MOLECULAR ECOLOGY

Molecular Ecology (2010) **19**, 4892–4905 doi: 10.1111/j.1365-294X.2010.04845.x

Migration, isolation and hybridization in island crop populations: the case of Madagascar rice

KRISTIE A. MATHER,*[1] JEANMAIRE MOLINA,*[1] JONATHAN M. FLOWERS,* SAMARA RUBINSTEIN,* BRAD L. RAUH,† AMY LAWTON-RAUH,† ANA L. CAICEDO,‡ KENNETH L. McNALLY§ and MICHAEL D. PURUGGANAN*

*Department of Biology and Center for Genomics and Systems Biology, 1009 Silver, 100 Washington Square East, New York University, New York, NY 10003, USA, †Department of Genetics and Biochemistry, Clemson University, 100 Jordan Hall, Clemson, SC 29634-0318, USA, ‡Department of Biology, University of Massachusetts, Amherst, MA 01003, USA, §T.T. Chang Genetic Resources Center, International Rice Research Institute, DAPO Box 7777, Metro Manila, 1301 Philippines

Abstract

Understanding how crop species spread and are introduced to new areas provides insights into the nature of species range expansions. The domesticated species *Oryza sativa* or Asian rice is one of the key domesticated crop species in the world. The island of Madagascar off the coast of East Africa was one of the last major Old World areas of introduction of rice after the domestication of this crop species and before extensive historical global trade in this crop. Asian rice was introduced in Madagascar from India, the Malay Peninsula and Indonesia approximately 800–1400 years ago. Studies of domestication traits characteristic of the two independently domesticated Asian rice subspecies, *indica* and tropical *japonica*, suggest two major waves of migrations into Madagascar. A population genetic analysis of rice in Madagascar using sequence data from 53 gene fragments provided insights into the dynamics of island founder events during the expansion of a crop species' geographic range and introduction to novel agro-ecological environments. We observed a significant decrease in genetic diversity in rice from Madagascar when compared to those in Asia, likely the result of a bottleneck on the island. We also found a high frequency of a unique *indica* type in Madagascar that shows clear population differentiation from most of the sampled Asian landraces, as well as differential exchange of alleles between Asia and Madagascar populations of the tropical *japonica* subspecies. Finally, despite partial reproductive isolation between *japonica* and *indica*, there was evidence of *indica/japonica* recombination resulting from their hybridization on the island.

Keywords: domestication, hybrid, indica, isolated population, japonica, linkage disequilibrium, recombination

Received 12 May 2010; revision received 12 August 2010; accepted 16 August 2010

Introduction

Domestication is a key evolutionary transition of many plant species in response to cultivation and can be regarded as a unique form of plant/animal mutualism. The origins of domesticated species and agriculture have occurred numerous times, with humans (Diamond 2002), ants (Schultz & Brady 2008) and beetles (Farrell *et al.* 2001) evolving cultivation as a new behavioural response to enhancing food supplies. In humans, domesticated plant species appear to have evolved in 24 distinct geographic regions in both the Old and New World, and several of these have subsequently spread outside of their centres of origins into new agro-ecological environments (Purugganan & Fuller 2009). Human migrations and adoption by trade spurred the introduction of crop

Correspondence: Michael Purugganan, Fax: 212 995 4986; E-mail: mp132@nyu.edu

[1]These authors have contributed equally to this work.

species to new geographic areas, while cultural, climatic and other ecological constraints have limited domesticated species migrations.

Understanding how crop species spread and are introduced to new areas may provide additional insights into the nature of species range expansions. The spread of domesticated species also possesses certain features that are fairly unique, including; (i) the role of humans in species introductions and adaptations, (ii) the recency of migrations, since domestication of most species occurred during the Neolithic Period <10 000 years ago; and (iii) the availability of archaeological, historical and ethnographic data on species spread. Founder events as a result of domesticated species introductions, and in some cases subsequent isolation, also provide an opportunity to develop new mapping populations to identify genes underlying organismal variation and adaptation (Varilo & Peltonen 2004).

The domesticated species *Oryza sativa*, or Asian rice, is one of the key domesticated crop species in the world (Khush 1997) and is emerging as a model for studying the origins and spread of domesticated taxa as well as in mapping genes associated with plant variation (Ashikari & Matsuoka 2006). At least two major domestication events of Asian rice have been identified, which resulted in the establishment of two major variety groups (Londo *et al.* 2006). *Japonica* varieties were initially domesticated in South China and spread northward, giving rise to the *temperate* varieties, as well as southward to Southeast Asia and from there to West Africa and Brazil, giving rise to the tropical varieties (Oka 1988; Khush 1997; Garris *et al.* 2005; Londo *et al.* 2006). The *indica* varieties were initially domesticated in the foothills of the Himalayas in Eastern India and spread primarily to the lowland environments throughout the Asian tropics and subtropics. From India, *indica* rice also travelled to Madagascar and East Africa and then to countries in West Africa (Khush 1997). These two major variety groups are derived from differentiated gene pools in the wild ancestor *O. rufipogon*, and there is some reproductive isolation between *indica* and tropical *japonica* that results in sterility of their hybrids (Li *et al.* 1997; Harushima *et al.* 2002).

The island of Madagascar off the coast of East Africa, which was settled by humans only 2000 years ago (Burney *et al.* 2004), was one of the last major Old World areas of introduction after the domestication of rice and before extensive historical global trade in this crop. Asian rice was introduced into Madagascar from India, the Malay Peninsula and Indonesia (Ahmadi *et al.* 1988, 1991; Rabary *et al.* 1989), possibly at the height of ancient trading in the Indian Ocean approximately 800–1400 years ago. Studies of domestication traits characteristic of the two independently domesticated Asian rice subspecies, *indica* and tropical *japonica*, suggest two major waves of migrations into Madagascar (Koji 1997).

The introduction of rice and rice cultures to the island of Madagascar can be initially traced to the westward expansion of Austronesian speaking peoples, specifically the Malays and Javanese from Southeast Asia, between 400 and 1000 A.D. (Vaughan *et al.* 2004), whose cultural practices are found primarily on the east coast of Madagascar. Tropical *japonica* varieties are believed to have been introduced first by these peoples from the Malay archipelago in the 5th and 6th century by the Malays migrating from Indonesia (Dewar & Wright 1993; Khush 1997). *Indica* varieties are believed to have been introduced to Madagascar in a subsequent migration from the Indian subcontinent, the timing of which is not well understood. This second migration of rice into Madagascar introduced *indica* to a wider range of geographical areas and resulted in hybridization with the tropical *japonica* varieties previously brought to Madagascar by the Malay (Vaughan *et al.* 2004). Historical records suggest only limited introductions of rice into Madagascar, which together with their geographic isolation in the western Indian Ocean prevented gene flow with other Asian rice landraces.

Morphological and early genetic studies have found many *indica* and tropical *japonica,* with some temperate *japonica* landraces, in Madagascar as a result of these island introductions. Allozyme analysis indicates lower diversity of Madagascar rice landraces than from other Asian locales (Ahmadi *et al.* 1991), which suggests a genetic bottleneck associated with possible founder effects of island introduction, whereby some minor alleles present in Asian rice were lost in the Madagascar varieties. At the same time, other alleles less frequent in Asia appeared at a much higher frequency in Madagascar, possibly because of genetic drift or selection on the island (Ahmadi *et al.* 1991). A general maintenance of *indica/japonica* differentiation is also observed, but there is evidence of hybridization and recombination between *indica* and tropical *japonica* in the island. A unique rice group specific to the island and found in the high plateau region at altitudes of 1000–1500 m has been described (Ahmadi *et al.* 1991), and morphological data suggest it is intermediate between *indica* and tropical *japonica* for most traits; it has thus been characterized as a hybrid resulting from *indica/japonica* recombination and selection.

