

Genetic Diversity of Isolated Populations of Indonesian Landraces of Rice (*Oryza sativa* L.) Collected in East Kalimantan on the Island of Borneo

Michael J. Thomson · Nicholas R. Polato ·
Joko Prasetyono · Kurniawan R. Trijatmiko ·
Tiur S. Silitonga · Susan R. McCouch

Received: 23 October 2008 / Accepted: 2 February 2009 / Published online: 19 February 2009
© Springer Science + Business Media, LLC 2009

Abstract Although the genetic diversity of rice germplasm has been well characterized globally, few studies have taken an in-depth view of a large number of rice landraces on a local scale. To better understand the relationships between rice genetic diversity and associated geographic and cultural factors, we collected and characterized 183 rice landraces from 18 villages along the Bahau and Kayan rivers in the Indonesian province of East Kalimantan on the island of Borneo. A genetic diversity analysis using 30 microsatellite markers detected a clear distinction between the *indica* and *japonica* varietal groups ($F_{st}=0.59$), with 80% of the landraces identified as tropical *japonica* and

20% *indica*, which largely correlated with the field-level ecotypes: upland *japonica* and lowland *indica*. Indigenous knowledge from local farmers was gathered about the names, origins, and uses of the landraces, which provides a rich background to compare with the genetic relationships of these traditional varieties.

Keywords Rice (*Oryza sativa*) · Genetic diversity · Landrace · Ecotype · SSR markers

Introduction

The genetic diversity of cultivated rice germplasm on a global scale has been well characterized using molecular markers (Caicedo et al. 2007; Garris et al. 2005; Glaszmann 1987; Yang et al. 1994; Yu et al. 2003). To encompass the entire range of rice diversity, these studies have sampled rice varieties from many different rice-growing countries around the world. Likewise, efforts to designate core collections have employed sampling strategies to maximize the genetic diversity of the subset, while minimizing the numbers of accessions studied to a reasonable level. Consequently, while global studies will provide an excellent overview of the population structure of cultivated rice, they cannot provide an in-depth view of rice germplasm on a local scale, since each region will only be represented by a few varieties. Increasingly, studies have begun to characterize subsets of rice germplasm at the country-wide level, with molecular markers being used to describe the genetic diversity of rice within specific countries (Gao et al. 2005; Jain et al. 2004; Lu et al. 2005; Pessoa-Filho et al. 2007; Prashanth et al. 2002; Thomson et al. 2007). However, there is still a lack of data that describe molecular

Electronic supplementary material The online version of this article (doi:10.1007/s12284-009-9023-1) contains supplementary material, which is available to authorized users.

M. J. Thomson · J. Prasetyono · K. R. Trijatmiko · T. S. Silitonga
Indonesian Center for Agricultural Biotechnology and Genetic
Resources Research and Development,
Jl. Tentara Pelajar 3A,
Bogor 16111, Indonesia

N. R. Polato · S. R. McCouch (✉)
Department of Plant Breeding and Genetics, Cornell University,
162 Emerson Hall,
Ithaca, NY 14853, USA
e-mail: srm4@cornell.edu

Present address:

M. J. Thomson
Plant Breeding, Genetics and Biotechnology Division,
International Rice Research Institute,
Los Baños, Laguna, Philippines

Present address:

N. R. Polato
Department of Biology, Penn State University,
University Park, PA 16802, USA

diversity of rice on a local scale. The most complete study to date characterized a set of 692 rice landraces from the province of Yunnan, China using 20 simple sequence repeat (SSR) markers and found seven population sub-groups roughly corresponding to subspecies (*indica/japonica*), field-level ecotypes (lowland/upland), and seasonal ecotypes based on maturity (Zhang et al. 2007). Another study examined the diversity of 170 accessions from 14 villages in Guinea and found that diversity between accessions within the same village was the most important component of regional genetic diversity (Barry et al. 2007). Likewise, several studies have begun to explore plant genetic diversity at a local scale in other crops, such as sorghum and millet in Africa (Adoukonou-Sagbadja et al. 2007; Barnaud et al. 2007). As additional studies detail rice genetic diversity on a local scale, the complex interaction between rice diversity and human cultivation practices can be better understood.

Rice landraces have been shaped by the interplay between adaptation to the local environment and selection imposed by farmers who determine which varieties will be grown each year. This is in contrast to wild rice populations, which proliferate based on their ability to survive and compete under natural conditions. Natural populations tend to be more stable than cultivars in their geographic distributions because the forces of natural selection generally lead to gradual changes over evolutionary time. Several studies carried out on wild rice populations in China have characterized the local diversity of *Oryza rufipogon* and found varying degrees of differentiation within and between local populations (Song et al. 2003; Xu et al. 2006; Zhou et al. 2003). In contrast, the accelerated evolution provided by intense artificial selection and seed exchanges among local farmers has led to dramatic changes in the genetics of cultivated crops in a relatively short time frame (<10,000 years). While the story of rice domestication is still being unraveled based on interpretations of geographic and phylogenetic data (Caicedo et al. 2007; Londo et al. 2006) and analysis of cloned domestication genes (Kovach et al. 2007; Sweeney et al. 2007), it appears that rice was domesticated from diverse gene pools and that gene flow among sub-populations and with wild ancestral populations has created a complex series of events leading to the modern day fabric of rice diversity (Sang and Ge 2007; Sweeney and McCouch 2007), although the possibility of a single domestication event is still debated (Gao and Innan 2008; Vaughan et al. 2008). While the relationships between the earliest domesticates and indigenous landraces surviving in rice-growing regions today are still unclear, signatures of shared ancestry, selection, and introgression are written in the genomes of traditional varieties. Over many generations, rice has become an essential component of the local cultures and customs of

rice-growing people across Asia, where rice cultivation has been practiced for thousands of years. Yet there is still a great deal to learn concerning the impact of farmer practices on the preservation, exchange, and continuing evolution of traditional landraces today. This knowledge will provide an essential foundation for making decisions about the conservation and use of traditional germplasm in breeding programs.

The Green Revolution was felt throughout Asia and greatly impacted the varietal landscape due to rapid dissemination of high-yielding, semi-dwarf *indica* varieties that displaced many traditional landrace varieties. This study focused on an isolated region on the island of Borneo populated by traditional hunter-gatherers who also practice rice cultivation. We were interested to examine the genetic structure and diversity of the rice varieties that were being grown by these villagers and to determine whether they still cultivated exclusively traditional landraces or whether there was any evidence that Green Revolution varieties had been introduced in the region.

There are a number of important issues concerning crop genetic diversity and its relationship to local cultures that can only be answered by intensively studying traditional varieties collected from a geographically isolated region (one that has not been affected by the introduction of modern varieties). For example, how do the landraces in an isolated region reflect the ethnic history, cultural preferences, and production practices of local farmers? How long have these varieties been cultivated by people in the village? Are the genetic relationships among local landraces defined more by geographical proximity, or does the dynamic exchange of germplasm within and between regions play a greater role? Which specific plant and/or grain characters are selectively maintained and valued by local cultures? How do farmers decide what varieties to plant every year? Carefully focused germplasm collection trips are needed to explore the intricacies and dynamics of rice genetic diversity on a local scale and to provide a more detailed view of the rice varieties adapted to a specific set of local environments and cultural expectations.

