Breeding Science 68: 614–621 (2018)
doi:10.1270/jsbbs.16105

Research Paper

Diverse genetic variation in maternal lineages with high heterogeneity among *in situ*-conserved wild rice (*Oryza rufipogon* Griff.) developed in Thailand

Reunreudee Kaewcheenchai[1)], Udompan Promnart[2)], Kasem Soontrajarn[3)], Somsong Chotechuen[4)], Songkran Chitrakon[4)], Honami Yuki[5)], Seiya Saito[5)], Yo-Ichiro Sato[6)] and Ryuji Ishikawa*[5)]

[1)] *Suphan Buri Rice Research Center, Bureau of Rice Research and Development, Rice Department*, Suphan Buri 72000, Thailand
[2)] *Prachin Buri Rice Research Center, Bureau of Rice Research and Development, Rice Department*, Prachin Buri 25150, Thailand
[3)] *Pathum Thani Rice Research Center, Bureau of Rice Research and Development, Rice Department*, Rangsit, Thanyaburi, Pathum Thani 12110, Thailand
[4)] *Bureau of Rice Research and Development, Rice Department*, Ladyao Chatuchak, Bangkok 10900, Thailand
[5)] *Faculty of Agriculture and Life Science, Hirosaki University*, Hirosaki, Aomori 036-8561, Japan
[6)] *Kyoto Prefectural University*, Kyoto 606-8522, Japan

Wild rice, *Oryza rufipogon*, is a genetic resource that can be used to improve cultivated rice, but its populations are now decreasing in terms of both size and number. Extensive research on wild rice has been conducted in Thailand, where two *in situ* conservation sites have been preserved in natural areas where perennial wild rice predominates. The genetic structure of wild rice populations was investigated by examining both the chloroplast and nucleus genomes at sites of *in situ* conservation site in Thailand. One accession from an *in situ*-conserved site was re-sequenced against the chloroplast genome of *O. sativa* cv. 'Nipponbare' to develop chloroplast insertion/deletion (cpINDEL) markers. These cpINDEL markers revealed unique maternal lineages in the *in situ*-conserved populations upon comparison with other Asian wild rice accessions. Diverse genetic variation was also detected with SSR markers throughout the genome. Three populations differed from each other and also within single populations. The sub-populations within an *in situ*-conserved population showed a complex population structure due to their multiple maternal lineages and relatively higher number of haplotypes when they maintained a relatively large population size. Such a heterogeneous population would serve as a unique gene pool for rice breeding.

Key Words: *Oryza*, AA genome, chloroplast genome, next-generation sequencing, maternal lineage.

Introduction

Wild grasses maintain wide genetic diversity and are regarded as valuable gene pools (Buckler *et al.* 2001). However, details of the diversity in wild populations of individual grass families or in single populations under different environmental and geographical conditions remain unclear. Representative samples have been compared using a variety of molecular tools (Hao *et al.* 2015, Huang *et al.* 2012, Londo *et al.* 2006, Nonomura *et al.* 2010, Sun *et al.* 2002). Some studies have focused on natural populations or randomly sampled collections representing the original natural populations. It is difficult to know whether natural net diversity can be recovered using any of the sampling methods currently available. In addition, there are so many severe situations for wild rice to survive under natural condition they did ever due to quick development of rural areas. The diversity of any given gene bank system may not include all the current wild populations include in nature.

Thailand is one of the countries where rice is the staple food. Recently acquired genomic tools have enabled rice breeders to apply wild rice species, especially *Oryza rufipogon*, for improvement of traits in cultivated rice (Brar and Khush 1997). In fact, some famous varieties they are now distributed worldwide, including the IR series released from the IRRI (International Rice Research Institute), some of which have been developed using wild rice as the parental line. Wild rice species are distributed from China to India, and are further expanding to Australia (Henry *et al.* 2010, Sotowa *et al.* 2013, Vaughan 1994). Wild rice accessions collected up to now are also regarded as parental donors for improvement of cultivated rice, and for conferring biotic and abiotic tolerance (Brar and Khush 1997).

Communicated by Hideshi Yasui
Received June 24, 2016. Accepted September 18, 2018.
First Published Online in J-STAGE on November 28, 2018.
*Corresponding author (e-mail: ishikawa@hirosaki-u.ac.jp)

Breeding Science
Vol. 68 No. 5 **BS**

Several related species such as *O. ridleyi*, *O. officinalis*, and *O. rufipogon* are found in Thailand, and huge natural populations have been observed especially for *O. rufipogon*. The Thailand Rice Department has recognized the importance of these species, and along with Japanese rice geneticists has started to conserve valuable genetic resources including not only landraces but also wild relatives (Morishima *et al.* 1983). In the past, abundant wild rice populations were known to be present, even in central areas of Bangkok. However, the recent development of infrastructure has driven many local populations to extinction (Akimoto *et al.* 1999, Nonomura *et al.* 2010), and it is feared that such resources may be lost within the next few decades. Her Royal Highness Princess Maha Chakri Sirindhorn of Thailand has encouraged geneticists to conserve wild resources *in situ* within rice research centers. The first wild rice conservation site was established at Pathum Thani Rice Research Center, and later other populations were preserved, such as the perennial population at Prachin Buri Rice Research Center. Even now it is still possible to see wild populations in natural lakes in northeastern areas of Thailand, such as Nong Han Lake in Sakon Nakhon.

Laos is another country where an *in situ*-conserved population has been developed through collaboration between local and Japanese researchers (Kuroda *et al.* 2007). The site in Laos has been examined for its genetic structure at the molecular level (Wang *et al.* 2012). Annual and perennial rice plants grow together in a shallow area, and only perennial plants are present in the deep water area at the center. As annual and perennial types represent different lifestyles and cycles, gene flow between the two types is limited to the periphery of the pond at the *in situ* conservation site in Laos. In addition, several alleles are restricted to some sites within the pond. This uneven distribution of alleles may help to maintain the genetic structure that has been estimated in one Chinese population (Qian *et al.* 2005, Song *et al.* 2003).