Here, we report on a population genetic analysis of rice in Madagascar to study the dynamics of island founder events during the expansion of a crop species' geographic range and introduction to novel agro-ecological environments. We observe a significant decrease in genetic diversity in rice from Madagascar

when compared to those in Asia, likely the result of a bottleneck on the island. We also find a high frequency of a unique *indica* type in Madagascar that shows clear population differentiation from most of the sampled Asian landraces. Finally, we provide evidence of *indica/japonica* recombination on the island that results from limited hybridization.

Materials and methods

Plant material

We used a sample set composed of 45 *O. sativa* landraces from Madagascar (*indica* and tropical *japonica*, subsequently referred to as Madagascar *indica* (MI) and Madagascar tropical *japonica* (MTJ), respectively) and compared them to 39 Asian accessions of *indica* (21, abbreviated AI) and tropical *japonica* (18, ATJ) that we already had sequences for from previous studies (Caicedo *et al.* 2007; Mather *et al.* 2007; Supplement Table S1). The list of Madagascar landraces, their accession numbers and origins are in Supplementary Table S2. These represent random samples with the most complete locality information taken from Madagascar accessions that are maintained at the International Rice Genebank and were collected in collaboration with the Centre National de la Recherche Appliquée au Développment Rural of Madagascar. Landrace material from Madagascar was chosen to represent traditional varieties cultivated in the island representing various ecological types (irrigated, rainfed-lowland, and upland, etc.; Supplementary Table S2), and more recent elite cultivars were not used so as to limit the analysis to strains that could plausibly have descended from the early introductions to the island. Duplicate variety names were avoided. *Oryza meridionalis,* the most divergent among the AA genome species, which also includes *O. sativa*, was also included to provide an outgroup. We also included in our analysis the whole-genome sequence of Nipponbare, a temperate *japonica* variety (International Rice Genome Sequencing Project 2005).

Sequenced gene fragments

Variation was assessed with a sequence tag approach that focused on four genomic regions, each spanning approximately 500 kb in length. The fragments examined here were previously sequenced in *indica, japonica* and *O. rufipogon* accessions for the purpose of assessing the decay of linkage disequilibrium in rice (Mather *et al.* 2007). These same fragments were re-sequenced in Madagascar varieties in this study primarily for the purpose of comparison with this data set. The physical locations of the sequenced fragments included three regions on chromosome 1, referred to as genomic regions E, B and F, and one on chromosome 4, which we refer to as region D (Mather *et al.* 2007). In regions E and B, twelve ~500-bp gene fragments spaced ~40 kb apart were sequenced to provide coverage across the 500-kb genomic regions (Mather *et al.* 2007). In genomic region F and genomic region D, seven ~500-bp fragments spaced ~80 kb apart were sequenced. The majority of these gene fragments were previously sequenced in Asian rice (Mather *et al.* 2007; Genbank nos. EU225527- EU231597; EU227898- EU227993). Fifteen gene fragments throughout from the genome were chosen from the 111 STS markers used in a previous study (Caicedo *et al.* 2007). Sequence data from the current study have been deposited in GenBank under accession nos. HQ015746- HQ018039.

Primers for amplification and sequencing were designed using Primer 3 (Rozen & Skaletsky 2000). Amplification and sequencing reactions were conducted by Cogenics (Houston, Texas) as previously described (Olsen *et al.* 2006; Caicedo *et al.* 2007).

Population structuring and genetic relationships

STRUCTURE 2.3 (Pritchard *et al.* 2000) was used to determine the genetic affinities of Madagascar accessions (33 MI, 12 MTJ) and to assign individuals to variety groups for downstream population genetic analyses. STRUCTURE runs were conducted with the focal accessions of *indica* and tropical *japonica* from Madagascar and Asia (84 total) along with additional Asian *O. sativa* including six aromatics, six *aus*, 21 temperate *japonica* and 21 members of the wild progenitor to Asian rice, *O. rufipogon,* which were sequenced for an overlapping set of fragments in a previous study (Caicedo *et al.* 2007). In the STRUCTURE analysis, we included the 15 STS markers and 38 gene fragments from the targeted genomic regions.

STRUCTURE runs were executed by treating accessions as haploid (Caicedo *et al.* 2007). SNPs in the analysis were limited to the site with the highest minor allele frequency in each fragment to minimize the effects of linkage. STRUCTURE was run five times for each *K* value with an admixture model with correlated allele frequencies (Falush *et al.* 2003) and a burn in of 50 000 iterations followed by a run length of 100 000 iterations. Physical map positions were entered to explicitly take linkage into account, although runs with and without map positions yielded similar results.

Individuals were assigned to a cluster if the STRUCTURE membership coefficient was at least 60%. While STRUCTURE does not take inbreeding into account, our previous work on rice (Caicedo *et al.* 2007)

demonstrated that this approach can recover the known genetic structure of *O. sativa*. However, to verify our groupings, we also ran Instruct (Gao *et al.* 2007), which, like STRUCTURE, assigns individuals to populations using a Bayesian clustering method and infers the optimal number of subpopulations that have the lowest Deviance Information Criterion (DIC). Unlike STRUCTURE, Instruct does not assume Hardy–Weinberg equilibrium and takes into account inbreeding, which is predominant in rice. For Instruct runs, we ran two chains with 1 000 000 iterations each including a burn in of 500 000 for $K = 1$ to $K = 10$ with mode (-v) set to 2 to infer population structure and selfing rates simultaneously.

Clustering analyses according to genetic similarity were conducted in MEGA4 (Tamura *et al.* 2007). We inferred the evolutionary relationships of *indica* varieties and tropical *japonica* varieties from Asia and Madagascar using the neighbour-joining method (Saitou & Nei 1987) with 500 bootstrap replicates on the concatenated data set. The evolutionary distances were computed as the number of base substitutions per site using the maximum composite likelihood method (Tamura *et al.* 2004). All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons. Among site rate variation was modelled with a gamma distribution with $\alpha = 1.0$ and heterozygous sites resolved by randomly selecting an allele.

Diversity analyses

A perl script was written to map the annotation of the Rice Genome Annotation Project v.6.0 release of the Nipponbare genome (ftp://ftp.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_6.0/) onto the sequenced fragments. Population genetic statistics were then calculated with the libsequence package (Thornton 2003), and diversity measures obtained independently for silent, synonymous and nonsynonymous sites. However, given the small number of nonsynonymous SNPs (28 out of 224 total SNPs) in the data set, subsequent analysis and interpretation are based on analysis of all sites. Statistics were calculated separately for AI and ATJ, and post-hoc partitions of MI varieties and MTJ varieties based on our STRUCTURE analysis. For the Madagascar accessions, nucleotide diversity measures were calculated separately for tropical *japonica* and *indica* after removing accessions inferred to be recombinant. Heterozygotes are rare in both Madagascar and Asian domesticates but were resolved by randomly selecting one of the two alleles. Heterozygotes were treated in this fashion for all analyses unless stated otherwise.

Differences in population summary statistics (i.e. π, θ_W, Tajima's D) between Asia and Madagascar were determined by nonparametric paired Mann–Whitney tests in R (http://www.R-project.org). Differences in diversity between these regions were also tested using a permutation approach, in which 10 000 data sets were created by randomly assigning accessions to each of two populations and determining the difference in nucleotide diversity between the populations. Significance was assessed by comparing the observed difference between Asia and Madagascar to the permutated distribution.

Linkage disequilibrium and recombination

Recombinant genotypes between *indica* and *japonica* have previously been reported in Madagascar (Ahmadi *et al.* 1991). To identify probable recombinant genotypes in our Madagascar sample, we used 55 SNPs that were fixed for alternate alleles between *indica* and *tropical japonica* in Asia. We considered the allelic configurations at these 55 sites to represent 'parental' genotypes, and accessions from Madagascar with 'nonparental' allelic configurations were inferred to be recombinant. This approach likely underestimates the number of recombinant genotypes as recombination events can only be assessed between sites that are fixed between subspecies in Asia.

