Euphytica (2007) 154:317–339
DOI 10.1007/s10681-006-9210-8

Through the genetic bottleneck: *O. rufipogon* as a source of trait-enhancing alleles for *O. sativa*

Susan R. McCouch · Megan Sweeney · Jiming Li · Hui Jiang · Michael Thomson · Endang Septiningsih · Jeremy Edwards · Pilar Moncada · Jinhua Xiao · Amanda Garris · Tom Tai · Cesar Martinez · Joe Tohme · M. Sugiono · Anna McClung · Long Ping Yuan · Sang-Nag Ahn

Received: 23 February 2006 / Accepted: 9 June 2006 / Published online: 17 October 2006

Abstract This paper summarizes results from a decade of collaborative research using advanced backcross (AB) populations to a) identify quantitative trait loci (QTL) associated with improved performance in rice and to b) clone genes underlying key QTLs of interest. We demonstrate that AB-QTL analysis is capable of (1) successfully uncovering positive alleles in wild germ-

S. R. McCouch (✉) · M. Sweeney · H. Jiang · M. Thomson · E. Septiningsih · J. Edwards · P. Moncada · J. Xiao · A. Garris
Department of Plant Breeding & Genetics, Cornell University, 162 Emerson Hall, Ithaca, NY 14853, USA
e-mail: srm4@cornell.edu

J. Li
Pioneer Hybrid International, Johnston, Iowa 50131, USA

Present Address:
M. Thomson · E. Septiningsih
International Rice Research Institute, Los Baños, Laguna, Philippines

Present Address:
J. Edwards
University of Arizona, Tucson, AZ, USA

Present Address:
P. Moncada
Cenicafe, Apartado aereo 2427, Manizales, Caldas, Colombia

Present Address:
J. Xiao
Monsanto Corp, St. Louis, MO, USA

Present Address:
A. Garris
USDA/ARS, Geneva, NY, USA

Present Address:
T. Tai
University of California, Davis, CA, USA

C. Martinez · J. Tohme
CIAT, Cali, Colombia

M. Sugiono
ICABIOGRAD, 16111 Bogor, Indonesia

A. McClung
USDA-ARS, DBNRRC & Beaumont Rice Research Unit, Beaumont, TX 77713, USA

L. P. Yuan
China National Hybrid Rice Research & Development Center, Changsha, Hunan 41025, P.R. China

S.-N. Ahn
Dept of Agronomy, College of Agriculture, Chungnam National University, Daejeon, Republic of Korea

plasm that were not obvious based on the phenotype of the parent (2) offering an estimation of the breeding value of exotic germplasm, (3) generating near isogenic lines that can be used as the basis for gene isolation and also as parents for further crossing in a variety development program and (4) providing gene-based markers for targeted introgression of alleles using marker-assisted-selection (MAS). Knowledge gained from studies examining the population structure and evolutionary history of rice is helping to illuminate a long-term strategy for exploiting and simultaneously preserving the well-partitioned gene pools in rice.

Keywords Inter-specific cross · Transgressive variation · Quantitative trait loci (QTL) · Rice (*Oryza sativa* L.) · Marker assisted selection · Molecular breeding

Abbreviations

AB	Advanced backcross
QTL	Quantitative trait loci
MAS	Marker-assisted-selection
NIL	Near isogenic line

Introduction

Evolutionary history and population sub-structure of *O. sativa*

Rice, *Oryza sativa*, is a cultivated, inbreeding species. It is estimated to have been domesticated in Asia approximately 10,000 years ago from a complex ancestral species, *O. rufipogon*, that inhabits diverse environments throughout Asia (Oka 1988; Vaughan et al. 2005). Two major subspecies within *O. sativa* have been recognized since ancient times, namely *indica* and *japonica* (Matsuo et al. 1997). The time since divergence of the *indica* and *japonica* gene pools has been estimated as approximately 0.44 million years based on comparisons of genome sequence between Nipponbare (*japonica*) and cv 9311 (*indica*) (Ma and Bennetzen 2004). This time estimate, in conjunction with evidence from fossilized remains of rice grains found in tombs in China, India and Vietnam (Matsuo 1997) suggests that differentiation of the *indica* and *japonica* gene pools pre-dates the domestication of *O. sativa* by several hundred thousand years. This is consistent with the belief that rice was independently domesticated on at least two occasions from a pre-differentiated ancestral pool (Cheng et al. 2003; Garris et al. 2005; Roy 1921; Second 1991, 1982).

In a recent study by Garris et al. (2005), five sub-populations were clearly identified in a collection of 234 landrace varieties of *O. sativa* using 169 simple sequence repeat (SSR) markers and two chloroplast sequences. The basic number and identity of the sub-populations identified by Garris et al. (2005) using 169 SSRs were consistent with those detected by Glaszmann (1987) using only 15 isozyme markers. The same groups were also identified by intronic SNP's assayed at 112 random genetic loci (A. Caicedo, NCSU and S. Williamson, Cornell Univ, pers. comm.). Consistent with Glaszmann's earlier study (1987), nuclear and chloroplast SSR data supported a close evolutionary relationship between the *indica* and *aus* groups, and between the *tropical japonica*, *temperate japonica* and *aromatic* groups (Fig. 1). A more complete evaluation of the population sub-structure of *O. sativa* is currently underway as part of the Generation Challenge Program (McNally pers. comm., IRRI) using a diverse set of molecular markers and several thousand accessions. Preliminary results suggest that the same groups identified by Garris et al. (2005) and Glaszmann (1987) are emerging in this study, providing strong support for the underlying population sub-structure described to date.

Population structure of *O. rufipogon*

O. rufipogon is considered to be the progenitor of *O. sativa* and this is consistent with the fact that it is genetically more diverse than *O. sativa* (Sun et al. 2002). It is found dispersed throughout a wide range of habitats across most of tropical and sub-tropical Asia. *O. rufipogon*, like *O. sativa*, is comprised of genetically identifiable sub-populations that show strong geographical and ecological differentiation (Edwards 2005; Lu et al. 2002; Morishima et al. 1984; Oka and Morishima 1967).

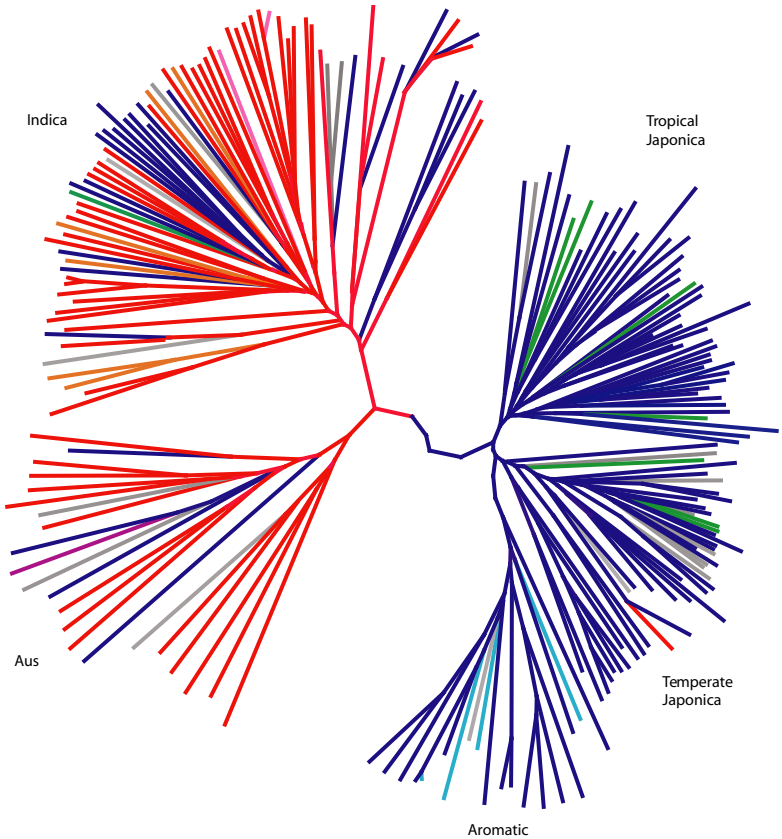

Fig. 1 Modified from Garris et al. (2005) showing unrooted Neighbor-joining (NJ) tree (Cavalli-Sforza 1967) developed using 169 nuclear SSRs and 234 landrace accessions of *O. sativa*; the intensity of the grey shading (color < in the online version of this figure >) of the branches of the tree corresponds to the observed chloroplast haplotype. Copyright permission granted by the Genetics Society of America

There is recent evidence that the ancestry of the *indica* and *japonica* subgroups of *O. sativa* can be traced to specific geographically-defined *O. rufipogon* populations (Cheng et al. 2003; Edwards 2005; Londo et al. 2006). However, admixture and introgression hybridization among sub-populations and between cultivated and wild populations can obscure the picture (Cheng et al. 2003; Edwards 2005). Further, movement of wild rice seed accompanying that of landrace or elite *O. sativa* varieties complicates our understanding of ancestor–descendent relationships (Lu et al. 2002).

Intra-specific breeding in rice

A majority of the world's rice is produced from inbred (pure line) varieties and historically inbred variety development has focused almost exclusively on crosses between members of the same sub-population (*indica–indica*) or between related sub-populations (i.e., *tropical japonica* × *temperate japonica*) (Lu et al. 2004; Ni et al. 2002; Sano 1993). This is largely due to the prevalence of sub-population incompatibilities that make it difficult to obtain a random array of fertile recombinants from *indica–japonica* crosses (Harushima et al. 2002; Oka 1988; Sano 1993). Breeding programs based on *indica–japonica* hybridization boast a few success stories, such as the *Tongil* rice in Korea (Choi 1978; Choi et al. 1974) but are inevitably plagued by genome-wide incompatibilities that lead to hybrid breakdown and a persistent loss of recombinants in successive generations (Oka 1988). The phenomenon is reminiscent of Dobzhansky–Muller incompatibilities (Dobzhansky 1936; Muller 1942) in which there is a gradual accumulation of mutations that

cause incompatible epistatic interactions when two species (or sub-species) are hybridized. The long-term consequences for rice breeders of restricting crossing and population development to within sub-population materials is that there is a limited pool of genetic variation available for identifying new and useful combinations of genes. Over time, this aggravates the genetic bottlenecks that were initially created during the process of domestication and leads to a cryptic form of genetic erosion that dramatically slows the rate of genetic gain.

In contrast to the programs devoted to inbred variety development, hybrid rice breeders have explored a much larger portion of the genetic variability of *O. sativa*. In China, hundreds of thousands of inter-sub-population crosses have been generated, aimed at identifying productive heterotic combinations (Li and Yuan 2000). While this has been very productive, hybrid rice breeding and seed delivery is a difficult, costly and technology-intensive enterprise, in large part due to the enclosed floret morphology and the naturally inbreeding habit of the species. It is difficult to achieve reliable and economically viable levels of out-crossing during F1 hybrid seed production and as a result, hybrid rice breeding as a commercial enterprise is still in its infancy. Nonetheless, experience in hybrid rice breeding programs around the world highlights the value of *indica* × *japonica* combinations as a way of maximizing levels of heterosis, with the caveat that the associated sterility barriers must be overcome (Li and Yuan 2000; Zheng et al. 1994).

