Theor Appl Genet (2004) 109: 1247–1260
DOI 10.1007/s00122-004-1732-7

ORIGINAL PAPER

K. A. Steele · G. Edwards · J. Zhu · J. R. Witcombe

Marker-evaluated selection in rice: shifts in allele frequency among bulks selected in contrasting agricultural environments identify genomic regions of importance to rice adaptation and breeding

Received: 3 November 2003 / Accepted: 13 May 2004 / Published online: 5 August 2004

Abstract Conventional methods for quantitative trait locus (QTL) mapping require the selection of particular traits to be measured based on assumptions as to their importance. We have tested an alternative approach for the location of QTLs—marker-evaluated selection—that makes no prior assumptions as to which traits are important. The results of phenotype selection were evaluated in the products of modified bulk-population breeding that was replicated across a range of rice ecosystems. Selection was carried out in close collaboration with farmers in bulk populations that were all derived from a cross between an Indian upland variety (Kalinga III) and a high-yielding semi-dwarf variety (IR64). Twenty-seven diverse bulks were produced that were screened with molecular markers in order to determine whether shifts could be detected in marker allele frequency as a result of selection and if such changes varied by genomic region across ecosystems. Marker loci linked to important traits for adaptation to specific environments were identified without making any prior assumptions about which traits might be important. Genomic regions from Kalinga III were strongly selected in the upland environments and regions from IR64 in the lowland ones. However, exceptions occurred where the upland parent contributed positively to lowland adaptation and vice versa. The results can be used as a basis for the development of second-cycle varieties, using marker-assisted selection to produce genotypic ideotypes for specific target environments. The very strong selection for genomic regions from the adapted parents of the wide (upland × lowland) cross indicates that, in non-marker-assisted breeding, where genetically distant parents have been used, modified backcross breeding should be efficient. A single backcross to the adapted parent for a specific ecosystem will result in a higher frequency of segregants with the desired high genetic contribution from the adapted parent.

Communicated by D.J. Mackill

K. A. Steele (✉) · G. Edwards · J. R. Witcombe
Centre for Arid Zone Studies, University of Wales,
Bangor, Gwynedd, LL57 1AY, UK
e-mail: k.a.steele@bangor.ac.uk
Tel.: +44-1248-382285
Fax: +44-1248-364717

J. Zhu
Comparative Genetics Unit, John Innes Centre, Norwich
Research Park,
Colney,
Norwich, NR4 7UH, UK

SelecGen, Norwich Research Park,
Colney Lane,
Norwich NR4 7UH, UK

Introduction

Quantitative trait locus (QTL) mapping requires analysis of both molecular markers and pre-defined, measurable phenotypic traits. Traditionally this method involves a full evaluation of many recombinant inbred lines (RILs) derived from a cross between two inbred lines that differ for a quantitative trait. It is expensive because very large numbers of RILs must be evaluated in order to detect QTLs with moderate effects. QTL mapping for complex traits, such as yield under stress, is difficult when phenotypic evaluation has to be carried out in typical field conditions because of uncontrolled environmental variation and the effects of genotype × environment interaction. The efforts of breeders using marker-assisted selection (MAS) to select for QTLs controlling complex traits such as drought resistance or yield have been hampered by these limitations (Reyna and Sneller 2001; Shen et al. 2001).

Various adaptations of conventional QTL analysis have been suggested whereby the size of the population to be genotyped is reduced. Selective genotyping reduces the population of RILs for genotyping to only those with the most extreme phenotypes of a pre-defined trait (Stuber et al. 1980; Lander and Botstein 1989). Lebowitz et al. (1987) estimated that selective genotyping was a powerful alternative to conventional QTL analysis, but noted that it

is limited to circumstances where a single trait is of interest. Michelmore et al. (1991) developed bulk segregant analysis (BSA) where DNA from RILs with contrasting expression of a trait is combined into two pools (one high expressing and the other low expressing). This reduces even more than selective genotyping the number of samples that are genotyped. In both BSA and selective genotyping, selection for a trait is expected to change allelic frequencies of segregating alleles at QTLs under selection leading to an increased frequency of positive alleles in the high tail and a decreased frequency of negative alleles in the low tail. BSA is often used as a complementary method to map-based QTL analysis; an example was the analysis of maize bulks selected for yield under stress (Quarrie et al. 1999).

We have tested a novel method for evaluation of the effects of selection in rice on marker frequency: marker-evaluated selection (MES). The selection was not for a pre-defined trait but for a wide range of agronomic characteristics that determine adaptation to a particular environment. No prior assumptions were made about which traits were likely to be important agronomically; hence, our approach was not trait based. It was carried out in bulk populations of an upland × lowland rice cross, in close collaboration with farmers and in several ecosystems with replication of the selection within each ecosystem. We genotyped representative samples of all promising

Table 1 Details of the parents (Kalinga III and IR64) and 27 products of modified bulk selection carried out by farmers in India and Nepal for different seasons and ecosystems. Plant height and number of days to anthesis were recorded under glasshouse conditions for four replicates and their standard errors are given

Name of bulk or variety	Derived from bulk [a]	Country	Season	Ecosystem	Number of individuals in DNA extract	Height on day 111 (cm)	Days to antheseis
IR64	(Parent)	All Asia	Any	Irrigated lowland (4)	11	73±3.0	160±3.0
Kalinga III	(Parent)	India	Main	Upland (1)	10	120±5.5	85±0.6
Judi 102 (ET)	ET	Nepal	*Chaite*[b]	Upland (1)	1	109±10.1	107±14.3
Judi 101	MT	Nepal	*Chaite*	Upland (1)	1	127±1.0	84±0.7
Judi 103	MT	Nepal	*Chaite*	Upland (1)	1	125±0.9	104±3.4
Judi 111 (MT1)	MT1	Nepal	*Chaite*	Upland (1)	7	110±9.2	120±15.6
Barkhe 1027	MT	Nepal	Main	Upland (1)	1	79±2.7	101±1.0
Ashoka 200F	Bulk	India	Main	Upland (1)	10	115±3.2	88±1.8
Ashoka 228	BF	India	Main	Upland (1)	11	116±3.6	86±1.3
Ashoka 238	BF	India	Main	Upland (1)	9	122±4.0	87±0.5
Ashoka 18-2	BF	India	Main	Upland (1)	1	122±3.4	93±2.5
Ecosystem mean						115±2.6	95±2.6
Judi 523	MD	Nepal	*Chaite*	Long duration (2)	1	83±1.9	99±2.7
Judi 566	MD	Nepal	*Chaite*	Long duration (2)	11	74±2.3	172±7.8
Judi 567	MD	Nepal	*Chaite*	Long duration (2)	13	80±7.6	153±2.1
Ecosystem mean						79±6.4	141±9.6
Barkhe 2003	MT4	Nepal	Main	Medium-low (3)	1	86±7.6	193±46.6
Barkhe 2004	MT4	Nepal	Main	Medium-low (3)	1	102±3.3	164±14.9
Barkhe 2012	MT4	Nepal	Main	Medium-low (3)	1	85±1.2	157±71.5
Barkhe 2014	MT1	Nepal	Main	Medium-low (3)	1	69±3.5	203±9.3
Ashoka 281	BF	India	Main	Medium-upland (3)	1	111±4.2	102±1.1
Ashoka 125	BF	India	Main	Medium-upland (3)	1	109	122
Ashoka 165	BF	India	Main	Medium-upland (3)	10	113±1.9	99±1.9
Ashoka 157	BF	India	Main	Medium-upland (3)	10	113±9.1	90±5.6
Ecosystem mean						98±3.2	136±11.0
Barkhe 3004	LD	Nepal	Main	Lowland (4)	1	74±3.9	199±34.2
Barkhe 3005	LD	Nepal	Main	Lowland (4)	1	75±2.3	176±9.8
Barkhe 3008	LD	Nepal	Main	Lowland (4)	1	72±5.1	193±11.9
Barkhe 3010	LD	Nepal	Main	Lowland (4)	4	75±6.0	182±13.3
MD	MD	Nepal	Main	Lowland (4)	10	76±4.7	106±5.8
LD	LD	Nepal	Main	Lowland (4)	9	71±3.9	143±6.6
Ashoka 900F	Bulk	India	Main	Lowland (4)	10	70±0.6	162±4.0
Ecosystem mean						73±1.4	168±8.5

[a] *BF* F_3:F_4 bulk family, *ET* early tall, *MT* medium tall, *MD* medium dwarf, *LD* late dwarf
[b] *Chaite* Early- or spring-season rice grown in Nepal

bulks produced by the selection in these environments. We evaluated whether these bulks differed for alleles at QTLs controlling agronomic traits by examining the significance of any shifts in allelic frequencies as a result of selection.