Linkage disequilibrium (LD) was calculated as the correlation coefficient r^2 between each single nucleotide polymorphism (SNP) pair within and between chromosomes (Hartl & Clark 1997). Only SNPs of at least 10% frequency in each group were considered. Both tropical *japonica* groups, from Asia and Madagascar, did not meet this criterion, because of low polymorphism levels; we therefore focused our LD analyses primarily on *indica* from Asia and Madagascar (AI, MI). In calculations of r^2, heterozygous accessions were retained if only one SNP in the pair was a heterozygous because it was possible to infer gametic phase. However, if both SNPs in a pair were heterozygous, the pair was excluded because gametic phase is unknown (Mather *et al.* 2007). Inferred recombinant accessions in Madagascar were omitted from the LD analysis.

Analysis of molecular variance (AMOVA)

Analysis of molecular variance was performed in Arlequin ver. 3.5 (Excoffier *et al.* 2005) to assess the significance of population substructure between Madagascar and Asian varieties of *indica* and *japonica*. Because *indica* and tropical *japonica* have repeatedly been shown to be genetically differentiated here and elsewhere (Garris *et al.* 2005; Caicedo *et al.* 2007), we conducted a hierarchical analysis with *indica* and *japonica* as groups and Asian and Madagascar samples of each cultivar as pop-

ulations within groups. An additional set of analyses were conducted on *japonica* and *indica* separately. All AMOVA analyses were run with a distance matrix generated with the Kimura-2-parameter model (Kimura 1980) and the input data treated as haplotypic. In addition, all inferred recombinants between *japonica* and *indica* in Madagascar were removed prior to analysis.

Gene flow analysis

Pairwise nonlinked isolation-with-migration models were tested for best fit for all population combinations using the Clemson University Condor Computing Cluster (Hey & Nielsen 2007; Lawton-Rauh, von Oehsen, Rauh and Duffy, submitted). The co-estimated parameters for the best fit models of each population pair were current effective and ancestral population sizes, nucleotide substitution rates and time of population splitting (coalescent time to most recent common ancestor). Identical sequence alignments for 10 core STS loci were edited for length and longest nonrecombining block. All insertion–deletion polymorphisms (indels) were removed to avoid confounded coalescence issues, and all loci were then combined into a composite multilocus file. All pairwise comparisons utilized the exact same alignments across all 10 core STS loci following this editing. The lengths of each STS locus ranged from 316 to 549 bps. For this analysis, we divided the data into the Madagascar and Asian *indica* (MI and AI) and tropical *japonica* (MTJ and ATJ) groups. Depending on sequence quality, the number of samples utilized per group varied.

For these analyses, the Hasegawa-Kishino-Yano (HKY) mutation model was assumed with an inheritance scalar of 1. As we do not have a mutation rate estimate for rice, we used a rate of 3×10^{-8} obtained from maize (Clark *et al.* 2005). As this estimate is based upon analysis of a single gene rather than a genome-wide estimate, we used an initial range of 0.1–15×10^{-10} for all loci to test for sensitivity of estimates to minor mutation rate differences across loci. A 100 000 step burn-in period was employed, and five Markov chains with 10 chain swap attempts per step were utilized in every model testing run and re-iteration. The random number seed was derived from a 'C' rand function, with 49999951 steps in the chain following the run before termination. Alternative models of isolation and migration were optimized using all possible iterations of input prior ranges for each parameter, resulting in 1779 total simulations. The best fit set input range was determined by the effective sample size estimate, and the stability of the posterior distribution curve was then re-run five times for each pairwise comparison to ensure estimate stability using approximately 10 simultaneous runs. Comparisons of different isolation-migration models were then performed using likelihood ratio tests, using models that allowed fluctuations in migration rate vs. models setting migration to zero. Together, these isolation-migration models (IMa) analyses utilized approximately 168-K core hours of simulation time on the web-based Condor computational system interface system that we designed and implemented (von Oehsen, Duffy, Rauh and Lawton-Rauh, submitted).

Results

Population structure of rice in Madagascar

We sequenced a total of ~25.7 kb of sequence across 53 gene fragments, 38 of which are found within four ~500-kb genomic regions on chromosomes 1 and 4. The other 15 are referred to as STS markers and scattered across the rice genome. The 45 Madagascar landraces we sequenced included accessions that are described as either tropical *japonica* or *indica*, which is representative of the rice diversity on the island. In discussing these island landraces, we will use these variety designations although our analysis indicates that some of these may represent hybrids.

We used a model-based Bayesian clustering approach implemented in STRUCTURE (Pritchard *et al.* 2000) to examine population stratification within Madagascar and determine the relationships between the landraces on this island with other Asian *O. sativa* varieties and the wild rice ancestor *O. rufipogon*. Using the Evanno *et al.* (2005) method to detect the correct number of clusters (*K*), we found that delta *K* was highest between $K = 1$ and $K = 2$, which simply groups the accessions into predominantly *indica* and *japonica* groups, with *rufipogon* accessions showing an approximately equal proportion of their alleles from the two main cultivars (Fig. S1, Supporting information). The second highest delta K was obtained when accessions were classified into three clusters: *indica, japonica* and *rufipogon*. However, plotting the mean log probability of the data, we find that it plateaus at $K = 6$ or 7. This largely supports previous STRUCTURE analysis (Caicedo *et al.* 2007), and we conclude that the data are consistent with 6 or 7 clusters, but we caution that this is a relatively small amount of data (53 gene fragments, one of which was invariant in this data set) and can only identify large-scale differentiation.

Considering the STRUCTURE runs with six clusters (see Fig. S1, Table S2, Supporting information), we find that each cluster generally corresponds to the variety groups/species determined previously (Caicedo *et al.* 2007). However, the *indica* accessions from Asia and

