Population structure in Japanese rice population

Yamasaki, Masanori
Ideta, Osamu

(Citation)
Breeding Science,63(1):49-57

(Issue Date)
2013-03

(Resource Type)
journal article

(Version)
Version of Record

(URL)
https://hdl.handle.net/20.500.14094/90002871

Breeding Science 63: 49–57 (2013)
doi:10.1270/jsbbs.63.49

Population structure in Japanese rice population

Masanori Yamasaki*[1] and Osamu Ideta[2,3]

[1] *Food Resources Education and Research Center, Graduate School of Agricultural Science, Kobe University*, 1348 Uzurano, Kasai, Hyogo 675-2103, Japan

[2] *National Agriculture and Food Research Organization, Institute of Crop Science*, 2-1-18 Kannondai, Tsukuba, Ibaraki 305-8518, Japan

[3] *Present address: National Agriculture and Food Research Organization, Western Region Agricultural Research Center*, 6-12-1 Nishifukatsu, Fukuyama, Hiroshima 721-8514, Japan

It is essential to elucidate genetic diversity and relationships among even related individuals and populations for plant breeding and genetic analysis. Since Japanese rice breeding has improved agronomic traits such as yield and eating quality, modern Japanese rice cultivars originated from narrow genetic resource and closely related. To resolve the population structure and genetic diversity in Japanese rice population, we used a total of 706 alleles detected by 134 simple sequence repeat markers in a total of 114 cultivars composed of 94 improved varieties and 20 landraces, which are representative and important for Japanese rice breeding. The landraces exhibit greater gene diversity than improved lines, suggesting that landraces can provide additional genetic diversity for future breeding. Model-based Bayesian clustering analysis revealed six subgroups and admixture situation in the cultivars, showing good agreement with pedigree information. This method could be superior to phylogenetic method in classifying a related population. The leading Japanese rice cultivar, Koshihikari is unique due to the specific genome constitution. We defined Japanese rice diverse sets that capture the maximum number of alleles for given sample sizes. These sets are useful for a variety of genetic application in Japanese rice cultivars.

Key Words: Japanese rice cultivars, simple sequence repeat, genetic diversity, pedigree, population structure, Japanese rice diverse cultivars.

Introduction

In modern breeding of domesticated species, once cross breeding started, breeders obtained an infinitely large number of lines involved in elite cultivars. Consequently, the genetic relationships among newly developed lines and parental or historic cultivars were remarkably complicated, hence it is not easy to demonstrate how cultivars and lines inherited the parental alleles (Sakai 1957). However, the intelligent utilization of cultivars and lines for genetic analysis and plant breeding requires a detailed knowledge of genetic and historical relationship among these plant materials and understanding of the partitioning of genetic diversity among them. For example, crop geneticists exploit these data for genetic analyses (Buckler *et al.* 2009, Liu *et al.* 2003, McMullen *et al.* 2009, Yu *et al.* 2008) and analyses of the domestication and evolutional history (Doebley *et al.* 2006, Wright *et al.* 2005, Yamasaki *et al.* 2005). Breeding program would like to predict which biparental cross populations are appropriate for developing improved lines.

To understand the relationship among related lines and cultivars, given the pedigree records, one of the classical methods is coefficient of parentage (Wright 1922). Coefficient of parentage is simple but does not necessarily coincide with the genetic similarity inferred from the genome level due to artificial selection and genetic drift. Furthermore, pedigree information for some cultivars is incomplete, inaccurate or conflicting. Other methods, molecular markers such as restriction fragment length polymorphisms (RFLPs) and simple sequence repeats (SSRs) are widely used to estimate the relationship among accessions whether the pedigree information is recorded or not. Marker-based relationship such as genetic distance estimated by multilocus genotype data have so far yielded phylogenetic analyses, i.e. distance-based clustering and multivariate analyses such as principal coordinate analyses in rice (Akagi *et al.* 1997, Ebana *et al.* 2008, Garris *et al.* 2005, Hashimoto *et al.* 2004, Ideta *et al.* 2012, Kojima *et al.* 2005, Tabuchi *et al.* 2007). The distance-based clustering is powerful, easy to apply and understand visually. However, most of the results could not satisfactorily elucidate the genetic relationships among parental cultivars and their relatives. The clusters identified could be heavily dependent on both the genetic distance and graphical representation (Pritchard *et al.* 2000). The

---

Communicated by T. Hayashi
Received August 25, 2011. Accepted November 21, 2011.
*Corresponding author (e-mail: yamasakim@tiger.kobe-u.ac.jp)

pairwise genetic distance estimated is usually averaged and unified by all genotype data, and thereby difficult to assess the data of each locus. For phylogenic representation of the relative populations, it is troublesome to evaluate how confident the clusters obtained in this method are meaningful: the bootstrap method proposed by Felsenstein (1985) usually could not support the major clusters in literatures, at times resulting in arbitrary clustering regardless of whether information of predefined populations and geographical sampling locations of lines are available. Thus, distance-based clustering of the relative population has difficulties in identifying the actual subpopulations or subgroups (i.e. population structure) and in assigning lines into these clusters.

Another method to detect cryptic population structure is model-based clustering by Bayesian approach established by Pritchard *et al.* (2000) and Falush *et al.* (2003). Based on the number of groups assumed, each group is characterized by a set of allele frequencies at each locus. The highest likelihood among the number of groups can infer the best estimate. This method demonstrates the presence of population structure, assigning lines to groups or populations and identifying admixed lines. Many crop populations are applied in maize (Liu *et al.* 2003, Yan *et al.* 2009), rice (Garris *et al.* 2005, Zhang *et al.* 2009) and barley (Saisho and Purugganan 2007).

