Comparative Genomics and Evolutionary Studies of *Oryza* Wild Species Using Expressed Sequences

Yukie Sano

20041807

Department of Genetics, School of Life Science, Graduate University for Advanced Studies

Plant Genetics Laboratory, National Institute of Genetics

Index

General introduction

Rice is one of the most important staple food of people more than half in the world. However, production of rice is continuously threatened by various unfavorable environmental factors such as high salinity conditions, drought, and undesirable growing temperature, and various pathogens (Cho *et al.* 2004). Efforts have been made to advance our understanding of rice genetics through numerous international collaborations and a completioj of rice genome sequencing in 2004 by The International Rice Genome Sequencing Project (IRGSP). This provides a powerful tool for identification of agronomically important genes in rice, as well as for identification of orthologous genes in other crop species and their wild relatives.

The genus *Oryza*, which has a widely diverged range of habitats, comprises 25 species in 10 different genome types (AA, BB, BBCC, CC, CCDD, EE, FF, GG, HHJJ and HHKK) including *O. sativa* that has the AA genome type (Vaughan *et al.* 2003). These species have been grouped into five species complexes, *O. sativa* complex (*O. sativa*, *O. rufipogon*, *O. nivara*, *O. glaberrima*, *O. barthii*, *O. longistaminata*, *O. meridionalis* and *O. glumaepatula*), *O. officinalis* complex (*O. officinalis*, *O. minuta*, *O. rhizomatis*, *O. eichingeri*, *O. punctata*, *O. latifolia*, *O. alta*, *O. grandiglumis* and *O. australiensis*), *O. ridley* complex (*O. ridley* and *O. longiglumis*), *O. granulata* complex (*O. granulata* and *O. meyeriana*), and others (*O. brachyantha*, *O. schlechteri* and *O. coacrata*) (Tateoka *et al.* 1963). Wild rice species can be found in Central and South America, Australia, Africa, and Asia, while a cultivated species *O. sativa* is distributed worldwide, in both temperate and tropical climate areas (Vaughan *et al.* 1994). Wild relatives of rice are important as a genetic resource to study domestication, speciation, polyploidization and evolution (Lu *et al.* 2000), and also to reveal agronomically useful traits including resistance to various stresses.

A wealth of knowledges from morphological and cytological studies (Nayer *et al.* 1973, Lu *et al.* 2000) and data from molecular markers such as isozyme (Second 1982), nuclear DNA restriction fragment length polymorphism (RFLP) (Bautista *et al.* 2001, Lu *et al.* 2002, Wang *et al.* 1992), randomly amplified polymorphic DNA (RAPD) (Ishii *et al.* 1996, Ren *et al.* 2003, Aggarwal *et al.* 1999), amplified fragment length polymorphism (AFLP) (Aggarwal *et al.* 1999, Park *et al.* 2003), simple sequence repeat (SSR) (Ishii *et al.* 2001, Ren *et al.* 2003), short interspersed element (SINE), miniature inverted-repeat transposable element (MITE) insertions (Mochizuki *et al.* 1993, Iwamoto *et al.* 1999, Cheng *et al.* 2002) and nucleotide sequences (Yoshida *et al.* 2004, Zhu *et al.* 2005, Londo *et al.* 2006) has significantly enhanced our understanding of phylogenetic relationships. However, taxonomy of the genus *Oryza* is particularly obscure, because some of them lack clear morphological traits characteristics (Second 1982, Vaughan 1989) and their phylogenetic relationships have not been fully resolved, and have been inconsistent among studies (Bao *et al.* 2004, Ge *et al.* 1999, Ge *et al.* 2002, Guo *et al.* 2005, Zhu *et al.* 2005).

To create a clear view of genome comparison in the genus *Oryza,* the Oryza Map Alignment Project (OMAP), has been performing to align DNA fragments of twelve wild genome species of rice to the *O. sativa* genome sequences (Wing *et al.* 2005, Ammiraju *et al.* 2006). BAC libraries of the wild relatives have been constructed and their end sequences were read to use them for constructing alignment physical maps of 12 species (Wing *et al.* 2005, Ammiraju *et al.* 2006). In contrast to the progress of comparative genomics in wild relatives, identification and characterization of genes remain to be studied. Only gene expression and EST analysis of *O. minuta* (BBCC genome) has been reported (Cho *et al.* 2004, Cho *et al.* 2004 (a), Shim *et al.* 2004, Cho *et al.* 2005)

In this study, I examined diversity of *Oryza* species by comparing expressed sequences of AA genome with those of BB and CC genomes, together with a comparison of several genomic sequences of unique genes among AA, BB, CC, DD, EE and FF genomes. Three questions were focused on this study. First, how many nucleotides have been substituted after divergence of AA, BB and CC genomes? Second, when did the divergence of nine genomes of the genus *Oryza* has been occurred? Third, what unique genes have been evolved in the wild relative of rice? Answers to these questions would provide new information on the phylogenetic relationship of rice species, and help to discover useful genes lost in the cultivated rice.

Chapter I

Expressed gene diversity among AA, BB and CC genome species of rice

Introduction

Three cultivated species and five wild species comprise the *O. sativa* complex of the AA genome group in the genus *Oryza*. Prenial type wild species *O. rufipogon* and annual type *O. rufipogon* (*O. nivara*) have been reported to be the origin of cultivated rice *O. sativa, japonica* and *O.sativa, indica*, respectively (Cheng *et al.* 2003, Li *et al.* 2006). As for the speciation and evolution of cultivated rice of *O. sativa, japonica* and *indica*, there have been many kinds of studies. In this area much interest has been focused on the relationship between domestication and evolution. Recent studies have revealed phylogenic relationships among AA genome species as relatively close species of cultivated rice by using several different markers of SSR (Chen *et al.*, 2002), SINEs (Cheng *et al.*, 2002) and of multiple genes (Tang *et al.*, 2006). Especially, Tang *et al.* (2006) discovered the presence of high divergence regions between *O. sativa, japonica* and *indica* genome sequences that covers above 6% of the genome. In the high divergence regions, they reported the presence of genes showing different haplotypes between *O. sativa, japonica* and *indica*. Examination in the distribution of genes of two haplotypes was carried out using 60 accessions covering all AA genome species and with 20 genes in the regions. Results revealed that all eight AA genome species have genes of both *japonica* and *indica* haplotypes together in the high divergence regions, suggesting two different origins of the AA genome.

In the stream of comparative genomics, cereal crops have shown to have genome-wide syntenic relationships among their genus. Rice, maize, wheat, barley, sorghum and sugarcane were probed to possess region by region similarity through the whole genome (Devos, 2005). Using this close correspondence many phylogenic analyses have been carried out among different genus of crop plants, in the tribe *Oryzeae* as well as in the genus *Oryza* and established clear view of phylogenic relationships among them (Ge *et al.*, 2002,

Swigonova *et al.* 2004, Guo and Ge, 2005). Thus the evolutionary and phylogenic studies among different genus have revealed the degree and the time of divergence among genus, and those studies within the AA genome species have also indicated origins of cultivated rice and AA genome species. However, the phylogenic relationships among remote relatives of rice carrying different genome types from AA (BB~HHKK) in the genus *Oryza* have not been clarified well yet. A few analyses using different *Oryza* species have been carried out and showed different phylogenetic trees depending on a few genes or sequences used in the analyses (Guo and Ge, 2005, Aggarwal *et al.*, 1999). As for studies on genome diversity in the genus *Oryza*, recent works of the Oryza Alignment Map Project (Wing *et al.* 2005, Ammiraju *et al.* 2006) and the EST analysis of a tetraploid species of *O. minuta* (BBCC) (Shim *et al.* 2004, Cho *et al.* 2005) have begun to reveal characteristics of diverged genomes and genes.

In this chapter, I would like to focus on the sequence diversity of the expressed genes of the different genome species in the genus *Oryza*. To built up fundamental view of interspecific divergence among wild and cultivated species, over 3,000 expressed sequence tags (ESTs) from two wild species of BB and CC genomes were analyzed by comparing them with the orthologous sequences of *O. sativa, japonica*, cv. Nipponbare and *O. sativa, indica*, cv. 93-11. How many nucleotide and amino acid substitutions have been occurred among genomes in the genus *Oryza*? Are there specific genes that evolved rapidly or differently? If there were, what is the reason for rapid evolution? Answers to these questions would provide a new clue for the phylogenic studies of wild genome species of rice and a discovery of missing and evolving interesting genes, such as a high performance stress tolerance genes and reproductive barrier genes.

Materials and Methods

Plant materials

Wild species and accessions of rice used in this study were *O. punctata* (accession number W1514) for BB genome species and *O. officinalis* (W0002) for CC genome species. The seeds of each accession maintained in the National Institute of Genetics were sown and grown in the paddy field in the city of Mishima, Shizuoka, Japan. Two organs, shoots of meristem major and young panicles were used for mRNA extraction. Around five leaves stage, shoots were collected from W1514 (BB) and young panicles of about 2-4 cm long were harvested from W0002 (CC) plants and frozen in liquid nitrogen. These plant materials were prepared by Ms. Miyabayashi (Expermintal farm, National Institute of Genetics, Japan).

cDNA library construction

Total RNA was extracted from the frozen plant materials by the oligotex^{TM}-MAG mRNA extraction kit (TAKARA, Japan) following the instruction of the manufacturer. Poly(A)+ RNA was purified by adding the Dynabeads oligo $(dT)_{25}$ (Dynal, USA) into total RNA solution, washing and eluting the polyA RNA from the beads. All procedure and buffers used in the RNA purification followed the manufacture's instruction (Dynal, USA). Poly(A)+ RNA of about 5 mg each was used for the first and then the second strand synthesis of cDNA using a ZAP cDNA synthesis kit (Stratagene, USA). Synthesized cDNA was blunted with DNA polymerase, ligated with EcoRI adaptors, digested with an EcoRI and then the adaptor ends were phospholyrated by a T4 polynucleotide kinase. After digesting with XhoI, cDNAs were fractionated through Shepharose CL-2B column and collected cDNAs of 500 bp to 5kbp fragments. Collected cDNAs of about 100 ng each was ligated to 1 mg Uni-ZAP XR vector and packaged with the lambda phage Gigapack III XL packaging extract (Stratagene, USA) and infected immediately to the *E. coli* strain Xl-Blue MRF'. Total of 5×10^4 and 3.5×10^5

packaged phages were obtained for BB genome shoot (BBS) library and CC genome panicle (CCP) library, respectively. A part of the libraries was used for excision reaction to obtain plasmid clones using a SOLR™ strain. From cDNA synthesis to the excision culture, all procedures and buffers followed strictly the manufacturer's instruction (Stratagene, USA). Total of 4000 excised clones, 2000 each from the BBS and the CCP library were picked up, cultured in a LB medium and extracted plasmid DNA from individual clones. Library construction and characterization were mainly performed by Ms. Sasaga (Plant Genetics Laboratory, National Institute of Genetics, Japan).

Sequencing of cDNA clones

Sequencing of 4000 randomly selected clones, 2000 from *O. punctata* (BB) and another 2000 from *O. officinalis* (BB), were performed in ABI 910 sequencer (Aplied Biosystems). All clones were sequenced from both 3' and 5' ends with M13 primer (5'-GTAAAACGACGGCCAGTG-3') and from the 3' end with M13 reverse primer (5'-GGAAACAGCTATGACCATGA-3') using Big Dye terminator Kit (Aplied Biosystems). Sequences difficult to be determined or shorter less than 100 bp were excluded from the analysis. The sequence profile was checked manually, and regions including anonymous nucleotide (N) were deleted from the analysis. All sequences were registered in the DDBJ as EST sequences with numbers of BB984512~ BB985121, BB994797~BB996458 for singletons, and as full insert sequencies cDNAs with numbers AK223775~AK224590. This part was conducted by Drs. Kanamori and Namiki (Society for Techno-innovation of Agriculture, Forestry and Fisheries, Japan).

Similarity search, detection of nucleotide substitution and insertion-deletion (indel), and

estimation of Ks and Ka

Similarity of all sequenced clones from BBS and CCP libraries was searched against the ESTs, cDNAs and genome sequences of *O. sativa, japonica* (February, 2005) and Genbank nonredundant database (release 151) by the BLAST program (Altschul *et al.* 1997). The *japonica* genes that showed the highest similarity were defined as orthologues of the clones in this study.

The nucleotide substitution was detected also using the BLAST program against the *O. sativa* (data based on the IRGSP genome sequence build 3). The all coding sequence (CDS) files and the all UTR files were retrieved from Rice Annotation Project database (RAP-db) and used for the alignment analysis with the ESTs sequenced in this study. Gaps on either sequences were deleted as indels and all nucleotide mismatches were counted as substitutions for the 3' and 5' UTR and the CDS regions respectively. Gap number was also counted as indel number in all the 3' and 5' UTR and the CDS. These BLAST results in the alignment analysis were also used to calculate Ks and Ka with a K-estimator (Comeron 1999). K-estimator applies the method described in Comeron (1995) to estimate Ks and Ka. This method is a modification of method of Li's (1993) and Pamilo and Bianchi's (1993) (LPB) method. Before using the K-estimator, putative codonframes were adopted for the CDSs of wild relatives and the gaps were eliminated from either the BB, CC or the AA genomes. The *indica* orthologues were obtained using the BLAT program against the *indica* genome sequences (Yu *et al.* 2005) with the identified *O. sativa, japonica* sequences as a query. Ks and Ka between *O. sativa, japonica* and *indica* and between BB or CC and *indica* CDS were calculated using the results of this blat analysis. Computer analysis was performed by Ms. Mochizuki (Plant Genetics Laboratory, National Institute of Genetics, Japan).

Results

EST sequences in the BBS and the CCP libraries

Manually confirmed sequences of 1222 from the *O. punctata* BBS library and 1865 from the *O. officinalis* CCP library were used for the analysis of ESTs. Average clone length of the library was 630 bp calculated from insert sizes of 100 clones for two libraries. Sequences from same genes were assembled by BLAST search and minimum contiguous clone sequences, as well as stand alone sequences were collected for further analysis. Similarity search of the assorted sequences were performed against the DNA data of all organisms in release 151 of GenBank. Genomic DNA sequences that contained putative introns or vector sequences and DNA sequence less than 100 bp long were deleted from the analysis. Total of 973 clones composed of 1222 sequences from the *O. punctata* BBS library and 1425 clones composed of 1866 sequences from the *O. officinalis* CCP library were selected for comparative analysis with the AA genome sequences. Similarity search at the nucleotide level using BLASTX revealed that 86.2 and 77.3% plant genes and 14.6% genes of other organisms showed similarity to BBS and CCP clones, respectively (Table I-1). 2.7% and 8.1% of BBS and CCP clones did not show similarity with any protein, respectively (Table I-1). Table I-2 shows similarities of the BB and the CC clones with the cDNA or genome DNA sequences of AA genome *O. sativa, japonica* rice at the nucleotide level using BLASTN. The members that showed no similarity to the genomic sequences of AA genome suggested the presence of highly diverged specific genes/sequences in both BB and CC genomes. Meanwhile, novel expressed gene sequences, which appered in *O. sativa* genome sequence but not in any cDNAs or ESTs from AA species, were 20 of 2398 clones in the BBS and CCP libraries. This indicates that at least 0.8% of the BBS and CCP genes/sequences analyzed in this study has not been identified yet in the cultivated rice.

Substitution finding in the ESTs of BB and CC genomes against AA genome sequences

Nucleotide substitutions of all BBS and CCP clones against AA genome sequence were investigated by the BLAST in this study. The genome sequences both of *O. sativa, indica* and *japonica* were used for substitution analysis. After excluding indel regions or non-homologous sequences, total of 465107 bp of *O. punctata* (BB) and 493511 bp of *O. officinalis* (CC) were used for substitution detection, and 17294 and 20267 substitutions were detected against *O. sativa, japonica* sequences, respectively (Table I-3). This showed that the base substitution rate in the ESTs was 37 bp/kb in *O. punctata* (BB) and 41 bp/kb in *O. officinalis* (CC) compared to the *O. sativa* (AA) sequence (Table I-3). In Table I-3, numbers of substitutions were counted separately in the untranslated regions (UTRs) and in the CDSs, indicating that substitutions in the CDSs were occurred more frequently in the CC genomes than in the BB (38 bp/kb in CC and 32 bp/kb in BB). On the other hand, the substitutions in the UTRs were occurred evenly in the BB and CC genomes (46 bp/kb in both genomes). Comparison of genomic DNA sequences of *O. punctata* (BB) and *O. officinalis* (CC) against *O. sativa, japonica* (AA) genome has also been carried out using each 100 BAC end sequences retrieved from the OMAP dataset in the Arizona Genomics Institute (AGI). The end sequences of BB and CC BAC clones showed high substitution rate of 82 bp/kb and 98 bp/kb against *O. sativa, japonica* genome. The average substitution rate in the genome sequence against *japonica* was about three-fold higher (82/32 and 98/38 bp/kb in BB and CC) than that in the CDS and two-fold higher (82/46 and 98/46 bp/kb in BB and CC) than that in the UTR as shown in Table I-3. These results showed that the CDS region is highly conserved and the UTR is also conserved in their sequence than the genomic region.

Comparison of Ka/Ks ratio among AA, BB, and CC genomes

The synonymous substitution rate Ks and non-synonymous substitution rate Ka was calculated for all combinations among ESTs of *O. punctata* (BB genome), *O. officinalis* (CC genome), and corresponding genomic sequences of *O. sativa, japonica* and *indica* (AA genome). Average Ks, Ka and Ka/Ks values in three genomes against *O. sativa, japonica* genome are shown in Table I-4. All Ks, Ka and Ka/Ks values for *O. sativa, japonica* AA genome were slightly higher in the CC genome of *O. officinalis* (0.10178, 0.01852 and 0.18196, respectively) than that of BB genome of *O. punctata* (0.09312, 0.01646 and 0.17676, respectively). Ks values against *indica* AA genome were slightly higher from those against *O. sativa, japonica*, while Ka and Ka/Ks values against *indica* were slightly lower from those against *O. sativa, japonica.*

Genes that show high Ka/Ks ratio should be expected to evolve rapidly during speciation. The scatterd plot of Ka and Ks values for all analyzed genes were shown in Fig. I-1. To detect rapidly evolving genes with 1.0 or higher than 1.0 Ka/Ks value were picked up. Number of ESTs with 1.0 or higher than 1.0 Ka/Ks was 31 in BB and 41 in CC when compared with *O. sativa, japonica* genes. Thirty-three ESTs with high Ka/Ks value beyond 1.5 were listed in Table I-5. Of them, 15 were hypothetical proteins. Eight ESTs of them (Accession No. AK224183, BB995654, BB984978, BB996395, AK224391, BB984792, BB995562 and AK223902) showed lower Ka/Ks values against *indica* sequence.