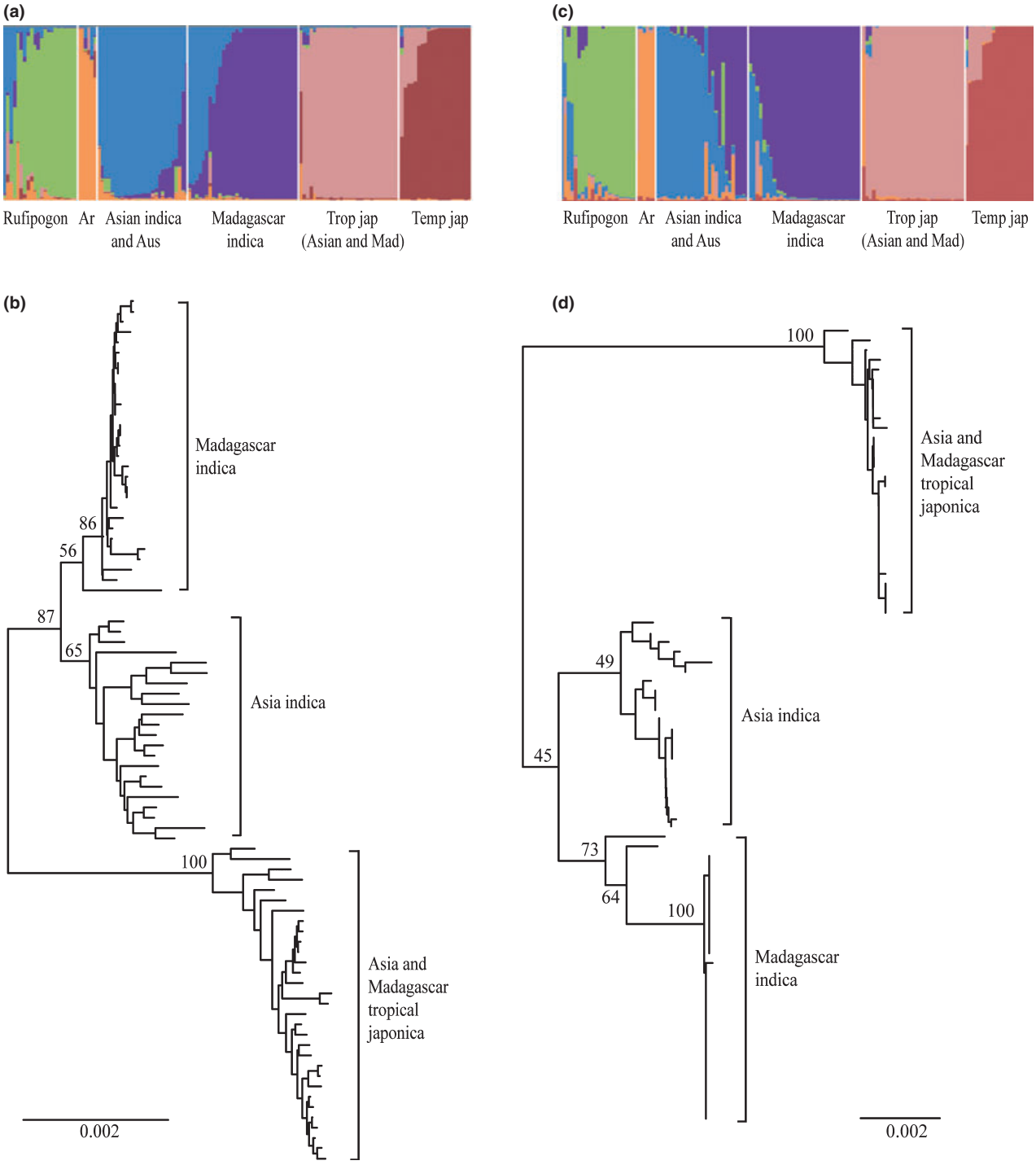

Fig. 1 Population structuring and neighbour-joining analyses of Madagascar and Asian rice landraces. (a) The inferred population structuring for all accessions based on data from all regions. STRUCTURE with $K = 6$; (b) NJ tree for concatenated data; only relationships among Asian and Madagascar *indica* and tropical *japonica* accessions are shown. The major clades and their bootstrap support are indicated. Bootstrap support was conducted using 500 replicates of the data; (c) STRUCTURE output *without* region B SNPs; (d) NJ analysis using data from genomic region B only, in which the major clades and their bootstrap support are indicated using data only from genomic region B. Only relationships among Asian and Madagascar *indica* and tropical *japonica* accessions are shown. Bootstrap support was conducted using 500 replicates of the data. Ar, aromatics; Mad, Madagascar; Trop jap, tropical *japonica*; Temp jap, temperate *japonica*.

most of those from Madagascar form separate clusters in all $K = 6$ runs. The appearance of two distinct *indica* clusters begins in runs starting at $K = 4$, whereas tropical and temperate *japonicas* do not appear as separate populations until $K = 6$ (Fig. 1; Fig. S1, Supporting information). Moreover, the MTJ landraces clustered with ATJ and could not be separated in our analyses. In the Instruct analyses, which take into account inbreeding, the mean log likelihood for the two chains starts to plateau at $K = 8$ but recovers similar groupings as STRUCTURE, except that *O. rufipogon* splits into two subpopulations corresponding to accessions from China and mostly Nepal, and Aus also becomes distinct from Asian *indica* (results not shown). The lowest DIC was recovered with $K = 9$, which just further splits the Chinese *rufipogon* cluster into a heterogeneous group; though, this DIC was only marginally higher (i.e. 8 units) from that of $K = 8$.

Clustering analysis using the neighbour-joining method also supports the STRUCTURE analysis (see Fig. 1b). In the neighbour-joining analysis using the concatenated data set including only island and continental tropical *japonica* and *indica* varieties, we find that the MTJ are interdigitated among the Asian landraces. The MI varieties, however, show two types of *indica* on the island. A minor fraction, comprising 3 of the 31 Madagascar *indica* varieties, are found scattered among the Asian *indica* landraces. The majority of island *indica* varieties, however, form a distinct clade with high bootstrap support (see Fig. 1b).

An alternate set of tests for population structure were conducted with AMOVA. A hierarchical analysis with *indica* and *japonica* as groups explained 71.6% of the genetic structure in the data but is not significant, whereas Asian and Madagascar samples as populations within groups significantly accounted for 13.1% of the genetic structure ($P < 0.01$, Table 1). Additional tests for geographic differentiation between Madagascar and Asia conducted separately for *indica* and *japonica* indicated significant geographic structure in both groups ($P < 0.01$). AMOVA of each cultivar separately revealed that the Asia-Madagascar split explained almost 50% ($P < 0.01$) of the genetic variance in *indica*, but only accounted for 18.3% ($P < 0.01$) of the genetic structure in tropical *japonica*. Analysis of *indica* and tropical *japonica* separately for each chromosomal region, indicated that region B is the most differentiated between Madagascar and Asia in *indica* (among population variance = 67.7%, $P < 0.01$), while region F is the most differentiated in tropical *japonica* (35.8%, $P < 0.01$).

Reduced levels and patterns of nucleotide variation in island populations

The mean nucleotide diversity level for MTJ was lower than that observed in ATJ ($\pi = 0.0006$ vs. 0.0010) [see

Table 1 Analysis of Molecular Variance (AMOVA) for each genomic region and for all data

Data type	Partitions	Number of groups	Variance components %		
			Within populations	Among populations within groups	Among groups
All data	*indica* populations vs. trop. *japonica* populations	2	15.29**	13.14**	71.56
All data	AI vs. MI	1	50.55	49.45**	—
All data	ATJ vs. MTJ	1	81.75	18.25**	—
Chr 1-E	AI vs. MI	1	75.73	24.27**	—
	ATJ vs. MTJ	1	85.22	14.78*	—
Chr 1-B	AI vs. MI	1	32.31	67.69**	—
	ATJ vs. MTJ	1	84.88	15.12**	—
Chr 1-F	AI vs. MI	1	83.81	16.19**	—
	ATJ vs. MTJ	1	64.22	35.78**	—
Chr 4	AI vs. MI	1	95.85	4.15	—
	ATJ vs. MTJ	1	95.24	4.76	—
STS	AI vs. MI	1	61.39	38.61**	—
	ATJ vs. MTJ	1	92.79	7.21	—

There was only one group in all analyses except in the first, where we conducted a hierarchical analysis with *indica* and *japonica* as groups and Asian and Madagascar samples of each cultivar as populations within groups. In subsequent analyses, *indica* and tropical *japonica* were analysed separately. The significance of the covariance components associated with the different possible levels of genetic structure (within individuals, within populations, within groups of populations, among groups) is tested using nonparametric permutation procedures as described in the Arlequin documentation (Excoffier *et al.* 2005). P values: **≤0.01, *$0.01 < P < 0.05$

Table 2 Mean summary statistics for nucleotide variation

Group	*n*	θ_W	π	TajD
Indica				
Asia	21	0.0022	0.0020	−0.2712
Madagascar				
R	32	0.0015**	0.0009**	−0.6347
NR	30	0.0012**	0.0008**	−0.5290
Tropical *japonica*				
Asia	18	0.0013	0.0010	−0.7061
Madagascar				
R	11	0.0008	0.0006	−0.446
NR	7	0.0002**	0.0002**	0.3373†

†Based on the mean of 7 values as TajD cannot be calculated for many fragments because of the lack of SNPs. Note: Asterisks (**) indicate diversity measure is significantly different ($P < 0.01$) between the Madagascar population and its corresponding Asian population using a Mann–Whitney test (see Methods). Diversity measures were computed for Madagascar populations with recombinants (R) and without recombinants (NR).

Table 2]. This difference, however, was not statistically significant (Mann–Whitney Test, $P > 0.05$). The pattern of reduced variation in Madagascar was also observed in *indica*, where mean nucleotide diversity levels for the island landraces were less than half of those in Asia ($\pi = 0.0020$ vs. 0.0009), a difference that was statistically significant (Paired Mann–Whitney Test, $P < 0.05$). This reduction in diversity in Madagascar rice is expected, given the bottleneck associated with founder events during the establishment of Madagascar rice; this reduction in nucleotide diversity can be used to estimate the size of the bottleneck on island migration.

