



Comparative linkage mapping of *Oryza glumaepatula* and *Oryza sativa* interspecific crosses based on microsatellite and expressed sequence tag markers

Priscila Nascimento Rangel¹, Rosana Pereira Vianello Brondani¹, Alexandre Siqueira Guedes Coelho², Paulo Hideo Nakano Rangel³ and Claudio Brondani¹

¹Laboratório de Biotecnologia, Embrapa Arroz e Feijão, Santo Antônio de Goiás, GO, Brazil.

²Instituto de Ciências Biológicas, Universidade Federal de Goiás, Goiânia, GO, Brazil.

³Melhoramento de Arroz, Embrapa Arroz e Feijão, Goiânia, GO, Brazil.

Abstract

Molecular linkage maps representing the rice genome have been an important tool for breeding programs because they allow the elucidation of polygenic traits and are an efficient tool for monitoring wild introgressions in interspecific crosses. Common markers among rice genetic maps are important in defining the homology of chromosomes and the synteny between genomic target regions. We used 148 markers (expressed sequence tags, microsatellites and single nucleotide polymorphisms) to construct a molecular linkage map based on co-dominant markers for an interspecific backcross population using a wild rice (*Oryza glumaepatula*) from Brazil and performed a comparative analysis with other interspecific maps. The comparative analysis revealed a Spearman correlation index of 0.86 for marker order conservation to a previous map constructed for an interspecific cross using the same wild parent. Approximately 90% of markers common to other interspecific maps kept the same order. These results indicate that it will be possible to generate a unique genetic map using the wild donor and that it may be a helpful tool for breeding programs because plants derived from different interspecific populations can be rapidly scanned using markers associated with useful wild traits.

Key words: comparative linkage map, molecular markers, *Oryza glumaepatula*, rice.

Received: March 23, 2006; Accepted: January 22, 2007.

Introduction

The development of linkage maps covering the whole rice genome has provided new opportunities for the elucidation of complex polygenic traits, enabling rice geneticists to identify and manipulate the genetic information necessary to obtain superior genotypes in breeding programs. Genetic maps from interspecific crosses are efficient at detecting genomic regions associated with the control of yield-related traits and monitoring wild parental genomic fragments containing alleles of interest which have been introgressed into cultivated lines, such maps providing the basis for the identification of positive effect alleles coming from the wild parent (Brondani *et al.*, 2001). Wild rice species are well-known to be suitable alternatives to the use of wide crosses for the introduction of new allelic variation in cultivated rice lines. Previous work using *Oryza glumaepatula*, the wild relative of cultivated rice (*Oryza*

sativa), as gene donor has demonstrated that *O. glumaepatula* has the potential to successfully transfer favorable yield-related alleles, promoting positive increases in cultivated rice architecture and grain yield (Brondani *et al.*, 2002; Rangel *et al.*, 2005).

The integration in a linkage map of co-dominant molecular markers such as microsatellites (also known as simple sequence repeats or SSRs), expressed sequence tags (ESTs) and single nucleotide polymorphisms (SNPs) has allowed the definition of a framework map with “anchor” loci for rapid genome scanning, the better characterization and ranking of quantitative trait loci (QTL) alleles and the evaluation of allele interactions in QTL effects. Microsatellite markers are especially attractive for genomic analysis because hundreds of such markers are already available for rice and they are co-dominant and multi-allelic, widely distributed in genomes and easily detectable by PCR assays (Temnykh *et al.*, 2000; McCouch *et al.*, 2002). The availability of published rice functional and entire genome sequences provides an unlimited source of genetic markers, resulting in extensive public collections of ESTs.

Rice EST databases can be employed, free of cost, in the design of locus-specific primers, ESTs having the advantage of being abundant, present in gene rich regions and, despite being less polymorphic than microsatellites, showing a high frequency of transferability which supports more extensive comparative analysis than microsatellite markers because ESTs are based on the conserved coding regions of the genome (Kota *et al.*, 2003). Another factor is that markers derived from ESTs reveal the direct association between genes and quantitative traits, making EST markers candidates to aid in the identification of complex traits through the construction of linkage maps (Cato *et al.*, 2001; Zhang *et al.*, 2004). In addition, EST databases can be used in the search for SNPs, which are increasingly being used in mapping studies because they are very abundant in the genome, with an average of one SNP every 170 bp in rice (Hayashi *et al.*, 2004), providing new opportunities for the association of gene regions to quantitative traits.

Common markers for use in different rice linkage maps are important to define the homology of chromosomes, and are particularly useful for the rapid establishment of synteny between genomic target regions, including those related to QTLs for important agronomic traits. Comparative genetic studies between linkage maps generated using advanced backcross (BC) interspecific populations (BC₁F₁) have shown remarkable conservation of QTL alleles introduced from wild species to the cultivated rice gene pool across genetic backgrounds and environments. The wild parental, *Oryza rufipogon*, has been extensively employed as a common donor parent in combination with different rice cultivars as recurrent parents, and a number of QTLs have been identified at similar regions of different linkage maps (Thomson *et al.*, 2003; Septiningsih *et al.*, 2003).

The objective of this work was to develop a molecular map based exclusively on co-dominant markers (microsatellites, ESTs and SNPs), using a BC₁F₁ mapping population derived from an interspecific cross between a wild rice *Oryza glumaepatula* accession collected in a population located in the Brazilian Amazon Basin and an *Oryza sativa* cultivated rice cultivar. We also tested the hypothesis that there are enough significantly conserved linkage groups and marker order arrangement between maps constructed from interspecific crosses to allow the construction of a unique BC₁F₁ genetic map when using a common wild donor parent in combination with different cultivars as recurrent parents.

