the grasses, unique to finger millet and, possibly, the Chloridoideae subfamily. Gene orders between rice and finger millet were highly conserved, with rearrangements being limited largely to single marker transpositions and small putative inversions encompassing at most three markers. Only some 10% of markers mapped to non-syntenic positions in rice and finger millet and the majority of these were located in the distal 14% of chromosome arms, supporting a possible correlation between recombination and sequence evolution as has previously been observed in wheat. A comparison of the organization of finger millet, Panicoideae and Pooideae genomes relative to rice allowed us to infer putative ancestral chromosome configurations in the grasses.

Communicated by A. Kilian.

Srinivasachary · M. M. Dida · M. D. Gale
John Innes Centre, Norwich Research Park,
Colney, Norwich NR4 7UH, UK

Srinivasachary · M. M. Dida · K. M. Devos
Department of Crop and Soil Sciences,
University of Georgia, 3111 Miller Plant Sciences,
Athens, GA 30602, USA

K. M. Devos (✉)
Department of Plant Biology, University of Georgia,
3111 Miller Plant Sciences, Athens, GA 30602, USA
e-mail: kdevos@uga.edu

M. M. Dida
Department of Botany and Horticulture, Maseno University,
P.O. Box 333, Maseno, Kenya

Introduction

Finger millet, *Eleusine coracana* (L.) Gaertn. subsp. *coracana*, is a tetraploid ($2n = 4x = 36$) belonging to the subfamily Chloridoideae. The finger millet crop is cultivated mainly in East Africa and Southern India where it makes an important contribution to food security due to the high nutritional value and good storage ability of its grain. Finger millet improvement has, however, lagged behind that of other crops. In particular, in Africa, farmers still rely largely on landraces for finger millet production and as a result obtain yields that are 60–90% below the yield potential of improved varieties (National Research Council 1996; FAOSTAT Data 2006).

The construction of a genetic map (Dida et al. 2007) was a first step in the development of finger millet tools that could assist breeders in their endeavor to improve the agricultural performance of this crop. To enhance the value of the linkage map and allow finger millet researchers and breeders to draw on information from the well-studied major cereals, in particular rice for which the whole-genome sequence is available (International Rice Genome

Sequencing Project 2005), we aimed to establish the relationship between the finger millet genome and that of rice. Comparative analyses at the map level between genomes of species belonging to the grass subfamilies Ehrhartoideae, Panicoideae and Pooideae have shown an overall high conservation of colinearity despite some 60 million years of independent evolution of the major grass lineages (Gale and Devos 1998, 2005). Species belonging to the same subfamily are characterized by a few common rearrangements, indicating that these particular chromosome configurations were acquired soon after the split of the subfamilies. In addition, rearrangements specific to individual lineages or species within a subfamily can be observed. The number of rearrangements a species accumulates appears to be independent of evolutionary distance (Zhang et al. 1998; Devos et al. 2000).

In addition to providing information on the relationship between the finger millet and rice genomes, the inclusion of finger millet in the comparative grass consensus map is important in a wider evolutionary context. The Chloridoideae and Panicoideae subfamilies belong to the PACCAD clade (Grass Phylogeny Working Group 2001) and may carry rearrangements that predate the Panicoideae–Chloridoideae divergence. Limited comparative information has previously been published for *Eragrostis tef*, also a Chloridoideae species, with rice but, in most cases, regions of orthology were defined by too few markers to draw conclusions on the evolutionary history of this Chloridoid genome (Zhang et al. 2001a). Finger millet is the first member of the Chloridoideae subfamily for which a comprehensive comparative map has been produced that can be analyzed for structural similarity with other grass genomes.

Materials and methods

The genetic map

The map was constructed in an F_2 population of 151 progeny generated from an intersubspecific cross between *E. coracana* subsp. *africana* acc. MD-20, the wild progenitor of finger millet, and the cultivated *E. coracana* subsp. *coracana* cv. Okhale-1, and was published in Dida et al. (2007). The markers consisted of finger millet *Pst*I genomic clones and expressed sequenced tags (ESTs), heterologous genomic clones and cDNAs, finger millet simple sequence repeats (SSRs), a few amplified fragment length polymorphism (AFLP) markers and two resistance gene analogs (RGAs). A total of 332 loci, detected by 266 probes or primer pairs, were organized into 26 linkage groups. Thirteen A-genome linkage groups and nine B-genome linkage groups were assembled into nine homoeologous groups.

Comparative markers

Fifty-one markers from the Nipponbare×Kasalath rice map generated by the Rice Genome Research Program (referred to as the RGP rice map) (Harushima et al. 1998; http://rgp.dna.affrc.go.jp/publicdata/geneticmap2000/index.html) had been mapped to 64 loci in finger millet and represented a first source of comparative data points. A second source was the clones mapped in finger millet for which sequence information was available. This included all finger millet ESTs and SSRs, 25 end-sequenced finger millet *Pst*I clones, and the cDNA clones RZ421 (rice) and CDO127 (oats). The sequences were used in a BLASTn search with default settings against a TIGR database that consisted of all rice BAC and PAC sequences present in GenBank (http://tigrblast.tigr.org/euk-blast/index.cgi?project=osa1). The threshold for detecting homology was initially set at E^{-10}. Finger millet–rice sequence pairs that showed homology at an E-value $> E^{-20}$ were aligned using the 'BLAST 2 SEQUENCES' tool at NCBI (http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi). Sequences within a pair for which no significant homology could be found (E-value threshold of E^{-5}) were considered to be non-orthologous and were omitted from the comparative analysis (Supplementary Table 1).

As a starting point for building the comparative maps, the rice BAC/PAC clone that displayed the highest homology to a finger millet query sequence was assumed to contain the ortholog of this query sequence. The genetic map position of each rice BAC/PAC clone was obtained from the integrated rice physical and genetic maps (http://rgp.dna.affrc.go.jp/cgi-bin/statusdb/irgsp-status.cgi). For markers that were located in apparently non-colinear positions in finger millet and rice, the stringency of our criterion for defining orthology was reduced from highest hit only to accept secondary and, in some cases, tertiary hits against the rice genomic sequence provided that the E-values of the secondary and tertiary hits were $\leq E^{-10}$ and $\leq E^{-\text{best hit}/2}$. Finger millet sequences that displayed high homology with eight or more rice chromosomes were considered to contain repeats and were not used in the comparative analysis.