The results are discussed with a view to breeding second-cycle products, using MAS (genetic ideotypes). We also discuss the potential of MES in relation to developments in genomics and the potential of modified backcross breeding for breeding programmes when markers are not employed.

Materials and methods

Plant material and breeding strategy

Crosses were made between an Indian upland variety, Kalinga III, and a high-yielding semi-dwarf, IR64 (cross ID: IR74263). The crosses were made by Dr. Brigitte Courtois at the International Rice Research Institute (IRRI) in 1996, at the request of the Centre for Arid Zone Studies (CAZS), University of Wales, Bangor, UK. A large quantity of F_2 seed was produced and supplied by CAZS to Birsa Agricultural University (BAU), India, and the Gramin Vikas Trust (GVT) through the National Bureau of Plant Genetic Resources, New Delhi. An F_2 population of about 10,300 plants was grown at the GVT/BAU upland farm, Kanke, Ranchi, eastern India, during the main (rainy) season of 1997. In later generations, modified bulk selection through participatory plant breeding (described in detail below) was conducted in a range of situations in India and Nepal: upland (coded 1), early season in Nepal (coded 2), medium-upland and medium-lowland (both coded 3) and lowland (irrigated) conditions (coded 4).

Farmers were closely involved in the selection, but the process was led, in India, by scientists from BAU, GVT and CAZS and, in Nepal, by scientists from the Local Initiatives for Biodiversity in Research and Development (LI-BIRD) and CAZS. Two types of farmer selection were employed: collaborative selection, where farmers selected within bulk populations in their own fields grown under traditional management and consultative selection, where farmers made selections within populations (usually bulk lines) grown on research stations but under fertility and moisture regimes close to those of farmers' fields. As a control population, 26 remnant F_1 seeds from IRRI of the same cross were sown in 2002 under glasshouse conditions at CAZS, and all F_2 seed was harvested from two F_1 plants. Fifty non-selected F_2 seeds were sown under glasshouse conditions and DNA was extracted from them at the seedling stage to form the F_2 control population.

Selection in India

The selection in India is described in detail in Virk et al. (2003). All the selected bulks in India from the cross Kalinga III × IR64 were given the name 'Ashoka' with a unique identification number (Table 1). At CRRI, Cuttak, Orissa, 600 F_3 bulk lines derived from all of the seed of individual F_2 plants were grown in the off-season (December–April) of 1997–1998. For collaborative selection farmers in three states were given an F_4 bulk made up of all of the lines in the 1998 main season. One farmer (RD) mass selected in the F_4 and the resultant selected bulk was advanced on station to F_6 and entered into trials in 1999 as Ashoka 200F. Another farmer (SKM) mass selected over four generations and returned seed of the F_8 bulk to breeders in 2002; this was entered into trials as Ashoka 900F. For consultative selection by farmers, 177 F_4 family bulks (from the harvest of breeder-selected F_3 lines) were grown in 10-m^2 plots at the GVT/BAU upland farm in the 1998 main season. The F_5 seed from 12 of these family bulks (representing selections for upland, medium land and lowland) was advanced to the F_6 at CRRI, Cuttack in the 1998–1999 off-season (Kumar et al. 2001). Seven of the family bulks from the consultative selection that had the best performance in replicated trials in subsequent years under upland or medium-land environments were included in the study.

Selection in Nepal

Selection in Nepal commenced with 290 F_2:F_3 families in a farmer's field in Chitwan district, in the spring (*Chaite*) season of 1998 (Witcombe et al. 2001). There was much variation among and within the families, and the seed harvested from individual plants was grouped into six bulks based on height and maturity (early <110 days seed to seed, medium 110–125 days and late >125 days). These bulks were made up from all of the F_4 seed from all of the healthy F_3 plants. Resultant F_4 bulks were grown in the main season of 1998 and designated as: early dwarf (ED), early tall (ET), medium dwarf (MD), medium tall (MT), late dwarf (LD) and late tall (LT). Of these, the performance of ED proved to be unsatisfactory and it was dropped. The highly variable MT bulk was further divided into four to produce F_5 bulk seed for: MT1 = earlier shorter, MT2 = earlier taller, MT3 = later shorter and MT4 = later taller. In the *Chaite* season of 1999 three bulks (ET, MD and LT) were advanced from F_5 to F_6 without further division. However, in five of the bulks (MT1 to MT4 and LD), further division was made at harvest by selecting among the F_6 plants, into grain type, i.e. long, intermediate or short in length. In the main season 1999, the bulks were multiplied and advanced from the F_6 to the F_7. Some bulks were rejected in the main season of 1999 on the basis of maturity and grain type.

In the F_7 generation, in *Chaite* 2000, four selected bulks (ET, MT1 long slender grains, MT2 long slender grains and MT3 shorter grains) were distributed to 24 farmers (six farmers per bulk) in the Chitwan district for collaborative selection. More than 40 bulks or family bulks selected from the modified bulks were obtained in Nepal through collaborative selection. For bulks there

were always at least 100 plants selected and sown in the next generation; for family bulks the random effects were larger, but there was no substantial difference in selection pressure between the two methods. Selection pressure was less strong for F_5 bulks ED, ET, MD, MT, LD and LT than for the bulks selected in collaborative selection. Early season (*Chaite*) selections were given the name 'Judi,' followed by a three-digit code, and main season selections were given the name 'Barkhe,' followed by a four-digit code. Eighteen bulks were used for the molecular analysis (Table 1). The bulk named Judi 102 was the entire ET bulk selected by farmers for upland conditions in the *Chaite* season. The bulks MD and LD were appropriate for the lowland environments even though they were not selected for this ecosystem by farmers. They were included in this study for molecular analysis (Table 1).

Glasshouse screen for height and maturity

Plants of 26 selected bulks (Ashoka 125 was not included), Kalinga III (two samples) and IR64 (two samples) were grown in 5-l pots filled with John Innes Potting Compost No. 1 mixed with 20 g Osmocote patterned release fertiliser (15 %N:9% P:9% K:3% MgO, plus trace elements). They were grown to maturity in a randomised block layout with four replicates under glasshouse conditions (minimum temperature 25°C) in daylight supplemented by 150 μmol m^{-2} s^{-1} PAR. They were sown on 26 February 2003 and watered twice daily. Total plant height on day 111 and the number of days to first anthesis were recorded for each plant.

DNA extraction

Twenty-seven selected bulks and 50 F_2 control plants (29 progeny from F_1 line number 5 and 21 progeny from F_1 line number 8) were grown under glasshouse conditions for DNA extractions at the seedling stage. DNA was extracted from 50 individual F_2 seedlings (un-selected) and from 27 selected bulks (Table 1), using leaf material from either one or several (maximum 13) individual plants. DNA extraction was carried out using DNeasy plant kits (Qiagen, Crawley, West Sussex, UK), following the recommended protocol for 100 mg of fresh leaf tissue, frozen and ground in liquid nitrogen, using a chilled pestle and mortar. The DNA concentration of samples was estimated on 1% (w/v) agarose gels alongside λ DNA of known concentration.

Microsatellite analysis

Microsatellite primer pairs for amplification of simple sequence repeats (SSRs) in rice were used (http://www.gramene.org/microsat/). Fifty-four SSR marker loci were chosen because many of them mapped to regions where QTLs (influencing maturity, plant height and roots) had been identified in previous studies of IR64 mapping populations. SSRs were assigned to nine panels of six markers for pooled fragment analysis with fluorescence detection on the CEQ 8000 DNA Analysis System (Beckman Coulter, Fullerton, Calif., USA). Panels were designed so that fragments in different size ranges could be detected with the same fluorescent dye, and fragments with overlapping size ranges were detected with different-coloured dyes. One labelled primer (Invitrogen, Paisley, UK) was required for each SSR locus, and in four of the panels one primer of the pair (either forward or reverse) was end labelled with one of three Beckman Coulter dyes (D2, D3 and D4 made using active ester (AE)-linked dye labels). In order to reduce costs in the remaining five panels, the forward primers each had an additional 19-base tail sequence (CACGACGTTGTAAAACGAC) at the 5′ end (Oetting et al. 1995). The PCR mix for tailed primers contained three primers: (1) the forward-tailed primer, (2) the reverse primer and (3) a universal 19-base tail primer labelled with a Beckman Coulter dye. Multiplex PCR was possible for a maximum of three SSRs from each panel. For tailed primers multiplex PCR was only possible for markers which amplified fragments of different size ranges because only one dye-labelled tail could be used in each PCR. When an SSR marker was not informative for distinguishing IR64 and Kalinga III, it was not included in the panel for analysis of progeny samples.