Inter-specific breeding in rice

Inter-specific breeding offers a way of expanding the gene pool of cultivated rice and enlarging the range of genetic variation available for plant improvement. Wide-hybridization has been used for many years to introduce qualitative characters from wild species into elite breeding material, such as disease and insect resistance and male sterility (Brar and Khush 1997; Dalmacio et al. 1995). More recently, quantitative traits have been targeted in inter-specific crossing programs (Tanksley and McCouch 1997). Some inter-specific crosses may present challenges that are reminiscent of the incompatibilities discussed for intra-specific crosses; when useful genes are transferred from genetically distant (non AA-genome) species, the process is often encumbered by serious sterility barriers and anomalous patterns of recombination (Ishii et al. 1994; Jena et al. 1992; Wang et al. 2005). However, crosses within the AA genome species are relatively easy to make and avoid the major problems of sterility encountered in more divergent crosses. In cases where linkage drag is a problem, the use of molecular markers to identify selected recombinant individuals that retain only a very small piece of wild chromosome can greatly facilitate the transfer of desirable genes.

A major challenge is the difficulty in identifying genes from wild or unadapted materials that are likely to enhance the performance of elite cultivars without disrupting favorable gene complexes. Favorable allele combinations have been identified and painstakingly accumulated in elite varieties over the last 50–100 years; they provide local adaptation and help satisfy essential grain quality and other production requirements. It is understandable that breeders are reluctant to lose the immediate yield and quality advantages of elite germplasm in efforts to identify rare new genes that might help expand the genetic base or break an existing ''yield barrier'' in the future. Furthermore, because wild or exotic germplasm is likely to contain many genes that reduce yield and quality and only a few genes that could improve these traits, investment in inter-specific crosses has generally been considered a poor risk by commercial plant breeders. Thus, the substantial ''pre-breeding'' effort required to identify and transfer favorable alleles from wild or unadapted sources into elite rice varieties has been left to a few national and international programs focusing on genetics, genomics and germplasm enhancement.

Exploring the genetic potential of wild and exotic germplasm for rice improvement

Despite the challenges, plant breeders have long recognized the value of inter-specific backcross breeding strategies to harness useful genetic variation (Frey et al. 1975; Harlan 1976, 1975; Hawkes 1958; Peloquin 1983; Rick 1983; 1967). These early plant breeders did not have the advantage of the high quality sequence informa-

tion, genetic maps and molecular markers that are now available for many crop species. This made it difficult to identify and select for favorable alleles contributing to a quantitative trait. Often, the genetic potential of an individual was obscured by the presence of a few deleterious alleles that confounded the breeder's ability to accurately determine its breeding value. Today, we can identify regions of the genome that are associated with specific components of a phenotype and determine which parent contributes the favorable allele at a particular locus. This critical information is required to selectively introduce useful components of quantitative trait variation from a wild or exotic gene pool into an elite cultivar. In doing so, rice breeders are able to easily access novel forms genetic variation for use in variety development and simultaneously help expand the gene pool of *O. sativa*.

Over the course of this work, we aimed to (a) identify quantitative trait loci (QTLs) derived from *O. rufipogon* that contributed positively to the performance of *O. sativa* in both inbred and hybrid combinations, (b) understand the molecular basis of the *O. rufipogon*-derived QTLs associated with domestication and transgressive variation, (c) utilize the ''wild QTLs'' in rice improvement and (d) use information about genetic diversity and population sub-structure in both *O. sativa* and *O. rufipogon* to make predictions about which parental combinations are likely to give the maximum genetic gain in a plant breeding program. The observations summarized in this paper were obtained from a decade of research exploring the potential of using low-performing wild and exotic germplasm as donors to enhance the performance of elite cultivars of *O. sativa*. Details of each study can be found in (Cho et al. 2003; Moncada 2001; Septiningsih et al. 2003a; Thomson et al. 2003; Xiao et al. 1998).

Materials and methods

Population development

An inter-specific crossing program was initiated in the early 1990's with collaborating institutions in China, Korea, Indonesia, Colombia, Ivory Coast and the US (McCouch et al. 2002). One or two elite cultivars from each collaborating national or international program were selected for hybridization with a wild/weedy accession of *O. rufipogon* from Malaysia (IRGC 105491) (Table 1). The inter-specific hybrid plants were crossed back to the elite, cultivated varieties for two generations to produce a series of 200–300 BC_2F_2 families or, in the case of hybrids, a series of BC_2F_1 testcross families. A flow diagram outlining the basic strategy is presented in Fig. 2. The approach is a modification of the ''advanced backcross (AB) QTL strategy as first described by (Tanksley and Nelson 1996). All of the crossing and population development was accomplished by collaborating scientists in the country where each of the *O. sativa* varieties was grown commercially, and all of the genotyping and QTL analysis was performed at Cornell University by students or visiting scientists from the participating institutions.

Plant material used for fine-mapping and QTL cloning

Two BC_2F_2 populations were used as the basis for fine-mapping and cloning genes underlying QTLs for plant stature, flowering time, grain weight and red-pericarp. One was a BC_2F_2 population consisting of 258 families derived from a cross between an elite U.S. tropical *japonica* cultivar, Jefferson, and the *O. rufipogon* accession, IRGC105491, from Malaysia (Thomson et al. 2003). This population provided material used to fine-map flowering time (Thomson et al. 2006, seed size (Li et al. 2004a) and red-pericarp (Sweeney et al. 2006). The other was a BC_2F_2 population derived from a cross between the tropical *indica* variety, IR64, and the same *O. rufipogon* accession described above. This population was used to clone a gene underlying plant stature (Septiningsih 2003a) and to confirm the cloning of the red-pericarp gene, *Rc* (Sweeney et al. 2006).

Jefferson is a semi-dwarf, early flowering, high-yielding cultivar that was developed for production in the Southern U.S. It is resistant to multiple diseases and has good grain quality (McClung

Table 1 Donor and recurrent parent (RP) lines used in AB QTL studies

Parent	References	Participating Institution	Origin	Notes
Donor				
O. rufipogon		IRRI[1]	Malaysia	(IRGC 105491) Ancestor of *O. sativa*
RP				
Ce64	Xiao et al. 1998	CNHRRDC[2]	China	*Indica* (hybrid) (irrigated)
Caiapo	Moncada et al. 2001	CIAT[3], Embrapa[4]	Brazil	*Tropical japonica* (upland)
Hwacheong	Cho et al. 2004	RDA[5], Chungnam Nat'l Univ.	Korea	*Temperate japonica* (irrigated)
Jefferson	Thomson et al. 2003	USDA-ARS[6]	USA	*Tropical japonica* (irrigated)
IR64	Septiningsih et al. 2003a, b	ICABIOGRAD[7]	IRRI	*Indica* (irrigated)

[1] Internationation Rice Research Institute

[2] China National Hybrid Rice Research and Development Center

[3] Centro Internacional de Agricultura Tropical

[4] Empresa Brasileira de Pesquisa Agropecuária

[5] Rural Development Administration

[6] United States Department of Agriculture-Agricultural Research Service

[7] Indonesian Center for Agricultural Biotechnology and Genetic Resources Research and Development

et al. 1997). The *O. rufipogon* donor, as described above, has no agronomic traits of interest but crosses well to both *indica* and *japonica* varieties. From this population, one BC_2F_2 family, C126-3, was selected as the starting material for fine mapping and near isogenic line (NIL) development of the grain weight QTL, *gw3.1* (Li et al. 2004a), five BC_2 families were selected for fine-mapping of the complex flowering time QTL, *dth1.1* (Thomson et al. 2006) and 18 BC_2F_2 families were selected for cloning of the red grain QTL, *rg7.1* (Sweeney et al. 2006).

Selection of families as the basis for fine-mapping in all cases was based on (a) confirmation that each contained an *O. rufipogon* introgression in the target region and also showed a phenotype that was distinguishable from BC_2F_2 individuals that did not carry the introgression, and (b) the fact that these families contained relatively few non-target background introgressions. NILs were

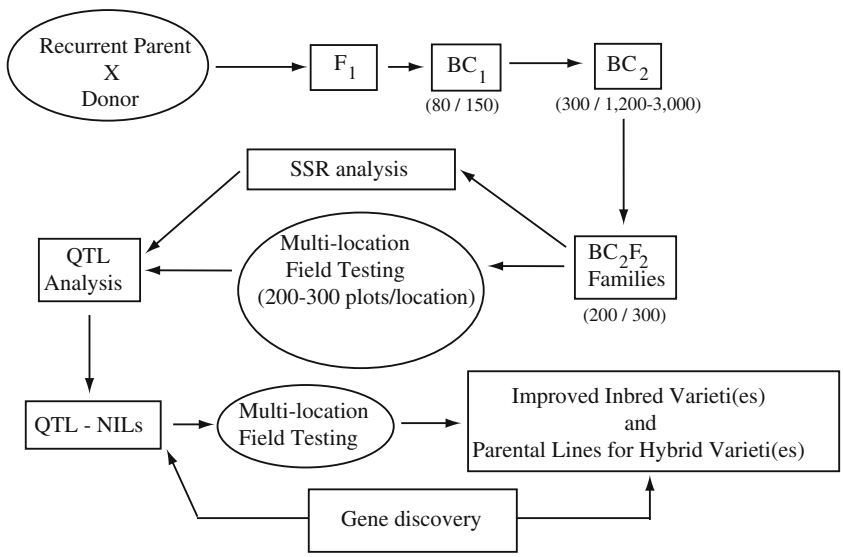

Fig. 2 Flow diagram outlining the basic strategy used to develop advanced backcross (AB) populations (BC_2F_2 families or BC_2F_1 testcross families) used for QTL analysis

developed by backcrossing to the Jefferson parent followed by selfing to eliminate non-target genomic regions as described in (Li et al. 2004a; Sweeney et al. 2006; Thomson et al. 2006).

IR64 is one of the most widely planted rice varieties in the world. It is grown throughout tropical Asia and is known for its long, fine grains, excellent eating quality, semi-dwarf stature, multiple disease resistance and good combining ability. From the 285 BC_2F_2 families used for QTL analysis (Septiningsih et al. 2003a, b), 15 families with overlapping *O. rufipogon* introgressions in the region defined by a plant height QTL, *ph1.1* were selected for fine-mapping. Selected BC_2F_2 individuals were backcrossed to produce BC_3F_2 families and a total of 1,300 plants from 25 segregating BC_3F_2 families were used for cloning the plant height QTL, *ph1.1* (Septiningsih, 2003b).

Phenotypic evaluation and QTL analysis

BC_2F_2 families were grown in replicated field trials in each participating program and evaluated using common protocols for yield and yield components, grain quality and other traits of importance to rice production (Table 2). While trait evaluation was standardized as much as possible, we consciously sought to maximize the diversity of both the genetic backgrounds and the abiotic environments in which our trials were conducted. In this regard, it is worth noting that the group in China developed BC_2F_1-test cross families, and was the only group evaluating the impact of *O. rufipogon* introgressions on hybrid performance. The groups in Colombia and in the US both used *tropical japonica* cultivars, cv Caiapo and cv Jefferson, respectively, but Caiapo is a Brazilian cultivar that was developed as an upland (non-irrigated) variety for use in low-input environments while Jefferson was developed for the irrigated, high-input conditions appropriate to the US rice industry. The group in Korea evaluated their BC_2F_2 families' populations in an irrigated temperate growing environment (using a temperate *japonica* RP), while the group in Indonesia evaluated families in an irrigated tropical growing environment (using an *indica* RP).

