Inconsistent diversities between nuclear and plastid genomes of AA genome species in the genus *Oryza*

Hao Yin¹, Masahiro Akimoto², Reunreudee Kaewcheenchai³, Masahiro Sotowa⁵,
Takashige Ishii⁴ and Ryuji Ishikawa^{5*}

¹United Graduate School of Agricultural Sciences, Iwate University, Morioka, Iwate 020-8550, Japan

²Obihiro University of Agriculture and Veterinary Medicine, Obihiro, Hokkaido 080-8555, Japan

³Prachinburi Rice Research Center, Bureau of Rice Research and Development,
Rice Department, Prachinburi 25150, Thailand

⁴Graduate School of Agricultural Science, Kobe University, Kobe 657-0013, Japan

⁵Faculty of Agriculture and Life Science, Hirosaki University, Hirosaki, Aomori 036-8561, Japan

(Received 6 September 2014, accepted 8 May 2015; J-STAGE Advance published date: 18 December 2015)

AA genome species in the genus *Oryza* are valuable resources for improvement of cultivated rice. *Oryza rufipogon* and *O. barthii* were progenitors of two domesticated rice species, *O. sativa* and *O. glaberrima*, respectively. We used chloroplast single-nucleotide repeats (RCt1-10) to evaluate genetic diversity among AA genome species. Higher diversity was detected in the American species *O. glumaepatula* and the Asian species *O. rufipogon*. Other chloroplast sequences indicated that *O. glumaepatula* shares high similarity with *O. longistaminata*. Insertions of retrotransposable elements, however, showed a close relation between *O. barthii* and *O. glumaepatula*. To clarify phylogenetic relationships among AA genomes, whole-genome sequences obtained from different species were used to develop chloroplast INDEL markers. The INDEL patterns clearly showed multiple maternal origins of *O. glumaepatula*. The complicated origins have resulted in high genetic diversity in this species. In contrast, the Australian endemic species *O. meridionalis* tended to show narrower diversity than the other species. High variation in *O. rufipogon*, reconfirmed using the chloroplast INDELs, covered the variation in *O. meridionalis* and part of the variation in *O. glumaepatula*. Maternal lineages including *O. barthii*, *O. longistaminata* and the remainder of *O. glumaepatula* were phylogenetically close to each other and carried low genetic diversity. They were separated from independent lineages, suggesting that they had diverged from a single ancestral maternal lineage, but diverged later to keep gene flow within respective species, as SSR compositions suggested. Genetic relationships among AA genome species indicate how these species have evolved and become distributed across four continents.

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

INTRODUCTION

The evolution of the genus *Oryza*, especially AA genome species, has been a major focus of rice research (Oka, 1988). AA genome species include two cultigens: *O. sativa*, which is cultivated worldwide, and *O. glaberrima*, an endemic species distributed along the River Niger. The progenitors of these cultivated species were once regarded as the *O. perennis* complex, consisting of four

geographical races – Asian, African, American and Oceanian – inhabiting different regions (Morishima, 1969). These were subsequently renamed and defined as different species: *O. rufipogon*, *O. barthii*, *O. longistaminata*, *O. glumaepatula* and *O. meridionalis* (Oka, 1988). *Oryza sativa* was domesticated in East Asia from *O. rufipogon* (Oka, 1988; Fuller and Sato, 2008; Huang et al., 2012), and *O. glaberrima* was independently domesticated from *O. barthii* (Wang et al., 2014).

The major cultivated species, *O. sativa*, has a deep population structure due partly to introgression from various varieties of *O. rufipogon*, and also from other varietal groups (Tang and Morishima, 1997; Ishikawa et al.,

Edited by Kiichi Fukui

* Corresponding author. E-mail: ishikawa@hirosaki-u.ac.jp

DOI: <http://doi.org/10.1266/ggs.14-00063>

2002a, b; Garri et al., 2005; McNally et al., 2009; Molina et al., 2011; Huang et al., 2012). Wide variation had already occurred 86 to 440 thousand years ago between subspecies in *O. sativa*, *indica* and *japonica*, and 2 million years ago among AA genome wild species (Zhu and Ge, 2005; Molina et al., 2011). The deep divergence between subspecies in *O. sativa* is evident in genome divergence (Huang et al., 2012). This variation can be seen in various measurements of divergence (McNally et al., 2009; Molina et al., 2011). *japonica* was classified into Temperate-*japonica* and Tropical-*japonica* (Oka, 1988; Sato, 1991). Divergence between Tropical-*japonica* and *indica* was estimated to be as deep as about 3,900 years ago (Molina et al., 2011). However, these varietal groups shared monophyletic origin. Because they shared identical alleles at several loci regulating major agronomic traits such as non-shattering, both varietal groups were created through past introgression to share the same genetic components (Tang and Morishima, 1997; Ishikawa et al., 2002a, b). Past introgression events among wild forms and cultivars have also been inferred from genomic data (Huang et al., 2012). Such introgressions can still be observed as weedy rice, and in parts of genome sequences where selective sweeps have occurred (Molina et al., 2011). In fact, strong selective sweeps can be seen to have occurred at several key domesticated genes such as the non-shattering gene and white pericarp gene (Li et al., 2006; Lin et al., 2007; Sweeney et al., 2007). In contrast, the endemic species *O. glaberrima* was domesticated from the annual *O. barthii*, which is widespread in west tropical Africa (Oka, 1988; Khush, 1997; Semon et al., 2005). Recent genomic data have suggested that *O. glaberrima* was domesticated from a narrow gene pool in *O. barthii* (Wang et al., 2014). Because the domesticated species occurred independently from *O. rufipogon*, *O. glaberrima* has diverged sufficiently from *O. sativa* to develop multiple reproductive barriers (Morishima, 1969). The related perennial species, *O. longistaminata*, shared the same habitat as African species, but was more widely distributed in Africa and was independent from domestication.

Two other wild species in different continents, *O. meridionalis* and *O. glumaepatula*, have shown no domestication events. *O. meridionalis* is a species endemic to Oceania, including New Guinea and Australia, and is characterized by an annual life history and morphologically by a short anther (Ng et al., 1981; Vaughan, 1994; Lu, 1999). Its distribution partly overlaps with that of *O. rufipogon*. Recently, a possible new species has been reported in the same area, which appears to have diverged from the endemic species but acquired a perennial life history (Sotowa et al., 2013; Brozynska et al., 2014). In addition, a conventional perennial type, the so-called Australian *O. rufipogon*, was found to share a highly similar chloroplast genome with this possible new

species and also *O. meridionalis*. However, the Australian *O. rufipogon* did not show any reproductive barrier against Asian *O. rufipogon*. Nuclear markers also suggested that the Australian type had not shown any divergence from the Asian type. However, there is still insufficient understanding of divergence among AA genome species. Another example is the American species, *O. glumaepatula*, which is distributed in the New World ranging from Cuba to Brazil. The life form of this taxon is still uncertain: an ecotype distributed in Central America and the northern region of South America seems to have a perennial habit (Oka, 1988), while another ecotype in tropical Brazil shows an annual-perennial intermediate form (Akimoto et al., 1998). This means that there are still opportunities to detect *de novo* diversity in natural populations for this and other species.

