Theor Appl Genet (2010) 120:563–572
DOI 10.1007/s00122-009-1176-1

ORIGINAL PAPER

A universal core genetic map for rice

Julie Orjuela · Andrea Garavito · Matthieu Bouniol · Juan David Arbelaez · Laura Moreno · Jennifer Kimball · Gregory Wilson · Jean-François Rami · Joe Tohme · Susan R. McCouch · Mathias Lorieux

Received: 22 December 2008 / Accepted: 30 September 2009 / Published online: 22 October 2009

Abstract To facilitate the creation of easily comparable, low-resolution genetic maps with evenly distributed markers in rice (*Oryza sativa* L.), we conceived of and developed a Universal Core Genetic Map (UCGM). With this aim, we derived a set of 165 *anchors*, representing clusters of three microsatellite or simple sequence repeat (SSR) markers arranged into non-recombining groups. Each anchor consists of at least three, closely linked SSRs, located within a distance below the genetic resolution provided by common, segregating populations (<500 individuals). We chose anchors that were evenly distributed across the rice chromosomes, with spacing between 2 and 3.5 Mbp (except in the telomeric regions, where spacing was 1.5 Mbp). Anchor selection was performed using in silico tools and data: the *O. sativa* cv. Nipponbare rice genome sequence, the CHARM tool, information from the Gramene database and the OrygenesDB database. Sixteen AA-genome accessions of the *Oryza* genus were used to evaluate polymorphisms for the selected markers, including accessions from *O. sativa*, *O. glaberrima*, *O. barthii*, *O. rufipogon*, *O. glumaepatula* and *O. meridionalis*. High levels of polymorphism were found for the tested *O. sativa* × *O. glaberrima* or *O. sativa* × wild rice combinations. We developed Paddy Map, a simple database that is helpful in selecting optimal sets of polymorphic SSRs for any cross that involves the previously mentioned species. Validation of the UCGM was done by using it to develop three interspecific genetic maps and by comparing genetic SSR locations with their physical positions on the rice pseudomolecules. In this study, we demonstrate that the UCGM is a useful tool for the rice genetics and breeding community, especially in strategies based on interspecific hybridisation.

Communicated by E. Guiderdoni.

J. Orjuela · A. Garavito · M. Bouniol · J. D. Arbelaez · L. Moreno · J. Tohme · M. Lorieux (✉)
Agrobiodiversity and Biotechnology Project,
International Center for Tropical Agriculture (CIAT),
AA 6713 Cali, Colombia
e-mail: m.lorieux@cgiar.org

Present Address:
J. Orjuela · A. Garavito · M. Bouniol · M. Lorieux
Plant Genome and Development Laboratory (UMR 5096,
CNRS, IRD, UPVD), Institut de Recherche pour le
Développement (IRD), 911 Av. Agropolis,
34394 Montpellier Cedex 5, France

J. Kimball · G. Wilson · S. R. McCouch
Department of Plant Breeding and Genetics, Cornell University,
162 Emerson Hall, Ithaca, NY 14853-1901, USA

J.-F. Rami
Centre de Coopération Internationale en Recherche
Agronomique pour le Développement (CIRAD), UMR DAP,
TA A-96/03, Avenue Agropolis, Montpellier, France

Introduction

Integrating information about gene and quantitative trait locus (QTL) locations on different genetic maps has been a constant issue since molecular markers were discovered. One of the main difficulties lies in the fact that maps of the populations under study do not share the same set of markers. Hence, mathematical projections are the only methods of comparing genetic maps that were constructed using different marker sets. However, this is not an optimal method, considering the variation of recombination frequency and the difference in marker density between populations. In the case of rice, the situation has greatly improved since the genomes of the Nipponbare and 93-11

cultivars have been sequenced (IRGSP 2005; Wu et al. 2002; Yu et al. 2002). It is now possible to accurately compare two genetic maps, looking at the physical positions of the markers in the Nipponbare and 93-11 pseudomolecules. Public databases such as Gramene (http://www.gramene.org), Oryzabase (http://www.shigen.nig.ac.jp/rice/oryzabase/top/top.jsp) or Rice-BRCdb (http://rice-brcdb.cines.fr) are tools that make it easier to compile QTL/gene location information on the rice physical map. However, some drawbacks still remain with this approach when developing a new genetic map: (1) finding a set of polymorphic markers that are evenly dispersed on the genome still requires a fair amount of work; (2) direct comparison between QTL locations on different genetic maps is still cumbersome.

Microsatellites are repeats of short DNA motifs (1–6 bp in length) arranged in tandem, exhibiting high variability in the number of repeats at a given locus. These simple sequence repeats (SSR) are a valuable source of genetic markers (Temnykh et al. 2001). They are codominant and highly polymorphic, making them ideal markers for use in a wide variety of parental combinations. In rice, a high-density SSR map with genome coverage of approximately 2 SSRs per centimorgan (cM) was reported by the International Rice Microsatellite Initiative (IRMI) (McCouch et al. 2002). This map can be exploited for developing markers that are tightly linked to traits of interest. Subsequently, a total of 18,828 Class 1 (di-, tri- and tetra-nucleotide) SSRs, representing 47 distinctive motif families, were identified and annotated from the complete sequence of the rice genome (IRGSP 2005). More recently, a set of 52,000 SSRs polymorphic between *indica*/*japonica* were reported (Zhang et al. 2007) (however, the URL provided in the publication does not appear to be functional and we could not obtain information from the authors). More than 80% of the primers designed for *O. sativa* successfully amplify in other AA-genome species (Lorieux et al. 2000), and find immediate application in the field of diversity and genomic studies.