Unusual haplotype structure in genomic region B of Madagascar indica

Examining the data from the genomic regions and STS gene fragments, the AMOVA indicated that differentiation between AI and MI is most pronounced in genomic region B on chromosome 1. This region has more high-frequency SNPs than any other region in our analysis, and the overall nucleotide diversity in this region ($\pi = 0.0016$) is 4- to 8-fold higher than those of the other three genomic regions ($\pi = 0.0003$–0.0010; Table S3, Supporting information).

To examine whether genomic region B is solely responsible for the observed population structuring, we conducted STRUCTURE analysis of the other three genomic regions and the STS data without SNPs from region B. This analysis still shows the separate and distinct Madagascar group (see Fig. 1c), indicating that the differentiation of this group is not driven solely by variation in genomic region B.

The presence of several SNPs in region B in near-complete linkage disequilibrium results in a group of haplotypes that are near identical to one another and found only in the distinctive Madagascar *indica* group. Neighbour-joining analysis using *only* data from genomic region B recovers the two distinct *indica* clades in Madagascar (see Fig. 1d). Data from the three other genomic regions and the STS data, however, also show a close relationship between many of the landraces identified as belonging to the Madagascar *indica* group, albeit with less differentiation from (Asian) *indica* landraces (results not shown).

Geographic patterns in island and continental rice

The Madagascar landraces used in our sample appear to be distributed in two different parts of the island (see Fig. 2) with the upper cluster of accessions mostly collected in upland altitudes (>1000 m), whereas the bottom cluster collected in lowland altitudes. The distinctive MI clade observed in the neighbour-joining analysis and supported as a unique group in the STRUCTURE analysis is found in the upper cluster and contains some AI varieties. A map of the geographic origins of these landraces (see Fig. 2) shows that these *indica* accessions are found in the Indian subcontinent.

Migration within and between Asia and Madagascar populations

We examined migration between pairwise groups (AI, ATJ, MI, MTJ) using IMa (Hey & Nielsen 2007). Joint point estimates of the immigration rate ($M_1 = 2N_1m_1$), relative pairwise population migration rate (M_2/M_1), contemporary and ancestral effective population sizes (N_1, N_2, N_A) and divergence times ($T = t/2N$ in generations) inferred with multilocus IMa are given in Table 3 and Fig. 3, along with 95% credibility intervals per population pair. The point estimates for each parameter are the bins with the highest estimate after curve smoothing.

The relative allele migration rate between each population pair, M_2/M_1, indicates the extent of allele introgression between groups. M_1 and M_2 are the estimated effective number of migrants per generation for each direction. Because these are pairwise population estimates, the relevant parameter for comparison is M_2/M_1, which is the relative population migration rate between each population pair, comparing immigration into Pop 1 from Pop 2 vs. immigration into Pop 2 from Pop 1. The relative migration rates are illustrated in Fig 3. The greatest skew in gene flow is between Asian and Madagascar tropical *japonica*, ($M_2/M_1 = 59.3$), with the direc-

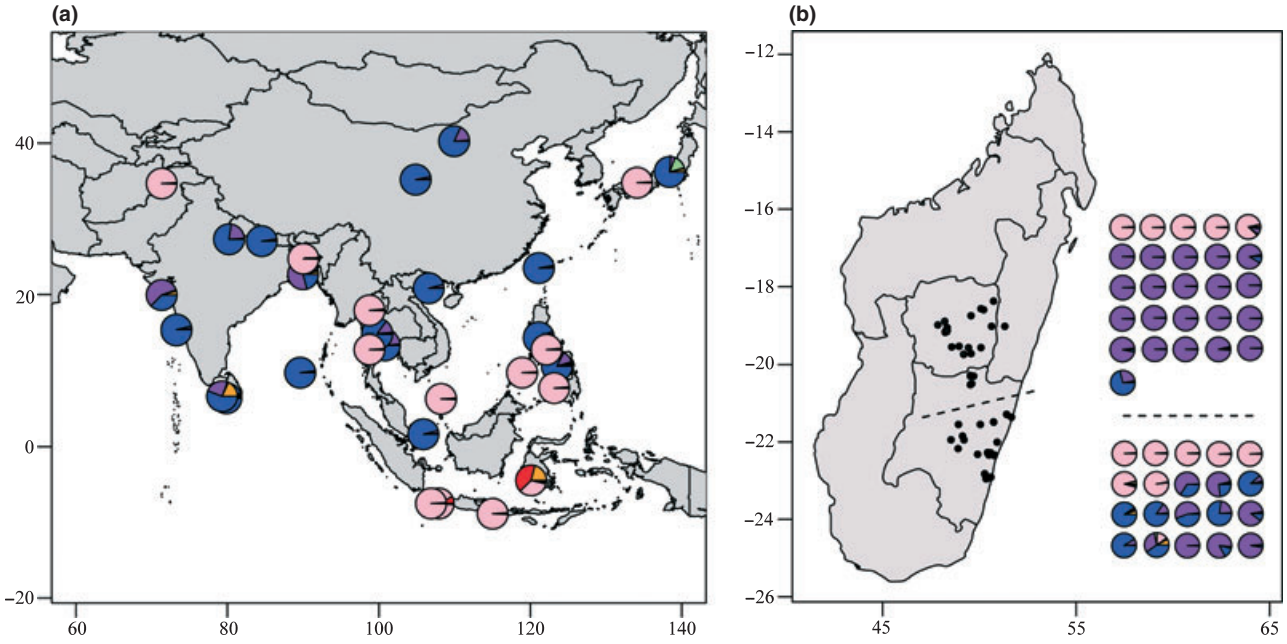

Fig. 2 Geographic distributions of (a) Asian and (b) Madagascar landraces. The locations of rice accessions are shown. The colours in the pie charts indicate membership in a specific population cluster with the colours corresponding to populations assigned in Fig. 1. In the Madagascar map (b), accessions are represented by black dots with their corresponding pie charts indicated to the right of the map. The upper cluster of accessions was mostly collected in upland altitudes (>1000 m) and separated by a dashed line from the bottom cluster of accessions, which were collected in lowland altitudes.

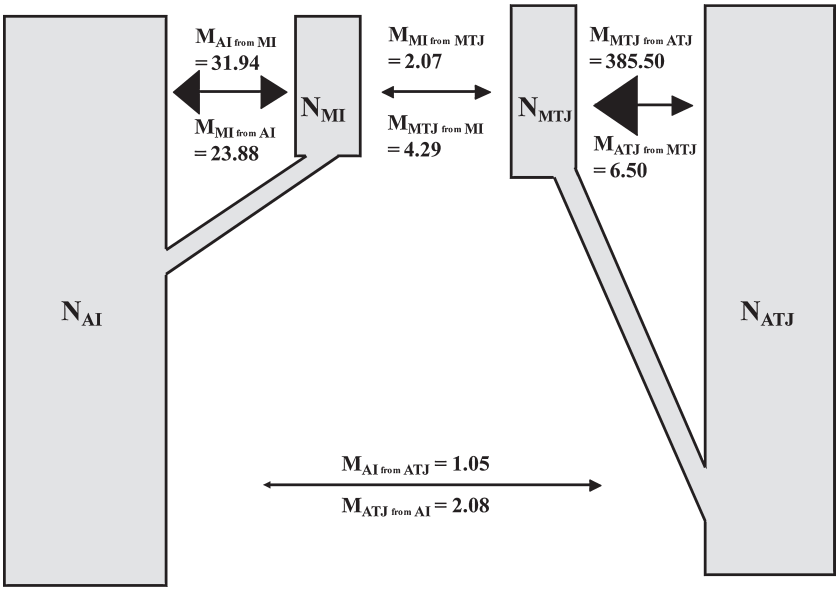

Fig. 3 Allele immigration rates between populations. The rates are for pairwise comparison among Asia *indica* (AI), Asia tropical *japonica* (ATJ), Madagascar *indica* (MI) and Madagascar tropical *japonica* (MTJ). The sizes of the arrows reflect relative migration rates.

tion of flow strongly biased from the continent to the island. In contrast, the gene flow estimates suggest a significant amount of back-migration from Madagascar to Asia in *indica* rice, which may result from continued contact between this island and South Asia. Moreover, there appears to be greater gene flow into tropical *japonica* from *indica* within Madagascar (M_2/M_1 = 2.076). It should be noted, however, that only in the case of tropical *japonica* between Asia and Madagascar is there a substantial bias in the direction of gene flow.