Chinese natural wild rice populations have also been intensively studied and some of them have been transferred for *ex situ* conservation (Gao 2004, Gao *et al.* 2002). Although high genetic diversity was observed, reduction of the population resulted in loss of genetic diversity (Gao *et al.* 2000). Therefore, some parts of these natural populations have been transferred to *ex situ*-conserved site to maintain their genetic diversity. Song *et al.* (2003) and Qian *et al.* (2005) have suggested that conserved populations do not retain sufficient variation in some cases because of the way in which they have been collected, and that recovery of genetic diversity can only be achieved with careful sampling that ensures genetic diversity is sufficiently representative. In order to conduct such efficient sampling, it is necessary to know the details of each population structure and to develop tools for precise assessment of genetic resources. Wild rice in Cambodia has also been studied to clarify the natural population structure (Orn *et al.* 2015). Different habitats characterized by annual and perennial wild rice plants were shown to have resulted in different ranges of genetic diversity. Ten years after the first observation, some sites had vanished due to development of rural infrastructure. As this type of situations will likely occur in many South-Asian countries, it will be necessary to acquire a better understanding using efficient analytic tools in order to conserve a sufficient degree of variation in *ex situ* conditions such as gene banks or *in situ*-conserved sites that are publicly maintained.

In the present study, we focused on wild rice populations in Thailand, in order to compare them with another mixed population of annual and perennial plants in Laos using the molecular markers. Perennial populations predominated in Thailand. Generally, perennial populations tend to maintain a higher degree of heterozygosity and genetic diversity, and show few differences in diversity among populations relative to annual populations (Oka 1988). These characteristics allow perennial populations to serve as a buffer, maintaining diverse genetic compositions including traits that are valuable for rice breeding. However, the genetic structures within such perennial populations have not yet been studied in detail.

Materials and Methods

Plant materials

Wild rice accessions were collected from Nong Hang Lake in Sakhon Nakon (NHL, sample size n = 32), Prachin Buri (PCB, n = 28), and Pathum Thani (PTT, n = 24), in 2013 (**Fig. 1**, **Table 1**). All three populations were composed of perennial individuals. Two of these sites, Prachin Buri Rice Research Center and Pathum Thani Rice Research Center, belonging to the Rice Department of Thailand, are *in situ*-conserved sites. PTT is a small swamp 5 m × 50 m in size. NHL is a large natural swamp, but only 32 accessions were collected from small areas within it, each at least 1 m distant from the others. PTB is a relatively larger natural water reservoir 600 m × 600 m in size. Among the PCB populations, 28 sub-populations were partitioned as plots measuring 2 m × 2 m. From each subpopulation, 32 accessions were collected, and one of each was randomly selected as a representative of the subpopulations. Three of the 28 sub-populations (Sites 1, 14, and 28) were further analyzed to clarify their genetic composition. Detailed geological information based on GPS is shown in **Supplemental Table 1**. A core collection in the National BioResource Project in Japan (Nonomura *et al.* 2010) was applied as a control population in order to evaluate genetic polymorphism as NBR (**Supplemental Table 2**). Individuals in the core collection are composed of representative *O. rufipogon* accessions collected from South and South-east Asia.

Novel chloroplast INDEL markers determined on the basis of whole-genome sequence data

One perennial *O. rufipogon* wild rice accession 45-2 belonging to the *in situ* population at PCB, has been shown

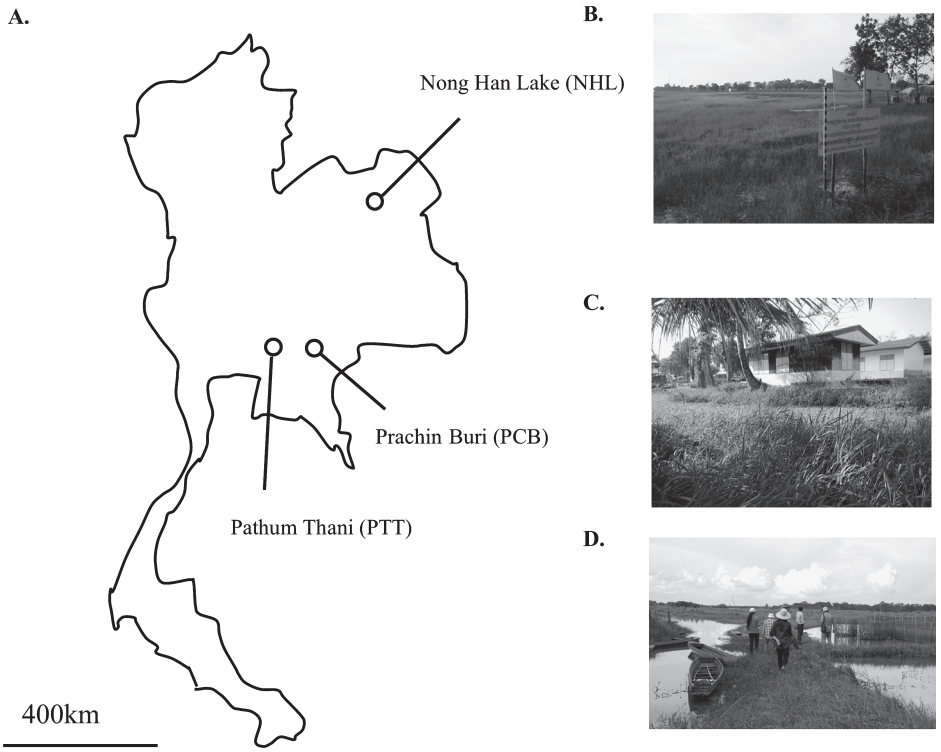

Fig. 1. Three natural wild rice populations composed of *Oryza rufipogon*. (A) Three sites in Thailand. (B) Prachin Buri *in situ* conservation site preserved at Prachin Buri Rice Research Center, where the natural population is deep water rice. In the rainy season, more than 3 m of water covers the population. (C) Pathum Thani *in situ* conservation site preserved at Pathum Thani Rice Research Center. (D) Natural wild rice population in Non Hang lake. Wild accessions were collected at the lake side.