The genomic information gathered about rice during the past 15 years makes it possible to optimise genetic mapping strategies, which in turn facilitates the search for genes or QTLs that control traits of interest for breeding. A genetic linkage map is a linear representation of the positions of various loci (e.g. markers, genes) along chromosomes. Distances between loci depend on the recombination rates observed in a particular segregating population. Hence, the distance between two specific loci may vary considerably from one population to another. This makes it difficult to compare QTL/gene locations between different genetic linkage maps. Moreover, genetic maps built on sets of "core" markers frequently suffer from poor genome saturation, particularly in chromosomal regions with low recombination frequency. Therefore, it is desirable to identify a set of well-distributed "core" markers for whole genome mapping based on physical rather than on genetic position.

The objective of our study was to develop a UCGM for rice that would be useful in our efforts to simultaneously develop several sets of chromosome segment substitution lines (CSSL), where each line carried a single or few chromosomal fragments of a donor genotype in a common genetic background (Eshed and Zamir 1995; Jeuken et al. 2008). A complete set of CSSLs represents the whole donor genome divided into small overlapping segments, where interactions between donor alleles are restricted to those falling within substituted regions within a given line. This makes CSSL populations useful materials for identification of QTLs with minor effects (Howell et al. 1996). Several CSSL populations have been developed for rice and found to be widely useful (Doi et al. 1997; Ebitani et al. 2005; Ghesquière et al. 1997; Jena et al. 1992; Kubo et al. 2002; Li et al. 2005; Lorieux et al. 2000; Sobrizal et al. 1999; Tian et al. 2006; Wan et al. 2004; Yu et al. 2005; Zhang et al. 2006; Tan et al. 2007).

By developing a UCGM for rice, we aimed to improve the efficiency of our efforts to develop interspecific rice CSSL populations within the framework of a collaborative Generation Challenge Programme project (http://www.generationcp.org/). By providing a common set of markers that could be used on multiple populations, we were able to reliably trace the introgression of chromosomal fragments in different populations in different laboratories. In this paper, we explain the concept of Universal Core Genetic Map (UCGM), and demonstrate its value in a distributed research program. This tool represents a set of well-dispersed, polymorphic markers that can be used for mapping in almost any interspecific combination of *O. sativa* × wild rice, enhancing the ability to compare genetic maps developed on different populations, and facilitating collaboration among research groups. The concept is valid for intraspecific crosses as well, though additional markers may be necessary to ensure polymorphism in more closely related materials. We also created a computer program, Paddy Map, which facilitates the selection of standard, well-distributed and universal marker sets for any kind of segregating population.

Materials and methods

Concept

The basic concept of Universal Core Genetic Map is the following: starting from the list of SSRs identified in the rice genomic sequence (IRGSP 2005), we extracted a set of

anchors. We defined an anchor as a group of closely linked SSRs located in a specific genomic position. The distance that separates the first and last SSRs within an anchor was chosen to be below the genetic resolution of common segregating populations (<500 individuals), based on the mean kbp/cM ratio calculated from the genetic distances of a saturated genetic map. This allowed us to consider the anchor as a single, or discrete, locus. Anchors were chosen to be evenly distributed along the genome, with spacing between anchors defined by physical distance (in kb). Taking the high recombination rate in telomeric regions into account, a higher density of anchors was required to avoid large gaps between anchors in the subsequently derived genetic maps. Similarly, the low recombination rate in most centromeric regions required a lower density of anchors in those regions.

The main property of an anchor is to be more polymorphic than a single SSR. Let us rewrite the event "At least one SSR belonging to an anchor is polymorphic" as "An anchor is polymorphic", since the two events are indeed identical. It is plain to see that the probability of an anchor being polymorphic between two given genotypes is higher than that of a single SSR being polymorphic. The difference between the two probability values is proportional to the number of SSRs per anchor; higher the number of SSRs per anchor, greater the difference.

This polymorphism property of the anchors is exploited in generating genetic maps that are no longer based on markers, but on anchors. As we will see in the case of interspecific crosses between AA-genome rice species, anchors made of triplets of SSRs make it possible to generate complete genetic maps that share the same set of polymorphic anchors, with very few exceptions. In this way, the maps can be directly compared, without the need for mathematical projection.

Probabilities of polymorphism

Let P_1 be the observed probability of an SSR being polymorphic, P_2 the probability of an anchor being polymorphic, and n the number of SSRs per anchor. Then:

$$P_2 = 1 - (1 - P_1)^n$$

According to our observations, P_1 was >0.8 in every interspecific combination we have tested so far. Taking the value of P_1 at its minimum, 0.8, gives $P_2 = 0.992$ with $n = 3$. In the case of an *O. sativa* × *O. meridionalis* combination, which we are working on in our laboratory, we got $P_1 = 0.976$ and $P_2 \approx 1.0$.