Asian cultivated rice (*Oryza sativa* L.) is a critical crop staple and an autogamous species. Most Japanese rice cultivars belong to the *temperate japonica* group in Asian rice (Garris *et al.* 2005). The progenitors have probably been introduced directly from China or through Korea since no wild rice has been found in Japan (Tabuchi *et al.* 2007). Japanese rice breeding has been conducted based on related cultivars. Since Japanese rice breeding has improved agronomic traits such as yield and eating quality to suit the needs and preference of the Japanese people, the modern Japanese rice cultivars originated from a narrow genetic resource and are closely related. The pedigree information for Japanese rice cultivars is accurately traceable. The pedigree records revealed that most of the cultivars are somewhat related. For future breeding of Japanese rice, a full knowledge of diversity and relationship of the extant Japanese rice population is important. However, the population structure in Japanese rice population is still poorly understood: several phylogenetic analyses could not clearly demonstrate whether or not there is population structure in Japanese rice population (Akagi *et al.* 1997, Hashimoto *et al.* 2004, Ideta *et al.* 2012, Tabuchi *et al.* 2007).

A set of diverse lines and core collection is a limited number of lines/accessions that represent the genetic spectrum in the entire population (Brown 1989, Frankel 1984). For example, several bioassays, DNA polymorphism screening and initial/preliminary tests, researchers need to reduce the number of samples analyzed because of cost considerations (e.g. time, money and labor). Thus, the diverse set and core collection is more efficient in these examination and genetic study. Rice core collections (Ebana *et al.* 2008, Kojima *et al.* 2005, Vaughan 1991) and the sets of diverse lines in maize (Liu *et al.* 2003, Yan *et al.* 2009) are established and widely used.

In the present study, we characterized the genetic diversity and the population structure using SSR markers in Japanese rice population (Ideta *et al.* 2012) to clarify the genetic relationship among Japanese rice cultivars as an illustration of a closely related population. Using the pedigree records, we compared the results by phylogenic (Ideta *et al.* 2012) and our population structure analyses. With the results by population structure analyses and the pedigree records, we examined the leading Japanese cultivar “Koshihikari” as an example. We defined several diverse sets of Japanese rice cultivars.

Materials and Methods

Japanese rice germplasm and simple sequence repeat alleles

We used 114 Japanese rice cultivars composed of 94 improved varieties and 20 landraces (Supplemental Table 1) and a total of 706 alleles at the 134 SSRs (Ideta *et al.* 2012). This population represents a wide geographical range and diversity available among current and historic cultivars used in Japanese rice breeding. Based on their pedigree records, we define an improved cultivar as a developed line by cross breeding and a landrace as an accession of which the progenitor is not recorded on the pedigree and line developed by pure line selection, which is not due to cross breeding. Most Japanese rice cultivars have their pedigree information accurately traceable to progenitorial accessions (Supplemental Table 1). To date, some cultivars ranked higher in the rice planted acreage of Japan. Several cultivars were used for a parental crossing. The pedigree records revealed that there is a relationship in the Japanese rice population. There are two landraces Asahi (no. 99 and 100), which are distinct in Chinese letters.

Data analysis

We used PowerMarker 3.25 (Liu and Muse 2005) to calculate the genetic diversity in the Japanese rice population: number of SSR alleles and gene diversity (expected heterozygosity). Allelic richness per locus was also estimated by Fstat 2.3.9 (Goudet 1995) based on the fixed sample size of 7 cultivars. The significant difference of these estimators measuring genetic diversity was statistically assessed by paired *t*-test.

The software STRUCTURE v2.2 (Falush *et al.* 2003, Pritchard *et al.* 2000) was used to detect population structure and to assign cultivars to subgroups by model-based clustering method. Because these cultivars are highly selfing, we relaxed the assumption of Hardy-Weinberg equilibrium and treated the data as haploid. We allowed for linkage model, considering linkage disequilibrium with admixture and correlated allele frequencies (Falush *et al.* 2003). The posterior probabilities were estimated using Markov chain Monte Carlo method based on 4×10^6 iterations following a burn-in period of 1×10^6 iterations. At least ten runs were performed

Table 1. Summary statistics for all cultivars and each subgroup

	Overall	Kirara 397	Reimei	Nipponbare	Koshihikari	Asahi	Kamenoo	Improved	Landrace
Sample Size	114	11	8	18	34	24	19	94	20
Total number of SSR alleles	706	321	299	285	360	378	562	591	555
Number of alleles per locus	5.3	2.4	2.2	2.1	2.7	2.8	4.2	4.4	4.1
Gene Diversity	0.44	0.35	0.33	0.27	0.32	0.33	0.49	0.41	0.49
Allelic richness		2.3	2.2	1.9	2.2	2.4	3.5	3.5	4.1

by setting number of populations (*K*) from 1 to 8. The final subgroups were determined based on the likelihood plots of these models. The coefficients of each membership determined by this population structure analysis were estimated and compared with pedigree records. The subgroup assignment of the cultivars is based on their highest membership coefficient.

The analysis of molecular variance (AMOVA, Excoffier *et al.* 1992) was performed to partition molecular variance at two levels by Arlequin 3.11 (Excoffier *et al.* 2005): (1) between subpopulations (improved lines and landraces) or among subgroups inferred by population structure analysis and (2) within subpopulations or subgroups. To evaluate genetic differentiation pairwise PhiPT values and Nei's minimum genetic distance between subgroups were respectively calculated by GenAlEx 6.2 (Peakall and Smouse 2006) and Populations 1.2.30 (Langella 2007). The PhiPT value, a Fst analog suppresses within-population variance and simply calculate population differentiation based on the genotypic variance. The probability values estimated by 1000 permutations were used to determine whether the partitioning of variance components was significant.