The availability of various molecular markers now allows researchers to acquire a comprehensive grasp of divergent evolution among the AA genome species described above, including chloroplast sequences, chloroplast single-nucleotide repeats and nuclear simple sequence repeats (SSRs), although this has not been completely applied to genomes other than those of some *O. sativa* cultivars. Since the complete sequencing of the chloroplast genome of *O. sativa* cv. Nipponbare (Hiratsuka et al., 1989), the known presence of conserved sequences has made it possible to compare divergent accessions and also the mitochondrial and nuclear genomes (Sotowa et al., 2013; Brozynska et al., 2014). Re-sequencing is also feasible for any genome belonging to the genus *Oryza*, allowing possible alignment against the reference sequence (Waters et al., 2012). In the present study, we utilized these materials and tools to evaluate the phylogenetic relationships of AA genome species among Asian, African, American and Oceanian species.

MATERIALS AND METHODS

Plant materials Wild rice accessions composed of ranks 1 to 3 and comprising five AA genome species were provided by the National BioResource Project in Japan (Nonomura et al., 2010). The accessions formed a core collection selected from typical representatives in historical collections stored at the National Institute of Genetics, Mishima, Japan. Twenty accessions for *O. barthii*, 20 accessions for *O. glumaepatula*, 19 accessions for *O. longistaminata*, 39 accessions for *O. rufipogon* and 18 accessions for *O. meridionalis* were examined to clarify their genetic diversity (Table 1). Seven *O. rufipogon* accessions were categorized as Oceanian *O. rufipogon*, because two of them originated in Australia and the other five originated in Papua New Guinea. In a previous report, we characterized these accessions as distinct from Asian *O. rufipogon* (Sotowa et al., 2013). To compare genetic diversity among the cultivars, 20 *indica* and 20

Table 1. Wild rice accessions examined in this study

Species, name of accession	Origin	Rank (1-2-3)			
<i>O. glumaepatula</i>			W0708	Guinea	2
W1169	Cuba	1	W1004	Ghana	3
W1171	Cuba	3	W1232	Unknown	3
W1183	Guyana	3	W1413	Sierra Leone	1
W1185	Suriname	2	W1420	Mali	3
W1187	Brazil	2	W1423	Mali	3
W1189	Manaus, Brazil	3	W1444	Ivory Coast	3
W1191	Brazil	3	W1448	Ivory Coast	3
W1196	Colombia	2	W1454	Burkina Faso	3
W1477	Brazil	3	W1460	Dahomey	3
W2140	Brazil	3	W1465	Nigeria	3
W2145	Brazil	1	W1504	Tanzania	3
W2149	Brazil	3	W1508	Unknown	1
W2160	Brazil	3	W1540	Congo	2
W2165	Brazil	3	W1570	Nigeria	3
W2173	Brazil	3	W1573	Nigeria	3
W2184	Brazil	3	W1624	Cameroon	2
W2192	Brazil	3	W1650	Tanzania	3
W2199	Brazil	1	<i>O. meridionalis</i>		
W2201	Brazil	3	W1297	Darwin, Australia	2
W2203	Brazil	3	W1300	Darwin, Australia	3
<i>O. barthii</i>			W1625	Darwin, Australia	1
W0042	No description	3	W1627	Australia	2
W0652	Sierra Leone	1	W1631	Kununurra area, Australia	3
W0698	Guinea	2	W1635	Darwin, Australia	1
W0720	Mali	2	W1638	Queensland, Australia	3
W0747	Mali	2	W2069	Kununurra area, Australia	2
W1050	Gambia	3	W2071	Kununurra area, Australia	3
W1063	No description	3	W2077	from Darwin to Normanton, Australia	3
W1410	Sierra Leone	3	W2079	from Darwin to Normanton, Australia	2
W1416	Sierra Leone	3	W2080	from Darwin To Normanton, Australia	3
W1443	Mali	3	W2081	Matarauka, Australia	3
W1467	Cameroon	3	W2100	Queensland, Australia	3
W1473	Chad	3	W2103	Queensland, Australia	2
W1574	Nigeria	3	W2105	Queensland, Australia	3
W1583	Chad	3	W2112	Queensland, Australia	3
W1588	Cameroon	1	W2116	Queensland, Weipa, North Point, Australia	3
W1605	Nigeria	3	Asian <i>O. rufipogon</i>		
W1642	Botswana	3	W0106	Phulankara, near Cuttack, Orissa, India	1
W1643	Botswana	3	W0107	Pahala, Orissa, India	3
W1646	Tanzania	2	W0108	Cuttack, Orissa, India	3
W1702	Mali	3	W0120	Cuttack, Orissa, India	1
<i>O. longistaminata</i>			Continued		
W0643	Gambia	2			

Table 1. Continued

Species, name of accession	Origin	Rank (1-2-3)
W0137	Kadiam, Andhra, India	3
W0180	Ngao, Lamphang, Thailand	3
W0593	Binjai Rendah, Malaysia	3
W0610	Rangoon, Myanmar	3
W0630	Magwe, Myanmar	2
W1294	Musuan, Mindanao, Philippines	1
W1551	Saraburi, Thailand	3
W1666	Siliguri, India	3
W1669	Orissa, India	3
W1681	Orissa, India	3
W1685	Orissa, India	3
W1690	Chiengrai, Thailand	3
W1715	China	3
W1807	Sri Lanka	2
W1852	Chiang Saen, Thailand	3
W1865	Saraburi, Thailand	3
W1866	Saraburi, Thailand	1
W1921	Saraburi, Thailand	1
W1939	Bangkoknoi, Thailand	3
W1945	No description	2
W1981	Palembang, Indonesia	3
W2003	from Pajani to Bombay, India	1
W2014	India	3
W2051	Hobiganji, Bangladesh	2
W2263	Cambodia	2
W2265	Laos	3
W2266	Laos	3
W2267	Laos	3
Australian <i>O. rufipogon</i>		
W2078	from Darwin to Normanton, Australia	2
W2109	Queensland, Australia	3
New Guinean <i>O. rufipogon</i>		
W1230	Baad, Koembe, Papua New Guinea	3
W1235	Taram, Papua New Guinea	3
W1236	Madang, Papua New Guinea	2
W1238	Koembe River, Papua New Guinea	3
W1239	Sakor River, Papua New Guinea	3

japonica accessions were also genotyped (Supplementary Table S1). The *japonica* cultivars had already been classified into Tropical-*japonica* (Tr-J) and Temperate-*japonica* (Tm-J) using a method described by Sato (1991). These wild rice accessions were supplied as DNA samples from the National Institute of Genetics, Japan. DNA samples of cultivars and additional material were extracted with

the general urea method.