Let P_3 be the probability of all anchors being polymorphic and Na the total number of anchors along the genome. Then:

$$P_3 = P_2^{\mathrm{Na}}$$

In the case of *O. sativa* × *O. meridionalis*, and taking Na = 165 and $n = 3$, we get $P_3 = 0.997$ with $n = 3$, which is quite a high value.

In Table 1, we show how P_1, n and Na modify the P_3 value.

Selection of SSR markers

We used the following criteria to include an SSR marker in the list of candidates for anchor building (1) it has a unique

Table 1 P_3 probability that all Na anchors are polymorphic, given n markers per anchor and P_1 the observed probability that a marker is polymorphic

Na P_1	100	150	165	200	250
$n = 3$					
0.5	0.000002	0.000000	0.000000	0.000000	0.000000
0.705	0.074217	0.020219	0.013688	0.005508	0.001501
0.8	0.447886	0.299744	0.265721	0.200602	0.134251
0.85	0.713145	0.602236	0.572457	0.508575	0.429481
0.9	0.904792	0.860643	0.847824	0.818649	0.778703
0.95	0.987577	0.981424	0.979585	0.975308	0.969231
0.99	0.999900	0.999850	0.999835	0.999800	0.999750
$n = 4$					
0.5	0.001574	0.000062	0.000024	0.000002	0.000000
0.705	0.467565	0.319715	0.285259	0.218617	0.149487
0.8	0.852035	0.786477	0.767811	0.725963	0.670105
0.85	0.950623	0.926856	0.919843	0.903684	0.881091
0.9	0.990049	0.985111	0.983635	0.980198	0.975309
0.95	0.999375	0.999063	0.998969	0.998751	0.998439
0.99	0.999999	0.999999	0.999998	0.999998	0.999998
$n = 5$					
0.5	0.041800	0.008546	0.005308	0.001747	0.000357
0.705	0.799584	0.714984	0.691394	0.639334	0.571689
0.8	0.968502	0.953126	0.948562	0.937995	0.923105
0.85	0.992435	0.988674	0.987548	0.984927	0.981194
0.9	0.999000	0.998501	0.998351	0.998002	0.997503
0.95	0.999969	0.999953	0.999948	0.999938	0.999922
0.99	1.000000	1.000000	1.000000	1.000000	1.000000
$n = 10$					
0.5	0.906917	0.863677	0.851112	0.822499	0.783284
0.705	0.999501	0.999252	0.999177	0.999002	0.998753
0.8	0.999990	0.999985	0.999983	0.999980	0.999974
0.85	0.999999	0.999999	0.999999	0.999999	0.999999
0.9	1.000000	1.000000	1.000000	1.000000	1.000000
0.95	1.000000	1.000000	1.000000	1.000000	1.000000
0.99	1.000000	1.000000	1.000000	1.000000	1.000000

The P_1 values were chosen to represent different situations. $P_1 = 0.5$: *indica* × *japonica* (general case); $P_1 = 0.705$: *indica* × *japonica* (our study); $P_1 = 0.8$–0.99: *O. sativa* × wild

position in the rice genome, (2) it belongs to the Class 1 microsatellite family (McCouch et al. 2002), and (3) it shows a high polymorphism information content (PIC) in *indica–japonica* crosses or in interspecific crosses. In the event of several equivalent candidates for an anchor according to the three above-mentioned criteria, we gave priority to SSRs that were previously used for genotyping in the literature and/or in our laboratory.

The candidate markers were selected based on the list of 18,828 SSRs published by the International Rice Genome Sequencing Project (IRGSP 2005). To verify the position of each SSR on the rice genome sequence, we performed a Blast alignment of each primer pair against the rice pseudomolecules (TIGR, v. 5; http://rice.plantbiology.msu.edu/). The blast searches were carried out using a command-line version of the Primer Blaster tool available on the OrygenesDB database Web page (http://orygenesdb.cirad.fr). The search for anchors was performed using both the CHARM database (http://sliver.plbr.cornell.edu/ssr) and a Visual Basic script that we designed for this purpose.

Genetic materials

To evaluate the polymorphism levels of the defined anchors, parental lines from the populations developed within the framework of our research projects were selected. These included three accessions of *O. sativa* ssp. *indica*, four of *O. sativa* ssp. *japonica*, three of *O. glaberrima* and six wild-rice biotypes of *O. meridionalis*, *O. rufipogon*, *O.barthii*, and *O. glumaepatula* (Table 2).