Remnant seed from each of the families was used for DNA extraction and the DNA was then subjected to analysis with a common set of molecular markers (McCouch et al. 2002; Temnykh et al. 2001). A molecular linkage map

Table 2 Descriptions of the 12 agronomic traits measured in wild QTL populations

Trait	Unit	Description
Days to heading	day	Days from sowing in field (greenhouse) to 10% panicles heading averaged from the whole plot.
Days to maturity	day	Days from sowing in field (greenhouse) to 80% grains reaching golden yellow averaged from the whole plot.
Plant height	centimeter	Averaged from 10 plants measured from ground to the tip of the tallest panicle (excluding awn).
Panicle length	centimeter	Averaged from all panicles of the 10 plants measured from panicle neck to panicle tip (excluding awn).
Panicles per plant	No. per plant	Averaged panicle number as counted from 10 plants (panicles having less than five seeds are not counted).
Spikelets per panicle	No. per panicle	(Number of spikelets from the 10 plants)/(number of panicles from the 10 plants).
Grains per panicle	No. per panicle	(Number of filled spikelets from the 10 plants)/(number of panicles from the 10 plants).
Seed set rate	percent (%)	(Number of grains per panicle)/(Number of spikelets per panicle).
Spikelets per plant	No. per plant	Averaged spikelet number as counted from the 10 plants.
Grains per plant	No. per plant	Averaged grain number as counted from the 10 plants.
1000-grain weight	gram	Averaged from three samples of 1000 fully filled grains.
Yield per plant	gram per plant	(Weight of bulked grains harvested from the 10 plants)/10.

was constructed for each population. The segregation data from the molecular markers was used in combination with the phenotypic scores for each family to identify quantitative trait loci (QTL) associated with the traits that had been evaluated in the field (Cho et al. 2003; Moncada 2001; Septiningsih et al. 2003a, b; Thomson et al. 2003; Xiao et al. 1998).

Quantitative trait loci were identified using single-point analysis (SPA), interval mapping (IM) and composite interval mapping (CIM) (Zeng 1994). Automatic cofactor selection using a forward/backward regression (forward $P < 0.01$, backward $P < 0.01$) was performed with QTL Cartographer v. 1.21 (Basten et al. http://www.stat.wisc.edu/biosci/qtlcart/qtlcart.html). To identify an accurate significance threshold for each trait, an empirical threshold was determined for IM using 10,000 permutations and for CIM using 1000 permutations for each trait for each trait across all 12 chromosomes (Churchill and Doerge 1994).

Results and discussion

Transgressive variation

QTL analysis in the diverse BC_2F_2 populations described above identified *O. rufipogon* introgressions (QTLs) that were associated with positive transgressive variation for days to heading (*dth*), yield (*yld*) and yield components, including seed size or grain weight (*gw*) and number of grains per plant (*gpp*) in each genetic background. Transgressive variants are offspring (segregants) whose performance exceeds that of either parent (in a positive or negative direction) and positive transgressives are those whose performance exceeds that of either parent in a positive direction. In this study, 30–50% of QTLs identified in each of the AB (BC_2F_2) populations showed improved performance for the target traits (Table 3) and 'wild QTLs' improved RP performance by 5–20% for most of the characters examined. QTLs associated with unwanted traits were also identified in all the RP backgrounds, including tall plant stature, red pericarp, dormancy and shattering, all of which are associated with wildness or weediness (Thomson et al. 2003). A summary of *O. rufipogon*-derived QTLs with both positive and negative effects in the different *O. sativa* backgrounds is available in the Gramene database (http://www.gramene.org).

The frequency of positive transgressive variation was higher than expected and we were interested to determine whether *O. rufipogon* introgressions in the same chromosomal regions contributed positively to performance in the diverse RPs. Because a single accession of *O. rufipogon* (IRGC 105491) had been used as a donor and standard phenotyping procedures were used in all experiments (Cho et al. 2003; Moncada 2001; Septiningsih et al. 2003a, b; Thomson et al. 2003; Xiao et al. 1998; Xiao et al. 1996), it was possible to address this question by aligning the QTL-introgressions reported for each study (using the CMAP and QTL modules in the Gramene database, http://www.gramene.org). In several cases, specific *O. rufipogon* introgressions were found to be associated with superior performance across several genetic backgrounds and environments (Fig. 3) (Lee et al. 2005; Moncada 2001; Septiningsih et al. 2003a, b; Thomson et al. 2003; Xiao et al. 1998). On the other hand, some yield and flowering time QTLs were associated with a positive effect in one genetic background and/or environment, but not in others (Fig. 3a, b). In other cases, *O. rufipogon*-derived QTLs with opposing effects (i.e., one contributing to large seed size and another contributing to small seed

Table 3 Percent of favorable QTLs from *O. rufipogon* detected in BC_2F_2 populations derived from crosses between diverse recurrent parents (RPs) and *O. rufipogon* (IRGC 105491)

RP	Sub-population	Country where evaluated	Percent favorable QTLs from *O. rufipogon*	Ref
Ce64	*Indica* (hybrid)	China	51%	Xiao et al. 1998
Caiapo	*Tropical japonica* (inbred)	Colombia	56%	Moncada et al. 2001
Jefferson	*Tropical japonica* (inbred)	USA	53%	Thomson et al. 2003
IR64	*Indica* (inbred)	Indonesia	33%	Septiningsih et al. 2003

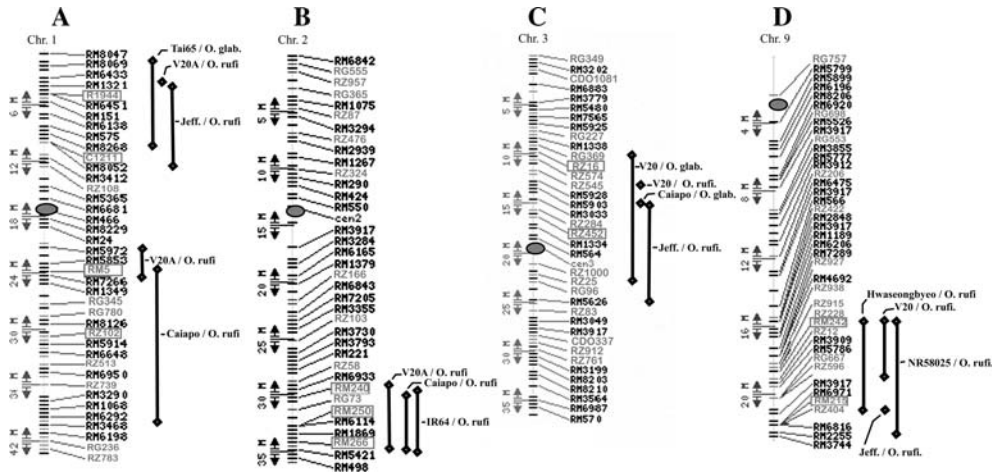

Fig. 3 Alignment of QTLs from different AB-QTL studies along the rice pseudomolecules. Numbers to left of chromosomes indicated Mb position on pseudomolecules; greyed ovals represent centromeres. (**A**) Chromosome 1 showing co-localization of QTLs for days to heading on the short arm, where the wild allele was transgressive for earliness in the Jefferson/*O. rufipogon* population evaluated in the US (Thomson et al. 2003) but conferred lateness in the V20/*O. rufipogon* population evaluated in China (Xiao et al. 1998) and in the Taichung 65/*O. rufipogon* population evaluated in Japan (Doi et al. 1998). On the long arm of chromosome 1, this figure also shows co-localization of two QTLs for yield where the wild allele was transgressive for yield enhancement in the V20A/*O. rufipogon* in China (Xiao et al. 1998) and in the Caiapo/*O. rufipogon* population in Colombia (Moncada et al. 2001); (**B**) Chromosome 2 showing a yield QTL where the effect of the wild *O. rufipogon* allele was positive and transgressive in the V20/*O. rufipogon* population evaluated in China (Xiao et al. 1998) and in the Caiapo/*O. rufipogon* population evaluated in Colombia, but was associated with decreased yield in the IR64/*O. rufipogon* population evaluated in Indonesia (Septiningsih et al. 2003); (**C**) Chromosome 3 near the centromere showing a seed size QTL where the wild allele was dominant for small seed size detected in four different studies: V20/*O. glaberrima* (Li et al. 2004); V20/*O. rufipogon* (Xiao et al. 1998); Caiapo/*O. glaberrima* (Aluko et al. 2004) and Jefferson/*O. rufipogon* (Thomson et al. 2003 and Li et al. 2004a); and (**D**) Chromosome 9 showing a yield and seed size QTL where the wild allele was dominant for yield and/or large seed size in the Hwaseongbyeo/*O. rufipogon* population evaluated in Korea (Cho et al. 2003), the V20/*O. rufipogon* population evaluated in China (Xiao et al. 1998), the IR58025/*O. rufipogon* population evaluated in India (Marri et al. 2005) and the Jefferson/*O. rufipogon* population evaluated in the US (Thomson et al. 2003)

size) were identified in the same elite genetic background, with similar allele effects evident in multiple RPs (Fig. 3c, d).

Choice of donor

This project clearly demonstrated that selected introgressions from one *O. rufipogon* accession could confer a consistent and agronomically relevant advantage in multiple genetic backgrounds grown in diverse environments. Furthermore, the *O. rufipogon* donor crossed readily with the diverse array of RPs (representing temperate *japonica*, *tropical japonica* and *indica* sub-populations) (Table 1) as summarized by (Xiao et al. 1998). What was not clear from this work was whether these results were specific to the particular accession of *O. rufipogon* that we happened to select at the beginning of the project, or whether other, randomly selected accessions of *O. rufipogon* or other species would produce similar results.

To address this question, we first looked at independent work undertaken by other researchers using different *O. rufipogon* donors. In parallel experiments using RPs from India, Korea, China and the Philippines and with similar crossing and population development schemes, all reported essentially identical results, with similar levels of transgressive variation and comparative percentages of ''favorable wild QTLs'' coming from other wild/weedy accessions of *O. rufipogon* (Lee et al. 2005; Marri et al. 2005; Nguyen et al. 2003; Xiong et al. 1999). These results confirm that AB-QTL analysis using diverse *O. sativa* × *O. rufipogon* crosses offers an effective strategy

for broadening the genetic base of *O. sativa* and enhancing the performance of a wide array of elite cultivars. We thus conclude that it was not simply a 'lucky' choice of donor in the first instance, but that almost any accession of *O. rufipogon* can be as a donor with similar results.

We did, however, note that the accession of *O. rufipogon* used in our work (IRGC 105491, from Malaysia) was genetically more similar to the *Aus* sub-population than to other *O. sativa* sub-populations (Edwards 2005). This may provide an explanation why a lower proportion of positive transgressive QTLs were identified in the cross between the *indica* cultivar, IR64, and this accession of *O. rufipogon* (Septiningsih et al. 2003a) than in crosses involving any of the *temperate* or *tropical japonica* RPs. It would be consistent with the studies by Glaszmann (1987) and Garris et al. (2005) that clearly demonstrate that the *aus* cultivar group is genetically more similar to *indica* than to *japonica*. This suggests that knowing something about the population substructure of *O. rufipogon* might offer predictive value in selecting wild/weedy parents for crossing with particular elite *O. sativa* varieties.

