

# Comparative Analysis of Rice Genome Sequence to Understand the Molecular Basis of Genome Evolution

Jianzhong Wu · Hiroshi Mizuno · Takuji Sasaki · Takashi Matsumoto

Received: 11 June 2008 / Accepted: 4 November 2008 / Published online: 22 November 2008  
© Springer Science + Business Media, LLC 2008

**Abstract** Accurate sequencing of the rice genome has ignited a passion for elucidating mechanism for sequence diversity among rice varieties and species, both in protein-coding regions and in genomic regions that are important for chromosome functions. Here, we have shown examples of sequence diversity in genic and non-genic regions. Sequence analysis of chromosome ends has revealed that there is diversity in both sequences and distribution in the region of telomere repeat arrays, from chromosome to chromosome, within a plant. Detailed study has allowed us to speculate the mechanism of generation of these arrays. Sequence analysis using various cultivated and wild rice of the *sd1* gene, which contributed to the “Green Revolution” in rice varieties and their wild progenitors, has also demonstrated sequence diversity, which is correlated with taxonomic classification. These results indicate that detailed analysis of sequence diversity and comparison might give us a clue in elucidating mechanism of the evolution of rice genome.

**Keywords** *Oryza* · Sequence diversity · Telomere repeats · Comparative genomics

---

Jianzhong Wu and Hiroshi Mizuno contributed equally on this article

J. Wu · H. Mizuno · T. Sasaki · T. Matsumoto (✉)  
National Institute of Agrobiological Sciences,  
2-1-2, Kannondai,  
Tsukuba, Ibaraki 305-8602, Japan  
e-mail: mat@nias.affrc.go.jp

J. Wu  
e-mail: jzwu@nias.affrc.go.jp

H. Mizuno  
e-mail: hmizuno@nias.affrc.go.jp

T. Sasaki  
e-mail: tsasaki@nias.affrc.go.jp

## Introduction

Rice is one of the three mega-crops (rice, maize, and wheat) on which more than half the world's population relies as major sources of calories and protein. Genetic improvement based on molecular biotechnology requires accurate genome sequence, which contributes to the establishment of genome-wide DNA markers for tagging and delimitation of the genetic regions in which genes and quantitative traits locus (QTLs) are located. Global identification of protein-coding genes in the rice genome could enhance the discovery of the genes that are responsible for agronomically desirable traits.

The International Rice Genome Sequencing Project (IRGSP, 1997–2004), which was run by a collaborative research consortium of ten countries, succeeded in establishing the nucleotide sequence of the Nipponbare cultivar of rice's *japonica* ssp. to a high standard [10]. The 370 million nucleotides from the 12 chromosomes of rice have now been widely utilized as molecular coordinates for investigating rice genomics and genetics. Although most of the euchromatic regions (95%) of the rice genome were covered by the published sequences, 62 gaps and heterochromatic regions, centromeres, and telomeres, corresponding to 5% of the total genome, remained unrevealed. To acquire the sequences of these missing regions and thus improve the public rice genome sequence, new genomic libraries recently constructed from physically fragmented genomic DNA have been utilized, and these efforts have succeeded in revealing some of the junction sequences between the euchromatic and heterochromatic regions of rice telomeres. This information, which we will describe here, should give us clues to understanding the molecular diversity of, and mechanisms responsible for, the generation of telomere structures.

The genus *Oryza* comprises 23 species [38], but one of the mysteries of rice history is that most of the modern

varieties of rice, derived from *Oryza sativa* and *Oryza glaberrima*, are the descendants of only a specific lineage (the AA genomes). *Oryza* emerged about 20 to 22 Mya [9]. There are geographic, physiological, and genetic diversities among *Oryza* species, including among rice varieties, landraces, and wild accessions. In recognition of the fact that this variation is indispensable for maintaining the vast genetic resources that should help in developing a sustainable future for the human race, rice is collected, evaluated, and stored in either national or international gene banks (e.g., NIAS Genebank, [http://www.gene.affrc.go.jp/about\\_en.php](http://www.gene.affrc.go.jp/about_en.php); [18]; Japanese National Bioresource Project, <http://www.shigen.nig.ac.jp/rice/oryzabase/top/top.jsp>; International Rice Research Institute; Genetic Resource Center, <http://www.irri.org/GRC/GRChome/Home.htm>). Global genome and information resources for the investigation of genome evolution among *Oryza* species have been facilitated by The Oryza Map Alignment Project (OMAP, <http://www.omap.org/>; [17]).

Now that we are armed with these resources, it should be interesting to know how each gene or genomic region has evolved in the course of rice evolution and domestication. The elucidation of molecular diversity, as revealed by detailed sequence analysis, should be a fundamental product of such research.

Here, we present a review of comparative genomics based on information on the sequences of the genic region. A detailed molecular diversity analysis of both the exon and intron regions within the “Green Revolution” gene would not only present information on protein diversity but also give us clues to genomic conservation and development.