Table 2 List of the AA-genome *Oryza spp.* accessions used for evaluation of SSR polymorphism

Species	Accession
O. sativa ssp. *indica*	93-11
O. sativa ssp. *indica*	IR36
O. sativa ssp. *indica*	IR64
O. sativa ssp. *japonica*	Nipponbare IRGC12731
O. sativa ssp. *tropical japonica*	Azucena
O. sativa ssp. *tropical japonica*	Caíapo
O. sativa ssp. *tropical japonica*	Curinga
O. glaberrima	TOG5681
O. glaberrima	IRGC103544 (MG12)
O. glaberrima	CG14
O. barthii	IRGC105613 (3232)
O. barthii	IRCG101937 (3239)
O. rufipogon	IRGC105491
O. rufipogon	IRGC105890
O. glumaepatula	GEN1233
O. meridionalis	OR44

The seeds were provided by the CIAT Rice Product Line research program, Embrapa-CNPAF (Goiania, Brazil) and IRD (Montpellier, France). Twenty seeds of each biotype were sown under greenhouse conditions for subsequent DNA extraction.

Total genomic DNA extraction

Fresh leaf tissue was collected 16 days after sowing. Total DNA was extracted on 96-well plates using the cetyltrimethylammonium bromide (CTAB) method (Murray and Thompson 1980) with some modifications. The DNA quality was verified by means of visualization of each sample in a 0.8% agarose gel stained with ethidium bromide, and then quantified using a DyNA Quant 200 fluorometer (Hoefer). Finally, DNA was diluted to 4 ng/µl and stored at −20°C.

PCR optimization

The DNAs of four control rice accessions were used as a template to choose the best amplification conditions for each primer pair: two *O. sativa* (accessions IR64 and Caíapo) and two *O. glaberrima* (accessions TOG5681 and MG12). These accessions correspond to the progenitors of two CSSL interspecific populations that are being developed in our laboratory. The PCR conditions were optimised in varying the annealing temperature of the primers and the magnesium concentration in the PCR. When no or too faint an amplification was obtained, the PCR was repeated using Nipponbare accession as control to discard the possibility of PCR failing in the control accessions.

Evaluation of SSR primer pairs

PCR was performed in a final volume of 15 µl containing 20 ng of template DNA, 1× buffer (Tris–HCL 100 mM, 500 mM of KCl, Triton X-100 1%), 2 ng/µL of $MgCl_2$, 0.3 µM d-NTP, 2 pmol of each primer, and one unit of Taq polymerase. DNA amplification was performed on an MJ Research PTC-220^{TM} thermocycler, programmed for an initial 3 min at 94°C, followed by 30 cycles of 30 s at 94°C, 45 s from 50 to 61°C according to the optimal temperature for each SSR primer, and 1 min at 72°C, with a final extension of 5 min at 72°C. The amplified products were separated electrophoretically using PAGE 4% (Sambrook and Rusell 2001) silver stained and photographed. Molecular weights were estimated using a 10-pb DNA ladder (GibcoBRL, Invitrogen). For some primer combinations, the polymorphism was also evaluated using a Li-cor sequencer 4200. For this purpose, 0.5 pmol of M13 IRD700 or IRD800 primers was added by reaction.

Polymorphism information content

We determined the PIC of 489 SSRs belonging to the anchor list, using the 16 AA-genome rice accessions listed in Table 2. Allelic information was compiled in a database, Paddy Map. The PIC value was calculated with the help of a small program written in Visual Basic for Applications running in MS Excel 2003.

Testing the UCGM

The UCGM was used to generate genetic maps of three interspecific BC_1F_1 populations derived from the following crosses: (1) *O. sativa* ssp. *indica* cv. IR64 × *O. glaberrima* acc. TOG5681, (2) *O. sativa* ssp. *japonica* cv. Curinga × *O. rufipogon* acc. IRGC105491, and (3) *O. sativa* BRSMG Curinga × *O. meridionalis* acc. OR44 (or W2112). The genetic maps were compared both for marker order and map distances with the expected map based on the genomic sequence. To get equivalent distance units between the physical and genetic maps, we used a conversion rate of 200 kbp/cM, which is a mean value obtained from the comparison of several rice genetic maps with the chromosome physical sizes.

Results

Finding anchor sets

A set of 165 well-distributed anchors was identified, each containing at least three, tightly linked SSR markers and representing a total of 517 SSR loci. The average distance covered by each triplet of SSR is 178,501 bp and the mean genomic distance between two anchors was approximately 2.5 megabases (Mb), except in the telomeric regions, where it was approximately 1.5 Mb (Fig. 1). Using this strategy, we were able to cover the whole rice genome with no significant gap between adjacent anchors.

The precise genomic positions of the 512 SSRs and their 165 anchors, together with the primer sequences, are available as electronic resources at (http://mapdisto.free.fr/PaddyMap/UCGM/).

Anchor polymorphism in selected germplasm

Using 16 diverse accessions (described in "Materials and methods"), a total of 3,207 alleles were obtained based on evaluation of 489 of the 512 SSRs markers. The number of alleles per locus varied widely, ranging from 2 (RM6233, RM1063, RM12923, RM15065, RM14327, RM16382, RM20155, RM6652, RM172, RM5999, RM23769, RM26001, RM6327) to 16 (RM8131), with an average of 6.55 alleles/locus. As expected, these values are higher than the values reported in other studies based on isoenzymes (Glaszmann 1987; Second 1982) and restriction fragment length polymorphisms (RFLP) (Zhang et al. 1997). They were similar to the values found in other studies based on SSR markers (Ni et al. 2001; Garris et al 2005; Blair et al. 2002). Inclusion of wild species in this study induced the presence of a high number of rare alleles.