PCR was carried out using Reddy Mix PCR Master Mix 3.0 mM $MgCl_2$ (ABgene, Epsom, Surrey, UK) in a total reaction volume of 16 μl containing 2.5 ng/μl genomic DNA. For non-tailed primers 0.25 μM of both forward and reverse primer was used. For tailed primers the concentrations were 0.025 pM of the forward-tailed primer and 0.25 pM of both the reverse primer and the labelled universal tail primer. PCR was carried out using an MJ Research PTC-100 96-well programmable heat/cool block. The PCR-cycling profile was: 94°C for 5 min; followed by 35 cycles of denaturing at 94°C for 1 min; annealing at decreasing temperatures (57°C for 2 cycles, 56°C for 2 cycles, 55°C for 5 cycles and 54°C for 26 cycles) for 1 min; extension at 72°C for 1 min and a final extension for 10 min at 72°C.

PCR products were diluted in deionised formamide and pooled in the ratio: 32 (D2):16 (D3):1 (D4) for each panel. The exact volume was adjusted for individual markers according to amplification intensity, or if different dyes were used in multiplexed PCR (panels 1–4). A maximum of 1 μl pooled PCR product and 1 μl 400-bp size standard (Beckman Coulter) was added to 40 μl deionised formamide in each well of a 96-well sample plate. Fragment analysis in the CEQ 8000 DNA Analysis System used automated sample injection with capillary temperature at 50°C, denaturing for 120 s at 90°C, injection at 2.0 kV for 30 s and 6.0-kV separation for 35 min. CEQ 8000 fragment analysis software (version 2) was used to determine fragment sizes with dye mobility calibration parameter AE-V2. The actual size of fragments detected depended on the dye synthesis method, the calibration parameters used for analysis, the presence of stutter and

±A fragments and the presence of tailed primers. When stuttering occurred, the fragment with the highest concentration (i.e. the highest peak on the chromatogram) was used to call the fragment size to the nearest whole number of bases. Manual re-scoring of all data was carried out; this was essential to detect the presence of three alleles, possible in bulk DNA samples if non-parental alleles were amplified.

AFLP analysis

The AFLP procedure used was based on that described by Vos et al. (1995) with *Eco*RI and *Mse*I restriction of genomic DNA. Selective primers were combinations of E primers with two selective nucleotides (E12, E22, E23) and M primers with three selective nucleotides (M35, M48, M63). The fragments amplified with these primers have been mapped in two mapping populations and assigned to DNA fingerprint linkage blocks (Zhu et al. 1999).

Single-nucleotide polymorphism analysis

Single-nucleotide polymorphism (SNP) analysis was carried out on 18 selected bulks (samples Judi 566 and 567 were pooled together), the two parents and negative controls. Each sample was genotyped four times and the percentage of concordance was 99%. Twelve SNP loci were tested, and the assays were run as three pooled reactions with four markers in each. An initial 5-μl PCR reaction was carried out to amplify the region containing the SNP site of interest from genomic DNA. Shrimp alkaline phosphatase (SAP) was added to the PCR product to dephosphorylate residual nucleotides; this was followed by heat inactivation of SAP. Extension primer, DNA polymerase and dideoxy- and deoxynucleotides were added for the primer extension reaction which was carried out using homogeneous MassEXTEND (Sequenom, San Diego, Calif., USA). Following this reaction SpectroCLEAN resin (Sequenom) was added to remove extraneous salts. For each sample 15 nl of extension product were spotted onto the pad of a 384 SpectroCHIP (Sequenom) array for mass spectrometry using a MALDI-TOF (matrix-assisted laser desorption/ionization—time of flight mass spectrometer). Genotypes were called with SpectroTYPER-RT software (Sequenom, San Diego, Calif., USA).

Data analysis

Data from 27 SSR loci and 1 SNP locus were summarised in a data matrix containing a single code for each SSR/variety combination: K = Kalinga III allele, H = Kalinga III and IR64, I = IR64 allele, M = IR64 and non-parental allele (NPA), N = NPA (one or more), O = Kalinga III and NPA, and P = NPA and Kalinga III and IR64). For the 13 bulks where bulk DNA from several plants was sampled, all alleles detected at a locus were assumed to have equal frequencies (0.5 for two alleles, 0.33 for three alleles). A two-way ANOVA in SPSS (version 11) was carried out on overall allele frequency scores (percentage) for 27 selected bulks and 50 F_2 lines. This analysis used country and ecosystem as factors and Tamhane's T2 post-hoc tests to determine where significant differences occurred. Tamhane's T2 is a conservative, pairwise comparisons test based on a *t*-test, and it assumes unequal variances. The frequency data were also arcsin transformed, and the ANOVA results were the same.

For each locus, the chi-squared (χ^2) test of goodness of fit to a 1:1 ratio was used to test the allele frequency of the two parental types (to detect segregation distortion) in the whole population of 27 selected bulks and also in the 50 F_2 controls. To test for significant difference at each locus between the F_2 and all the selected bulks, and between each set of bulks selected for different ecosystems, countries and seasons, the *G*-test of independence (log-likelihood ratio test) was used with a model I design (Sokal and Rohlf 1995). The three bulks to which the least selection pressure had been applied and so were the least well adapted [ET (Judi 102), MD and LD] were not included in the *G*-tests of independence between ecosystems. For cluster analysis squared Euclidian distance was calculated in SPSS (data were entered as: 1 = Kalinga III, 2 = a mixture of both parent alleles, 3 = IR64), and dendrograms were constructed using average linkage between groups. Linear regression was used to test the relationship between days to maturity and genotype at the marker RM248.

Results

Phenotypes of selected bulks

Farmers with varying selection criteria selected or identified a wide range of successful phenotypes for the target ecosystems from this one cross. Each selected bulk had many of the agronomic traits required for the specific target environment in which it was selected (Table 1). Only two phenotypic characteristics were measured in the glasshouse: height and number of days to anthesis. Both of these traits varied significantly among ecosystems: height, $P>0.001$; days to anthesis, $P>0.001$ ($df=15, 104$); the bulks selected in the lowlands were shorter and later than those from the upland (Table 1). For both traits, the bulks selected for the non-upland ecosystems (2, 3 and 4) differed significantly from upland Kalinga III, and the bulks selected for the upland ecosystem 1 differed significantly from lowland IR64.

Molecular polymorphism of selected and control populations

Twenty-seven SSR loci, located on ten chromosomes, were polymorphic and showed segregation in the F_2

population (Table 2). NPAs were detected at 18 loci (5% of total) in the 27 bulks but not in the F_2 population. Twelve SNP loci were tested on 18 bulks. Only one SNP marker (C10B2L2) showed segregation in the selected population (Kalinga III = nucleotide C and IR64 = nucleotide T or both T/C), and it was the only SNP locus used for allele frequency analysis. SNPs were not tested in the F_2 population.

In the 27 bulks 83 alleles were detected at 28 loci (Table 3). Individual results from 50 control F_2 plants are not shown; the F_2 population had a mean of 54% Kalinga III alleles, with a range in individual F_2 plants from 31% to 81%. Overall, the 27 selected bulks had 64% Kalinga III alleles, 30% IR64 alleles and 6% NPAs.

The SSR loci were each tested for departure from the expected parental allele frequency of 1:1. In the F_2 control population four loci showed segregation distortion; three were on chromosome 5 (RM163, χ^2=7.35, RM440, χ^2=6.76, RM188, χ^2=8.00), and one was near the end of the long arm of chromosome 7 (RM248, χ^2=28.17). In the 27 selected bulks deviations from a 1:1 ratio could be due to segregation distortion, or they could be the result of selection for an allele in the majority of the environments tested. Fifteen loci had ratios that differed from 1:1 in the selected bulks, with Kalinga III alleles always the most frequent (RM5, RM110, RM233, RM279, RM262, RM349, RM153, RM593, RM163, RM188, RM334, RM2, RM234, RM258). Of these, only RM163 and RM188 also showed segregation distortion in the F_2, again with Kalinga III alleles more frequent. However, *G*-tests to compare frequencies between bulks (with NPAs excluded) and the control population revealed that only at the locus RM279 was there a higher frequency of Kalinga III alleles in the selected populations than in the control population. This is evidence that selection has favoured Kalinga III alleles at this locus. The test for deviation from a 1:1 ratio is more likely to be significant than a test for change in frequency from the control population. For example, at RM5 the 1:1 ratio is tested as 21%:79%, and the change from the control as 50%:50%. It is unsurprising, therefore, that there was an overall significant shift in frequency in favour of Kalinga III alleles, compared to the control only at one locus.