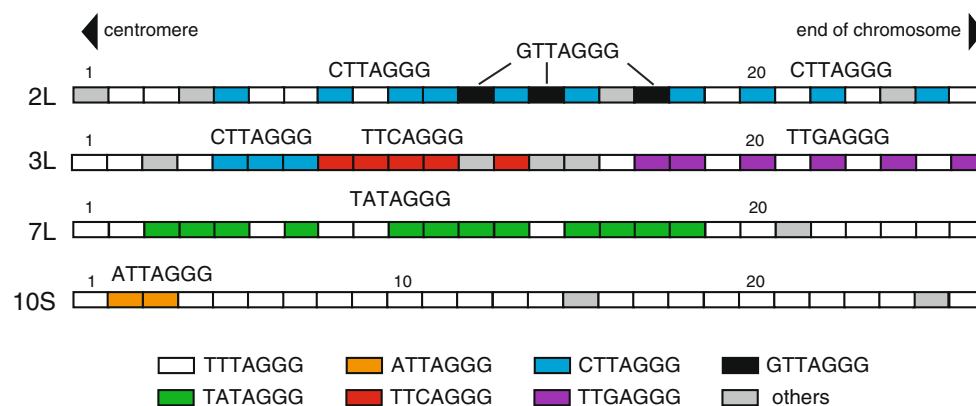
#### Diversity of the telomere region among rice chromosomes

Although IRGSP attempted clone-by-clone genomic sequencing to cover the whole genome, clone gaps remained in the

chromosomal ends. As the restriction enzymes used in the construction of PAC/BAC libraries could not cut the canonical telomere array, (TTTAGGG)<sub>n</sub>, the libraries did not contain the clones derived from telomeric sequences [5, 39]. To capture the telomere sequences, a rice fosmid library constructed by the cloning of random mechanically sheared DNA [1] was screened [24]. The library enabled telomeric sequences to be obtained without the constraints imposed by enzyme site preferences. We describe here the characteristics of the telomeric regions on the basis of their sequence and length diversity among chromosomes.

The rice chromosomal end has tandemly repeated blocks of the sequence 5'-TTTAGGG-3' [40]. These telomeric repeats are organized in the order of 5'-TTTAGGG-3' from the chromosome-specific region [24, 42]. The seven-nucleotide unit has deletions, insertions, or substitutions of single nucleotides near the junction between the telomere and the chromosome-specific region. The rate of accumulation of telomeric variants is higher in the proximal region than in the distal region [25], suggesting that the proximal region has rarely been reconstructed by telomerase on an evolutionary time scale.

This expansion of telomeric variants makes it possible to characterize the rice chromosomal end. Copies of ATTAGGG, CTTAGGG, GTTAGGG, TATAGGG, TTCAGGG, or TTGAGGG are arrayed in tandem, or the same subtypes are close to each other, at the ends of chromosomes 2L, 3L, 7L, and 10S (Fig. 1; [25]). Inversion of telomeric repeats is observed adjacent to the beginning of the telomere array on the ends of chromosomes 4L, 7S, and 9S. Therefore, the proximal telomeric sequences are composed of blocks of at least six types of TTTAGGG variants and the canonical sequence in a chromosome-specific manner. This distribution suggests that telomeric variants might have arisen from the rapid expansion of a single mutation rather than from the gradual accumulation of random mutations. The telomere of



**Fig. 1** Distribution of TTTAGGG substitution variants. Each box represents the 7-nt unit of the telomere repeat TTTAGGG (white) and the different variants (ATTAGGG, CTTAGGG, GTTAGGG, TATAGGG, TTCAGGG, and TTGAGGG), as shown in the key.

Gray box represents other variants, including deletion and insertion variants. Numbers indicate positions of telomere sequences from the junction between the chromosome-specific region and the telomere array.

rice contains a nucleotide deletion of one T in TTAGGG. Rice has a 4.9% content of deletion variants, TTAGGG, dispersed throughout the whole of the sequenced region. The telomeric sequence in the Asparagales is similar to that of rice but not identical: the deletion type in rice, TTAGGG, is present in the Asparagales [35]. The partial or full replacement of the telomeric sequences by these variants might have been due to evolutionary changes in the genomic sequence that codes the RNA template or to structural changes in the catalytic subunit.

The telomere lengths vary among various accessions of rice. The telomeres of 31 rice accessions (both cultivars and wild species, which belong to AA, BB, BBCC, CC, CCDD, GG, or HHJJ species of *Oryza*) are 5 to 20 kb in length [24]. Marked variation in telomere length is also observed among cultivated rice of the AA genome: the *japonica* cultivar Nipponbare shows a relatively low MW pattern and the *indica* cultivar Kasalath shows a relatively high MW pattern. Moreover, variation in telomere length is observed among chromosomes in Nipponbare. Use of the fiber-fluorescent *in situ* hybridization (FISH) method has revealed the diversity of telomere length of each chromosome. Seven telomeres in Nipponbare range from 5.1 to 10.8 kb in length, corresponding to about 730 to 1,500 copies of the TTAGGG telomeric repeat. The chromosome-dependent variation might be a consequence of genetic or epigenetic differences among the sequences of subtelomeres; these differences might affect the balance between telomere shortening and telomere elongation.

Telomere length in various plants has been reported: 2.5 kb in *Arabidopsis thaliana* [20]; 4.5 kb at most in *Melandrium album* [28]; 60 to 160 kb (in most cases 90 to 130 kb) in *Nicotiana tabacum* [6]; and 1.8 to 40.0 kb in maize [4]. Does telomere length change in different cells? In barley (*Hordeum vulgare*), wide variation in telomere length is observed during the differentiation or ageing of cells. The cells that develop in long-term callus cultures have very long telomeres [16]. Thus, it is possible that telomere length in rice varies with different tissue or developmental stages.