Several sets of the Japanese rice diverse cultivars (JRDCs) in the defined populations were determined to capture the maximum number of SSR alleles using the Core Set function based on a simulated annealing algorithm in PowerMarker 3.25 (Liu and Muse 2005). In the resultant JRDCs, we allowed conserved cultivars that are agronomicaly and historically important, or representatives of the subgroups. A total of 1000 replicates with different initial subsets were performed.

Results

SSR diversity in Japanese rice population

Based on the 134 SSR markers using 114 Japanese rice cultivars (Ideta *et al.* 2012, Table 1), the overall average values were 5.3 alleles per locus with a range of 2 to 16 alleles/locus and a gene diversity of 0.44. The landraces exhibited greater genetic diversity than improved lines (0.49 versus 0.41, paired *t*-test, $P < 10^{-10}$).

Model-based clustering

We hope to investigate the degree of relatedness among Japanese rice cultivars genetically similar and to identify the population structure. We conducted the model-based clustering approach: The maximum likelihood revealed that the optimum number of population (*K*) inferred to be six (Table 2). The posterior probability at $K = 6$ was by far the highest among the models assuming $K = 1$ to 8. Thus, Japanese rice population was firstly differentiated into six subgroups, designated as important and historic cultivars: Kirara 397, Reimei, Nipponbare, Koshihikari, Asahi and Kamenoo subgroups (Fig. 1). We used Asahi no. 100 as a representative because the membership coefficient characteristic of Asahi type was higher than that of Asahi no. 99. The population could mainly be classified into improved cultivar group composed of four subgroups, Kirara 397, Reimei, Nipponbare and Koshihikari and landrace group comprised of two subgroups, Asahi and Kamenoo. Kamenoo subgroup included 13 landraces whereas Asahi subgroup contained four landraces. The representative improved lines for Asahi and Kamenoo subgroups selected Kinmaze and Yamadanishiki, respectively. We tried further model-based clustering to divide six subgroups into several clusters but obtained no more clusters (data not shown). These six subgroups reflect geographical distribution, breeding history and the differentiation between improved lines and landraces. The admixture situation in many cultivars was observed (Fig. 1).

Table 2. Inferring the *K* value for 114 Japanese rice cultivars with 134 SSRs by model-based clustering method

K	$\ln P(X\|K)$	$P(K\|X)$
1	−13648.1	~0
2	−12450.5	~0
3	−11563.7	~0
4	−11238.0	$1.21E^{-243}$
5	−10877.9	$2.97E^{-87}$
6	−10678.7	0.999
7	−10800.9	$8.19E^{-54}$
8	−11385.1	$1.58E^{-307}$

Admixture situations in related cultivars based on the pedigree of Koshihikari

We confirmed that the defined subgroups and the admixture estimation of the cultivars by model-based clustering (Fig. 1) are largely consistent with the pedigree information (Supplemental Table 1). Since the degrees of the progenitorial alleles inherited to Koshihikari, which is the most popular in Japan, are unclear, we assessed the pedigree of Koshihikari and its progenitorial cultivars and these admixture situations (Fig. 2). Norin 1, in which membership

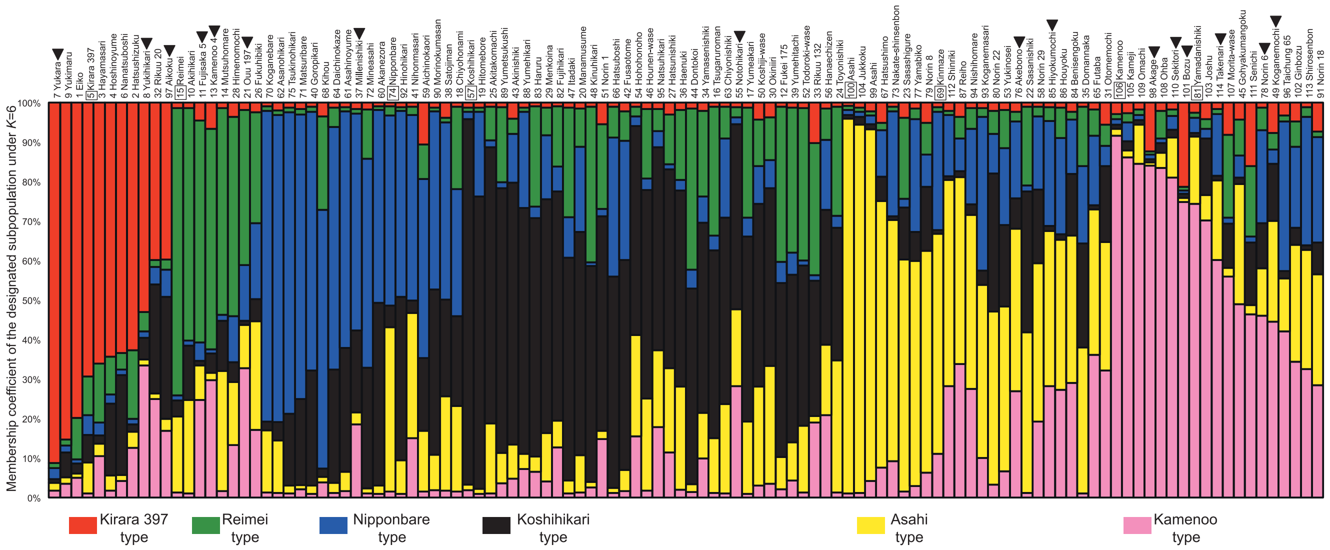

Fig. 1. Model-based clustering ($K = 6$) in 114 Japanese rice cultivars with 134 SSR markers. Color codes indicate typical genotypes of the inferred subgroups: Kirara 397, red; Reimei, green; Nipponbare, blue; Koshihikari, black; Asahi, yellow; Kamenoo, pink. The numbers and cultivar names in Supplemental Table 1 are indicated above the color code. The order of the cultivars is based on the highest membership coefficient. The squared numbers show representatives in the subgroups and also members in 24 Japanese rice diverse cultivars. The cultivars pointed by black triangles indicate members in 24 Japanese rice diverse cultivars.