Table 1. cpINDEL markers and nuclear SSRs examined in this study

			Genome position (bp)*	Forward primer	Reverse primer	Remarks
cpINDEL						
	cpINDEL1	cp genome	12670..12673	GGATTCACCGAAACAAACAACC	GCCAAATTGAGCAGGTTGCG	Nipponbare-45-2
	cpINDEL2	cp genome	14012..14013	TTTGGGGAAGAAAACATCTTCC	TAAACGGAGAGAATCGACTAAG	Nipponbare-45-2
	cpINDEL3	cp genome	17380..17385	AATTGCTCTCACCGCTCTTTC	TAGTCGAATTGTTGTATCAACTC	Nipponbare-45-2
	cpINDEL4	cp genome	46087..46091	TAATTTGATATGGCTCGGACG	TGCTATGATTCTATGTTCTCC	Nipponbare-45-2
	cpINDEL5	cp genome	46534..46539	AGATGGAGGAAATTGCACAAGG	CAAAACATGGATTTGGCTCAGG	Nipponbare-45-2
	cpINDEL8	cp genome	57644^57645	TTTTACAGGAGTATCTAGTTGG	ATTACCTCTTTTTCGAGAACC	Nipponbare-45-2
	cpINDEL9	cp genome	60865^60866	AAATCCTTTTAGGAGGGATTG	TCCACTACATCGCCTGAACC	Nipponbare-45-2
	cpINDEL12	cp genome	77735^77736	TGTCTTTCCAGAAAGAAGAACC	TTGTTAAACCAGGTCGAATAC	Nipponbare-45-2
SSR	RM3604	1	5139420	ATGTCAGACTCCGATCTGGG	TCTTGACCTTACCACCAGGC	McCouch (2002)
	RM1347	2	5314190	TCTTGACCTTACCACCAGGC	GTCTTATCATCAGAACTGGA	McCouch (2002)
	RM3180	3	18264873	GGGTCGGATAGCCACACAC	GAGGTAATCTCGCGGAGTTG	McCouch (2002)
	AL606650	4	31892264	CACATAGACCGAAATCGGGG	GACGGTAGGTAAAGTACAATC	Wang *et al.* (2012)
	+29CAT	6	30917713	CACGATCTAGAAGACGAGAG	CCAAATTACGCCTTCCTACC	Wang *et al.* (2012)
	RM125	7	5478776	TCAGCAGCCATGGCAGCGACC	ATGGGGATCATGTGCCGAAGGCC	McCouch (2002)
	SSR-chr8	8	2882902	CAGATATTCCGAAAATCTCAGG	CTCATTGTGAACTCCTCAAC	Flanking to *Pi36* locus
	SSR-chr9	9	9688857	AATGCACTATGCATATGGTC	AACAAGAGCAATTTTAGCAC	Flanking to *Pi5* locus
	RM311	10	9487264	TGGTAGTATAGGTACTAAACAT	TCCTATACACATACAAACATAC	McCouch (2002)
	SSR-Chr12	12	10619152	ATGGATTAGAGCGTAATTG	TGTGTATGGATGGATGCATCA	Flanking to *Pita* locus
	RM17	12	26954668	TGCCCTGTTATTTTCTTCTCTC	GGTGATCCTTTCCCATTTCA	McCouch (2002)

* Deletion site between two positions was shown as "..". Insertion between nucleotides was revealed by "^".
# cpINDELs were detected in chloroplast genomes of Nipponbare and Acc45-2 (Thai wild rice accession).

to have high tolerance to acid sulfate soil and resistance to bacterial leaf streak (data not published). The accession 45-2 accession was subjected to next-generation sequencing (NGS) in order to reveal any novel INDELs. The genomic DNA was extracted from mature leaves using a DNAeasy Plant Mini Kit (QIAGEN Co., Japan) for sequencing with Illumina-Hiseq as 100-bp pair-end-reads. In total, 62,762,232 raw reads were subjected to re-sequencing using CLC-workbench genomics (CLC Bio Japan Inc., Japan) against the complete chloroplast genome of 'Nipponbare'

(GU592207.1). Insertions or deletions (INDELs) were then screened with an average coverage of 2000x. Variants were detected when there were over 100 counts of each variant, and the frequencies were more than 50%. INDELs of more than 2~6 bp were reconfirmed by amplification and electrophoresis. Eight INDELs were constantly amplified and the presumed polymorphisms were confirmed between 45-2 and 'Nipponbare'. These eight INDELs were applied as cpINDELs to clarify the plastid polymorphisms (**Table 1**, **Supplemental Table 3**).

Molecular markers

Eight SSR primer pairs were used to genotype wild rice in order to characterize the wild populations (**Table 1**). The PCR conditions employed were the same as those reported by Wang *et al.* (2012). In addition to the SSR markers, three SSR loci were developed and applied. These PCR products were mixed with loading dye, separated on 6% denaturing polyacrylamide gel, and subjected to electrophoresis at 1500 V for 2 hours in 0.5 × TBE. The gels were finally subjected to silver staining to visualize the DNA fragments (Promega Co., Japan).