To begin to answer some of these questions, we initiated a collection trip to a remote region in the interior of Borneo with the goal of collecting indigenous landraces along with corresponding information from local farmers (Fig. 1). We selected the Bahau River to focus the collection trip due to its isolated location and because a previous survey in 1993 had interviewed local farmers and identified 35 distinct rice landraces grown in the village of Apau Ping (Setyawati 2003). This region of East Kalimantan, which is partially encompassed by the Kayan Mentarang National Park covering 1.6 million hectares, is sparsely populated by several thousand indigenous people collectively called the Dayak. The

Fig. 1 Location of the three geographic regions from which rice accessions were collected on the Indonesian province of East Kalimantan on the island of Borneo: including five villages from the Upper Bahau river, four villages from the lower Bahau river, and the remaining from villages along the lower Kayan river.



Bahau river area is settled by a number of related Kenyah tribes who had migrated from nearby Sarawak during the seventeenth and eighteenth centuries (Anau 2003). These indigenous people rely on hunting, gathering forest products, and rice farming for their subsistence. Most of the rice is grown through the traditional practice of swidden, or shifting, cultivation where areas on the mountain sides are cleared and sown with upland rice for 1 year and then left fallow for a long period (usually more than 7 years) before the farmers return to the same location (Supplementary Fig. 1). More recently, several villages along the Bahau river have developed irrigated lowland rice fields which allow for more intensive cultivation. Our objectives were (1) to gather traditional landraces from an isolated region; (2) to collect information from farmers concerning the meanings of the variety names, origins, and special uses; and (3) to use genetic diversity data to characterize the accessions, compare the genetic relatedness with factors such as geographic isolation, ecotype, grain characters, and names, and ultimately to make conclusions that are relevant for the conservation and genetic use of these landraces.

Results

Overall SSR diversity

Seeds from 183 rice varieties were collected from 18 villages along the Bahau and Kayan rivers in the Indonesian province of East Kalimantan (Table 1). Fluorescent SSR genotyping by capillary electrophoresis was performed on the 183 landraces plus seven controls across 30 SSR markers (Supplementary Tables 1 and 2). A total of 166 alleles were detected at the 30 microsatellite markers across the 190 rice accessions, ranging from two alleles up to 15 alleles per locus, with an average of 5.5 alleles across the study. The polymorphism information content (PIC) values ranged from 0.04 (RM271) to 0.87 (RM154), with an average of 0.45, and the mean heterozygosity was 0.03. Rare alleles, defined as those alleles with a frequency less than 5%, were identified at 70% of the loci, with an average of two rare alleles per locus, while the frequency of the most common allele at each locus ranged from 19% (RM162) to 98% (RM271). On average, 62% of the 190 rice accessions shared a common major allele at any given locus (Supplementary Table 2).

Table 1 East Kalimantan Villages Where Rice Varieties were Collected in March 2005

Village	Sub-district	District	River	GPS coordinates ^a	Elevation (feet) ^b	No. ^c	% Upland
Apau Ping	Punjungan	Malinau	Upper Bahau	03° 06.155 N 115° 49.116 E	1,482 (vil.), 1683 (field)	19	79
Long Berini	Punjungan	Malinau	Upper Bahau	n/a	n/a	13	100
Long Kemuat	Punjungan	Malinau	Upper Bahau	02° 57.281 N 115° 51.443 E	1,323 (vil.)	4	75
Long Alango	Punjungan	Malinau	Upper Bahau	02° 54.781 N 115° 50.826 E	1,228 (vil.), 1,950 (field)	22	55
Long Tebulo	Punjungan	Malinau	Upper Bahau	02° 50.607 N 115° 50.404 E	1,111 (vil.), 1,308 (field)	21	48
Long Uli	Punjungan	Malinau	Lower Bahau	02° 46.358 N 115° 48.178 E	941 (village)	9	100
Long Peleran	Punjungan	Malinau	Lower Bahau	02° 42.359 N 115° 47.109 E	906 (village)	8	63
Long Pujungan	Punjungan	Malinau	Lower Bahau	02° 35.459 N 115° 47.527 E	779 (village)	24	100
Long Aran	Punjungan	Malinau	Lower Bahau	02° 32.533 N 115° 48.979 E	751 (village)	4	100
Long Bia	Long Peso	Bulungan	Kayan	02° 41.912 N 116° 47.627 E	n/a	10	90
Long Peso	Long Peso	Bulungan	Kayan	02° 41.412 N 116° 47.077 E	118 (village)	12	83
Long Lembu	Long Peso Hilir	Bulungan	Kayan	n/a	n/a	6	100
Penisir	Tanjung Palas	Bulungan	Kayan	n/a	n/a	2	100
Teras Baru	Tanjung Palas	Bulungan	Kayan	n/a	n/a	15	0
Merukau	Tanjung Palas Barat	Bulungan	Kayan	n/a	n/a	2	100
Salangketo	Tanjung Palas Tengah	Bulungan	Kayan	n/a	n/a	5	0
Sabanar Baru	Tanjung Selor	Bulungan	Kayan	n/a	n/a	6	50
Sabanar Lama	Tanjung Selor	Bulungan	Kayan	n/a	n/a	1	100

^a GPS coordinates taken using map data WGS 84

^b Approximate elevation using a handheld GPS unit at the village or nearby rice fields

^c Number of rice varieties actually collected from each village (not a comprehensive survey)

Genetic distance-based analysis

The genetic distance-based results using UPGMA clustering revealed two major groups in the Indonesian germplasm (Fig. 2). Using the control varieties to assign groups, the larger group of 146 Kalimantan accessions corresponds to *japonica*, while the smaller group of 36 accessions corresponds to *indica*. No Indonesian accessions cluster with the *aus* variety Kasalath. Similarly, the principal coordinate analysis revealed a clear separation between the *indica* and *japonica* groups, although five Kalimantan accessions (P. Kancat, P. Pui, P. Ketan Adong Adong, P. Mbau Enan Bulu, and P. Putuk) seemed to lie outside the primary *japonica* cluster (Fig. 3). Upon labeling the accessions with their origin from the Upper Bahau river, lower Bahau river, and Kayan river, there is one cluster of 25 landraces that was found only in the more remote Bahau river region, while the remainder of the clusters contained a mixture of accessions from all three regions (Fig. 2). The Kayan river subset had a slightly higher proportion of *indica* accessions (29% *indica*/71% *japonica*) compared with the more remote Bahau river (17% *indica*/83% *japonica*), combining to give an overall proportion of 20% *indica* and 80% *japonica* across all 183 Kalimantan accessions. When farmer designations for ecotype (dry upland cultivation versus irrigated lowland) are overlaid on an unrooted neighbor-joining tree, it becomes clear that the

indica group largely corresponds to lowland varieties while the *japonica* group consists mostly of upland varieties (Fig. 4). In fact, only 5% of the *indica* accessions were declared to be upland varieties, while 14% of the *japonica* accessions were labeled by the farmers as irrigated lowland varieties.