Molecular markers Chloroplast single-nucleotide repeat markers, RCt1-RCt10, were applied to evaluate the genetic diversity of the maternal origins of AA genome species (Ishii and McCouch, 2000; Ishii et al., 2001). Nuclear SSR markers were also applied to evaluate genetic diversity (Table 2). These fragments were amplified with Thermopol Taq (New England Biolabs, Tokyo, Japan) and electrophoresed on a 6% denaturing polyacrylamide gel, and genotyped after staining the gel with silver nitrate (Wang et al., 2012). The plastid genotypes were also subjected to principal component analysis (PCA) to evaluate their relationships. The chloroplast genes *rpl16* and *matK* were amplified to clarify sequence diversification using the respective primers listed in Table 2. Fragments of *rpl16* and *matK* were amplified to detect SNPs in order to determine phylogenetic relations among African and American species. All fragments were amplified with Thermopol Taq under the following cycle conditions: 94 °C for 3 min for pre-heating; 30 rounds of 94 °C for 10 s, 55 °C for 30 s and 72 °C for 30 s; and 72 °C for 5 min. These PCR fragments were purified with a FastGene Gel/PCR extraction kit (Nippon Genetics, Tokyo, Japan), and then sequenced with a BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Tokyo, Japan) using an ABI3500 Genetic Analyzer. Retrotransposon insertions were re-examined, as reported previously by Cheng et al. (2002). Three insertions, *p-SINE1-r705*, *p-SINE1-r801* and *p-SINE1-r806*, were selected to clarify relationships among *O. barthii*, *O. longistaminata* and *O. glumaepatula*. PCR was performed with *Ex Taq* (Takara Bio, Kusatsu, Japan) using the following protocol: 94 °C for 3 min for pre-heating; 30 rounds of 94 °C for 10 s, 60 °C for 30 s and 72 °C for 30 s; and 72 °C for 5 min. Amplified fragments were electrophoresed on 1.5% agarose gels to clarify whether insertions were carried by each accession.

Novel INDEL markers based on whole-genome sequence data W1171 (*O. glumaepatula*) and a single Thai wild rice, 45-2, grown at Prachinburi Rice Research Center (Rice Department, Thailand), were subjected to next-generation sequencing (NGS) to identify novel insertions or deletions (INDELs) which may be detected only from non-model accessions. DNA samples of these accessions were extracted from mature leaves with a DNeasy Plant Mini Kit (QIAGEN, Tokyo, Japan). The whole genome sequences were obtained with Illumina-Hiseq as 100-bp pair-ends, with 75,832,764 reads for W1171 and 62,762,232 for 45-2. The raw reads were re-sequenced using a CLC Genomics Workbench (CLC bio Japan, Tokyo, Japan) against the complete chloroplast genome of *O. sativa* cv. Nipponbare (GU592207.1). INDELs were then screened. When the variations showed over 1,000-

Table 2. List of primers used in this study

Type of marker	Locus	Forward	Reverse	Chloroplast genome region (INDELs)	INDEL type	Reference	Original data to develop chloroplast INDELs
RCt	RCt1	CATCCTTTTCAATCCAAATCA	TGCCTGATGTAGGGAAAAGC		(A)10	Ishii and McCouch. (2000)	
	RCt2	CTGGGGGGGATTATACCTGT	ATATCTCTCATTTCGCACGCA		(A)11	Ishii and McCouch. (2000)	
	RCt3	TAGGCATAATTCCCAACCCA	CTTATCCATTTCGAGCATAGGG		(A)10	Ishii and McCouch. (2000)	
	RCt4	ACGGAATTGGAACCTCTTTGG	AAAAGGAGCCTTGGAAATGGT		(T)12	Ishii and McCouch. (2000)	
	RCt5	ATTTGGAATTTGGACATTTTCG	ACTGATTCGTAGGCGTGGAC		(T)10	Ishii and McCouch. (2000)	
	RCt6	GAATTTTGAAGCTTTGAAATTTTACCC	AAGCGTACCGAAGACTCGAA		(A)10	Ishii and McCouch. (2000)	
	RCt7	GTGTCAATCTCTAGGCGAAC	AAATATGACAGAAAAGAAAATAGG		(T)10	Ishii and McCouch. (2000)	
	RCt8	ATAGTCAAGAAAGAGGATCTAGAAT	ACCGGATTCATAAGAGTA		(T)17	Ishii and McCouch. (2000)	
	RCt9	ATAAGGTTATTCCCGCTTACC	AAATTGGGGGAATTCGTACC		(T)10	Ishii and McCouch. (2000)	
	RCt10	TCTTCATTTGGAATCTGGGC	CTATTGATGCAAACGCTGTAC C		(T)10	Ishii and McCouch. (2000)	
rpl16	rpl16-exon1	ATGCTTAGTGTGTGACTCGTTAG				This study	
	rpl16-336f	GGTCTATGAATTACATCATAAAAAG				This study	
	rpl16-500f	TTTTTGAAGCTCCATTGCGAG				Sotowa et al. (2013)	
	rpl16-1kb	ATGAGAAGAACTCTCATGTCC				This study	
	rpl16-486r		CAATTCTCAGTTTTTATTAECTCGG			This study	
	rpl16-5Preverse		TGTTTACGAAATCTGGTCTCTTTTG			This study	
	3P		ATCTGCTCATTTTAAAGGGT			Nakamura et al. (1997)	
INDEL (cp genome)	glum-cpINDEL1	CTCGGACGAATAATCTAATACATGG	CTATGATTCTATGTCTCCTTAGTG	46087..46091	TATAT Deletion	This study	W1171 genome data
	glum-cpINDEL2	ATATATAGTCAAGAAAGAGGATC	ATGAATTAAACAAATAAGACAGG	78424..78429	TTTTTT Deletion	This study	W1171 genome data
	glum-cpINDEL3	CAAAAATTTTCTCATTTGAAACAATC	CAATTTGAGTTACGAAACAAGGGAG	103827*103828	GTTTT Insertion	This study	W1171 genome data
	glum-cpINDEL4	TGGCGGCAGTCTCGAAAAAG	CAAGTTCACGAACTAATAAGG	105208*105209	ATTCA Insertion	This study	W1171 genome data
	rufi-cpINDEL1	GGATTACCGGAAACAAACAACC	GCCAAATTGAGCAGGTTGCG	12670..12673	AGGG Deletion	This study	45-2 genome data
	rufi-cpINDEL2	TTTGGGGAAGAAACATCTTTC	TAAACGGGAGAAATCGACTAAG	14012..14013	AC Deletion	This study	45-2 genome data
	rufi-cpINDEL3	AATTGCTCTCACCGCTCTTTC	TAGTCGAATTGTTGTATCAACTC	17380..17385	ATAGAA Deletion	This study	45-2 genome data
	rufi-cpINDEL4	TAAATTGATATGGCTCGGACG	TGCTATGATTCTATGTTCTCC	46087..46091	TATAT Deletion	This study	45-2 genome data
	meri-cpINDEL2	GCCTTGTTCAGGAACCTCGACAG	TTGGTTGTACCATTGCATTTTCAG	5852..5856	CAATC Deletion	This study	GU592208
	meri-cpINDEL3	AATGGCGCAATGATCTTGGAGA	GAATGGCGATGGCTCGATTTC	8192*8193	AGAAA Insertion	This study	GU592208
	meri-cpINDEL5	AAGTGTGCCTTGCAACCGAG	AAGCAGCAGAACACCTGAAAC	13566..13567	T Deletion	This study	GU592208
	meri-cpINDEL8	GATATATTGTGTGCTGGCATTCTC	TTCCAGTGAATCATATGCAC	17379..17383	ATAGAA Deletion	This study	GU592208
matK	matK	TTGATGCAAGAATTGCCTTTTC	AAAATGCAACACCCTGTTCTGACC			This study	
SINE-INDEL	p-SINE1 (r705)	TGTTCGCGAACTTGCAATTGT	AGAATCAAACCTGACCTGTC			Cheng et al. (2002)	
	p-SINE1 (r801)	CTTGGCTTATTATTACTGATT	ATGAAGAATAGCGTAACAAAT			Cheng et al. (2002)	
	p-SINE1 (r806)	ATGCAGCTGTAAGAAGAGT	CAAGATTAAGGCTCATCTGA			Cheng et al. (2002)	
SSR	AP003436	GCAGCGAAGCCAACGTAGTCC	CTGCCTTCCCAACATCTTCTC			Wang et al. (2012)	
	AP004212	GGAGGCTCTACTACATATGG	TGGGAAACTATGCATCAGTC			Wang et al. (2012)	
	+29Cat	CACGATCTAGAAGACGAGAG	CCAAATTACGCCTTCTCTACC			Wang et al. (2012)	
	RM3204	GCAACCTTTTCTCCTCCTC	CCAAGGAGAGCGCACTAGC			McCouch et al. (2002)	
	RM257	CAGTCCGAGCAAGAGTACTC	GGATCGGACGTGGCATATG			Chen et al. (1997)	
	RM3577	CCGATCCCATTCACAGATTC	CAGTGCCTTGATCGATGTTG			McCouch et al. (2002)	
	RM17	TGCCCTGTATTCTTCTCTCTC	GGTGATCCTTTCCCATTTCA			Panaud et al. (1996)	