Allele frequency in selected bulks

No IR64 alleles were detected at the 28 loci tested in three upland selections of Judi 523, Ashoka 200F and Ashoka 238 (Table 3). However, analysis at additional loci with SSRs and AFLPs revealed the presence of IR64 alleles (see below). In contrast to the upland selections, the

Table 2 Polymorphic simple sequence repeat markers used for evaluation of products derived from a cross between varieties Kalinga III and IR64, showing the fragment sizes detected. (NB markers with tailed primers amplify fragments with an extra 18 bp.) *NPAs* Non-parental alleles

Marker	Chromosome	Approximate marker position (cM)	Primer type	Fragment size of Kalinga III allele (bp)	Fragment size of IR64 allele (bp)	NPAs (bp)
RM2	7	36	Tailed	168	166	163
RM5	1	95	F-D3	114	110	108
RM11	7	47	F-D4	144	144, 146	158, 160
RM13	5	30	Tailed	159	163	
RM51	7	0	Tailed	149	159	
RM105	9	32	Tailed	152	145	155
RM110	2	7	Tailed	157	174	
RM153	5	1	Tailed	222	219	212, 214, 225
RM163	5	75	Tailed	159	194, 196	150, 179
RM188	5	101	Tailed	230	232	
RM201	9	81	F-D4	159	157	
RM213	2	186	F-D4	149	137	
RM229	11	78	F-D3	119	117	123
RM233	2	16	Tailed	162	158	176
RM234	7	88	F-D3	155	165	151
RM237	1	115	F-D2	123	135	131, 148
RM242	9	73	F-D4	221	229	219
RM243	1	57	Tailed	131	139	133, 137
RM248	7	100	F-D4	93	103	83
RM258	10	62	Tailed	158	155	150, 162, 166
RM262	2	103	Tailed	174	176	166
RM279	2	17	Tailed	186	195	
RM334	5	142	Tailed	210	195	191, 201
RM349	4	138	F-D2	133	131	
RM440	5	93	Tailed	228	235	
RM593	5	31	Tailed	238	241	
WAXY	6	0	F-D4	108	123	110, 115

Table 3 Genotypes at 28 loci of 27 products of modified bulk selection, and the mean of 50 control F_2 lines, derived from a cross between varieties Kalinga III and IR64. The number of individual plants represented in each DNA sample is shown, and bulks are grouped according to ecosystem code. The chromosome to which the marker maps is shown in parenthesis. Allelic classes are: K=Kalinga III homozygote, H=Kalinga III and IR64, I=IR64 homozygote, M=IR64 and NPA, N=NPA homozygote, O=Kalinga III and NPA, P=NPA and Kalinga III and IR64,—missing data,—percentage of Kalinga III alleles at each locus are shown for 50 F_2 control, un-selected lines

Variety ID	Ecosystem	Number of individuals	RM243 (1)	RM5 (1)	RM237 (1)	RM110 (2)	RM233 (2)	RM279 (2)	RM262 (2)	RM213 (2)	RM349 (4)	RM153 (5)	RM13 (5)	RM593 (5)
Ashoka 200F	1	10	K	K	K	K	K	K	K	K	K	K	K	K
Ashoka 228	1	11	K	K	K	K	K	K	K	K	K	K	K	K
Ashoka 238	1	9	K	K	K	K	K	K	K	K	K	K	K	K
Ashoka 18-2	1	1	K	K	K	K	K	K	K	K	K	K	K	K
Barkhe 1027	1	1	K	K	N	K	I	I	N	I	K	K	K	K
Judi 103	1	1	I	K	N	K	K	K	K	K	I	K	K	K
Judi 101	1	1	K	K	K	K	K	K	K	K	K	N	K	K
Judi 111 (MT1)	1	7	H	K	O	H	H	H	O	H	H	O	H	K
Judi 102 (ET)	1	10	H	K	H	K	H	H	H	H	H	P	K	K
Judi 523	2	1	K	K	N	K	K	K	K	K	K	K	K	K
Judi 566	2	11	I	H	I	K	K	K	H	K	K	K	K	K
Judi 567	2	13	I	I	I	K	K	K	I	K	K	K	K	H
Ashoka 281	3	1	H	K	K	K	I	K	K	K	K	O	I	K
Ashoka 125	3	1	H	I	H	H	K	K	I	H	K	K	K	K
Ashoka 165	3	10	H	K	I	K	K	K	K	K	K	N	K	K
Ashoka 157	3	10	H	K	I	K	K	H	K	K	I	O	K	K
Barkhe 2003	3	1	K	K	K	H	K	–	H	K	I	K	I	K
Barkhe 2004	3	1	O	K	K	I	K	K	K	K	K	K	I	K
Barkhe 2012	3	1	O	N	K	K	K	K	K	K	H	K	I	K
Barkhe 2014	3	1	O	K	M	K	N	I	K	I	K	I	I	K
Ashoka 900F	4	10	I	I	I	H	H	K	I	I	I	I	I	I
Barkhe 3004	4	1	N	I	I	K	K	K	K	I	I	I	I	I
Barkhe 3005	4	1	N	K	K	I	I	I	I	I	K	I	I	I
Barkhe 3008	4	1	I	K	I	K	K	K	K	K	K	I	I	I
Barkhe 3010	4	4	I	H	H	K	K	K	K	K	K	H	K	K
LD	4	9	H	H	M	H	H	H	O	H	I	H	H	H
MD	4	10	H	K	P	H	H	H	H	H	K	N	K	K
F_2 control	–	50	52	50	47	56	58	47	53	49	44	48	54	58

Variety ID	RM163 (5)	RM440 (5)	RM188 (5)	RM334 (5)	Waxy (6)	RM51 (7)	RM2 (7)	RM11 (7)	RM234 (7)	RM248 (7)	RM105 (9)	RM242 (9)	RM201 (9)	C10B2L2 (10)	RM258 (10)	RM229 (11)
Ashoka 200F	K	O	K	K	K	K	K	K	K	K	N	K	K	K	K	K
Ashoka 228	K	O	K	K	K	K	K	K	K	K	I	K	K	K	K	K
Ashoka 238	K	O	K	K	K	K	K	K	K	K	K	K	K	K	K	K
Ashoka 18-2	K	K	K	K	K	K	K	K	K	K	I	K	K	K	K	K
Barkhe 1027	K	K	K	K	K	K	K	K	K	K	K	K	K	–	K	K
Judi 103	K	K	K	N	K	H	K	I	K	K	H	I	K	K	N	N
Judi 101	K	K	K	K	K	K	K	N	K	K	K	K	K	K	O	K
Judi 111 (MT1)	K	H	K	H	K	H	H	H	P	H	H	H	I	H	O	O
Judi 102 (ET)	K	H	K	H	H	H	H	P	K	K	H	H	H	H	P	K
Judi 523	K	K	K	K	K	K	K	K	K	K	K	K	K	–	K	K
Judi 566	K	K	K	K	I	I	K	O	K	I	K	H	K	I	I	H
Judi 567	K	K	K	K	K	K	K	N	K	I	I	K	K	I	I	K
Ashoka 281	K	N	K	K	O	K	I	K	K	K	K	I	I	K	K	K
Ashoka 125	K	K	K	K	H	K	I	I	I	K	H	I	I	H	K	K
Ashoka 165	H	I	H	I	I	K	K	N	K	K	I	H	H	I	K	H
Ashoka 157	H	H	I	M	H	K	K	O	K	K	H	H	I	–	K	H
Barkhe 2003	N	N	I	K	I	K	K	N	K	I	K	N	I	–	M	I
Barkhe 2004	K	K	I	K	I	I	K	O	I	I	I	N	I	–	K	I
Barkhe 2012	N	N	K	N	M	K	K	N	I	N	I	N	I	I	K	I
Barkhe 2014	N	I	K	K	N	K	N	I	K	N	I	N	I	–	I	I
Ashoka 900F	I	I	I	I	I	I	I	I	I	I	I	I	I	–	I	I
Barkhe 3004	I	I	I	I	K	I	I	I	I	I	–	I	I	–	K	I
Barkhe 3005	I	I	K	K	K	I	I	I	K	I	–	K	K	I	K	K
Barkhe 3008	K	K	K	K	K	I	K	K	I	I	–	I	K	K	K	I
Barkhe 3010	K	K	K	K	K	I	K	K	H	I	K	H	K	H	H	K
LD	K	H	K	K	H	H	I	O	M	H	K	I	I	H	K	P
MD	K	K	K	O	H	K	H	P	K	K	H	H	I	–	O	O
F_2 control	64	63	65	58	50	51	62	51	50	77	45	47	51	–	54	43

lowland selections, Ashoka 900F and Barkhe 3004, had a high proportion of IR64 alleles.

Although five varieties (Ashoka 900F, Ashoka 18-2, Ashoka 125, Barkhe 3010 and Barkhe 3008) had no NPAs, at least one NPA was detected in the other 22 varieties. Barkhe 2012 had the greatest proportion of NPAs (26%), four of which were not detected in any other bulk. NPAs could have several possible origins that could vary in importance across the selected populations: variation existing in the parent lines used for the cross not identified in the molecular genotyping with subsequent selection in favour of these NPAs, out-crossing in the field, or new mutations arising at the SSRs.