Genetic distance and breeding value

At what genetic distance would it be most advantageous to look for useful variation in rice? Should we focus on *O. sativa* landrace relatives, closely related species such as *O. rufipogon* or *O. nivara* or more distantly related species such as *O. glaberrima or O. barthii* from Africa, or *O. glumaepatula* from Brazil? Does the degree of genetic divergence provide any prediction about the likelihood of identifying positive transgressive variation in the offspring? Such questions are highly relevant to plant breeders who seek to identify the most productive materials to use as parents before investing in AB-QTL analysis.

To consider these questions, we studied the results of projects that involved crosses between elite *O. sativa* parents and the more diverse accessions of *O. glaberrima* or *O. glumaepatula* using the AB-QTL strategy (Aluko et al. 2004; Brondani et al. 2002; Li et al. 2004a). In these crosses, there were considerably higher rates of sterility and a lower proportion of detected QTL were associated with positive effects than in the crosses with *O. rufipogon*. This suggests that crosses between *O. sativa* and its immediate ancestor(s), *O. rufipogon* (and/or *O. nivara*) are more likely to generate useful variation for plant breeding than are more genetically divergent crosses, but further experiments are needed to confirm these early reports.

Finally, we were interested to determine whether a breeder would be more likely to find useful genetic variability in domesticated cultivars which have undergone human selection for favorable alleles (i.e., landraces of *O. sativa* or *O. glaberrima*) than in wild relatives that are the result of natural selection. Work initiated at IRRI has reported success using AB QTL analysis with intraspecific populations derived from crosses between elite cultivars and divergent landraces of *O. sativa* (Li et al. 2005). Numerous transgressive segregants with favorable phenotypes were identified in that study, but the experimental design did not allow us to compare the frequency of detecting favorable transgressive QTLs or the magnitude of the allele effects with our work using *O. rufipogon*.

It is counter-intuitive but empirically demonstrated that in several cases, fewer crossing barriers are encountered in *O. rufipogon* × *O. sativa* crosses than in crosses between the *indica* and *japonica* sub-populations of *O. sativa*. This is consistent with the idea that the genetic isolation and differentiation between these subgroups has been accentuated over the course of domestication. The value of using *O. rufipogon* as a donor is further underscored by the fact that it harbors a higher level of genetic variation than *O. sativa* (Sun et al. 2002). Thus, the use of interspecific crosses can be expected to increase the rate at which useful, novel allelic variation (i.e., variation not formerly present in the cultivated gene pool) is discovered and used to expand the genetic base of cultivated rice. On the other hand, it could be argued that *O. sativa* landraces might contain fewer deleterious ''wild'' alleles and thus might be expected to generate less linkage drag in crosses with elite *O. sativa* cultivars than would *O. rufipogon*. They also might be less disruptive of gene complexes that plant breeders seek to maintain. However, since elite varieties of *O. sativa* originate as selections from landrace

varieties (Dilday 1990; Lu et al. 2004) they tend to be genetically more similar than either is to *O. rufipogon*, making it theoretically less likely to identify transgressive segregants among the offspring. Opportunities to quantify genetic differences between sub-populations of both *O. sativa* and *O. rufipogon* and empirical observations about the relative value of each as a donor in an introgression breeding strategy await further research.

Identifying genes underlying QTLs

While the body of work summarized above demonstrates that useful levels of positive transgressive variation are observed among the offspring of divergent crosses and that QTLs associated with the transgressive segregants can be readily identified, the identity of most of the genes underlying these QTLs has not yet been determined and only a few of the *O. rufipogon*-derived QTLs have been incorporated into commercial varieties (Deng et al. 2004). We were interested in developing near isogenic lines (NILs) for several of these QTLs as the basis for releasing pre-breeding lines to interested plant breeders and also as the basis for cloning some of the genes underlying the QTL. To develop NILs, we undertook a backcrossing program that allowed us to fine-map and positionally clone some of the target genes.

Fine mapping provides an opportunity to identify key recombination events that break linkage drag, separating favorable from deleterious alleles along a chromosome and to deliver high quality NILs for applications in plant breeding. Cloning of genes underlying QTL provides essential information that is needed to effectively link QTL analysis with gene-based studies in other organisms. The cloning of target genes simultaneously provides the basis for developing gene-based, or 'perfect' markers for use in marker-assisted breeding and provides the basis for understanding how these genes interact in biochemical and regulatory pathways. Once the identity of genes underlying complex traits is known, reverse genetics strategies can be employed to more efficiently predict how certain genotypes are likely to perform and to select parents and offspring in an applied breeding program.

To date, we have isolated several genes underlying ''wild QTLs''. In the next section we will review three examples related to grain weight, plant stature and red pericarp.

Grain size

Seed and grain weight are important components of yield in the cereals, and in rice, grain length and shape are especially valuable because the physical appearance of the grain is important to consumers and it affects the cooking quality. Seed size or weight is also important in the evolution of cereal crops because humans have tended to select for large seed size during domestication while small seeds that are easily dispersed are important in the wild (Harlan 1992). Several independent studies in rice have identified QTL associated with grain weight or grain shape (length/width).

As reported by Thomson et al. (2003), *O. rufipogon* introgressions in the cv Jefferson (*tropical japonica*) background decreased grain weight (*gw*) at six QTL locations on chromosomes 2, 3, 9, 10, 12, whereas two QTL on chromosomes 1 and 5 increased grain weight. In a follow-up study, we aimed to refine the position of the, *gw3.1*, a QTL that had been mapped to the peri-centromeric region of rice chromosome 3, and to develop a set of NILs that would provide the foundation for map-based isolation of the gene underlying this QTL. We developed large, segregating populations to identify informative recombinants in the target region and subsequently developed NILs to characterize the magnitude and behavior of the *O. rufipogon*-derived allele in a domesticated *tropical japonica* background. As can be seen in Fig. 4, the range of phenotypes observed among BC_3 segregants shows transgressive variation for grain size, C126-3-6-4 contains an *O. rufipogon* introgression in the *gw3.1* target region while C133-1-1-5 does not. At *gw3.1*, the *O. rufipogon* allele was completely dominant over the Jefferson allele, and the dominant allele conferred small grain weight. The fact that multiple QTLs with opposing effects originally resided in the

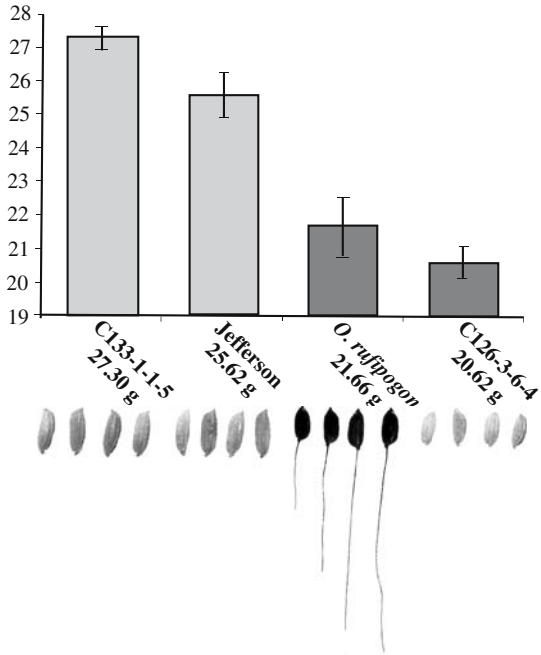

Fig. 4 Modified from Li et al. (2004a) showing transgressive variation for grain size and grain weight in two rice lines derived from the same BC_3 backcross generation. C133-1-1-5 is an individual with Jefferson DNA across the QTL region on chromosome 3 (illustrated in Fig. 3); C126-3-6-48 is an NIL with an *O. rufipogon* introgression across the grain weight QTL region on chromosome 3 where the wild allele is dominant for small seed size; *O. rufipogon* = wild donor parent (IRGC 105491); Jefferson = cultivated recurrent parent. Copyright permission granted by the Genetics Society of America

same parental line explains why ''mendelizing'' the trait using NILs can give rise to lines that show a more extreme phenotype than either parent. Additional details are described in Li et al. (2004a).

The *gw3.1* gene was recently cloned and found to encode a protein consisting of 232 amino acids with a putative phosphatidylethanolamine-binding protein (PEBP)-like domain, a trans-membrane region and a Von Willebrand Factor, type C (VWFC) module (Fan et al. 2006) that behaves as a dominant-negative regulator of grain size. A single nucleotide polymorphism (SNP) that causes a pre-mature stop in the gene gives rise to a non-functional allele which is associated with longer, heavier grains.

Now that the functional nucleotide polymorphism is known, it can be used to determine whether the same allele confers a similar phenotype in different genetic background. QTLs for grain size and grain shape have been reported in the same location on rice chromosome 3 in at least ten different inter and intra-specific populations (Brondani et al. 2002; Kubo T. et al. 2001; Li et al. 1997; Moncada 2001; Redona 1998; Thomson et al. 2003; Xiao et al. 1998; Xing et al. 2002; Xing et al. 2001; Yu et al. 1997). Yet the seed size and shape characteristics varied widely among the varieties used in these different studies. Thus, it can be concluded that allelic variation at multiple loci in the seed size pathway, in addition to *gw*3.1, determines the specific size and shape of rice grains. Nonetheless, the information at hand already provides plant breeders with a new tool for selecting seed size and seed shape in variety development programs. From our own observations of the NILs in our study, we infer that the *gw3.1* gene is likely to represent an early actor in the gene network controlling seed size in rice, this conclusion is based on the fact that the functional allele has the power to depress grain weight/seed size, even when combined with 5 QTL alleles from Jefferson at other loci that contribute to increased grain weight/seed size.

Plant stature

Using the IR64 × *O. rufipogon* population, a major plant height QTL, *ph1.1*, was identified on the long arm of chromosome 1 (Septiningsih et al. 2003a). The QTL co-localized with the *semi dwarf 1* (*sd1*) locus that was associated with increased harvest index during the Green Revolution (Aquino 1966; Jennings 1964). In our study, a total of 1300 plants from 25 segregating BC_3F_2 families were used in a positional cloning experiment in which the target region was narrowed from 47 cM to 7 cM in the first round of marker screening and then to 100 kb when additional markers were used to improve the mapping resolution (Septiningsih, 2003b). Once the QTL had been narrowed down to 100 kb, the region was searched for candidate genes and gene models in rice were BLASTED against all known genes associated with plant height in *Arabidopsis*, including the gibberellin biosynthesis and

gibberellin signal transduction genes. The only hit in the target region was weakly associated with the Arabidopsis gene *GA5* (56% amino acid similarity to the *gibberellin (GA) 20-oxidase* gene). A stronger hit to *GA5* was identified at a locus on rice chromosome 3 (72% amino acid similarity to *(GA) 20-oxidase*).

To test whether the *GA 20-oxidase* gene identified in the QTL region on chromosome 1 was the causative locus underlying *ph1.1*, additional markers were added to increase the map resolution. In the BC_3F_2 population of 1,300 individuals, 13 recombinants were identified in the 100 kb region and when single nucleotide polymorphism (SNP) markers were designed within and around the candidate *GA 20-oxidase* locus, two recombination breakpoints, defining a 2,157 bp interval, were identified within the *GA 20-oxidase* candidate. These recombinants confirmed that *GA 20-oxidase* was the gene underlying the *ph1.1* QTL and when the gene was sequenced, the IR64 allele differed from *O. rufipogon* by a 382 bp deletion that eliminated parts of the first and second exons (280 bp coding sequence and a 102 bp intron). The deletion causes a frame-shift, resulting in a novel termination codon (TGA) and effectively knocks out gene activity, resulting in the semi-dwarf phenotype (Septiningsih 2003a,b). The cloning of the *sd1* gene was simultaneously reported by multiple groups (Ashikari et al. 2002; Monna et al. 2002; Sasaki et al. 2002; Spielmeyer et al. 2002).