Most of the markers produced successful amplification in all tested accessions, with 98.5 and 98.2% of positive reactions for *japonica* and *indica* accessions, respectively and 78.9% for *O. meridionalis* (accession OR44) and 95.7% for *O. glaberrima* (accession TOG 5681) (Table 3). These values are similar to those found by (Song et al. 2003), who evaluated the genetic diversity in *O. rufipogon* populations with 23 SSR markers, and where all of the markers did amplify. However, they are higher than those found in a study of the population structure of *O. glumaepatula* (Brondani et al. 2005).

The observed rate of polymorphic SSRs (RPS) was 85.4 and 70.4% for inter- and intra-specific crosses, respectively. The RPS ranged from 79.8 to 90.5% for *O. sativa* × wild rice crosses, with a mean of 85.8%, while RPS was slightly lower in *O. sativa* × *O. glaberrima* crosses, ranging from 83.2 to 88.8%. The average RPS between the parents of the six CSSL populations under development was 86.2%, including two *O. sativa* × *O. glaberrima*, one *O. sativa* × *O. rufipogon*, one *O. sativa* × *O. barthii*, one *O. sativa* × *O. glumaepatula*, and one *O. sativa* × *O. meridionalis* crosses.

The PIC ranged from 0.110 (RM6233, RM15065, RM16382, RM5999) to 0.908 (RM6738, RM8131), with an average of 0.691. This is evidence of high mutation frequencies in the tested accessions, confirming the usefulness of the tested SSRs for diversity analysis and mapping purposes (Olufowote et al. 1997; Garris et al. 2005). However, the relation between the number of alleles and the PIC is not a direct one, an observation that was already reported in other studies (Huang et al. 2003; Matus and Hayes 2002).

Database

We developed a simple database, Paddy Map, to provide the scientific community with a user-friendly tool for access to the information described in this paper. Polymorphic SSR markers for any cross between the species analysed in this study can be displayed, as well as their attached anchor and position in the genome. Several other features, such as searching for markers within a specific anchor, or computing various types of statistics, are implemented. Paddy Map is available for download at (http://mapdisto.free.fr/PaddyMap/).

Fig. 1 **a** Distribution of SSR markers that form the Universal Core Genetic Map, arranged in anchors throughout the rice genome (chromosomes 1–6). Graphical map obtained with Mapdisto v 1.7. **b** Distribution of SSR markers that form the Universal Core Genetic Map, arranged in anchors throughout the rice genome (chromosomes 7–12)

Application to the genetic mapping of three interspecific crosses

To validate the utility of the UCGM, we used it to derive genetic maps of three BC_1F_1 populations representing the first generation development of CSSL libraries. We built these maps to validate the SSR positions, keeping in mind that some of them may be duplicated in certain *Oryza* wild species and could potentially tag unexpected and unknown chromosomal fragments. This would result in misleading interpretations of chromosomal position of introgressions and would ultimately lead to confusion when analyzing the phenotypic performance of a number of CSSLs. The three crosses we used were *O. sativa* ssp. *indica* cv. IR64 × *O. glaberrima* acc. TOG5681, *O. sativa* ssp. tropical *japonica* cv. Curinga × *O. rufipogon* acc. IRGC105491, and *O. sativa* BRSMG Curinga × *O. meridionalis* acc. OR44 (or W2112). Backcrosses were made to the *O. sativa* parent in the three populations. Population sizes ranged from 88 to 125 individuals. We used the Paddy Map database to extract one polymorphic SSR per anchor for each cross. The populations were then genotyped for the