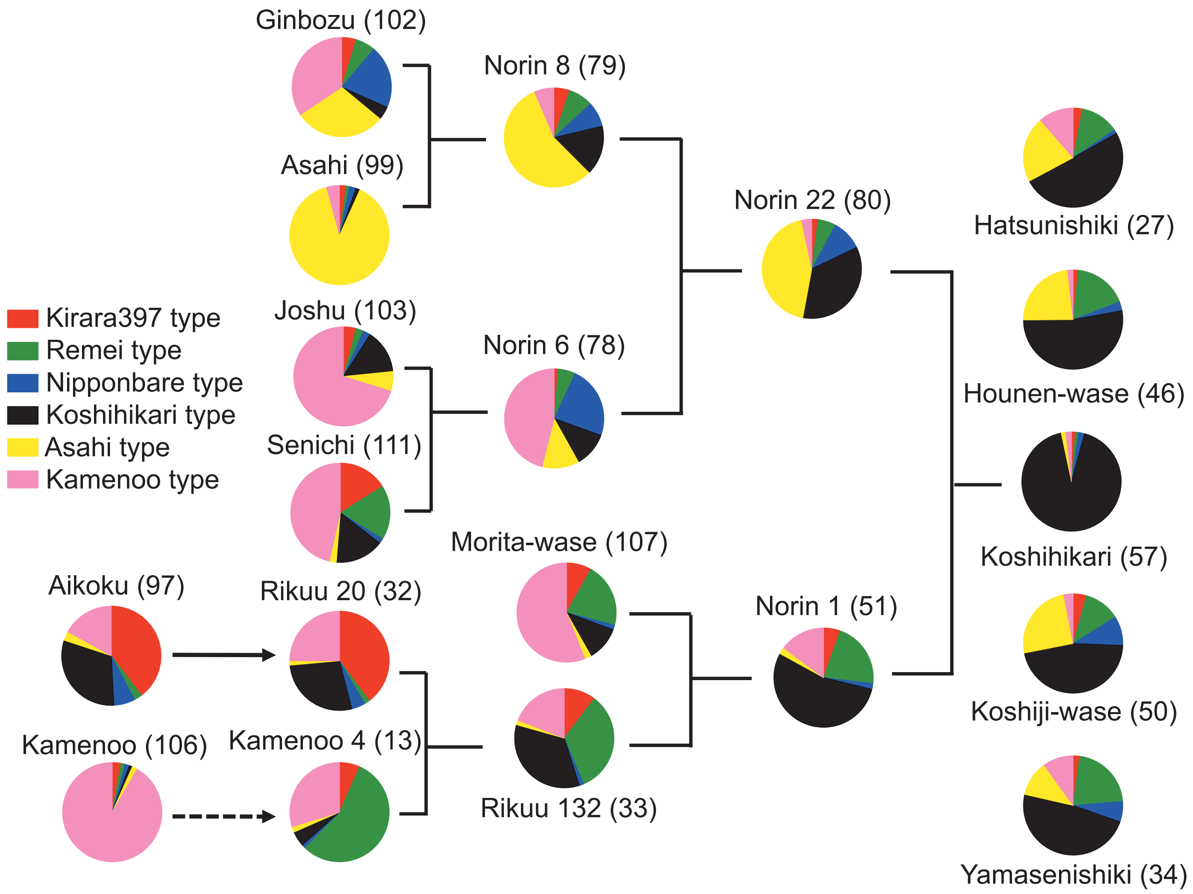

Fig. 2. The pedigree and admixture situation in Koshihikari, the sister cultivars and the progenitorial cultivars.

Table 3. Summary of the hierarchinical AMOVA based on SSR data

	Source of variation	Degree of freedom	Sum of square	Variation (%)
Improved cultivar and landrace	among subpopulation	1	188.8	7.1
	within subpopulation	112	6000.8	92.9
Six subgroups	among subgroups	5	1331.7	21.3
	within subgroups	108	4857.9	78.7

Table 4. Pairwise PhiPT values (above diagonal) and Nei's minimum genetic distance (below diagonal) between subgroups

	Kirara 397	Reimei	Nipponbare	Koshihikari	Asahi	Kamenoo
Kirara 397		0.258	0.356	0.279	0.329	0.157
Reimei	0.173		0.283	0.209	0.203	0.117
Nipponbare	0.201	0.152		0.211	0.166	0.194
Koshihikari	0.154	0.116	0.098		0.219	0.187
Asahi	0.196	0.118	0.078	0.106		0.117
Kamenoo	0.121	0.106	0.119	0.108	0.075	

coefficient of Koshihikari type was 54.2%, largely contributed to Koshihikari than Norin 22 (35.0%). The order of the high value of Koshihikari-type coefficient among the progenitorial landraces was Aikoku (30.9%), Rikuu 20 (27.6%), Senichi (15.9%), Joshu (14.4%) and Morita-wase (11.2%). Meanwhile, Asahi (0.9%), Kamenoo (1.0%), Kamenoo 4 (4.9%) and Ginbozu (4.3%) indicated low membership coefficient of Koshihikari type. The membership coefficients of Kamenoo and Kamenoo 4 are quite different although Kamenoo 4 is a cultivar selected from Kamenoo population in the pedigree. On the other hand, Koshihikari has four sister cultivars, Hatsunishiki, Hounen-wase, Koshiji-wase and Yamasenishiki (Fig. 2). The four admixture situations are similar but different from that of Koshihikari. After we removed several SSRs due to mutation, the percentages of SSR alleles inherited from Norin 1 and Norin 22 in the developed cultivars are not biased: 40 to 51% from Norin 1 and 46 to 58% from Norin 22.