For rice and wheat RFLP markers that had been mapped by hybridization onto the RGP map (Harushima et al. 1998; http://rgp.dna.affrc.go.jp/publicdata/geneticmap2000/index.html) and detected loci in non-syntenic positions in finger millet, the corresponding end-sequences were retrieved from GenBank and used as queries against the TIGR rice BAC/PAC database. Primary, secondary or tertiary hits that had E-values $\leq E^{-10}$ were assumed to be true orthologs of the locus mapped in finger millet if their rice location fitted the overall pattern of rice–finger millet synteny displayed by the majority of the comparative loci.

Results

Comparative markers

Sequence information was available for 218 (80%) out of the 272 markers present on the finger millet genetic map (Table 1). Seventy-eight percent of the sequences showed homology with rice at an E-value $\leq E^{-10}$ and with pairwise homology $\leq E^{-5}$. In total, 62% of the mapped clones represented comparative data points. For each type of clone, the number of clones with sequence information and the percentage of sequenced clones with homology to a rice BAC or PAC clone are given in Table 1.

Comparative maps

The rice BAC/PAC clones that carried sequences with putative orthology to mapped finger millet markers were located on the RGP rice map using the integrated physical and genetic map data (http://rgp.dna.affrc.go.jp/cgi-bin/statusdb/irgsp-status.cgi) (Supplementary Table 2). The resulting comparative maps are shown in Fig. 1. Six and three of the nine finger millet homoeologous groups each correspond to a single and two rice chromosomes, respectively. The individual finger millet–rice comparative maps are discussed below. Overall, marker orders are highly conserved. It should be noted that we were conservative in the identification of rearrangements. Shuffled orders of closely linked markers are often the result of mapping errors. Therefore, an inversion needed to encompass at least three markers in order to be classified as a rearrangement.

Finger millet 1A, 1B: rice 1

The group 1 finger millet maps carry 29 comparative loci, 21 of which are colinear with and span 85% of rice chromosome 1. One locus, *Xrgc1456*, is duplicated on LG 1B. The proximally located finger millet locus *Xrgc1456.1* is orthologous to the rice locus and most likely the original copy. The most likely position of the centromere on LG 1B can be extrapolated from the centromere location in rice and is located in the 5.2 cM interval between markers *Xlfo11* and *Xinf73*.

Of the seven non-syntenic loci, one is located on LG 1Aa, which consists of only two markers, *Xrgr569*, located on rice chromosome 5, and *Xpse167.2*. Because Southern analysis has shown that the RFLP probe PSE167 is single copy in the finger millet genome, the *Xpse167.1* and *Xpse167.2* loci on LG 1B and LG 1Aa, respectively, are expected to be homoeologous. If this is indeed the case, then the rice and finger millet *Xrgr569* loci are non-syntenic. The alternative is that the *Xrgr569* loci on finger millet LG 1Aa and rice chromosome 5 are orthologous, and that LG 1Aa is, in fact, part of finger millet linkage group 5A. A second non-syntenic locus, *Xrgc58.1* is duplicated on finger millet LGs 1B and 6A. The LG 6A locus is probably the original copy as it is syntenic with its rice ortholog on rice chromosome 6. The other five non-syntenic markers identify loci on rice chromosomes 2 (2 markers), 7, 9 and 10. Four of them are located in the distal region of finger millet linkage group (LG) 1B for which orthology with rice could not be established.

Finger millet 2A, 2B: rice 2 and 10

The group 2 chromosomes of finger millet are orthologous partly to the long arm of rice chromosome 2 and partly to the long arm of rice chromosome 10. No markers from the short arm of rice 10 have been mapped in finger millet. This is not surprising as most comparative analyses carried out to date have failed to establish a relationship between rice 10S and other grass genomes (Devos 2005). Finger millet LG 10, which spans 16.5 cM and had not previously been assigned to a homoeologous group (Dida et al. 2007), carries four markers with homology to sequences on rice chromosome arm 2S. We speculate that LG 10 is a part of either

Table 1 Number of clones on the finger millet genetic map with sequence information and showing homology to rice BAC/PAC clones at an E-value $\leq E^{-10}$ and pairwise homology at an E-value $\leq E^{-5}$

Type of clone	No. of clones mapped	No. of clones with sequence information	No. of sequences with homology to rice BAC/PAC
Finger millet *Pst*I clones (*Xpse*)	68	25 (37%)	7 (28%)
Finger millet EST (*Xlfo, Xlfm, Xrtm, Xinf*)	96	96	87 (91%)
Finger millet SSR (*Xugep*)	31[a]	31	10 (32%)
Rice RFLP (*Xrgc, Xrgg, Xrgr, Xrgl, Xrz*)	52	52	52 (100%)
Other cereal RFLP (*Xpsr, Xcdo*)	14	14	14 (100%)
AFLP	9	0	0
Cereal RGA clones	2	0	0
Total	272	218 (80%)	170 (78%)