Loci having both parent alleles were detected in all varieties where bulk DNA samples were tested in accordance with the expectation that these bulks were heterogeneous, i.e. are made up of different homozygous lines. However, in 12 out of 18 DNA samples from individual plants, heterozygous loci were observed. Of these, six were heterozygous for parental alleles (H) and eight were heterozygous for a parental allele and an NPA (M or O) (Table 3).

Allele frequency in contrasting ecosystems

The mean parental allele frequencies for four populations of bulks varied according to selection environment (Table 4; Fig. 1a). With mean frequencies of all loci there was significant difference ($P<0.001$, $df=3, 69$) for parental allele frequency among ecosystems, but no significant difference for parental alleles between countries. There was significant interaction between country and ecosystem ($P<0.001$, $df=2, 69$).

At individual loci, the allele frequencies in each ecosystem group were compared with the unselected F_2, and significant allele shifts could be detected (Table 5). In ecosystem 1, nine loci showed shifts in allele frequency towards Kalinga III, relative to the control population. The allele frequency at a single locus shifted towards IR64, relative to the control, in ecosystem 3 and at another locus in ecosystem 4.

At 12 individual loci, there were significant shifts for parental alleles between ecosystems (Table 5). In ecosys-

Table 4 Mean allele frequencies and their standard errors for 28 marker loci for bulks selected for adaptation to four different ecosystems (1–4), the whole selected population and the F_2 control population

Bulks from ecosystem	Number of bulks	Kalinga III alleles (%)	IR64 alleles (%)	NNPAs (%)
1	9	83±5	11±4	6±2
2	3	76±10	21±10	3±1
3	8	56±4	32±2	12±4
4	7	46±9	50±10	4±2
All (1–4)	27	58±2	40±2	2±1
F_2 control	50 lines	53±2	46±2	0

Table 5 Marker loci with significant shifts in parent allele frequency (percentage) between the control F_2 population and bulks selected in different ecosystems (1–4). The *G*-test was used for all comparisons and the direction of shift according to Kalinga III alleles is indicated. The chromosome to which the marker maps is shown in parenthesis

Comparison	RM243 (1)	RM5 (1)	RM237 (1)	RM279 (2)	RM262 (2)	RM153 (5)	RM13 (5)	RM593 (5)	RM163 (5)	Waxy (6)	RM51 (7)	RM234 (7)	RM248 (7)	RM201 (9)	RM258 (10)	RM229 (11)
1 v F_2		+50**	+22**		+28**	+33**		+42**		+50**		+42**			+21**	+38**
3 v F_2								+42**						−45**		
4 v F_2													−77***			
All v F_2				+36**												
1 v 2			+69**													
1 v 3										+81***				+81***		+44**
1 v 4		+81**	+39**			+71***	+74**	+80***	+60**		+87***	+62**	+94***			
2 v 3														+94**		
2 v 4						+90**										
3 v 4								+80***			+87***		+50**			

Significance levels: ** $P=0.01$, *** $P=0.001$

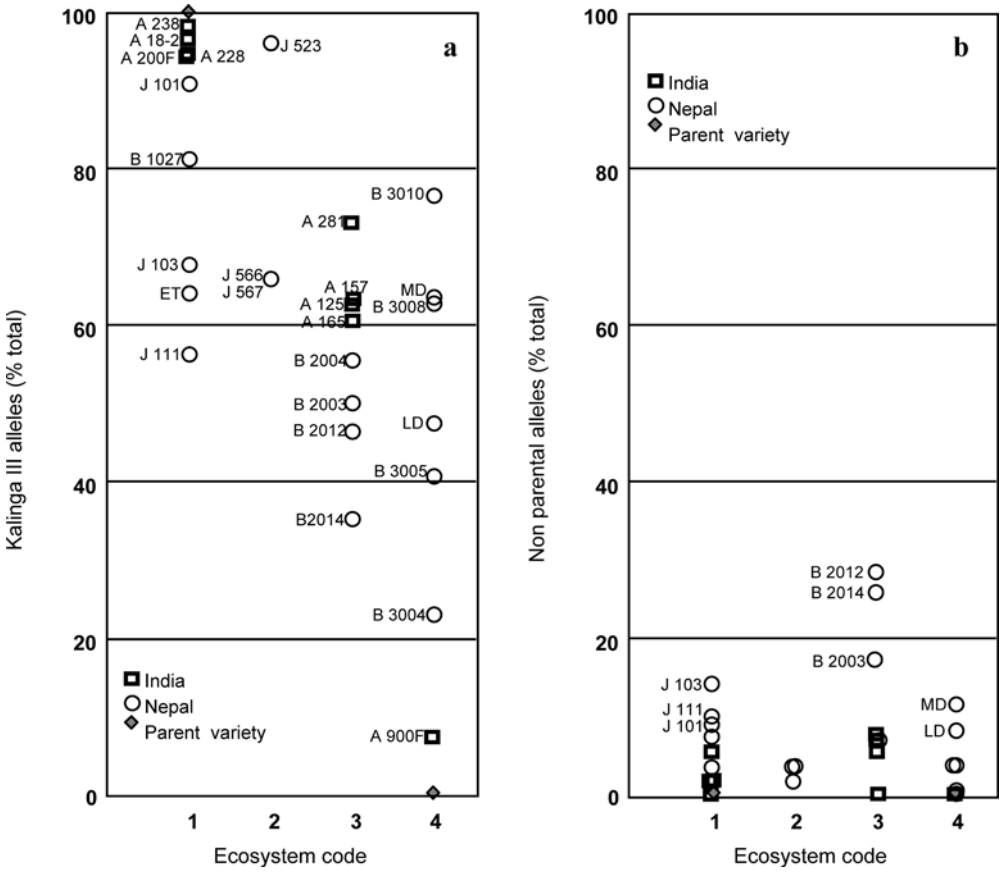

Fig. 1a, b Percentage Kalinga III allele types (**a**) and non-parental allele (NPA) allele types (**b**) according to ecosystem: *1* upland; *2* long-duration *Chaite* (Nepal early season), rainfed; *3* rainfed medium land types; *4* irrigated lowland. Parent varieties Kalinga III and IR64 are represented as *grey diamonds*, selected bulks labelled in **b** when % NPA>8%. *A* Ashoka, *B* Barkhe, *J* Judi. Variance ratios (*F*) and probability (*P*) for each allele type were as follows: NPA F=3.05, P=0.047; Kalinga III F=8.66, P<0.001; IR64 F=9.37, P<0.001

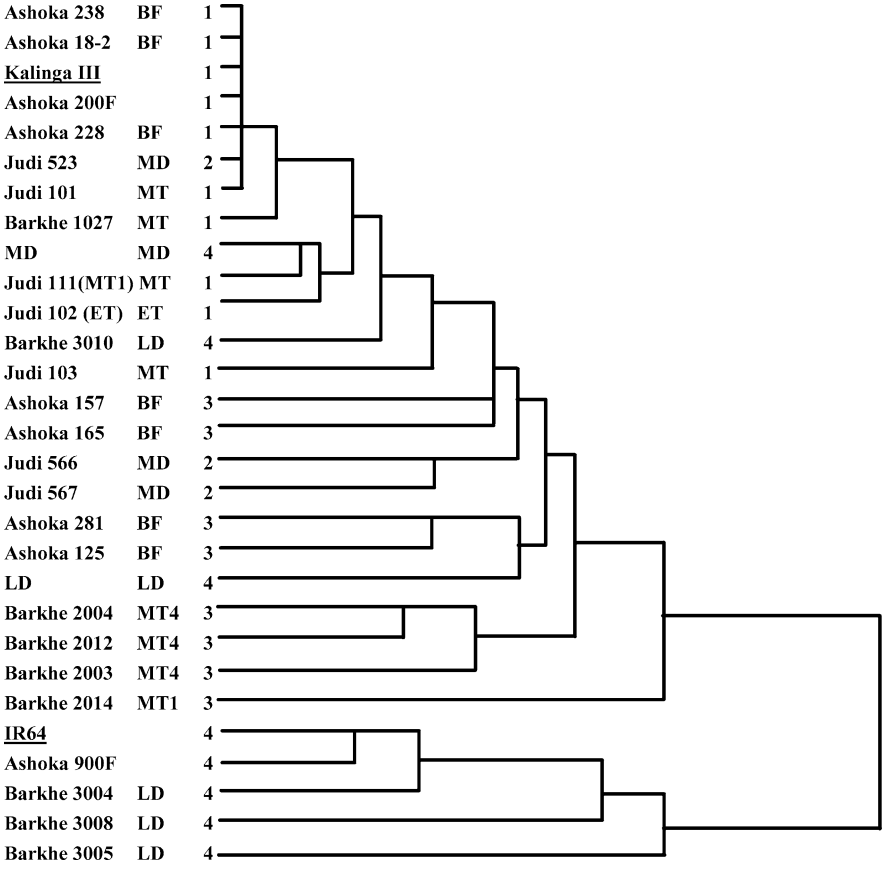

Fig. 2 Dendrogram drawn from analysis of 25 simple sequence repeat (SSR) marker genotypes with one allele type only using squared Euclidian distance with average linkage between groups. Branch lengths are proportional to the combined cluster analysis distances. Varieties are shown with ecosystem codes. Abbreviations are the same as for Table 1

tem 1, all the significant shifts in allele frequency were towards Kalinga III, the parent adapted to this upland ecosystem. Conversely at nine loci, more IR64 alleles were detected in the lowland environment. No shifts between ecosystems were detected at RM279 (chromosome 2), which had a high frequency of Kalinga III alleles in all ecosystems, including the lowland, and a significant overall shift to Kalinga III when compared to the control population. The SSR locus at Waxy (chromosome 6) had a greater proportion of IR64 alleles in medium-land conditions than in the uplands. Both upland and lowland selected bulks had a high proportion of Kalinga III alleles. RM248 alleles shifted towards IR64 in ecosystem 4 compared with the control and with ecosystems 1 and 3. The number of days to anthesis was significantly different between selected bulks having different allelic status at marker RM248 (in the single-marker regression analysis r^2=83%, P<0.001).