fold coverage, over 100 counts, and frequencies of more than 50%, these INDELs were used in “wet experiments”. To confirm these variations in comparison with the Nipponbare chloroplast genome, four INDELs from each accession were developed (Table 2). Another complete chloroplast genome of *O. meridionalis* (GU592208) was aligned with the Nipponbare genome and four INDELs

were developed.

Data analysis Data of RCt microsatellite and nuclear SSRs were used to perform PCA and to calculate genetic distance among accessions with GenAlEx (<http://biology-assets.anu.edu.au/GenAlEx/Welcome.html>). The number of alleles (*Na*) and effective alleles (*Ne*), and expected

heterozygosity (*He*), were also calculated by GenAlEx. The *Ne* will be significantly lower than the actual number if the frequencies for some alleles are much smaller than those for others.

He was calculated using the formula,

$$He = 1 - \sum(1 - x_i^2).$$

Phylogenetic trees based on RCt genotypes, INDELs and SSRs were constructed by the neighbor joining method using Populations 1.2.31 (<http://bioinformatics.org/~tryphon/populations/>). Phylogenetic trees were drawn with MEGA 5.0 (Tamura et al., 2011).

RESULTS

Genetic diversity in chloroplast microsatellites
Polymorphism in chloroplast single-nucleotide repeats was found in six of ten loci examined (Supplementary Table S2). The average expected heterozygosity (*He*) ranged from 0.176 to 0.542 among species (Table 3). The highest score was for *O. glumaepatula*, followed by 0.528 for *O. rufipogon*. Our previous study showed that *O. rufipogon* accessions were composed of two diverged maternal lineages (Sotowa et al., 2013). Therefore, these accessions were divided into Asian and Oceanian groups, which were found to have distinct *He* scores of 0.511 and 0.150, respectively. *O. sativa* showed relatively higher genetic diversity when two varietal groups (*indica* and *japonica*) were mixed together. These varietal groups diverged at the subspecies level and were strongly influenced by different groups of *O. rufipogon*. When the genetic diversity of each varietal group was calculated

separately, their scores were lower than that of *O. rufipogon*. Subgroups in *japonica*, Tm-J and Tr-J, were also diverged to some extent, and their individual scores were also calculated. The two subgroups showed lower scores than the *japonica* group. These trends suggested that varietal groups tended to carry different RCt genotypes among *O. sativa*.

To obtain an overview of diversity among AA genome species, PCA was adopted. Allelic combinations revealed particular groups (Fig. 1A). The subgroups in Asian cultivated species were clearly separated from each other, but the variation of the cultivars as a whole was included within that of *O. rufipogon*. This showed that *O. rufipogon* in the core collection covered the fundamental variation inherited by the varietal accessions. PCA showed that the Oceanian group was located in a position intermediate between *O. rufipogon* and *O. meridionalis* (Fig. 1B). After excluding *O. sativa*, the relationships of all wild AA genome species were confirmed by PCA (Fig. 1C). *O. glumaepatula* overlapped partly with *O. longistaminata* and *O. rufipogon*, and *O. barthii* was placed outside the others (Fig. 1D). The distribution graphs obtained by PCA suggested that *O. longistaminata* shared a closely related chloroplast genome with *O. glumaepatula*.

A phylogenetic tree was constructed with RCt genotypes (Supplementary Fig. S1). *O. meridionalis*, *O. barthii* and Oceanian *O. rufipogon* formed distinctive clades. *O. glumaepatula* tended to form two major clades with either *O. rufipogon* or *O. longistaminata*. *O. rufipogon* accessions were widely distributed. Ambiguous divergence was due to the mutable nature of single-sequence repeats. Except for RCt6, polymorphic markers carried multiple alleles (Supplementary Table S2). In

Table 3. Expected heterozygosity (*He*) evaluated with chloroplast microsatellite markers among AA genome species

Species	No.	He						Average
		RCt1	RCt3	RCt5	RCt6	RCt8	RCt9	
Cultigen								
<i>O. sativa</i>	40	0.049	0.635	0.495	0.489	0.569	0.644	0.480
<i>indica</i>	20	0.000	0.180	0.180	0.095	0.095	0.000	0.092
<i>japonica</i>	20	0.095	0.480	0.000	0.320	0.530	0.480	0.318
Tm-J	10	0.180	0.180	0.000	0.000	0.460	0.480	0.217
Tr-J	10	0.000	0.420	0.000	0.480	0.540	0.320	0.293
Wild								
<i>O. rufipogon</i>	39	0.650	0.537	0.264	0.375	0.570	0.670	0.528
Asia	32	0.671	0.554	0.273	0.387	0.588	0.692	0.511
Oceania	7	0.000	0.245	0.000	0.000	0.245	0.408	0.150
<i>O. meridionalis</i>	18	0.444	0.401	0.000	0.105	0.204	0.105	0.210
<i>O. glumaepatula</i>	20	0.810	0.585	0.515	0.255	0.445	0.640	0.542
<i>O. barthii</i>	20	0.345	0.335	0.000	0.000	0.000	0.375	0.176
<i>O. longistaminata</i>	19	0.637	0.465	0.000	0.188	0.283	0.548	0.354

Fig. 1. PCA among *O. sativa* (*indica*, Tm-japonica and Tr-japonica) and AA genome species. (A) Varietal groups and *O. rufipogon* were compared. (B) *O. sativa*, Asian and Oceanian *O. rufipogon*, and *O. meridionalis* as an endemic species in Australia were compared. (C) All wild rice species carrying AA genomes were compared. (D) African and American species were compared.

one case, RCt1 carried seven alleles in a single species. Although mutable alleles of RCt markers caused phylogenetic relations to be ambiguous, PCA with RCt repeats demonstrated a novel phylogenetic relationship of *O. glumaepatula*.