Data analysis

Parameters for evaluating genetic variations in loci and populations were calculated, including the number of different alleles per locus (Na) and the expected and observed heterozygosity (*He* and *Ho*), using the GenAlEx 6.501 software package (Peakall and Smouse 2006). Genetic distances among populations determined using principal component analysis (PCA) were also calculated using GenAlEx 6.501, and a dendrogram were constructed using the neighbor joining method based on Nei's unbiased genetic distances employing the Population program, which is a free software package that can be downloaded at http://bioinformatics.org/~tryphon/populations/#ancre_bibliographie (Nei *et al.* 1983). A dendrogram was viewed in TreeExplorer software, where can show and edit a population dendrogram, supplied as a free software with MEGA at http://www.megasoftware.net/ (Tamura *et al.* 2011).

Results

Chloroplast genome

One of the wild rice accessions in the PCB *in situ* conservation site (45-2) was subjected to re-sequencing against the 'Nipponbare' chloroplast genome. In total, 37 INDELs were presumed, varying from one to six nucleotides (**Supplemental Table 3**). Replacement with size differences was also listed among the INDELs. Eight INDEL markers among those expected were chosen for evaluation of diversity in maternal lineages. The expected INDEL genotypes were reconfirmed by gel electrophoresis followed by PCR amplification. Genotyping results were matched to the expected size variations between 'Nipponbare' and 45-2. These INDEL markers were used to genotype 32 accessions in the NBR core collection and three populations in Thailand.

In total, there were 14 INDEL combinations for the eight INDEL markers examined (**Table 2**). These were regarded as plastid types and used to evaluate maternal diversity and also trace maternal lineages (**Table 3**). Reads obtained from local wild rice accession made it possible to distinguish various forms of plastid diversity as INDEL differences. 'Nipponbare' was used as a control for determination of plastid genotype, carrying Plastid type-13. All presumed INDELs were directly compared with 45-2 and 'Nipponbare'. Another control accession, 45-2, carried Plastid type-1. The

Table 2. Plastid types detected with eight cpINDELs

Locus	Plastid types													
	–1	–2	–3	–4	–5	–6	–7	–8	–9	–10	–11	–12	–13	–14
cpINDEL1	1	1	1	1	1	1	1	1	2	2	2	2	2	2
cpINDEL2	1	1	1	1	1	1	2	2	1	1	2	2	2	2
cpINDEL3	1	1	2	2	2	2	2	2	1	2	1	1	2	3
cpINDEL4	1	1	1	1	1	1	1	1	2	2	1	2	2	2
cpINDEL5	1	2	1	2	2	2	1	2	2	2	2	2	2	2
cpINDEL8	2	2	2	1	2	2	2	2	1	1	1	1	1	1
cpINDEL9	2	2	2	2	2	2	2	2	1	1	1	1	1	1
cpINDEL12	2	1	2	1	1	2	2	1	1	1	1	1	1	1

*Genotype codes 1 and 2 corresponded to deletion or insertion type, respectively, against Nipponbare, based on the whole chloroplast data obtained from 45-2. Code 3 showed longer insertion expected.

Table 3. Plastid types among wild rice populations compared to NBR core collection

Population/ accession	No. of accessions	Plastid types														No. of plastid types	Genetic diversity*
		–1	–2	–3	–4	–5	–6	–7	–8	–9	–10	–11	–12	–13	–14		Diversity ± SE
PCB	28	0	1	2	0	11	0	2	2	7	0	1	2	0	0	8	0.394 ± 0.034
PTT	24	0	0	12	0	0	0	0	0	0	0	0	0	0	12	2	0.500 ± 0.000
NHL	32	0	0	29	0	0	2	0	0	0	0	0	0	1	0	3	0.067 ± 0.017
Total	84	0	1	43	0	11	2	2	2	7	0	1	2	1	12		
Control#																	
NBR	32			4	1	20				2	2			3		6	0.266 ± 0.034
Nipponbare	1													1		–	–
Accession 45-2	1	1														–	–

* Genetic diversity calculated with the same formula of Expected heterozygosity.

'Nipponbare' was used as a control for determination of plastid genotype, carrying Plastid type-13. Another control accession, 45-2, carried Plastid type-1.

NBR core collection comprised six plastid types. Plastid type-5 predominated, and three of the 32 accessions shared the same plastid type (Plastid type-13). Eight of the total 14 plastid types were not detected in the NBR core collection. The PCB, PTT, and NHL populations comprised eight, two, and three plastid types, respectively. Five unique plastid types were found in PCB, one unique type (Plastid type-14) was found in PTT, and another unique type (Plastid type-6) was found in NHL. Plastid type of 45-2 itself was not detected in the PCB population, although 45-2 was selected from the population. This appeared to reflect its high maternal diversity in the PCB population.

Genetic diversity among wild rice populations in Thailand

Eleven SSR markers dispersed across the genome were used to assess genetic diversity in Thai wild rice populations (**Table 4**). The number of multiple alleles in Thai populations tended to be low: the average number of alleles ranged from 2.5 to 7.4, being relatively lower than in the NBR core collection (*Na* = 7.8) when 11 SSR markers were genotyped. However, all individuals in Thailand revealed *Na* = 10.8, reflecting the fact that the three populations were composed of diverse individuals.

Zero scores for the observed heterozygotes (*Ho*) were obtained at two loci in PTT and NHL populations, meaning that they were all homozygotes. Only SSR-chr9 in PTT showed a zero score as *He*. The lower score may have been due to the small size of the PTT population. In contrast, the PCB population showed higher scores at these loci. The observed heterozygosity (*Ho*) scores of the three populations ranged from 0.406 (NHL) to 0.523 (PTT) for the 11 markers, and from 0.500 (PTT) to 0.568 (PTT) for eight markers (**Table 4**). *Ho* in total was *Na* = 0.524 when *Ho* in Laos was 0.281. These relatively higher scores were due to their population structure, as these examined Thai populations examined in this study were composed of only perennials. *Ho* in NBR showed the most lowest score than others because of the limited propagation method of this population in *ex situ* conditions: their self-crossing seeds were maintained.