Indels in the ESTs among AA, BB, and CC genome species

In all ESTs analyzed here, many indels were identified between AA and BB, and AA and CC genomes. Total indel events between *O. sativa, japonica* (AA) and BB were 214 of 1222 ESTs analyzed, and between *O. sativa, japonica* and CC were 290 in 1865 ESTs analyzed (Table I-6). The length of each indel ranged

from one to hundreds of nucleotides. The frequency of indel in the EST sequences was also analyzed in UTR and CDS. The rate of indel occurrence was 17 bp/kb in UTR and 2 bp/kb in CDS. This showed that the indel was accumulated much in the UTR but not in the CDS. Three typical examples of comparative gene structures between AA and BB genomes are shown in Fig. I-2. Fig. I-2(A) showed orthologues with an inframe indel and nucleotide substotutions. The genes with a frame shift indel and with putative termination site were shown in (B) and (C), respectively. Rate of nucleotide substitution was almost same between the CDS and UTR. However, indel was accumulated much in UTR, while small number of indel was detected in the CDS.

Discussion

Substitutions and indels among three genomes

Substitutions counted separately for the 3' and 5'untranslated regions (UTRs) and for the coding sequences (CDSs) showed different rate and was slightly more accumulated in the CC than in the BB genome in the CDS, whereas UTR of the CC genome showed just same rate as that of the BB genome (Table I-3). The results indicate a little far relation of the CC genome than the BB genome from the AA genome. The substitution rate in the noncoding region between *O. sativa, japonica* and *indica* was 15/kb, while in the CDS and UTR the rate were 3.0 bp/kb and 4.6 bp/kb, respectively (Yu *et al.* 2005). Comparison of the results from Yu *et al.* (2005) and that of this study showed that in the noncoding region, accumulation of substitution was about six times higher in the BB and the CC genomes than that of the *indica* genome. Meanwhile, the substitution rate was about 10 times higher in the CDS of the BB and CC genomes than that of the *indica* genome aginst japonica genome. The reason why the substitution was accumulated higher in the CDS region

than the whole genome sequences in the wild genome species of BB and CC genomes compared to the *indica* genome is not known. Some selective pressures for domestication might suppress the gene variation between two cultivated subspecies of *O. sativa, japonica* and *indica* by selecting common characteristics convenient for cultivation. On the contrary, the higher substitution rate in the CDS might allow the wild species of BB and CC genomes to survive in the wild habitat under various natural selective pressures. However, a conclusion should be awaited for the results of the Ka rate of these different genomes.

The indel rate in the EST of *O. punctata* (BB) and *O. officinalis* (CC) were very similar with each other and is about 2 bp/kb in the CDS, 17 bp/kb in the UTR and 11~14 bp/kb in the whole genome (Table I-6). This showed that the rate of indel in the CDS of BB and CC genome species is one third of that in the UTR and one fifth of that in the whole genome region. While in the case of *O. sativa indica*, the indel was accumulated five times more in the UTR and ten times more in the genome region than that accumulated in the CDS (Yu *et al.* 2005). The indel rate showed that the indel was highly accumulated in the UTR of two wild species compared to that of the *indica*. On the contrary, the indel rate in the CDS is slightly higher in the wild species than in the cultivated subspecies. These results are different from the case of substitution. However, in many cases the indels accumulated in the CDS as triplet nucleotide insertion or deletion (Fig. I-2), suggesting that almost indels occurred inframe and many deleterious mutations should be discarded from the population. Comparison of the rate of non-frame-shift and frame-shift mutations in CDS would be necessary to analyze the correlation of it with the genetic variation.

In most of the cases of the comparison among three different genomes, the substitution and the indel rates are a little higher in the UTR than in the CDS but about one third of that in the whole genome average (Table I-6). This indicated that the UTR has been evolved under positive selection and involved in the formation of

genetic variation through expression specificities or protein stabilities of the genes in rice.

Ks, Ka and Ka/Ks values and their relation to genetic diversity and evolution

The distribution of Ka/Ks between BB or CC genomes and *O. sativa, japonica* showed many genes above 1.0 in Fig. I-1(A) and (B). The gene with a Ka/Ks value 1.0 is assumed under neutrality, above one under adaptive selection and below one under purifying selection. This study showed a total of 31 and 41 ESTs that Ka/Ks was above one, and 12 and 21 ESTs of above 1.5 value in the BB and the CC genomes, respectively (Fig. I-1, Table I-5). Have these genes evolved rapidly under positive selection? A hypothetical protein of clone 08G05, an RNA dependent RNA polymerase family protein and a no apical meristem protein moreover showed different Ka/Ks values against *indica* genomes. These genes with highly variable sequences might have evolved specific functions and so might take some roles in the evolutionary process. To confirm the evolutionary characteristics of these genes, Ka/Ks values in the full length cDNAs should be estimated. Additional Ka/Ks examination of orthologous genes in the different genome species would also be needed to reveal evolutional features of these genes.

Comparison of expressed sequences

The novels expressed sequences found in the BBS and the CCP libraries were 3% and 8%, respectively. The comparison of gene sequences between subspecies of *indica* and *japonica* in the *O. sativa* reported that the unique gene sequences for either the two subspecies were 2-3% (Yu *et al.* 2005). The ratio of species specific expressed sequences was not high in the BB and CC genome species compared to the unique gene ratio between *O. sativa, japonica* and *indica* genomes. The difference of unique gene ratio in each species was not

comparable to the difference of sequence variations in the CDS, but observed in each species in UTR and non coding regions against *O. sativa* sequence. Therefore, it is not so simple to search and identify species specific useful genes from wild relatives of rice. More questions arose in the results of this analysis. To resolve complicated evolutional relationships among rice genomes, species and genes, comparative genomics approach for both genome-wide and gene-to-gene level is needed.

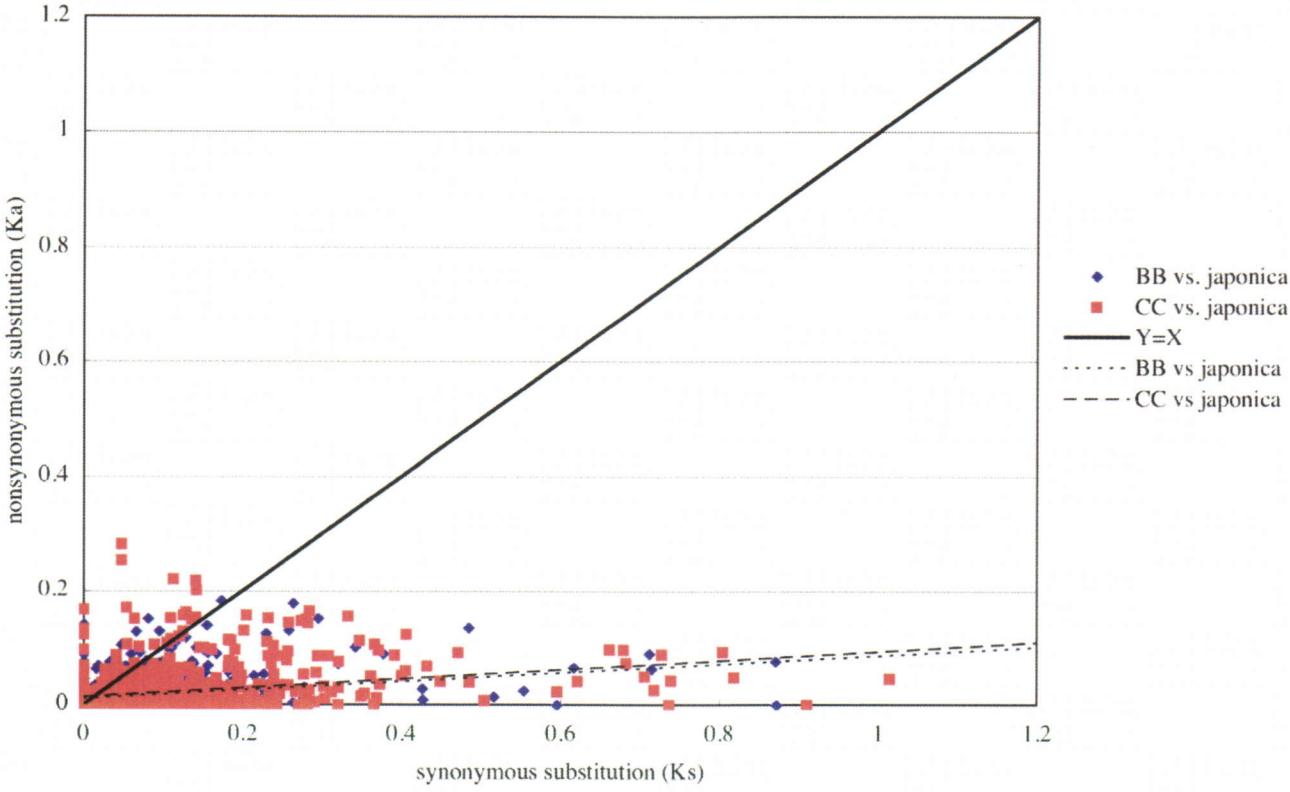


Fig. I-1. Distribution of genes with various Ka/Ks ratio between AA and BB, and AA and CC genome, respectively.

Blue dots showed Ka/Ks of BB genes against AA genes, red dots showed Ka/Ks of CC genes against AA genes. Broken lines showed the linearized line.

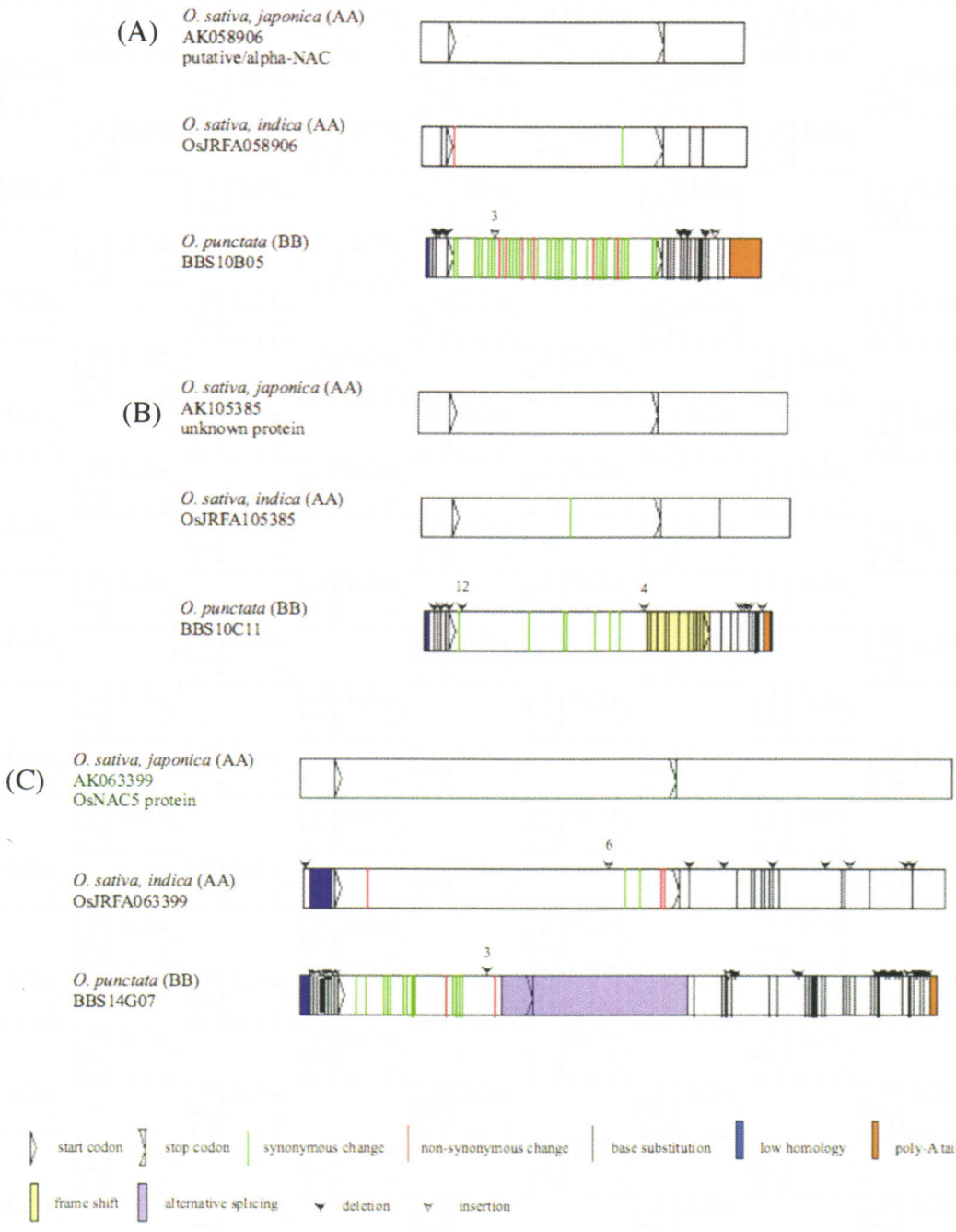


Fig. I-2. Structural comparison of three genes between AA and BB genomes.

Comparison between *O. sativa, japonica* and *indica* or *O. punctata* showed in (A) for AK058906, in (B) for AK105385 and in (C) for AK063399. Explanation of symbols used in this figure was showed at the bottom of the figure. The nucleotide numbers of indel were shown on the indel marks,except that marks with no figure were indels of single nucleotide.

Table I-1. Similarity search of BBS and CCP clone sequences at the amino acid level using BLASTX

Organisms showing the highest similarity	BBS		CCP	
	clones*	% of total	clones*	% of total
Plants	839	86.2	1102	77.3
Oryza	742		977	
Other Plants	85		103	
Animals, Micro Organisms etc.	108	11.1	207	14.6
No Hit	26	2.7	116	8.1
Total	973		1425	

* The number of clones that showed highest similarity to the protein of indicated organisms

Table I-2. Similarity search of BBS and CCP clones at the nucleotide level using BLASTN

Organisms showing the highest similarity	BBS clones	CCP clones
BLASTN for *japonica* cDNA		
Hit	894	1242
No hit*	79	183
BLASTN for *japonica* genome DNA		
Hit	863	1370
No hit*	110	55
Total	973	1425

*No hit shows sequences with E-value of below 1E-05

Table I-3. Substitution number and rate of the BB and CC EST sequences against AA *O. sativa, japonica* genome

Substitutions in the EST	vs AA *japonica* (Nipponbare)		
	O. punctata (BB)	*O. officinalis* (CC)	*O. sativa indica* (AA)**
analyzed clone number	1222	1865	49088 annotated genes
sequence length*			
(1) CDS	297862 bp	301199 bp	49088 annotated genes
(2) UTR	167245 bp	192312 bp	49088 annotated genes
Substitutions number			
(3) CDS	9534 bp	11369 bp	
(4) UTR	7760 bp	8898 bp	
Substitution rate			
CDS [(3)/(1)]	32 bp/kb	38 bp/kb	3.0 bp/kb
UTR [(4)/(1)]	46 bp/kb	46 bp/kb	(4.6 bp/kb)
EST [(3+4)/(1+2)]	**37 bp/kb**	**41 bp/kb**	**(3.2 bp/kb)**
Substitutions of the genome	vs AA *japonica* (Nipponbare)		
(BAC end sequence)	*O. punctata* (BB)	*O. officinalis* (CC)	*O. sativa indica (AA)*
clone number	100	100	whole genome sequence
sequence length*	31127 bp	38641 bp	466.3 Mb
Substitutions number in the genome	2555 bp	3774 bp	
Substitutions rate in the genome	**82/kb**	**98/kb**	**15 bp/kb**

* Sequence length indicates total nucleotide length in bp matched with EST and genome sequences of *japonica* or *indica* in the blast analysis.

** Data was quoted from Yu *et al.* (2005)

Table I-4. Ka, Ks and Ka/Ks rates of the CDS in the BB and CC genomes against AA *O. sativa, japonica* or *indica* genomes, and between *japonica* and *indica* genomes

		Average substitution rate in the CDS		
		vs O.sativa, japonica (AA)	*vs O. sativa, indica* (AA)	substitution rate between *japonica* and *indica*
O. punctata (BB)	Ks	0.09312	0.09847	0.01242
	Ka	0.01646	0.01603	0.00174
	Ka/Ks	0.17676	0.16279	0.14010
O. officinalis (CC)	Ks	0.10178	0.10473	0.01323
	Ka	0.01852	0.01790	0.00221
	Ka/Ks	0.18196	0.17092	0.16703

Ks: rate of substitutions per silent site, Ka: rate of nonsynonymous substitutions per silent site,

Ka/Ks: ratio of nonsynonymous to synonymous substitution

Table I-5. CDS of BB and CC genomes that showed Ka/Ks value above 1.5

Genome	Accession No.	Clone	CDS annotation	vs *japonica*				vs *indica*			
				NT No.	Ks	Ka	Ka/Ks	NT No.	Ks	Ka	Ka/Ks
BB	AK224220	17A09	Hypothetical protein.	402	0.00745	0.07186	9.646	114	0	0.06858	-
	BB995130	10G08F	Conserved hypothetical protein.	168	0.01808	0.06887	3.809	150	0.02034	0.0783	3.850
	AK224005	05A10	PAK-box/P21-Rho-binding domain containing protein.	54	0.06908	0.2568	3.717	54	0.06908	0.2568	3.717
	AK223951	01F07	Hypothetical protein.	288	0.01748	0.05758	3.294	69	0	0.06968	-
	BB995101	10A10F	Phosphoesterase PHP, N-terminal domain containing protein.	426	0.01436	0.03424	2.384	390	0.01592	0.03532	2.219
	AK224141	12C07	RNA-binding protein.	126	0.04745	0.10495	2.212	126	0.04745	0.10495	2.212
	AK224183	14E03	Multi antimicrobial extrusion protein MatE family protein.	261	0.01308	0.02671	2.042	123	0.02916	0.01164	0.399
	AK223775	01B08	Hypothetical protein.	252	0.09695	0.19326	1.993	252	0.09695	0.19326	1.993
	BB995327	18B03F	Cyclin-like F-box domain containing protein.	360	0.06575	0.12861	1.956	129	0.0867	0.14569	1.680
	AK224084	09E05	Phytosulfokines 2 precursor	240	0.0247	0.04519	1.830	240	0.0247	0.04519	1.830
	BB995067	09D09R	RNA-binding region RNP-1 domain containing protein.	312	0.0099	0.01797	1.815	312	0.0099	0.01797	1.815
	BB994801	01B07R	Heat shock protein DnaJ family protein.	321	0.02054	0.03078	1.499	321	0.02054	0.03078	1.499
CC	BB996059	13F10R	Conserved hypothetical protein.	120	0.05766	0.2269	3.935	120	0.05766	0.2269	3.935
	BB995472	01H05F	Peptidase S14, ClpP family protein.	138	0.05412	0.17042	3.149	138	0.05412	0.17042	3.149
	BB995911	11A03R	TB2/DP1 and HVA22 related protein family protein.	66	0.05421	0.1547	2.854	66	0.05388	0.12522	2.324
	BB995654	05F06R	Riboflavin biosynthesis protein RibD.	270	0.02181	0.05716	2.621	231	0.02544	0.03041	1.195
	AK224547	18E04	Conserved hypothetical protein.	252	0.02673	0.06007	2.247	171	0	0.02532	-
	BB984812	08F04R	Splicing factor, CC1-like family protein.	99	0.03383	0.07168	2.119	99	0.03383	0.07168	2.119
	AK223880	12D10	Hypothetical protein.	204	0.01463	0.02986	2.041	39	0	0.0423	-
	BB984978	16B05R	Conserved hypothetical protein.	54	0.05655	0.113	1.998	51	0.06056	0.08284	1.368
	BB984781	06H05F	TPR-like domain containing protein.	78	0.11316	0.22144	1.957	78	0.11316	0.22144	1.957
	AK224477	13A08	Harpin-induced 1 domain containing protein.	591	0.05461	0.10627	1.946	588	0.05484	0.10975	2.001
	BB996395	20E09F	Hypothetical protein.	96	0.06934	0.13335	1.923	57	0.0761	0.04902	0.644
	AK224463	12B09	Hypothetical protein.	237	0.01294	0.0233	1.801	237	0.01294	0.0233	1.801
	BB995634	05C06F	RNA dependent RNA polymerase family protein.	354	0.02712	0.04707	1.736	354	0.02711	0.04343	1.602
	AK224391	07G06	No apical meristem (NAM) protein domain containing protein.	243	0.0126	0.02167	1.720	141	0.0216	0.01936	0.896
	BB984792	07E07F	Hypothetical protein.	162	0.01832	0.03129	1.708	153	0.01928	0.01491	0.773
	AK224486	13G11	Conserved hypothetical protein.	249	0.01288	0.02196	1.705	249	0.01288	0.02196	1.705
	BB995562	03H09R	Nucleic acid-binding OB-fold domain containing protein.	165	0.0407	0.06924	1.701	75	0.1004	0.09535	0.950
	AK223902	16B08	Conserved hypothetical protein.	408	0.02285	0.03813	1.669	156	0.04154	0.01975	0.475
	BB995890	10E06R	Conserved hypothetical protein.	153	0.04215	0.0661	1.568	153	0.04215	0.0661	1.568
	BB984870	11D07R	Glycine-rich RNA-binding protein PsGRBP.	162	0.01767	0.02687	1.521	156	0.01836	0.02041	1.112
	BB984809	08E11F	Conserved hypothetical protein.	102	0.04002	0.0607	1.517	102	0.04002	0.0607	1.517