Fig. 1 continued

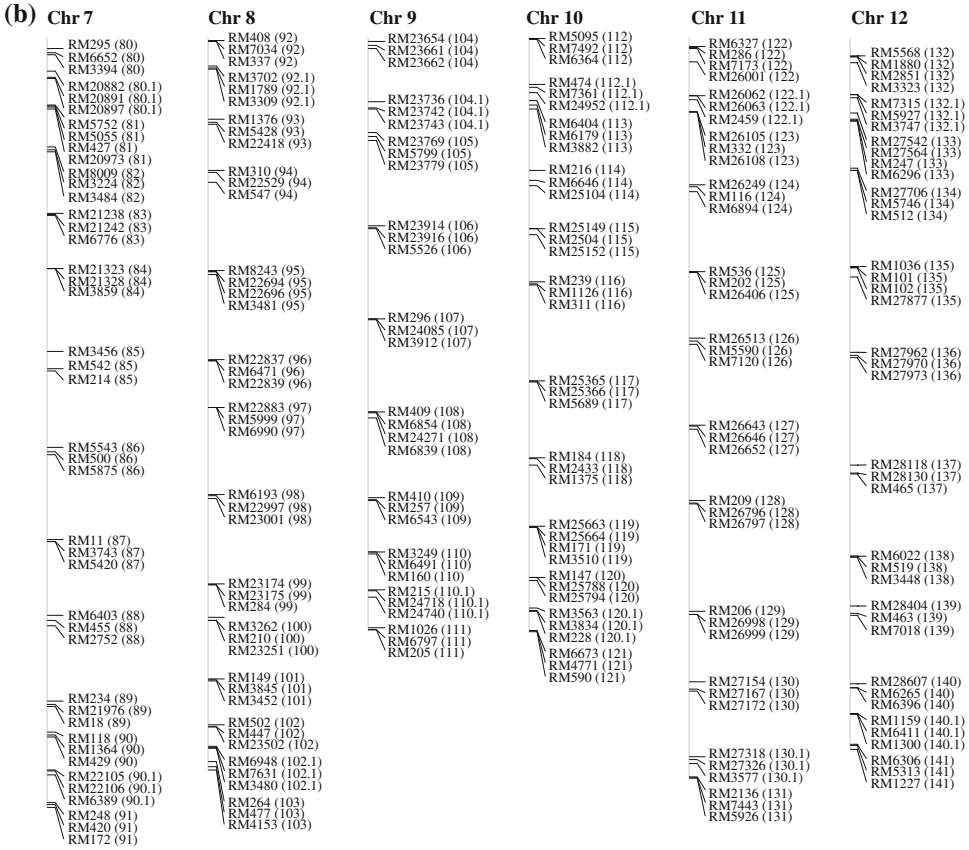

Table 3 Percentages of successful amplification in the cultivated and wild species, calculated on the 489 SSRs used in this study

Accession groups	Percentage of positive PCR (%)
Cultivated species	
O. sativa ssp. *indica*	98.2
O. sativa ssp. *japonica*	98.5
O. glaberrima	95.7
O. sativa + *O. glaberrima*	97.5
Wild species	
O. barthii	95.7
O. rufipogon	94.6
O. glumaepatula	87.7
O. meridionalis	78.9
All wild	89.2

selected markers, and the genetic maps were built using the MapDisto v. 1.7 software (Lorieux 2007) (http://mapdisto.free.fr/). The obtained genetic maps were compared with the expected map using the "Compare Maps" tool implemented in MapDisto. All maps showed a strong colinearity with the expected map, except in the centromeric regions, where a few inversions were observed due to the poor precision of the estimation of the map distances in these low recombining regions. The total map sizes were well correlated with the predicted sizes (Table 4).

Discussion

On the concept of UCGM

The strategy used to develop the UCGM and the particular marker sets utilized in this study make it possible for different laboratories to readily generate genetic maps on different populations using a common set of markers. This avoids many of the problems encountered by collaborative projects where different sets of markers are used and comparisons among subsequently generated maps can be difficult. The concept behind the UCGM is that by arranging a series of SSR markers into *anchors* that are represented by triplets of closely linked SSRs, the probability of at least one SSR being polymorphic within a given anchor is much higher than that of a single SSR being polymorphic. By comparing anchor positions instead of marker positions, the UCGM enables a direct comparison of any objects (genes, QTLs, introgressed fragments, etc.) that are positioned on any map generated from the set of anchors.

Table 4 Comparisons of physical and genetic maps sizes in three BC1F1 interspecific populations

	O. sativa ssp. *indica* acc. IR64 × *O. glaberrima* acc. TOG5681	*O. sativa* ssp. *japonica* acc. curinga × *O. rufipogon* acc. IRGC105491	*O. sativa* ssp. *japonica* acc. curinga × *O. meridionalis* acc. OR44
Genetic map	1,838	1,884	1,886
Physical map	1,932	1,839	1,843

Genetic distances were computed using Mapdisto v 1.7. Physical distances were converted to centimorgans (cM) using the conversion rate of 200 kbp/cM

Another common issue in the construction of genetic maps is the choice of sets of "core" markers that are evenly distributed along a reference genetic map. Due to the uneven distribution of recombination events along the chromosomes, the use of genetic map as the reference results in poor distribution of the markers along the physical genome, leading to a lower number of markers per gene in genomic regions with a low recombination rate. To address this issue, we chose the positions of the UCGM anchors on the basis of the physical map of rice.

Case of interspecific crosses

The polymorphism observed in 16 accessions representing 6 AA-genome *Oryza* species showed high to very high levels of allelic variation between the cultivated Asian rice, *O. sativa*, and its relatives *O. glaberrima*, *O. meridionalis*, *O. rufipogon*, *O. barthii*, and *O. glumaepatula*. This allowed us to successfully apply the concept of a universal map, which is based on the polymorphism probability of a set of anchors (Table 1), to the case of three interspecific crosses. The three genetic maps that we generated were validated for both the marker orders/positions and the total map sizes. We therefore continued the development of CSSL libraries with the same anchor set applied to the BC2 and BC3 generations. This way, it will be possible to locate all QTLs or major genes identified using these libraries on the virtually same genetic map.