Sequence-based phylogenetic analysis with *rpl16* and *matK* Parts of the chloroplast genome including *rpl16* were sequenced among all species. As the *rpl16* sequence included RCt8 and other multiple single-nucleotide repeats, its phylogenetic tree did not show clear relationships (data not shown). Thus, single-nucleotide repeats were excluded from the *rpl16* sequence data and SNP combinations were summarized as haplotypes to confirm relationships among *O. glumaepatula*, *O. barthii* and *O. longistaminata* (Table 4). *O. longistaminata* shared haplotypes (Haplotypes 2 and 7) with *O. rufipogon*, one of which (Haplotype 2) was also a major haplotype in *O. glumaepatula*. Two haplotypes (Haplotypes 3 and 4) in the latter species carried SNPs against Haplotype 2. Haplotypes 3 and 4 carried an A to G substitution at 1430 nt against Haplotype 2. Haplotype 4 carried another substitution, an A to T substitution at 655 nt. They were identical to Haplotype 2 except for these two substitutions. None of the haplotypes was identical to the two haplotypes in *O. barthii*, because Haplotypes 5 and 6 carried a T insertion between 779 and 780 nt and Haplotype 6 carried another A substitution at 1180 nt when com-

pared to Haplotype 2. These variations resulted in a unique feature in *O. barthii*.

To confirm relationships among *O. glumaepatula*, *O. barthii* and *O. longistaminata*, an additional chloroplast gene, *matK*, was also sequenced. In each case, *O. glumaepatula* and *O. longistaminata* shared the same sequence, which was identical to *O. sativa* cv. Nipponbare, except that one accession, W1448, carried a C to T substitution at 684 nt (Supplementary Table S3). All *O. barthii* accessions possessed a unique substitution at 516 nt.

Genetic distance evaluated by nuclear DNA It has previously been reported that *O. glumaepatula* shares particular insertions of retrotransposons with *O. barthii* (Cheng et al., 2002). However, chloroplast markers did not show any evidence for this. Nuclear SSR markers and other markers were therefore applied to clarify the phylogenetic relationships and thus to resolve the apparent inconsistency regarding genetic relationships between American and African species. Twenty-eight SSR markers were randomly chosen and seven of them showed sufficient polymorphism. Other markers represented monomorphism over all species or were not amplified with DNA templates in multiple species, probably because of sequence divergence among different species or genomic rearrangement. The averaged *He* scores for the species overall ranged from 0.335 at AP004212 to 0.704 at RM257 (Supplementary Table S4). *O. meridionalis*

Table 4. SNPs except single-nucleotide repeats detected in *rpl16* sequence among wild accessions with AA genomes

Species	No. of accessions	Intron 1				Exon 2				Haplotype	Accessions	
		655 nt	676 nt	770^771 nt	779^780 nt	1180 nt	1288 nt	1397–1398 nt	1430 nt			1450 nt
<i>O. sativa</i>	1	A	T	–	–	G	C	CG	A	G	1	Nipponbare
<i>O. glumaepatula</i>	8	A	T	–	–	G	C	GC	A	G	2	W1171, W1183, W1185, W1196, W1477, W2199, W2201, W2203
	7	A	T	–	–	G	C	GC	G	G	3	W1191, W1169, W1187, W2140, W2145, W2149, W2192
	5	T	T	–	–	G	C	GC	G	G	4	W1189, W2160, W2165, W2173, W2184
<i>O. barthii</i>	16	A	T	–	T	G	C	GC	A	G	5	W0042, W0652, W0698, W0720, W0747, W1050, W1063, W1410, W1416, W1443, W1467, W1473, W1583, W1588, W1605, W1702
	4	A	T	–	T	A	C	GC	A	G	6	W1574, W1642, W1643, W1646
<i>O. longistaminata</i>	18	A	T	–	–	G	C	GC	A	G	2	W0643, W0708, W1004, W1232, W1413, W1420, W1423, W1444, W1448, W1454, W1460, W1465, W1504, W1508, W1540, W1570, W1624, W1650
	1	A	T	–	–	G	C	GT	A	G	7	W1573
<i>O. rufipogon</i>												
Asia	23	A	T	–	–	G	C	GT	A	G	7	W0106, W0108, W0137, W0180, W1294, W1551, W1666, W1669, W1681, W1685, W1690, W1715, W1807, W1852, W1865, W1866, W1921, W1939, W1981, W2051, W2263, W2265, W2266
	7	A	T	–	–	G	C	GC	A	G	2	W0107, W0120, W0593, W0610, W0630, W1945, W2267
	2	A	T	–	–	G	C	GC	A	C	8	W2003, W2014
Oceania	2	A	T	–	–	G	C	GT	A	G	7	W1230, W1236
	2	A	T	C	–	G	C	GC	A	G	9	W2078, W2109
	3	A	T	C	–	G	T	GC	A	G	10	W1235, W1238, W1239
<i>O. meridionalis</i>	2	A	T	C	–	G	C	GC	A	G	9	W1635, W2116
	10	A	T	C	–	G	T	GC	A	G	10	W1631, W1638, W2069, W2071, W2080, W2081, W2100, W2103, W2105, W2112
	6	A	A	C	–	G	C	GC	A	G	11	W1297, W1300, W1625, W1627, W2077, W2079

showed the lowest diversity among the species. A high *He* score was found in Oceanian *O. rufipogon*, next to Asian *O. rufipogon*. This relatively high value for Oceanian *O. rufipogon* would have been due to the mixture of Asian and Australian factors described in our previous paper (Sotowa et al., 2013).

These genotypes were then used to calculate genetic dis-

tances in order to construct a phylogenetic tree. Species-specific clades were clearly recognized for *O. meridionalis* and *O. glumaepatula*. *O. rufipogon* showed a scattered distribution and partly formed clades with *O. longistaminata*. Oceanian *O. rufipogon* accessions W1235 and W1239 formed a clade with the Australian endemic species *O. meridionalis* (Fig. 2). The past classification of Oceanian

Fig. 2. Phylogenetic tree obtained by the neighbor-joining method based on seven SSR loci. The diverse distribution of Oceanian accessions is shown using bold letters. One *O. barthii* accession, W1416, was located next to Indian *O. rufipogon*, and is underlined. The scale indicates genetic distance.