The rice telomere has diversity in both sequence and length. The mosaics of blocks of telomere variants might have resulted from slips during DNA synthesis, a high frequency of DNA recombination, or rapid deletion in the telomeric region, suggesting that the areas near the distal chromosome ends are dynamic and variable.

#### Diversity analysis of rice functional genes

Growing in a wide range of environments, the genus *Oryza* contains 23 species; rich in genomic diversity, they could serve not only as potential genetic resources for improvement of rice production but also as good research materials

for studies of the evolutionary history and functionality of genes related to speciation, domestication, polyploidy, ecological adaptation, and human selection of rice [37]. The public rice genome sequences obtained from two rice cultivars, Nipponbare (by the IRGSP) and 93-11 (by the Beijing Genomics Institute, BGI), as well as the wild rice BAC library resources established from the AA to HHKK genomes of *Oryza* species at Arizona Genomics Institute (AGI) provide a good opportunity to carry out such studies [46, 10, 2]. For example, analyses of BAC end sequences and preliminary generation of BAC contigs by using the above libraries have been conducted. These studies suggested that repeat sequences play a role in genome size evolution and found the physical evidence of changes in genomic composition and structure between the different genomes of *Oryza* species [17]. Materials on all BAC libraries and information on BAC end sequences and BAC contigs are available through the AGI BAC/EST Resource Center (<http://www.genome.arizona.edu/orders>).

Belonging to the *Oryza* genus, *Oryza sativa*, also called Asian cultivated rice, is thought to have originated from the Asian wild rice *Oryza rufipogon* only about 10,000 years ago [14]. Growing now throughout the world, *Oryza sativa* has two subspecies, *indica* and *japonica*. Knowledge of the differences in phenotype variations among rice species or subspecies at the level of molecular biology would widen future rice breeding possibilities. With this purpose, the Rice Genome Research Program (RGP) has constructed nine novel BAC libraries from species that carry the AA genome, as an important resource for comparative analysis of rice genomes. These include the three Asian rice varieties Kasalath (*indica*), Shuusoushu (*indica*), and Kha Mac Kho (*japonica*) from *O. sativa*, one accession from the African cultivated rice *O. glaberrima*, and one accession from each of the wild rice species *Oryza rufipogon*, *Oryza barthii*, *Oryza glumaepatula*, *Oryza meridionalis*, and *Oryza longistaminata* [41]. By chromosomal *in silico* mapping of 78,427 high-quality BAC end sequences, 450 Kasalath BAC contigs that consisted of 12,170 clones and covered 308.5 Mbp of the genome were generated [13]. These resources are freely accessible through the RGP homepage (BAC end sequences at <http://rgp.dna.affrc.go.jp/blast/runblast.html>, BAC contigs at [http://rgp.dna.affrc.go.jp/E/publicdata/kasala\\_thendmap/index.html](http://rgp.dna.affrc.go.jp/E/publicdata/kasala_thendmap/index.html)) for researchers to perform comparative analyses of the genomes of the two subspecies of *O. sativa* and to generate single nucleotide polymorphism (SNP) or indel markers for genetic studies.

Both basic and applied research on rice genes has been carried out in the past decade, and especially after the completion of sequencing of the two rice genomes (Nipponbare and 93-11), genomic and genetic analyses have greatly increased our understanding of the function of

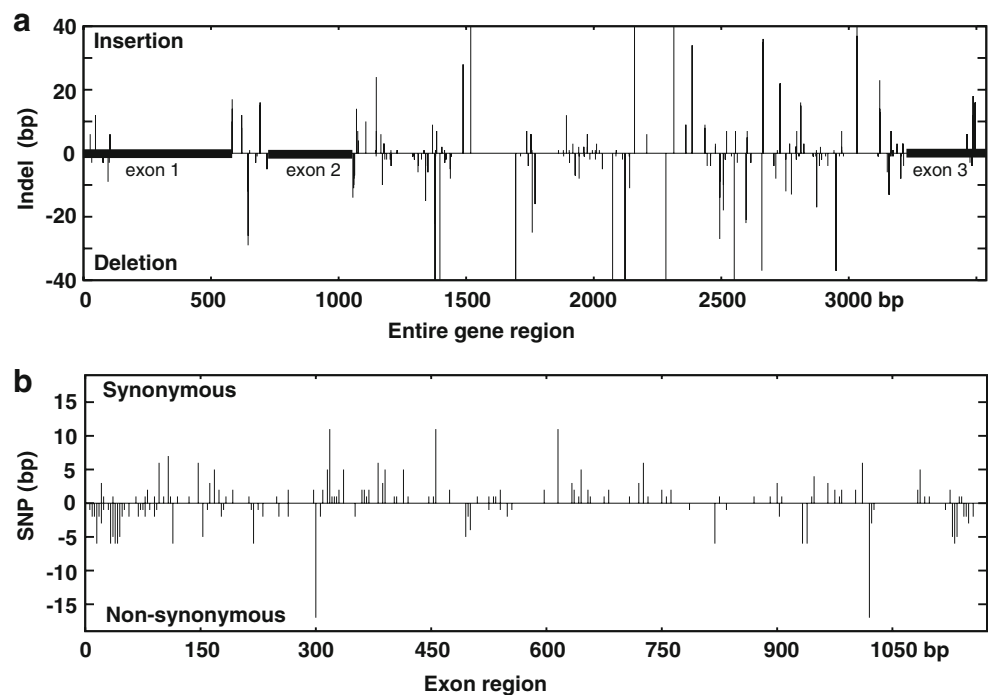
the rice genome. Among the most important achievements are the current use of advanced QTL mapping and genomic sequencing techniques for successful cloning and functional analysis of the rice genes controlling agriculturally important traits. For example, the structure and function of the genes involved in spikelet shattering, grain number, grain shape (width and length), photoperiod sensitivity, tillering, and plant architecture have been reported [3, 7, 19, 21, 22, 29, 33, 36, 43]. It will be both scientifically interesting and agriculturally important to investigate the sequence diversity of these genes among different varieties and species; this information could not only provide valuable information on evolutionary history of a crop but also lead to the discovery of new alleles for the improvement of rice breeding.