Material and Methods

Rice populations

The wild rice *Oryza glumaepatula* accession RS-16, selected from an inbred population collected from a Solimões River population in the Amazon Basin during a germplasm collection expedition (Morishima, 1994), was used

as the female parent in crosses with the irrigated rice (*Oryza sativa*) cultivar Cica-8. The F₁ plants, whose hybrid nature was confirmed with RAPD and microsatellite markers (Brondani *et al.*, 2001), were backcrossed to Cica-8, used as the recurrent parent, to obtain 186 BC₁F₁ segregating plants. These plants were genotyped with molecular markers and the information was used to construct the linkage map.

Selection of molecular markers

For the linkage analysis, 108 microsatellite markers developed from rice genome enriched libraries and previously mapped in the *O. glumaepatula* (RS-16) x *O. sativa* (BG90-2) interspecific cross (Brondani *et al.*, 2001) were selected for linkage analysis. Marker selection was performed in order to maximize the rice genome coverage and was based on the amplification quality of each marker. Of the 108 markers, 34 were from the *O. glumaepatula* (OG) series (Brondani *et al.*, 2001), 14 from the *O. sativa* (OS) series (Akagi *et al.*, 1996), 59 from the rice markers (RM) series (Chen *et al.*, 1997) and one (O484) from the sequence of the rice *Waxy* gene (Ayres *et al.*, 1997). In addition, rice genome sequences were searched in GenBank (NCBI) and Gramene web sites for all tri-nucleotide microsatellite motifs and these sequences were used to design 105 primer pairs using the Primer 0.5 software (Lincoln *et al.*, 1991).

EST-derived markers

Rice functional genome sequencing databases available at GenBank (NCBI) and Gramene were searched to develop EST-derived markers. Sequences of enzymes involved in photosynthetic and metabolic pathways, carbohydrate transportation, and grain quality and rice production traits were identified and selected. All sequences were searched for the presence of EST-derived microsatellites using the Tandem Repeats Finder (TRF) software (Benson, 1999). A total of 152 primers were designed using the Primer 0.5 software (Lincoln *et al.*, 1991). The EST-derived markers that produced high-quality amplified products but which did not detect polymorphism among the genitors were searched for EST-derived SNPs. The PCR amplified products from each genitor being sequenced at least three times in an automated ABI 3100 DNA sequencer (Applied Biosystems) and the sequences aligned using the BioEdit Sequence Alignment Editor software (Hall, 1999). All progeny were genotyped using the same strategy described above for the detection of polymorphisms on the EST-derived SNP markers.

Molecular marker analysis

The amplification reactions were performed in a final volume of 15 µL containing 15 ng of DNA, 0.3 µM of each primer, 0.25 mM of each dNTP, 5% (v/v) dimethyl sulfoxide, 10 mM of Tris-HCl (pH 8.3), 50 mM of KCl,

1.5 mM of MgCl₂ and one unit of Taq DNA polymerase. The reactions were conducted in a 96-well PT-100 Thermal Controller (MJ Research) with a pre-cycle of 94 °C for 5 min followed by 30 cycles of 94 °C for 1 min, specific primer annealing temperature (48 °C to 56 °C) for 1 min and 72 °C for 1 min, with a final extension at 72 °C for 7 min. Amplification products were separated by electrophoresis on 5% (w/v) agarose gels stained with ethidium bromide (0.1 µg/µL) or, when more resolution was needed to detect allele differences 6% (w/v) denaturing acrylamide gels stained with silver nitrate (Bassam *et al.*, 1991).

Linkage analysis

A chi-square test (χ^2) was performed for each marker to test for deviation of genotypic classes from the expected Mendelian inheritance ratios of 1:1 ($p < 0.05$). The false discovery rate (FDR) was determined from the observed p-value distribution and was used to correct multiple comparisons by controlling the probability of false positives (Benjamin & Hocheberg, 1995). Linkage analysis was performed using the Mapmaker 2.0 software for Macintosh (Lander *et al.*, 1987). Markers were placed into linkage groups using the “group” command with a minimum logarithm of the odds (LOD) score of 3.0 and a maximum recombination fraction (theta) of 0.40. The “first-order” and “compare” commands were used to identify the most probable marker order within a linkage group. The “try” command was used to place unlinked individual markers in the framework followed by the “ripple” command that was used to verify the order and establish a log likelihood support. Recombination fractions were transformed to estimate map distances by the Kosambi mapping function.

Comparative analysis

Comparative mapping analysis was carried out between the developed linkage map and the previous map obtained for the *O. glumaepatula* (RS-16) x *O. sativa* (BG90-2) interspecific cross (Brondani *et al.*, 2001). A total of 108 markers, common to both maps, were evaluated for the conservation of linkage groups and locus order arrangement. The markers order conservation was tested using a Spearman correlation test using Excel[®]. The difference between recombination fractions on the two maps was analyzed using Mantel’s test for correlation between matrices (Mantel, 1967). The recombination fraction matrices were obtained for each population and a correlation coefficient was calculated according to Mantel. This test is a randomization procedure that compares the correlation between two matrices with the correlation between one of these and randomizations of the other. The pairwise comparison was carried out with 9999 permutations using the Ntsys software version 2.0 (Rohlf, 1989). A visual comparative analysis was performed against the Cornell University *indica* x *japonica* reference map (Temnykh *et al.*, 2001; Cheng *et al.*, 2001; Temnykh *et al.*, 2000; Singh *et al.*, 1996) and the Cornell University *O. sativa* V20A x *O. glaberrima* map (Li *et al.*, 2004) using the comparative map tools comparative map viewer (Cmap) available at Gramene. In addition, two reference maps constructed for interspecific crosses with the wild relative *O. rufipogon* were also used for comparative analysis (Thomson *et al.*, 2003; Septiningsih *et al.*, 2003).

et al., 1996) and the Cornell University *O. sativa* V20A x *O. glaberrima* map (Li *et al.*, 2004) using the comparative map tools comparative map viewer (Cmap) available at Gramene. In addition, two reference maps constructed for interspecific crosses with the wild relative *O. rufipogon* were also used for comparative analysis (Thomson *et al.*, 2003; Septiningsih *et al.*, 2003).