The expected heterozygosity (*He*) of the PCB, PTT, and NHL populations ranged from 0.452 to 0.674 for the 11 markers, and 0.473 to 0.681 for the eight markers. The highest score was found in the PCB population. *He* in Thailand was 0.773 for the 11 markers and 0.772 for the eight markers, the score being higher than in Laos and also in NBR. Although the core collection was composed of 32 accessions originating from various countries such as Thailand, India, Malaysia, Myanmar, the Philippines, China, Sri Lanka, Indonesia, Bangladesh, Cambodia, and Laos, there was no significant difference among the *He* scores for the Thai populations, suggesting that wild rice in Thailand still retains much diversity as natural populations. Populations originating from different areas showed lower scores for *He*, reflecting the fact that they have diverged genetically from each other.

Principle component analysis

In order to obtain an overview of wild rice in Thailand, PCA was conducted based on the SSR data matrix of the 11 markers for 116 accessions including all Thai wild rice and NBR accessions. The scatter plot of the first and second principle components showed a clear divergence of wild rice populations in Thailand, compared to the core collection (**Fig. 2**). The observed difference for the Thai wild rice populations was probably due to limited exchange of genetic materials among the populations because of differences in their habitats. Accessions in both the PCB and NHL populations showed wide variation in the plot, suggesting their high level of genetic diversity. Some individuals in the PCB population partly overlapped with the NBR core collection, but most of them were distributed independently.

Table 4. No. of alleles (*Na*), Observed heterozygosity (*Ho*), and Expected Heteozygosity (*He*) in Thai wild rice populations compared with NBR core collection collected from various South-Asian countries and reference data reported by Wang *et al.* (2012)

Marker	Chr.	Thai wild rice populations												Control data			Reference data*	
		Thailand total			PCB			PTT			NHL			NBR			Laos	
		Na	*Ho*	*He*	*Na*	*Ho*	*He*	*Na*	*Ho*	*He*	*Na*	*Ho*	*He*	*Na*	*Ho*	*He*	*Ho*	*He*
RM3604	1	7	0.476	0.782	3	0.321	0.603	2	0.500	0.375	6	0.594	0.619	7	0.188	0.556	0.665	0.774
RM1347	2	16	0.714	0.820	11	0.750	0.786	2	0.500	0.375	9	0.844	0.797	15	0.188	0.840	0.492	0.703
RM3180	3	11	0.548	0.781	9	0.571	0.835	3	0.542	0.405	4	0.531	0.465	12	0.594	0.865	0.414	0.658
AL606650	4	15	0.500	0.818	10	0.464	0.776	2	0.500	0.375	9	0.531	0.839	9	0.125	0.760	0.016	0.691
+29CAT	6	6	0.452	0.776	4	0.357	0.474	2	0.500	0.375	4	0.500	0.595	6	0.125	0.615	0.000	0.604
RM125	7	9	0.631	0.778	5	0.464	0.476	4	1.000	0.750	4	0.500	0.571	9	0.656	0.819	0.424	0.543
SSR-chr8	8	7	0.310	0.800	5	0.750	0.643	3	0.208	0.565	3	0.000	0.580	4	0.156	0.729	–	–
SSR-chr9	9	4	0.060	0.385	3	0.179	0.645	3	0.000	0.000	1	0.000	0.119	4	0.000	0.600	–	–
RM311	10	7	0.238	0.648	5	0.357	0.615	2	0.000	0.500	3	0.313	0.529	9	0.094	0.807	0.239	0.696
SSR-chr12	12	6	0.679	0.696	4	0.643	0.680	4	1.000	0.625	3	0.469	0.588	5	0.125	0.499	–	–
RM17	12	15	0.631	0.776	12	0.821	0.879	3	1.000	0.625	2	0.188	0.170	10	0.125	0.876	0.000	0.198
Average																		
8 SSR		10.8	0.524	0.772	7.4	0.513	0.681	2.5	0.568	0.473	5.1	0.500	0.573	9.6	0.262	0.767	0.281	0.608
11 SSR		9.4	0.476	0.733	6.5	0.516	0.674	2.7	0.523	0.452	4.4	0.406	0.534	8.2	0.216	0.724	–	

* Wang *et al.* (2012).

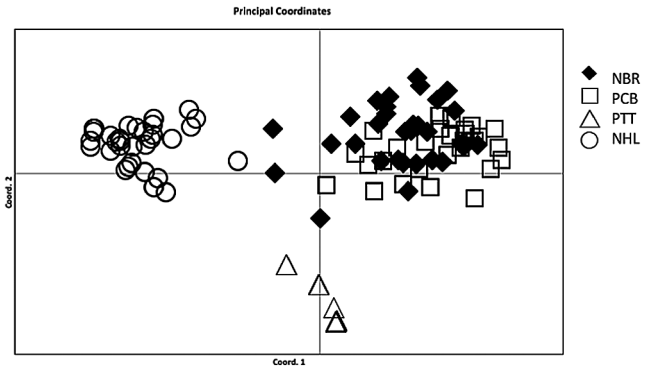

Fig. 2. Principal component analysis (PCA) performed with nuclear SSRs among a core collection (NBR: solid diamonds), the Prachin Buri population (PCB: clear squares), the Pathum Thani population (PTT: clear triangles), and the Nong Han Lake population (clear circles).