O. rufipogon W1235 and W1239, based on a field observation by Katayama (1968), would be mistranslated to annotate the species classification between annual types belonging to either *O. rufipogon* or *O. meridionalis*, because *O. meridionalis* had not been reported at that time; Ng et al. (1981) reported the species. They should now be referred to as *O. meridionalis*. Detailed phenotypic observations should be performed to determine the taxonomic classification. Other Oceanian *O. rufipogon* accessions formed three clades with some of the Asian *O. rufipogon* accessions. In one of these clades, two Australian *O. rufipogon* accessions (W2078 and W2109) and one Papua New Guinean *O. rufipogon* accession (W1238) formed a single clade with Cambodian, Indian and Malaysian *O. rufipogon* accessions. W1230 was grouped with accessions from India, Laos, Myanmar and Thailand. W1236 was grouped with accessions from India, Indonesia, the Philippines and Thailand. *O.*

glumaepatula and *O. barthii* were not genetically close and tended to form distinct clades with other species except for one *O. barthii* accession (W1416) close to one Indian *O. rufipogon* accession, W0137.

The previously reported close relationship between *O. glumaepatula* and *O. barthii* has been confirmed on the basis of SINE insertions (Cheng et al., 2002). Accessions used in this experiment were re-examined using the three known SINE insertions (Supplementary Table S5). *O. barthii* and *O. glumaepatula* shared the *p-SINE1-r806* insertion, while only *O. glumaepatula* carried the *p-SINE1-r801* insertion. *O. longistaminata* carried the *p-SINE1-r705* insertion but other species did not. These insertions suggested that these species diverged and that all accessions in single species originated from single ancestral populations. Some *O. rufipogon* accessions that were included in the same clades with *O. longistaminata* did not carry the *p-SINE1-r705* insertion (data not

shown). The genetic relationships inferred from SSR genotypes were imprecise beyond the species level, but SSR markers provided an overview of relationships among species.

NGS data Two accessions, W1171 (*O. glumaepatula*) and Thai wild rice 45-2 (*O. rufipogon*), were newly re-sequenced against the chloroplast genome of *O. sativa* cv. Nipponbare. 45-2 was wild rice from the Prachinburi Rice Research Center, Rice Department, Thailand. The two accessions were subjected to NGS. To understand polymorphism in wild rice populations, such a native wild rice accession was selected. Based on the data of Basic Variant Detection, several rearrangements, insertions and deletions were assumed to exist. Another chloroplast genome from *O. meridionalis* was also used to develop INDELs. All INDELs except rufi-cpINDEL2 and meri-cpINDEL5 were presumed to be simple four- to six-nucleotide insertions or deletions (Table 2). These INDELs were expected not to be mutable like SSR (Table 5). One INDEL marker, rufi-cpINDEL3, was denoted as an ATAGAA deletion (Table 2). However, because the site was flanked by incomplete inverted repeat units similar to ATAGAA, this INDEL marker generated a high number of alleles and high *He* scores (Table 5, Supplementary Table S6). *O. glumaepatula* and *O. barthii* carried six and seven alleles at rufi-cpINDEL3. These seemed to behave as markers with high mutability such as SSR markers.

A phylogenetic tree drawn with the 11 INDELs except

for rufi-cpINDEL3 is shown in Fig. 3. *O. meridionalis* accessions tended to form species-specific clades. Four Oceanian *O. rufipogon* accessions (W1235, W1238, W1239 and W2078) were included in a single clade with *O. meridionalis*. Two of them, W1235 and W1239, were grouped with *O. meridionalis* in a clade constructed with nuclear SSR genotypes (Fig. 2). As mentioned above, these may belong taxonomically to the species *O. meridionalis*. In contrast, other Oceanian *O. rufipogon* accessions, W1238 (Papua New Guinea) and W2078 (Australia), were grouped with Asian *O. rufipogon* accessions originating from Cambodia (W2263), India (W1681) and Malaysia (W0593), along with an accession from Australia (W2109), into another clade constructed with nuclear genotypes. *O. barthii* also showed narrow genetic diversity: the species formed a single clade except for W1702, which was close to the remaining *O. barthii* accessions in the tree. This clade was relatively close to the clade comprising *O. longistaminata* and several *O. glumaepatula* accessions. In overview, *O. glumaepatula* formed several subgroups which were scattered between *O. rufipogon* and *O. longistaminata*.

Phylogenetic trees obtained from INDEL markers based on different species are shown in Supplementary Fig. S2. The phylogenetic tree obtained with the glum-cpINDEL data was similar to the tree in Fig. 3. *O. glumaepatula* carried multiple maternal lineages shared partly with *O. rufipogon* and partly with *O. longistaminata*. The tree with the data of rufi-cpINDEL was ambiguous, due to the high mutability in rufi-

Table 5. Expected heterozygosity (*He*) detected among AA genome species using cpINDELs generated from NGS data

Markers		<i>He</i>					
		<i>O. glumaepatula</i>	<i>O. barthii</i>	<i>O. longistaminata</i>	<i>O. rufipogon</i> (Asia)	<i>O. rufipogon</i> (Oceania)	<i>O. meridionalis</i>
glum-cpINDEL	cpINDEL1	0.500	0.095	0.000	0.389	0.000	0.000
	cpINDEL2	0.375	0.000	0.100	0.498	0.000	0.105
	cpINDEL3	0.375	0.000	0.100	0.000	0.000	0.000
	cpINDEL4	0.375	0.000	0.100	0.000	0.000	0.000
rufi-cpINDEL	cpINDEL1	0.000	0.000	0.000	0.342	0.490	0.000
	cpINDEL2	0.000	0.000	0.000	0.170	0.000	0.000
	cpINDEL3	0.750	0.759	0.100	0.117	0.000	0.000
	cpINDEL4	0.000	0.000	0.000	0.342	0.000	0.000
meri-cpINDEL	cpINDEL2	0.495	0.000	0.000	0.000	0.408	0.000
	cpINDEL3	0.000	0.000	0.188	0.000	0.000	0.000
	cpINDEL5	0.000	0.000	0.000	0.000	0.000	0.000
	cpINDEL8	0.000	0.000	0.488	0.117	0.408	0.000
Average <i>He</i>	over 11 loci*	0.193	0.009	0.089	0.169	0.119	0.010
Average <i>He</i>	over 12 loci	0.239	0.071	0.090	0.165	0.109	0.009

* rufi-cpINDEL3 was excluded from the scores because of its high mutability.

Fig. 3. Phylogenetic tree obtained by the neighbor-joining method based on 11 chloroplast INDEL markers developed using chloroplast genomes. The scale indicates genetic distance.

cpINDEL3 (Supplementary Fig. S3). Thus, INDEL markers will help to understand maternal lineages easily and efficiently.