To date, there are a few genes whose sequences within the different *Oryza* species have been extensively investigated and compared to elucidate molecular and evolutionary mechanisms. The genes most analyzed for sequence comparisons in *Oryza* are probably the alcohol dehydrogenases (*Adh*). Ge et al. [8] were the first to sequence two genes (*Adh1* and *Adh2*) from 31 accessions representing all 23 rice species; they reported the phylogenetic relationships among the different *Oryza* species that are determined from the sequence polymorphisms. Yoshida et al. [44, 45] have investigated the nucleotide diversity in the *Adh1* and *Adh2* gene regions of *O. rufipogon* in order to clarify the mechanisms by which DNA variation is maintained.

We have started to perform comparative genomics on these functional genes. Here, we introduce the current results coming from the molecular and evolutionary analysis of the semidwarfing gene *sd1*, one of the most

important genes used for the development of high-yielding rice varieties. The semidwarfing gene (*sd1*) is located on the long arm of chromosome 1 in rice and encodes gibberellin 20-oxidase (GA20ox2). In the 1960s, a dramatic increment of rice production throughout Asia was obtained by the development of a high-yielding semidwarf *indica* rice cultivar known as IR8. This so-called rice Green Revolution depended largely on the introduction of the *sd1* gene, because the recessive character of the gene results in a shortened culm with improved lodging resistance and a great harvest index, allowing for the increased use of nitrogen fertilizers to improve yield [12, 15]. Using the AGI and RGP BAC libraries, we obtained and sequenced the entire regions of *sd1* genes from 17 cultivated and wild rice species by screening and chromosomal *in silico* mapping of the positive BAC clones that covered the target region in each species. For comparison of genome diversity and divergence within and among the species, the genomic region of the *Adh1* gene within the same accessions was also sequenced as controls in this study. Sequences obtained in this study have been submitted to the DNA Data Bank of Japan (acc. no. AB469048–AB469082). GA20ox2 differed in length among the species examined, ranging from 389 to 407 amino acids, with the exception of the *indica* cultivar 93-11, which contained only 341 amino acids because of the presence of an SNP creating a stop codon within the third exon. When the Nipponbare sequence was used as a reference, the indels detected in the other species were found to be distributed only on the N- and C-terminal regions within the coding sequence (Fig. 2a). Nucleotide substitutions, on the other hand, could

**Fig. 2** Distribution of indels (a) and SNPs (b) detected within the *sd1* gene between the Nipponbare and other rice species. Upper bar indicates synonymous base substitutions while lower bar indicates non-synonymous base substitutions.



be detected throughout the coding region, although, as was the case for the indels, more non-synonymous substitutions seemed to have occurred in the two terminal regions than in the internal regions of the gene (Fig. 2b). It is clear that the sequence of the gene encoding GA20ox2 is conserved within all the species examined, particularly within the AA-genome species, in which only between 0 and 5 non-synonymous sites are present, giving  $\geq 99.2\%$  identity at the amino acid level (Table 1). Even between the two most distant species—*O. sativa* and *Oryza granulata*—the gene encoding GA20ox2 had an identity of 88.0%.

The *sd1* gene was first identified in the Chinese variety Dee-geo-woo-gen (DGWG) and was crossed at the International Rice Research Institute (IRRI) in the early 1960s with Peta (tall) to develop the semidwarf cultivar IR8 [11]. Genetic and molecular analyses have demonstrated that the *sd1* gene in DGWG contains a 383-bp deletion spanning parts of the first and second exon and resulting in a frameshift that gives a stop codon within the coding sequence [26, 29]. A similar deletion (280-bp) was detected in the semidwarf indica rice cultivar Doongara [34]. Additional alleles that carry a single mutation causing changes in the amino acid residues in the semidwarf *japonica* rice cultivars Jikkoku (in exon 1), Calrose76 (in exon 2), and Reimei (in exon 3) have also been found [29, 34]. Interestingly, two accessions of wild rice *O. rufipogon* (W1944 and W1718) are reported to carry the DGWG

allele, suggesting the preservation and human use of natural alleles from the wild progenitor [27]. However, our examination of the *sd1* gene sequence within the 17 cultivated and wild rice species revealed none of these types of alleles. The rice cultivar 93-11 seems to encode a truncated protein owing to the presence of a premature stop codon; this codon could, however, be considered as a null allele, because 93-11 does not have a semidwarf phenotype. We also surveyed the presence of alleles as reported above within 60 accessions of *O. sativa* and 34 accessions of *O. rufipogon* by using the world core collections from the National Institute of Agrobiological Sciences, National Institute of Genetics, and IRRI. Along with the two modern indica cultivars IR58 and Milyang 23, another rice cultivar, Rexmont, from the USA, contains the DGWG type of allele. No other varieties within the above collections carry any of the remaining types of known alleles.