Results and Discussion

Molecular marker polymorphism

Of the 365 markers analyzed, 109 (29.8%) either did not amplify or resulted in multiple banding patterns but the other 255 (69.8%) markers produced easily interpretable amplification products, with 157 (61.6%) displaying polymorphism. From the polymorphic markers, 120 (76.4%) were detected on high-resolution acrylamide gels. All 108 previously mapped microsatellite markers (Brondani *et al.*, 2001), were submitted to linkage analysis. From the 105 newly developed tri-nucleotide microsatellite markers, 59 (56.2%) produced amplified products, 36 (34.3%) showed polymorphism between the parents and 28 (32.5%) were selected for linkage analysis based on their broad distribution on the rice chromosomes (Table 1). Of the 152 EST markers, 55 (36.2%) were EST-derived microsatellites and 3 (1.97%) were EST-derived SNPs, the remaining 94 (61.8%) were derived from expressed sequences without tandem repeats (Table 2). The 152 EST markers produced 89 (58.5%) amplified products and these were searched for polymorphism between the genitors and 10 (11.2%) useful markers for linkage analysis were detected, 5 of which (50%) were EST-derived microsatellite markers. In the screening for EST-derived SNPs, out of a set of 29 markers that produced high quality amplified products, but were initially classified as monomorphic, sequencing analysis showed three markers with single nucleotide polymorphism, indicating that more polymorphic markers could still be derived (Table 2). Only one SNP (an adenosine to cytosine (A/C) substitution) was found along the 470 bp fragment corresponding to EST 90. Four SNPs were found for EST 26 (470 bp) and three SNPs for EST 111 (370 bp), all due to transitions.

From a total of 257 GenBank rice-sequences markers, consisting of 105 microsatellite markers derived from tri-nucleotide repeats and 152 EST-derived markers, we found 148 (57.6%) produced amplification. The low marker output of these sequences, especially for those related to ESTs, was probably due to primer extension across splice site, presence of large introns in genomic DNA sequence and primers derived from chimerical sequences (Varshney *et al.*, 2005).

Our search for microsatellite markers from EST sequences produced a higher rate of useful markers, as shown by the fact that of 55 EST-derived microsatellite markers 36 (65.5%) produced amplified fragments. The EST-

Table 1 - Microsatellite markers based on tri-nucleotide repeats obtained from sequences available on public databases that showed polymorphism between the *Oryza glumaepatula* and *Oryza sativa* parents.

Marker	GenBank number	Flanking primer sequences (F = forward, R = reverse)	Annealing temperature (°C)	Linkage group map position
4599	AY022274	F: AGAGGCGGTAGTGAATCAAC R: GTCAGCGTGGTGAATGTGTT	50	11
4611	AY022286	F: CAACTTAACCACCAAGCCAC R: CACGCCTGACGATGATTAGT	52	4
4614	AY022289	F: TCTCTCCGGACCAAGAAG R: TCACTCCACCATCCACC	50	2
4635	AY022310	F: CGCGTCGTAGCTGAACACTG R: GCGACGTGCTCGGTAGCTAA	52	4
4653	AY022328	F: CTCGGACAAGCATGATCT R: AACCGATGCAGATCAGAG	52	12
4689	AY022364	F: ATGCACATGGTGAGACGTCC R: AGTAGGTCACCCGACGGTT	50	10
4712	AY022387	F: ACAGGCTCGTGAATGGTA R: CTCATCTTCGCCTTCTTG	48	1
4749	AY022424	F: CGCCGTGACCACATCTAT R: GGCAAGGTTGGATCAGTC	52	4
4752	AY022427	F: TCCTGACCATCTCAACCTGC R: GCCGGAGAGATGATCGAGTA	50	7
4756	AY022431	F: CAGACCGATGAACTGATACG R: GGAAGAAGAGGAGCTACCAT	52	10
4770	AY022445	F: CCTTCACAGCATCCATCGTC R: GCGTGCTGCAGGATAGTTCT	48	5
4771	AY022446	F: GAGAGGAAGGAGTCGTAGGA R: GGCTCCTCCTCTATATGTC	52	4
4797	AY022472	F: GGAGAAGGCAATGCAACACG R: GCCATTGCCGCAAGTACTA	52	4
4814	AY022489	F: ACTCACAGCTCATCACTCCA R: CTACATCAAGGCTCGCTACT	50	6
4879	AY022554	F: CAGAGATCGATTGGTAGC R: CCTTGTACTIONCAGCTCCAT	52	4
4914	AY022589	F: GACTCACGGACGCTATCAAT R: TTGGCCTCCATCTCTCTAAC	54	1
4950	AY022625	F: CGGAAGAAGGCCATCGAGGT R: TGCTCGTGGTGGTGGTGTG	52	4
4961	AY022636	F: CCACTTGTCTCCTGTATGCT R: GTGATGTGAACGCCTCTACT	56	11
4968	AY022643	F: ACTACGGCATCGGCGTCACT R: GCAGCAGCAGTGGTGGTGAA	52	5
5023	AY022698	F: CGAAGGAAGCAGCGATGTGA R: AGAGATCACCTGGCGCATCA	48	2
5094	AY022769	F: GGTGATGCATGCTTGATCTG R: TAGCAAGCATTAGCAGCACC	52	11
5132	AY022807	F: CGCCATCTCTTTCAGTTCC R: GGAGAGGACGAGGAGAAGAA	52	10
5274	AY022949	F: CTCGACCTCCATGGCGAACT R: ATTGCAGACGCTCGGAGAGG	48	6
5335	AY023010	F: TTACGGCAGCTAGGCAAGAG R: GTCGAGTGGAGACAAGGAA	52	11
5370	AY023045	F: AGTCTTGACAAGCCACACAC R: AGTGCCTAGTCTCTTCTTC	56	10
5371	AY023046	F: GTACGGTCTCAGCGACGAAG R: AGCCTCTCCACCTCCTCCTT	56	12
5411	AY023086	F: GGTATTGTCCGGTGTTCAGG R: GTGAAGGCTGTACCATCCA	52	12
5623	AY023298	F: CATCTCTGCCACTGCACAC R: GGAGGAGGAGGTGTCGATCT	48	6