DISCUSSION

Diploid AA genome species are unique among other diploid and tetraploid species belonging to *Oryza* because two domesticated forms, *O. sativa* and *O. glaberrima*, belong to AA genome species. As the major cultigen, *O. sativa*, is composed of the diverged *indica* and *japonica*, *O. rufipogon* was presumed to carry high diversity, next to *O. glumaepatula*. RCT markers have helped to distinguish accessions at the subspecies level (Ishii and McCouch, 2000). When compared at the species level, they could not give precise resolution. On the other hand, species distributions detected by PCA provided an overview of species divergence. However, phylogenetic relations obtained by data of RCT markers were not as precise as INDEL- or sequence-based relations because of the high mutability of RCT markers compared with SNPs. The divergence speed of chloroplast DNA was inferred as being slower than that of nuclear DNA, but high mutation rates can occur in single-nucleotide repeats included in RCT markers. Sequences of the chloroplast gene *rpl16* and its flanking region also demonstrated a similar result. When we excluded single-nucleotide repeats from the *rpl16* sequences, a phylogenetic tree obtained from the data was relatively clear. The *matK* sequence did not include such repeats. Thus, phylogenetic relationships estimated from *matK* sequence data were simple. *Oryza glumaepatula* shared high similarity with *O. longistaminata* based on the chloroplast sequences of *matK* and *rpl16*. In contrast, *O. glumaepatula* shared the same SINE insertions in the nuclear genome with *O. barthii*, but not with *O. longistaminata* described by Cheng et al. (2002). Compared to the relationships obtained with chloroplast data and the SINE insertions, nuclear SSR genotypes demonstrated that *O. glumaepatula* was independent from both *O. barthii* and *O. longistaminata*. This was due to the nature of SSR markers, which are easily mutated. These data suggested that all species have diverged as biological species, but the phylogenetic relationships were more complicated than we had anticipated.

SSR markers possess higher mutability, as noted above. They offer good resolution within species or even among landraces (Garris et al., 2005; Ootsuka et al., 2014). In this report, they were applied to determining how species are related to each other. Some accessions belonging to different species showed close relatedness. In the case of *O. longistaminata* and *O. rufipogon*, some accessions from both species were included in the same clades, although an insertion of *p-SINE1-705* was detected only in *O. longistaminata* and not in *O.*

rufipogon. This suggested that resolution beyond the species level with SSR markers was not reliable, due to the mutability of SSRs. Species-specific clades were recognized for *O. glumaepatula*, *O. barthii* and *O. meridionalis*. SSR markers gave an overview of relations among these species, which have diverged as independent species.

As single-nucleotide repeats tend to mutate at a higher rate than other kinds of DNA sequences, we developed INDEL markers from chloroplast genome information. NGS techniques allowed us to obtain genome information from non-model species such as *O. glumaepatula* and *O. meridionalis*. Although there is no complete nuclear genome sequence, re-sequencing against the chloroplast genome is available. The higher copy number of chloroplast genomes resulted in highly reliable polymorphism. Several INDELs were developed and relatively larger INDELs were selected, which were expected to yield obvious differences. Most INDELs resulted from a simple insertion or deletion event except for rufi-cpINDEL3, which offered six and seven alleles in *O. glumaepatula* and *O. barthii*, respectively. Several identical motifs inside a region targeted by designed primers might generate such relatively high numbers of alleles.

Except for the highly mutable rufi-cpINDEL3, other INDEL markers made it possible to evaluate the polymorphism within species. Notably, INDELs in a particular accession in *O. glumaepatula* identify heterogeneity in the species, and in fact, high heterogeneity was detected. The degree of resolution allowed us to define subgroups in this species. These subgroups of *O. glumaepatula* distinguished with chloroplast data are likely to reflect a complex evolutionary history, in which *O. glumaepatula* probably carries multiple maternal origins shared partly with *O. longistaminata* and partly with *O. rufipogon*. This complexity may account for *O. glumaepatula* having the highest *He* score estimated from RCt markers. In the process of divergence, the presumed proto-*O. glumaepatula* population as a maternal donor apparently had high gene-flow from a proto-*O. barthii* population before they diverged as different species in different continents. Thus, some *O. glumaepatula* accessions shared higher similarity of their maternal lineages with *O. longistaminata*, but they shared higher similarity with *O. barthii* at the nuclear level as estimated from SINE insertions. This inconsistency of the compositions of nucleus and cytoplasm arose before these species had completely diverged from each other. All *O. barthii* accessions were separated from other maternal origins, as estimated from chloroplast INDEL data. The different maternal origin of *O. barthii* from other species suggested that a minor maternal lineage carried by the ancestral population evolved as *O. barthii*.

When compared to *O. glumaepatula*, Australian *O. meridionalis* accessions showed the lowest heterogeneity.

In the case of *O. meridionalis*, this may have been attributable to a founder effect of earlier migration to the Australian continent from Asia. The maternal lineage, however, has not been found in Asia; the lineage has probably become extinct in Asia. All accessions of *O. rufipogon* in Australia and Papua New Guinea shared the same or a similar maternal lineage with *O. meridionalis*. This suggested that *O. meridionalis* had diverged from Asian *O. rufipogon* in the past, and later the progeny carrying the same maternal lineage and belonging to *O. rufipogon* might extend to the Oceania area. *O. rufipogon* accessions were scattered in various clades and covered various maternal lineages of other AA genome species. It could be that *O. rufipogon* donated those maternal lineages. Further data using chloroplast genome INDELs obtained from NGS data should allow a more detailed understanding of which species might be the progenitor species among AA genome species. The current data suggest that this is *O. rufipogon*. Recent studies of complete sequences of chloroplast genomes suggest that *O. longistaminata* is closely related to some *O. glumaepatula* accessions (Kim et al., 2015; Wambugu et al., 2015). Thus, we need to know more precisely about the diversity among AA genome species. As field work on native wild rice accessions is still being conducted, more intense field work will reveal precise phylogenetic relations with the help of NGS as a tool.

This work was funded by a Grant-in-Aid B (Overseas project, No. 25304021) from JSPS. The valuable wild rice accessions used in this study were distributed by the National Institute of Genetics supported by the National BioResources Project, MEXT, Japan. The sequencing facility in the Gene Research Center, Hirosaki University was used.

REFERENCES

- Akimoto, M., Shimamoto, Y., and Morishima, H. (1998) Population genetic structure of wild rice *Oryza glumaepatula* distributed in the Amazon flood area influenced by its life-history traits. *Mol. Ecol.* **7**, 1371–1381.
- Brozynska, M., Omar, E. S., Furtado, A., Crayn, D., Simon, B., Ishikawa, R., and Henry, R. J. (2014) Chloroplast genome of novel rice germplasm identified in Northern Australia. *Trop. Plant Biol.* **7**, 111–120.
- Chen, X., Temnykh, S., Xu, Y., Cho, Y. G., and McCouch, S. R. (1997) Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.) *Theor. Appl. Genet.* **95**, 553–567.
- Cheng, C., Tsuchimoto, S., Ohtsubo, H., and Ohtsubo, E. (2002) Evolutionary relationships among rice species with AA genome based on SINE insertion analysis. *Genes Genet. Syst.* **77**, 323–334.
- Fuller, D., and Sato, Y.-I. (2008) *Japonica* rice carried to, not from, Southeast Asia. *Nat. Genet.* **40**, 1264–1265.
- Garris, A. J., Tai, T. H., Coburn, J., Kresovich, S., and McCouch, S. R. (2005) Genetic structure and diversity in *Oryza sativa* L. *Genetics* **169**, 1631–1638.
- Hiratsuka, J., Shimada, H., Whittier, R., Ishibashi, T., Sakamoto,