Genetic diversity, molecular variance and relationship between subpopulations and among subgroups

The AMOVA analyses (Table 3) indicated that substantially more genetic variation within subpopulations (92.9%) and subgroups (78.7%) was observed than the genetic variation between improved cultivars and landraces variations (7.1%), and among six subgroups (21.3%). All the subgroup-pairwise PhiPT values were significant ($P < 0.001$, Table 4). The PhiPT value between landrace subgroups Asahi and Kamenoo was smallest whereas the PhiPT value between Kirara 397 and Nipponbare subgroups was largest. Kirara 397 subgroup was highly differentiated from other groups. A similar pattern of differentiation among subgroups was observed using Nei's minimum distance. The PhiPT value and Nei's minimum distance between improved cultivars and landraces were smaller (0.068 and 0.045 respectively, both $P < 0.001$).

The estimates for genetic diversity were calculated for each subpopulations and subgroups defined by population structure analysis (Table 1). Koshihikari subgroup has the largest number of cultivars ($n = 34$) whereas Reimei subgroup has the smallest ($n = 8$). The amount of genetic diversity differed from subpopulations and subgroups. Kamenoo subgroup was the most diverse with 4.2 alleles per locus, gene diversity of 0.49 and allelic richness of 3.5, whereas Nipponbare subgroup showed the smallest with 2.1 alleles, 0.27 (gene diversity) and 1.9 (allelic richness). Landrace subpopulation had more diversity than improved subpopulation.

Japanese rice diverse cultivars sets

We determined several sets of Japanese diverse rice cultivars (JRDCs) that capture the maximum number of alleles for a given sample size (Table 5). In these sets, we constrained the selection to include six cultivars: Kirara 397, Reimei, Nipponbare, Koshihikari, Kinmaze and Yamadanishiki representative of six subgroup in improved cultivars and Asahi and Kamenoo for two landrace subgroups due to population structure results (Fig. 1) and historic/high agronomic importance. The diverse set of six representative-improved cultivars accounted for 56% alleles in improved cultivars and 47% alleles in all cultivars. Other diverse set of the eight JRDCs (six cultivars for improved cultivars and two for landraces) explained 52% alleles in all cultivars. For a balanced set from improved cultivars and landraces, we aimed similar percentage of accounted alleles in improved cultivars and landraces. We set 24 JRDCs, which is approximately one-fifth of all cultivars. Simulated annealing on many data sets revealed that the combination of 16 improved cultivars and 8 landraces covered the same percentage (80%) of accounted alleles in improved cultivars and landraces, respectively. The set of 24 JRDCs accounted for 82% of alleles in all cultivars (Table 5). This set include a large proportion of Kamenoo subgroup (eight cultivars), which is expected since the subgroup have the greatest allelic richness. A total of 69 cultivars composed of 50 improved cultivars and 19 landraces recovered all 706 SSR alleles.

Table 5. Japanese rice diverse cultivars

Population	Improved cultivars		Landraces	Improved cultivars and landraces	
Number of diverse set	Diverse set of 6	Diverse set of 16	Diverse set of 8	Diverse set of 8	Diverse set of 24
	5 Kirara 397	5 Kirara 397		5 Kirara 397	5 Kirara 397
	15 Reimei	15 Reimei		15 Reimei	15 Reimei
	57 Koshihikari	57 Koshihikari		57 Koshihikari	57 Koshihikari
	69 Kinmaze	69 Kinmaze		69 Kinmaze	69 Kinmaze
	74 Nipponbare	74 Nipponbare		74 Nipponbare	74 Nipponbare
	81 Yamadanishiki	81 Yamadanishiki		81 Yamadanishiki	81 Yamadanishiki
		7 Yukara			7 Yukara
		8 Yukihikari			8 Yukihikari
		11 Fujisaka 5			11 Fujisaka 5
		21 Ouu 197			21 Ouu 197
		37 Millenishiki			37 Millenishiki
		49 Koganemochi			49 Koganemochi
		55 Notohikari			55 Notohikari
		76 Akebono			76 Akebono
		78 Norin 6			78 Norin 6
		85 Hiyokumochi			85 Hiyokumochi
			100 Asahi	100 Asahi	100 Asahi
			106 Kamenoo	106 Kamenoo	106 Kamenoo
			97 Aikoku		97 Aikoku
			98 Akage		98 Akage
			101 Bozu		101 Bozu
			13 Kamenoo 4		13 Kamenoo 4
			110 Sekitori		110 Sekitori
			114 Takenari		114 Takenari
Percentage of accounted alleles in improved cultivars	56	80			
Percentage of accounted alleles in landraces			80		
Percentage of accounted alleles in all cultivars	47	67	64	52	82

Discussion

SSR diversity in Japanese rice population

The SSR markers can highly detect polymorphism among rice accessions (Garris *et al.* 2005) and even closely related cultivars such as Japanese rice (Akagi *et al.* 1997). We have analyzed a larger number of Japanese rice cultivars (114) that encompass Japanese rice diversity by 134 SSR markers covering rice genome. Consequently, genome-wide genetic variation in Japanese rice population was clarified in the present study. The overall genetic diversity with a mean gene diversity of 0.44 was lower than that in world rice population (0.7, Garris *et al.* 2005) as previously reported (Akagi *et al.* 1997, Monna *et al.* 2006). On the other hand, our diversity estimate was slightly higher than that in *temperate japonica* derived from other countries as well as Japan (0.39, Garris *et al.* 2005). Our estimate was also larger than the previous report in the Japanese rice population comprised of 171 cultivars mainly involved in the brewing of Japanese rice wine (0.33, Hashimoto *et al.* 2004) using 32 SSRs. One of the reasons is that the sake-brewing cultivars exhibited lower genetic diversity than the cooking cultivars (Hashimoto *et al.* 2004), resulting in the smaller overall diversity than our estimate (0.44). Another reason is likely due to our long-term strenuous searching of a number of polymorphic SSRs specific for the Japanese rice population (Ideta *et al.* 2012). The 134 SSR data will be useful for future Japanese rice breeding and authentication of cultivars.