Table 2 - Expressed sequence tag (EST) markers showing polymorphism between the *Oryza glumaepatula* and *Oryza sativa* parents. The annealing temperature was 56 °C for EST 30 and EST 69 but 48 °C for all the other markers. An asterisk indicates EST-derived single nucleotide (SNP) markers.

Marker	GenBank number	Primer (F = forward, R = reverse)	Linkage group map position	Sequence protein	Protein identification
EST 12	AF378184	F: CCGCCTCGAGAACATGTGCT R: GCTGTGCCAATTCACCGACG	1	Stejaree45 sucrose phosphate synthase mRNA membrane protein	Protein for transmembrane sucrose transportation
EST 20	AF443603	F: GTACGACTATTGCGCCGA R: TTCACACTCCATTCTTTAAATCT	4	Xyloglucan endotransglycosylase mRNA	Enzyme from the cell wall growth
EST 21	AF467731	F: GGAGGAATCGACCGATCAAT R: CTTCATCGACACGGTGGCTA	9	Putative drought protein mRNA	Protein putatively expressed during water stress
EST 26*	AF288196	F: CCAGAGGTGGTATCAGCA R: CACTCGTTGAACTCATCG	-	mRNA violaxanthin de-epoxidase precursor, a nuclear gene for the chloroplast	Enzyme from photosynthesis pathway
EST 30	E09636	F: TTCAAGCGCACCGTTCAC R: CACGATTCTGCGGCTCA	5	Pyruvate orthophosphate dikinase (PPDK) gene	Enzyme from carbohydrates metabolism pathway
EST 41	AC097278	F: AGGTCCTTAGCTTCTGAG R: AGACTTCACTGGTGGAG	1	Sequence related to rice protein quality (unidentified)	Unidentified protein related to grain protein quality
EST 69	AB037664	F: TCTGCTGATCACGATCAC R: GCAATTCACCAGATCGTC	2	Gene for cytosolic glutamine synthetase	Enzyme from the glutamine synthesis pathway
EST 85	E09408	F: GTCTCCATCGGCACTGTT R: GGCATCAGGTTCTCTGTTG	10	Alpha-Amylase gene	Enzyme from carbohydrates metabolism pathway
EST 88	AB053295	F: GCCACCAGCAATCGTATC R: AAGGACCACGGAAGCTGT	10	mRNA for NADP dependent malic enzyme	Enzyme from photosynthesis pathway
EST 90*	AB041773	F: GTTGTCACCGTCTCACCA R: CGAGGAGAGCGGAAGAA	12	Adk-a gene for adenylate kinase	Enzyme from ADP to ATP conversion pathway
EST 94	AF136268	F: CCGCAATGCTGTGTCTCA R: TTGTGGCACCATCAACTG	2	Starch-branching enzyme I (Rbe1) mRNA	Enzyme from carbohydrates metabolism pathway
EST 111*	L18914	F: CAGCGACAAGCCTTGATG R: CATGACGTGACGCAATTC	12	Calmodulin gene	Protein, from diverse metabolic pathways, which associates with a Ca ²⁺ molecule and activates other proteins
EST 144	AP004395	F: GGTTCACCTTCTCCTTCC R: AACTTGCCGTAGTGCTCG	12	Hexokinase	Enzyme from glucose metabolism pathway

derived microsatellite markers were less polymorphic than those derived from genomic regions, in accordance with a previous report (Thiel *et al.*, 2003). Intra-genic derived markers should be less variable than markers obtained from non-coding genomic regions, as described in rice by Cho *et al.* (2000). Coding regions of the genome tend to be highly conserved due to the fact that deleterious and/or non-favorable mutations that occur on alleles of such regions are usually naturally or artificially eliminated by selection. This is true especially for cultivated crops such as rice where high selective pressure is exerted on the coding regions through selective breeding (Fraser *et al.*, 2004). Consequently, mutated alleles are usually not fixed in such populations, making the coding regions more conserved. However, despite the low level of polymorphism, EST-derived markers can be used to sample different portions of the genomes, particularly those that could not be represented using the highly polymorphic microsatellite markers derived from enriched genomic libraries, thus enriching the map with transcriptional information and opening up the

perspective of co-localization of QTLs and candidate genes in regions of higher recombination.