Case of intrasubspecific crosses

We see from Table 1 that, for a value of 0.5 of the observed probability of an SSR being polymorphic, P_1, which is close to the commonly observed value in the case of intrasubspecific *indica* × *japonica* crosses, it is necessary to increase the number of SSRs per anchor to about $n = 10$, to obtain a fair probability P_3 of all anchors being polymorphic. This objective is feasible, due to the availability of increasing amounts of sequence information on diverse genomes. Moreover, the present study revealed a significant polymorphism level in *indica* × *japonica* crosses (70.4%). We see from Table 1 that a new set of anchors made of five SSR loci, an easily reachable goal, should produce a very useable framework, at least for the accessions we tested.

Conclusion

Usefulness of the UCGM

AA-genome *Oryza* species represent an invaluable genetic reservoir. The development of introgression line libraries is a powerful means of taking advantage of this genetic diversity. The UCGM represents an efficient way to build directly comparable genetic maps and graphical genotyping representations, which will eventually allow us to project all genes and QTLs detected using these populations in the same framework of physically located anchors. We are currently applying this strategy to the development and analysis of six CSSL libraries.

The UCGM is also valuable for marker-assisted backcrossing (MAB) in a breeding program. In a separate study, we are using 108 of the 165 anchors to detect and introgress three QTLs for tolerance to the Rice Hoja Blanca Virus (RHBV) and its insect vector, *Tagosodes orizicolus*, in two intrasubspecific populations (*O. sativa* spp. *indica* acc. Fedearroz 50 and acc. Fedearroz 2000 × *O. sativa* spp. *indica* acc. WC366).

Widespread use of the UCGM for genotyping enables us to integrate all of the gene discovery outcomes in a centralized database (Laboratory Information Management System), facilitating the breeders' work in choosing appropriate donors in their selection schemes. This integration will also help us in drastically reducing the time spent in polymorphism screening and it will allow us to make our gene discovery and breeding programs more cost-efficient, as the same set of primers can then be shared by the various rice genetics and breeding projects.

Applicability to other crops

Now that sequencing a whole genome has become a relatively economical and easy task, developing UCGM for other crops can be envisioned.

Acknowledgments We are grateful to the Generation Challenge Programme for their financial support through the project No G3005.10 "Exploring natural genetic variation: Developing genomic resources and introgression lines for four AA-genome rice relatives".

References

Blair MW, Hedetale V, McCouch SR (2002) Fluorescent-labeled microsatellite panels useful for detecting allelic diversity in cultivated rice (*Oryza sativa* L.). Theor Appl Genet 105:449–457

Brondani RPV, Zucchi MI, Brondani C, Rangel PHN, Borba T, Rangel PN, Magalhaes MR, Vencovsky R (2005) Genetic structure of wild rice *Oryza glumaepatula* populations in three Brazilian biomes using microsatellite markers. Genetica 125:115–123

Doi K, Iwata N, Yoshimura A (1997) The construction of chromosome substitution lines of African rice (*Oryza glaberrima* Steud.) in the background of japonica rice (*Oryza sativa* L.). Rice Genet News 14:39–41

Ebitani T, Takeuchi Y, Nonoue Y, Yamamoto T, Takeuchi K, Yano M (2005) Construction and evaluation of chromosome segment substitution lines carrying overlapping chromosome segments of indica rice cultivar "Kasalath" in a genetic background of japonica elite cultivar "Koshihikari". Breed Sci 48:395–399

Eshed Y, Zamir D (1995) An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. Genetics 141:1147–1162

Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S (2005) Genetic structure and diversity in *Oryza sativa* L. Genetics 169:1631–1638

Ghesquière A, Séquier J, Second G, Lorieux M (1997) First steps towards a rational use of African rice, *Oryza glaberrima*, in rice breeding through a "contig line" concept. Euphytica 96:31–39

Glaszmann JC (1987) Isozymes and classification of Asian rice varieties. Theor Appl Genet 74:21–30

Howell PM, Marshall DF, Lydiate DJ (1996) Towards developing intervarietal substitution lines in *Brassica napus* using marker-assisted selection. Genome 39:348–358

Huang XQ, Wang LX, Xu MX, Roder MS (2003) Microsatellite mapping of the powdery mildew resistance gene Pm5e in common wheat (*Triticum aestivum* L.). Theor Appl Genet 106:858–865

IRGSP (2005) The map-based sequence of the rice genome. Nature 436:793–800

Jena K, Kochert G, Khush G (1992) RFLP analysis of rice (*Oryza sativa* L.) introgression lines. Theor Appl Genet 84:608–616

Jeuken MJ, Pelgrom K, Stam P, Lindhout P (2008) Efficient QTL detection for nonhost resistance in wild lettuce: backcross inbred lines versus F_2 population. Theor Appl Genet 116:845–857

Kubo T, Aida Y, Nakamura K, Tsunematsu H, Doi K, Yoshimura A (2002) Reciprocal chromosome segment substitution series derived from japonica and indica cross of rice (*Oryza sativa* L.). Breed Sci 52:319–325