[a] This does not include two SSRs that were located in a LINE-like retroelement (Dida et al. 2007)
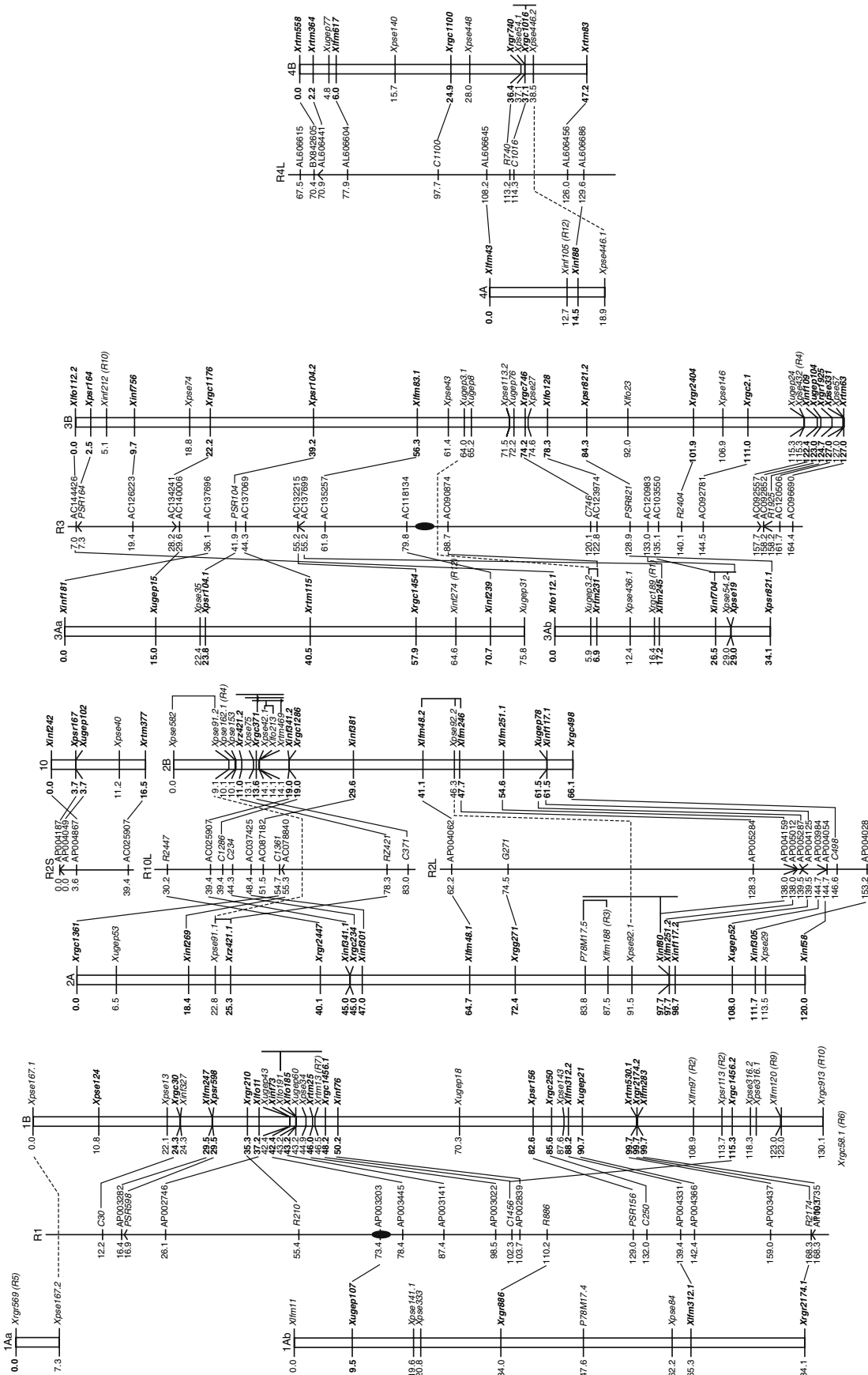
Fig. 1 Comparative finger millet–rice genetic maps. Rice genetic maps are indicated with R followed by the chromosome number and are flanked on the left-hand-side (LHS) by the orthologous finger millet A genome linkage group(s) and on the right-hand-side (RHS) by the orthologous finger millet B genome linkage group. Vertical bars spanning some loci on the finger millet maps indicate ambiguous marker orders. Genetic distances are given in centiMorgan (Kosambi) on the LHS of each genetic map. Black ovals on the rice genetic map indicate the position of centromeres. Comparative loci between rice and finger millet are connected with full lines and are in bold on the finger millet genetic map. Homoeologous loci between the finger millet A and B genomes are connected by dashed lines. The location in rice of finger millet loci that are non-syntenic in rice and finger millet is given in parenthesis
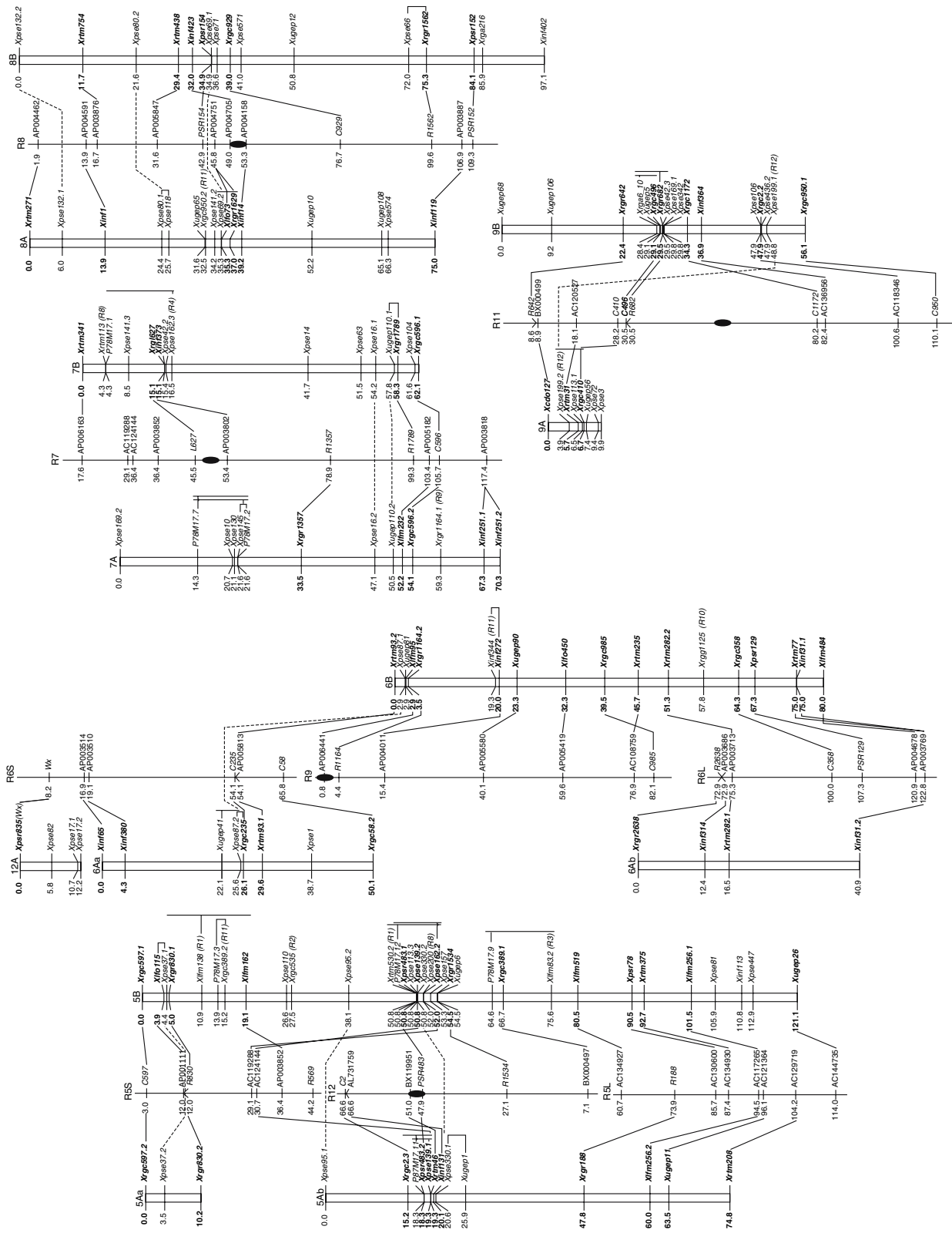
Fig. 1 continued