Other strategies to detect EST marker polymorphisms are available, such as hybridization methodologies using probes obtained from the cDNA library (Zhang *et al.*, 2004) and other methodologies such as the use of cleaved amplified polymorphic sequences (CAPS) and single-strand conformation polymorphism (SSCP) (Schneider *et al.*, 2002). Additionally, microsatellites and SNPs derived from EST sequences are valuable alternatives for enhancing EST polymorphism (Scott *et al.*, 2000; Fraser *et al.*, 2004). The large-scale EST sequencing projects that have been conducted for several species have resulted in extensive databases which can be used for *in silico* identification and the development of non-redundant EST-derived microsatellite markers (Thiel *et al.*, 2003) and EST-derived SNP markers (Hayashi *et al.*, 2004). Markers derived from EST sequences are attractive because they present higher levels of transferability to related species, making them useful for assaying the functional diversity in natural populations or germplasm collections (Gupta *et al.*, 2003). Additionally,

markers derived from genes allow the target genome region to be efficiently explored during the detection of allelic variation, especially if such markers are related to the expression of important agronomic traits used in QTL mapping. In this case, the sampled region can be used in QTL mapping experiments, potentially applied to marker-assisted selection and dissected in fine mapping or association studies.

Linkage map construction and map colinearity

The linkage analysis data set had 149 molecular markers, consisting of 108 microsatellite markers derived from genomic libraries, 28 microsatellite markers derived from GenBank sequences, 10 EST-derived markers and 3 EST-derived SNP markers. The EST-derived SNP marker EST26 could not be placed on the map using the standard parameters for linkage analysis. The remaining 148 markers were distributed over 13 linkage groups, out of the 12 existing rice chromosomes, covering 1656.9 cM of the rice genome, with an average marker distance of 11.2 cM (Figure 1). The average number of markers on each group was 11.38, ranging from 4 (group 12a) to 22 (group 2). Overall, marker distributions on the linkage groups were in agreement with previous maps (Temnykh *et al.*, 2000; Brondani *et al.*, 2001) and also were in accordance with their chromosomal location. As expected, the O484 marker, derived from the sequence of the rice *Waxy* gene, mapped at chromosome 6 (Ayres *et al.*, 1997). The extra linkage group (chromosome 12b) represents part of chromosome 12 that was fragmented into two groups at a minimum LOD score of 3.0. The inclusion of an additional set of markers would probably help to reduce unlinked markers and fragmented groups. It is also probable that the high level of linkage information available for rice could be used as a guide to identify and select strategic markers to be mapped on less saturated genomic regions and thus improve the accuracy of the linkage map.

With the inclusion of a new set of microsatellite and EST markers in the interspecific *O. glumaepatula* x *O. sativa* linkage map, the total map length increased by 11.3% as compared to the previous map constructed for the same interspecific cross (1500.4 cM; Brondani *et al.*, 2001). Several factors could have contributed for the increased map distance seen in our analysis. The integration of molecular markers of different types, such as EST-derived markers, could have resulted in a larger area of the genome being sampled as previously described in rice by Maheswaran *et al.* (1997). The new set of tri and tetra microsatellites allowed better coverage of genomic regions on chromosomes 4, 10 and 12. However, some gaps remained, including nine gaps of more than 30 cM, considering all chromosomes. In addition, the linkage of some markers to distal regions of the linkage groups and the inclusion of skewed markers (reported below) probably contributed to the increase in map length. These variations in

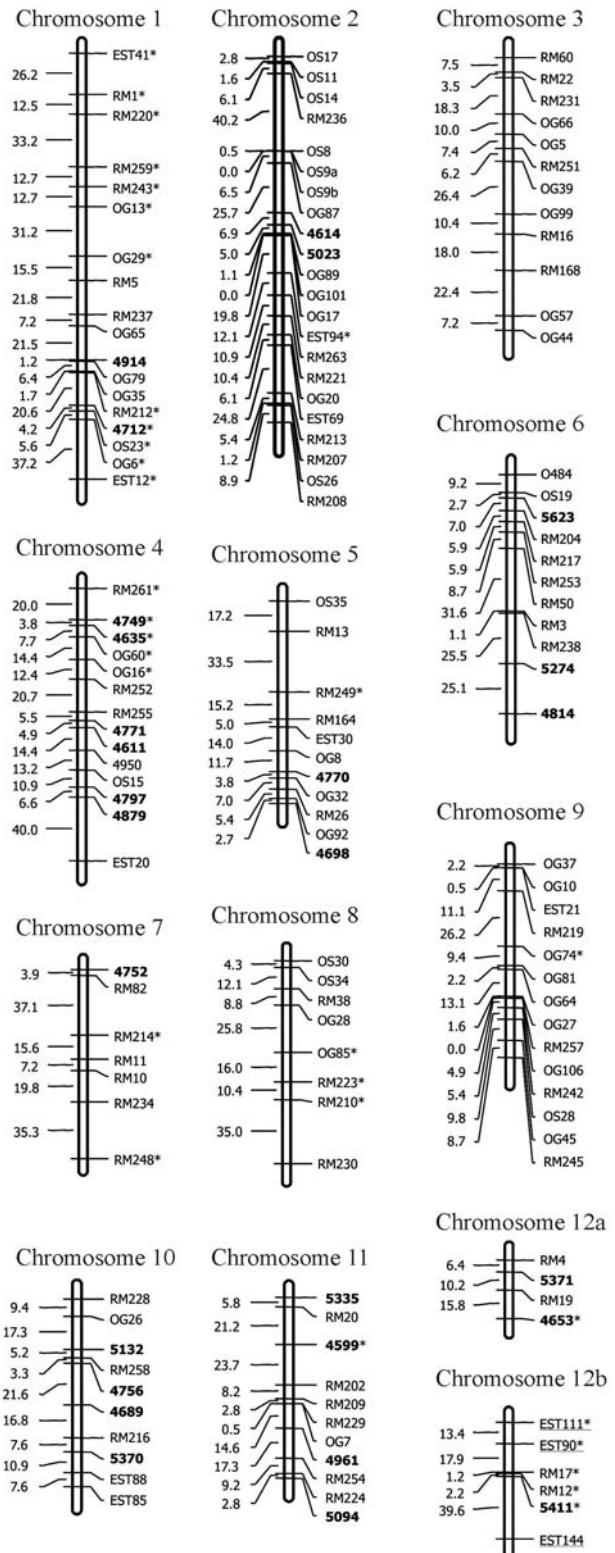


Figure 1 - Molecular linkage map of the *Oryza glumaepatula* x *Oryza sativa* interspecific population. Marker distances are expressed in centimorgans. The Kosambi mapping function was used to adjust recombination values. Distorted markers are indicated with an asterisk. Microsatellite markers based on tri-nucleotide repeats derived from the GenBank database are in bold and expressed sequence tag (EST) markers are in italics. The EST-derived SNP markers are in italics and underlined. Marker O484 (*Waxy* gene) is underlined.]