Table I-6. Indel number and rate of the BB and CC EST sequences against AA *O. sativa, japonica*

Indels in the EST	vs AA *japonica* (Nipponbare)		
	O. punctata (BB)	*O. officinalis* (CC)	*O. sativa indica* (AA)**
clone number	1222	1865	49088 annotated genes
sequence length*			
(1) CDS	297862 bp	301199 bp	49088 annotated genes
(2) UTR	167245 bp	192312 bp	49088 annotated genes
Indels number			
(3) CDS	604 bp	743 bp	
(4) UTR	2770 bp	3311 bp	
Indel rate			
CDS [(3)/(1)]	2.0 bp/kb	2.5 bp/kb	(0.22 bp/kb)
UTR [(4)/(2)]	16.6 bp/kb	17.2 bp/kb	(1.01 bp/kb)
EST [(3+4)/(1+2)]	**7.3 bp/kb**	**8.2 bp/kb**	**(0.79 bp/kb)**
Indels of the genome	vs AA *japonica* (Nipponbare)		
(BAC end sequence)	*O. punctata* (BB)	*O. officinalis* (CC)	*O. sativa indica (AA)*
clone number	100	100	whole genome sequence
sequence length* (bp)	31127 bp	38641 bp	466.3 Mb
Indels number in the genome	2555 bp	3774 bp	
Indels rate in the genome	**14.1 bp/kb**	**11.2 bp/kb**	**2.89 bp/kb**

* Sequence length indicates total nucleotide length in bp matched with EST and genome sequences of *japonica* or *indica* in the blast analysis.

** Data was quoted from Yu *et al.* (2005)

Chapter II

Phylogenetic analysis of wild relatives of rice

Introduction

The genus *Oryza* comprises 25 species in ten different genome types (AA, BB, BBCC, CC, CCDD, EE, FF, GG, HHJJ and HHKK) and these species have been grouped into five species complexes (see general introduction). Due to the wild distribution over the world of many different species, wild relatives of rice are important as genetic resources just as shown by useful traits including resistance to various stresses and diseases. They have been used as genetic resources for breeding to acquire the hybrid rice revolution, yield enhancing traits (Zhang *et al.* 2006, Ashikari *et al.* 2005) and tolerance to biotic and abiotic stresses (Chang *et al.* 1976, Vaughan 1989).

The cultivated rice has been analyzed well, and genome sequencing of two *O. sativa* subspecies *japonica* and *indica* has been completed (IRGSP November 18, 2004). It becomes a powerful tool to employ wild relatives of rice for functional, comparative and evolutionary analyses of the genus *Oryza* and the cereal genomes.

For the academic and practical importance of rice, classification and phylogeny of *Oryza* have been studied in many papers as described in the general introduction. Recently, nucleotide sequence data have been used for phylogenetic investigation (Soltis *et al.* 1998, Sang 2002). Intron sequences with sufficient variations are also useful for phylogenetic analysis in close taxonomic relationship (Doyle *et al.* 1996, Oh *et al.* 2003, Zhu *et al.* 2005). A recent literature has demonstrated that genes with the low-copy number have a great potential to improve the resolution of plant phylogenetic reconstruction (Yoshida *et al.* 2004, Zhu *et al.* 2005, Londo *et al.* 2006, Doyle *et al.* 1999, Grob *et al.* 2004). An alcohol dehydrogenase (*Adh*) gene and *GPA1* that encodes the G protein subunit were low-copy number nuclear genes were used to investigate the phylogeny of *Oryza* (Bao *et al.* 2004), whereas the chloroplast *matK* gene, *trnL* intron and the mitochondrial nad1 intron 2 were

also used in phylogenetic studies (Soltis *et al.* 1998, Sang 2002, Ge *et al.* 1999, Ge *et al.* 2002, Guo *et al.* 2005, Gugerli *et al.* 2001, Hass *et al.* 2003). Guo et al. (Guo *et al.* 2005) studied the phylogeny of *Oryza* using two chloroplast (*matK* and *trnL* intron), one mitochondrial (*nad1* intron 2), and two nuclear (*Adh2* and *GPA1*) sequences from each genome. They verified that the tribe was a monophyletic group consisting of two strongly supported monophyletic groups corresponding to the two traditionally recognized subtribes. Based on the magnitude of *matK* and *GPA1* sequence divergence, they suggested that the most basal split in the *Oryza* (*O. granulata*, GG genome vs. remaining species) branched ~9 million years ago (mya).

Although many phylogenetic analyses of the genus *Oryza* have been reported, only two nuclear genes and two chloroplast DNAs (cpDNA) were used in the studies and results were not coincided well with each other. In the present study, I reconstructed the phylogeny of *Oryza* using nine nuclear DNAs; three *O. sativa* full length cDNA clones (AK058507, AK059353 and AK098919) (Kikuchi *et al.* 2003), RAFTIN 1 (AJ575667) (Wang *et al.* 2003), MAP kinase 5 (AF479884) (Xiong *et al.* 2003), fructokinase 1 (AF429948) (Jiang *et al.* 2003), teosinte branched 1 (*tb1*), tryptophane synthase (*orp*) (Swiginiva *et* al. 2006) and *Adh1* 3rd intron (*Adh1*-i3)(Zhu *et al.* 2005)(Table II-1). The divergence time was estimated based on a molecular clock approach using nucleotide sequences. These results should give additional information about phylogenetic relationship of the genus *Oryza*.

Materials and methods

Plant materials

Genome and species used in this study are shown in Table II-2. These accessions have been maintained in the National Institute of Genetics as self-pollinated seeds.

PCR amplification and DNA sequencing

The comparison was carried out on AA-FF genome *Oryza* species (Table II-2). The corresponding genome DNA fragments of the *O. sativa* full length cDNA clones (AK058507, AK059353 and AK098919), RAFTIN 1 (AJ575667), MAP kinase 5 (AF479884), fructokinase 1 (AF429948), teosinte branched 1 (*tb1*), tryptophane synthase (*orp*) and *Adh1* 3rd intron (Table II-1) were amplified from genomic DNA of each species with gene-specific primers shown in Table 3 using Ex Taq polymerase (Takara Bio) according to the manufacture's instruction. Some amplified products were cloned into a pT7Blue-T vector (Novagen). DNA sequencing was carried out using a Big Dye terminator Kit (Aplied Biosystems) with an automated DNA capillary sequencer ABI 3130xl.

Data analysis

In the phylogenetic analysis, sequences were aligned with clastalW (Higgins *et al.* 1994) and refined manually. Sequence divergence and phylogenetic analysis were performed using the Kimura's two-parameter model (Kimura 1980), the Neighbor-joining method (Saitou *et al.* 1987) and bootstrap analysis with 1000 replicates by MEGA 3.1 (Kumar *et al.* 2004). In the intron sequences, mononucleotide repeats were excluded from phylogenetic analysis. Rates of synonymous and non-synonymous substitution were calculated by K-estimator (Comeron 1999). For an estimation of the divergence time of each gene in the genus *Oryza*, a simple molecular clock assumption was employed and clock calibration was carried out based on the assumption that maize and rice diverged 50 million years ago [mya] (Gaut 2002). The divergence time of every genes was calculated by MEGA 3.1 (Kumar *et al.* 2004). Protein domains were searched by the

InterProScan in the EBI database (http://www.ebi.ac.uk/InterProScan/).

Results

Phylogenetic analysis using coding sequences and genome sequences

Phylogenetic analyses were carried out to resolve the relationship of species and genomes in the genus *Oryza*. The analysis was performed with coding sequences (CDSs) and genome sequences. Eight genes and one intron used in this study are presented in Table II-1 and the genome structures of used genes were shown in Fig. II-1. They have introns except for Os02g0121300 (AK098919) and *tb1*. Length of CDS were 591 bp (AK058507), 240 bp (AK059353), 519 bp (AK098919), 1239 bp (RAFTIN1), 798 bp (MAP kinase 5), 972 bp (fructokinase 1), 1167 bp (*tb1*), 1416 bp (*orp*) and those of the introns were 716 bp (*Adh1*-i3, excluding mononucleotide repeats), 651 bp (1st-4th introns of genome sequences annotated AK058507), 1164 bp (1st-4th introns of MAP kinase 5), 1063 bp (1st and 2nd introns of fuructokinase1), 697 bp (2nd, 3rd and 4th intron of *orp*). AK058507, AK059353 and AK098919 were selected from the ESTs in chapter I, because the full length cDNAs of the three genes were available in both *O. punctata* and *O. officinalis*. AK058507, AK059353 and AK098919 were annotated LIM domain-containing protein, ozone-responsive stress-related protein and cyclophilin, respectively. LIM domain-containing protein is specific to pollen. Ozone-responsive stress-related protein is a stress response protein. Cyclophilin is involved in protein folding. RAFTIN 1 (AJ575667), MAP kinase 5 (AF479884) and fructokinase 1 (AF429948) were reported to be single genes in *O. sativa, japonica*. RAFTIN1 is important for pollen development. MAP kinase plays in a signal transduction. Fructokinase is an enzyme that catalyzes fructose into fructose-1-phosphate. teosinte branched 1 (*tb1*) and tryptophan synthase (*orp*) were reported in the phylogenetic analysis of Sorghum and Maize

(Swigonova *et al.* 2006).

The NJ phylograms of each CDS and genome sequence were shown in Figs. II-2 and 3, respectively. In Fig. II-2, relationship of *O. brachyantha* (FF genome) and other *Oryza* species was supported by >90 bootstrap values in the analysis of AK058507, AK059353, AK098919 and MAP kinase 5. Relationship of AA genome species (*O. sativa, japonica*, *O. rufipogon* and *O. barthii*) and other *Oryza* species was also supported by >90 bootstrap values in AK059353, RAFTIN 1, MAP kinase 5, *tb1* and *orp*. On the other hand, relationships of BB, CC, CCDD and EE genomes, which were contained in the *O. officinalis* complex, were confounded. Genome sequences including introns of AK058507, MAP kinase, fructokinase1 and *orp*, were used for resolving *O. officinalis* complex, because longer sequences were contained more information about phylogenetic relationships. The bootstrap values of genome sequenes showed higher bootstrap values than those of CDS as shown in Fig. II-3. The phylogenetic trees were different among the results obtained from different genes but the results could not resolve the entangled relationships among species of the *O. officinalis* complex. These results showed that the relationships among *O. brachyantha*, *O. officinalis* complex and *O. sativa* complex were clear, however the relationships in *O. officinalis* complex were complicated. These results also showed close relationships among species in *O. officinalis* complex. Then, reconstruction of phylogenetic tree using combined sequence of five CDSs (AK058507, AK059353, AK098919, MAP kinase and *tb1*) was performed. The results showed higher bootstrap values (Fig. II-4A). This tree showed clear relationships even in *O. officinalis* complex. The results indicated that construction of a phylogenetic tree for the complicated taxisonomic group requires longer sequence information.

Estimation of divergence time

The divergence time of *Oryza* was calculated based on the assumption that rice and maize diverged from a common ancestor at 50 mya (Gaut 2002) except for *Adh-i3*. In the case of *Adh1*-i3, the substitution rate, which was estimated to be 9.0 $\times 10^{-9}$ substitution per site per year (Zhu *et al.* 2005), was used. Table II-4 shows divergence time for eight CDSs and one intron between indicated genomes. Through divergence time was different from gene to gene, the estimated time and their context seems corresponding fairly well to other evolutional events. In this analysis, *O. brachyantha* (FF) vs. other *Oryza* species revealed to be branched at 11.7~21.6 mya. The divergence within the *O.* officinalis complex containing BB, CC, CCDD, and EE genomes was estimated later than 2.1~12.1 mya. AA genome was diverged from *O. officinalis* complex at 2.2~14.9 mya and speciation within the AA genome was estimated to occur at 0 ~5.6 mya. Fig. II-4 was a phylogenetic tree of six genomes in the genus *Oryza* that was reconstructed with combined sequence of five DNAs (AK058507, AK059353, AK098919, MAP kinase and *tb1*) excepting three genes with lacking sequences in any species (RAFTIN1, frutokinase 1 and *orp*). The phylogenetic tree indicates highly probable relationships and split time of seven *Oryza* species. The *O. officinalis* complex and *O. brachyantha* (FF) were split at around 16.3 mya, next two ancestor genomes of BB and CC complex, and DD and EE complex were split at around 12.8 mya and then three successive divergence were occurred at around 8.0 mya spliting *O. sativa* complexs and at around 4.0 mya spliting DD and EE genomes.

Synonymous and non-synonymous substitution rates

Synonymous and non-synonymous substitution rates (Ks and Ka, respectively) of eight genes used for phylogenetic analysis were shown between *O. sativa, japonica* and wild relatives of rice in Table II-5 and Table II-6. Table II-6 showed synonymous and non-synonymous substitution rates for domain region. Codon

based Z-test using MEGA3.1 showed that each gene was not operated by positive selection. The averages of Ka/Ks ratio were 0.053 in AK058507, 0.177 in AK059353, 0.084 in AK098919, 0.763 in RAFTIN1, 0.108 in MAP kinase 5, 0.062 in fructokinase 1, 0.475 in *tb1* and 0.181 in the *orp*. RAFTIN1 and *tb1* showed higher values than others. AK058507 and fructokinase 1, AK059353 and *orp*, and AK098919 and MAP kinase 5 had a similar rate, respectively. Comparison of synonymous and non-synonymous rates between entire genes and domain regions showed that functional domains of Zn-binding protein (8th to 64th codons of AK058507), peptidyl-prolyl cis-trans isomerase (24th to 39th codons of AK098919), protein kinase domain (1st to 218th codons of MAP kinase 5), carbohydrate kinase (257th to 270th codons of fructokinase 1) and beta subunit of pyridoxal-5'-phosphate-dependent enzyme (123rd to 457th codons of *orp*) were conserved than the whole gene region. Peptidyl-prolyl cis-trans isomerase (103rd to 146th codons of AK098919), BURP domain (188th to 406th codons of RAFTIN1) and carbohydrate kinase (38th to 62nd fructokinase 1) had lower non-synonymous rates than those of entire gene. These results totally showed that the eight genes used in the phylogenetic analysis for eight (nine including DD genome) species have non-viased structural characters suitable for phylogenetic tree construction.

Discussion

Relationship of species in O. officinalis complex

Nine genes in Table II-1 were used in this study and these genes were not positively evolved genes as revealed in the analysis of the codon based Z-test. The phylogenetic tree indicateed highly probable relationships of seven *Oryza* species. The *O. officinalis* complex and *O. brachyantha* (FF) were split first, next two ancestor genomes of BB and CC complex, and DD and EE complex were split and then three

successive divergence occurred in the process to *O. sativa* complex, to BB and CC genomes and to DD and EE genomes. However, in *O. officinalis* complex, the relationship was still confused after application of the results using nine genes. The phylogenetic analyses using nine genes showed five different pattern of the evolutional relationships of five genome types summarized in Fig. II-5. AA genome and *O. officinalis* complex were branched in case of AK058507, MAP kinase, RAFTIN and *orp*, and then diverged within the *O. officinalis* complex. While in case of AK098919, *tb1* and *Adh1*-i3, first CCDD and EE genomes were branched from the rest of genomes and then *O. officinalis* and *O. punctata* branched each other. In case of AK059353 and fructokinase1, CC genome was first branched, followed by branching of CCDD and EE genomes, and AA and BB genome in order. These results showed that genes were categorized into two groups. One is the genes first differentiated within officinalis complex (Fig. II-5c, d and e) and another is the genes differentiated between the AA genome and the *officinalis* complex (Fig. II-5a and b). This confusion showed that species in the *O. officinalis* complex had close relationships with each others. The phylogenetic analysis using indepenent genome sequences of four genes slightly improved bootstrap values (Fig.II-3), however the relationships among *O. officinalis* complex were still confused. On the contrary, the combined genome sequence of five CDSs showed a high bootstrap value and clear relationships among species of *O. officinalis* complex (Fig. II-4). These results showed that genome sequences of independent genes were not enough to resolve *O. officinalis* complex relationships. More gene sequences or longer combined sequences are needed to reconstract the phylogenetic relationship of the complicated evolutional group like *O. officinalis* complex.