DnaSP (<http://www.ub.edu/dnasp/>) analysis based on the aligned sequences within the entire region of the *sd1* gene among different species revealed 507 polymorphic sites (in total, 1,925 available sites), including 117 synonymous sites and 78 non-synonymous sites within the exon region that enabled us to estimate the genome diversity ( $\pi$ ) as well as divergence ( $K$ , genetic distance) within and between the *Oryza* species (Table 2). The differences in genome diversity and divergence between the two genes *sd1* and *Adh1* are very interesting. The  $\pi$  value of the *sd1* gene in *O.*

**Table 1** Summary of sequence comparison in the entire region of *sd1* gene among *Oryza* species using Nipponbare as a reference

Species/subspecies variety/ accession genome		Origin	Entire region (bp)	Intron (bp)			Exon (bp)			Peptide (a. a.)		
				Total	SNPs	Indel	Total	SNPs	Indel	Total	Matching (%)	
<i>O. sativa</i> ssp.	Nipponbare	AA	Japan	2,743	1,573	–	–	1,170	–	–	389	–
<i>japonica</i>	Khau Mac Kho	AA	Vietnam	2,744	1,574	0	1	1,170	0	0	389	100
<i>O. sativa</i> ssp. <i>Indica</i>	Kasalath	AA	India	2,717	1,547	8	30 (7)	1,170	2 (2)	0	389	99.5
	Shuusoushu	AA	China	2,719	1,549	8	28 (6)	1,170	3 (2)	0	389	99.5
	93-11	AA	China	2,720	1,550	13	27 (5)	1,170	4 (3)	0	341 <sup>a</sup>	99.2
<i>O. rufipogon</i>	W0106	AA	India	2,740	1,570	7	5 (3)	1,170	2 (2)	0	389	99.5
	W0574	AA	Malaysia	2,720	1,550	8	29 (7)	1,170	2 (2)	0	389	99.5
	W1943	AA	China	2,744	1,574	8	1	1,170	2 (2)	0	389	99.5
<i>O. glaberrima</i>	IRGC104038	AA	Senegal	2,683	1,513	10	64 (4)	1,170	6 (2)	0	389	99.5
<i>O. barthii</i>	W1588	AA	Cameroon	2,682	1,512	10	65 (5)	1,170	6 (2)	0	389	99.5
<i>O. glumaepatura</i>	IRGC105688	AA	Brazil	2,681	1,511	26	72 (8)	1,170	8 (3)	0	389	99.2
<i>O. meridionalis</i>	W1625	AA	Australia	2,793	1,605	116	188 (26)	1,185	13 (5)	15 (1)	394	97.7
<i>O. longistaminata</i>	W1413	AA	Sierra Leone	2,752	1,567	65	88 (18)	1,188	7 (2)	18 (1)	395	98.0
<i>O. punctata</i>	IRGC105690	BB	Bangladesh	2,466	1,290	145	695 (47)	1,176	51 (17)	26 (5)	391	94.4
<i>O. minuta</i>	IRGC101141	BBCC	Philippines	2,987	1,802	255	555 (51)	1,185	55 (23)	29 (5)	394	92.4
<i>O. officinalis</i>	IRGC100896	CC	Thailand	2,831	1,634	199	667 (46)	1,197	52 (21)	33 (4)	398	92.7
<i>O. alta</i>	IRGC105143	CCDD	S. America	2,546	1,322	153	691 (49)	1,224	63 (27)	62 (6)	407	89.2
<i>O. brachyantha</i>	IRGC101232	FF	Africa	2,787	1,590	456	577 (60)	1,197	100 (35)	39 (5)	398	88.5
<i>O. granulata</i>	IRGC102118	GG	Thailand	2,464	1,294	292	671 (54)	1,170	103 (41)	42 (4)	389	88.0

Numbers of indels and SNPs in parentheses represent the total sites of insertion/deletion and non-synonymous substitutions, respectively

<sup>a</sup> 93-11 contains a non-synonymous SNP creating a stop codon within the exon 3 region

**Table 2** Estimated nucleotide diversity within the *sd1* gene region of *O. sativa* and its divergence with other *Oryza* species in comparison with *Adh1* gene region