Table 5. Genetic diversity evaluated by eight SSR markers of three sub-populations in Prachinburi *in situ*-conserved population

Marker	Site 1			Site 14			Site 28		
	Na	*Ho*	*He*	*Na*	*Ho*	*He*	*Na*	*Ho*	*He*
RM3604	4.0	0.219	0.694	2.0	0.000	0.305	4.0	0.281	0.431
RM1347	3.0	1.000	0.529	3.0	0.063	0.398	3.0	0.844	0.557
RM3180	4.0	0.875	0.584	2.0	0.000	0.305	3.0	0.219	0.251
AL606650	5.0	0.688	0.569	6.0	0.813	0.713	8.0	0.906	0.809
+29CAT	5.0	1.000	0.736	4.0	0.094	0.444	6.0	0.563	0.726
RM125	6.0	0.906	0.770	5.0	0.656	0.718	8.0	0.625	0.836
RM311	7.0	1.000	0.793	3.0	0.063	0.354	6.0	0.781	0.600
RM17	3.0	0.688	0.471	3.0	0.719	0.630	4.0	0.344	0.300
Average	4.6	0.797	0.643	3.5	0.301	0.483	5.3	0.570	0.564

Divergence among three sub-populations

Three sub-populations in PCB were further genotyped using the eight SSRs (**Table 5**). *Na* scores ranged from 3.5 to 5.3. These higher scores of more than 2.0 indicated that these sub-populations were composed of different individuals carrying multiple alleles. High *Ho* scores were observed at three loci in at Site 1, suggesting a higher outcrossing rate and also frequent vegetative propagation. The lowest scores were found for RM3604 and RM3180 at Site 14 as a result of self-crossing and vegetative growth. Relatively high *Na* and *Ho* values at other loci suggested they resulted from fully self-crossed progeny after many generations. The *Ho* value for representatives of the 28 sub-populations was 0.513 (SE = 0.066). *Ho* at Site 1 was 0.797 (SE = 0.094), which was significantly higher than that of representatives at the 5% level. *Ho* at Site 14 was 0.301 (SE = 0.127), and that at Site 28 was 0.570 (SE = 0.094). The average *Ho* for the three sub-populations was 0.556 (SE = 0.072). The relatively lower score at Site 14 was due to the use of samples of vegetative tissues from single individuals that had propagated vegetatively (**Fig. 3**). *He* among representatives of the 28 sub-populations was 0.681 (SE = 0.056). *He* at Site 1 was 0.643 (SE = 0.043) and that at Site 28 was 0.564 (SE = 0.079) (**Fig. 3A**). *He* at Site 14 was 0.483 (SE = 0.0.063), which was significantly lower than that of representatives at

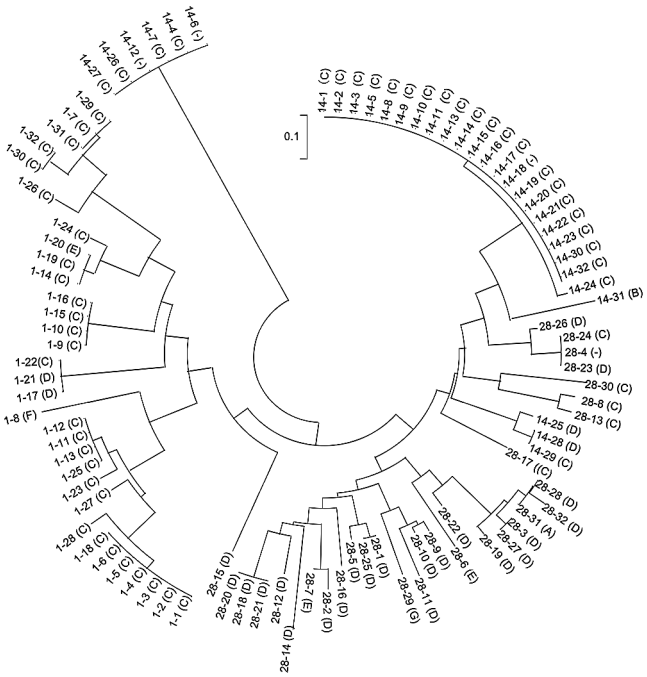

Fig. 3. Phylogenetic tree of three sub-populations in the Prachin Buri mother population, constructed by the neighbor joining method, with genetic distance calculated using nuclear SSRs. Plastid types were distinguished with cpINDEL1, 3, 8, and 9. Hyphens indicated the plastid types that were not determined.

the 5% level. The average *He* of the three sub-populations was 0.563 (SE = 0.037), due to the high number of individuals with the same genotypes (**Fig. 3B**). Overall genotypes were used for evaluation of the haplotypes. Even in small 2 m × 2 m plots, multiple nuclear haplotypes and maternal lineages were observed (**Fig. 3**). The number of haplotypes was 14 at Site 1, five at Site 14, and 26 at Site 28. Maternal lineages were distinguished by cpINDELs1, 3, 8, and 9, which showed a genetic diversity of more than 0.459. In total, seven different types from A to G types were distinguished based on the INDEL combinations (**Supplemental Table 4**). The number of plastid types was four at Site 1, three at Site 14, and five at Site 28. Plastid type C was predominated at both Sites 1 and 14, and type D at Site 28. These small sub-populations were composed of multiple maternal lineages.

Discussion

In order to preserve the genetic diversity of rice, some aspects of which are not completely understood, *in situ* conservation trials for future rice breeding would be beneficial for future rice breeding. This has been planned in Thailand and Laos from the 1990'. In previous studies, the genetic diversity of wild rice at the *in situ*-conserved site in Laos was evaluated using molecular markers, annual and perennial types being present together (Wang *et al.* 2012). Partial gene flow between these two types was assumed. In

addition, unique alleles were found in parts of the population. These inner structural differences resulted from failures to recover the natural population structure through *ex situ* conservation (Qian *et al.* 2005, Song *et al.* 2003).