Population structure analysis

An analysis of population structure of the 183 Kalimantan landraces identified the most appropriate grouping with six sub-populations: one group corresponding to the 36 *indica* landraces and five other sub-groups for the *japonica* varieties ($K=6$ using the program STRUCTURE; Fig. 5). 149 accessions had >75% of their inferred ancestry derived from one of the six groups, while 34 landraces showed varying degrees of admixture between sub-groups. Most of the subgroups were composed of varieties from both the Bahau and Kayan regions, with only one subgroup differentiated geographically: Pop2 had 23 out of 24 accessions coming from the Bahau river (Supplementary Table 3). There was also differentiation between sub-groups based on proportions of upland/lowland and glutinous/non-glutinous varieties. Pop2 and Pop4 consisted largely of glutinous upland varieties (Pop2: 96% glutinous and 96% upland; Pop4: 69% glutinous and 100% upland), Pop5 was largely non-glutinous upland (14% glutinous, 83% upland), Pop 6 consisted of half glutinous and

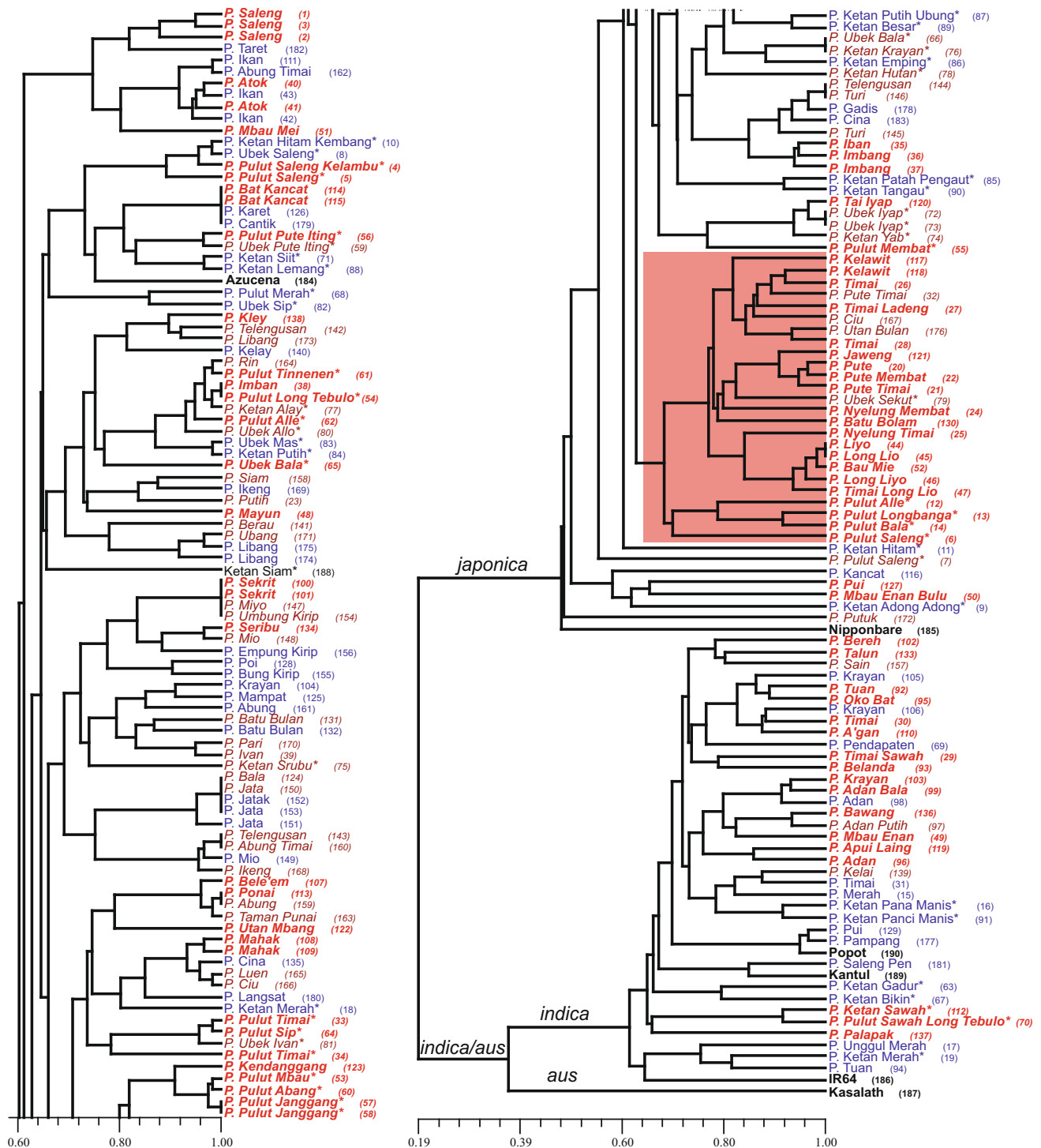


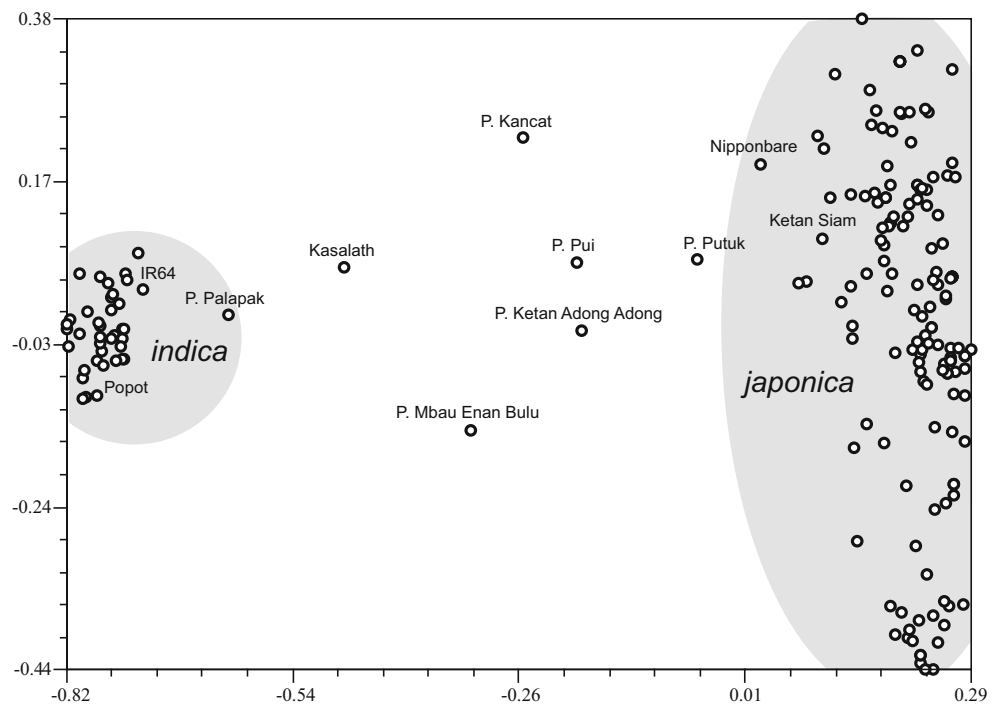
Fig. 2 A UPGMA cluster dendrogram showing the genetic relationships between the 190 rice accessions based on 30 microsatellite markers. Each accession is identified by name (with an asterisk for glutinous varieties), accession number, and origin (Upper Bahau river accessions in *bold italics*, lower Bahau river accessions in *italics*, Kayan river accessions in *normal font*, and control accessions in *bold*). The stems of the major groups for *indica* and *japonica* are labeled, and a cluster of traditional landraces only found in the Bahau river region is shaded.

mostly upland (50% glutinous and 93% upland), while Pop3 consisted of mostly non-glutinous and half upland (14% glutinous and 50% upland). Lastly, an analysis at $K=2$ split the 183 landraces into the two *indica* and *japonica* varietal

italics, lower Bahau river accessions in *italics*, Kayan river accessions in *normal font*, and control accessions in *bold*). The stems of the major groups for *indica* and *japonica* are labeled, and a cluster of traditional landraces only found in the Bahau river region is shaded.

groups, with 36 landraces having >75% *indica* ancestry and 143 landraces with >75% *japonica* ancestry. This is in agreement with the PCA analysis (Fig. 3). Furthermore, the fact that all the Indonesian *japonica* varieties cluster more

Fig. 3 Principal coordinate analysis (PCA) with the two *indica* and *japonica* groups outlined. Several control varieties and the five accessions showing admixture are also labeled.



closely with Azucena, a tropical *japonica* variety, than with Nipponbare, a temperate *japonica* variety in both the PCA and the dendrogram (Fig. 2) confirms that the Indonesian *japonica* materials all belong to the tropical *japonica* sub-population, as expected. In addition to the *indica* and tropical *japonica* groups, five landraces showed significant *indica/japonica* admixture (P. Kancat, P. Pui, P. Ketan Adong Adong, P. Mbau Enan Bulu, and P. Putuk). Although these five landraces were designated as *japonica* varieties based on the UPGMA analysis (Fig. 2), their level of *japonica* ancestry ranged from just 44% for P. Mbau Enan Bulu to 77% for P. Putuk (Supplementary Table 3).