Divergence time of Oryza

Some paper reported the estimation of divergence of species in the genus *Oryza*. Buso *et al.* (2001) estimated divergence time of CCDD and CC 20 mya. Meanwhile, Guo *et al.* (2005) showed that the remote relatives of *O. granulata* (GG) and remaining *Oryza* species was separated 10.2 mya based on the estimation of *matK*, or 8.8 mya based on *GPA1*. Zhu *et al.* (2005) estimated that AA genome split 2.0 mya and *O. sativa, indica* and *japonica* separated 0.4 mya. This study estimated the divergence time of eight *Oryza* species at each split point. The divergence time between BB, CC, CCDD and EE genome was about 12.8 mya (range 2.1~12.1 mya). *O. brachyantha* (FF) and remaining *Oryza* species was separated about 16.3 mya (range11.7~21.6 mya). AA genome and *O.* officinalis complex including BB, CC genomes were separated about 8.0 mya (range 2.2~14.9 mya), and speciation in AA genome was occurred at about 0.6 mya (range 0~5.6 mya). AA genome separation time was consistent with the report of Zhu *et al.* (2005). The divergence time of *O. officinalis* complex was shorter than the previous studies, while that of AA genome or FF genome split was longer than the results of previous studies. This study used nine genes of seven different species, those were much more genes than other studies, for estimating more accurate split time of the species in the genus *Oryza*. Investigation of a set of additional nuclear genes, which are under neutral selection and other species in the same genome type, is needed to confirm the relationship and divergence time of the genomes and species in the genus *Oryza*.

Comparison of synonymous and non-synonymous substitutions rates

Synonymous (Ks) and non-synonymous (Ka) substitution rates of eight genes were investigated (Table II-5). Ka/Ks >1 indicates genes under positive selection pressure. Ka/Ks of the eight genes were not over 1, however RAFTIN 1 and *tb* showed higher Ka/Ks than others (0.763 and 0.475, respectively). RAFTIN 1 is

an anther-specific single copy gene. RNA interference of RAFTIN resulted in elimination of pollen (Wang *et al.* 2003). RAFTIN is essential for the late phase of pollen development. *tb* gene belongs to the TCP gene family whose members encode putative basic helix-loop-helix DNA-binding proteins. *tb* regulates the number and length of axillary branches and contributed to the architectural differences between maize and teosinte (Hubbard *et al.* 2002). These genes play their roles in a reproductive phase, which is the most important phase for speciation. This may reflect the higher Ka/Ks ratio. The domain regions were known well conserved among species. Synonymous and non-synonymous substitutions of entire genes and domain regions in this study also showed that the domain regions were well conserved as expected. The peptidyl-prolyl cis-trans isomerase (24th to 39th codons of AK098919), and the carbohydrate kinase (38th to 62nd and 257th to 270th codons of fructokinase 1) had no non-synonymous substitution, and the beta subunit of pyridoxal-5'-phosphate-dependent enzyme (123rd to 457th codons of *orp*) showed about half non-synonymous substitution rate than its entire gene sequences. These domains would be needed to protein function, because they were conserved than other regions.

This study showed the relationship and divergence time of *Oryza*, and synonymous (Ks) and non-synonymous (Ka) substitution rates of eight genes. Through, the relationship of species in the *O. officinalis* complex had been confused yet, this study could reconstruct a phylogenetic tree of the genus *Oryza* that correlated with the evolution time of other plant genus (Gaut 2002). Additional phylogenetic analysis of many genes or other molecular analyses using many accessions of *O. officinalis* complex will give more useful information to clarify the relationship.

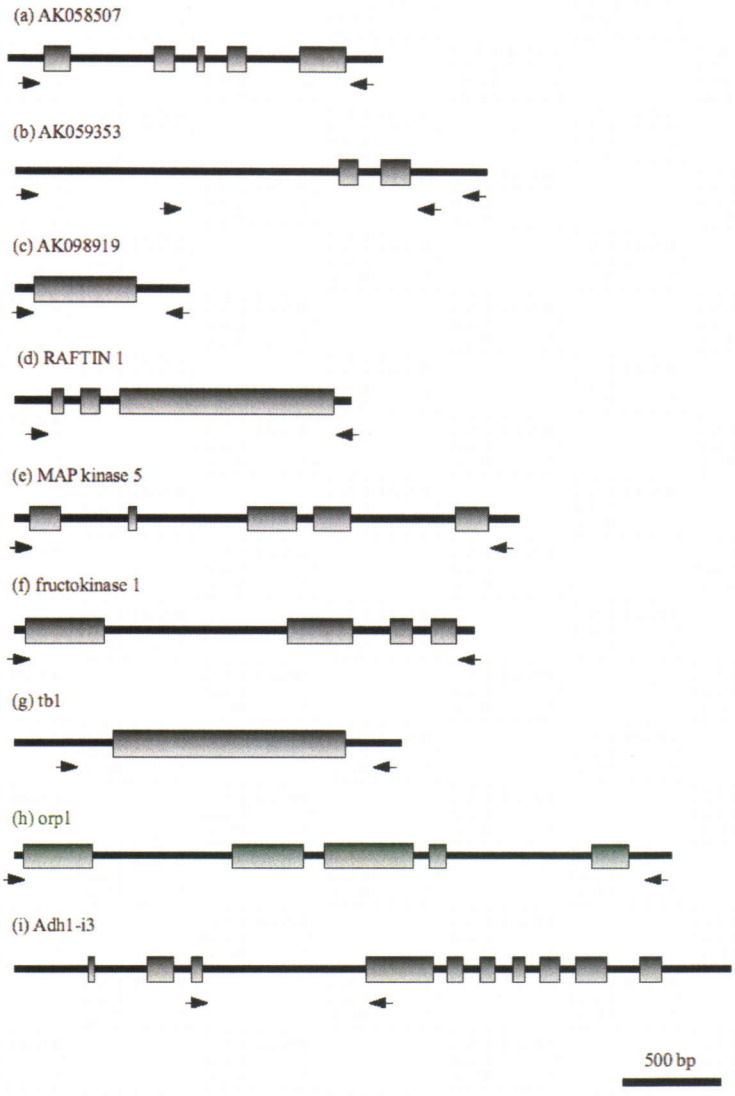


Fig. II-1. Genome structures of nine genes used in the analysis.

(a): AK058507, (b) AK059353, (c) AK098919, (d) RAFTIN 1, (e) MAP kinase 5, (f) Fructokinase 1, (g) Tb1,

(h) Orp1 and (i) Adh1 3rd intron. Boxes: exons. Arrows indicate position and directions of primers.

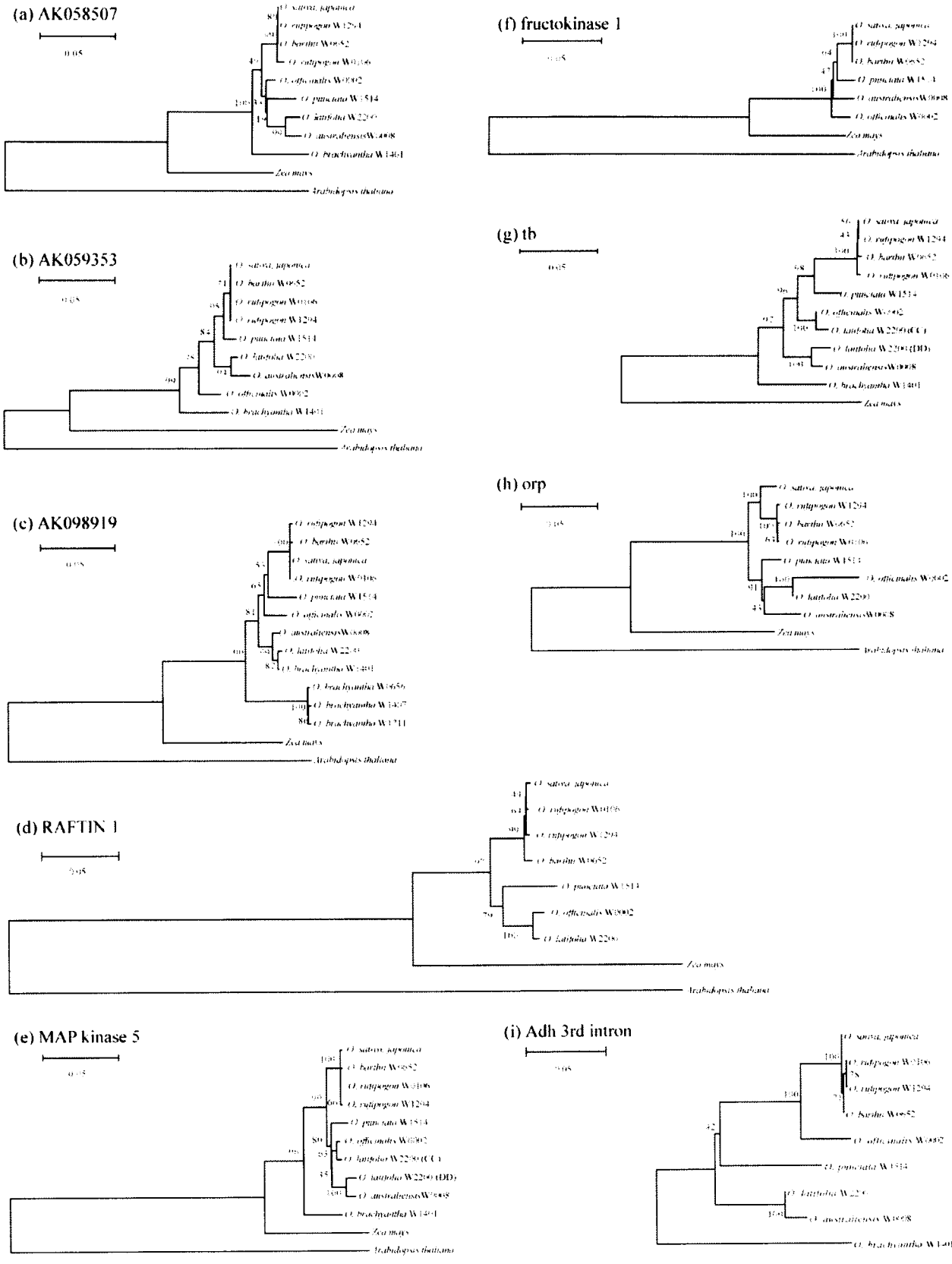
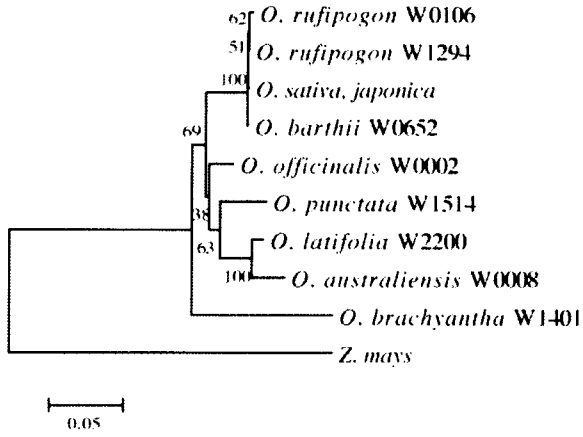
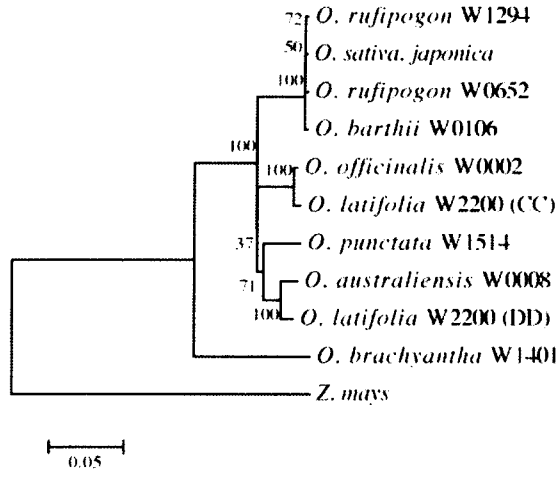


Fig. II-2. NJ phylogram based on the Kimura two-parameter distance model using CDSs. Numbers on branches are bootstrap values from 1000 replicates. Branch lengths are proportional to distance.

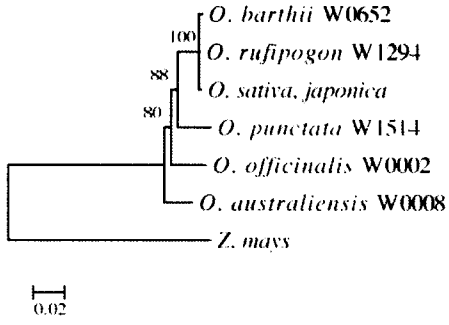
(a) AK058507



(b) MAP kinase 5



(c) fructokinase 1



(d) *orp 1*

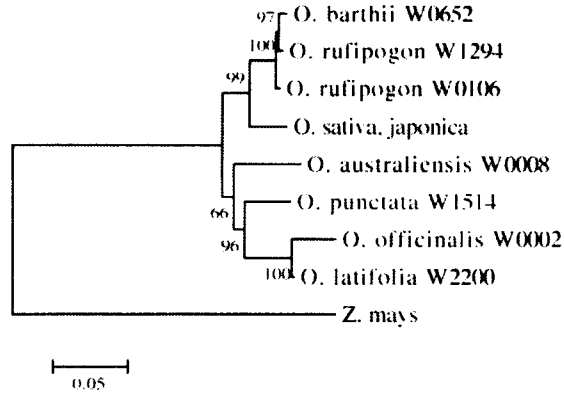


Fig. II-3. NJ phylogram based on the Kimura two-parameter distance model using genome sequences.

Numbers on branches are bootstrap values from 1000 replicates. Branch lengths are proportional to distance.

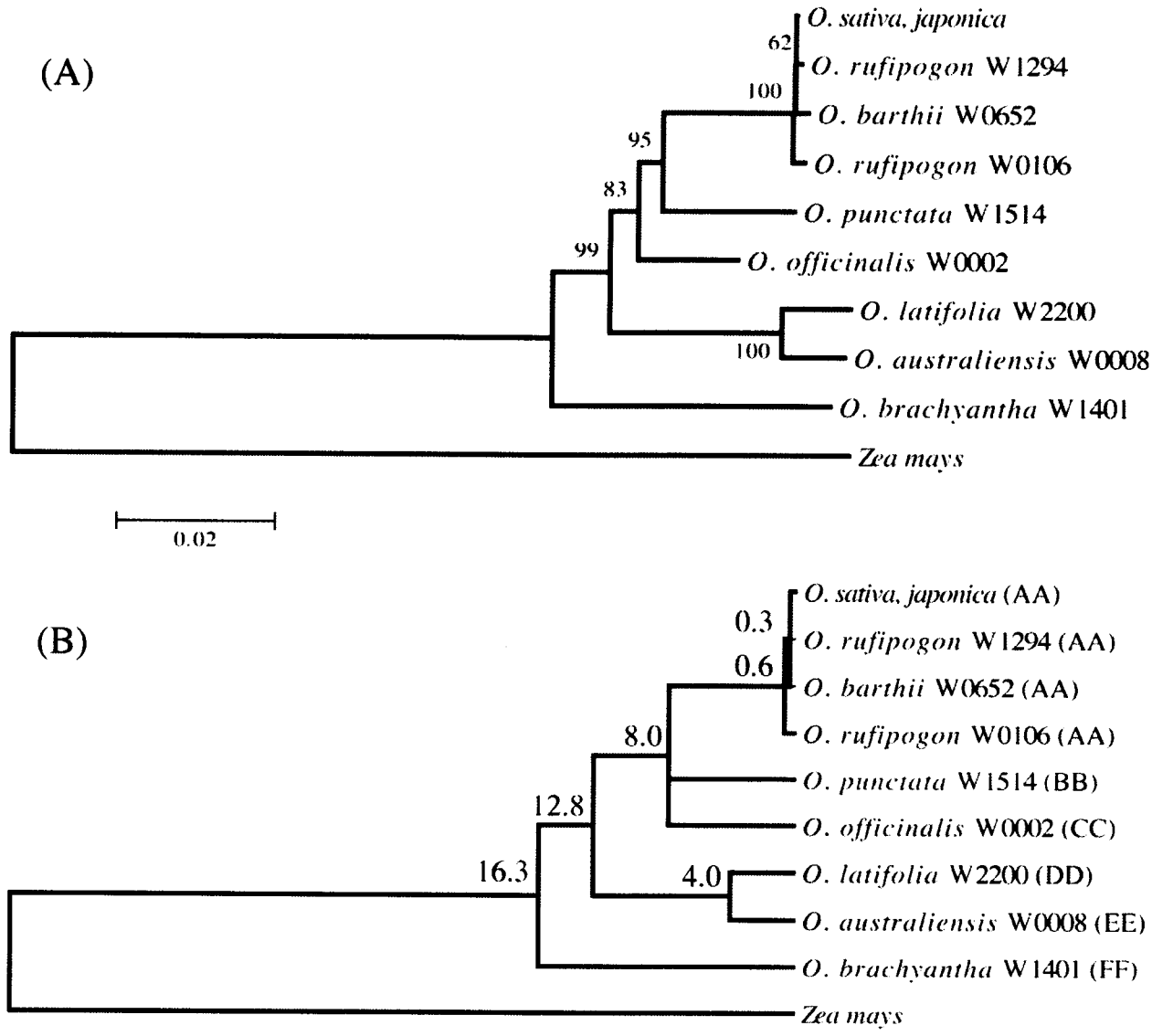


Fig. II-4. A phylogenetic tree of six genomes in the genus *Oryza*.

(A) NJ phylogram based on the Kimura two-parameter distance model using combined sequences of five CDSs. Numbers on branches are bootstrap values from 1000 replicates. Branch lengths are proportional to distance. (B) The divergent time of species in the genus *Oryza*. The number shows the divergence time (mya).

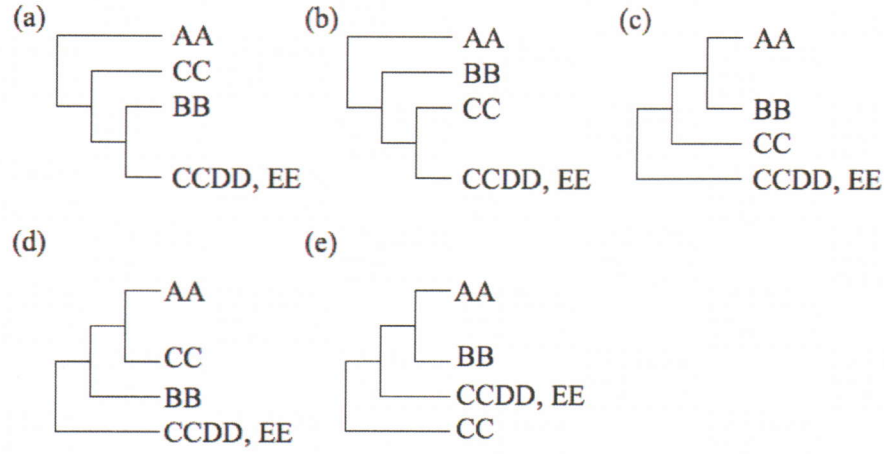


Fig. II-5. Evolutional relationship among gnomes of *O. officinalis* complex (BB, CC, CCDD, EE) and *O. sativa* complex (AA).

(a) inferred from AK058507. (b) inferred from MAP kinase 5, RAFTIN 1 and orp. (c) inferred from AK098919 and tb. (d) inferred from Adh1-i3. (e) inferred from AK059353 and fructokinase 1.