finger millet LG 2A or LG 2B. Including LG 10 in the comparison, finger millet homoeologous group 2 covers and is highly colinear with the entire rice chromosome 2.

Finger millet 3A, 3B: rice 3

Finger millet LGs 3A and 3B are orthologous to and cover 94% of rice chromosome 3. The centromere can be putatively assigned to the 17.9 cM *Xlfm83.1–Xrgc746* interval. There is evidence of three possible rearrangements in finger millet LG 3A relative to LG 3B and rice. The location of markers *Xinf704, Xpse19* and *Xpsr821.1* suggest that the distal region of LG 3A may have undergone an inversion. The second rearrangement is identified by the non-colinear position of *Xlfo112.1* and *Xlfo112.2*. When comparing LGs 3A and 3B, it was not possible to establish whether this rearrangement involved a large chromosome segment or a single marker (Dida et al. 2007). The complete colinearity of rice chromosome 3 with finger millet LG 3B, however, suggests that the location of *Xlfo112.1* is due to a single gene translocation. This assumes, of course, that *Xlfo112.1* and *Xlfo112.2* are truly homoeologous. The non-colinear position of *Xlfm245* in LG 3A and rice chromosome 3 may also be the result of a single locus translocation that occurred either in finger millet or in rice. However, because *Xlfm245* shows homology with sequences on rice chromosomes 5, 11 and 3 (Supplementary Table 2) and detects at least two copies in finger millet, it is possible that the loci mapped in finger millet and rice are paralogs rather than orthologs.

Finger millet 4A, 4B: rice 4

Finger millet homoeologous group 4 covers only 48% of rice chromosome 4 and is incomplete. This is most likely due to the limited number of markers mapped that have orthology to rice chromosome 4, rather than to a specific characteristic of the finger millet homoeologous group 4 chromosomes. There is no evidence of rearrangements between finger millet homoeologous group 4 and rice 4 chromosomes.

Finger millet 5A, 5B: rice 5 and 12

Finger millet homoeologous group 5 is, from top to bottom, colinear with the short arm of rice chromosome 5, rice 12 long to short, and the long arm of rice 5. The finger millet maps cover 93 and 54% of rice 5 and 12, respectively. An association between rice chromosomes 5 and 12 has also been observed in the organization of foxtail millet chromosome III and pearl millet LG 1 relative to rice (Devos et al. 1998, 2000). The rice 5–12 arrangements seen in the foxtail and pearl millet maps are rather complex, but suggest that these chromosomes are characterized by the insertion of

rice 5 into 12 rather than rice 12 into rice 5 as seen in finger millet.

Two closely linked finger millet markers, *Xpse139* and *Xpse162.2*, detect closely linked copies in rice that are duplicated on rice chromosomes 4 and 5. A small duplication involving approximately 400 kb on rice 4 and 200 kb on rice 5 has previously been observed (http://www.tigr.org/tdb/e2 k1/osa1/segmental_dup/500 kb/ segdup_500 kb.shtml). However, although the rice 4 homologs of *Xpse139* and *Xpse162.2* fall with the 400 kb region that is duplicated on rice 5, this is not the case for the rice 5 copies. A possible scenario is that the group 5 copies were part of the larger rice 4–5 duplication and subsequently translocated to a different position on rice 5. If this is the case, their post-duplication but pre-translocation position would be at ~87 cM on the RGP rice chromosome 5 map. This location is also non-colinear with the map position of *Xpse139* and *Xpse162.2* in finger millet.

Three markers, *Xpsr483.2, Xrtm46* and *Xinf131*, that span a region of 1.8 cM in finger millet LG 5Ab are located in inverted orientation relative to their orthologs on rice chromosome 12. Although the inversion is supported by three markers, it is equivocal because the marker order is based on only few recombinants and thus highly prone to errors.

Finger millet 6A, 6B: rice 6 and 9

The homoeologous group 6 chromosomes of finger millet correspond, from top to bottom, with 100% of the short arm of rice chromosome 6, 85% of rice 9 (oriented from short to long), and 85% of the long arm of rice 6. Marker orders are highly colinear. Linkage group 12, which had previously been shown to belong to the A genome but did not link to any of the linkage groups for which homoeology had been established (Dida et al. 2007), carries *Xpsr835(Wx)* which defines the *waxy* locus. In rice, the *waxy* locus has been mapped to the distal region of the short arm of rice chromosome 6, suggesting that LG 12 is part of LG 6A.