Li ZK, Fu BY, Gao YM, Xu JL, Ali J, Lafitte HR, Jiang YZ, Rey JD, Vijayakumar CH, Maghirang R, Zheng TQ, Zhu LH (2005) Genome-wide introgression lines and their use in genetic and molecular dissection of complex phenotypes in rice (*Oryza sativa* L.). Plant Mol Biol 59:33–52

Lorieux M (2007) MapDisto: a free user-friendly program for computing genetic maps. Computer demonstration given at the Plant and Animal Genome XV conference, 13–17 Jan 2007, San Diego, CA. Available at http://mapdisto.free.fr/

Lorieux M, Ndjiondjop M-N, Ghesquière A (2000) A first interspecific *Oryza sativa* × *Oryza glaberrima* microsatellite-based genetic linkage map. Theor Appl Genet 100:593–601

Matus IA, Hayes PM (2002) Genetic diversity in three groups of barley germplasm assessed by simple sequence repeats. Genome 45:1095–1106

McCouch SR, Teytelman L, Xu Y, Lobos KB, Clare K, Walton M, Fu B, Maghirang R, Li Z, Xing Y, Zhang Q, Kono I, Yano M, Fjellstrom R, DeClerck G, Schneider D, Cartinhour S, Ware D, Stein L (2002) Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). DNA Res 9:257–279

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321–4325

Ni J, Colowit PM, Oster JJ, Xu K, Mackill DJ (2001) Molecular markers linked to stem rot resistance in rice. Theor Appl Genet 102:511–516

Olufowote JO, Xu YB, Chen XL, Park WD, Beachell HM, Dilday RH, Goto M, McCouch SR (1997) Comparative evaluation of within-cultivar variation of rice (*Oryza sativa* L) using microsatellite and RFLP markers. Genome 40:370–378

Sambrook J, Russell D (2001) Molecular cloning: a laboratory manual, 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor

Second G (1982) Origin of the genic diversity of cultivated rice (*Oryza spp.*): study of the polymorphism scored at 40 isozyme loci. Jpn J Genet 57:25–57

Sobrizal K, Sanchez P, Doi K, Angeles E, Khush GS, Yoshimura A (1999) Development of *Oryza glumaepatula* introgression lines in rice, *O. sativa* L. Rice Genet News 16:107–108

Song ZP, Xu X, Wang B, Chen JK, Lu BR (2003) Genetic diversity in the northernmost *Oryza rufipogon* populations estimated by SSR markers. Theor Appl Genet 107:1492–1499

Tan L, Liu F, Xue W, Wang G, Ye S, Zhu Z, Fu Y, Wang X, Sun C (2007) Development of *Oryza rufipogon* and *O. sativa* introgression lines and assessment for yield-related quantitative trait loci. J Integr Plant Biol 49:871–884

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res 11:1441–1452

Tian F, Li DJ, Fu Q, Zhu ZF, Fu YC, Wang XK, Sun CQ (2006) Construction of introgression lines carrying wild rice (*Oryza rufipogon* Griff.) segments in cultivated rice (*Oryza sativa* L.) background and characterization of introgressed segments associated with yield-related traits. Theor Appl Genet 112:570–580

Wan XY, Wan JM, Su CC, Wang CM, Shen WB, Li JM, Wang HL, Jiang L, Liu SJ, Chen LM, Yasui H, Yoshimura A (2004) QTL detection for eating quality of cooked rice in a population of chromosome segment substitution lines. Theor Appl Genet 110:71–79

Wu J, Maehara T, Shimokawa T, Yamamoto S, Harada C, Takazaki Y, Ono N, Mukai Y, Koike K, Yazaki J, Fujii F, Shomura A, Ando T, Kono I, Waki K, Yamamoto K, Yano M, Matsumoto T, Sasaki T (2002) A comprehensive rice transcript map containing 6591 expressed sequence tag sites. Plant Cell 14:525–535

Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Liu Z, Li L, Liu J, Qi Q, Liu J, Li L, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Zhang J, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Ren X, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Wang J, Zhao W, Li P, Chen W, Wang X, Zhang Y, Hu J, Wang J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Li G, Liu S, Tao M, Wang J, Zhu L, Yuan L, Yang H (2002) A draft

sequence of the rice genome (*Oryza sativa* L. ssp. indica). Science 296:79–92

Yu C, Wan J, Zhai H, Wang C, Jaing L, Xiao Y, Liu Y (2005) Study on heterosis of inter-subspecies between indica and japonica rice (*Oryza sativa* L.) using chromosome segment substitution lines. Chin Sci Bull 50:131–136

Zhang Q, Liu KD, Yang GP, Maroof MAS, Xu CG, Zhou ZQ (1997) Molecular marker diversity and hybrid sterility in indica–japonica rice crosses. Theor Appl Genet 95:112–118

Zhang X, Zhou S, Fu Y, Su Z, Wang X, Sun C (2006) Identification of a drought tolerant introgression line derived from Dongxiang common wild rice (*O. rufipogon* Griff.). Plant Mol Biol 62:247–259

Zhang ZH, Deng YJ, Tan J, Hu SN, Yu J, Xue QZ (2007) A genome-wide microsatellite polymorphism database for the indica and japonica rice. DNA Res 14:37–45