Population structure in Japanese rice population

Japanese rice cultivars are closely related and have a complex breeding history. These situations made it difficult so far to classify the cultivars into realistic groups that reflect their degree of genetic similarity. We demonstrated the presence of population structure in Japanese rice population and significant differentiation into six subgroups: Kirara 397, Reimei, Nipponbare, Koshihikari, Asahi and Kamenoo subgroups. The differentiation between improved cultivars and landraces was also verified. Recently Yamamoto *et al.* (2010) and Nagasaki *et al.* (2010) supported our results and revealed several population subgroups: Nipponbare, Koshihikari, cultivar group in Hokkaido and Asahi, using 151 Japanese rice cultivars overlapping our sample. Previous studies in Japanese rice population suggested two or more distinct and progenitorial genomes based on sequence and RFLP data (Tabuchi *et al.* 2007) and several clusters using prior population information (Hashimoto *et al.* 2004). Depending on a matter of opinion (e.g. bootstrap value), one may conclude no major cluster in their phylogenetic trees (Akagi *et al.* 1997, Hashimoto *et al.* 2004, Tabuchi *et al.* 2007). Tabuchi *et al.* (2007) revealed the two distinct genomes were originated from Asahi and Jukkoku and from Kamenoo and Aikoku. This result was almost consistent with our classification Asahi and Kamenoo subgroups, but their classification of other Japanese cultivars was unclear.

Six subgroups in Japanese rice population

These six subgroups inferred in the present study well reflect independent breeding history in Japanese rice and geographical distribution. The population structure in Japanese rice may also be due to the autogamous breeding system. Garris *et al.* (2005) predicted a partitioning of diversity among rather than within populations in the absence of human-mediated (e.g. breeding) gene flow between populations. Indeed, the diversity variation estimated among groups in rice (21.3% in this study and 37.5% in Garris *et al.* (2005)) was greater than that in maize inbreds (8.3%, Liu *et al.* 2003). Other lower estimate of 7.1% accounting for the variation between improved cultivars and landraces can be attributed to much utilization for breeding.

We enlarge on the defined representative cultivars in this study (Hashimoto *et al.* 2004, Imbe 2002, Kushibuchi 1997) and the amount of genetic diversity within each subgroup was not equivalent. For landraces, Kamenoo was cultivated in eastern part of Japan, mainly in Tohoku region for 20 years or more, since 1900; Asahi was popular in western part of Japan for 30 years or more, since 1915. Several cultivars belonging to Kamenoo and Asahi subgroups have recently been cultivated. The diversity estimates in landrace are higher than those in improved cultivars, and Kamenoo subgroup exhibited by far the highest diversity. These data strongly suggest that landraces especially belonging to Kamenoo subgroup are important sources of diversity for broadening the genetic base for Japanese rice breeding. The admixture data suggest that a total of 20 landraces consists of at least four landrace-groups: Aikoku (Aikoku and Rikuu 20), Kamenoo 4, Asahi (four cultivars) and Kamenoo groups (13 cultivars) largely contributed towards the development of Japanese rice.

Koshihikari, developed and named by Fukui Prefecture in 1956, has been the most popular for 30 years for its high eating quality and recently accounts for ~40% of Japanese rice acreage. Recent Japanese rice breeding using Koshihikari as one of crossing parents has conducted to introduce its higher eating quality into the developed line. Thus, Koshihikari subgroup composed of its relatives and descendants is the largest and widely distributed over the region except Hokkaido. Tradition has it that high eating quality originated from Asahi (a western representative) and Kamenoo (an eastern one). However, the admixture situation in Koshihikari cannot support the tradition compared with those in Asahi and Kamenoo (Fig. 2). Rather, the other landraces, Aikoku, Rikuu 20, Senichi, Joshu and Morita-wase could highly contribute to the development of Koshihikari. This is the first finding that Aikoku is the top contributor to Koshihikari genome. Koshihikari is a unique cultivar and the breeders' observant eyes and their efforts should be praiseworthy: the admixture situation in Koshihikari is different from those in the four sister cultivars (Fig. 2) although all the sister cultivars equally inherited SSR alleles from Norin 1 and Norin 22. We speculate that the specific genome constitution contributed to the high eating-quality in Koshihikari.

The cultivars in Nipponbare subgroup were mainly developed in Aichi Prefecture and distributed around western and eastern part of Japan. Nipponbare was developed and designated by Aichi Prefecture in 1963. This cultivar was the top ranking from 1970 to 1978 and popular in rice acreage of Japan, within the fifth ranking, until 1995. Several cultivars successfully contained disease resistance to rice blast and rice stripe virus (Sugiura *et al.* 2004). These cultivars together with Koshihikari played essential roles in Japanese rice breeding, but the noteworthy fact is the smallest genetic diversity in the six subgroups. The admixture situation and pedigree records suggest that Nipponbare group benefited from three landraces, Shinriki, Takenari and Shirosenbon.