Finger millet 7A, 7B: rice 7

Finger millet homoeologous group 7 is highly colinear with and covers 85% of rice chromosome 7. Extrapolating from its location in rice, the centromere in LG 7B can be placed between markers *Xrgl627* and *Xinf373* that co-map in finger millet.

Finger millet 8A, 8B: rice 8

Marker orders are highly conserved between finger millet LGs 8A and 8B, and rice chromosome 8. The finger millet maps cover 90% of the rice genetic map. Putative centromere

positions can be assigned to the 2.2 cM *Xrgr1629–Xinf14* interval on LG 8A and the 7 cM *Xpsr154–Xrgc929* interval on LG 8B.

Finger millet 9A, 9B: rice 11

The homoeologous group 9 chromosomes are highly colinear with and cover 86% of rice chromosome 11. The centromere on LG 9B can be putatively allocated to the 4.8 cM *Xrgr682–Xrgc1172* interval.

Unlinked groups

Linkage group 13A consists of two loci, *Xrtm301* and *Xinf345*, and has previously been assigned to the A genome (Dida et al. 2007). *Xrtm301* shows homology to multiple sugar transporter genes with the highest hits on rice chromosomes 10 and 2, and *Xinf345* has homology to a locus on rice chromosome 4 (Supplementary Table 2). Neither the rice chromosome 2 nor the rice chromosome 4 regions that are potentially orthologous to *Xrtm301* and *Xinf345*, respectively, are covered on our finger millet maps. It remains unclear where LG 13A is located in finger millet.

Linkage group 11A, also belonging to the A genome and consisting of four markers (Dida et al. 2007), also remains unassigned. Comparative data from marker *Xlfo201* suggests that this linkage group is orthologous to either rice chromosome 9 or 2 (Supplementary Table 2).

Discussion

Comparative markers

Some 48% of the markers, including all finger millet RFLPs and SSRs, and around 20% of ESTs, were mapped without prior information on sequence identity with rice. Approximately 73% of finger millet ESTs (data not shown) and 30% of end-sequenced finger millet genomic clones identify putative rice orthologs when used as queries against the rice genomic sequence. The majority of the finger millet genomic clones were generated using the methylation-sensitive restriction enzymes *Pst*I and *Sal*I. Based on the proportions of wheat *Pst*I clones that hybridize to all three homoeologous chromosomes in hexaploid wheat and those that display genome-specific hybridization signals, we have estimated that some 50% of *Pst*I genomic clones represent coding regions (Devos et al. 1993). In finger millet, the percentage of *Pst*I and *Sal*I genomic clones that identify rice orthologs is 41% that of finger millet ESTs, suggesting that at least 41% of the clones generated with methylation-sensitive restriction enzymes comprise genic DNA. The extent of homology between rice

and finger millet is likely to be higher than suggested by the 73% of finger millet ESTs for which a rice ortholog could be identified. The homology between finger millet and rice is mostly limited to exons by the criteria we use to define orthology and no homology will be found for short ESTs that consist mainly of 3′ untranslated regions. The fact that finger millet ESTs for which no rice ortholog could be detected are, on average, 28% shorter than those that do have a rice homolog suggests that, at least in some cases, the lack of homology may be due to the absence of exon regions in the short sequence reads.

The remaining 52% of mapped markers were chosen specifically to provide comparative information and consist of RFLP probes previously mapped in rice and finger millet unigenes from leaf, root and inflorescence cDNA libraries (Dida et al. 2007) for which a rice homolog could be identified at an *e*-value $\leq E^{-10}$. In total, 62% of the mapped markers represent comparative datapoints.

Establishing orthology versus paralogy

When generating comparative genetic maps, it is important to ascertain to the best extent possible that the markers used as comparative datapoints are true orthologs. Hybridization of the same probe in different species requires high levels of conservation between probe and target, and is expected to lead to the identification of either orthologous or paralogous loci. *In silico* analyses are much more sensitive than hybridization assays to detect orthologs, and consequently are also more prone to error. Searching for homologs in incomplete databases and homology limited to domains may confound the identification of true orthologs. Some 37% of the comparative datapoints on the rice–finger millet comparative maps were obtained by hybridizing RFLP probes with known location in rice to finger millet genomic DNA. The remaining 63% represent sequences mapped in finger millet for which a rice ortholog was identified *in silico*.

The criteria for determining true orthology were established empirically. A BLASTn search using the finger millet sequences as queries against the TIGR database that houses all rice BAC and PAC sequences identified a first set of orthologous pairs at an *E*-value threshold of E^{-10}. When these pairs were used to evaluate the extent of colinearity between finger millet and rice, it was noted that 25% of the markers mapped to non-syntenic positions. This number does not include markers for which multiple loci were mapped in finger millet and for which at least one locus was located in a syntenous position in rice. Considering not only the highest hit ($E^{-\text{best hit}}$), but all hits with *E*-values $\leq E^{-10}$ and $\leq E^{-\text{best hit}/2}$ to correct for the identification of paralogs, the amount of non-syntenic pairs decreased to 16%. Four of the 17 non-syntenic pairs were identified at *E*-values $\geq E^{-20}$, and were considered false

positives as they could not be aligned at an E-value $\leq E^{-5}$ using NCBI's BLAST 2 SEQUENCES tool.

The percentage of non-syntenic loci identified in hybridization assays was 20%. However, 54% of the non-syntenic pairs could be explained by the inclusion of paralogs rather than orthologs in the comparison, as syntenic locations could be found *in silico* in the rice genome at E-values $\leq$ E^{-10} (Supplementary Table 2). For marker *Xrgg271*, a putative orthologous locus was identified with an E-value of $1.2e^{-9}$ on rice chromosome 2 (Fig.1). Although this was above the threshold of E^{-10}, the finger millet and rice sequences could be aligned at an *e*-value of $2e^{-24}$ using NCBI's 'BLAST 2 SEQUENCES' tool. A total of 13 (13%) and 6 (9%) marker pairs identified *in silico* and by hybridization, respectively, remain in non-syntenic positions. For these 11% of non-syntenic pairs, it may be that the rice orthologous sequence is not yet available in GenBank, differential deletion of duplicated gene copies has led to single paralogous loci remaining in rice and finger millet (discussed in more detail under genome duplications) or, more likely, that the locus has been rearranged in either rice or finger millet.