- M., Mori, M., Kondo, C., Honji, Y., Sun, C. R., Meng, B. Y., et al. (1989) The complete sequence of rice (*Oryza sativa*) chloroplast genome: Intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Mol. Gen. Genet.* **217**, 185–194.
- Huang, X., Kurata, N., Wei, X., Wang, Z. X., et al. (2012) A map of rice genome variation reveals the origin of cultivated rice. *Nature* **490**, 497–501.
- Ishii, T., and McCouch, S. R. (2000) Microsatellites and microsynteny in the chloroplast genomes of *Oryza* and eight other Gramineae species. *Theor. Appl. Genet.* **100**, 1257–1266.
- Ishii, T., Xu, Y., and McCouch, S. R. (2001) Nuclear- and chloroplast-microsatellite variation in A-genome species of rice. *Genome* **44**, 658–666.
- Ishikawa, R., Yamanaka, S., Kanyavong, K., Fukuta, Y., Sato, Y.-I., Tang, L., and Sato, T. (2002a) Genetic resources of primitive upland rice in Laos. *Econ. Bot.* **56**, 192–197.
- Ishikawa, R., Sato, Y.-I., Tang, T., and Nakamura, I. (2002b) Different maternal origins of Japanese lowland and upland rice populations. *Theor. Appl. Genet.* **104**, 976–980.
- Katayama, T. C. (1968) Scientific Reports on the Rice-Collection-Trip to the Philippines, New Guinea, Borneo and Java. *Memories of Fac. Agri., Kagoshima Univ.* **6**, 89–134.
- Khush, G. S. (1997) Origin, dispersal, cultivation and variation of rice. *Plant Mol. Biol.* **35**, 25–34.
- Kim, K., Lee, S.-C., Lee, J., Yu, Y., et al. (2015) Complete chloroplast and ribosomal sequences for 30 accessions elucidate evolution of *Oryza* AA genome species. *Sci. Rep.* **5**, 15655.
- Li, C., Zhou, A., and Sang, T. (2006) Rice domestication by reducing shattering. *Science* **311**, 1936–1939.
- Lin, Z., Griffith, M. E., Li, X., Zhu, Z., Tan, L., Fu, Y., Zhang, W., Wang, X., Xie, D., and Sun, C. (2007) Origin of seed shattering in rice (*Oryza sativa* L.). *Planta* **226**, 11–20.
- Lu, B. R. (1999) First record of the wild rice *Oryza meridionalis* in Indonesia. *Intern. Rice Res. Notes* **24**(3), 28.
- McNally, K. L., Childs, K. L., Bohnert, R., Davidson, R. M., et al. (2009) Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *Proc. Nat. Acad. Sci. USA* **106**, 12273–12278.
- McCouch, S. R., Teytelman, L., Xu, Y., Lobos, K., Clare, K., Walton, M., Fu, B., Maghirang, R., Li, Z., Xing, Y., et al. (2002) Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). *DNA Res.* **9**, 199–207.
- Molina, J., Sikora, M., Garud, N., Flowers, J. M., Rubinstein, S., Reynolds, A., Huang, P., Jackson, S., Schaal, B. A., Bustamante, C. D., Boyko, A. R., and Purugganan, M. D. (2011) Molecular evidence for a single evolutionary origin of domesticated rice. *Proc. Natl. Acad. Sci. USA* **108**, 8351–8356.
- Morishima, H. (1969) Phenetic similarity and phylogenetic relationships among strains of *Oryza perennis*, estimated by methods of numerical taxonomy. *Evolution* **23**, 429–443.
- Nakamura, I., Kameya, N., Kato, Y., Yamanaka, S., Jomori, H., and Sato, Y.-I. (1997) A proposal for identifying the short ID sequence which addresses the plastid subtype of higher plants. *Breed. Sci.* **47**, 385–388.
- Ng, N. Q., Hawkes, J. G., William, J. T., and Chang, T. T. (1981) The recognition of a new species of rice (*Oryza*) from Australia. *Bot. J. Linn. Soc.* **82**, 327–330.
- Nonomura, K.-I., Morishima, H., Miyabayashi, T., Yamaki, S., Eiguchi, M., Kubo, T., and Kurata, N. (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.
- Ootsuka, K., Takahashi, I., Tanaka, K., Itani, T., Tabuchi, H., Yoshihashi, T., Tonouchi, A., and Ishikawa, R. (2014) Genetic polymorphisms in Japanese fragrant landraces and novel fragrant allele domesticated in northern Japan. *Breed. Sci.* **64**, 115–124.
- Panaud, O., Chen, X., and McCouch, S. R. (1996) Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). *Mol. Gen. Genet.* **252**, 597–607.
- Sato, Y.-I. (1991) Variation in spikelet shape of the *indica* and *japonica* rice cultivars in Asian origin. *Japan. J. Breed.* **41**, 121–134.
- Semon, M., Nielsen, R., Jones, M. P., and McCouch, S. R. (2005) The population structure of African cultivated rice *Oryza glaberrima* (Steud.): Evidence for elevated levels of linkage disequilibrium caused by admixture with *O. sativa* and ecological adaptation. *Genetics* **169**, 1639–1647.
- Sotowa, M., Ootsuka, K., Kobayashi, Y., Hao, Y., Tanaka, K., Ichitani, K., Flowers, J. M., Purugganan, M. D., Nakamura, I., Sato, Y.-I., et al. (2013) Molecular relationships between Australian annual wild rice, *Oryza meridionalis*, and two related perennial forms. *Rice* **6**, 26.
- Sweeney, M. T., Thomson, M. J., Cho, Y. G., Park, Y. J., Williamson, S. H., Bustamante, C. D., and McCouch, S. R. (2007) Global dissemination of a single mutation conferring white pericarp in rice. *PLoS Genet.* **3**, e133.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739.
- Tang, L.-H., and Morishima, H. (1997) Genetic characterization of weedy rices and the inference on their Origins. *Breed. Sci.* **47**, 153–160.
- Vaughan, D. (1994) The Wild Relatives of Rice -A Genetic Resources Handbook. pp. 48–49. IRRI, Manila, Philippines.
- Wambugu, P. W., Brozynska, M., Furtado, A., Waters, D. L., and Henry, R. J. (2015) Relationships of wild and domesticated rices (*Oryza* AA genome species) based upon whole chloroplast genome sequences. *Sci. Rep.* **5**, 13957.
- Wang, M., Yu, Y., Haberer, G., Reddy, M., et al. (2014) The genome sequence of African rice (*Oryza glaberrima*) and evidence for independent domestication. *Nat. Genet.* **46**, 982–988.
- Wang, Y. P., Bounphanousay, C., Kanyavong, K., Nakamura, I., Sato, Y.-I., Sato, T., Zhang, H.-S., Tang, L.-H., and Ishikawa, R. (2012) Population structural analysis of an *in-situ* conservation site for wild rice in Laos. *Genes Genet. Syst.* **87**, 311–322.
- Waters, D. L. E., Nock, C. J., Ishikawa, R., Rice, N., and Henry, R. J. (2012) Chloroplast genome sequence confirms distinctness of Australian and Asian wild rice. *Ecol. Evol.* **2**, 211–217.
- Zhu, Q. H., and Ge, S. (2005) Phylogenetic relationships among A-genome species of the genus *Oryza* revealed by intron sequences of four nuclear genes. *New Phytologist* **167**, 249–265.