Despite the fact that the *sd1* locus has been reported to be a simple recessive mutation, results of our study are consistent with a previous report by (Foster and Rutger 1978) in which the locus was reported to be partially recessive. This suggests that either the phenotype is sensitive to the level of GA-20 oxidase in the system (dosage response) or that the recessive allele retains some level of activity, possibly promoting interaction(s) with other genes in the gibberellin biosynthesis pathway. In the BC_2F_2 segregating population between IR64 and *O. rufipogon*, plant height was normally distributed. If the phenotype were controlled by a single gene, we would expect to see a bi or tri-modal distribution with clearly differentiated peaks. Understanding how *sd1* interacts with other genes in the GA pathway(s) determining plant stature is critical for plant breeders interested in altering plant stature in a plant breeding program. Though the *ph1.1* QTL had a large effect on this trait (explaining up to 74% of the variance in one environment), four other loci in the genome were significantly associated with plant height in this population (Septiningsih et al. 2003a).

QTL studies such as those summarized here provide the opportunity to directly test for epistasis among interacting loci. In our study, we were interested in understanding the interactions and roles of the different GA biosynthesis and signal transduction genes involved in determining plant height in rice. In *Arabidopsis, GA1, GA2*, and *GA3* are known to be single-copy genes, and the loss-of-function alleles for these mutants cause severe dwarfism. On the other hand, GA5 (GA 20-oxidase) and GA4 (GA 3b-hydroxylase) are known to be encoded by at least four genes each in *Arabidopsis*; consequently, a knockout mutation at any one locus only causes semi-dwarfism, presumably due to the enzyme activity of other members of the gene family (Hedden and Phillips 2000). A test for digenic interactions for the plant height trait in the IR64 by *O. rufipogon* population identified nine loci that showed a significant two-way interaction with the *sd1* locus in a field environment in Bogor, Indonesia, and two other loci showed interaction in the Sukamandi field environment. Two other QTL studies have also identified potential digenic interactions between a plant height QTL located in the *sd1* region and other loci across the rice genome: one in a cross between IR64 and Azucena (Cao et al. 2001) and another in a cross between Zhenshan 97 and Minghui 93 (Yu et al. 2002). These networks of interactions can be further characterized by comparing the phenotypes of NILs or transgenics lines containing different alleles at each of these loci, as well as by crossing different NILs or transgenics in different combinatorial experiments and evaluating them in multiple environments to help unravel the genetic pathways affecting plant height in rice.

It was of interest to determine whether the underlying genetic cause of the semi-dwarf phenotype was associated with the same gene in rice and wheat, given that semi-dwarf varieties of both

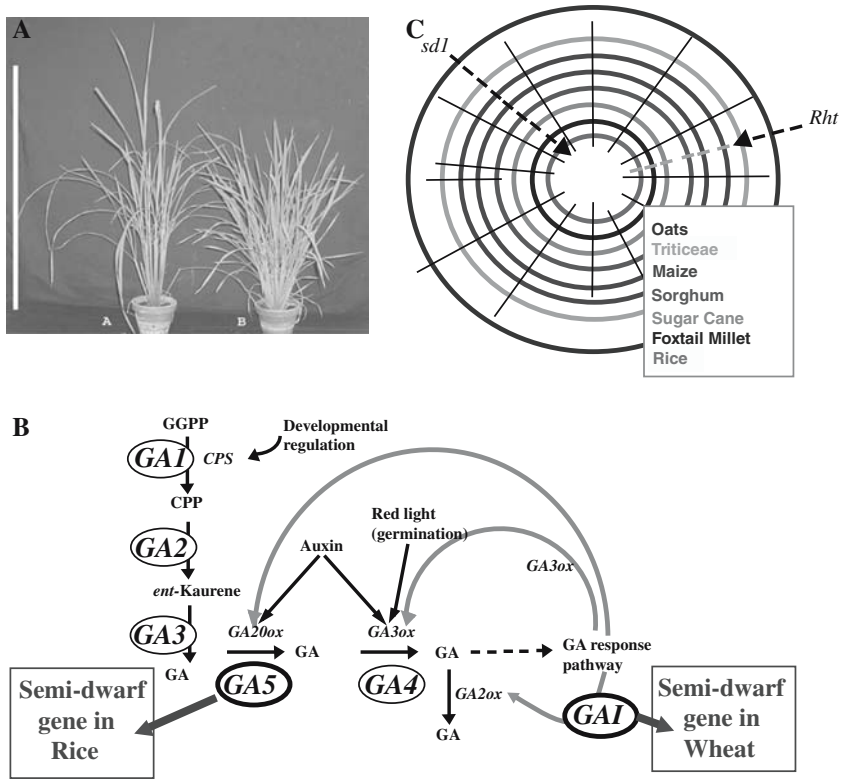

Fig. 5 (A) Photo of young *O. rufipogon* (Parent A) and the semi-dwarf IR64 (Parent B); (B) Pathway for *Arabidopsis* plant stature adopted from (Hedden and Phillips 2000) showing the GA20 oxidase gene (*sd1* = GA5) responsible for the Green Revolution semi-dwarf phenotype in rice and the GA regulatory protein (*Rht* = GAI) responsible for the semi-dwarf phenotype in wheat; (C) Concentric circles adapted from (Gale and Devos 1998) showing the comparative chromosome alignments and relative genome sizes of seven grass species, where rice is the smallest (innermost) circle and oats is the largest (outermost) circle (note: maize is depicted with two rings to illustrate the recent polyploidy of its genome). Positions of the two semi-dwarf genes on the comparative map show no relationship, but positioning them on the GA biosynthesis pathway does illustrate a meaningful functional relationship between them

cereals were associated with the higher yields of the Green Revolution (Hedden 2003; Khush 2001). Cloning of the *Rht* gene in wheat was reported by Peng et al. (1999). When *sd1* was aligned with *Rht*, there was no sequence similarity, demonstrating that breeders had not selected for mutations at the same locus in the two crops. However, when the genes were positioned on the biochemical pathway associated with plant height in Arabidopsis, it was obvious that they were both members of the same pathway (Fig. 5). The fact that different genes in the same pathway were associated with the semi-dwarf phenotype in rice and wheat suggests that comparative genome analysis may offer new candidates for allele mining. In cases where a gene is responsible for phenotypic variation in one species but no natural variation is known for that gene in another, scientists can use reverse genetics to search for novel alleles in exotic germplasm and, if they discover something useful, it can be harnessed immediately.

Red pericarp

Most modern rice varieties have white grain, but red pericarp is ubiquitous among the wild ancestors of cultivated rice (*Oryza rufipogon*). White grain is associated with domestication and is a target of strong selection in most rice breeding programs today. The red pigment in wild rice grains, proanthocyanidin (Oki et al. 2002), has been shown to have important deterrent effects

on pathogens and predators (Shirley 1998), making the pigmented pericarp a favorable attribute in the wild. Pigmented rices are also associated with nutritional and medicinal benefits (Ling et al. 2001), and are gaining in popularity as a specialty market. However, red rice is most familiar as a pernicious weed and represents the most economically important pest and grain quality problem in direct seeded rice systems throughout the US and Asia (Gealy et al. 2002). Weedy rices interbreed freely with cultivated, white-grained types and exhibit increased dormancy and shattering, making them difficult to eradicate from farmer's fields.

Two loci have been identified with red pericarp using classical genetic analysis, *Rc* (brown pericarp and seed coat) and *Rd* (red pericarp and seed coat). When present together, these loci produce red seed color (Kato et al. 1928). Both loci have been mapped using standard two-point analysis on the morphological map of rice; *Rc* on chromosome 7 and *Rd* on chromosome 1 (Kinoshita 1998). QTL mapping in several BC_2F_2 populations identified a single, significant QTL associated with red grain (*rg7.1*) near the centromere on chromosome 7. The LOD score associated with the *rg7.1* QTL peaks was 99 and 33, respectively, in two of the populations and the QTL was consistently detected in multiple environments (Lee et al. 2005; Septiningsih et al. 2003a; Thomson et al. 2003). QTL peaks corresponded to the previously mapped position of the mutant locus, brown pericarp, *Rc* (Kinoshita 1998). All of the BC_2F_1 plants in our population had red seeds, indicating that the *rg7.1* locus was dominant for red color, with the dominant allele donated by the *O. rufipogon* parent.

We have recently reported the cloning of a basic domain Helix-Loop-Helix (bHLH) gene underlying the QTL *rg7.1* for rice pericarp color (Sweeney et al. 2006). The QTL co-localizes with the mutant *Rc*. The functional allele conferring red pericarp was found in *O. rufipogon*, and a frame shift deletion before the bHLH domain in cv Jefferson resulted in a knockout of proanthocyanidin production, leading to white rice.

Red pericarp has long been used as marker for the cluster of domestication traits associated with weedy rice, including dormancy and shattering (Gu et al. 2005a; Gu et al. 2005b). Several studies have placed QTLs for dormancy and shattering in the pericentromeric region of chromosome 7 encompassing the *Rc* locus (Gu et al. 2005b). Now that *Rc* has been cloned, it is possible to ask if this association is the result of linkage or pleiotrophy. Given the reduced rate of recombination within the *rg7.1* QTL reported by Sweeney et al. (2006), it is logical to assume that genes for shattering, dormancy and pericarp color have hitchhiked together in a linkage block. It is also possible that *Rc* acts pleiotropically, as do many other bHLH proteins associated with pigmentation (Payne et al. 2000; Spelt et al. 2002). The NILs generated as part of this work will allow us to test these different hypotheses and, if the association among these weedy characters is the result of linkage, to provide recombinant lines to plant breeders interested in the possibility of breeding specialty rices with red pericarp but avoiding alleles that would increase dormancy and shattering.

The cloning of this gene in rice makes it possible to ask whether the same gene controls red pericarp in maize or wheat. Comparative mapping shows no homoeology between the position of the *Rc* gene in rice and the *R* gene controlling red pericarp in wheat. A reverse genetics approach also failed to locate any ESTs from wheat that show sequence similarity to the *Rc* locus in rice. Although the phenotypes are similar, molecular genetics experiments suggest that the mutation leading to white pericarp in the two genera probably occurred at different points in the pathway. Phylogenetic comparisons suggest that *Rc* is in a different clade than other bHLH proteins from rice that are known to regulate anthocyanin but that it is orthologous to *In1* in maize and to *tt8* in *Arabidopsis* (Sweeney et al. 2006). Despite the putative orthology inferred from sequence analysis, these proteins appear to have evolved different functions (Sweeney et al. 2006). *In1* is known to be a negative regulator of anthocyanin, with mutations leading to an intense purple color (Burr et al. 1996), while *Rc* is a positive regulator of proanthocyanidin synthesis, leading to no color when function is lost (Sweeney et al. 2006).