Breakdown of synteny

Wheat-rice comparative analyses have demonstrated that loss of synteny is more frequent towards the chromosome ends and that this is correlated with the increase in recombination along the centromere–telomere axes in wheat chromosomes (Akhunov et al. 2003; See et al. 2006). To investigate whether this is also the case in finger millet, we compared the percentage of non-syntenic loci in the distal chromosome regions compared to the rest of the genome. This could be done only for finger millet linkage groups that covered chromosomes relatively completely ($\geq$85%), that is LGs 1B, 3B, 5B, 8A, 8B and 9B. Of the 15 non-syntenic loci identified on these six chromosomes, including the loci that are duplicated in finger millet and for which one, presumably the ancestral, copy has retained synteny with rice, 10 (67%) are located in the distal 25% of a linkage group. It should be noted that the distal 25% of a linkage group represents physically almost certainly much less than half of a chromosome arm as recombination in these regions is higher compared to centromeric regions. Although precise knowledge of the distribution of recombination requires a combination of physical and genetic mapping, which is not available for finger millet, we can nevertheless estimate the physical size of finger millet linkage group intervals using rice as a reference. Transfer of linkage intervals from the finger millet map to the TIGR rice pseudomolecules release four estimates that the distal 25% of LGs 1B, 3B, 8B and 9B correspond to, on average, 13.3% of a chromosome. Thus, about two-third of non-syntenic loci are located in telomeric regions representing less

than one-third of the genome. This suggests that, as in wheat, breakdown in synteny between finger millet and rice is correlated with increased recombination in the distal chromosome regions and could be caused by the preferential translocation and retention of duplicated gene copies in high-recombinant regions (Akhunov et al. 2003; See et al. 2006).

The finger millet: rice comparative map

Overall, colinearity is highly conserved between finger millet and rice (Fig.1). Three finger millet homoeologous groups, 2, 5 and 6, show colinearity with two rice chromosomes each, 2 and 10, 5 and 12, and 6 and 9, respectively. In all the three cases, one of the rice chromosomes is inserted into the centromeric region of another rice chromosome to give the finger millet arrangement.

The insertion of rice chromosome 10 into rice 2 is a configuration not previously seen in the grasses. In the Panicoideae, represented by foxtail millet, pearl millet, sorghum and maize, rice 10 has inserted in rice 3 to form specific Panicoideae chromosomes (Devos 2005). Comparative data from wheat, oats, fescue and rye grass reveals that the insertion of rice 10 into 5 characterizes species within the Pooideae subfamily. Similarly, the insertion of rice 9 into 7 and 6, respectively, is required to explain the structure of specific Panicoideae chromosomes and finger millet LG 6. It is intriguing that some chromosomes seem to be involved in rearrangements more often than others. Two rearrangements, rice 3S-10-3L and 7S-9-7L have been identified that characterize all Panicoideae species analyzed to date (Devos 2005). Similarly, the rice configurations 5S-10-5L and 6S-8-6L appear to predate the divergence of species within the Pooideae subfamily. Of the three two-chromosome arrangements identified in finger millet, rice 2S-10-2L, 5S-12-5L and 6S-9-6L, rice chromosomes 5, 6, 9 and 10 have been involved in rearrangements in both finger millet and in the Panicoideae or Pooideae. Because no information is available on what drives chromosomal rearrangements in plants, it is not possible to speculate whether the repeated involvement of these chromosomes if fortuitous or mechanistic. However, it seems reasonable to assume that the individual rice chromosomes 5, 6, 9 and 10 represent the more ancestral chromosome form as each chromosome has been involved in at least two different evolutionary rearrangements.

Also interesting is the rearrangement of rice chromosome 10 in both the Panicoideae rice 10–rice 3 and Chloridoideae rice 10–rice 2 configuration (Fig.1). The Panicoideae and Chloridoideae subfamilies are both members of the PACCAD Clade (Grass Phylogeny Working Group 2001), and it is therefore conceivable that the rearrangement in the segment with orthology to rice 10

occurred before the divergence of the Panicoideae and Chloridoideae lineages. To trace the evolutionary history of this rearrangement, we precisely compared the breakpoint regions on rice chromosome 10 relative to maize chromosome 1, foxtail millet LG IX and finger millet homoeologous group 2 (Table 2). Based on the comparative map between rice 10 and maize chromosome 1 displayed in Gramene (http://www.gramene.org/Oryza_sativa/syntenyview), the breakpoint on rice 10 relative to maize chromosome 1 is between 19.3 and 19.6 Mb. A BLASTn search of the rice chromosome 10 markers mapped on the foxtail millet genetic map (Devos et al. 1998) against the TIGR Rice 10 pseudomolecule release 4 (http://tigrblast.tigr.org/euk-blast/index.cgi?project=osa1) shows that the breakpoint is located in the region 18.8–20.6 Mb on rice chromosome 10. A similar approach was used to establish that the breakpoint relative to finger millet is located in the region 18.0–18.8 Mb on rice chromosome 10. However, the relative location of marker *Xrgc1361* in the finger millet and foxtail millet genetic maps suggests that these rearrangements occurred independently. Although this conclusion is based on only one marker, the mapping data for *Xrgc1361* was scrutinized in both foxtail and finger millet, and found to result in an unambiguous map position. The breakpoints relative to foxtail millet and maize appear to be identical within the limits of our analysis, indicating that this rearrangement may have occurred early in the Panicoideae lineage.

