Molecular characterization of the European rice collection in view of association mapping

Brigitte Courtois[1]*, Raffaella Greco[2], Gianluca Bruschi[2], Julien Frouin[1], Nourollah Ahmadi[1], Gaëtan Droc[1], Chantal Hamelin[1], Manuel Ruiz[1], Jean-Charles Evrard[1], Dimitrios Katsantonis[3], Margarida Oliveira[4], Sonia Negrao[4], Stefano Cavigiolo[5], Elisabetta Lupotto[5] and Pietro Piffanelli[2]

[1]CIRAD, 34098 Montpellier, France, [2]Parco Technologico Padano, 26900 Lodi, Italy, [3]NAGREF, 57001 Thermi-Thessaloniki, Greece, [4]ITQB, 2784-505 Oeiras, Portugal and [5]CRA-RIS, 13100 Vercelli, Italy

Abstract

In South Europe, rice is grown as an irrigated crop in river deltas where it plays an important role in soil desalinization. Specific varieties are needed for these tough conditions. We analyzed the genetic structure of a set of 305 varieties coming from the European Rice Germplasm Collection (ERGC) with 90 single nucleotide polymorphisms and compared it with a reference set representative of the diversity of *Oryza sativa* (mini-Germplasm Bank (GB)). These accessions had been characterized for their grain type and growth cycle duration. The polymorphism information contents of the ERGC were lower than those of the mini-GB, indicating a narrower genetic basis. Indeed, almost all ERGC accessions belong to the japonica group. Within the japonica group, both a dendrogram and a Bayesian clustering identified two major clusters. The first cluster encompassed tropical japonicas and American varieties from USA and Argentina characterized by long and narrow grains and medium to long duration. On a finer level, tropical japonicas appear separated from the other accessions. The second cluster is composed of European varieties mostly early or medium in duration and Asian temperate accessions, with a subgrouping based on grain format. A set of 200 accessions was composed for association mapping studies on traits such as salt tolerance.

Keywords: association mapping; genetic diversity; *O. sativa*; single nucleotide polymorphisms; temperate japonica

Introduction

Grown in the South of Europe since the 15th century, rice occupies presently approximately 400,000 ha for a production of 2.7 Mt in 2008 (http://faostat.fao.org). The main producers are Italy (224,000 ha) and Spain (96,000 ha) and, to a lesser extent, Greece (31,000 ha), Portugal (25,000 ha) and France (16,000 ha) with an average yield of 6.5 t/ha. All the rice area is irrigated. Rice is grown mostly in river deltas, in areas ecologically fragile, where it plays an important environmental role by limiting salinity problems through field flooding. Rice is at the northern limit of its natural cultivation zone and suffers of short cropping seasons (May to September) and low temperatures at both extremities of the cycle.

To face these constraints, specific rice varieties are needed. The varieties grown in Europe are mainly but not exclusively temperate japonica types. In the framework of two European projects (RESGEN 1999–2002 and EURIGEN 2006–2010), public breeding institutions decided to pool their working collections and characterize the resulting European Rice Germplasm Collection

---

* Corresponding author. E-mail: brigitte.courtois@cirad.fr

(ERGC) of 455 accessions on phenotypic and molecular bases. Part of the entries originates from Asia or from other temperate areas. Most of the collection, however, is composed of cultivars created by European breeding programmes, which are original genetic combinations potentially useful for the Mediterranean and the overall temperate areas.

European breeding programmes target salinity tolerance, blast resistance, grain quality and adaptation to water-saving strategies. The understanding of the genetic basis of these traits has made large progress through the use of mapping populations for quantitative trait loci detection. The decision to undertake association studies that bring a better resolution without the constraint of having to develop mapping populations motivated the teams to define a common association panel from the ERGC and establish marker-assisted breeding strategies.

The objective of this study was to characterize a set of 305 accessions from the ERGC with single nucleotide polymorphism (SNP) markers and extract a representative panel for association studies.

Materials and methods

The 305 accessions included in the genotyped collection are listed in Supplementary Table S1 (available online only at http://journals.cambridge.org). Accessions were classified according to their growth cycle (early, medium or late) and grain type (round, medium, long and large called 'long A' or long and narrow called 'long B') based on data obtained through several field experiments and will soon be available on http://eurigen.cirad.fr. A set of 62 additional accessions representative of the four main rice varietal groups (indica, japonica, aus and basmati) extracted from a core collection known as mini-Germplasm Bank (GB) (Glaszmann *et al.*, 1995) was genotyped at the same time to be able to position the ERGC accessions in comparison with this reference set.