		No of sites	S	$\pi$	$K_{\text{sat-ruf}}$	$K_{\text{sat-gla}}$	$K_{\text{sat-bar}}$	$K_{\text{sat-glu}}$	$K_{\text{sat-mer}}$	$K_{\text{sat-lon}}$	$K_{\text{sat-pun}}$	$K_{\text{sat-min}}$	$K_{\text{sat-off}}$	$K_{\text{sat-alt}}$	$K_{\text{sat-bra}}$	$K_{\text{sat-gra}}$
Entire genic region	<i>sd1</i>	2,713	19	0.0036	0.0022	0.0058	0.0058	0.0074	0.0223	0.0175	0.0772	0.0778	0.0749	0.0812	0.1635	0.1446
	<i>Adh1</i>	2,891	7	0.0014	0.0013	0.0061	0.0019	0.0058	0.0160	0.0182	0.0752	0.0685	0.0676	0.0690	0.1655	0.1381
Exons	<i>sd1</i>	1,170	4	0.0019	0.0012	0.0047	0.0047	0.0065	0.0109	0.0065	0.0454	0.0482	0.0454	0.0528	0.0844	0.0942
	<i>Adh1</i>	1,140	2	0.0011	0.0008	0.0016	0.0007	0.0016	0.0042	0.0042	0.0257	0.0339	0.0348	0.0385	0.0487	0.0603
Synonymous	<i>sd1</i>	300.6	1	0.0020	0.0014	0.0154	0.0154	0.0190	0.0326	0.0218	0.1229	0.1036	0.1036	0.1228	0.2198	0.1889
	<i>Adh1</i>	265.4	2	0.0045	0.0035	0.0068	0.0030	0.0068	0.0183	0.0183	0.1085	0.1487	0.1534	0.1720	0.2040	0.2373
Non-synonymous	<i>sd1</i>	869.4	3	0.0018	0.0012	0.0012	0.0012	0.0024	0.0038	0.0014	0.0214	0.0261	0.0236	0.0300	0.0396	0.0585
	<i>Adh1</i>	874.6	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0023	0.0023	0.0023	0.0023	0.0074	0.0139
Introns	<i>sd1</i>	1,543	15	0.0049	0.0037	0.0074	0.0074	0.0087	0.0416	0.0336	0.1300	0.1255	0.1240	0.1255	0.3001	0.2512
	<i>Adh1</i>	1,751	5	0.0015	0.0013	0.0104	0.0030	0.0097	0.0272	0.0314	0.1246	0.1023	0.0995	0.0986	0.2938	0.2189

S and  $\pi$ : number of substitutions (polymorphic sites) and nucleotide diversity in *O. sativa* (five varieties).  $K$ : nucleotide divergence calculated by Jukes and Cantor method between two species as shown in subscript; sat=*O. sativa*, ruf=*O. rufipogon*, gla=*O. glaberrima*, bar=*O. barthii*, glu=*O. glumeapatula*, mer=*O. meridionalis*, lon=*O. longistamonata*, pun=*O. punctata*, min=*O. minuta*, off=*O. officinalis*, alt=*O. alta*, bra=*O. brachyantha*, gra=*O. granulata*

*sativa* is higher than that of the *Adh1* gene, except at synonymous sites within the exon region, and the change in  $K$  value between the two species for the two genes is well correlated with the current taxonomic classification of *Oryza* species on the basis of crossing ability [37]. The *Adh1* gene has a lower level of variation than the average heterozygosity in *O. sativa* and *O. rufipogon*; this might be related to the adaptive importance of this gene in the face of anaerobic environmental and stress in the tropics and subtropics [30, 31, 23]. The species of *O. sativa* complex (AA genome) and *Oryza officinalis* complex (BB–EE genomes) are >1 m tall; examples are *Oryza alta* (CCDD), *Oryza latifolia* (CCDD), and *Oryza australiensis* (EE), which can grow to a height of 2 to 4 m [32]. In contrast, two small species, *Oryza brachyantha* (FF) and *O. granulata* (GG), are shorter than 1 m. Nobody knows what causes the difference in the plant height between these species. Doubtless, however, the gibberellin hormone family is involved in many aspects of plant growth and development. Although many more rice varieties and accessions of wild rice species might be needed for this kind of genomic analysis, the higher genomic diversity and larger divergence in the *sd1* region than in the *Adh1* region, particularly in terms of non-synonymous sites within the exon region, should provide primary information for us to understand the evolutionary mechanism of genes involved in the control of plant architecture. Through further comparative genomic and genetic studies, it should be possible to determine how phenotypic variations are induced by DNA mutations. This information could facilitate the exploration of natural alleles for future breeding of rice.

## Conclusion

Genome sequences from many plant species have been published, and more than 150 projects aiming to sequence plant genomes have been either completed or ongoing (Genomes OnLine Database v2.0, <http://genomesonline.org>). Only the rice and *Arabidopsis* genomes have been sequenced completely. As the conversion of draft sequences to “finished” ones takes huge amounts of time, effort, and funding, these two plants will serve as reference genomes for the study of monocot and dicot plants, respectively. The emerging ultra-high-throughput sequencing technology will enable us to obtain whole-genome information, which will be mapped and compared with these references, in less time. Studies of genome sequences within and among *Oryza* species will produce a concrete database for comparative genomics. We will be able to use this database to investigate both the evolution and function of regions, genes, motifs, and sequences within the genome.

**Acknowledgments** We thank all the members of the Rice Genome Research Program for joining our research and discussion. We also thank Dr. N. Kurata (National Institute of Genetics), Drs. D. A. Vaughan, K. Ebana, and T. Izawa (National Institute of Agrobiological Resource Sciences), and Dr. R. A. Wing (the Arizona Genomics Institute) for providing the plant materials and BAC libraries used in this study. This work was supported by a grant from the Ministry of Agriculture, Forestry, and Fisheries of Japan (Green Technology Project GD-2007).