Table II-1. Summary of analyzed genes

Gene	Accession	RAP-DB	CDS Length	Inton length	Reference
LIM domain-containing protein	AK058507	Os03g0266100	591	651	Kikuchi *et al.* 2003
oaone-responsive stress related proetin	AK059353	Os06g1145000	240	-	Kikuchi *et al.* 2003
cyclophilin	AK098919	Os02g0121300	519	-	Kikuchi *et al.* 2003
RAFTIN 1	AJ575667	Os08g0496800	1239	-	Wang *et al.* 2003
MAP kinase 5	AF479884	Os03g0285800	798	1164	Xiong *et al.* 2003
fructokinase 1	AF429948	Os01g0894300	972	1063	Jiang *et al.* 2003
teosinte branched 1	AK072535	Os03g0706500	1167	-	Swigonová *et al.* 2004
tryptophane synthase	AK107083	Os08g0135900	1416	697	Swigonová *et al.* 2004
Adh1 3rd intron	AC123521	–	-	716	Zhu *et al.* 2005

RAP-DB referenced RAP1 annotation (http://rapdb.lab.nig.ac.jp/index.html)

Intron length showed only used intron in this analysis.

Table II-2. Plant materials

Species	Genome type	Accession number
O. sativa, japonica	AA	-
O. rufipogon	AA	W0106
O. rufipogon	AA	W1294
O. barthii	AA	W0652
O. punctata	BB	W1514
O. officinalis	CC	W0002
O. latifolia	CCDD	W2200
O. australiensis	EE	W0008
O. brachyantha	FF	W1401
O. brachyantha	FF	W0656
O. brachyantha	FF	W1407
O. brachyantha	FF	W1711

Table II-3. The DNA fragments and primers used in this study

Gene		Sequence	Reference
AK058507	Fw	AGGAGATCAGATCGGGCT	This study
	Rv	CTGGAACATGGCATACAAGC	This study
AK059353	Fw	CCAAGAACACCAAGAACTTG	This study
	Rv	AGCACAAAGTAGAGGTTG	This study
	Fw2	GGATTGTGTGAGCTTCTCTT	This study
	Rv2	GCTTTAGCATCAGCACCTCA	This study
AK098919	Fw	TTCCCCATCTGTGAAATCGC	This study
	Rv	ACCAACACGACACGAACTCA	This study
RAFTIN 1	Fw	CTGTTTCCTGTCACCGTCG	This study
	Rv	CAGCAGCTGACACGAATAACAT	This study
MAP kinase 5	Fw	TTAGGTTGGTCAATTCGGCT	This study
	Rv	TGGGTTTCGAACATCAACAA	This study
fructokinase I	Fw	CTTTCATCGCCTCTTGTGTTC	This study
	Rv	AAGGCAGGCATCTTGGAGTA	This study
teosinte branched 1	Fw	CTTCCTTTCTTCGATTCCCC	This study
	Rv	GCGATGACCAAACCAAAGTT	This study
tryptophane synthase	Fw	AATTACTCGCTTCGTGGTCG	This study
	Rv	TCCGCCAATAACATCATTCA	This study
Adh1 3rd intron	F5	TCCCGTGTTCCCTCGGATCTTC	Zhu *et al*. 2005
	R3	GTCACACCCTCTCCAACACTCT	Zhu *et al*. 2005

Fw : forward, Rv : reverse

Table II-4. Divergence time of six different genomes calculated with nine gene sequences

	Combination of divergence	Time range	AK058507	AK059353	AK098919	RAFTIN 1	MAP kinase 5	fructokinase 1	*tb*	*orp*	*Adh1*-i3	Average time
AA	AA (*O. rufipogon*)	0-1.7	1.7	0.0	0.4	1.7-0.8	0.9-0.5	0.0	0.6	5.6-0.5	0.1	1.2
	O. officinalis complex	2.2-14.9	10.9	2.2	10.0	12.1	10.0	8.3	7.7	14.9	3.2	8.8
FF	AA and *O. officinalis* complex	11.9-21.6	19.6	11.9	21.0-10.78	-	21.6	-	16.3	-	12.8	17.2
O. officinalis complex												
BB	CC	6.8-12.1	9.4	6.8	10.0	11.9	9.2	10.0	8.9	12.1	7.1	8.5
	(CC)DD	5.4-12.1	9.4	5.4	10.8	11.9	9.2	-	11.5	12.1	7.1	9.7
	EE	5.4-12.1	9.4	5.4	10.8	-	9.2	10.0	11.5	12.1	7.1	9.4
CC	(CC)DD	2.1-11.5	9.4	6.8	10.8	2.1	8.9	-	11.5	10.9	7.1	8.5
	EE	6.8-12.1	9.4	6.8	10.8	-	8.9	10.0	11.5	12.1	7.1	9.6
(CC)DD	EE	0.7-12.1	6.8	3.0	2.9	-	5.2	-	3.3	12.1	0.7	4.9

Table II-5. Rates of Synonymous and Non-synonymous substitution in eight CDSs

		AK058507		A059353		AK098919		RAFTIN 1		MAP kinase 5		fructokinase I		*tb*		*orp*	
Accession	Genome type	Ka	Ks	Ka	Ks	Ka	Ks	Ka	Ks	Ka	Ks	Ka	Ks	Ka	Ks	Ka	Ks
W0106	AA	0.003	0.010	0.000	0.000	0.000	0.000	0.005	0.000	0.002	0.000			0.000	0.013	0.005	0.057
W1294	AA	0.000	0.000	0.000	0.000	0.000	0.006	0.004	0.005	0.002	0.004	0.000	0.000	0.000	0.005	0.005	0.064
W0652	AA	0.000	0.000	0.000	0.000	0.000	0.006	0.007	0.005	0.002	0.004	0.000	0.000	0.001	0.011	0.007	0.055
W1514	BB	0.009	0.132	0.000	0.046	0.006	0.156	0.058	0.079	0.010	0.069	0.003	0.079	0.037	0.078	0.012	0.105
W0002	CC	0.003	0.079	0.016	0.079	0.013	0.105	0.055	0.069	0.006	0.058	0.008	0.078	0.036	0.106	0.061	0.180
W2200	CCDD									0.006	0.070			0.042	0.127	0.016	0.136
		0.007	0.122	0.011	0.063	0.011	0.159	0.052	0.069	0.008	0.106			0.066	0.155		
W0008	EE	0.010	0.115	0.011	0.098	0.008	0.160			0.012	0.092	0.004	0.101	0.066	0.132	0.025	0.164
W1401	FF	0.016	0.219	0.032	0.156	0.008	0.159			0.017	0.175			0.080	0.269		

Ka and Ks were calculated between wild relatives of rice and *O. sativa, japonica.*

Table II-6. Rates of synonymous and non-synonymous substitutions in domain regions

		AK058507 Zn-binding protein, LIM				AK098919 Peptidyl-prolyl cis-trans isomerase, cyclophilin type					
		8-64 codons		107-163 codons		24-39 codons		60-72 codons		103-146 codons	
Accession	Genome type	Ka	Ks	Ka	Ks	Ka	Ks	Ka	Ks	Ka	Ks
W0106	AA	0.009	0.000	0.000	0.018	0.000	0.000	0.000	0.000	0.000	0.000
W1294	AA	0.000	0.000	0.000	0.000	0.000	0.072	0.000	0.000	0.000	0.000
W0652	AA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
W1514	BB	0.000	0.075	0.009	0.232	0.000	0.218	0.000	0.000	0.000	0.115
W0002	CC	0.000	0.035	0.009	0.107	0.000	0.065	0.031	0.000	0.000	0.115
W2200	CCDD										
		0.000	0.035	0.014	0.172	0.000	0.097	0.031	0.000	0.011	0.193
W0008	EE	0.016	0.038	0.009	0.188	0.000	0.097	0.031	0.235	0.000	0.197
W1401	FF	0.000	0.178	0.016	0.232	0.000	0.097	0.031	0.000	0.000	0.191

		RAFTIN 1		MAP kinase 5		fructokinase I				*orp*	
		BURP domain		protein kinase domain		Carbohydrate kinase, PfkB				Pyridoxal-5'-phosphate-dependent enzyme, beta subunit	
	region	188-406 codons		1-218 codons		38-62 codons		257-270 codons		123-457 codons	
Accession	Genome type	Ka	Ks	Ka	Ks	Ka	Ks	Ka	Ks	Ka	Ks
W0106	AA	0.005	0.000	0.000	0.000					0.001	0.059
W1294	AA	0.005	0.006	0.000	0.005	0.000	0.000	0.000	0.000	0.001	0.064
W0652	AA	0.009	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.056
W1514	BB	0.049	0.095	0.006	0.065	0.000	0.262	0.000	0.091	0.003	0.067
W0002	CC	0.045	0.084	0.004	0.058	0.000	0.330	0.000	0.091	0.037	0.137
W2200	CCDD			0.004	0.061					0.004	0.106
		0.040	0.084	0.006	0.103						
W0008	EE			0.009	0.087	0.000	0.403	0.000	0.000	0.004	0.107
W1401	FF			0.019	0.166						

Chapter III

Unique genes in BB and CC genome wild species of rice

Introduction

The cultivated rice *Oryza sativa* (AA genome) has been genetically analyzed well, and its genome sequencing was almost completed (IRGSP November 18, 2004, Yu *et al.* 2005). A number of full length cDNAs and tag lines are also available (Kikuchi *et al.* 2003, Miyao *et al.* 2003, Jeong *et al.* 2003). Several genes related to agricultural traits, such as dwarf, seed shattering and waxy, have been identified and characterized (Zhang *et al.* 2006, Konishi *et al.* 2006, Ashikari *et al.* 2005). These materials will help researches on the genus *Oryza* for comparative genomics studies as well as studies of domestication and ecological adaptation.

Wild relatives of rice are another important genetic resources and have been shown to possess useful traits including resistance to various biotic and abiotic stresses. For example, *O. officinalis* (CC genome) has been reported to show resistance to vermin such as yellow stem borer, planthopper and leafhopper (Vaughan 1994, Brar *et al.* 1997, Xiao *et al.* 1996, Xiao *et al.* 1998). In spite of the importance of the wild relatives of rice as genetic resources, their studies have only recently started on a molecular and a genetic base. The Oryza Map Alignment Project (OMAP) is now conducting an alignment of wild relative genomes with the *O. sativa* genome sequence to create physical maps of twelve wild genomes of rice. BAC libraries of the wild relatives have been constructed and their end sequences were reported (Wing *et al.* 2005, Ammiraju *et al.* 2006). In contrast to the progress of our understanding on the genome structure of wild relatives through comparative genomics studies, identification and characterization of wild relative genes remain to be studied. Only gene expression and EST analyses of *O. minuta* (BBCC genome) was reported (Cho *et al.* 2004, Cho *et al.* 2004 (a), Shim *et al.* 2004, Cho *et al.* 2005).

In the study of chapter I, I report comparative cDNA analysis using 973 clones of *O. punctata* shoot ESTs

and 1,425 clones of *O. officinalis* panicle ESTs. My study showed that 88-96% of cDNA clones of BB and CC genomes were highly homologus (90-100% identity) to *O. sativa* sequences at a nucleotide level, and that about 8% of *O. punctata* ESTs and 4% of *O. officinalis* ESTs did not match with any of the rice ESTs, full length cDNA and genome sequence.

In this chapter, I identified and analyzed genes specific to the BB genome or the CC genome. Twenty-six clones from the BB genome and sixty-seven clones from the CC genome did not find their homologues in GenBank nonredundant and unigene databases. These ESTs were candidate transcripts unique to wild relatives. Among them 20 genes were shown to be unique to wild relatives by Southern blot analysis. Four full length cDNAs were obtained which encode a ribosomal protein, a cyclin D, a cold response protein and an unknown protein. Cyclin D and the unknown protein would be absent in *O. sativa*. The ribosomal protein had an identical protein in *O. sativa*, but 5' and 3' UTRs of their cDNAs were greatly different from each other. The cold response protein had a homologue in *O. sativa*, but 5' and 3' UTRs of its cDNA was also greatly different from each other. *O. punctata* and *O. officinalis* may contain unique features such as stress resistance, disease resistance and morphology that do not exist in the cultivated rice. Some of these transcripts might be related to these unique features. My results provide new information on what genes have been deleted from the wild relatives and what unique genes were evolved in the wild species of rice.

Materials and Methods

Plant materials

Plant materials used in this study are shown in Table III-1. These accessions have been maintained in National Institute of Genetics as self-pollinated seeds. Cold induction was treated by transferring plants to 4℃

from 28℃ 7 days after germination. Plants were harvested prior to and after 2, 4, 6, 8, and 24 hr induction.

Database search and Phylogenetic analysis

O. punctata and *O. officinalis* EST sequences (accession numbers; BB984512-BB985121, BB994797-BB995924, BB995926-BB996458, AK223775-AK223829, AK223936-AK224590) analysed in the chapter I were used. Homology search of all these sequences was carried out against Rice genome database, Rice EST/cDNA database and GenBank non-redundant databases using BLAST and BLASTX, with these EST sequences as queries. The sequences were aligned with clastalW (Jeanmougin *et al.*1998, Thompson *et al.* 1994) and refined manually. Sequence divergence and phylogenetic analysis of the cyclin clones were performed using the Kimura's two-parameter model (Kimura 1980) and the Neighbor-joining method with MEGA 3.1(Kumar *et al.* 2004), respectively.

5'RACE and RT-PCR

Shoot apex of *O. punctata* (W1514) and young panicles (ca 2-4 cm in length) *of O. officinalis* (W0002) were used. Total RNAs were isolated from shoot apex of *O. punctata* (W1514) and young panicles (ca 2-4 cm in length) *of O. officinalis* (W0002) using CONCERT™ Plant RNA Purification Reagent (Invitrogen) according to the manufacture's instruction. Poly(A)$^+$ RNA was extracted from the total RNA using Dynabeads mRNA Direct™ Kit (DYNAL). 5'RACE (rapid amplification of cDNA ends) was performed using Marathon™ cDNA Amplification Kit (BD Biosciences). Sequencing was performed using BigDye terminator Kit (Applied Biosystems) with an automated DNA capillary sequencer ABI 3130xl.

For RT-PCR, poly(A)$^+$ RNA was isolated from panicles (ca 2 cm and 10 cm in length), shoot apex, leaf and

ovary as described above. Poly(A)$^+$ RNA was reverse-transcribed using SuperscriptIII RNaseH– reverse transcriptase (Invitrogen) with an oligo dT primer (5'-GAGCTCTTTTTTTTTTTT-3'). PCR was carried out using primers shown in Table III-2.

Southern blot and BAC library screening

Genomic DNAs of AA, BB, CC, CCDD, EE and FF genome species were digested with *Hind*III, electrophoresed in an agarose gel, blotted onto nylon membrane and hybridized with ESTs using an ECL Direct Nucleic Acid Labelling and Detection System (GE Healthcare) following the manifacture protocol.

BAC clone library filters were obtained from the Arizona Genomics Institute. The BAC filters were hybridized with full length cDNA clones (KOME: Knowledge-based Oryza Molecular biological Encyclopedia) or genome DNA fragments (amplified by PCR) of all the members of *O. sativa* cyclin D2/4 and D3 genesas described above.

Complementation test in yeast

A full-length cDNA of BBS13D01 was obtained from *O. punctata* shoot apex cDNA using primers BBS13D01- *BamH*I (5'- CGGGATCCATGGTTCTGAGCTATGACTG-3') and BBS13D01- *Xho*I (5'-CCGCTCGAGTCAGCTGCAAAGCTTCCTTC-3'). The amplified DNA was digested with *BamH* I and *Xho*I and cloned into the *BamH*I/*Xho*I site of pYES2 (Invitrogen). The pYES2 vector has a URA3 marker for selection in yeast. Transformation of yeast K3413 was performed using Frozen-EZ Yeast Transformation II^{TM} (Zymo Research). K3413 (cln1Δ::HisG cln2Δ cln3Δ::LEU2 MET-CLN2) was maintained on an SD plate (2% glucose, 0.67% yeast nitrogen base, amino acids except methionine, and 1.5% Bactoagar). Transformed yeast

cells were cultured on the SD plate supplemented with 20 mM methionine or with 20 mM methionine and 2% galactose.

Results

Identification of genes unique to wild relatives

Twenty-six clones from the BB genomes and 67 clones from the CC genomes were predicted to be unique genes to wild relatves as described in chapter I. To examine whether the candidate genes of 26 clones from the BB genome and 67 clones from the CC genome were unique to wild relatives or not, Southern hybridization was carried out using all 93 EST clones as probes. The results showed that bands were detected only in either one or a few genomes of BB to FF genomes, and no positive band was detected in *O. sativa, japonica* DNA for twenty EST probes out of 93 clones (Fig. III-1, Table III-3). This indicated that the ESTs were specific to the wild relatives.

To obtain the full length cDNAs of these genes, 5'RACE was carried out. The full length cDNAs of four clones were obtained, and they covered complete ORFs of 336 bp (BBS02H03), 1050 bp (BBS13D01), 543 bp (CCP17F01) and 387 bp (CCP19B09) in length.

A database search showed that BBS13D01 was most similar to *Zea mays* D2 type cyclin in any of the deposited sequences (81% identity at a nucleotide level and 75% at an amino acid level). A rice full-length cDNA AK103765 was most similar to BBS13D01 among *O. sativa, japonica* sequences, but its 5' sequence was shorter than that of BBS13D01 (Fig. III-2A). BBS13D01 protein contained a cyclin box, which is conserved among all known cyclins, and its possible function is binding cyclin dependent kinases (CDKs). It was also found that the BBS13D01 contained an LxCxE motif at its N-terminus, which is a binding site for

plant Rb and Rb related proteins (Oakenfull *et al.* 2002). The presence of all the domains conserved among cyclinDs suggests that BBS13D01 encodes cyclin D.

The genome sequence of BBS13D01 was analysed with from an *O. punctata* BAC clone from Arizona Genomics Institute (Fig. III-2C). The genome sequence of its orthologue in *O. officinalis* (CC genome) was obtained in the same way (Fig. III-2C). Both of them consisted of six exons and five introns and highly similar sequence with each other. About 4000 bp genome sequence covering the entire ORFs was used as the query for database search with the BLAST program. No highly homologous sequence to the 4000 bp long region was identified, consistent with the result of Southern blot analysis. This results confirmed that *O. sativa* lacks an orthologue of the cyclin, BBS13D01.

To further examine the phylogenetic relationship of the BBS13D01 to other cyclin D family genes, I identified all the cyclin D genes of *O. sativa* and *O. punctata.* By an extensive database search seven *O. sativa* cyclin D genes were identified. *O. punctata* cyclin D genes were identified by screening of BAC library using the orthologous *O. sativa* gene clones as probes followed by amplification and sequencing of the selected BAC clones. As a result, eight cyclin D genes of *O. punctata* were identified. The phylogenetic analysis of the identified cyclin Ds based on their amino acid sequences using the N-J method showed that each of *O. punctata* cyclin Ds has an orthologue in *O. sativa* except BBS13D01, which is present only in *O. punctata* (Fig. III-5). This confirmed that BBS13D01 was a wild relative specific gene.