Estimated effective population sizes are listed in columns N_1, N_2, and N_A and refer to contemporary populations ('Pop 1' and 'Pop 2') in the relevant comparison followed by the ancestral population from which both

Table 3 Demographic modeling of rice

Pop 1	Pop 2	$M_1 = 2N_1m_1$	$M_2 = 2N_2m_2$	M_2/M_1	N_1	N_2	N_A	T
AI	ATJ	1.046	2.078	1.986	0.223	0.037	0.089	27.500
		(0.322, 3.322)	(0.758, 9.473)		(0.119, 0.482)	(0.016, 0.132)	(0.157, 5.999)	(11.5, 973.5)
AI	MI	31.938	23.875	0.748	0.039	0.008	0.705	0.425
		(5.813, 120.813)	(8.375, 236.625)		(0.019, 0.241)	(0.004, 0.082)	(0.065, 2.492)	(0.675, 48.675)
ATJ	MTJ	6.500	385.500	59.30	0.003	0.001	0.000	31.250
		(11.5, 970.5)	(76.5, 2812.5)		(0.002, 0.104)	(0.001, 0.051)	(0.001, 0.054)	(16.75, 484.25)
MI	MTJ	2.065	4.288	2.076	0.030	0.008	0.035	164.500
		(0.825, 9.325)	(1.788, 22.913)		(0.016, 0.130)	(0.004, 0.055)	(0.569, 22.629)	(34.5, 975.5)

Multilocus IMa joint point estimates of the immigration rate ($M_1 = 2N_1m_1$), relative pairwise population migration rate (M_2/M_1), contemporary and ancestral effective population sizes (N_1, N_2, N_A), divergence times ($T = t/2N_e$ = generations) and relevant credibility intervals per population pair. Each pairwise population comparison is listed in the first two columns and refers to AI, Asia *indica*, ATJ, Asia tropical *japonica*, MTJ, Madagascar tropical *japonica* and MI, Madagascar *indica*. The point estimates for each parameter are the bins with the highest estimate after curve smoothing. Ninety-five percentage credibility intervals for the posterior distributions for each parameter are indicated in parentheses and are the upper and lower highest posterior density intervals of the posterior probability surface. M_1 and M_2 are the estimated effective number of migrants per generation for each direction. M_2/M_1 is the relative population migration rate between each population pair, comparing immigration into Pop 1 from Pop 2 vs. immigration into Pop 2 from Pop 1. Estimated effective population sizes are listed in columns N_1, N_2, and N_A and refer to contemporary populations ('Pop 1' and 'Pop 2') in the relevant comparison followed by the ancestral population from which both contemporary populations are derived.

contemporary populations are derived (Table 3). The divergence time estimate between AI and MI is significantly shorter than the number of generations estimated as the divergence of ATJ and MTJ. The longest divergence time estimate in this analysis is for the population ancestral to MI and MTJ. These results, however, may not reflect true divergence times but instead suggest that associated migration rate estimated between these two subspecies is the result of very recent allele sharing. It should be noted, moreover, that all these estimates should be treated with caution, as the credibility intervals for all the point estimates are fairly large.

Recombination and linkage disequilibrium in Madagascar rice

To identify possible recombinants between *indica* and tropical *japonica* landraces on Madagascar, we used SNPs that were fixed between the two major rice variety groups from Asia. Nearly one-fifth of Madagascar landraces (eight accessions) show evidence of recombination between *indica* and tropical *japonica* using this approach. Most of the recombinant landraces in Madagascar may be advanced generation hybrids, given that recombinant segments of the genome are found in some but not all the chromosomes (see Fig. 4). Evidence of double recombination is also found in the Madagascar landrace Tsimatahopaosa, where two clear recombination events approximately 76 and 5.549 Mb apart are apparent on chromosome 1. Finally, there is evidence of segregation between chromosomes in this and other putatively hybrid landraces. Despite evidence of recombination on Madagascar, there is no appreciable decay of LD with distance in the island landraces as a group (results not shown).

Discussion

Domesticated *Oryza sativa* evolved twice in the last 9000 years—in the Indian subcontinent, giving rise to the *indica* variety group and in the Yangtze Valley in China, giving rise to tropical *japonica* (Londo *et al.* 2006)—and has since fed more people than any other crop in human history. The movement of *O. sativa* outside of its centres of domestication has contributed to its widespread distribution in the Old World, and dissecting the nature of introduction of rice to new areas provides clues to the dynamics of the spread of domesticated species and the evolutionary genetics accompanying domesticated species introductions.

The introduction of rice into Madagascar is believed to have occurred at least twice in the last ~1000 years from Asia. Anthropological (Koji 1997) and genetic studies (Ahmadi *et al.* 1988, 1991; Rabary *et al.* 1989) have suggested that rice in Madagascar may exhibit population substructuring corresponding to the large-scale genetic differentiation between *indica* varieties and tropical *japonica* varieties. The introduction of distinct *O. sativa* landraces corresponds to the separate movement of Malay/Indonesian archipelago and Indian subcontinent peoples to Madagascar, both of whom