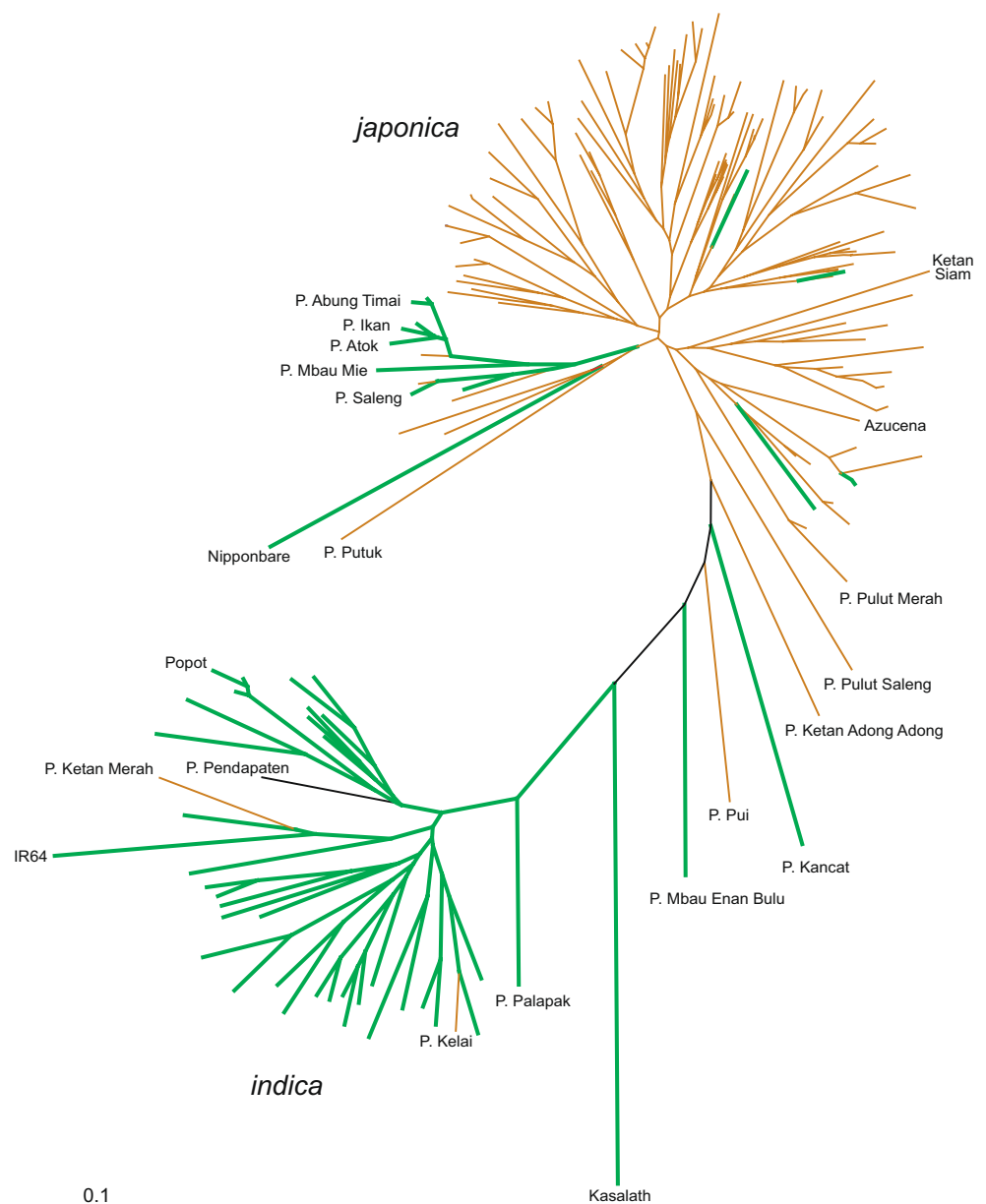
Population differentiation and diversity

When the 183 Kalimantan accessions are analyzed by subgroup, a high degree of population differentiation was seen between the Kalimantan *indica* and *japonica* groups with an F_{st} of 0.59. Furthermore, the *japonica* group had a slightly higher gene diversity at 0.35 compared with 0.30 for the *indica* group, and the *japonica* group had an average of 4.5 alleles per locus, compared with 2.6 alleles for the *indica* group (Table 2). In contrast, an analysis of the Kalimantan landraces divided into three geographical groups (upper Bahau, lower Bahau, and Kayan rivers) showed no population differentiation based on their geographic origin ($F_{st}=0.03$). Likewise, these three groups showed similar values for gene diversity, though slightly lower for the lower Bahau at 0.38 compared with the upper Bahau at 0.49 and Kayan at 0.53 (Table 2).

Duplicates and naming conventions

During the collection trip, seed samples from varieties having the same name but from different villages were treated as separate accessions to allow for an analysis of the relationship between variety names and their genetic similarity. Of the 183 collected landraces, there were 31 sets of varieties having the same name (22 pairs, seven triplets, and two sets of four accessions with identical names), which represents 23% of the collection as being potentially duplicated. An analysis of the genetic similarity, however, revealed only five pairs of accessions with the same name as being genetically identical for the loci surveyed, indicating that only 3% of the collection is actually duplicated (P. Bat Kancat, P. Sekrit, P. Jata, P. Pulut Janggang, and P. Ubek Iyap; Fig. 2). Of those five duplicate accessions, four were found in the same region, while the pair of P. Jata duplicates had one accession from the lower Bahau river and the other accession from the Kayan river region. Another 12 sets of accessions had the same name and were genetically similar, though not identical, suggesting that they were members of heterogeneous landraces. In contrast, ten sets of accessions with the same name were in completely different clusters within the *japonica* or *indica* subspecies and another four sets had accessions of the same name being split across the *japonica* and *indica* division (P. Ketan Merah, P. Krayan, P. Pui, and P. Timai; Fig. 2). In summary, of the 73 accessions being found in 31 sets of same-named varieties, 60% were genetically similar while 40% were completely different. Information obtained from

Fig. 4 An unrooted neighbor-joining tree showing the genetic relationships between the 190 rice accessions, with the upland varieties shown as *narrow lines* and the irrigated lowland varieties as *heavy lines*. Seven control varieties are labeled (Nipponbare, Azucena, IR64, Kasalath, and three Indonesian varieties previously studied: Popot, Ketan Siam, and Kantul), along with several varieties showing *indica/japonica* admixture and a cluster of irrigated *japonica* landraces. Out of the 147 *japonica* landraces, 126 were upland, while out of 36 *indica* landraces only two were upland varieties. Note that sets of identical varieties are shown by a *single line*.