Cluster analyses used only the 25 markers for which there was a full data set. Clustering related most clearly to ecosystem groups when segregating loci were classified as neutral in the analysis (Fig. 2). The groupings suggest that there are differences in allele frequencies between ecosystem classes, and that the selection pressure is greatest amongst the varieties selected for ecosystem 1 in India because they show the least variation and clustered together along with Kalinga III. The clustering by ecosystem is even clearer if bulks to which the least selection has been applied are excluded, i.e. LD and MD (ecosystem 4) and Judi 111 and Judi 102 (ecosystem 1).

Additional loci tested in three upland bulks

Three bulks that were selected in upland environments in India (Ashoka 200F, Ashoka 228 and Ashoka 238) were evaluated at 170 additional loci in order to increase the proportion of the genome sampled. Out of 104 AFLP loci, only 33 were polymorphic, and 35 additional SSR loci were polymorphic (Fig. 3). With this wider coverage of genotyping, these varieties were no longer entirely, or almost entirely, Kalinga III. There were seven regions, on chromosomes 2, 4, 5, 6, 9, 10 and 11, where IR64 alleles exist at AFLP loci in all three of these upland varieties. However, the map positions of AFLP loci are less precise than the SSR positions, and this could explain the apparent double recombinants, for example, on chromosome 10.

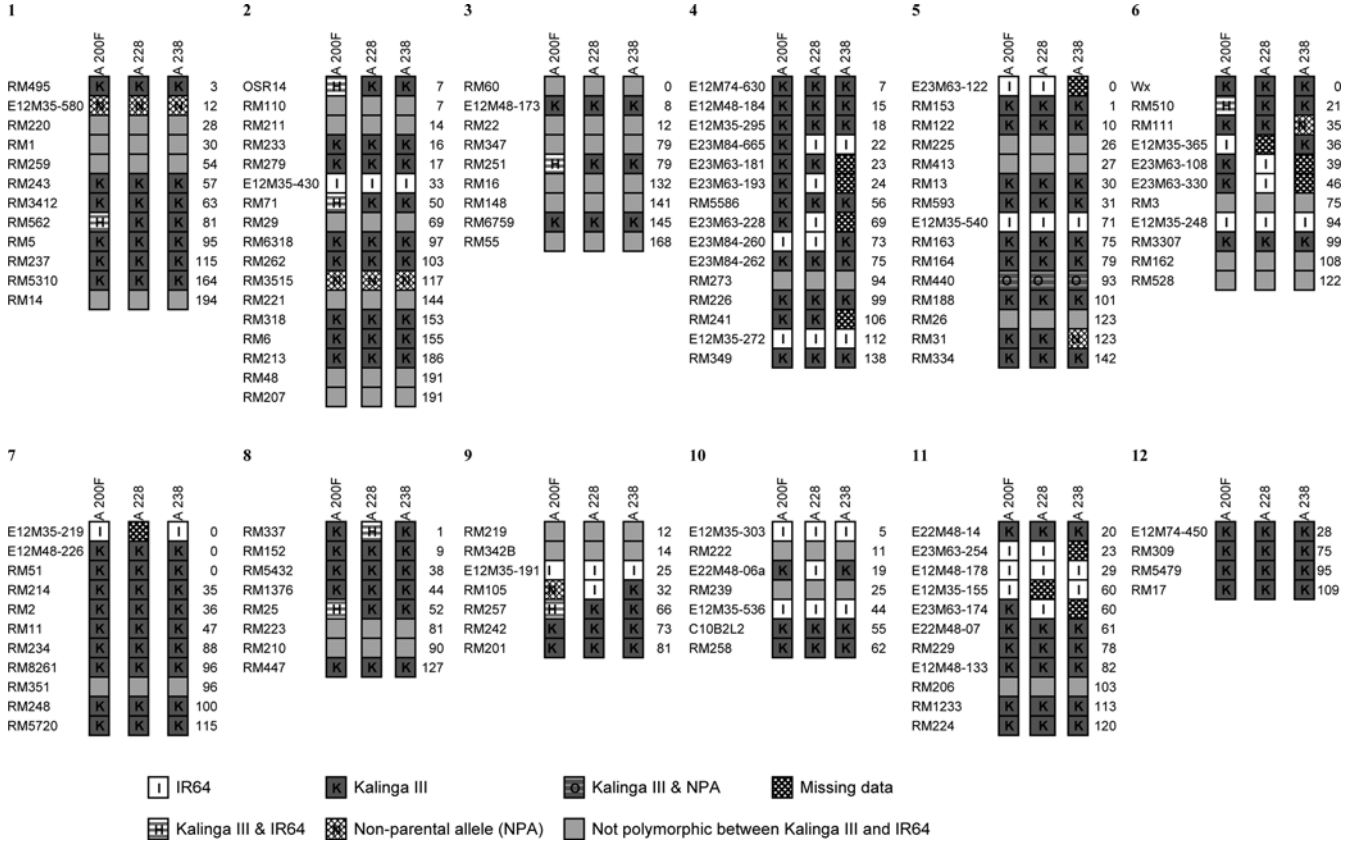

Fig. 3 Graphical genotypes of three upland bulks: Ashoka 200F (A 200F), Ashoka 228 (A 228) and Ashoka 238 (A 238) at 31 polymorphic AFLP loci (coded E_M_), 93 SSRs (coded RM_ and Waxy) and one SNP (C10B2L2). *Shaded blocks* indicate the following allelic classes: *K* Kalinga III homozygote (*dark shading*), *H* Kalinga III and IR64 (*light grey*), *I* IR64 homozygote (*unshaded*), *N* NPA homozygote (*cross-hatched*), *O* Kalinga III and NPA (*dark shading*). Missing data are highlighted with *very dark cross-hatching*, and non-polymorphic SSR loci are indicated with *light grey shading*. The approximate map positions (in centiMorgans) of SSR markers are from the Web site: http://www.gramene.org. The AFLP loci were mapped by Zhu et al. (1999) in two populations: IR20/63-83 and H359/8558

Discussion

Shifts in frequency between ecosystems

MES in a single cross identified genomic regions that are under strong selection. Sixteen loci had shifted in allele frequency significantly in the selected bulks when compared to the F_2 population, or when populations of bulks from contrasting ecosystems were compared. At eight of these loci, the shifts occurred in both comparisons. The shifts were greatest in the bulks that had, over time, proven to be the best performing in their ecosystem. Hence, in both Nepal and India, the upland bulks that were the best adapted to upland conditions, such as the released varieties Ashoka 200F and Ashoka 228, had the highest frequencies of Kalinga III alleles (over 95% of the 27 SSR alleles), while varieties that were the best adapted to lowland conditions (Barkhe 3004 and Ashoka 900F) had the highest frequencies of IR64 alleles (>70%). Bulks selected for intermediate conditions also had intermediate frequencies of alleles compared to the upland- or lowland-adapted bulks.

The two parents chosen for this cross were adapted to highly contrasting environments, and this was a crucial factor for the range of adaptation observed in the selected progeny. Breeders often use many different parents in their breeding programme, but parents are usually from a more narrow range of adaptation than the two we have chosen, and smaller populations for selection are more common. Although MES could be conducted to evaluate the products of a more conventional breeding programme, we expect the power to detect QTLs would be lower because random effects would be higher.

The very high selection pressures for the adapted parent genotype indicate that modified backcross breeding would be a highly appropriate method for breeding for specific environments, particularly when wide crosses are used. Backcrossing to a recurrent adapted parent will increase the frequency of genotypes having a high proportion of the genome from the adapted parent. This increases the effective population size as fewer genotypes, including those having even moderate amounts of unadapted genome, need to be eliminated from the population. Such a strategy would benefit the selection of varieties for adaptation to specific ecosystems either with or without MAS.