Rice breeding in Hokkaido, located in northern Japan has been uniquely conducted in Japanese rice breeding: rice cultivars adapted to Hokkaido exhibit extremely early heading and weak or no photoperiod sensitivity to escape cool weather damage. Therefore, the divergence between Hokkaido cultivars and other cultivars was relatively clear because of subgroup pairwise values (Table 4). The representative cultivar, Kirara 397 was registered in 1990 and was the top ranking for more than 15 years in Hokkaido region. In admixture results of Kirara 397 subgroup and pedigree records, Aikoku largely contributed to Hokkaido cultivars as well as Koshihikari. Bozu also benefited Hokkaido cultivars.

Most cultivars in Reimei subgroup are no longer recommended for cultivation in Tohoku region, northern part of Japan. Reimei was developed and named by Aomori Prefecture in 1966. This cultivar was derived from irradiated Fujiminori by gamma-ray; its plant height successfully reduced. Since the genetic diversity was not so small in spite of the smallest number, this subgroup should be important to conserve the genetic diversity of Japanese rice population. Although Kamenoo 4 has a high coefficient of Reimei type, there is no direct relationship between their pedigree: Kamenoo 4 and the improve cultivars in Reimei subgroup.

A representative for improved cultivars of the Asahi group, Kinmaze was developed by Aichi Prefecture in 1948. It was the top ranking from 1958 to 1961 in Japanese rice acreage because of its high eating quality. Yamadanishiki, a representative for improved cultivars of the Kamenoo group, has long been a leading cultivar for Japanese rice wine production. This cultivar was developed and named by Hyogo Prefecture in 1936.

Model-based clustering and distance-based clustering based on pedigree records

Ideta *et al.* (2012) demonstrated that only the distance-based clustering such as neighbor-joining tree was likely to indicate monophyletic clade and could not fully resolve relationships among cultivars. We also underwent the limitation in the distance-based clustering of a related population. Artificial crossing and hybridization among Japanese rice cultivars resulted in complex constitution of inherited alleles, that is to say admixture situation. It is difficult for distance-based clustering to identify the admixed lines and to infer their constitution. It is highly possible that even a related population have a clear and cryptic differentiation. We would like to propose that the model-based clustering method is superior in classifying a related population. Although model-based clustering may consume long time, this method can provide detailed information on the population structure that distance-based clustering cannot support.

The admixture situations in cultivars estimated by model-based clustering are able to support whether to be consistent or conflicting with pedigree records. Our model-based clustering analysis estimated the admixture situation in Japanese rice cultivars, showing good agreement with pedigree information but there are some inconsistencies. We clarified that Kamenoo and Kamenoo 4 are different whereas Aikoku and Rikuu 20 are similar in genome constitution by population structure analysis (Fig. 2). Now that things have come to this pass, the reason for the discrepancy is unclear. The molecular markers such as SSRs are useful for cultivar identification and to discriminate redundant or different accessions with the same variety name in germplasm bank (Ebana *et al.* 2008).

Japanese rice diverse cultivars sets

We have defined sets of JRDCs to represent maximum diversity in a given set of improved lines and landraces. These JRDCs absolutely include representative cultivars in Japanese rice population and are well balanced from improved lines and landraces. Each JRDCs set is convenient for 96 or 384-well PCR plate in DNA polymorphism screening. We can determine new JRDCs sets suitable for users. These sets are also useful for future research targets such as allele mining and genome-wide association mapping.

Acknowledgment

We are grateful to National Agriculture and Food Research Organization: National Agricultural Research Center, National Agricultural Research Center for Hokuriku Research Center, National Institute of Crop Science, National Agricultural Research Centers for Hokkaido Region, Tohoku Region, Western Region and Kyushu Okinawa Region; National Institute of Agrobiological Sciences, Kyushu University and Prefecture Agricultural Experiment Centers or Stations: Hokkaido, Aomori, Miyagi, Akita, Yamagata, Ibaraki, Gunma, Chiba, Niigata, Ishikawa, Fukui, Gifu, Aichi, Okayama, Yamaguchi, Fukuoka, Kumamoto, Miyazaki and Kagoshima for rice seeds. We also thank Dr. Hiroyoshi Iwata for much advice on statistical analyses, Dr. Takeshi Honda for their facility for the use of a high-performance computer and Dr. Paulino Laude Sanchez for his help in correcting the English manuscript. This work was supported by Hyogo Science and Technology Association.

Literature Cited

Akagi,H., Y.Yokozeki, A.Inagaki and T.Fujimura (1997) Highly polymorphic microsatellites of rice consist of AT repeats, and a classification of closely related cultivars with these microsatellite loci. Theor. Appl. Genet. 94: 61–67.

Brown,A.H.D. (1989) The case for core collection. *In*: Brown,A.H.D., O.H.Frankel, D.R.Marshall and J.T.Williams (eds.) The Use of Plant Genetic Resources, Cambridge University Press, Cambridge, pp. 136–156.

Buckler,E.S., J.B.Holland, P.J.Bradbury, C.B.Acharya, P.J.Brown, C.Browne, E.Ersoz, S.Flint-Garcia, A.Garcia, J.C.Glaubitz *et al.* (2009) The genetic architecture of maize flowering time. Science 325: 714–718.

Doebley,J.F., B.S.Gaut and B.D.Smith (2006) The molecular genetics of crop domestication. Cell 127: 1309–1321.

Ebana,K., Y.Kojima, S.Fukuoka, T.Nagamine and M.Kawase (2008) Development of mini core collection of Japanese rice landrace. Breed. Sci. 58: 281–291.

Excoffier,L.G., P.E.Smouse and J.M.Quattro (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131: 479–491.

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

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

Felsenstein,J. (1985) Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39: 783–791.