the estimation of genome size could also be due to the size of the mapping population. Distances between adjacent markers can be significantly reduced with increasing population size and the higher the number of specimens sampled the more precise is the map (Ferreira *et al.*, 2006). Our map was constructed with 147 markers and 186 plants, while the previous map was constructed with 162 molecular markers in a population of 93 plants (Brondani *et al.*, 2001). Despite the occurrence of map inflation even with a higher number of plants in our mapping population, we have strong support to believe that the incorporation of a new set of markers with a wider distribution in the genome resulted in an increase of map length.

Deviations from the expected 1:1 segregation ratio for Cica-8 homozygotes:Cica-8/*O. glumaepatula* heterozygotes (75% Cica-8 alleles/25% *O. glumaepatula* alleles) were observed for 33 markers (22.76%) with the FDR test ($p < 0.05$), which was very similar to the results for the chi-squared test performed alone. The p-value for most distorted markers was close to zero, indicating that the null hypothesis of marker segregation at the expected Mendelian inheritance ratios was false and that, therefore, distortion was probable. Most of the skewed markers were located on chromosomes 1, 4 and 8 (Figure 1), of which 21 markers (63.64%) were skewed towards Cica-8 and 12 (36.36%) were skewed towards *O. glumaepatula*. All skewed markers were retained for the linkage analysis and tended to be clustered on chromosomes 1, 4, 8 and 12 (Figure 1), which may have contributed to altering the distance and the linear order of the other markers in the linkage groups. When the skewed markers were removed from the linkage analysis, there was a decrease in the number of linked markers and in the total length of some chromosomes, resulting in a lower coverage of the genome. Marker segregation distortions have been frequently observed in rice mapping studies using interspecific populations (Causse *et al.*, 1994; Lorieux *et al.*, 2000; Brondani *et al.*, 2001). Segregation distortions may be the result of reduced recombination due to the genetic distance between the parental lines, gamete competition or abortion of gametes and zygotes (Causse *et al.*, 1994; Thomson *et al.*, 2003). In our study, most markers were skewed toward the cultivated rice parent, showing an over-representation of Cica-8 alleles. Frequently, some plants from interspecific backcrosses are discarded due to an excess of the wild-type phenotype (*e.g.* excessive height, long cycle, etc.) Consequently, it could be expected that Cica-8 alleles would be over-represented in the population. Marker deviations toward *O. glumaepatula* can be interpreted as a consequence of linkage drag in some regions, possibly influenced by sterility genes.

Comparative analysis

The Spearman correlation analysis performed using the map data from this study and the data from the map previously constructed using *O. glumaepatula* (Brondani *et al.*

et al., 2001) was limited to the common DNA sequences of the rice genome on both maps and revealed a correlation index of 0.86, indicating high marker order conservation. The Mantel test revealed a correlation index of 0.79 between recombination fraction matrices. These results indicate that markers distances and orders were well-conserved between the two maps with only slight differences. Out of 108 microsatellite markers common to both maps, 96 (88.9%) kept the same order, 12 (11.1%) showed order inversions and only one marker (OG32) mapped to different linkage groups (Chromosome 5 in the present work; Chromosome 6 in the previous work by Brondani *et al.* (2001) (Figure 2). The lack of conservation found for some marker orders and distances was most probably due to the different population sizes used in map construction in the two studies. The higher number of meiotic events sampled in our study produced more robust recombination fraction estimates, a lar-