Recent field trials in Nepal have shown that Barkhe 3010 is extremely early for lowland environments, and it could be more suited to ecosystem 3. Likewise, Judi 523 has been dropped from advanced trials because it is too early for long-duration *Chaite*. The definition of ecosystems could be made more stringent so that future studies of MES can have greater precision because allele shifts would be even more significant.

Ecosystem-specific markers

Ecosystem-specific loci were identified. Alleles at RM51 and RM248 (both located on chromosome 7 but not linked) and in the region RM153–RM593 (chromosome 5) were more frequently IR64 in lowlands and more frequently Kalinga III in upland ecosystems. Genes linked to these alleles could contribute towards one or more traits determining adaptation to ecosystem.

Earliness is an important adaptive trait for ecosystem 1, while long duration is important in ecosystem 4, and this was reflected in an allele frequency difference of 94% at RM248 between these two ecosystems. Farmers in upland conditions preferentially selected for early maturity, and upland bulks lost this IR64 allele completely.

Some Kalinga III alleles were selected in all ecosystems

Not all shifts were in favour of the adapted genome for its respective environment. Kalinga III alleles were the most common class, present in all ecosystems. At RM279 (chromosome 2), they were detected at higher frequencies in the whole selected population compared with the F_2 population. Kalinga III alleles at such loci are likely to confer a significant improvement in traits important for adaptation across all ecosystems.

Some IR64 alleles may be selected in upland conditions

In the small proportion of the genome tested by the 27 SSR markers, no loci were found to have IR64 alleles favoured in upland conditions. No IR64 alleles were found in Ashoka 200F or Ashoka 228, and in other upland bulks there were few, none of them common across bulks. IR64 alleles were identified in three of the upland bulks when more loci were examined. There were eight genomic regions where IR64 alleles were present at AFLP loci in all three of the upland varieties examined, and hence these are possible regions where the IR64 genome is selected in upland conditions. A sample size larger than three varieties is required to draw definitive conclusions, and it is not yet known if these IR64 alleles are specifically selected in upland conditions, or whether they are selected across all ecosystems.

A factor influencing the high frequency of Kalinga III alleles detected with SSR markers may be that many of them were chosen because they map in regions found to control agronomic traits in QTL maps. The AFLPs used were completely random, and selection was apparently less strong for the alleles of the adapted parent at AFLP loci than at SSRs.

NPAs

The existence of NPAs is most likely to be due to genetic variation in the parents used for the original cross. They may not have been completely pure breeding, and this could explain the segregation in the progeny for some alleles not detected in the parent lines that we tested. The IR64 sample tested was heterozygous at 6 out of 54 SSR markers and at 3 of the 12 SNP loci tested (of these, only two SSR and one SNP were used in the MES of all the selected bulks). In a previous study, ten individual Kalinga III plants were tested at 25 SSR loci, and 20 % of loci were shown to be segregating between individuals (Bajracharya 2003). NPAs could also be derived from outcrossing in the farmers' fields, and the high level of NPAs in Barkhe 2012 and Barkhe 2014 (Fig. 1b) suggests that this occurred in these bulks. Some of these alleles may then have had a selective advantage over one, or both, of the parent alleles, possibly at RM163 and RM242 in medium-lowland conditions in Nepal and at RM11 across the whole of ecosystem 3. It is unlikely that the NPAs were due to new mutations.

Within-bulk variation

As expected, the three bulks ET (Judi 102), MD and LD had many (37–57%) loci still segregating for both parent alleles. The only other bulk with more than 50% segregating loci was Judi 111, which is bulk MT1, although this result could have been caused by the sampling error involved in testing only one plant. When cluster analysis was carried out without data for loci having a mixture of both parental alleles, a distinct cluster containing IR64 together with four of the bulks adapted to the lowland ecosystem (4) became clear.

Linking traits to markers

In the MES, significant shifts in allele frequency are most likely to result from selection for traits that contribute to overall performance and adaptation. Markers at loci that had significant shifts in frequency are useful for selection for overall performance. This performance can be either ecosystem specific (markers such as RM51 and RM248) or impart improved performance across ecosystems (e.g. the Kalinga III allele at RM279). It was possible to identify a highly heritable trait (days to anthesis) linked to an ecosystem-specific marker. The bulks that were fixed for the Kalinga III allele at RM248 were significantly earlier than bulks with an IR64 allele or a NPA; therefore, a recessive allele for earliness was donated by Kalinga III. There was much variation for maturity in the bulks where neither allele was fixed (Tables 1, 3). Grain yield is positively correlated with increased duration in favourable (non-water-limited) environments, but there is a negative association under drought stress. This association is apparent in the most stressed environment, ecosystem 1, where the Kalinga III allele at RM248 was most common. Selection pressure against Kalinga III alleles was also strong in ecosystem 4.

The numbers of days to anthesis were recorded in the glasshouse under long days so the phenotypes could be influenced by loci controlling photoperiod sensitivity. However, the data do agree with field observations. It should be noted that Lander and Botstein (1989) warned that single-marker regression is unreliable for selective genotyping when only the extreme progeny have been genotyped. However, the MES differed from selective genotyping because selection was replicated across a range of environments, and the locus was shown to differ significantly for allele frequency across environments in *G*-tests (Table 5). A QTL for heading date has previously been mapped to the same region of chromosome 7 in a population derived from the *japonica– indica* cross 'Nipponbare' × 'Kasalath' where the *japonica* allele increased the number of days to heading (Ishimaru et al. 2001). It is possible that IR64 shares this genomic region.

Do the results from QTL mapping and MES agree?

IR64 has been a common parent in several populations analysed for QTLs controlling many different, potentially agronomically important traits, including root traits, panicle number, plant height and grain quality.

Generally there is a positive correlation between improved root traits and performance under drought; genotypes with thicker and deeper roots have a higher biomass and grain yield under stress. Yadav et al. (1997) used a population derived from the cross IR64 × 'Azucena' and found that IR64 contributed alleles towards increased root weight at two QTLs on chromosome 5. The marker RM163 is in the region of a QTL where an IR64 allele increases the total root-weight mean by 7%. It might therefore be expected that RM163 would be linked to a genomic region from IR64 under positive selection in upland and medium upland ecosystems where drought stress is common. However, the Kalinga III allele was more frequent in ecosystem 1 than in the control population at this locus, showing that the QTL from IR64 was not selected in this cross.

IR64 alleles on chromosome 9 at RM201 have been shown to decrease maximum root length (Venuprasad et al. 2002). Eight of the nine upland varieties had the Kalinga III allele at RM201, so Kalinga III alleles could have been selected for at this QTL in uplands. RM201 appears to be under strong selection since IR64 alleles were less frequent in ecosystems 1 and 2 and more frequent in ecosystem 3. Although not significantly more common in ecosystem 4, it could be a valuable allele for irrigated ecosystems as both Barkhe 3004 (the best lowland selection from Nepal) and Ashoka 900F (the only lowland selection from India) have the IR64 allele at this locus.

QTL analysis for panicle number in the IR64 × 'Azucena' cross (Liao et al. 2001) showed that

IR64 contributed positively at three QTLs, including the regions containing RM349 (chromosome 4), which was tested on all bulks, and RM17 (chromosome 12), which was only tested on four bulks. The MES did not support the importance of this contribution from IR64 at RM349. At RM17 three upland bulks were fixed for Kalinga III (Fig. 3), whereas Ashoka 900F (lowland) had the IR64 allele (data not shown); this could be due to ecosystem selection for a QTL at this locus.

Farmers can easily select for plant height. At least 60 dwarfing genes have been identified in rice by classical genetic analysis, and they are expressed selectively at different growth stages and are likely to have important epistatic interactions. IR64 carries the *sd-1* gene on chromosome 1. However, no polymorphic marker near to the *sd-1* locus was tested in all of the bulks, the nearest (RM237) being 80 cM from it. The marker RM5310 is nearer to *sd-1*, and it was fixed for Kalinga III in the three upland varieties examined (all >110 cm), while Ashoka 900F (70 cm) was fixed for IR64 at this locus (data not shown).

There are a number of regions that have been identified as influencing traits involved with grain cooking and eating quality. The *Waxy* locus (chromosome 6) has been genetically well characterised and the SSR marker Waxy, located in the sequence coding for granule-bound starch synthase (Bligh et al. 1995), is polymorphic between IR64 and Kalinga III. The published *Waxy* allele has 11 CT repeats at this SSR, and our results show that Kalinga III also has 11 repeats, while IR64 has 17. The *Waxy* locus has a large contribution to grain quality which can potentially mask the effects of other grain quality QTLs in some crosses (Bao et al. 2002). Although there was little difference in grain quality between the two parents, there was a significant shift towards Kalinga III alleles at Waxy in bulks selected for ecosystem 1, and towards IR64 alleles in ecosystem 3. One possible explanation is linkage drag with another trait. It is also possible that grain quality is less important to farmers in upland ecosystems than it is to farmers in more lowland ecosystems.