A set of 90 SNPs was extracted from the *Oryza* SNP database (www.oryzasnp.org) on the basis of (1) their polymorphism among the japonica accessions of the database; (2) a good coverage of the genome with a minimum distance of 1 Mb between SNPs (Supplementary Table S2, available online only at http://journals. cambridge.org). DNA was extracted from one plant/ accession. The SNPs were genotyped using an Illumina Veracode assay. Appropriate controls (duplicated accessions and artificial heterozygotes) were included in the plates. The allele calling, performed with Bead Studio, was manually controlled for each SNP.

A weighted neighbour-joining (NJ) tree was built based on a dissimilarity matrix using a Sokal and Michener index with DARwin software (http://darwin.cirad.fr).

The number of sub-populations, K, was assessed using the software Structure. We run 50,000 burn-ins, 500,000 iterations and 10 replications/K value, with K varying from 2 to 8. An accession was classified into a sub-population when its component from this population was above 75%.

Results

The genotyping was successful for the 90 markers with only 0.39% missing data. The heterozygosity rate was low (0.19%) as expected for accessions conducted as pure lines and represented by a unique plant. The average polymorphism information content (PIC) value of the ERGC was 0.11 against 0.19 for the mini-GB with 40 and 3% of the markers with a PIC below 0.05 respectively, indicating a narrower genetic basis for the ERGC.

The NJ tree assembling ERGC and mini-GB accessions shows the bipolar structure characteristic of *O. sativa* with indica (in red) and aus/boro (in yellow) accessions clearly separated from sadri/basmati (in green) and japonica (in blue) accessions (Supplementary Figs. S1 and S2, available online only at http://journals.cambridge.org). Most of the ERGC accessions (in black) clustered with the japonica accessions. The exceptions were limited to three accessions from Asia, such as Zhenshan 97 from China, which clustered with the indica group and one accession, Fragrance from Italy, derived from a basmati × japonica cross which clustered with the sadri/basmati group.

Among the japonica group, two subgroups could be distinguished. The first subgroup encompassed most of the tropical japonicas and accessions from America, which could be further subdivided into accessions from USA and Argentina. Almost all EGRC accessions belonging to this subgroup have long B grains and medium to late duration. The second subgroup of accessions clustered almost all accessions from Europe together with the temperate japonicas accessions from the reference set. No specific pattern linked to geographic origin could be detected within this subgroup. However, a trend to sub-grouping by grain type and, to a lesser extent, by duration, was noticed. The upper part of the cluster corresponds to varieties with round to medium grain. The intermediate part corresponds to a mixture of grain types. The lower part corresponds to accessions with mostly long A grains. Most varieties were early or medium in duration with a specific cluster for the very early Eastern Europe varieties.

The results of Structure confirmed almost perfectly the split between American and European accessions for $K = 2$ (Supplementary Table S1, available online only at http://journals.cambridge.org). For $K = 3$, the American

accessions stayed together while the subgroup with long and narrow grains separated from the European accessions. Higher values of *K* were unlikely.

Discussion

We have characterized 305 ERGC accessions and shown, as expected, that most of them were japonica, with a narrow genetic basis, since many of them were derived from the others by hybridization. A first subgroup was composed of American accessions and tropical japonicas which constituted three separate clusters at a finer level. The proximity between US accessions, except those from California, the Argentinian accessions and tropical japonica has already been demonstrated (Mackill, 1995; Giarrocco *et al.*, 2007), but the fact they represented three separate clusters show that modern varieties have specific allelic combinations, which is a new information. The European accessions which represented a distinct subgroup were organized mainly by grain type and growth cycle duration. Both traits have been the target of strong selection: the grain type, because it determines the marketability of the varieties with a lasting importance of round and medium grains, often needed for local preparations (e.g. risotto or paella), and the growth cycle duration, because of the need for adaption to European latitudes.

From these data, we composed an association panel of 200 accessions (see passport data) using a linkage disequilibrium minimization method available in DARwin based on allelic combinations. Because of the overlapping between the genetic and phenotypic structure, association mapping is not likely to be successful for traits such as duration and grain type but other important phenotypic traits such as salinity tolerance do not show the same organization (Ahmadi *et al.*, in preparation) and will be amenable to association mapping.

Acknowledgements

References

Giarrocco LE, Marassi MA and Salerno GL (2007) Assessment of the genetic diversity in Argentine rice cultivars with SSR markers. *Crop Science* 47: 853–860.

Glaszmann J-C, Mew T, Hibino H, Kim CK, Mew TI, Vera Cruz CH, Notteghem J-L and Bonman JM (1995) Molecular variation as a diverse source of disease resistance in cultivated rice. *Rice Genetics III*. Los Baños: IRRI, pp. 460–466.

Mackill DJ (1995) Classifying japonica rice cultivars with RAPD markers. *Crop Science* 35: 889–894.