An immediate application of this work involves the use of 'functional markers' that specifically target the 14 bp functional nucleotide polymorphism (FNP) within the bHLH gene. FNP markers will facilitate screening for red rice contamination within certified seed lots, but more importantly, they can be used to help catch contaminants before plants set seed. Because the red pericarp is not visible until after harvest when grains are dehulled, it is difficult to eradicate in the field. The fact that the pericarp is maternal tissue, so that its color is dependent on the maternal genotype, means that seeds pollinated by red rice can be white (if the maternal plant carries the *rc* allele), but plants grown from these seeds will produce red seeds. Thus, the ability to detect carriers of the red allele long before the phenotype is visible will facilitate the utilization of wild relatives by allowing breeders to conclusively select against progeny carrying *Rc*.

Discussion

A decade of collaborative AB-QTL analysis in rice has demonstrated that this approach is productive in terms of both breeding applications and scientific discovery. Lines containing wild introgressions have been used as parents in the development of new hybrid varieties in China and are being tested in variety trials in Korea, Colombia and India. Many of the same lines have been used as the basis for isolating genes underlying QTLs, providing information about gene function, allelic variation, genetic systems and serving as the template for the development of functional markers.

Functional markers are a valuable tool for implementing marker-assisted breeding strategies. They increase the power and reliability of marker assisted selection by eliminating the possibility that recombination will separate a favorable allele from its marker and by ensuring that the target allele is, in fact, present in a breeder's lines. Functional markers are being used to select for several major genes in rice breeding programs, including the *Pi-ta* gene conferring blast resistance (Jia et al. 2000; 2004), the *waxy* gene for amylose content (Bergman et al. 2001; Sato and Nishio 2003; Yamanaka et al. 2004), the BAD2 gene for aroma (Bradbury et al. 2005) and the *xa5* gene for bacterial blight resistance (Iyer-Pascuzzi), and as genes underlying QTLs are cloned, allele-specific markers can also be used to select for favorable components of quantitative variation. The development of these markers depends on knowing the specific functional nucleotide polymophism(s) that distinguishes alleles at a particular locus. In many cases, such as for the *sd1* or *waxy* genes, multiple alleles are available in germplasm resources and breeders will want markers that allow them to distinguish which of several alleles are present in a given line. Thus, cloning genes underlying QTLs is not merely a scientifically motivated activity; it provides information that is increasingly important to plant breeders.

AB-QTL analysis has been used productively by researchers working on a variety of crop species. These include the pioneering work in tomato (Bernacchi et al. 1998), as well as reports in pepper (Rao et al. 2003), wheat (Huang et al. 2003) and barley (Pillen et al. 2003), among others. When AB-QTL analysis is used in maize or other out-crossing species, it does not identify the same frequency of positive transgressive QTLs that is seen in the inbreeding species (M. Smith, Cornell Univ, pers. comm.). Presumably, this is because breeders have already accessed most of the favorable alleles present in Teosinte relatives, faciliated by the outcrossing habit of maize, making it less likely that useful novel variation will be discovered through this approach. However, as a basic introgression breeding strategy, AB-QTL analysis is capable of identifying and selectively introgressing useful components of quantitative variation from any source, and thus is relevant as a molecular breeding strategy in many plant breeding programs.

AB populations consist of lines that are generally around 85–95% recurrent parent DNA, making it relatively efficient to select BC_2F_2 lines of interest, backcross them for one or two generations, and create NILs containing only one or two selected QTLs (QTL-NILs). QTL-NILs (or testcross hybrids produced from QTL-NILs) are useful to breeders as parents for further crossing, or as candidates for new varieties

containing one or more enhanced attributes. This approach is in contrast to two-step strategies that first identify a QTL in a population derived from a set of divergent mapping parents and then seek to introgress favorable QTL alleles into an entirely different parental background (Steele et al. 2005). The two-step approach requires many more generations of marker assisted backcross selection and often generates lines where the target QTL does not behave as expected in the new genetic background. This is due to the fact that QTLs interact epistatically with other genes in the genetic background and without a good understanding of the identity of the interacting alleles in critical pathways, it is difficult to predict what impact a QTL introgression will have. Thus, the fact that QTL discovery is accomplished in a genetic background that is of immediate use to plant breeders is part of what makes this approach so valuable. Further, information about pedigrees and population sub-structure can help a breeder determine the likelihood that a QTL introgression will have the desired phenotypic impact in a varietal background that is not the one in which the QTL was discovered.

Perspectives and outlook

Gene-based approaches to plant improvement

QTL studies, association analysis and the use of large mutant populations as the basis for gene identification are providing plant breeders and biologists with new opportunities to link genotype with phenotype and to understand the genetic causes of superior or inferior performance in a plant breeding program. Information about functional Nucleotide polymorphisms opens the door to exploring new types of genetic variation and using that variation to construct varieties that offer a range of novel traits. Understanding what genes and alleles underlie the traits most important to plant improvement shifts the paradigm that plant breeders use when searching for novel sources of useful genetic variation. It is increasingly feasible to use gene-based rather than phenotype-based approaches for selecting parents or superior offspring in a breeding program (Tanksley and McCouch 1997).

To increase the power of the gene-based approach, favorable alleles and allele combinations must be mapped onto the pathways that drive a plant's metabolism and responses to environment so that the consequences of variation at any one locus can be predicted in the context of the other genes in the pathway (European-Plant-Science-Organization-(EPSO) 2005). To predict the phenotypic consequences of allelic variation at a particular QTL, it is helpful to understand what regulatory or biosynthetic pathway is affected, how the regulation of the various genes in the pathway is coordinated and how the gene product(s) interacts with other genes or gene products (protein–protein or protein–DNA interactions) to confer a phenotype. Knowing what other genes are likely to be affected by an allele substitution at one locus offers a plant breeder powerful insight regarding the phenotypic consequences of a particular genetic change. All of the bits of information currently being generated by evolutionary studies, pathway analysis and systems biology are beginning to intersect with gene-based information derived from studies linking regions of the genome to components of quantitative trait variation for numerous species.

Linking bits of information using ontologies and controlled vocabularies

QTL analysis in agricultural crops is generally conducted in a field environment that is relevant to commercial production and therefore offers critical information about gene or trait expression in a context that needs no further translation. However, to make inferences about trait expression across environments requires that the environment as well as the phenotype be described and quantified in a way that everyone can understand. What is often missing from reports describing QTL experiments is the use of ontologies or standard controlled vocabulary terms to quantify variables related to the environment (i.e., day and night temperature, relative humidity and moisture regime, soil type and pH, nutrient availability, biotic stress factors, etc.) and the phenotype of the plant (description of how a trait

is assayed and, in particular, the anatomical part of the plant and developmental stage when the assay is conducted).

The development of ontologies and controlled vocabularies is a rapidly evolving science (Clark et al. 2005; Jaiswal et al. 2005; Pujar et al. 2006) requiring interdisciplinary collaboration among biologists, agriculturalists, computer scientists and software engineers. The objective is to develop vocabularies (sets of terms with defined relationships to each other) that reflect essential biological relationships in a clear and logical way and to use the vocabularies to annotate observations when entering data into a database. The requirement is that the computer must be able to ''read'' the terms used for annotation and then use those terms to flag bits of information from among the millions of data points stored in the database to identify meaningful relationships. The development and use of ontologies and appropriate controlled vocabularies is thus an essential component of modern biological inquiry that must go hand in hand with the development of genome databases.

Genome databases and plant improvement

From the perspective of the plant breeder or biologist, a genome database makes it possible to efficiently leverage information from an experiment conducted on one genome or in one environment to make inferences about traits, genes and pathways in another genome evaluated in a different environment. This helps narrow down the range of hypotheses that need to be tested. Genome databases offer sophisticated tools for data mining and data browsing that make it possible to find and bring together disparate pieces of information that have not necessarily been brought together before. This is only possible if the computer is able to accurately identify and catalogue the bits of information entered. The rapid accumulation of biological information today means that our understanding of the living world is changing quickly, and with it comes the requirement to constantly update and refine the ontologies that underlie our annotation efforts. While many database users will not choose to participate in this endeavor, it is imperative that they understand the importance of using appropriate terminologies and controlled vocabularies when reporting data. These vocabularies provide the essential infrastructure that allows the computer to link information about traits, environments, pathways, genes and alleles from diverse organisms and experiments and offers the breeder and the biologists insight about what genes do, how they interact, how they have evolved and how they are inherited in the context of phenotypes and plant performance (Bruskiewich et al. 2003; Jaiswal et al. 2006b).

Assembling the pipeline of information linking genotype and phenotype

Numerous genes and QTLs associated with traits of interest have been identified over the last 15 years in diverse species. In rice, the availability of genome sequence and the ability to clone genes underlying QTL makes it possible to identify genes that are critical to agricultural performance and map them to biochemical and regulatory pathways. Information about the specific genes and alleles that contribute favorably to the performance of a variety also make it possible for plant breeders and geneticists to implement a reverse genetics approach to the selection of parents for use in crosses. Understanding the natural sub-population structure of *O. sativa* offers a framework for investigating how specific alleles are distributed among the different gene pools of rice, and offers a guide to more efficient use of exotic germplasm in plant improvement. It also allows us to start to make predictions about which combinations of parents are most likely to give rise to positive transgressive variation and to begin to harness that variation in a breeding program.

Getting the pipeline to deliver new varieties

There is still a gap between our ability to generate information about which genes and alleles are associated with phenotypic variation of interest and a plant breeder's ability to utilize that information to develop a new variety. To help bridge this gap, we need an efficient system for targeted allele replacement that does not introduce addi-

tional copies of genes, does not require selectable markers and does not disrupt the genome beyond the minor change that is needed to alter an allele. With current approaches to genetic transformation, we are altering the genome in ways that are ultimately destabilizing. Allele modification based on site-specific base conversion (Iida and Terada 2005) looks promising when very small changes are needed to modify a target gene. This is essentially 'gene therapy' for plants. Homologous recombination in plants (Terada et al. 2002) is still inefficient and costly, but offers promise as a possible mechanism for allele replacement. New ways of enhancing and directing the natural process of meiotic recombination would be an alternative strategy that would facilitate the introgression of favorable alleles and would minimize the current requirement for screening large populations. Additional approaches that this author has not thought of are likely to emerge in response to the need to bridge the gap between our ability to generate information about genes and gene function, and the ability to generate productive new varieties that are an essential component of a sustainable food supply for the future.

This paper has focused on a molecular breeding strategy that is capable of generating lines with enhanced performance for use in breeding and simultaneously for identifying genes and alleles that underlie the variation of interest. While our work is focused on exploring and utilizing natural diversity found in *O. rufipogon* as a donor for rice improvement, the power of this approach is currently limited by our dependence on meiotic recombination as the mechanism by which novel alleles are introduced into our breeding material. New technologies that allow us to harness the process of recombination and direct it more efficiently, or that allow us to target allele modification directly would greatly enhance the power of this approach. As we learn more about the diversity of genes and alleles found in our germplasm resources, we will find ourselves increasingly limited in our ability to realize the potential of that information to drive variety development unless we find efficient ways to replace alleles. Enhancing the effective utilization of that variation in plant improvement is an essential objective of breeding programs around the world.