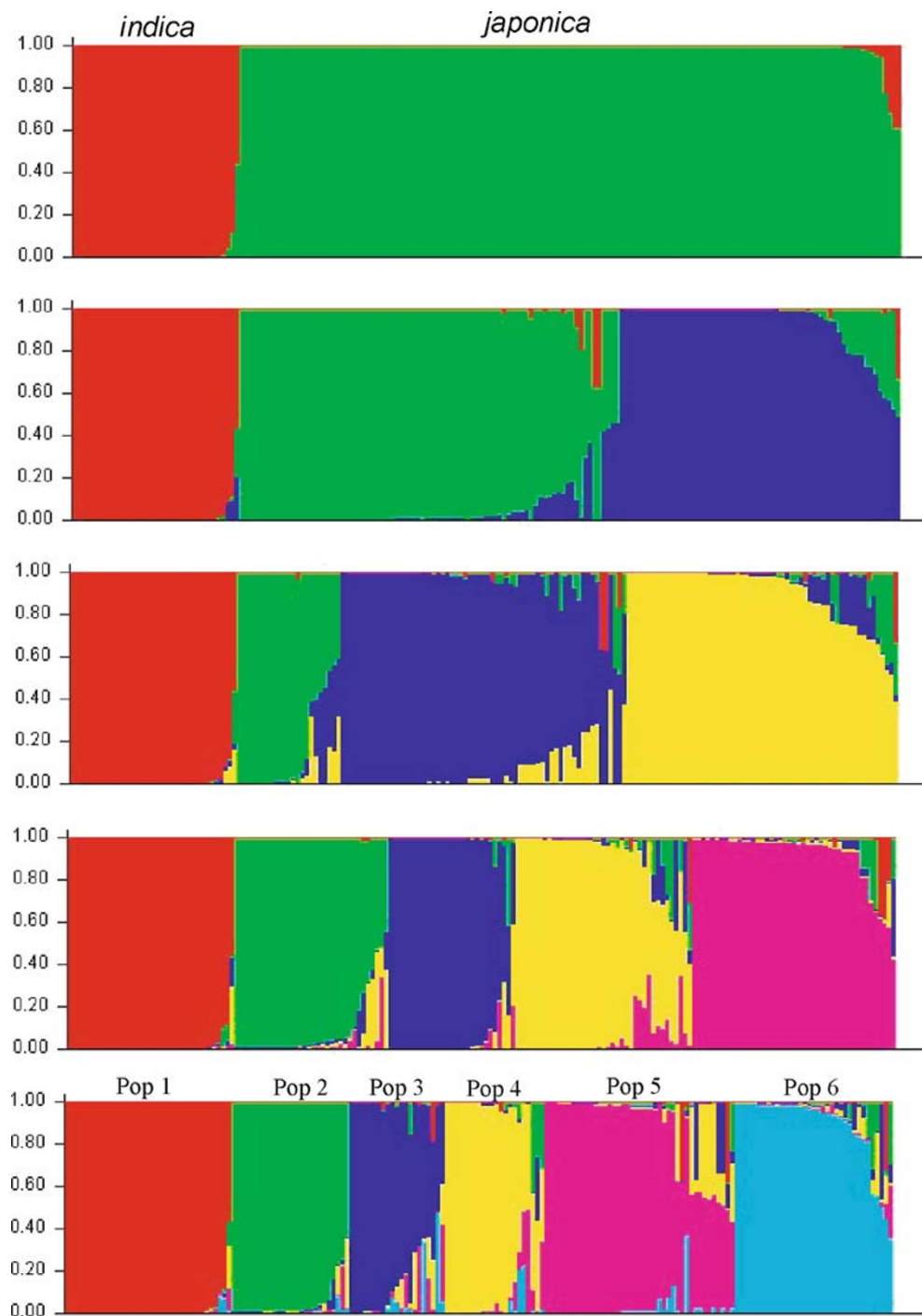
the farmers during the seed collection concerning the meanings and origins of the variety names was recorded and compared with the genetic relationships.

Discussion

The 183 Kalimantan landraces had a lower gene diversity value (0.49) when compared to the gene diversity of 0.68 across a set of 309 Indonesian varieties originating from 20 provinces on eight islands in Indonesia and a value of 0.70 across a set of 234 global rice accessions from 34 countries (Garris et al. 2005; Thomson et al. 2007). This is consistent with the expectation that varieties sampled from a local geographic region would be inherently less diverse since

they are grown in a similar set of environments, whereas germplasm from a wider range of geographic and ecologically distinct regions will contain greater genetic diversity due to divergent selection pressures. An alternate explanation is that there was a greater emphasis placed on including more diverse varieties in the subsets of germplasm chosen to represent larger geographic regions, as samples were selected to encompass the full range of genetic diversity from a much larger germplasm collection. The Kalimantan collection trip, on the other hand, attempted to sample all available varieties from each village with no selection, including varieties of the same name found in different villages. The fact that 17 of the landraces were represented by two or more genetically identical or similar accessions also tends to lower the average diversity across the study (Fig. 2).

Fig. 5 Population structure analysis plots with different K values using STRUCTURE. The top graph at $K=2$ shows inferred ancestries as a proportion of *indica* (darker grey in top panel) and *japonica* (lighter grey in top panel). The other graphs show the same *indica* group, but with further population subdivision within the *japonica* varieties from $K=3$, $K=4$, $K=5$, and $K=6$ (online figure appears in color). See Supplementary Table 3 for inferred ancestries for each variety.



Out of 183 Kalimantan landraces, 20% were *indica* and 80% were *japonica*, which reflects the fact that most of the growing environments in this region were upland fields and 98% of varieties grown in upland fields were *japonica*. In comparison, a study of 548 upland landraces from Brazil, which were previously thought to consist entirely of *tropical japonica* varieties, was found to contain a cluster of 63 accessions with an *indica* genetic background (Pessoa-Filho et al. 2007). A similar proportion was found in Yunnan,

China where 57 out of 692 landraces were designated as upland *indica* varieties (Zhang et al. 2007). While upland *indicās* seem to be quite rare, there are often greater numbers of tropical *japonica* varieties grown in irrigated lowland fields, as 39% of the lowland Kalimantan varieties were *japonica*, while 36% of the Yunnan lowland varieties were *japonica* (Zhang et al. 2007). Of note, nine of the 21 lowland *japonica* landraces clustered together, suggesting a special subgroup of Kalimantan *japonica* landraces that have

Table 2 SSR Diversity and Population Differentiation Across Different Sub-Groups of East Kalimantan Landraces using 30 SSR Loci

Sub-groups	Sample size	Mean no. alleles/locus	Major allele frequency	Mean gene diversity	Mean PIC value	F_{st}
All Kalimantan landraces (<i>indica</i> and <i>japonica</i>)	183	5.0	0.63	0.49	0.44	0.59
<i>Indica</i>	36	2.6	0.79	0.30	0.26	
<i>Japonica</i>	147	4.5	0.74	0.35	0.31	
All Kalimantan landraces (Bahau and Kayan rivers)	183	5.0	0.63	0.49	0.44	0.03
Upper Bahau landraces	82	4.3	0.63	0.49	0.44	
Lower Bahau landraces	45	3.9	0.71	0.38	0.35	
Kayan landraces	56	4.3	0.59	0.53	0.47	

been adapted to irrigated conditions (Fig. 4). In addition, there were no *indica* accessions from Kalimantan clustering near the improved Green Revolution cultivar IR64 (Fig. 2), in contrast to a large cluster of improved Indonesian varieties near IR64 from a previous study (Thomson et al. 2007), which supports the information from the farmers that all 183 accessions are traditional landraces.