CCP17F01 showed high homology to *Triticum aestivum* WCSP1 (68% identity at nucleotide level and 62% at amino acid level). A rice full-length cDNA AK101577 was most similar to CCP17F01 in *O. sativa, japonica*, but 3'-UTR of CCP17F01 had low identity to AK101577 (53% identity at nucleotide level). *T. aestivum* WCSP1 contained, from N-terminus to C-terminus, a cold shock domain, a large glycine rich region,

three zinc finger motives and two glycine rich domains (Karlson *et al.* 2002), whereas *O. officinalis* CCP17F01 and *O. sativa, japonica* AK101577 had only two zinc finger motives and a glycine rich domain in addition to a cold shock domain and a large glycine rich region (Fig. III-2B). The presence of all the domains conserved among cold shock proteins suggests that CCP17F01 encodes a cold shock protein.

BBS02H03 was shown to be homologous to a rice full length cDNA AK102775 (99% identity in their coding region), which encoded ribosomal RNA, but 3'-UTR of BBS02H03 had low identity to AK102775 (48% identity at a nucleotide level). These results showed that BBS02H03 encodes a ribosomal protein and their specificity presented in their UTRs.

No highly homologous sequence to CCP19B09 was identified by a database search. There was no known domain and specific feature of sequence in CCP19B09. CCP19B09 was unknown protein unique to *O. officinalis*.

5'RACE of remaining five clones in *O. punctata* and eleven clones in *O. officinalis* was not successful, although short sequences almost same length as cDNA clones were amplified frequently. All these clones could not identify homologues in the databases (Table III-3).

Expression analysis

The expression profile of the three unique genes to wild relatives was investigated by RT-PCR using RNAs isolated from panicles (ca 2 cm or 10 cm in length), shoot apex, leaf and ovary. Gene-specific primers used for the RT-PCR were shown in Table III-2. Actin was used as an internal control. The expression of BBS13D01 was observed in all organs examined, although lower expression in the leaf than in the others was detected (Fig. III-3A).

The expression of CCP17F01 was observed in all organs examined at a low level (Fig. III-3B). Because expression of WCSP1 was reported to be induced by low temperature (Karlson *et al.* 2002), I examined low temperature induced expression of CCP17F01. CCP17F01 expression was induced within 2~4 hours following low temperature treatment, and increased throughout the treatment of seedling (Fig. III-3C). This indicates that the CCP17F01 is a cold-inducible gene.

The expression of CCP19B09 was also observed in all organs examined with higher expression in panicles (Fig. III-3D).

Complementation of a yeast mutant for Cyclin D

To examine whether BBS13D01 encodes functional cyclin, yeast complementation test was carried out. The yeast strain K3413 shows a methionine dependent growth arrest phenotype due to deletion of CLN1 and CLN3 and replacement of wild type CLN2 with a CLN2 chimera gene in which a CLN2 coding region is under control of a methionine repressible promoter. This strain grew normally on an SD medium without methionine, but could not on a medium with 20 mM methionine. This strain was transformed with a plasmid containing BBS13D01 in an expression vector pYES2 (Invitrogen) under a GAL1 promoter, so that the expression of BBS13D01 can be induced by galactose. As shown in Fig. III-4, yeast cells transformed with BBS13D01 or CLN2 could grow on the medium containing galactose and methionine, but could not on the medium lacking galactose and containing methionine. Yeast cells transformed with an empty vector (pYES2) grew on neither plates. Because BBS13D01 complemented the yeast *cln* mutant, BBS13D01 revealed to encode a functional cyclin.

Discussion

In this study, I identified and analyzed genes unique to wild relatives of rice. Full-length cDNAs of four genes were obtained by 5'RACE. Southern hybridization and extensive database searches showed that they were wild relative-unique genes. They encode ribosomal protein, cyclin D, cold response protein and unknown protein. My results suggest that wild relatives have various types of unique genes such as a cell cycle regulator, component of a translational machinery and a stress response gene.

Cyclins are cell cycle stage specific activators of CDKs. Cyclin D and its associated CDKs respond positively to mitogenic signals and target their kinase activity to Rb related proteins which results in the activation of E2F transcription factors. CycD promotes the progression of G1-S phase through this pathway (Oakenfull *et al.* 2002). BBS13D01 shares high homology with *Z. mays* CycD2;1. Phylogenetic analysis of the cyclin D in plants indicated that BBS13D01 was closely related to cyclin D2/4. The BBS13D01 contained the conserved cyclin N-terminus and C-terminus motives that are binding sites of CDKs. Additionally, an LXCXE motif, which was a binding site to Rb related proteins (Oakenfull *et al.* 2002), was also identified. These results suggest that BBS13D01 encodes CycD2.

Some analyses of plant cyclin Ds have been reported. In tobacco plants overexpressing *Arabidopsis thaliana* CycD2;1, the length of a G1 phase is reduced and cell cycling was enhanced (Cockcroft *et al.* 2000). The transgenic plants had an enhanced rate of leaf initiation, accelerated development and increased root growth, but without alteration of morphology. In contrast, the overexpression of *A. thaliana* CycD3:1 resulted in curled leaves by the increased number of smaller cells and a partial loss of cellular organization in *Arabidopsis*, and led to inhibition of cell cycle exit and differentiation of leaf tissues (Meijer *et al.* 2001, Deqitte *et al.* 2003). *T. aestivum* (wheat) CycD2;1, which was placed in the same clade of the phylogenetic tree with BBS13D01, was

primarily expressed in the proliferating tissues. Ectopic expression of *T. aestivum* CycD2;1 in *Arabidopsis* affected plant morphology and retarded plant growth (Wang *et al.* 2006). These results showed that D type cyclin is related to cell cycle and consequently to morphology. BBS13D01 was more similar to *T. aestivum* CycD2;1 than *A. thaliana* CycD2;1. Thus, BBS13D01 might affect plant morphology and retarded plant growth like *T. aestivum* CycD2;1.

The cDNA clone complemented the *cln-* yeast strain. In RT-PCR, BBS13D01 expression in leaf was weaker than other organs. Leaf was the only well-differentiated organ used for the RT-PCR analysis. In another word, this cyclin D shows high expression in young organs or undifferentiated tissues. This supports a previous report in which *T. aestivum* CycD2;1 was only detecterble in young root and shoot by Northern hybridization (Wang *et al.* 2006). These results indicate that BBS13D01 is a cyclin D.

Gramene CMap showed that the OMAP contig containing BBS13D01 is syntenic to a 29.4 Mbp ~ 35.8 Mbp region of chromosome 2 of *O. sativa* (http://www.gramene.org/cmap/). Detailed comparison between *O. sativa* genome sequences of this region and the BAC end sequences in this OMAP contig showed that 87 kbp, from 31.9 Mbp to 32 Mbp, on chromosome 2 of *O. sativa* was altered and the BBS13D01 genome sequences was absent in the *O. sativa, japonica* genome sequence. Southern hybridization suggested that BB, CC, CCDD, EE and FF genome species have the homologues of BBS13D01. I actually obtained the genome sequence orthologous to BBS13D01 from a BAC library of *O. officinalis* (CC). From *O. sativa* sequence, seven other sequences of CycD2/4 and D3 family genes have been identified (Karlson *et al.* 2002). I obtained seven orthologues of *O. sativa, japonica* cycD2/4 and D3 from *O. punctata* (La *et al.* 2006). The phylogenetic analysis showed that BBS13D01 has no orthologue in *O.* sativa and is a unique additional cyclin D (Fig. 5). Southern hybridization and this result (Fig. III-5) suggested that the BBS13D01 orthologue in *O. sativa,*

japonica was deleted after divergence of AA genome from the others.

CCP17F01 has domains characteristic to cold shock proteins. CCP17F01 has two RNA binding motives, a large glycine-rich region, two zinc fingers and a small glycine-rich region. Homology searches indicated that CCP17F01 has highest homology to a cold response protein WCSP1 (Karlson *et al.*2002), which is similar to GRP-2 (de Oliveira *et al.* 1990) and cspA (Goldstein *et al.* 1990, Karlson *et al.* 2002). WCSP1 was known to be induced its expression by cold treatment. WCSP1 has a single strand DNA binding (ssDNA) activity in cold shock domain (CSD). WCSP1 has also double strand DNA (dsDNA) binding activity in the zinc fingers and the glycine-rich region (Karlson *et al.* 2002), and only a single set of a zinc finger and a small glycine rich region was shown to be sufficient for the dsDNA binding activity (Nakaminami *et al.* 2006). Because CCP17F01 has two zinc fingers and a small glycine rich region in addition to CSD, CCP17F01 might have activity to bind both ssDNA and dsDNA. RT-PCR analysis showed that the expression of CCP17F01 was induced by cold treatment. This expression pattern was similar to that of WCSP1 (Karlson *et al.*2002). These structural features and the expression pattern suggest that CCP17F01 is involved in a cold response through cold-induced gene expression and repression.

CCP17F01 had a homologue in *O. sativa, japonica* with 80% similarity in the CDS, but in 5' and 3'-UTRs with low similarities of only 61 % and 53%, respectively. Two clones, CCP17F01 and BBS02H03, had close homologues in *O. sativa*. Nucleotide sequences of CCP17F01 and its homologue were 80% identical to each other in their coding regions, but in their 5' and 3' UTRs the identity was dropped to only 61% and 53%, respectively. Similarly, the coding region of BBS02H03 was identical to its homologue of *O.* sativa, but its 3'UTR was only 48% identical to that of its homologue. Because promoters and terminators are important for mRNA transcription and 5' and 3'UTRs are often involved in the control of protein translation and stability, it

is speculated that the control of protein expression of these two wild relative unique genes were different from those of *O.* sativa homologues and unique to the wild relatives. Differential control of expression of functionally orthologous protein between cultivated rice and its wild relatives may be one of the typical differences between them.

CCP19B09 clone might be unique only to CC genome. RT-PCR shows that CCP19B09 expression almost same level in all organs examined. Many candidate genes other than above four clones were failed to clone by 5' RACE, although several short RNA sequences almost identical to those gene sequences were often amplified by the reaction. These sequences seemed not gene transcripts but might possibly non-coding short RNAs. Further studies will be needed to identify these sequences.

In this study, I reported four full length cDNAs (ribosomal protein, cyclin D, cold response protein and unknown protein). They were divided into two categories. One was unique sequence present only in wild relatives of rice, and the other has homologous gene in *O. sativa, japonica*, but its expression might be controlled differently. Unique sequences in wild relatives of rice like as the genes identified in this study might be related to unique features of wild relatives of rice such as morphology and stress response. Reverse genetic studies will clarify functions of these genes.

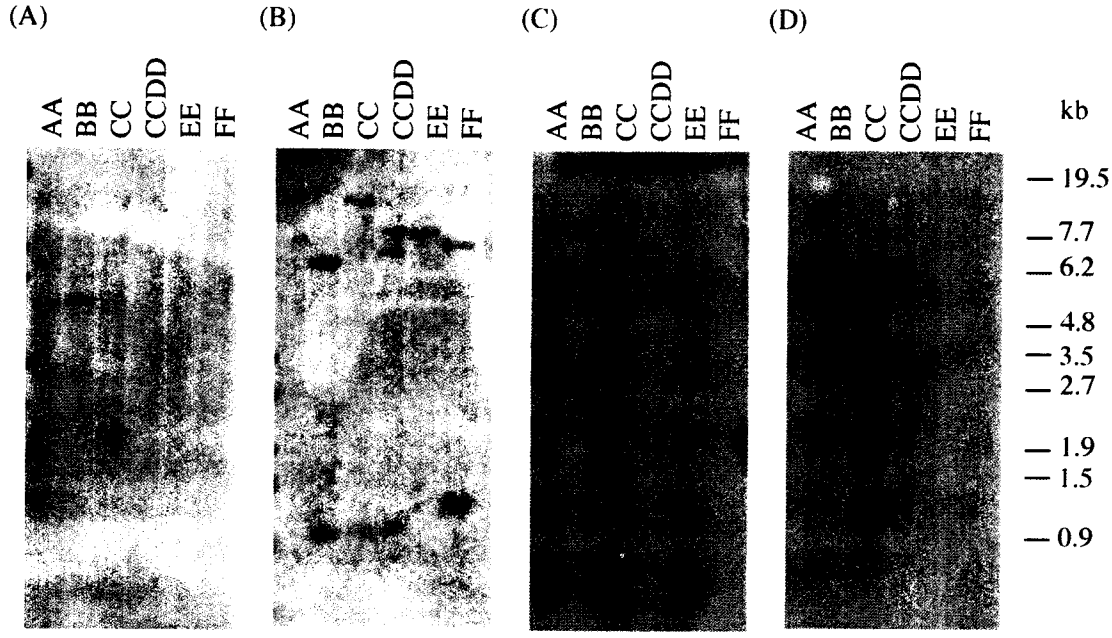


Fig. III-1. Southern hybridization of four genes unique to the wild relatives of rice.

Genomic DNAs of following species were digested with *Hind*III, electrophoresed, blotted onto the membrane, and hybridized with labelled probes of (A) BBS02H03, (B) BBS13D01, (C) CCP17F01 and (D) CCP19B09. Lanes on the blot are as follows. AA: *O. sativa, japonica*, BB: *O. punctata*, CC: *O. officinalis*, CCDD: *O. latifolia*, EE: *O. australiensis*, FF: *O. brachyantha*.

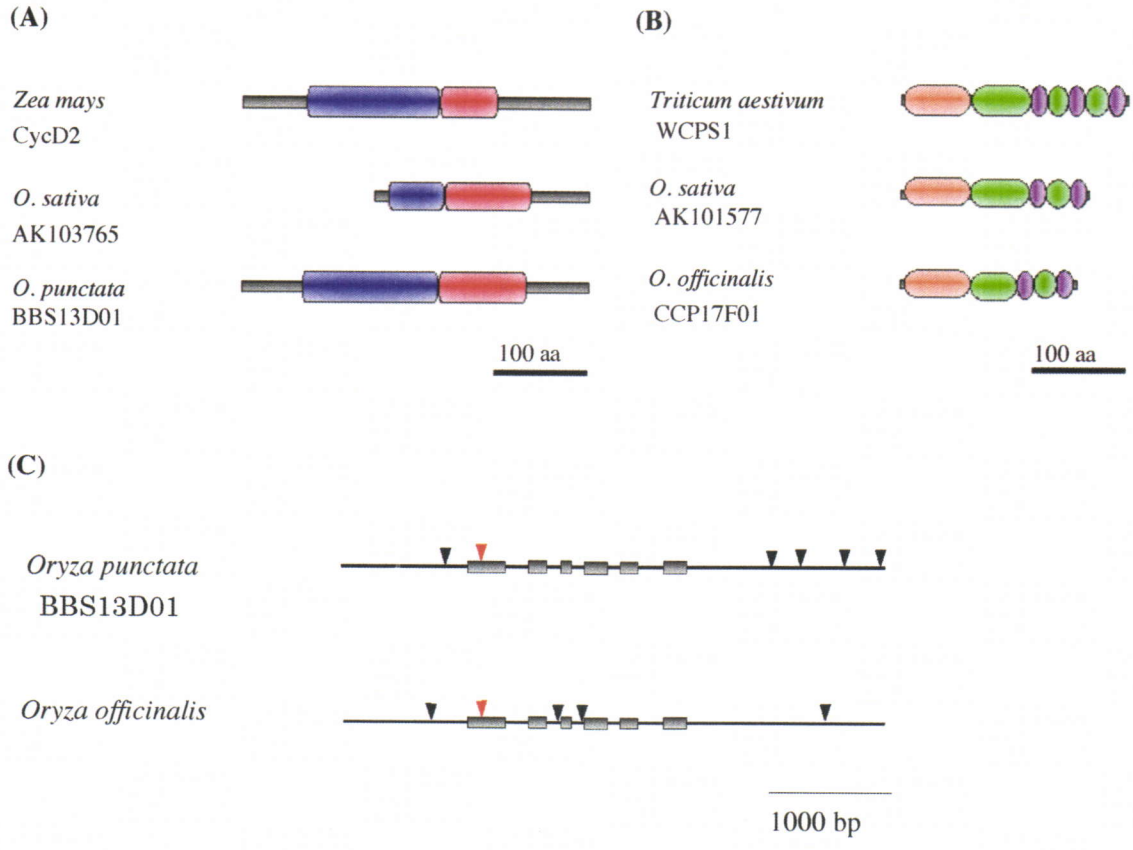


Fig. III-2. Domain structure of the unique proteins.

(A) Domain structure of BBS13D01 protein and its homologues. Blue: cyclin N-terminal domain, Red: cyclin C-terminal domain. (B) Domain structure of CCP17F01 and its orthologous proteins. Orange: cold shock domain, Green: Glycine-rich domain, Magenta: zinc-finger domain. (C) Genome structure of BBS13D01 and its homologue in *O. officinalis*. Boxes: exons, filled triangles: insertion, red triangle: insertion in an exon

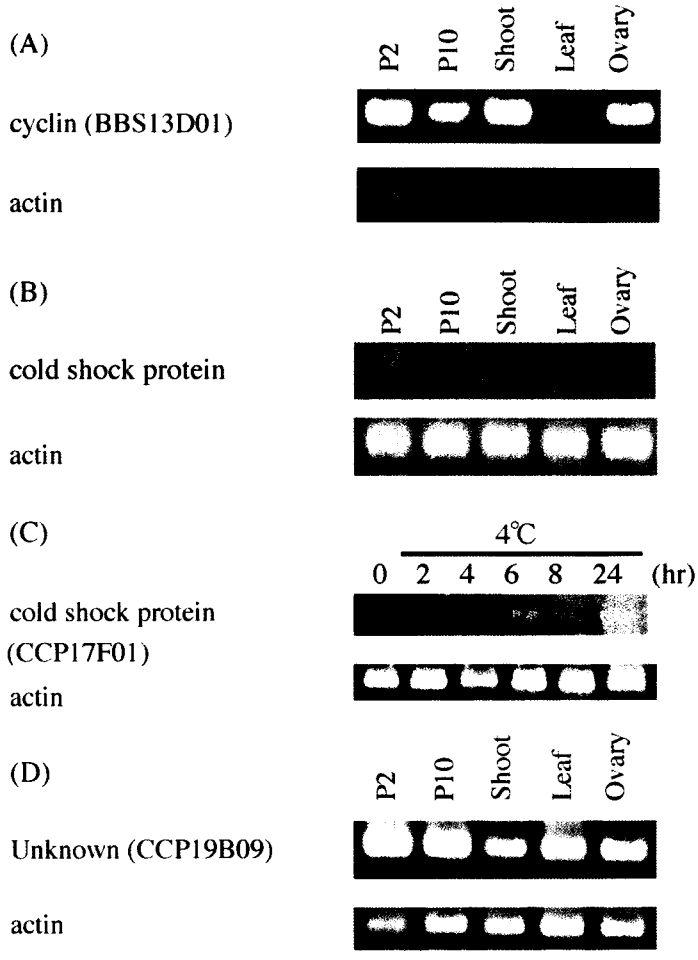


Fig. III-3. Expression patterns of the genes unique to the wild relatives of rice.