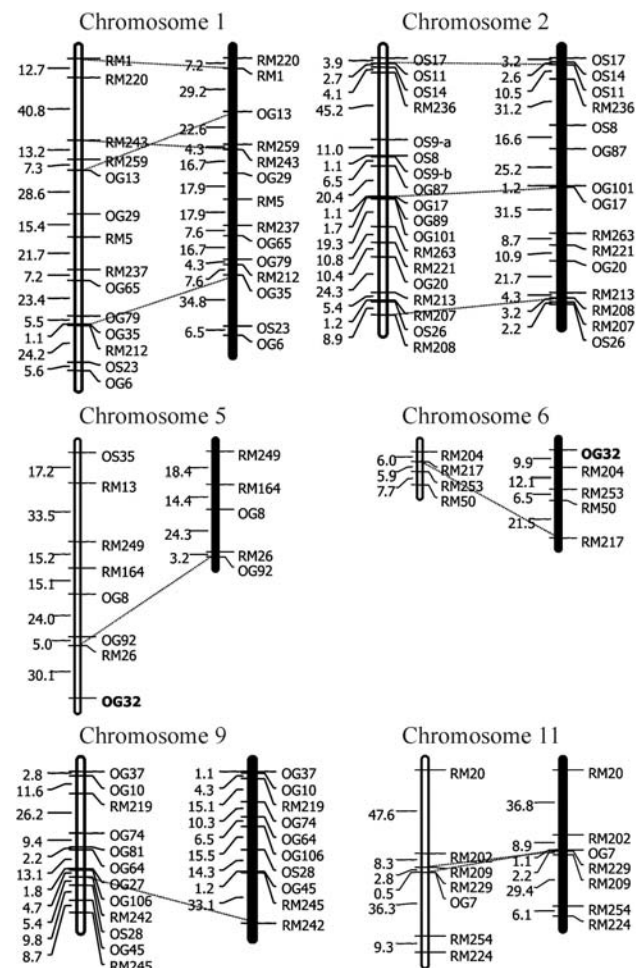


Figure 2 - Comparative map of the two populations derived from the *O. glumaepatula* x *O. sativa* crosses. The open bars are the RS-16 (*O. glumaepatula*) x Cica-8 (*O. sativa*) linkage groups and the filled bars are the RS-16 (*O. glumaepatula*) x BG90-2 (*O. sativa*) linkage groups. All markers in the groups were common to both populations. The discontinuous lines show the markers that had their orders inverted between the two maps. The OG32 marker that mapped to different linkage groups is in bold.

ger mapping population being more appropriate for ordering closely spaced markers with a higher confidence level. In general, the order and distance conservation of markers between the two maps were high in our study, indicating that our map may be used as a reference for future QTL studies involving the wild *O. glumaepatula* as genitor in interspecific crosses with cultivated rice.

The comparative analysis performed between our map and four other maps showed the following number of common markers: 33 for the Cornell University *indica* x *japonica* map (Temnykh *et al.*, 2001), 6 for the Cornell University *O. sativa* V20A x *O. glaberrima* map (Li *et al.*, 2004), and 29 and 26, respectively, for the maps constructed by Thomson *et al.* (2003) and Septiningsih *et al.* (2003) for interspecific crosses with the wild relative *O. rufipogon*. The analyses were done by visually comparing marker orders on linkage groups with more than three common markers. In general, marker orders were well conserved when comparing the present map and the other four maps, demonstrating that rice microsatellite marker positions are well established and that anchor markers can be selected for mapping studies using different types of populations for purposes of comparative studies with other grasses or with different rice populations.

In conclusion, our results indicate that it will be possible to align and generate a unique genetic BC₁F₁ map when using a common wild donor parent in combination with different cultivars as recurrent parent. However, it is important to point out that our comparative analysis was performed at the macro level, with markers spaced at millions of DNA base pairs in the genome. Thus, our analysis does not provide precise information about genomic regions related to traits of economical interest or whether the QTLs detected in different populations will be present at the same regions considering different backgrounds and environmental conditions under which the crosses are performed. However, the consensus map can be used to rapidly construct a genome framework using a set of anchor loci that will be useful for predicting the location of key regions and exploring the genome variation at this level. Furthermore, since our study established map distances it will be possible to explore the utilization of the wild genitor *O. glumaepatula* in a wide range of crosses with cultivated rice and at a lower cost and in less time. In addition, our map will be highly useful for the rice breeding program in Brazil, since the information obtained from this map has been used to guide the development of lines containing introgressed regions of the genome from the wild species *O. glumaepatula* that will be used as genitors in the program or directly as new cultivars.

Acknowledgments

We thank Gerusa Vaz de Souza and Airton Barbosa Marques from Embrapa Rice and Beans for providing the *O. glumaepatula* x *O. sativa* interspecific population, the

Brazilian Ministry of Education (CAPES) for the M.Sc. fellowship extended to PNR and the Brazilian National Research Council (CNPq – PROFIX) for the financial support.

References

- Akagi H, Yokozeki Y, Inagaki A and Fujimura T (1996) Microsatellite DNA markers for rice chromosomes. *Theor Appl Genet* 93:1071-1077.
- Ayres NM, McClung AM, Larking PD, Bligh HFJ, Jones CA and Park WD (1997) Microsatellites and a single nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germplasm. *Theor Appl Genet* 94:773-781.
- Bassam BJ, Cactano-Anolles G and Gresshoff PM (1991) Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal Biochem* 196:80-83.
- Benjamin Y and Hocheberg Y (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Jour Royal Stat Soc* 57:289-300.
- Benson G (1999) Tandem Repeats Finder: A program to analyze DNA sequences. *Nucleic Acids Res* 27:573-580.
- Brondani C, Brondani RPV, Rangel PHN and Ferreira ME (2001) Development and mapping of *Oryza glumaepatula*-derived microsatellite markers in the interspecific cross *Oryza glumaepatula* X *Oryza sativa*. *Hereditas* 134:59-71.
- Brondani C, Brondani RPV, Rangel PHN and Ferreira ME (2002) QTL mapping and introgression of yield related traits from *Oryza glumaepatula* to cultivated rice (*Oryza sativa*) using microsatellite markers. *Theor Appl Genet* 104:1192-1203.
- Cato SA, Gardner RC, Kent J and Richardson TE (2001) A rapid PCR-based method for genetically mapping ESTs. *Theor Appl Genet* 102:296-306.
- Causse MA, Fulton TM, Cho YG, Ahn SN, Chunwongse J, Wu K, Xiao J, Yu Z, Ronald PC, Harrington SE, *et al.* (1994) Saturated molecular map of the rice genome based on an interspecific backcross population. *Genetics* 138:1251-1274.
- Chen X, Temnykh S, Xu Y, Cho YG and McCouch SR (1997) Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.). *Theor Appl Genet* 95:553-567.
- Cheng Z, Presting GG, Buell R, Wing RA and Jiang J (2001) High-resolution pachytene chromosome mapping of bacterial artificial chromosomes anchored by genetic markers reveals the centromere location and the distribution of genetic recombination along chromosome 10 of rice. *Genetics* 157:1749-1757.
- Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Park WD, Ayres N and Cartinhour S (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). *Theor Appl Genet* 100:713-722.
- Ferreira A, da Silva MF, Costa e Silva L and Cruz CD (2006) Estimating the effects of population size and type on the accuracy of genetic maps. *Genet Mol Biol* 29:187-192.
- Fraser LG, Harvey CF, Crowhurst RN and De Silva HN (2004) EST-derived microsatellites from *Actinidia* species and their potential for mapping. *Theor Appl Genet* 108:1010-1016.
- Gupta PK, Rustgy ES, Sharma ES, Singh R, Kumar EN and Balyan EHS (2003) Transferable EST-SSR markers for the