Genome duplications

Recent studies, made possible by the availability of the rice genomic sequence, have revealed that the rice genome has undergone a whole-genome duplication event some 70 MYA (Paterson et al. 2004). This duplication is assumed to predate the divergence of the major grass lineages. This implies that the finger millet ancestors, together with other grasses, are ancient tetraploids, and that finger millet itself, which is considered an allotetraploid, is in fact an ancient octoploid. No evidence was found at the map level of this ancient tetraploidization event, which would have taken place in the diploid progenitors of current-day finger millet. This is not surprising considering the age of the duplication. However, some of the 11% of non-syntenic loci identified between finger millet and rice might be attributable to differential deletion of paralogous gene copies in the two species following the whole-genome duplication event. This would lead to the remaining paralogs being identified as orthologs. To investigate if any of the 19 non-syntenic finger millet–rice pairs were composed of paralogs, we checked whether synteny was maintained between finger millet and a putative duplicated gene copy in rice (Supplementary Table 2). Only the large segmental duplications between rice 1 and 5, rice 2, 4 and 6, rice 3, 7 and 10, rice 8 and 9, and rice 11 and 12 were considered (Paterson et al. 2004). Five markers, *Xrgr569* (LG 1Aa–rice 5), *Xlfm188* (LG 2A–rice 3), *Xinf212* (LG 3B–rice 10), *Xlfm138* (LG 5B–rice 1) and *Xpse199* (LGs 9A, 9B–rice 12), were identified for which the rice ortholog might have been deleted. The location of these markers was established on the TIGR rice pseudomolecules release 4 and this information, in turn, was used to determine the expected location of a duplicate generated during the whole genome duplication event (http://www.tigr.org/tdb/e2 k1/osa1/segmental_dup/500 kb/segdup_500 kb.shtml). *Xpse199* mapped to a region

Table 2 Homologous regions identified between rice chromosome 10 and rice 3, maize 1, foxtail millet IX and finger millet 2

Rice 10	Rice 3[1]	Maize 1[2]	Foxtail IX[3]	Finger millet 2[4]
13 – 14 Mb	4.0 – 3.6 Mb	~171 Mb	*Xrgc961* – 84.4 cM	
14 – 15 Mb	3.6 – 3.2 Mb	~170 Mb	*Xrgr2447* – 86.4 cM	*Xrgr2447* – 30.2 cM
15 – 16 Mb	3.2 – 2.6 Mb			*Xinf341*- 45.0 cM
16 – 17 Mb	2.6 – 2.1 Mb	~165 Mb	*Xrgc1286* – 92.4 cM, *Xrgc814* – 93.4 cM	*Xrgc1286* – 45 cM; *Xrgc234* – 45.0 cM
17 – 18 Mb		~162 Mb		*Xinf301* – 47.0 cM, *Xinf381* – 55.6 cM
18 – 19 Mb		~158 Mb	*Xrgc1361* - 98.4 cM	*Xrgc1361* – 0 cM, *Xinf269* – 18.4 cM[5]
19 – 20 Mb	0.6 – 1.0 Mb[5]	~61 Mb[5]		
20 – 21 Mb	1.0 – 2.1 Mb[5] 1.5 – 0.4 Mb	~64 Mb	*Xrgr716* – 75.5 cM[5]	
21 – 22 Mb	0.4 – 0.1 Mb	~70 Mb		
22 – 23 Mb		~75 Mb	*Xrgc239* – 79.5 cM, *Xrgc371*- 80.4 cM	*Xrz421* – 25.3 cM, *Xrgc371* – 27.9 cM

[a] Homologous regions between rice chromosomes 3 and 10 were obtained from TIGR (http://www.tigr.org/tdb/e2k1/osa1/segmental_dup/500kb/segdup_500kb.shtml)

[b] Homologous regions between rice chromosome 10 and maize chromosome 1 were obtained from Gramene (http://www.gramene.org/Zea_mays/syntenyview?otherspecies=Oryza_sativa&chr=1&loc=69859328)

[c] Genetic distances in foxtail millet are from Devos et al. (1998)

[d] Genetic distances are from LG 2A or extrapolated from LG 2B (this paper)

[e] Transitions between white and shaded regions indicate breakpoints of chromosomal rearrangements relative to rice chromosome 10

of rice chromosome 12 that was unique under the criteria used for identifying the rice segmental duplications. *Xlfm188* mapped to a region of finger millet group 2 that is orthologous to rice 2, and thus the rice 3 location is neither orthologous nor paralogous to the finger millet *Xlfm188* locus. The locations of *Xinf212* and *Xlfm138* in rice, however, fit within the pattern of segmental duplications in rice, and likely represent paralogs to the finger millet loci. This may also be the case for *Xrgr569*. Although the expected position of the deleted rice paralog is at position 5.8 Mb on rice chromosome 1, which is at least 2.2 Mb more proximal than the rice location extrapolated from the finger millet–rice group 1 comparative maps (Table 2), it should be noted that the location of *Xrgr569* in finger millet may be imprecise, due to its dominant nature and its linkage with only 1 other marker (Fig. 1).

A detailed analysis of the duplication between rice chromosomes 3 and 10 shows that rearrangements have taken place in either rice 10 or 3 since the duplication event (Table 2). The finger millet maps are not sufficiently detailed to establish whether the rearrangement occurred in rice 3 or 10. Colinearity between this region of rice chromosome 3 and the orthologous segment in maize chromosome 1 is, however, completely conserved (http://www.gramene.org/Oryza_sativa/syntenyview). This implies that the organization of rice chromosome 3 is most ancestral. The rearrangement that occurred in rice 10 does not, however, explain the altered configuration of rice 10 relative to the maize 1, foxtail millet IX and finger millet 2 maps. This suggests that rice 10 is prone to rearrangements, with some of the breakpoints mapping to similar regions. Translocation hot-spots have previously been observed on Triticeae chromosomes 4, 5 and 7 (Zhang et al. 1998, 2001b).

Another rearrangement that may have occurred multiple times independently is the duplication that has been observed between the short arms of rice chromosomes 11 and 12, the orthologous regions on foxtail millet chromosomes VII and VIII, and pearl millet linkage groups 1 and 4. Because the duplications extend over similar regions, we have always assumed that this was an ancestral duplication that occurred before the divergence of the grasses. However, the rice duplication has recently been dated to ca. 5–7.7 MYA (Choisne et al. 2005; Wang et al. 2005) and is thus considerably younger than the split of the Ehrhartoideae and Panicoideae lineages. This suggests that the duplication must have taken place at least twice, once in the rice lineage, and once in the foxtail millet/pearl millet lineage. No evidence was found of this duplication in finger millet.