RT-PCR analysis was carried out using specific primers for (A) BBS13D01, (B, C) CCP17F01 and (D) CCP19B09. Actin was used as an internal control. Signals were visualized with ethidium bromide staining. P2: panicles (ca. 2 cm in length), P10: panicles (ca. 10 cm in length), Shoot: shoot apex, Leaf: leaf, Ovary: ovary 0-5 days after flowering. The numbers in (C) indicate hours after cold treatment at 4℃.

(A) (B) (C) (D)

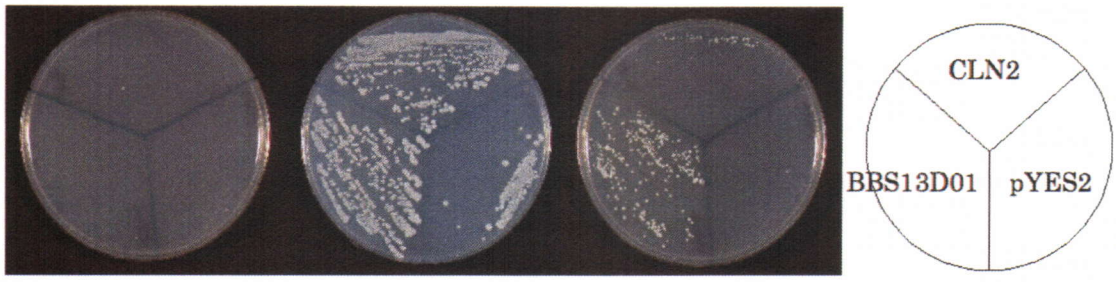


Fig. III-4. Complementation of the cyclin-deficient yeast mutant.

Yeast cells transformed with the construct indicated in (D) were spread onto the following media; (A) an SD plate containing 20 mM methionine and lacking galactose (B) an SD plate lacking galactose and methionine (C) an SGAL plate containing 20 mM methionine and 2% galactose.

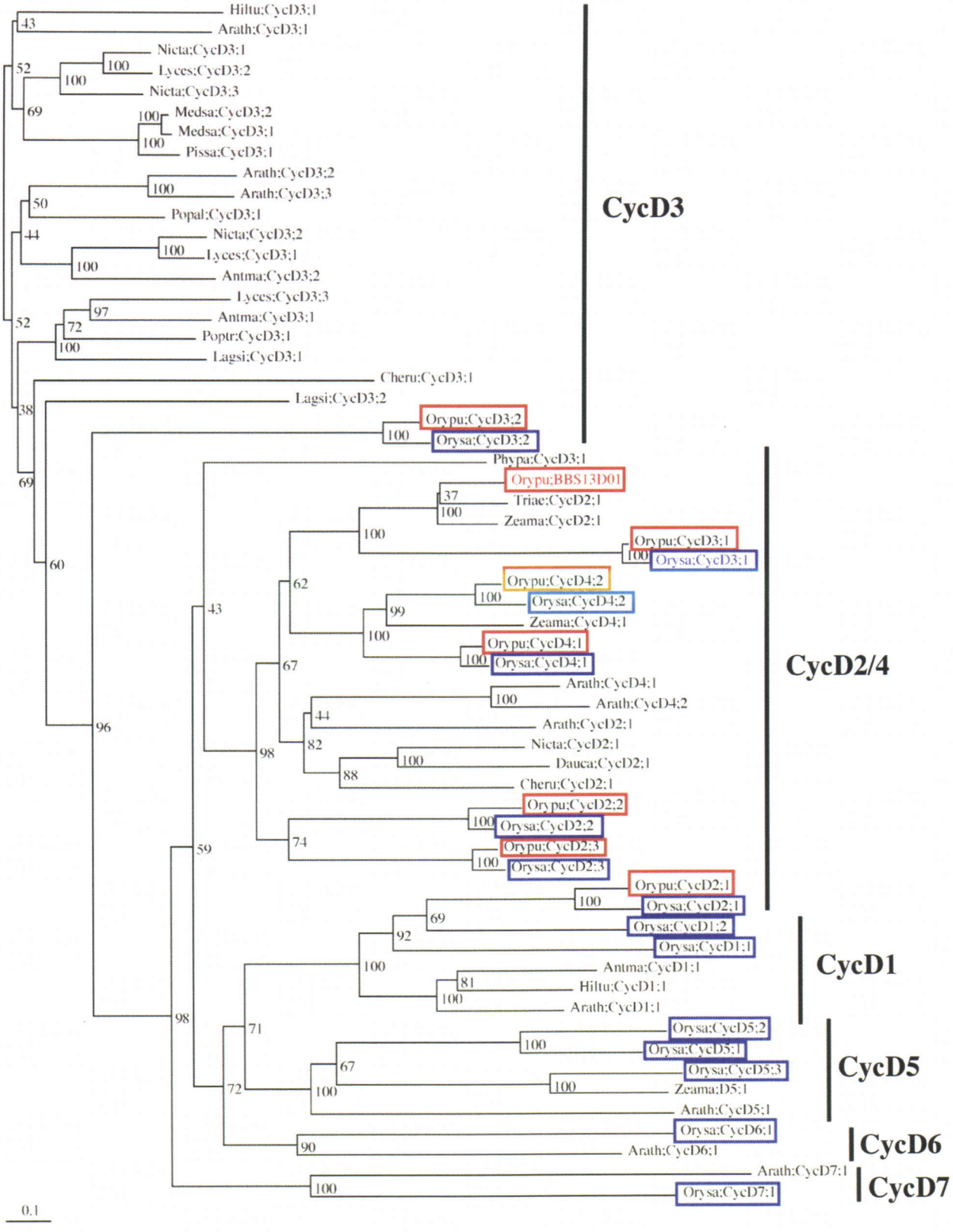

Hiltu;CycD3;1
Arath;CycD3;1
Nicta;CycD3;1
Lyces;CycD3;2
Nicta;CycD3;3
Medsa;CycD3;2
Medsa;CycD3;1
Pissa;CycD3;1
Arath;CycD3;2
Arath;CycD3;3
Popal;CycD3;1
Nicta;CycD3;2
Lyces;CycD3;1
Antma;CycD3;2
Lyces;CycD3;3
Antma;CycD3;1
Poptr;CycD3;1
Lagsi;CycD3;1
Cheru;CycD3;1
Lagsi;CycD3;2
Orypu;CycD3;2
Orysa;CycD3;2
Phypa;CycD3;1
Orypu;BBS13D01
Triae;CycD2;1
Zeama;CycD2;1
Orypu;CycD3;1
Orysa;CycD3;1
Orypu;CycD4;2
Orysa;CycD4;2
Zeama;CycD4;1
Orypu;CycD4;1
Orysa;CycD4;1
Arath;CycD4;1
Arath;CycD4;2
Arath;CycD2;1
Nicta;CycD2;1
Dauca;CycD2;1
Cheru;CycD2;1
Orypu;CycD2;2
Orysa;CycD2;2
Orypu;CycD2;3
Orysa;CycD2;3
Orypu;CycD2;1
Orysa;CycD2;1
Orysa;CycD1;2
Orysa;CycD1;1
Antma;CycD1;1
Hiltu;CycD1;1
Arath;CycD1;1
Orysa;CycD5;2
Orysa;CycD5;1
Orysa;CycD5;3
Zeama;D5;1
Arath;CycD5;1
Orysa;CycD6;1
Arath;CycD6;1
Arath;CycD7;1
Orysa;CycD7;1
CycD3
CycD2/4
CycD1
CycD5
CycD6
CycD7
0.1

Fig. III-5. The unrooted NJ tree of D type cyclins from plants.

O. punctata BBS13D01 and *O. sativa* AK103765 are shown in red and blue, respectively. Cyclin Ds clones and sequenced in this study from *O. punctata* and retrieved from cDNA sequences of *O. sativa. japonica* are enclosed in red and blue rectangles, respectively. Antma, *Antirrhinum majus*; Arath, *Arabidopsis thaliana*; Cheru, *Chenopodium rubrum*; Dauca, *Daucus carota*; Hiltu, *Helianthus tuberosus*; Lagsi, *Lagenaria siceraria*; Lyces, *Lycopersicon esculentum*; Medsa, *Medicago sativa*; Nicta, *Nicotiana tabacum*; Orysa, *Oryza sativa*; Orypu, *Oryza punctata*; Pissa, *Pisum sativum*; Phypa, *Physcomitrella patens*; Popal, *Populus alba*; Poptr, *Populus tremula* x *Populus tremuloides*; Triae, *Triticum aestivum*; Zeama, *Zea mays*.

Table III-1. Plant materials used in this study

Species	Genome type	Accession number
O. sativa, japonica	AA	Nipponbare
O. punctata	BB	W1514
O. officinalis	CC	W0002
O. latifolia	CCDD	W2200
O. australiensis	EE	W0008
O. brachyantha	FF	W1401

Table III-2. Primers used for the RT-PCR analysis

Gene		sequence
BBS13D01	Fw	TTCAGTGATCATGGTGCACC
	Rv	AATGTGATCAGTCAGCTGCA
CCP17F01	Fw	TCGAGTTCTCCATCAGCTCC
	Rv	TCCTCGCCACACTTGTAGCA
CCP19B09	Fw	ATGAGAAGAGGTTGCATTGC
	Rv	TGCTAGACTCAGTATGTGGC
actin	Fw	AACTGGGATGATATGGAGAA
	Rv	CCTCCAATCCAGACACTGTA

Fw: forword, Rv: reverse

Table III-3. Summary of wild relative unique genes

			5'RACE	Southern hybridization							
	Clone No	Length (bp)	Length (bp)	AA	BB	CC	CCDD	EE	FF	ORF (aa)	Indication
	14C06	553		+	++	++	+	++	++		
	02E05	466		+	++	+	+	+	+		
	02H03	267	776	-	++	-	-	-	-	111	ribosomal protein
BBP	02E09	613		-	++	++	++	-	-		
	09F10	1056		+	++	++	++	+	++		
	13D01	576	1628	-	++	++	++	++	++	349	cyclin D2
	09E08	1297		+	++	+	+	+	+		
	14C06	493		+	+	++	++	-	-		
	15A03	331		+	++	++	+	-	-		
	19B09	463	1045	-	-	++	-	-	-	128	unknown
	16A06	393		+	+	++	-	+	+		
	21C01	485		-	++	++	-	-	-		
	14F06	519	566	-	-	++	-	-	++	48	unknown
CCP	04H05	368		-	++	++	-	-	-		
	19E03	366	361	+	+	++	++	+	+	50	unknown
	14F08	180		-	++	++	-	-	+		
	17F10	517		+	-	++	++	-	-		
	17F01	289	900	+	+	++	++	-	-	180	cold response protein
	18D06	447		+	+	++	++	++	-		
	04D10	192	335	-	++	++	-	-	-	20	unknown

++; clear band by Southern hybridization, +; faint band by Southern hybridization, -; no band by Southern hybridization

General disccusion

In this study, I examined diversity of *Oryza* species by comparing expressed sequences of AA genome with those of BB or CC genomes. First subject was investigation of nucleotide substitution, insertion/deletion and synonymous or non-sysnonymous substitution between AA and BB, CC genome. The results showed that substitution rate between *japonica* and *O. punctata*, and between *japonica* and *O. officinalis* were 32 and 38 bp/kb in the coding region and 46 and 46 bp/kb in the untranslated region, respectively. They were three or four times higher than those between *japonica* and *indica*. The number of nonsynonymous and synonymous substitutions per site (Ka and Ks) was calculated 0.016 and 0.018 for Ka and 0.093 and 0.101 for Ks between *japonica* and *O. punctata* or *O. officinalis*, respectively. These results together with indel comparison indicate that the genetic structure of BB and CC genomes are very close with each other and indels are accumulated in the untranslated region.

Second subject was divergence of nine genomes of the genus *Oryza*. The phylogenic relationships of the species in the genus *Oryza* were examined using seven nuclear DNA sequences; *O. sativa* full length cDNA (AK058507, AK059353 and AK098919), RAFTIN 1 (AJ575667), MAP kinase 5 (AF479884), fructokinase 1 (AF429948), teosinte branched 1 (*tb1*) and tryptophan synthase (*orp*), and one intron sequence of *Adh1* 3rd intron. The results showed that the phylogenetic relationships among *O. brachyantha* (FF genome), AA genome species (*O. sativa, japonica, O. rufipogon* and *O. barthii*) and other *Oryza* species was supported by high bootstrap values, while in the *O. officinalis* complex being composed of BB, CC, CCDD, and EE genomes, pylogenetic tree was entangled among genes used for the analysis. Estimation of the divergence time based on the molecular clock approach was performed by assuming maize and rice diverged 50 million years ago [mya]. *O. brachantha* (FF) vs. remaining *Oryza* species revealed to be at 16.3 mya. The divergence time

within the *O. officinalis* complex containing BB, CC, CCDD and EE genomes was estimated later than 12.8 mya. AA genome was diverged from *O. officinalis* complex at 8.0 mya and speciation within the AA genome was estimated to occur at 0.6 mya. These values can explain well the phylogenetic relationships and the estimatied divergence time of these species in relation to other evolutional evidences. Synonymous (Ks) and non-synonymous (Ka) substitution rates showed that RAFTIN 1 and *tb* showed higher Ka/Ks than other genes in all genome types. These genes might contribute to speciation, or phenotypic diversity.

Last subject was analysis of unique genes in the wild relatives of rice. Thirteen clones identified as unique to the wild relatives, but not to *japonica* rice, were analyzed. Full length cDNAs of four genes were obtained by 5'-RACE of BB or CC genome RNAs. Althogh these four genes revealed to encod a ribosomal protein, a cyclin, a cold response protein and an unknown protein, their sequences were largely different from those or completely lacking in *O. sativa*. One of the BB genome clones, the BBS13D01 encodes a cyclin D2/4 like protein and its expression in yeast cells complemented cyclin mutations. This indicates that the BBS13D01 encodes a functional cyclin unique to the wild relatives of rice. Another clone CCP17F01 from CC genome has domains conserved among cold response proteins, and its expression was induced by a cold treatment. This suggests that CCP1701 encodes a cold response protein and is involved in the cold response of the wild relatives. The results suggest that wild relatives of rice have various unique genes which may be related to the characteristic features of the wild relatives in the morphogenesis and stress resistance. Unique genes of the wild relatives were devided into two categories: One was unique sequence present only in the wild relatives of rice, and the other has homologous genes in *japonica*, but its expression might be controlled differently.

Answers to the questions raised in this study would provide new insight on the phylogenetic relationship of *Oryza* species, useful genes lost in the cultivated rice and on the genetic diversity among genomes.

My approach to analyze molecular characteristics of the wild relatives of *Oryza* species brought new information about nucleotide substitution rate, in/del characteristics, unique genes of the wild relatives, phylogenetic relationships and divergence time of the wild species of rice. *Oryza* is one of the most valuable species in the studies of domestication, speciation, polyploidization and evolutional analysis, due to the accumulation of fundamental knowledge of genomics and a plenty of genetic resources. Further analysis of the wild relatives of rice should provide us a lot of fundamental and valuable resources to resolve questions on evolution and speciation, and on practical application of the genomes in the genus *Oryza*.

Acknowledgement

I thank Drs. Hiroyuki Kanamori and Nobuyuki Namiki (Society for Techno-innovation of Agriculture, Forestry and Fisheries) for EST sequencing, Dr. Yukiko Yamazaki (Genetic Informatics Laboratory, National institute of Genetics, Shizuoka, Japan) for BLAST search of EST, and Drs. Seiji Tanaka and Hiroyuki Araki (Division of Microbial Genetics, National institute of Genetics, Shizuoka, Japan) and Dr. Kim Nasmyth (Research Institute of Molecular Pathology, Austria) for kindly providing me the yeast strain K3413.

I thank Dr. Nori Kurata (Genetic Informatics Laboratory, National institute of Genetics, Shizuoka, Japan) for her advises and encouragement to pursue this research and Plant Genetic Laboratory members for their help of my experiments and for kind advises.

In closing, Iwould like to thank my parents for full support to me.

Literature Cited

General Introduction

Aggarwal, RK, Brar, DS, Nandi, S, Huang, N, Khush, GS, "*Phylogenetic relationships among Oryza species revealed by AFLP markers.*" , Theor Appl Genet **98** (1999) 1320-1328.

Ammiraju, JS, Luo, M, Goicoechea, JL, Wang, W, Kudrna, D, Mueller, C, Talag, J, Kim, H, Sisneros, NB, Blackmon, B, Fang, E, Tomkins, JB, Brar, D, MacKill, D, McCouch, S, Kurata, N, Lambert, G, Galbraith, DW, Arumuganathan, K, Rao, K, Walling, JG, Gill, N, Yu, Y, SanMiguel, P, Soderlund, C, Jackson, S,

Bautista, NS, Solis, R, Kamijima, O, Ishii, T, *"RAPD, RFLP and SSLP analyses of phylogenetic relationships between cultivated and wild species of rice."*, Genes Genet Syst **76** (2001) 71-9.

Bao, Y, Ge, S, "*Origin and phylogeny of Oryza species with the CD genome based on multople-gene sequence data.*" , Plant Syst Evol **249** (2004) 55-66.

Cheng, C, Tsuchimoto, S, Ohtsubo, H, Ohtsubo, E, *"Evolutionary relationships among rice species with AA genome based on SINE insertion analysis."*, Genes Genet Syst **77** (2002) 323-34.

(a) Cho, SK, Jeung, JU, Kang, KH, Shim, KS, Jung, KW, You, MK, Ok, SH, Chung, YS, Hwang, HG, Choi, HC, Moon, HP, Shin, JS, *"Identification of genes induced in wound-treated wild rice (Oryza minuta)."*, Mol Cells **17** (2004) 230-6.

Cho, SK, Jung, KW, Jeung, JU, Kang, KH, Shim, KS, You, MK, Yoo, KS, Ok, SH, Shin, JS, *"Analysis of differentially expressed transcripts from planthopper-infested wild rice (Oryza minuta)."*, Plant Cell Rep **24** (2005) 59-67.

Cho, SK, Ok, SH, Jeung, JU, Shim, KS, Jung, KW, You, MK, Kang, KH, Chung, YS, Choi, HC, Moon, HP, Shin, JS,*"Comparative analysis of 5,211 leaf ESTs of wild rice (Oryza minuta)."*, Plant Cell Rep **22** (2004)

839-47.

Ge, S, Li, A, Lu, BR, Ahang, SZ, Hong, DY, "*A phylogeny of the rice tribe Oryzeae (Poaceae) Based on matK sequence data.*" , American Journal of Botany **89** (2002) 1967-1972.

Ge, S, Sang, T, Lu, BR, Hong, DY, *"Phylogeny of rice genomes with emphasis on origins of allotetraploid species.",* Proc Natl Acad Sci USA **96** (1999) 14400-5.

Guo, YL, Ge, S, "*Molecular phylogeny of Oryzeae (Poaceae) based on DNA sequences from chloroplast, mitochondrial, and nuclear genomes.*" , American Journal of Botany **92** (2005) 1548-1558.

Ishii, T, Nakano, T, Maeda, H, Kamijima, O, *"Phylogenetic relationships in A-genome species of rice as revealed by RAPD analysis.",* Genes Genet Syst **71** (1996) 195-201.