Acknowledgments This research was supported by grants from the Rockefeller Foundation; the U.S. Department of Agriculture (USDA) Plant Genome Research Program (National Research Initiative grant No. 96-35300-3645 and No. 00-35300-9216); CRIS Project 6225-21000-006; NSF Plant Genome awards DBI-0110004 and DBI-0319553, subaward 2003-1054-01; RiceTec, Inc.; a graduate assistantship to Michael Thomson provided by the Cornell Plant Cell and Molecular Biology Program (DOE/NSF/USDA Interagency Training Grant); a graduate assistantship to Amanda Garris provided by USDA/CSRS Competitive Grant 97-35300-5101 representing Food and Agricultural Sciences National Needs Graduate Fellowship in Plant Biotechnology; Ph.D. support for Pilar Moncada from the National Federation of Coffee Growers (Cenicafe) of Colombia; support for QTL analysis in Korea from the Crop Functional Genomics Center of the of the 21st Century Frontier Research Program (Project code: CG3112) and from Bio Green 21 of the Rural Development Administration, Republic of Korea. We are grateful to the International Center for Tropical Agriculture (CIAT) in Cali, Colombia for population development and phenotypic evaluation of the Caiapo population; to the China National Hybrid Rice Research and Development Center, Changsha, China for population development and phenotypic evaluation of the Ce64 hybrid families; to ICABIOGRAD in Indonesia for population development and phenotypic evaluation of the IR64 population; and to the USDA-ARS Beaumont Rice Research Unit, Beaumont, TX for phenotypic evaluation of the Jefferson population. We are grateful to Junjian Ni for the curation of QTLs in the Gramene database and for contributing to the development of Fig. 3; to Lois Swales for formatting the manuscript and creating figures from rough sketches; to all the many scientists and staff in the national breeding programs and in the McCouch lab without whose help and dedication we could not have developed the populations, collected the phenotypic data or done the fine-mapping that has been so critical to the success of these projects.

References

Aluko G, Martinez C, Tohme J, Castano C, Bergman C, Oard J (2004) QTL mapping of grain quality traits from the interspecific cross *Oryza sativa* × *O. glaberrima.* Theor Appl Genet 109:630–639

Aquino RC, Jennings PR (1966) Inheritance and significance of dwarfism in an Indica rice variety. Crop Sci 6:551–554

Ashikari M, Sasaki A, Ueguchi-Tanaka M, Itoh H, Nishimura A, Datta S, Ishiyama K, Saito T, Kobayashi M, Khush GS, Kitano H, Matsuoka M (2002) Loss-of-function of a rice gibberellin biosynthesis gene,H *GA20 oxidase (GA20 ox-2)*, let to the rice 'green revolution'. Breed Sci 52:143–150

Bergman CJ, Delgado JT, McClung AM, Fjellstrom R (2001) An improved method for using a microsatellite in the rice waxy gene to determine amylose class. Cereal Chem 78:257–260

Bernacchi D, Beck-Bunn T, Emmatty D, Inai S (1998) Advanced backcross QTL analysis of tomato. II. Evaluation of near-isogenic lines carrying single-donor introgressions for desireable wild QTL alleles derived from *Lycopersicon hirsutum* and *L. pimpinellifolium*. Theor Appl Genet 97:170–180

Brar DS, Khush GS (1997) Alien introgression in rice. Plant Mol Biol 35:35–47

Bradbury LM, Henry RJ, Jin QS, Reinke RF, Waters DL (2005) A perfect marker for fragrance genotyping in rice. Mol Breed 16:279–283

Brondani C, Rangel N, Brondani V, Ferreira E (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

Bruskiewich, RM, Cosico AB, Eusebio W, Portugal AM, Ramos LM, Reyes MT, Sallan MA, Ulat VJ, Wang X, Mcnally KL, Sackville Hamilton R, Mclaren CG (2003) Linking genotype to phenotype: the International Rice Information System (IRIS). Bioinformatics 19(Suppl 1):i63–65

Burr FA, Burr B, Scheffler BE, Blewitt M, Wienand U, Matz EC (1996) The maize repressor-like gene intensifier1 shares homology with the r1/b1 multigene family of transcription factors and exhibits missplicing. Plant Cell 8:1249–1259

Cao G, Zhu J, He CX, Gao YJ, Yan J, Wu P (2001) Impact of epistasis and QTL x environmental interaction on the developmental behavior of plant height in rice (*Oryza sativa* L.). Theor Appl Genet 103:153–160

Cheng CY, Motohashi R, Tsuchimoto S, Fukuta Y, Ohtsubo H, Ohtsubo E (2003) Polyphyletic origin of cultivated rice: based on the interspersion pattern of SINEs. Mol Biol Evol 20:67–75

Cho YC, Suh JP, Choi IS, Hong HC, Baek MK, Kang KH, Kim YG, Ahn SN, Choi HC, Hwang HG, Moon HP (2003) QTLs analysis of yield and its related traits in wild rice relative *Oryza rufipogon*. Treat of Crop Res In Korea 4

Choi HC (1978) Recent advances in rice breeding in Korea. Korean J Breed 10:201–238

Choi HO, Bae SH, Chung GS, Cho CY (1974) A New Short-Statured Rice Variety ''Tongil''. Res. Rep. of the O.R.D. in Korea 16:1–12

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 139:963–971

Clark JI, Brooksbank C, Lomax J (2005) It's all GO for plant scientists. Plant Physiol 138:1268–1279

Dalmacio RD, Brar DS, Ishii T, Sitch TA, Virmani SS, Khush GS (1995) Identification and transfer of a new cytoplasmic male sterility source from *Oryza perennis* into *indica* rice (*O. sativa*). Euphytica 82:221–225

Deng QY, Yuan LP, Liang FS, Li J, Wang LG, Wang B (2004) Studies on yield-enhancing genes from wild rice and their marker-assisted selection in hybrid rice. Hybrid Rice 19:6–10

Dilday RH (1990) Contribution of ancestral lines in the development of new cultivars of rice. Crop Sci 30:905–911

Dobzhansky T (1936) Studies on hybrid sterility. II. Localization of sterility factors in Drosophila pseudoobscura hybrids. Genetics 21:113–135

Edwards JD (2005) Origins and distribution of allelic diversity in populations of wild and cultivated rice and phenotypic consequences of admixture at a complex flowering time locus. Cornell

European-Plant-Science-Organization-(EPSO) (2005) European plant science: a field of opportunities. J Exp Bot 56:1699–1709

Fan C, Xing YZ, Mao HL, Lu TT, Han B, Xu C, Li X, Zhang Q (2006) *GS3*, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. Theor Appl Genet DOI 10.1007/s00122-006-0218-1

Foster KW, Rutger JN (1978) Inheritance of semi dwarfism in rice *Oryza sativa*. Genetics 88:559–574

Frey KJ, Hammond EG, Lawrence PK (1975) Inheritance of oil percentage in interspecific crosses of hexaploid oats. Crop Sci 15:94–95

Gale MD, Devos KM (1998) Comparative genetics in the grasses. Proc Natl Acad Sci 95:1971–1974

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

Gealy DR, Tai TH, Sneller CH (2002) Identification of red rice, rice, and hybrid populations using microsatellite markers. Weed Sci 50:333–339

Glaszmann JC (1987) Isozymes and Classification of Asian Rice Varieties. Theor Appl Genet 74:21–30

Gu XY, Kianian SF, Foley ME (2005a) Phenotypic selection for dormancy introduced a set of adaptive haplotypes from weedy into cultivated rice. Genetics 171:695–704

Gu XY, Kianian SF, Hareland GA, Hoffer BL, Foley ME (2005b) Genetic analysis of adaptive syndromes interrelatd with seed dormancy in weedy rice (*Oryza sativa*). Theor Appl Genet 110:1108–1118

Harlan J (1992) Crops & man. In: 2nd Harlan J (ed) American society of agronomy, Inc.; Crop Science Society of America, Inc.: Madison, Wisconsin, p 288

Harlan JR (1976) Genetic resources in wild relatives of crops. Crop Sci 16:329–333

Harlan JR (1975) Our vanishing genetic resources. Science 188:618–621

Harushima Y, Nakagahra M, Yano M, Sasaki T, Kurata N (2002) Diverse variation of reproductive barriers in three intraspecific rice crosses. Genetics 160:313–322

Hawkes JG (1958) Significance of wild species and primitive forms for potato breeding. Euphytica 7:257–270

Hedden P (2003) The genes of the green revolution. Trend Genet 19:5–9

Hedden P, Phillips AL (2000) Gibberellin metabolism: new insights revealed by the genes. Trends Plant Sci 5:523–530

Huang XQ, Coster H, Ganal MW, Roder MS (2003) Advanced backcross QTL analysis for the identification of quantitative trait loci alleles from wild relatives of wheat (*Triticum aestivum* L.). Theor Appl Genet 106:1379–1389

Iida S, Terada R (2005) Modification of endogenous natural genes by gene targeting in rice and other higher plants. Plant Mol Biol 59:205–219

Ishii T, Brar DS, Multani DS, Khush GS (1994) Molecular tagging of genes for brown planthopper resistance and earliness introgressed from *Oryza australiensis* into cultivated rice, *O. sativa*. Genome 37:217–221

Iyer-Pascuzzi A, McCouch S (2006) Functional markers for xa5 mediated resistance in rice (*Oryza sativa*, L.). Mol Breed (in press)

Jaiswal J, Avraham S, Ilic K, Kellogg E, McCouch S, Pujar A, Reiser L, Rhee S, Sachs M, Schaeffer M, Stein L, Stevens P, Vincent L, Ware D, Zapata F (2006a) Plant Ontology (PO): a controlled vocabulary of plant structures and growth stages. Comp Funct Genom 6:388–397

Jaiswal P, Ni J, Yap IV, Ware D, Spooner W, Youens-Clark K, Ren L, Liang C, Zhao W, Ratnapu K, Faga B, Canaran P, Fogleman M, Hebbard C, Avraham S, Schmidt S, Casstevens T, Buckler E, Stein L, McCouch S (2006b) Gramene: a bird's eye view of cereal genomics. Nucleic Acids Res 34:D717–D723

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

Jennings PR (1964) Plant type as a rice breeding objective. Crop Sci 4:13–15

Jia Y, McAdams SA, Bryan GT, Hershey HP, Valent B (2000) Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. EMBO J 19:4004–4014

Jia Y, Redus MA, Wang Z, Rutger J (2004) Development of SNLP marker from the Pi-ta blast resistance gene by Tri-Primer PCR. Euphytica 138:97–105

Kato, S, Kosaka H, Hara S (1928) On the affinity of rice varieties as shown by the fertility of rice plants. Central Agricultural Inst Kyushu Imperial Univ 2:241–276

Khush GS (2001) Green Revolution: the way forward. Nat Rev Genet 2:815–822

Kinoshita T (1998) Linkage mapping using mutant genes in rice. Rice Genet Newslett 15:13–74

Kubo T, Takano-Kai N, Yoshimura A (2001) RFLP mapping of genes for long kernel and awn on chromosome 3 in rice. Rice Genet Newslett 18:26–28

Lee SJ, Oh CS, Suh JP, McCouch SR, Ahn SN (2005) Identification of QTLs for domestication-related and agronomic traits in an *Oryza sativa* x *O. rufipogon* BC1F7 population. Plant Breed 124:209–219

Li J, Thomson MJ, McCouch SR (2004a) Fine mapping of a grain weight QTL in the peri-centromeric region of rice chromosome 3. Genetics 168:2187–2195

Li J, Xiao J, Grandillo S, Jiang L, Wan Y, Deng Q, Yuan L, McCouch S (2004b) 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

Li J, Yuan L (2000) Hybrid rice: genetics, breeding, and seed production. Plant Breed Rev 17:150–158