The *indica* and *japonica* landraces from Kalimantan were highly differentiated, with an F_{st} of 0.59 between the subspecies. This is even greater than the *indica/japonica* F_{st} of 0.38 previously detected across 309 Indonesian varieties and an *indica/tropical japonica* F_{st} of 0.39 found in a set of global accessions (Garris et al. 2005; Thomson et al. 2007). When divided into two groups (STRUCTURE $K=2$), 35 out of 36 *indica* landraces shared $\geq 95\%$ *indica* ancestry (P. Palapak had 88% *indica*), while 141 out of 147 *japonica* landraces shared $\geq 95\%$ *japonica* ancestry (the remaining six accessions ranged from 44% to 92% *japonica*). In total, 96% of the 183 Kalimantan landraces contained $\geq 95\%$ inferred *indica* or *japonica* ancestry, which is similar to the study of 309 Indonesian-wide accessions, where 95% had contained $\geq 95\%$ inferred *indica* or *japonica* ancestry (Thomson et al. 2007). The low level of *indica/japonica* admixture in Indonesian varieties may be due to a combination of sterility barriers, field level ecotype differences, and visible differences in grain characters which enables farmers to differentiate subspecies and select pure seeds to plant for the next generation.

The Bahau river was originally selected for this collection trip to test if traditional landraces from a geographically isolated region also showed evidence of genetic isolation. The degree of genetic isolation would have implications for using this germplasm for association studies, since a higher amount of linkage disequilibrium (LD) would be predicted within an isolated group, making whole genome LD scans more feasible. The Bahau river region can be considered isolated due to the immense area of forested land in the surrounding area, the extremely low

population density, and the difficulty in traveling between villages. The only mode of transportation is the river: the Kayan river is wide and easily navigable from the ocean up to Long Bia for large boats, relatively easy to travel to Long Pujungan using a longboat, but only passable with small river boats upstream along the Bahau river (Supplementary Fig. 1). In particular the upper Bahau river is especially difficult to pass due to several rapids and variable river depths and traveling between villages is rare for the local villagers due to the high cost of motor fuel in this remote area. However, a test for geographical differentiation using the SSR data did not show any overall differences in the genetic makeup when the accessions were grouped according to the Kayan river, lower Bahau river, and upper Bahau river regions (Table 2). Furthermore, when local farmers were asked about the origins of their varieties, a number of instances of seed exchange between villages and even between regions were noted. Often the names of the varieties themselves reflect their origin: for instance a variety collected in Apau Ping that was named after another village along the Bahau river (P. Pulut Long Tebulo), a variety collected along the Bahau river named after a village in another region (P. Long Liyo), and a variety that was named after another region (P. Krayan). This evidence suggests that even if seed exchanges of traditional varieties between regions are infrequent, over time they have prevented any significant genetic differentiation from occurring. Thus, while isolated natural populations might be useful for association mapping due to higher levels of LD, our evidence suggests that more caution is needed when dealing with cultivated varieties, where geographic isolation does not necessarily lead to genetic isolation.

Nevertheless, there was a cluster on the UPGMA tree of 25 mostly upland, non-glutinous *japonica* accessions that originated only from the Bahau river and did not contain any varieties from the Kayan river (shaded in Fig. 2). This group of varieties may be unique to the more isolated Bahau river region and warrants more in-depth analysis to determine how

these varieties differ from others in this study in terms of genome-wide diversity, decay of linkage disequilibrium, and traits of interest to farmers. While there was little geographic differentiation within East Kalimantan, more research is needed to test if the Borneo rice landraces form a distinct gene pool compared to those of other islands, as would be expected if farmer seed exchanges occurred within the island but not between islands.

Although little or no geographic differentiation was observed, there was enough population genetic structure across the accessions to divide them into six groups, one for *indica* and five for the *japonica* accessions (Supplementary Table 3). One major determinant in dividing the groups seemed to be the field level ecotype, as four of the five *japonica* groups were almost entirely upland (83–100%), while one group was more mixed (50% upland). As the maturity times were not recorded, it is not known if seasonal ecotypes also led to part of the population structure, as was found in the study of rice landraces from Yunnan, China (Zhang et al. 2007). Another factor in the genetic grouping was the grain amylose content: three groups contained 50 to 96% glutinous varieties, while three groups consisted of less than 20% glutinous varieties. In this region the glutinous varieties were often reserved for preparing sweet rice foods for special occasions, and consequently each farmer usually cultivated at least one glutinous variety and three or four non-glutinous varieties each season.

One confounding issue when dealing with genetic resource collections is that the rice varieties are often studied in isolation, disconnected from their original environment and cultural surroundings. Instead of treating each accession as an independent entity, more interaction with the farmers who are the guardians of the traditional landraces can enable investigation into deeper relationships between the varieties and their original context. An essential starting point is knowledge about the names of the rice varieties, which are often in the local languages of the region. In the Bahau river region, each ethnic tribe speaks a Kenyah language variant specific to that group, including Leppo' Ké (in the villages of Apau Ping and Long Tebulo), Leppo' Ma'ut (in Long Alango, Long Kemuat, and Long Berini), and several other variants across the lower Bahau river villages (Eghenter and Sellato 2003). In addition to the diversity of local languages, most people in the region also speak the national Indonesian language, Bahasa Indonesia. Across most of East Kalimantan, the local name for rice is "Pa'dai" which has been added to all of the accessions names in this study (abbreviated as "P.") since that appendage is always used by local farmers when referring to the variety names. When analyzing the Kalimantan landraces, an understanding of the names often helped clarify the genetic relationships between them. For

example, there were a number of times when varieties with different names were tightly clustered, suggesting they were in fact the same variety. In one case the discrepancy is explained by the language difference: two varieties from the Bahau river named "P. Atok" clustered with two varieties from the Kayan region named "P. Ikan;" as it happens, the word for "fish" in the local language is "atok" while in Bahasa Indonesia it is "ikan." Similarly, the name for a glutinous variety in the local language is "pulut" while in Bahasa Indonesia it can be "ketan" or "ubek," which explains the tight linkage between pairs of varieties such as P. Pulut Pute Iting and P. Ubek Pute Iting (Fig. 2).

The collection of multiple accessions from 31 varieties having the same name revealed that surprisingly 40% of the same-named varieties were actually completely different on a genetic level. This has implications for the genetic conservation of rice germplasm, since it supports the idea that each accession is inherently valuable, even if there are multiple accessions with the same name. While varieties of the same name collected in nearby villages were often genetically similar and could be considered members of the same landrace, this was often not true between varieties of the same name from more distant villages. For example, there were four cases where varieties had the same name but different accessions within each variety split across the *japonica* and *indica* division (P. Ketan Merah, P. Krayan, P. Pui, and P. Timai). Again, knowledge obtained from the farmers during the seed collection can help explain such discrepancies. For example, several varieties were named after distant locations, which may indicate that these were recent introductions. The Krayan is another region in East Kalimantan, so it is possible that several different varieties were introduced from this region, but were given the same name of "P. Krayan" by the farmers in the Bahau and Kayan river regions. In addition, the variety names often reflected seed characters, such as "P. Ketan Merah" which means "red, sticky rice" and "P. Timai" which means "small-grained rice" in the local language. The fact that many genetically dissimilar varieties can share the same seed characters (such as red color or small grains) may help explain why multiple accessions of the same name were found to be genetically different. A similar finding has been noted in rice accessions collected in Guinea, where two generic names designating small- and long-grain varieties also showed low genetic consistency across accessions (Barry et al. 2007).