Conclusions

The finger millet–rice comparative maps demonstrate that, other than the rearrangements that are necessary to account for the difference in chromosome number between finger millet and rice, the finger millet genome has remained highly conserved since its divergence from a common ancestor with rice some 60 MYA. Some 10% of the markers were found in non-syntenic positions between finger millet and rice, most of which were located in the distal 14% of chromosome arms. The breakdown of finger millet–rice synteny in the distal regions might be correlated with high recombination rates, as has been previously observed in wheat.

The composite structures of finger millet LGs 2 (rice 10 into 2), 5 (rice 12 into 5) and 6 (rice 9 into 6) are unique to finger millet or possibly the Chloridoideae and are different from those that characterize the Panicoideae and Pooideae species. However, the recurrent involvement of orthologous chromosomes in different rearrangements suggests these rearrangements do not occur randomly but are driven by as yet unknown structural characteristics of the chromosomes involved.

The high conservation of colinearity between finger millet and rice will facilitate the exploitation of the information and resources available from rice and other grasses in this

Table 3 Finger millet loci for which orthology may be confounded by the whole genome duplication of rice

Locus	Region on rice pseudomolecule (Mb)[a]	Corresponding duplicated region in rice (Mb)[b]	Flanking markers	Finger millet LG	Region on rice pseudomolecule (Mb)
Xrgr569	R5; 6.7	R1; 5.8	*Telomere–Xrgc30*[c]	1Aa	R1; 0–2.6
Xlfm188	R3; 16.9	R7; 25.0	*Xrgg271–Xinf80*	2A	R2
Xinf212	R10; 15.9	R3; 2.6	*Xpsr164–Xinf756*	3B	R3; 1.6–4.2
Xlfm138	R1; 13.9	R5; 3.2	*Xrgr830–Xlfm162*	5B	R5; 1.1–5.2
Xpse199	R12; 14.9	–	*Xrgc2.2–Xrgc950*	9B	R11; 24.5–26

[a] TIGR rice pseudomolecules release 4

[b] Obtained from http://www.tigr.org/tdb/e2k1/osa1/segmental_dup/500kb/segdup_500kb.shtml

[c] Extrapolated from finger millet LG 1B

orphan crop. Traits that are of immediate interest to finger millet breeders are blast, caused by the fungus *Pyricularia oryzae*, and drought resistance, both of which have been the subject of much research in rice.

References

Akhunov ED, Goodyear AW, Geng S, Qi LL, Echalier B, Gill BS, Miftahudin, Gustafson JP, Lazo G, Chao SM, et al. (2003) The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. Genome Res 13:753–763

Choisne N, Demange N, Orjeda G, Samain S, D'Hont A, Cattolico L, Pelletier E, Couloux A, Segurens B, Wincker P, et al. (2005) The sequence of rice chromosomes 11 and 12, rich in disease resistance genes and recent gene duplications. BMC Biol 3:e20

Devos KM (2005) Updating the 'Crop Circle'. Curr Opin Plant Biol 8:155–162

Devos KM, Millan T, Gale MD (1993) Comparative RFLP maps of the homoeologous group-2 chromosomes of wheat, rye and barley. Theor Appl Genet 85:784–792

Devos KM, Wang ZM, Beales J, Sasaki T, Gale MD (1998) Comparative genetic maps of foxtail millet (*Setaria italica*) and rice (*Oryza sativa*). Theor Appl Genet 96:63–68

Devos KM, Pittaway TS, Reynolds A, Gale MD (2000) Comparative mapping reveals a complex relationship between the pearl millet genome and those of foxtail millet and rice. Theor Appl Genet 100:190–198

Dida MM, Srinivasachary, Ramakrishnan S, Bennetzen JL, Gale MD, Devos KM (2007) The genetic map of finger millet, *Eleusine coracana*. Theor Appl Genet 114:321–332

FAOSTAT Data (2006) FAO Statistical Databases (http://faostat.fao.org)

Gale MD, Devos KM (1998) Plant comparative genetics after 10 years. Science 282:656–659

Grass Phylogeny Working Group (2001) Phylogeny and subfamilial classification of the grasses (*Poaceae*). Ann Mo Bot Gard 88:373–457

Harushima Y, Yano M, Shomura A, Sato M, Shimano T, Kuboki Y, Yamamota T, Lin SY, Antonio BA, Parco A, Kajiya H, Huang N, Yamamoto K, Nagamura Y, Kurata N, Khush GS, Sasaki T (1998) A high density-rice genetic linkage map with 2275 markers using a single F_2 population. Genetics 148:479–494

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

National Research Council (1996) Lost crops of Africa; Volume I Grains. National Academy, Washington

Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Natl Acad Sci 101:9903–9908

See DR, Brooks S, Nelson JC, Brown-Guedira G, Friebe B, Gill BS (2006) Gene evolution at the ends of wheat chromosomes. Proc Natl Acad Sci 103:4162–4167

Wang XY, Shi XL, Hao BL, Ge S, Luo JC (2005) Duplication and DNA segmental loss in the rice genome: implications for diploidization. New Phytol 165:937–946

Zhang H, Jia J, Gale MD, Devos KM (1998) Relationship between the chromosomes of *Aegilops umbellulata* and wheat. Theor Appl Genet 96:69–75

Zhang D, Ayele M, Tefera H, Nguyen HT (2001a) RFLP linkage map of the Ethiopian cereal *tef* [*Eragrostis tef* (Zucc) Trotter]. Theor Appl Genet 102:957–964

Zhang H, Reader SM, Liu X, Jia J.Z., Gale MD, Devos KM (2001b) Comparative genetic analysis of the *Aegilops longissima* and *A. sharonensis* genomes with common wheat. Theor Appl Genet 103:518–525