Second-cycle breeding—genotypic ideotypes

Many genomic regions appear to have shifted in frequency through bulk-population selection, and the effect is repeated across genotypes of the same ecosystem. MES can be used to identify makers linked to traits of specific importance to farmers in particular ecosystems. These markers can then be used in MAS to improve existing lines by creating genetic ideotypes that should perform better in the target environment. From the limited number of loci tested, several backcrosses have been identified that will produce genetic ideotypes for specific ecosystems (Table 6). The bulks ET, MD and LD still contain valuable heterogeneity, and they can be subjected to further selection at important loci. The AFLP loci tested in the three upland varieties Ashoka 200F, Ashoka 228 and Ashoka 238 could be converted to a more repeatable marker system, such as STS, for use in MAS.

Future direction of MES

Rice genetics and breeding is entering a new era where genetic variation can be evaluated to the letter of the genetic code now that several versions of complete, or nearly complete, rice genome sequences exist. This has potential impacts on MAS and breeding. SNPs will be examined at the sequence level and, because of the availability of whole genome sequences, marker loci can be directly targeted to precise chromosome positions, using established genomic databases. SNPs that are polymorphic in specific crosses could be made available throughout the genome at any required marker density if new technologies such as genome partitioning (GB Patents ADP 109006) are used to reveal and validate SNP loci. Thus MES can be applied at the level of the entire genome. Because SNPs are always associated with sequence information, they are likely to become the most widely used marker type in future crop genetics and breeding programmes (Rafalski 2002)

The current MES study is still at an early stage because only a limited number of markers were used. The AFLPs used did not bear the sequence information with them, so they could not be accurately located into a genome sequence. The SNPs that were used in this study were originally detected in sequences from unrelated varieties to the parent varieties used in this study, and they showed a very low level of polymorphism in this cross. Although some genomic regions were identified to be important agronomically, it is likely that many important regions

Table 6 Putative target markers and recipient genotypes for genomic ideotype breeding

Recipient genotype to improve	Target loci	Chromosome	Method	Direction of selection	Target ecosystem
Judi 523	RM243	1	Backcross to Judi 567	IR64	2
	RM248	7			
	RM258	10			
Barkhe 3010	RM13- RM163	5	Backcross to Ashoka 900F	IR64	4
	RM234	7	Select from bulk	IR64	
	Waxy	6	Select from bulk	Kalinga III	
Judi 103	RM243	1	Backcross to Ashoka 228	Kalinga III	1
	RM51	7	Select from bulk	Kalinga III	

have been missed and will be revealed with more intense genomic coverage.

To develop MES further in this cross, we plan to test additional markers to increase the density of genome coverage. We will also evaluate the bulks selected from three other crosses with Kalinga III in farmer-oriented breeding programmes conducted at the same time as those on the Kalinga III × IR64 cross. This will allow comparisons to be made across genotypes to evaluate the effects of regions from Kalinga III in different genetic backgrounds. The MES could be extended to include more crosses and selection environments, with the resulting bulks evaluated at the same loci in order to build up a database of agronomically important genomic regions from different genotypes.

Conclusions

MES can detect agronomically important loci as shifts in marker allele frequency occurring as a result of selection. This strategy did not require prior assumptions as to which traits were important and did not involve the phenotyping of large populations of lines to determine the locations of QTLs. MES is potentially more powerful than conventional trait-based QTL analysis as it can measure the results of selection throughout the entire genome for all traits that have been under selection. The application phase of MES comes from a second cycle of breeding using MAS to develop lines that are genotypic ideotypes, with all the loci contributing to adaptation in a particular ecosystem.

Acknowledgements Without the genetic material produced from the rice breeding programmes in Nepal and India, the molecular studies reported here would not have been possible. For this work the authors are grateful to the many farmers who participated in the research and, for the PPB in Nepal, to Mr. Madhu Subedi, Ms. Sharmila Sunwar and Mr. Sanjaya Gyawali of LI-BIRD, Pokhara and, for the PPB in India, to Dr. Ravi Kumar and Dr. D.N. Singh, BAU, Ranchi; Mr. J.S. Gangwar and Dr. S.C. Prasad, GVT (East), Ranchi; and Dr. Daljit Virk, CAZS. The authors thank IRRI for the crosses and the supply of IR64, and Mr. Julian Bridges and Ms. Elly Rodriguez for technical assistance in CAZS. This document is an output from a project (Plant Sciences Research Programme R7434) funded by the UK Department for International Development (DFID) for the benefit of developing countries. The views expressed are not necessarily those of DFID.

References

Bajracharya J (2003) Genetic diversity study in landraces of rice (*Oryza sativa* L.) by agro-morphological characters and microsatellite DNA markers. PhD Thesis, University of Wales

Bao JS, Wu YR, Hu B, Wu P, Cui HR, Shu QY (2002) QTL for rice grain quality based on a DH population derived from parents with similar apparent amylose content. Euphytica 128:317–324

Bligh HFR, Till RI, Jones CA (1995) A microsatellite sequence closely linked to the *WAXY* gene of *Oryza sativa*. Euphytica 86:83–85

Ishimaru K, Yano M, Aoki N, Ono K, Hirose T, Lin SY, Monna L, Sasaki T, Ohsugi R (2001) Toward the mapping of physiological and agronomic characters on a rice function map: QTL analysis and comparison between QTLs and expressed sequence tags. Theor Appl Genet 102:793–800

Kumar R, Singh DN, Prasad SC, Gangwar JS, Virk DS, Witcombe JR (2001) Participatory Plant breeding in rice in eastern India. In: An exchange of experiences from South and Southeast Asia: proceedings of the international symposium of participatory plant breeding and participatory plant genetic resources enhancement. CGIAR Programme for Participatory Research and Gender Analysis (PGRA). Cali, Columbia, pp 233–236

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

Lebowitz RJ, Soller M, Beckmann JS (1987) Trait-based analyses for the detection of linkage between marker loci and quantitative trait loci in crosses between inbred lines. Theor Appl Genet 73:556–562

Liao CY, Wu P, Hu B, Yi KK (2001) Effects of genetic background and environment on QTLs and epistasis for rice (*Oryza sativa* L.) panicle number. Theor Appl Genet 103:104–111

Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci USA 88:9828–9832

Oetting WS, Lee HK, Flanders DJ, Wiesner GL, Sellers GL, King RA (1995) Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tailed primers. Genomics 30:450–458

Quarrie SA, Lazi-Janci V, Kovacevi D, Steed A, Peki S (1999) Bulk segregant analysis with molecular markers and its use for improving drought resistance in maize. J Exp Bot 50:1299–1306

Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94–100

Reyna N, Sneller CH (2001) Evaluation of marker-assisted introgression of yield QTL alleles into adapted soybean. Crop Sci 41:1317–1321

Shen L, Courtois B, McNally KL, Robin S, Li Z (2001) Evaluation of near-isogenic lines of rice introgressed with QTLs for root depth through marker-aided selection. Theor Appl Genet 103:427–437

Sokal RR, Rohlf FJ (1995) Biometry: the principles and practice of statistics in biological research, 3rd edn. Freeman, New York, pp 699–707

Stuber CW, Moll RH, Goodman MM, Schaffer HE, Weir BS (1980) Allozyme frequency changes associated with selection for increased grain yield in maize (*Zea mays* L.). Genetics 95:225–236

Venuprasad R, Shashidhar HE, Hittalmani S, Hemamalini GS (2002) Tagging quantitative trait loci associated with grain yield and root morphological traits in rice (*Oryza sativa* L.) under contrasting moisture regimes. Euphytica 128:293–300

Virk DS, Singh DN, Prasad SC, Gangwar JS, Witcombe JR (2003) Collaborative and consultative participatory plant breeding of rice for the rainfed uplands of eastern India. Euphytica 132:95–108

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

Witcombe JR, Subedi M, Joshi KD (2001) Towards a practical participatory plant-breeding strategy in predominantly self-pollinated crops. In: An exchange of experiences from South and Southeast Asia: proceedings of the international symposium of participatory plant breeding and participatory plant genetic resources enhancement. CGIAR Programme for Participatory Research and Gender Analysis (PGRA), Cali, Colombia pp 243–248

Yadav R, Courtois B, Huang N, McLaren G (1997) Mapping genes controlling root morphology and root distribution on a double-haploid population of rice. Theor Appl Genet 94:619–632

Zhu JH, Stephenson P, Laurie DA, Li W, Tang D, Gale MD (1999) Towards rice genome scanning by map-based AFLP fingerprinting. Mol Gen Genet 261:184–195