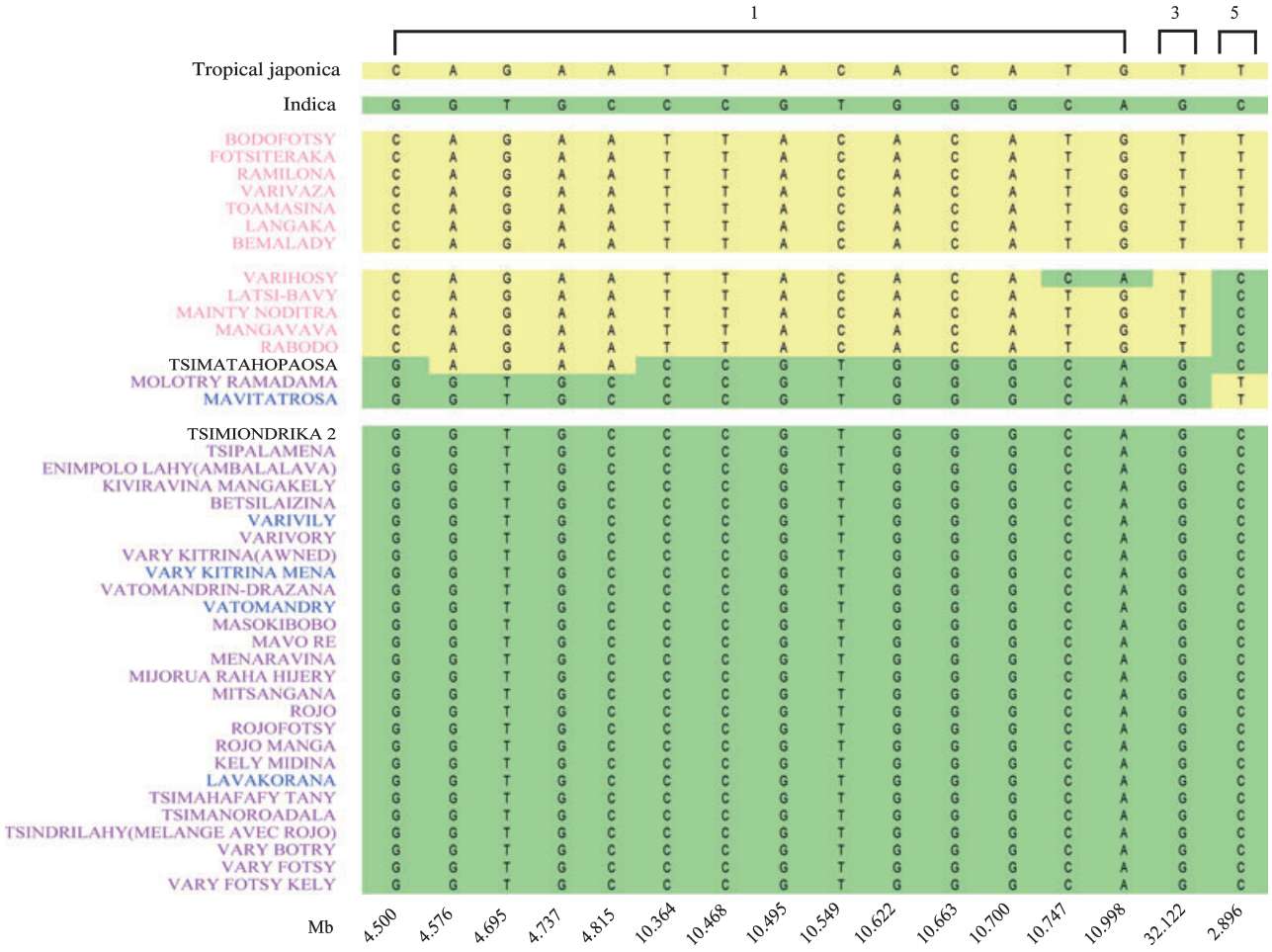

Fig. 4 Patterns of SNPs showing recombination and segregation in Madagascar accessions. Only fixed SNPs between tropical *japonica* and *indica* in Asia are used in the analysis. The cell colours indicate whether the SNP is diagnostic of *indica* (green) or tropical *japonica* (yellow), and their general locations in chromosomes 1, 3 and 5 are shown (bottom). Variety names of Madagascar accessions are coloured according to their predominant STRUCTURE membership assignment (≥60% membership coefficient; pink = tropical *japonica*; purple = unique Madagascar *indica* group; blue = Asian *indica*; compare with Fig. 1 and Table S2, Supporting information). Tsimatahopaosa and Tsimiondrika 2 are admixed accessions (i.e. <60% membership to a population).

introduced agriculture to the island (Dewar & Wright 1993; Allibert 2008; Regueiro *et al.* 2008). It should be noted that the levels of gene flow between Asia and Madagascar groups differ with tropical *japonica*, showing a strong bias in introgression from the continent to the island while there may be weak back-migration of *indica* back to Asia (see Table 3; Fig. 3).

The differentiation of Madagascar rice into *indica* and tropical *japonica* varieties as observed in a population structure analysis together with their sampling localities suggests that these two variety groups are also geographically structured on the island (see Fig. 2). Interestingly, the geographic structuring of Madagascar rice corresponds to cultural practices of agriculture on the island (Koji 1997). The cultural groups in the east coast of Madagascar, where tropical *japonica* is primarily found in our sample, traditionally employ agricultural techniques that have clear Malay Peninsular origins. In contrast, groups in the central areas of Madagascar where the *indica* varieties are primarily found use a unique combination of techniques that appear to be a Malay/Indian mix (Koji 1997).

It is likely that only a fraction of the genetic variation of the Asian landraces is represented in this offshore African island. This is observed in our data, where the mean levels of nucleotide diversity for both tropical *japonica* varieties and *indica* varieties are 40–60% lower in Madagascar than in Asia (Table 2). This reduction in diversity and more positive mean Tajima's D (and across all five genomic regions, see Supplement Table S3) for MTJ is consistent with population bottlenecks associated with founder effects in the island

andraces when compared to those in the source continental group. The pattern of landrace variation in Madagascar, however, differs between tropical *japonica* and *indica*. The reduction in molecular diversity in the tropical *japonica* landraces is less severe than the *indica* varieties, as supported by the relative effective population sizes of Asian vs. Madagascar groups (see Table 3). This suggests that human migration to Madagascar brought a more diverse set of tropical *japonica* landraces to the island. This was not the case for MI, which has a more negative Tajima's D than their continental counterparts that may be attributed to the sampling of an excess of rare haplotypes during severe bottlenecking (Depaulis *et al.* 2003) and/or positive selection.

The reduction in diversity of MI compared to AI is almost entirely because of the high frequency of a distinctive *indica* group observed in the island. AMOVA, STRUCTURE and neighbour-joining analyses confirmed that island *indica* populations are genetically differentiated from the Asian population, with this partition explaining as much as 68% of the genetic structure in this cultivar (Table 1). This was not the case for tropical *japonica* where Asian and Madagascar accessions could not be distinguished in the STRUCTURE and neighbour-joining analyses. However, AMOVA revealed significant differentiation between ATJ and MTJ, though with low overall variance (18.3%, Table 3).

The Madagascar *indica* landraces are almost entirely identical, and there are three possibilities that could account for the predominance of this distinct group: (i) The introduction of *indica* into Madagascar utilized a much narrower genetic base; (ii) there was increased genetic drift in this group; or (iii) there was selection for this group because of adaptation to the island agro-ecological environment. These possibilities are not mutually exclusive, and more than one possibility could be operative; unfortunately, there is insufficient data to determine the relative importance of one or more of these alternative possibilities. Moreover, our observation that three continental *indica* landraces in India belong to the same population cluster as the distinctive Madagascar *indica* group suggests that this predominant island group originated from India. Alternatively, the Indian accessions may represent reverse migration, but given the known direction of rice across the Indian Ocean (Koji 1997), this appears less likely.

The identification of a unique Madagascar rice group was also shown by previous allozyme studies, and it was thought that this distinctive landrace group might be an *indica*/tropical *japonica* hybrid (Ahmadi *et al.* 1991; Rabary *et al.* 1989). Our analysis, however, indicates that this distinct group is not of hybrid origins (except for Molotry Ramadama, Fig. 4), though our accessions belonging to this group also grew in high altitudes (*c.* >1000 m; Supplement Table S2; Fig. 2) and their average grain width (but not length) is intermediate to those of the *indica* and tropical *japonica* accessions collected from the island (groups G5 and G4, respectively, in Ahmadi *et al.* 1991; Supplement Table S2) like the distinct Madagascar type identified by Ahmadi *et al.* (1991; group G6) .

However, we do observe separate evidence of hybridization and recombination between the *indica* and tropical *japonica* variety groups on the island, which indicates a history of gene flow between these two major variety groups. At least eight landraces show evidence of recombination between these two variety groups from distinct evolutionary lineages (Fig. 4), despite genetic evidence that indicates partial sterility of *indica* and tropical *japonica* hybrids (Chen *et al.* 2008). It is unclear whether hybrids between these two variety groups are also widespread on continental and insular Asia where both *indica* and tropical *japonica* are cultivated, although Chen *et al.* (2008) demonstrated that a unique group of rice germplasm known as wide compatibility varieties, which have been found to produce highly fertile hybrids when crossed with both *indica* and *japonica* varieties, is in fact a widespread phenomenon. Our results suggest that the contact between *indica* and *japonica* varieties occurred via separate introductions to Madagascar and has led to novel hybrid landraces there.

The recent spread of domesticated rice outside its major range in Asia to the offshore African island of Madagascar shares similarities with other human-associated movement of species, including the movement of invasive species (Dlugosch & Parker 2008). These include reduction in diversity associated with founder events, histories of multiple introductions of species and hybridization of introduced genotypes in their novel environments (Dlugosch & Parker 2008). Domesticated species, however, are unique in that geographic spread and introduction occurs deliberately under human intervention, and the establishment in new environments is under human control.