Frankel,O.H. (1984) Genetic perspectives of germplasm conservation. *In*: Arber,W.K., K.Llimensee, W.J.Peacock and P.Stralinger (eds.) Genetic Manipulation: Impact on Man and Society, Cambridge University Press, Cambridge, pp. 161–170.

Garris,A.J., T.H.Tai, J.Coburn, S.Kresovich and S.McCouch (2005) Genetic structure and diversity in *Oryza sativa* L. Genetics 169: 1631–1638.

Goudet,J. (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. J. Heredity 86: 485–486.

Hashimoto,Z., N.Mori, M.Kawamura, T.Ishii, S.Yoshida, M.Ikegami, S.Takumi and C.Nakamura (2004) Genetic diversity and phylogeny of Japanese sake-brewing rice as revealed by AFLP and nuclear and chloroplast SSR markers. Theor. Appl. Genet. 109: 1586–1596.

Ideta,O., I.Kono, Y.Takeuchi, H.Hirabayashi, M.Hirayama, H.Ohta, H.Sato, I.Ando, H.Kato, H.Nemoto *et al.* (2012) Diversity and relationships between coefficient of parentage and genetic distance estimated by SSR markers in Japanese rice cultivars. Breed. Res. 14: 106–113.

Imbe,T. (2002) Rice breeding in Japan. Res. J. Food and Agric. 25: 35–41.

Kojima, Y., K. Ebana, S. Fukuoka, T. Nagamine and M. Kawase (2005) Development of an RFLP-based rice diversity research set of germplasm. Breed. Sci. 55: 431–440.

Kushibuchi, K. (1997) *In*: Science of the Rice Plant Volume 3 Genetics, Matsuo, T., Y. Futsuhara, F. Kikuchi and H. Yamagata (eds.) Nobunkyo, Tokyo.

Langella, O. (2007) Populations 1.2.30: Population genetic software (individuals or populations distances, phylogenetic trees). http://bioinformatics.org/~tryphon/populations/

Liu, K., M. Goodman, L. Muse, J.S. Smith, E. Buckler and J. Doebley (2003) Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. Genetics 165: 2117–2128.

Liu, K. and S.V. Muse (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21: 2128–2129.

McMullen, M.D., S. Kresovich, H. Sanchez Villeda, P. Bradbury, H. Li, Q. Sun, S. Flint-Garcia, J. Thornsberry, C. Acharya, C. Bottoms *et al.* (2009) Genetic properties of the maize nested association mapping population. Science 325: 737–740.

Monna, L., R. Ohta, H. Masuda, A. Koike and Y. Minobe (2006) Genome-wide searching of single-nucleotide polymorphisms among eight distantly and closely related rice cultivars (*Oryza sativa* L.) and a wild accession (*Oryza rufipogon* Griff.). DNA Res. 13: 43–51.

Nagasaki, H., K. Ebana, T. Shibaya, J. Yonemaru and M. Yano (2010) Core single-nucleotide polymorphisms-a tool for genetic analysis of the Japanese rice population. Breed. Sci. 60: 648–655.

Peakall, R. and P.E. Smouse (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Eco. Notes 6: 288–295.

Pritchard, J.K., M. Stephens and P. Donnelly (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945–959.

Saisho, D. and M.D. Purugganan (2007) Molecular phylogeography of domesticated barley traces expansion of agriculture in the old world. Genetics 177: 1765–1776.

Sakai, K.I. (1957) Theoretical studies on plant breeding technique. V. Coefficient relationship in the breeding of autogamous plants. Jpn. J. Breed. 7: 87–92.

Sugiura, N., T. Tsuji, K. Fujii, T. Kato, N. Saka, T. Touyama, Y. Hayano-Saito and T. Izawa (2004) Molecular marker-assisted selection in a recurrent backcross breeding for the incorporation of resistance to rice stripe virus and panicle blast in rice (*Oryza sativa* L.). Breed. Res. 6: 143–148.

Tabuchi, H., Y.I. Sato and I. Ashikawa (2007) Mosaic structure of Japanese rice genome composed mainly of two distinct genotypes. Breed. Sci. 57: 213–221.

Vaughan, D.A. (1991) Choosing rice germplasm for evaluation. Euphytica 54: 147–154.

Wright, S. (1922) Coefficients of inbreeding and relationship. Amer. Nat. 56: 330–338.

Wright, S.I., I. Vroh Bi, S.G. Schroeder, M. Yamasaki, J.F. Doebley, M.D. McMullen and B.S. Gaut (2005) The effect of artificial selection on the maize genome. Science 308: 1310–1314.

Yamamoto, T., H. Nagasaki, J. Yonemaru, K. Ebana, M. Nakajima, T. Shibaya and M. Yano (2010) Fine definition of the pedigree haplotypes of closely related rice cultivars by means of genome-wide discovery of single-nucleotide polymorphisms. BMC Genetics 11: 267.

Yamasaki, M., M.I. Tenaillon, I. Vroh Bi, S.G. Schroeder, H. Sanchez-Villeda, J.F. Doebley, B.S. Gaut and M.D. McMullen (2005) A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. Plant Cell 17: 2859–2872.

Yan, J., T. Shah, M.L. Warburton, E.S. Buckler, M.D. McMullen and J. Crouch (2009) Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. PLoS One 4(12): e8451.

Yu, J., J.B. Holland, M.D. McMullen and E.S. Buckler (2008) Genetic design and statistical power of nested association mapping in maize. Genetics 178: 539–551.

Zhang, D., H. Zhang, M. Wang, J. Sun, Y. Qi, F. Wang, X. Wei, L. Han, X. Wang and Z. Li (2009) Genetic structure and differentiation of *Oryza sativa* L. in China revealed by microsatellites. Theor. Appl. Genet. 119: 1105–1117.