## References

1. Ammiraju JS, Yu Y, Luo M, Kudrna D, Kim H, Goicoechea JL, Katayose Y, Matsumoto T, Wu J, Sasaki T, Wing RA. Random sheared fosmid library as a new genomic tool to accelerate complete finishing of rice (*Oryza sativa* spp. Nipponbare) genome sequence: sequencing of gap-specific fosmid clones uncovers new euchromatic portions of the genome. *Theor Appl Genet* 2005;111:1596–607.
2. Ammiraju JSS, Luo M, Goicoechea JL, Wang W, Kudrna D, et al. The *Oryza* bacterial artificial chromosome library resource: construction and analysis of 12 deep-coverage large-insert BAC libraries that represent the 10 genome types of the genus *Oryza*. *Genome Res* 2006;16:140–7.
3. Ashikari M, Sakakibara H, Lin S, Yamamoto T, Takashi T, Nishimura A, Angeles ER, Qian Q, Kitano H, Matsuoka M. Cytokinin oxidase regulates rice grain production. *Science* 2005;309:741–5.
4. Burr B, Burr FA, Matz EC, Romero-Severson J. Pinning down loose ends: mapping telomeres and factors affecting their length. *Plant Cell* 1992;4:953–60.
5. Chen M, Presting G, Barbazuk WB, Goicoechea JL, Blackmon B, Fang G, Kim H, Frisch D, Yu Y, Sun S, Higingbottom S, Pimphilai J, Pimphilai D, Thurmond S, Gaudette B, Li P, Liu J, Hatfield J, Main D, Farrar K, Henderson C, Barnett L, Costa R, Williams B, Walser S, Atkins M, Hall C, Budiman MA, Tomkins JP, Luo M, Bancroft I, Salse J, Regad F, Mohapatra T, Singh NK, Tyagi AK, Soderlund C, Dean RA, Wing RA. An integrated physical and genetic map of the rice genome. *Plant Cell* 2002;14:537–45.
6. Fajkus J, Kovarik A, Kralovics R, Bezdek M. Organization of telomeric and subtelomeric chromatin in the higher plant *Nicotiana tabacum*. *Mol Gen Genet* 1995;247:633–8.
7. Fan C, Xing Y, Mao H, Lu T, Han B, Xu C, Li X, Zhang Q. GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theor Appl Genet* 2006;112:1164–71.
8. Ge S, Sang T, Lu BR, Hong DY. Phylogeny of rice genomes with emphasis on origins of allotetraploid species. *Proc Natl Acad Sci U S A* 1999;96:14400–5.
9. Guo YL, Ge S. Molecular phylogeny of Oryzae (Poaceae) based on DNA sequences from chloroplast, mitochondrial, and nuclear genomes. *Am J Bot* 2005;2005 92:1548–58.
10. International Rice Genome Sequencing Project. The map-based sequence of the rice genome. *Nature* 2005;436:793–800.
11. International Rice Research Institute. 1967. Annual Report for 1966: Los Baños: 59–82.
12. Jennings PR. Plant type as a rice breeding objective. *Crop Sci* 1964;4:13–5.
13. Katagiri S, Wu J, Ito Y, Karasawa W, Shibata M, Kanamori H, Katayose Y, Namiki N, Matsumoto T, Sasaki T. End sequencing and chromosomal in silico mapping of BAC clones derived from an indica rice cultivar, Kasalath. *Breed Sci* 2004;54:273–9.
14. Khush GS. Origin, dispersal, cultivation and variation of rice. *Plant Mol Biol* 1997;35:25–34.
15. Khush GS. Green revolution: preparing for the 21st century. *Genome* 1999;42:646–55.
16. Kilian A, Stiff C, Kleinhofs A. Barley telomeres shorten during differentiation but grow in callus culture. *Proc Natl Acad Sci U S A* 1995;92:9555–9.
17. Kim H-R, Hurwitz B, Yu Y, Collura K, Gill N, et al. Construction, alignment and analysis of twelve framework physical maps that represent the ten genome types of the genus *Oryza*. *Genome Biol* 2008;9:R45.
18. Kojima, et al. Development of an RFLP-based rice diversity research set of germplasm. *Breeding Science* 2005;55:431–40.
19. Konishi S, Izawa T, Lin SY, Ebana K, Fukuta Y, Sasaki T, Yano M. An SNP caused loss of seed shattering during rice domestication. *Science* 2006;312:1392–6.
20. Kotani H, Hosouchi T, Tsuruoka H. Structural analysis and complete physical map of *Arabidopsis thaliana* chromosome 5 including centromeric and telomeric regions. *DNA Res* 1999; 6:381–6.
21. Li C, Zhou A, Sang T. Rice domestication by reducing shattering. *Science* 2006;311:1936–9.
22. Li X, Qian Q, Fu Z, Wang Y, Xiong G, et al. Control of tillering in rice. *Nature* 2003;422:618–21.
23. Matsumura H, Takano T, Takeda G, Uchimiya H. Adh1 is transcriptionally active but its translational product is reduced in a rad mutant of rice (*Oryza sativa* L.), which is vulnerable to submergence stress. *Theor Appl Genet* 1998;97:1197–203.
24. Mizuno H, Wu J, Kanamori H, Fujisawa M, Namiki N, Saji S, Katagiri S, Katayose Y, Sasaki T, Matsumoto T. Sequencing and characterization of telomere and subtelomere regions on rice chromosomes 1S, 2S, 2L, 6L, 7S, 7L and 8S. *Plant J* 2006;46:206–17.
25. Mizuno H, Wu J, Katayose Y, Kanamori H, Sasaki T, Matsumoto T. Chromosome-specific distribution of nucleotide substitutions in telomeric repeats of rice (*Oryza sativa* L.). *Mol Biol Evol* 2008;25:62–8.
26. Monna L, Kitazawa N, Yoshino R, Suzuki J, Masuda H, Maehara Y, Tanji Y, Sato M, Nasu S, Minobe Y. Positional cloning of rice semidwarfing gene, *sd-1*: rice “green revolution gene” encodes a mutant enzyme involved in gibberellin synthesis. *DNA Res* 2002;9:11–7.
27. Nagano H, Onishi K, Ogasawara M, Horiuchi Y, Sano Y. Genealogy of the “Green Revolution” gene in rice. *Genes Genet Syst* 2005;80:351–6.
28. Riha K, Fajkus J, Siroky J, Vyskot B. Developmental control of telomere lengths and telomerase activity in plants. *Plant Cell* 1998;10:1691–8.
29. Sasaki A, Ashikari M, Ueguchi-Tanaka M, Itoh H, Nishimura A, Swapan D, Ishiyama K, Saito T, Kobayashi M, Khush GS, Kitano H, Matsuoka M. Green revolution: a mutant gibberellin-synthesis gene in rice. *Nature* 2002;416:701–2.
30. Second G. Origin of the genic diversity of cultivated rice (*Oryza sativa*): study of the polymorphism scored at 40 isozyme loci. *Jpn J Genetic* 1982;57:25–57.
31. Second G. Evolutionary relationships in the sativa group of *Oryza* based on isozyme data. *Genetic Sel Evol* 1985;17:89–114.
32. Sharma SD. Species of genus *Oryza* and their interrelationships. In: Nanda JS, Sharma SD, editors. Monograph on genus *Oryza*. Enfield: Science Publishers; 2003. p. 73–111.
33. Song X-J, Huang W, Shi M, Zhu M-Z, Lin H-X. A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat Genet* 2007;39:623–30.
34. Spielmeier W, Ellis MH, Chandler PM. Semidwarf (*sd-1*), “green revolution” rice, contains a defective gibberellin 20-oxidase gene. *Proc Natl Acad Sci U S A* 2002;99:9043–8.