A previous study in 1993 had interviewed 42 farmer households in the village of Apau Ping over a period of 7 weeks and identified 35 distinct rice varieties (Setyawati 2003). The current study collected seed samples from 19 varieties, but did not have time to make a comprehensive survey of all of the farmers in Apau Ping. Seven of the 35

varieties found in 1993 were also collected in Apau Ping in 2005, while ten additional varieties from the 1993 survey were found in other villages along the Bahau river in 2005. Notably, 12 varieties collected in Apau Ping in 2005 were not found in 1993, indicating a rapid introduction of landraces from outside of the village. This supports previous evidence that the farmers in Apau Ping were eager to test new varieties, and would sometimes bring in seeds from outside Apau Ping from other villages (Setyawati 2003). This also helps explain the lack of genetic differentiation between the local regions. The 1993 study also found that farmers would replace a variety if a harvest failure occurred or if it did not perform to expectations, which could lead to a higher turnover rate (Setyawati 2003).

Nonetheless, there were still a few traditional varieties that were considered by the farmers as being very old (such as P. Pute and P. Nyelung Mambat) and were highly regarded as performing well in their environment. Since most of the rice is sown by broadcasting seed on the mountain sides with no subsequent human intervention (no fertilizer, pesticide or irrigation, and with little or no weeding), these varieties must be adapted to withstand a range of abiotic and biotic stresses that naturally occur. Thus, the genetic characterization of these landraces, which is itself a snapshot of a dynamic gene pool of traditional varieties adapted to the local environment and in harmony with cultural preferences, represents an important foundation towards conserving this valuable genetic diversity and applying the knowledge gained towards improving rice varieties for future generations.

Methods

Seed collection

Seeds from 183 rice varieties were collected from the districts of Malinau and Bulungan in the Indonesian province of East Kalimantan during a collection trip from March 15–30, 2005. The collection team, consisting of three researchers (M.J.T., N.R.P., T.S.S.) and a local guide (Samsu), first flew a chartered MAF Cessna from Tarakan to the grassy air strip at Long Alango (Pujungan sub-district, Malinau district). As there are no roads in the region, the team traveled by small river boats to visit villages along the Bahau river, from the most upstream village of Apau Ping to the downstream village of Long Aran (Fig. 1). In each village, seed samples were collected from farmers from bags of recently harvested and stored grain. In a few cases, fields were visited for direct collection of freshly harvested seeds; while farmers did not usually mix varieties in the same field, they often harvested multiple fields at the same time and could

identify each variety based on the seed and panicle appearance (Supplementary Fig. 1). After visiting nine villages along the Bahau river, the team traveled by longboat down the Kayan river, and collected samples from another nine villages along the lower Kayan river (Table 1). In each village, farmers were interviewed to gather information about the specific varieties: if they were grown in an upland field (“ladang” in Indonesian) or in a lowland field (“sawah”), if they were glutinous or non-glutinous, what is the meaning of the name, where the variety originated, and which varieties had special uses.

Seeds from the 183 landrace varieties are available for research purposes through the Indonesian Center for Agricultural Biotechnology and Genetic Resources Research and Development in Bogor, Indonesia and can be requested by contacting the head of the Rice Germplasm Collection (email: ts_silitonga@yahoo.com) and agreeing to an MTA.

SSR marker genotyping

One plant per accession was grown in the greenhouse for DNA extraction. While a single plant was used in this study to maximize the number of different accessions that could be characterized and to define the frequency of true heterozygotes (versus heterogeneity if a bulked sample was taken), additional seeds are available for each accession for future studies on diversity within accessions.

Total genomic DNA was extracted from young leaf tissue after crushing in liquid nitrogen using a Tris/SDS extraction buffer (100 mM Tris–HCl pH 8, 50 mM EDTA pH 8, 500 mM NaCl, 1.25% SDS, 0.38% sodium bisulfate) and chloroform extraction followed by ethanol precipitation. A DNA sample from each accession was then sent to Cornell University for SSR marker analysis. A total of 190 accessions were used for SSR genotyping, including the 183 landraces plus seven additional varieties as controls: one *temperate japonica* (Nipponbare), one *tropical japonica* (Azucena), one *indica* (IR64), one *aus* (Kasalath), and three Indonesian varieties previously studied (Popot, Ketan Siam, and Kantul; Thomson et al. 2007).

Thirty SSR markers were selected as a subset of markers previously used for genetic diversity analysis of *Oryza sativa* (Garris et al. 2005) and were organized into four multiplex panels (Supplementary Table 1). PCR reactions were run in a total volume of 15 μ L consisting of 10 \times PCR buffer, 10 mM dNTP mix, DMSO, 5 μ M forward and reverse primers, 1 U taq polymerase, and 15 ng genomic DNA on an MJ Research thermal cycler with a standard PCR program (initial denaturation 3 min at 94°C; 40 cycles of 30 s at 94°C, 45 s at 55°C, 1 min at 72°C, followed by a 10 min final extension at 72°C). For each SSR marker, the

forward primer was labeled with a fluorescent label and pooled PCR products for each panel were size separated by capillary electrophoresis on an Applied Biosystems 3730xl DNA analyzer. DNA fragment size analysis and allele calling was performed using GeneScan and Genotyper software, followed by manual allele binning. The average percent missing data across the 30 loci is less than 3%. The SSR genotype data for 190 accessions with the 30 SSR markers is available in Supplementary Table 2.

Data analysis

Summary statistics, including the number of alleles per locus, major allele frequency, gene diversity, PIC values, and classical F_{st} values, were determined using POWERMARKER version 3.25 (Liu and Muse 2005). For the unrooted phylogenetic tree, genetic distance was calculated using the “C.S. Chord, 1967” distance matrix (Cavalli-Sforza and Edwards 1967), followed by phylogeny reconstruction using neighbor-joining as implemented in POWERMARKER. The allele frequency data from POWERMARKER was used to export the data in binary format (allele presence=“1” and allele absence=“0”) for analysis with NTSYS-PC version 2.2 (Rohlf 2007). A similarity matrix was calculated with the SIMQUAL subprogram using the Dice coefficient, followed by cluster analysis with the SAHN subprogram using the UPGMA clustering method as implemented in NTSYS-PC. The similarity matrix was also used for principal coordinate analysis with the DCENTER, EIGEN, and MXPLOT subprograms in NTSYS-PC. A model-based cluster analysis was then performed using the program STRUCTURE version 2.2 (Pritchard et al. 2000). The optimum number of populations (K) was selected by testing for $K=1$ to $K=8$ using ten independent runs of 10,000 burn-in runs followed by 100,000 iterations with a model allowing for admixture and correlated allele frequencies (Falush et al. 2003). The average $\ln \Pr(X/K)$ values quickly increased between $K=1$ and $K=2$, then moderately increased up to $K=6$ and began to taper off at $K=7$ and $K=8$ (Supplementary Fig. 2). Since there was no clear indication of which K value provided the best fit for the data, the population subgroups were examined for biological relevance from $K=2$ up to $K=6$. The data set had a clear grouping of individuals into *indica* and *japonica* groups at $K=2$, while further sub-structure was found within the *japonica* varieties at higher values of K (Fig. 5). $K=6$ provided the best distinction between subgroups based on categories of upland/lowland and glutinous/non-glutinous varieties, therefore $K=6$ was used to determine inferred ancestries of the multiple *japonica* subgroups, while $K=2$ was used to determine ancestries when the population structure was limited just to the *indica* and *japonica* subspecies.