The finding of reduced diversity for Madagascar rice, coupled with evidence of limited hybridization and recombination, also provides an opportunity to use these populations to identify genes underlying plant variation and adaptation. Despite evidence of recombination, however, there still remains a large amount of LD in the genome of Madagascar rice. This is likely because of the largely selfing nature of *O. sativa*, the limited number and short history of hybridization and recombination events between the Madagascar *indica* and tropical *japonica* landraces, reinforced by the partial reproductive isolation between these two major rice variety groups, as well as the recent bottleneck on the

island associated with the founder introductions. Nevertheless, it may be possible to use these populations in structured association studies within *indica* and tropical *japonica*, as well as in studies of the differences between these two major rice groups.

Acknowledgements

This work was funded in part by a US National Science Foundation Plant Genome Research Program MCB 0701836. Computational support for the Isolation-with-migration analyses was provided by Edward Duffy and James von Oehsen from the Clemson University Computational and Information Technology group and utilized a framework funded by Clemson University startup funds and USDA Grant 2009-35320-05036 (ALR).

References

Ahmadi N, Becquer T, Larroque C *et al.* (1988) Genetic variability of rice in Madagascar. *Agronomie Tropicale*, **43**, 209–221.

Ahmadi N, Glaszmann JC, Rabary E (1991) Traditional highland rices originating from intersubspecific recombination in Madagascar. In: *Rice Genetics II*: Proceedings of the Second International Rice Genetics Symposium 14–18 May 1990. pp. 67–79, IRRI Press, Manila.

Allibert C (2008) Austronesian migration and the establishment of the Malagasy civilization: contrasted readings in linguistics, archaeology, genetics and cultural anthropology. *Diogenes*, **55**, 7–15.

Ashikari M, Matsuoka M (2006) Identification, isolation and pyramiding of quantitative trait loci for rice breeding. *Trends Plant Science*, **11**, 344–350.

Burney DA, Burney LP, Godfrey LR *et al.* (2004) A chronology for late prehistoric Madagascar. *Journal of Human Evolution*, **47**, 25–63.

Caicedo AL, Williamson SH, Hernandez RD *et al.* (2007) Genome-wide patterns of nucleotide polymorphism in domesticated rice. *PLOS Genetics*, **3**, 1745–1756.

Chen J, Jihua D, Yidan O *et al.* (2008) A triallelic system of S5 is a major regulator of the reproductive barrier and compatibility of indica–japonica hybrids in rice. *Proceedings of the National Academy of Sciences USA*, **105**, 11436–11441.

Clark RM, Tavare S, Doebley J (2005) Estimating a nucleotide substitution rate for maize from polymorphism at a major domestication locus. *Molecular Biology and Evolution*, **22**, 2304–2312.

Depaulis F, Mousset S, Veuille M (2003) Power of neutrality tests to detect bottlenecks and hitchhiking. *Journal of Molecular Evolution*, **57**, S190–S200.

Dewar RE, Wright HE (1993) The culture history of Madagascar. *Journal of World Prehistory*, **7**, 417–466.

Diamond J (2002) Evolution, consequences and future of plant and animal domestication. *Nature*, **418**, 700–707.

Dlugosch KM, Parker I (2008) Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology*, **17**, 431–449.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Farrell BD, Sequeira AS, O'meara BC *et al.* (2001) The evolution of agriculture in beetles (Curculionidae : Scolytinae and Platypodinae). *Evolution*, **55**, 2011–2027.

Gao H, Williamson S, Bustamante CD (2007) An MCMC approach for joint inference of population structure and inbreeding rates from multi-locus genotype data. *Genetics* (online), **176**: 1635–1651.

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

Hartl DL, Clark AG (1997) *Principles of Population Genetics*. 3rd edn, Sinauer Associates, Sunderland, Massachusetts.

Harushima Y, Nakagahra M, Yano M *et al.* (2002) Diverse variation of reproductive barriers in three intraspecific rice crosses. *Genetics*, **160**, 313–322.

Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov -chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences USA*, **104**, 2785–2790.

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature*, **436**, 793–800.

Khush GS (1997) Origin, dispersal, cultivation and variation of rice. *Plant Molecular Biology*, **35**, 25–34.

Kimura M (1980) A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**, 111–120.

Koji T (1997) Malayan cultivated rice and its expansion. *Agricultural Archaeology*, **1**, 73–78.

Li ZK, Pinson SRM, Paterson AH *et al.* (1997) Genetics of hybrid sterility and hybrid breakdown in an intersubspecific rice (*Oryza sativa* L) population. *Genetics*, **145**, 1139–1148.

Londo JP, Chiang YC, Hung KH *et al.* (2006) Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. *Proceedings of the National Academy of Sciences USA*, **103**, 9578–9583.

Mather KA, Caicedo AL, Polato NR *et al.* (2007) The extent of linkage disequilibrium in rice (*Oryza sativa* L.). *Genetics*, **177**, 2223–2232.

Oka HI (1988) *Origin of Cultivated Rice*. Elsevier/JSSP, Amsterdam/Tokyo.

Olsen KM, Caicedo AL, Polato N *et al.* (2006) Selection under domestication: evidence for a sweep in the rice *Waxy* genomic region. *Genetics*, **173**, 975–983.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Purugganan MD, Fuller D (2009) The nature of selection during plant domestication. *Nature*, **457**, 843–848.

Rabary E, Noyer JL, Benyayer P *et al.* (1989) Variabilité génétique du riz (*Oryza sativa* L.) à Madagascar: origine de types nouveaux. *L'Agrononomie tropicale*, **44**, 305–312.

Regueiro M, Mirabal S, Lacau H *et al.* (2008) Austronesian genetic signature in East African Madagascar and Polynesia. *J. Hum. Genetics*, **53**, 106–120.

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (eds Krawetz S, Misener S), pp. 365–386, Humana Press, Totowa, NJ.

Saitou N, Nei M (1987) The neighbor-joining method – a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4**, 406–425.

Schultz TR, Brady SG (2008) Major evolutionary transitions in ant agriculture. *Proceedings of the National Academy of Sciences USA*, **105**, 5435–5440.

Tamura K, Nei M, Kumar S (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences USA*, **101**, 11030–11035.

Tamura K, Dudley J, Nei M *et al.* (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596–1599.

Thornton K (2003) Libsequence: a C++ class library for evolutionary genetic analysis. *Bioinformatics*, **19**, 2325–2327.

Varilo T, Peltonen L (2004) Founder events and mapping: isolates and their potential use in complex gene mapping efforts. *Current Opinion in Genetics and Development*, **14**, 316–323.

Vaughan DA, Miyazaki S, Miyashita K (2004) The rice genepool and human migrations. In: *Biological Resources and Migration* (ed. Wener D), pp. 1–14. Springer, Berlin.

J. Molina, J. Flowers and K. Mather are postdoctoral research scientists in the laboratory of M. Purugganan. Research in the Purugganan laboratory is focused on the evolutionary genomics of domestication and other plant adaptations, as well as on socio-genomics. A. Caicedo, A. Lawton-Rauh and Brad Rauh are also interested in questions of plant molecular evolutionary genetics. K. McNally is a molecular geneticist whose research focuses on studying and exploiting the diversity of rice.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Non-Madagascar accessions included in this study (modified from Caicedo *et al.* 2007)

Table S2 *Oryza sativa* accessions from Madagascar used in this study, their ecological and morphological characteristics, geographic coordinates and provenance information available from The International Rice Genebank Collection Information System (IRGCIS)

Table S3 Diversity statistics for each genomic region in the study for Asian and Madagascar populations of *indica* and tropical *japonica* with (R) and without recombinants (NR)

Fig. S1 The inferred population structuring for all accessions based on data from all regions from $K = 2$ to $K = 7$.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.