The perennial type was predominant at the *in situ*-conserved site in Thailand. Perennial wild rice generally tends to show relatively higher genetic diversity because it prefers cross-hybridization and tends to propagate vegetatively. Thus, perennial wild rice is more competitive than annual wild rice in river estuaries and lakes that are always filled with water. Such areas are suitable as *in situ*-conserved sites. Perennial populations in Cambodia were located in ponds and deep water areas (Morishima *et al.* 1983, Orn *et al.* 2015). We also found perennials in canals and at road sides in Vietnam (unpublished data).

One of the perennial populations in Prachin Buri, Thailand (PCB population) has been designated an *in situ*-conserved site. The most distinct difference between the PCB population and the Laos *in situ*-conserved site was the proportion of perennial types. It is likely that 100% of accessions in the PCB population were perennial. The pond has been used as a water reservoir for long time and the water bed rarely has been dry. The areas of the two sites are also similar. However, the *Ho* score for the Laos site was lower than that of the PCB population except for one particular locus (**Table 4**). The average *Ho* score for the Laos site was 0.281 and that for the PCB population was 0.513 based on the loci examined at both sites.

One perennial wild rice accession selected for its tolerance to acid sulfate soil and bacterial leaf streak is now being applied to improve modern varieties. This individual, accession 45-2, was used to obtain NGS data for SNP analysis. In addition, the complete chloroplast sequence of this accession was obtained by the re-sequencing method in order to trace maternal lineage. Eight among 37 INDELs were chosen for plastid genotyping. This detected multiple maternal lineages in single populations. Although this accession was screened from the same population, none of the others carried the same plastid type. It is suggested that the Prachin Buri population maintains divergent maternal lineages. In fact, three of the multiple maternal lineages detected were shared with the NBR core collection but the other five were novel types and were not detected in the core collection. Because of this high variation, a unique accession showing tolerance to acids sulfate soils was successfully developed.

The Phathun Thani and Non Hang lake populations were small in size. However, they maintained variation to some extent. As a small population was conserved in PTT and a small part of the large population in Non Hang lake was collected, differences in the scores seemed to reflect the size of the area from which the accessions were collected, and thus the presumed net genetic diversity of NHL would be larger than we estimated here. In the PCB population, half of the markers carried more than five multiple alleles. These higher allele numbers allowed use distinguish diverse individuals within the populations and also inside sub-populations.

Three additional sub-populations were screened to evaluate genetic polymorphism in 2 m × 2 m plots in the Prachin Buri population. Perennials tended to prefer vegetative propagation and outcrossing. Vegetative propagation may result in particularly small spaces being occupied by single genotype. On the other hand, outcrossing, preferred by perennial wild rice, leads to polymorphism through seed propagation. In fact, the relatively high genetic variance observed in each plot suggested that perennial wild rice individuals had left descendants through seed propagation (**Fig. 3**). Some individuals belonged to the same haplotypes, but showed different maternal lineages. In contrast, the PTT population which was smaller than the PCB population, retained only two maternal lineages, each composed of a few haplotypes similar to each other. This suggested that disturbance in a natural pond may not increase uniformity if a sufficient population size is maintained. The preference of perennial wild rice for vegetative propagation may not lead directly to uniformity within populations. Identical genotypes were due to vegetative propagation or redundant sampling from small plots. In one case at Site 14, 20 individuals shared a nuclear lineage not only in terms of haplotypes but also shared a particular maternal lineage, suggesting that a particularly vigorous individual had expanded inside the sub-population. Many of the particular individuals at Site 14 tended to have propagated vegetatively. However, this trend was not observed in other sub-populations. Especially, the sub-populations at Sites 1 and 28 showed a relatively high number of haplotypes: 14 at Site 1 and 26 at Site 28. It is suggested that this sub-population had high gene flow from outside and a high potential to leave descendants through seed propagation simultaneously. These sub-population structures indicated that perennial populations maintained a complex population structure.

Plants cannot move by themselves, and are spread by factors such as human intervention, wind, animal movement, and water flow, and in addition wind can convey pollen to other individuals. Once gene flow occurs, the new traits are captured in a population. A preference to retain heterozygosity is another factor that can maintain high diversity. Thus, natural perennial populations might maintain high diversity including neutral polymorphism. In fact, the soil in the *in situ* habitat of the Prachin Buri population is not acid sulfate (pH 3.9–4.1). We grew randomly selected wild rice accessions using soils transferred from regions of low pH and succeeded in screening accession 45-2, which also showed high tolerance to bacterial leaf streak. A high rate of polymorphism has allowed this population to retain various genetic components. This may also be the case for blast fungus polymorphism.

In contrast, the Prachin Buri population showed synchronization of flowering time. Not all perennial populations in South-east Asia show such synchronization; our recent long-term field observations of perennial populations in the Mekong Delta, Vietnam, have not demonstrated synchronized flowering. However, synchronization of flowering

time results in frequent gene exchange through outcrossing. High genetic variance detected in maternal lineages is not due to this mechanism, but depends on the origin of a perennial population. Disturbance caused by frequent flooding and water buffalo may be other factors that introduce multiple maternal origins into single populations. Successive vegetative propagation may allow populations to maintain their heterogeneity. The wide range of genetic variations in these *in situ* populations will help rice breeders to screen valuable resources.

The present authors have been performing ongoing field studies of natural wild rice populations in Thailand. However, natural populations have been showing rapid extinction over the last few decades, as a result of increased economic development and the social changes (Akimoto *et al.* 1999, Nonomura *et al.* 2010). To conserve the rich genetic diversity of rice for future breeding programs, conservation of natural populations *in situ* is imperative. Furthermore, researchers must make efforts to understand the significance of the conservation by the local agricultural societies who wish to control wild rice populations. The authors consider that the Thai government should register the sites of *in situ* conservation as a Globally Important Agricultural Heritage Systems (GIAHS) to encourage economic growth in local communities.