- study of polymorphism and genetic diversity in bread wheat. *Mol Gen Genomics* 270:215-323.
- Hall TA (1999) BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids* 41:95-98.
- Hayashi K, Hashimoto N, Daigen M and Ashikawa I (2004) Development of PCR-based SNP markers for rice blast resistance genes at the *Piz* locus. *Theor Appl Genet* 108:1212-1220.
- Kota R, Rudd S, Facius A, Koselov G, Thiel T, Zhang H, Stein N, Mayer K and Graner A (2003) Snipping polymorphisms from large EST collections in barley (*Hordeum vulgare* L.). *Theor Appl Genet* 270:24-33.
- Lander ES, Green P, Abrahamson J, Barlow MJ, Daly MJ, Lincoln SE and Newburg L (1987) MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1:174-181.
- Li J, Xiao J, Grandillo S, Jiang L, Wan Y, Deng Q, Yuan L and McCouch SR (2004) QTL detection for rice grain quality traits using an interspecific backcross population derived from cultivated Asian (*O. sativa* L.) and African (*O. glaberrima* S.) rice. *Genome* 47:697-704.
- Lincoln SE, Daly MJ and Lander ES (1991) "Primer" software. Available from E. Lander, Whitehead Institute, Cambridge, MA.
- Lorieux M, Ndjiondjop N and Ghesquière A (2000) A first interspecific *Oryza sativa* X *Oryza glaberrima* microsatellite-based genetic linkage map. *Theor Appl Genet* 100:593-601.
- Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Res* 27:209-220.
- McCouch SR, Teytelman L, Xu Y, Lobos KB, Clare K, Walton M, Fu B, Maghirang R, Li Z, Xing Y, *et al.* (2002) Development and mapping of new 2240 SSR markers for rice (*Oryza sativa* L.). *DNA Res* 9:199-207.
- Maheswaran M, Subudhi PK, Nandi S, Xu JC, Parco A, Yang DC and Huang N (1997) Polymorphism, distribution, and segregation of AFLP markers in a doubled haploid rice population. *Theor Appl Genet* 94:39-45
- Morishima H (1994) Background information about *Oryza* species in tropical America. In: Morishima H and Martins PS (eds) *Investigations of Plant Genetic Resources in the Amazon Basin with the Emphasis on the Genus Oryza*. Report of 1992/1993, Amazon Project, p. 4-5.
- Rangel PHN, Brondani C, Brondani RPV, Rangel PN and Zimmermann JF (2005) Development of rice lines with gene introgression from the wild *Oryza glumaepatula* by the AB-QTL methodology. *CBAB* 5:10-19.
- Rohlf FJ (1989) NTSYS-PC: Numerical Taxonomy and Multivariate Analysis System. Exeter Publisher, New York, 236 pp.
- Schneider K, Schäfer-Pregl R, Borchardt DC and Salamini F (2002) Mapping QTLs for sucrose content in a sugar beet population fingerprinted by EST-related markers. *Theor Appl Genet* 104:1107-1113.
- Scott KD, Egger P, Seaton G, Rossetto M, Ablett EM, Lee LS and Henry RJ (2000) Analysis of SSR derived from grape ESTs. *Theor Appl Genet* 100:723-726.
- Septiningsih EM, Prasetyono J, Lubis E, Tai TH, Tjubaryat T, Moeljopawiro S and McCouch SR (2003) Identification of quantitative trait loci for yield and yield components in an advanced backcross population derived from the *Oryza sativa* variety IR64 and the wild relative *O. rufipogon*. *Theor Appl Genet* 107:1419-1432.
- Singh K, Ishii T, Parco A, Huang N, Brar DS and Khush GS (1996) Centromere mapping and orientation of the molecular linkage map of rice (*Oryza sativa* L.). *Proc Natl Acad Sci USA* 93:6163-6168.
- Temnykh S, Park WD, Ayres N, Cartinhour S, Hauck N, Lipovich L, Cho YG, Ishii T and McCouch SR (2000) Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). *Theor Appl Genet* 100:697-712.
- Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S and McCouch SR (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations and genetic marker potential. *Genome Res* 11:1441-1451.
- Thiel T, Michalek W, Varshney RK and Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet* 106:411-422.
- Thomson MJ, Tai TH, McClung AM, Lai XH, Hinga ME, Lobos KB, Xu Y, Martinez CP and McCouch SR (2003) Mapping quantitative trait loci for yield, yield components and morphological traits in an advanced backcross population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson. *Theor Appl Genet* 107:479-493.
- Varshney RK, Graner A and Sorrells ME (2005) Genic microsatellite markers in plants: Features and applications. *Trends Biotech* 23:48-55.
- Zhang WK, Wang YJ, Luo GZ, Zhang JS, He CY, Wu XL, Gai JY and Chen SY (2004) QTL mapping of ten agronomic traits on the soybean (*Glycine max* L. Merr.) genetic map and their association with EST markers. *Theor Appl Genet* 108:1131-1139.

Internet Resources

National Center for Biotechnological Information (NCBI), <http://www.ncbi.nlm.nih.com> (January 17, 2004)

Gramene: A resource for comparative grass genomics, <http://www.gramene.org>. (February 18, 2004)

Associate Editor: Márcio de Castro Silva Filho