Li Z, Pinson SR, Park WD, Paterson AH, Stansel JW (1997) Epistasis for three grain yield components in rice (*Oryza sativa* L.). Genetics 145:453–465

Li ZK, Fu BY, Gao YM, Xu JL, Ali J, Lafitte JR, Jiang YZ, Rey JD, Vijayakumar CHM, 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

Ling WH, Cheng ZX, Ma J, Wang T (2001) Red and black rice decrease atherosclerotic plaque formaion and increase antioxidant status in rabbits. J Nutr 131:1421–1426

Londo JP, Chiang Y-C, Hung K-H, Chiang T-Y, Schaal BA (2006) Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. Proc Natl Acad Sci 103:9578–9583

Lu BR, Zheng KL, Qian HR, Zhuang JY (2002) Genetic differentiation of wild relatives of rice as assessed by RFLP analysis. Theor Appl Genet 106:101–106

Lu H, Redus MA, Coburn JR, Rutger JN, McCouch SR, Tai TH (2004) Population structure and breeding patterns of 145 US rice cultivars based on SSR marker analysis. Crop Sci 45:66–76

Ma J, Bennetzen JL (2004) Rapid recent growth and divergence of rice nuclear genomes. Proc Nat Acad Sci USA101:12404–12410

Marri PR, Sarla N, Reddy LV, Siddiq EA (2005) Identification and mapping of yield and yield related QTLs from an Indian accession of *Oryza rufipogon*. BMC Genet UK 6:33 DOI: 10.1186/1471-2156-1186-1133

Matsuo T, Futsuhara Y, Kikuchi F, Yamaguchi H (1997) Science of the rice plant. In: Matsuo T, Futsuhara Y, Kikuchi F, Yamaguchi H (eds) Food and Agriculture Policy Research Center, Tokyo

McClung AM, Marchetti MA, Webb BD, Bollich CN (1997) Registration of 'Jefferson'' rice. Crop Sci 37:629–630

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 9:199–207

Moncada M, Martínez C, Tohme J, Guimaraes E, Chatel M, Borrero J, Gauch H, McCouch S (2001) Quantitative trait loci for yield and yield components in an *Oryza sativa* x *Oryza rufipogon* BC2F2 population evaluated in an upland environment. Theor Appl Genet 102:41–52

Monna L, Kitazawa N, Yoshino R, Suzuki J, Masuda H, Maehara Y, Tanji M, Sato M, Nasu S, Minobe Y (2002) Positional cloning of rice semidwarfing gene, sd-1: rice ''green revolution gene'' encodes a mutant enzyme involved in gibberellin synthesis. DNA Res 9:11–17

Morishima H, Sano Y, Oka HI (1984) Differentiation of perennial and annual types due to habitat conditions in the wild rice *Oryza perennis*. Plant Syst Evol 144:119–135

Muller HJ (1942) Isolating mechanisms, evolution and temperature. Biol Sump 6:71–125

Nguyen BD, Brar DS, Bui BC, Nguyen TV, Pham LN, Nguyen HT (2003) Identification and mapping of the QTL for aluminum tolerance introgressed from the new source, *Oryza rufipogon* Griff., into indica rice (*Oryza sativa* L.). Theor Appl Genet 106:583–593

Ni J, Colowit PM, Mackill DJ (2002) Evaluation of genetic diversity in rice subspecies using microsatellite markers. Crop Sci 42:601–607

Oka HI (1988) Origin of cultivated rice. In: Oka HI (ed) Elsevier Science/Japan Scientific Societies Press, Tokyo, p 254

Oka HI, Morishima H (1967) Variations in the breeding systems of a wild rice, *Oryza perennis*. Evolution 21:249–258

Oki T, Masuda M, Kobayashi M, Nishiba Y, Furuta S, Suda I, Sato T (2002) Polymeric Procyanidins as Radical-Scavenging Components in Red-Hulled Rice. J Agriculture Food Chem 50:7524–7529

Payne CT, Zhang F, Lloyd AM (2000) GL3 Encodes a bHLH Protein That Regulates Trichome Development in *Arabidopsis* Through Interaction With GL1 and TTG1. Genetics 156:1349–1362

Peloquin SJ (1983) Utilization of exotic germplasm in potato breeding: germplasm transfer with haploids and 2n gametes. In: Conservation and utilization of exotic germplasm to improve varieties.; Plant Breeding Research Forum: WHERE? pp 147–167

Peng J, Richards DE, Hartley NM, Murphy GP, Devos KM, Flintham JE, Beales J, Fish LJ, Worland AJ, Pelica F, Sudhakar D, Christou P, Snape JW, Gale MD, Harberd NP (1999) 'Green Revolution' genes encode mutant gibberellin response modulators. Nature 400:256–261

Pillen K, Zacharias A, Leon J (2003) Advanced backcross QTL analysis in barley (*Hordeum vulgare* L.). Theor Appl Genet 107:340–352

Rao GU, Ben Chaim A, Borovsky Y, Paran I (2003) Mapping of yield-related QTLs in pepper in an interspecific cross of Capisicum annum and *C. frutescens*. Theor Appl Genet 106:1457–1466

Redona ED, Mackill DJ (1998) Quantitative trait locus analysis for rice panicle and grain characteristics. Theor Appl Genet 96:957–963

Rick CM (1983) Conservation and use of exotic tomato germplasm. In: Conservation and utilization of exotic germplasm to improve varieties. Report of the 1983 Plant Breeding Research Forum pp 147–167

Rick CM XVII (1967) Int Hort Congr 3:217–229

Roy SC (1921) A preliminary classification of the wild rices of the Central Province and Berar. Agric J India 16:365–380

Sano Y (1993) Constraints in using wild relatives in breeding: lack of basic knowledge on crop gene pools. In: Buxton DR (ed) International crop science. Crop Science Society of America, Madison, WI, pp 437–443

Sasaki A, Ashikari M, Ueguchi-Tanaka M, Itoh H, Nishimura A, Swapan D, Ishiyama K, Saito T, Kobayashi M, Khush GS, Kitano H, Matsuoka M (2002) A mutant gibberellin-synthesis gene in rice. Nature 416:701–702

Sato Y, Nishio T (2003) Mutation detection in rice waxy mutants by PCR-RF-SSCP. Theor Appl Genet 107:560–567

Second G (1991) Molecular markers in rice systematics and the evaluation of genetic resources. Biotech Agric Forest 14:468–494

Second G (1982) Origin of the genetic diversity of cultivated rice (Oryza spp.): Study of the polymorphism scored at 40 isozyme loci. Japanese J Genet 57:25–57

Septiningsih EM, Prasetiyono J, Lubis E, Tai TH, Tjubaryat T, Moeljopawiro S, McCouch SR (2003a) 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

Septiningsih EM (2002) Identification of near-isogenic line development and fine mapping of quantitative trait loci from the rice cultivar IR64 and its wild relative *Oryza rufipogon*. PhD Thesis, Cornell

Shirley B (1998) Flavinoids in seeds and grains: physiological function, agronomic importance and the genetics of biosynthesis. Seed Sci Res 8:415–422

Spelt C, Quatrocchio F, Mol J, Koes R (2002) ANTHOCYANIN1 of petunia controls pigment synthesis, vacuolar pH, and seed coat development by genetically distinct mechanisms. Plant Cell 14:2121–2135

Spielmeyer W, Ellis MH, Chandler PM (2002) Semidwarf (sd-1), ''green revolution'' rice, contains a defective gibberellin 20-oxidase gene. Proc Natl Acad Sci USA 99:9043–9048

Steele KA, Price AH, Shashidhar HE, Witcombe JR (2005) Marker-assisted selection to introgress rice QTLs controlling root traits into an Indian upland rice variety. Theor Appl Genet 112:208–221

Sun CQ, Wang XK, Yoshimura A, Doi K (2002) Genetic differentiation for nuclear mitochrodrial and chloroplast genomes in common wild rice (*Oryza rufipogon* Griff.) and cultivated rice (*Oryza sativa* L.). Theor Appl Genet 104:1335–1345

Sweeney M, Thomson MJ, Pfeil B, McCouch S (2006) Caught red-handed: *Rc* encodes a basic helix-loop-helix protein conditioning red pericarp in rice. The Plant Cell 18:283–294

Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277:1063–1066

Tanksley SD, Nelson JC (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and tranfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor Appl Genet 92:191–203

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

Terada R, Urawa H, Inagaki Y, Tsugane K, Iida S (2002) Efficient gene tarteting by homologous recombination in rice. Nat Biotechnol 20:983–984

Thomson MJ, Edwards JD, Septiningsih EM, Harrington S, McCouch SR (2006) Substitution mapping of *dth1.1*, a flowering time QTL associated with transgressive variation in rice, reveals a cluster of QTLs. Genetics: Genetics: doi 10.1534/genetics.1105.050500

Thomson MJ, Tai TH, McClung AM, Hinga ME, Lobos KB, Xu Y, Martinez C, 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

Vaughan DA, Kadowaki KI, Kaga A, Tomooka N (2005) On the Phylogeny and Biogeography of the Genus *Oryza*. Breed Sci 55:113–122

Wang YM, Dong ZY, Zhang ZJ, Lin Y, Shen D, Zhou D, Liu B (2005) Extensive *de Novo* genomic variation in rice induced by introgression from wild rice (*Zizania latifolia* Griseb.). Genetics 170

Xiao J, Li J, Grandillo S, Ahn SN, Yuan L, Tanksley SD, McCouch SR (1998) Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. Genetics 150:899–909

Xiao J, Li J, Yuan L, Tanksley SD (1996) Identification of QTLs affecting traits of agronomic importance in a recombinant inbred population derived from a subspecific cross. Theor Appl Genet 92:230–244

Xing Y-Z, Tan Y-F, Hua J-P, Sun X-L, Xu C-G, Zhang Q (2002) Characterization of the main effects, epistatic effects and their environmental interactions of QTLs on the genetic basis of yield traits in rice. Theor Appl Genet 105:248–247

Xing YZ, Tan YF, Xu CG, Hua JP, Sun XL (2001) Mapping quantitative trait loci for grain appearance traits of rice using a recombinant inbred line population. Acta Botanica Sinica 43:721–726

Xiong LZ, Liu KD, Dai XK, Xu CG, Zhang Q (1999) Identification of genetic factors controlling domestication-related traits of rice using an F2 population of a cross between *Oryza sativa* and *O. rufipogon*. Theor Appl Genet 98:243–251

Yamanaka S, Nakamura I, Watanabe KN, Sato YI (2004) Identification of SNPs in the waxy gene among glutinous rice cultivars and their evolutionary significance during the domestication process of rice. Theor Appl Genet 108:1200–1204

Yu SB, Li JX, Xu CG, Li XH, Tan YF, Zhang Q (2002) Identification of quantitative trait loci and epistatic interactions for plant height and heading date in rice. Theor Appl Genet 104:619–625

Yu SB, Li JX, Xu CG, Tan YF, Gao YJ, Li XH, Zhang Q, Sagai X, Maroof MA (1997) Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. Proc Natl Acad Sci 94:9226–9231

Zeng W (1994) Precision mapping of quantitative trait loci. Genetics 136:1457–1468

Zheng HG, Qian HR, Shen BZ, Zhuang JY, Lin HX, Lu J (1994) RFLP-based Phylogenetic analysis of wide compatibility varieties in *Oryza sativa* L. Theor Appl Genet 88:65–69