Acknowledgments

This work was funded by a Grant-in-aid B (Overseas project, No. 16H05777) and partly by a Grant-in-Aid for Scientific Research on Innovative Areas (15H05968). The valuable wild rice accessions used in this study were distributed by the National Institute of Genetics, supported by the National BioResource Project, MEXT, in Japan. Sequencing was performed at the Gene Research Center, Hirosaki University.

Literature Cited

Akimoto, M., Y. Shimamoto and H. Morishima (1999) The extinction of genetic resources of Asian wild rice, *Oryza rufipogon* Griff.: A case study in Thailand. Genet. Resour. Crop Evol. 46: 419–425.

Brar, D.S. and G.S. Khush (1997) Alien introgression in rice. Plant Mol. Biol. 35: 35–47.

Buckler, E.S., J.M. Thornsberry and S. Kresovich (2001) Molecular diversity, structure and domestication of grasses. Genet. Res. 77: 213–218.

Gao, L., W. Chen, W. Jiang, S. Ge, D. Hong and X. Wang (2000) Genetic erosion in northern marginal population of the common wild rice *Oryza rufipogon* Griff. and its conservation, revealed by the change of population genetic structure. Hereditas 133: 47–53.

Gao, L., B. Schaal, C. Zhang, J. Jia and Y. Dong (2002) Assessment of population genetic structure in common wild rice *Oryza rufipogon* Griff. using microsatellite and allozyme markers. Theor. Appl. Genet. 106: 173–180.

Gao, L. (2004) Population structure and conservation genetics of wild rice *Oryza rufipogon* (Poaceae): a region-wide perspective from microsatellite variation. Mol. Ecol. 13: 1009–1024.

Hao, Y., M. Akimoto, R. Kaewcheenchai, M. Sotowa, T. Ishii and R. Ishikawa (2015) Inconsistent diversities between nuclear and plastid genomes of AA genome species in the genus *Oryza*. Genes Genet. Syst. 90: 269–281.

Henry, R.J., N. Rice, D.L.E. Waters, S. Kasem, R. Ishikawa, Y. Hao, S. Dillon, D. Crayn, R. Wing and D. Vaughan (2010) Australian *Oryza*: Utility and Conservation. Rice 3: 235–241.

Huang, X., N. Kurata, X. Wei, Z.-X. Wang, A. Wang, Q. Zhao, Y. Zhao, K. Liu, H. Lu, W. Li *et al.* (2012) A map of rice genome variation reveals the origin of cultivated rice. Nature 490: 497–501.

Kuroda, Y., Y.-I. Sato, C. Bounphanousay, Y. Kono and K. Tanaka (2007) Genetic structure of three *Oryza* AA genome species (*O. rufipogon*, *O. nivara* and *O. sativa*) as assessed by SSR analysis on the Vientiane Plain of Laos. Conserv. Genet. 8: 149–158.

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

McCouch, S.R., L. Teytelman, Y. Xu, K.B. Lobos, K. Clare, M. Walton, B. Fu, R. Maghirang, Z. Li, Y. Xing *et al.* (2002) Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). DNA Res. 9: 199–207.

Morishima, H., Y. Shimamoto, Y. Sano and Y.-I. Sato (1983) Observations on wild and cultivated rices in Thailand for ecological-genetic study. *In*: Reports of The Study-Tours for Investigation of Wild and Cultivated Rice Species, Part II. (Compiled by H. Morishima), http: //shigen.nig.ac.jp/rice/oryzabase/ricereport/, pp. 1–93.

Nonomura, K.-I., H. Morishima, T. Miyabayashi, S. Yamaki, M. Eiguchi, T. Kubo and N. Kurata (2010) The wild *Oryza* collection in National BioResource Project (NBRP) of Japan: History, biodiversity and utility. Breed. Sci. 60: 502–508.

Oka, H.-I. (1988) Origin of cultivated rice. Japan Scientific Societies Press, Tokyo, Elsevier Amsterdam, p. 254.

Orn, C., R. Shishido, M. Akimoto, R. Ishikawa, T.M. Htun, K. Nonomura, Y. Koide, M. Sarom, S. Vang, S. Sophany *et al.* (2015) Evaluation of genetic variation among wild rice populations in Cambodia. Breed. Sci. 65: 430–437.

Peakall, R. and P.E. Smouse (2006) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics 28: 2537–2539.

Qian, J., T. He, Z. Song and B. Lu (2005) Genetic evaluation of *in situ* conserved and reintroduced populations of wild rice (*Oryza rufipogon*: Poaceae) in China. Biochem. Genet. 43: 561–575.

Song, Z.P., X. Xu, B. Wang, J.K. Chen and B.R. Lu (2003) Genetic diversity in the northernmost *Oryza rufipogon* populations estimated by SSR markers. Theor. Appl. Genet. 107: 1492–1499.

Sotowa, M., K. Ootsuka, Y. Kobayashi, Y. Hao, K. Tanaka, K. Ichitani, J.M. Flowers, M.D. Purugganan, I. Nakamura, Y.-I. Sato *et al.* (2013) Molecular relationships between Australian annual wild rice, *Oryza meridionalis*, and two related perennial forms. Rice (N Y) 6: 26.

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

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei and S. Kumar (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28: 2731–2739.

Vaughan, D. (1994) The wild relatives of rice—A genetic resources handbook IRRI Manila, Philippines, pp. 64.

Wang, Y.P., C. Bounphanousay, K. Kanyavong, I. Nakamura, Y.-I. Sato, T. Sato, H.-S. Zhang, L.-H. Tang and R. Ishikawa (2012) Population structural analysis of an *in-situ* conservation site for wild rice in Laos. Genes Genet. Syst. 87: 311–322.