35. Sykorova E, Lim KY, Kunicka Z, Chase MW, Bennett MD, Fajkus J, Leitch AR. Telomere variability in the monocotyledonous plant order Asparagales. *Proc Biol Sci* 2003;270:1893–904.
36. Tamaki S, Matsuo S, Wong HL, Yokoi S, Shimamoto K. Hd3a protein is a mobile flowering signal in rice. *Science* 2007;316:1033–6.
37. Vaughan DV, Morishima H, Kadowaki K. Diversity in the *Oryza* genus. *Curr Opin Plant Biol* 2003;6:139–46.
38. Vaughan DA, Ge S, Kaga A, Tomooka N. Phylogeny and biogeography of the genus *Oryza*. In: Hirano H-Y, editor. *Rice biotechnology in the genomics era*. Berlin: Springer; 2008. p. 219–34.
39. Wu J, Mizuno H, Hayashi-Tsugane M, Ito Y, Chiden Y, Fujisawa M, Katagiri S, Saji S, Yoshiki S, Karasawa W, Yoshihara R, Hayashi A, Kobayashi H, Ito K, Hamada M, Okamoto M, Ikeno M, Ichikawa Y, Katayose Y, Yano M, Matsumoto T, Sasaki T. Physical maps and recombination frequency of six rice chromosomes. *Plant J* 2003;36:720–30.
40. Wu KS, Tanksley SD. Genetic and physical mapping of telomeres and macrosatellites of rice. *Plant Mol Biol* 1993;22:861–72.
41. Yamane H, Ito T, Ishikubo H, Fujisawa M, Yamagata H, Kamiya K, Ito Y, Hamada M, Kanamori H, Ikawa H, Katayose Y, Wu J, Sasaki T, Matsumoto T. Molecular and evolutionary analysis of the Hd6 photoperiod sensitivity gene within *Genus Oryza*. 2008; doi:10.1007/s12284-008-9019-2.
42. Yang TJ, Yu Y, Chang SB, de Jong H, Oh CS, Ahn SN, Fang E, Wing RA. Toward closing rice telomere gaps: mapping and sequence characterization of rice subtelomere regions. *Theor Appl Genet* 2005;111:467–78.
43. Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, Baba T, Yamamoto K, Umehara Y, Nagamura Y, Sasaki T. Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene CONSTANS. *Plant Cell* 2000;12:2473–83.
44. Yoshida K, Miyashita NT. Nucleotide polymorphism in the Adh2 region of the wild rice *Oryza rufipogon*. *Theor Appl Genet* 2005;111:1215–28.
45. Yoshida K, Miyashita NT, Ishii T. Nucleotide polymorphism in the Adh1 locus region of the wild rice *Oryza rufipogon*. *Theor Appl Genet* 2004;109:1406–16.
46. Yu J, Hu SN, Wang J, Wong GKS, Li SG, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* 2002;296:79–92.