Acknowledgments We thank Samsu from LSM-Pionir (Tanjung Selor, Indonesia) for serving as our guide and consultant on the trip, Bapak Anye Apui of Long Alango for hosting the collection team in the Upper Bahau river, Cristina Eghenter (WWF, Tarakan, Indonesia) and Lini Wollenberg (CIFOR, Bogor, Indonesia) for useful advice and discussions. We acknowledge conversations with K. McNally, M. Purugganan, K. Olsen, A. Caicedo, and K. Mather during the planning stage of the collection trip, and we are grateful to Lois Swales for help in formatting this manuscript and the anonymous reviewers for their useful comments. This research was funded in part by a US National Science Foundation (NSF) International Research Fellowship award 0301919 to M.J.T., NSF-Plant Genome Research Program (PGRP) Awards 0110004 and 0606461 to S.Mc., an NSF-Developing Country Collaboration grant through PGRP Award 0319553, and by the support of the Indonesian Center for Agricultural Biotechnology and Genetic Resources Research and Development (Bogor, Indonesia).

References

- Adoukonou-Sagbadja H, Wagner C, Dansi A, Ahlemeyer J, Dainou O, Akpagana K, et al. Genetic diversity and population differentiation of traditional fonio millet (*Digitaria* spp.) landraces from different agro-ecological zones of West Africa. *Theor Appl Genet* 2007;115:917–31.
- Anau N. A history of the Kenyah Leppo' Ké and Nyibun in Pujungan subdistrict. In: Cristina Eghenter, Bernard Sellato, G. Simon Devung, editors. Social science research and conservation management in the interior of Borneo. Bogor, Indonesia: CIFOR (Center for International Forestry Research); 2003. p. 153–174.
- Barnaud A, Deu M, Garine E, McKey D, Joly H. Local genetic diversity of sorghum in a village in northern Cameroon: structure and dynamics of landraces. *Theor Appl Genet* 2007;114:237–48.
- Barry MB, Pham JL, Noyer JL, Courtois B, Billot C, Ahmadi N. Implications for in situ genetic resource conservation from the ecogeographical distribution of rice genetic diversity in Maritime Guinea. *Plant Genet Resour* 2007;5:45–54.
- Caicedo AL, Williamson SH, Hernandez RD, Boyko A, Fledel-Alon A, York TL, et al. Genome-wide patterns of nucleotide polymorphism in domesticated rice. *PLoS Genetics* 2007;3:e163.
- Cavalli-Sforza LL, Edwards AWF. Phylogenetic analysis. Models and estimation procedures. *Am J Hum Genet* 1967;19:233–57.
- Eghenter C, Sellato B. Introduction. In: Eghenter C, Sellato B, Simon Devung G, editors. Social science research and conservation management in the interior of Borneo. Bogor: CIFOR (Center for International Forestry Research); 2003. p. 1–34.
- Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 2003;164:1567–87.
- Gao L-z, Innan H. Nonindependent domestication of the two rice subspecies, *Oryza sativa* ssp. *indica* and ssp. *japonica*, demonstrated by multilocus microsatellites. *Genetics* 2008;179:965–76.
- Gao LiZhi, Zhang ChiHong, Chang LiPing, Jia JiZeng, Qiu ZongEn, Dong YuShen. Microsatellite diversity within *Oryza sativa* with emphasis on *indica-japonica* divergence. *Genet Res* 2005;85:1–14.
- Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S. Genetic structure and diversity in *Oryza sativa* L. *Genetics* 2005;169:1631–8.
- Glaszmann JC. Isozymes and classification of Asian rice varieties. *Theor Appl Genet* 1987;74:21–30.
- Jain S, Jain R, McCouch S. Genetic analysis of Indian aromatic and quality rice (*Oryza sativa* L.) germplasm using panels of fluorescently-labeled microsatellite markers. *Theor Appl Genet* 2004;109:965–77.

- Kovach MJ, Sweeney MT, McCouch SR. New insights into the history of rice domestication. *Trends Genet* 2007;23:578–87.
- Liu K, Muse SV. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 2005;21:2128–9.
- Londo JP, Chiang Y-C, Hung K-H, Chiang T-Y, Schaal BA. Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. *Proc Natl Acad Sci* 2006;103:9578–83.
- Lu H, Redus MA, Coburn JR, Rutger JN, McCouch SR, Tai TH. Population structure and breeding patterns of 145 U.S. rice cultivars based on SSR marker analysis. *Crop Sci* 2005;45:66–76.
- Pessoa-Filho M, Beló A, Alcochete A, Rangel P, Ferreira M. A set of multiplex panels of microsatellite markers for rapid molecular characterization of rice accessions. *BMC Plant Biol* 2007;7:23.
- Prashanth SR, Parani M, Mohanty BP, Talame V, Tuberosa R, Parida A. Genetic diversity in cultivars and landraces of *Oryza sativa* subsp. *indica* as revealed by AFLP markers. *Genome* 2002;45:451–9.
- Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics* 2000;155:945–59.
- Rohlf F (2007) NTSYS-pc: Numerical Taxonomy System, ver. 2.20. In: Exeter Publishing, Ltd, Setauket, NY, USA
- Sang T, Ge S. The puzzle of rice domestication. *J Integr Plant Biol* 2007;49:760–8.
- Setyawati I. Biodiversity and traditional knowledge: rice varieties among the Leppo' Ké of Apau Ping. In: Eghenter C, Sellato B, Simon Devung G, editors. Social science research and conservation management in the interior of Borneo. Bogor, Indonesia: CIFOR (Center for International Forestry Research); 2003. p. 35–48.
- Song ZP, Xu X, Wang B, Chen JK, Lu BR. Genetic diversity in the northernmost *Oryza rufipogon* populations estimated by SSR markers. *Theor Appl Genet* 2003;107:1492–9.
- Sweeney M, McCouch S. The complex history of the domestication of rice. *Ann Bot* 2007;100:951–7.
- Sweeney MT, Thomson MJ, Cho YG, Park YJ, Williamson SH, Bustamante CD, et al. Global dissemination of a single mutation conferring white pericarp in rice. *PLoS Genetics* 2007;3:e133.
- Thomson M, Septiningsih E, Suwardjo F, Santoso T, Silitonga T, McCouch S. Genetic diversity analysis of traditional and improved Indonesian rice (*Oryza sativa* L.) germplasm using microsatellite markers. *Theor Appl Genet* 2007;114:559–68.
- Vaughan DA, Lu B-R, Tomooka N. Was Asian rice (*Oryza sativa*) domesticated more than once? *Rice J* 2008;1:16–24.
- Xu X, Lu B-R, Chen YH, Xu M, Rong J, Ye P, et al. Inferring population history from fine-scale spatial genetic analysis in *Oryza rufipogon* (Poaceae). *Mol Ecol* 2006;15:1535–44.
- Yang GP, Saghai Maroof MA, Xu CG, Zhang Q, Biyashev RM. Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. *Mol Gen Genet* 1994;245:187–94.
- Yu SB, Xu WJ, Vijayakumar CHM, Ali J, Fu BY, Xu JL, et al. Molecular diversity and multilocus organization of the parental lines used in the International Rice Molecular Breeding Program. *Theor Appl Genet* 2003;108:131–40.
- Zhang H, Sun J, Wang M, Liao D, Zeng Y, Shen S, et al. Genetic structure and phylogeography of rice landraces in Yunnan, China, revealed by SSR. *Genome* 2007;50:72–83.
- Zhou H-f, Xie Z-w, Ge S. Microsatellite analysis of genetic diversity and population genetic structure of a wild rice (*Oryza rufipogon* Griff.) in China. *Theor Appl Genet* 2003;107:332–9.