Ishii, T, Xu, Y, McCouch, SR, *"Nuclear- and chloroplast-microsatellite variation in A-genome species of rice.",* Genome **44** (2001) 658-66.

Iwamoto, M, Nagashima, H, Nagamine, T, Higo, H, Higo, K, *"pSINE1- like intron of the CatA catalase homologs and phylogenetic relationships among AA genome Oryza and related species."* , Theor Appl Genet **98** (1999) 853-861.

Londo, JP, Chiang, YC, Hung, KH, Chiang, TY, Schaal, BA, *"Phylogeography of Asian wild rice, Oryza rufipogon, reveals multiple independent domestications of cultivated rice, Oryza sativa.",* Proc Natl Acad Sci USA **103** (2006) 9578-83.

Lu, BR, Naredo, MBE, Juliano, AB, Jackson, MT, "*Preliminary studies on taxonomy and biosystematics of the AA-genome. Oryza species (Poaceae).*" , In: Jacobs SWL, Everett J, eds. Grasses, systematics and evolution. Melbourne, Australia:CSIRO, (2000) 51-58.

Lu, BR, Zheng, KL, Qian, HR, Zhuang, JY, *"Genetic differentiation of wild relatives of rice as assessed by*

RFLP analysis.", Theor Appl Genet **106** (2002) 101-6.

Mochizuki, K, Ohtsubo, H, Hirano, H, Sano, Y, Ohtsubo, E, *"Classification and relationships of rice strains with AA genome by identification of transposable elements at nine loci."*, Jpn J Genet **68** (1993) 205-17.

Nayer, NM, *"Origin and cytogenetics of rice."*, Advanced in Genetics Incorporating Genetic Medicine **17** (1973) 153-292.

Park, KC, Kim, NH, Cho, YS, Kang, KH, Lee, JK, Kim, NS, *"Genetic variations of AA genome Oryza species measured by MITE-AFLP."*, Theor Appl Genet **107** (2003) 203-9.

Ren, F, Lu, BR, Li, S, Huang, J, Zhu, Y, *"A comparative study of genetic relationships among the AA-genome Oryza species using RAPD and SSR markers."*, Theor Appl Genet **108** (2003) 113-20.

Second, G, "*Origin of the genetic diversity of cultivated rice: study of the polymorphism scored at 40 isozyme loci.* ", Japanese Journal of Genetics , **57** (1982) 25-57.

Shim, KS, Cho, SK, Jeung, JU, Jung, KW, You, MK, Ok, SH, Chung, YS, Kang, KH, Hwang, HG, Choi, HC, Moon, HP, Shin, J, *"Identification of fungal (Magnaporthe grisea) stress-induced genes in wild rice (Oryza minuta)."*, Plant Cell Rep **22** (2004) 599-607.

Tateoka, T, *"Taxonomic studies of Oryza III. Key to the species and their enumeration."*, Bot Mag Tokyo **76** (1963) 165-173.

Vaughan, DA. "*The genus Oryza L. Current status of taxonomy.*" , Manila, The Philippines: IRRI. (1989)

Vaughan, DA, *"The Wild relatives of Rice: A genetic Resources Handbook."*, Manila, Philippines: International Rice Research Institute (1994).

Vaughan, DA, Morishima, H, Kadowaki, K, *"Diversity in the Oryza genus."*, Curr Opin Plant Biol **6** (2003) 139-46.

Wang, ZY, Second, G, Tanksley, SD, "*Polymorphism and phylogenetic relationships among species in the genus Oryza as determined by analysis of nuclear RFLPs.*" , Theor Appl Genet **83** (1992) 565-581.

Wing, RA, Ammiraju, JS, Luo, M, Kim, H, Yu, Y, Kudrna, D, Goicoechea, JL, Wang, W, Nelson, W, Rao, K, Brar, D, Mackill, DJ, Han, B, Soderlund, C, Stein, L, SanMiguel, P, Jackson, S, *"The oryza map alignment project: the golden path to unlocking the genetic potential of wild rice species."*, Plant Mol Biol **59** (2005) 53-62.

Yoshida, K, Miyashita, NT, Ishii, T, *"Nucleotide polymorphism in the Adh1 locus region of the wild rice Oryza rufipogon."*, Theor Appl Genet **109** (2004) 1406-16.

Zhu, Q, Ge, S, *"Phylogenetic relationships among A-genome species of the genus Oryza revealed by intron sequences of four nuclear genes."*, New Phytol **167** (2005) 249-65.

Chpter I

Aggarwal, RK, Brar, DS, Nandi, S, Huang, N, Khush, GS, "*Phylogenetic relationships among Oryza species revealed by AFLP markers.*" , Theor Appl Genet **98** (1999) 1320-1328.

Altschul, SF, Madden, TL, Schaffer, AA, Zhang, J, Zhang, Z, Miller, W, Lipman, DJ,, *"Gapped BLAST and PSI-BLAST: a new generation of protein database search programs."*, Nucleic Acids Res **25** (1997) 3389-402.

Ammiraju, JS, Luo, M, Goicoechea, JL, Wang, W, Kudrna, D, Mueller, C, Talag, J, Kim, H, Sisneros, NB, Blackmon, B, Fang, E, Tomkins, JB, Brar, D, MacKill, D, McCouch, S, Kurata, N, Lambert, G, Galbraith, DW, Arumuganathan, K, Rao, K, Walling, JG, Gill, N, Yu, Y, SanMiguel, P, Soderlund, C, Jackson, S,

Chen, X, Cho, YG, McCouch, SR, "*Sequence divergence of rice microsatellites in Oryza and other plant species.*", Mol Genet Genomics **268** (2002) 331-43.

Cheng, C, Motohashi, R, Tsuchimoto, S, Fukuta, Y, Ohtsubo, H, Ohtsubo, E, *"Polyphyletic origin of cultivated rice: based on the interspersion pattern of SINEs."*, Mol Biol Evol **20** (2003) 67-75.

Cheng, C, Tsuchimoto, S, Ohtsubo, H, Ohtsubo, E, *"Evolutionary relationships among rice species with AA genome based on SINE insertion analysis."*, Genes Genet Syst **77** (2002) 323-34.

Cho, SK, Jung, KW, Jeung, JU, Kang, KH, Shim, KS, You, MK, Yoo, KS, Ok, SH, Shin, JS, *"Analysis of differentially expressed transcripts from planthopper-infested wild rice (Oryza minuta)."*, Plant Cell Rep **24** (2005) 59-67.

Comeron, JM, *"A method for estimating the numbers of synonymous and nonsynonymous substitutions per site."*, J Mol Evol **41** (1995) 1152-1159.

Comeron, JM, *"K-Estimator: calculation of the number of nucleotide substitutions per site and the confidence intervals."*, Bioinformatics **15** (1999) 763-4.

Devos, KM, "*Updating the 'Crop Circle' "*, Curr Opin Plant Biol **8** (2005) 155-162.

Ge, S, Li, A, Lu, BR, Ahang, SZ, Hong, DY, "*A phylogeny of the rice tribe Oryzeae (Poaceae) Based on matK sequence data.*" , American Journal of Botany **89** (2002) 1967-1972.

Guo, YL, Ge, S, "*Molecular phylogeny of Oryzeae (Poaceae) based on DNA sequences from chloroplast, mitochondrial, and nuclear genomes.*" , American Journal of Botany **92** (2005) 1548-1558.

Li, C, Zhou, A, Sang, T, "*Genetic analysis of rice domestication syndrome with the wild annual species, Oryza nivara.*", New Phytol **170** (2006) 185-93.

Li, WH, *" Unbiased Estimation of the Rates of Synonymous and Nonsynonymous Substitution."*, J Mol Evol **36**, (1993) 96-99.

Lu, J, Tang, T, Tang, H, Huang, J, Shi, S, Wu, CI, "*The accumulation of deleterious mutations in rice genomes: a hypothesis on the cost of domestication.*", Trends Genet **22** (2006) 126-31.

Pamilo, P, Bianchi, NO, *"Evolution of the Zfx and Zfy genes: rates and interdependence between the genes."*, Mol Biol Evol **10** (1993) 271-281.

Shim, KS, Cho, SK, Jeung, JU, Jung, KW, You, MK, Ok, SH, Chung, YS, Kang, KH, Hwang, HG, Choi, HC, Moon, HP, Shin, JS, *"Identification of fungal (Magnaporthe grisea) stress-induced genes in wild rice (Oryza minuta)."*, Plant Cell Rep **22** (2004) 599-607.

Swigonová, Z, Lai, J, Ma, J, Ramakrishna, W, Llaca, V, Bennetzen, JL, Messing, J, "*Close split of sorghum and maize genome progenitors*", Genome Res **14** (2004) 1916-1923.

Tang, T, Lu, J, Huang, J, He, J, McCouch, SR, Shen, Y, Kai, Z, Purugganan, MD, Shi, S, Wu, CI, "*Genomic variation in rice: genesis of highly polymorphic linkage blocks during domestication.*", PLoS Genet **17** (2006) e199.

Wing, RA, Ammiraju, JS, Luo, M, Kim, H, Yu, Y, Kudrna, D, Goicoechea, JL, Wang, W, Nelson, W, Rao, K, Brar, D, Mackill, DJ, Han, B, Soderlund, C, Stein, L, SanMiguel, P, Jackson, S, *"The oryza map alignment project: the golden path to unlocking the genetic potential of wild rice species."*, Plant Mol Biol **59** (2005) 53-62.

Yu, J, Wang, J, Lin, W, Li, S, Li, H, Zhou, J, Ni, P, Dong, W, Hu, S, Zeng, C, Zhang, J, Zhang, Y, Li, R, Xu, Z, Li, S, Li, X, Zheng, H, Cong, L, Lin, L, Yin, J, Geng, J, Li, G, Shi, J, Liu, J, Lv, H, Li, J, Wang, J, Deng, Y, Ran, L, Shi, X, Wang, X, Wu, Q, Li, C, Ren, X, Wang, J, Wang, X, Li, D, Liu, D, Zhang, X, Ji, Z, Zhao, W, Sun, Y, Zhang, Z, Bao, J, Han, Y, Dong, L, Ji, J, Chen, P, Wu, S, Liu, J, Xiao, Y, Bu, D, Tan, J, Yang, L, Ye, C, Zhang, J, Xu, J, Zhou, Y, Yu, Y, Zhang, B, Zhuang, S, Wei, H, Liu, B, Lei, M, Yu, H, Li, Y, Xu, H, Wei, S, He, X, Fang, L, Zhang, Z, Zhang, Y, Huang, X, Su, Z, Tong, W, Li, J, Tong, Z, Li, S, Ye, J, Wang, L, Fang, L, Lei, T, Chen, C, Chen, H, Xu, Z, Li, H, Huang, H, Zhang, F, Xu, H, Li, N, Zhao, C, Li, S, Dong, L, Huang, Y, Li, L, Xi, Y, Qi, Q, Li, W, Zhang, B, Hu, W, Zhang, Y, Tian, X, Jiao, Y, Liang, X, Jin, J, Gao, L, Zheng, W, Hao, B, Liu, S, Wang, W, Yuan, L, Cao, M, McDermott, J, Samudrala, R, Wang, J, Wong, GK, Yang, H, "*The Genomes of Oryza sativa: a history of duplications.*", PLoS Biol **3** (2005) e38.

Chpter II

Aggarwal, RK, Brar, DS, Nandi, S, Huang, N, Khush, GS, "*Phylogenetic relationships among Oryza species revealed by AFLP markers.*" , Theor Appl Genet **98** (1999) 1320-1328.

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 **309** (2005) 741-5.

Bao, Y, Ge, S, "*Origin and phylogeny of Oryza species with the CD genome based on multople-gene sequence data.*" , Plant Syst Evol **249** (2004) 55-66.

Bautista, NS, Solis, R, Kamijima, O, Ishii, T, *"RAPD, RFLP and SSLP analyses of phylogenetic relationships between cultivated and wild species of rice."*, Genes Genet Syst **76** (2001) 71-9.

Buso, GS, Rangel, PH, Ferreira, ME, *"Analysis of random and specific sequences of nuclear and cytoplasmic DNA in diploid and tetraploid american wild rice species (Oryza spp.)."*, Genome **44** (2001) 476-94.

Chang, TT, *"The origin, evolution, cultivation, dissemination, and diversification of Asian and Africa rice."*, Euphytica **25** (1976) 435-441.

Cheng, C, Motohashi, R, Tsuchimoto, S, Fukuta, Y, Ohtsubo, H, Ohtsubo, E, *"Polyphyletic origin of cultivated rice: based on the interspersion pattern of SINEs."*, Mol Biol Evol **20** (2003) 67-75.

Cheng, C, Tsuchimoto, S, Ohtsubo, H, Ohtsubo, E, *"Evolutionary relationships among rice species with AA genome based on SINE insertion analysis."*, Genes Genet Syst **77** (2002) 323-34.

Comeron, JM, *"K-Estimator: calculation of the number of nucleotide substitutions per site and the confidence intervals."*, Bioinformatics **15** (1999) 763-4.

Doyle, JJ, Doyle, JL, "*Nuclear protein-coding genes in phylogeny reconstruction and homology assessment: some examples grom Leguminosae.*" In: Hollogsworth PM, Bateman RM, Gornall RJ [eds.], Molecular

systematics and plant evolution, (1999) 229-254. Taylor and Francis, London, UK.

Doyle, JJ, Kanazin, V, Shoemaker, RC, *"Phylogenetic utility of histone H3 intron sequences in the perennial relatives of soybean (Glycine: Leguminosae).",* Mol Phylogenet Evol **6** (1996) 438-47.

Gaut, BS, *"Evolutionary dynamics of grass genomes",* New Phytol **154** (2002) 15-28.

Ge, S, Li, A, Lu, BR, Ahang, SZ, Hong, DY, "*A phylogeny of the rice tribe Oryzeae (Poaceae) Based on matK sequence data.*" , American Journal of Botany **89** (2002) 1967-1972.

Ge, S, Sang, T, Lu, BR, Hong, DY, *"Phylogeny of rice genomes with emphasis on origins of allotetraploid species.",* Proc Natl Acad Sci USA **96** (1999) 14400-5.

Grob, GBJ, Granvendeel, B, Eurings, MCM, "*Potential phylogenetic utility of nuclear FLORICAULA/LEAFY second intron: comparison with three chloroplast DNA regions in Amorphophallus (Araceae).*" , Molecular phylogenetics and Evolution **30** (2004) 12-23.

Gugerli, F, Senn, J, Anzidei, M, Madaghiele, A, Büchler, U, Sperisen, C, Vendramin, GG, *"Chloroplast microsatellites and mitochondrial nad1 intron 2 sequences indicate congruent phylogenetic relationships among Swiss stone pine (Pinus cembra), Siberian stone pine (Pinus sibirica), and Siberian dwarf pine (Pinus pumila).",* Mol Ecol **10** (2001) 1489-97.

Guo, YL, Ge, S, "*Molecular phylogeny of Oryzeae (Poaceae) based on DNA sequences from chloroplast, mitochondrial, and nuclear genomes.*" , American Journal of Botany **92** (2005) 1548-1558.

Hass, BL, Pires, JC, Porter, R, Phillips, RL, Jackson, SA, *"Comparative genetics at the gene and chromosome levels between rice (Oryza sativa) and wildrice (Zizania palustris).",* Theor Appl Genet **107** (2003) 773-82.

Higgins, D, Thompson, J, Gibson, T, Thompson, JD, Higgins, DG, Gibson, TJ, "*CLUSTAL W: improving the*

sensitivity of progressive multiple sequence alignment through sequence weighting,position-specific gap penalties and weight matrix choice.", Nucleic Acids Res **22** (1994) 4673-4680.

Hubbard, L, McSteen, P, Doebley, J, Hake, S, *"Expression patterns and mutant phenotype of teosinte branched1 correlate with growth suppression in maize and teosinte."*, Genetics **162** (2002) 1927-1935.

Ishii, T, Nakano, T, Maeda, H, Kamijima, O, *"Phylogenetic relationships in A-genome species of rice as revealed by RAPD analysis."*, Genes Genet Syst **71** (1996) 195-201.

Ishii, T, Xu, Y, McCouch, SR, *"Nuclear- and chloroplast-microsatellite variation in A-genome species of rice."*, Genome **44** (2001) 658-66.

Iwamoto, M, Nagashima, H, Nagamine, T, Higo, H, Higo, K, *"pSINE1- like intron of the CatA catalase homologs and phylogenetic relationships among AA genome Oryza and related species."* , Theor Appl Genet **98** (1999) 853-861.

Jena, KK, Khush, GS, Kochert, G, *"Comparative RFLP mapping of a wild rice, Oryza officinalis, and cultivated rice, O. sativa."*, Genome **37** (1994) 382-9.

Jiang, H, Dian, W, Liu, F, Wu, P, "*Isolation and characterization of two fructokinase cDNA clones from rice*", Phytochemistry **62** (2003) 47-52.

Kikuchi, S, Satoh, K, Nagata, T, Kawagashira, N, Doi, K, Kishimoto, N, Yazaki, J, Ishikawa, M, Yamada, H, Ooka, H, Hotta, I, Kojima, K, Namiki, T, Ohneda, E, Yahagi, W, Suzuki, K, Li, CJ, Ohtsuki, K, Shishiki, T, Otomo, Y, Murakami, K, Iida, Y, Sugano, S, Fujimura, T, Suzuki, Y, Tsunoda, Y, Kurosaki, T, Kodama, T, Masuda, H, Kobayashi, M, Xie, Q, Lu, M, Narikawa, R, Sugiyama, A, Mizuno, K, Yokomizo, S, Niikura, J, Ikeda, R, Ishibiki, J, Kawamata, M, Yoshimura, A, Miura, J, Kusumegi, T, Oka, M, Ryu, R, Ueda, M, Matsubara, K, Kawai, J, Carninci, P, Adachi, J, Aizawa, K, Arakawa, T, Fukuda, S, Hara, A,

Hashizume, W, Hayatsu, N, Imotani, K, Ishii, Y, Itoh, M, Kagawa, I, Kondo, S, Konno, H, Miyazaki, A, Osato, N, Ota, Y, Saito, R, Sasaki, D, Sato, K, Shibata, K, Shinagawa, A, Shiraki, T, Yoshino, M, Hayashizaki, Y, Yasunishi, A, Rice Full-Length cDNA Consortium, "*Collection, Mapping, and Annotation of Over 28,000 cDNA Clones from japonica Rice*" , Science **301** (2003) 376-379.

Kimura, M, (1980), "*A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences*", J. Mol. Evol **16** (1980) 111-120.

Kumar, S, Tamura, K, Nei, M, *"MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment."*, Brief Bioinform **5** (2004) 150-63.

Londo, JP, Chiang, YC, Hung, KH, Chiang, TY, Schaal, BA, *"Phylogeography of Asian wild rice, Oryza rufipogon, reveals multiple independent domestications of cultivated rice, Oryza sativa."*, Proc Natl Acad Sci USA **103** (2006) 9578-83.

Lu, BR, Naredo, MBE, Juliano, AB, Jackson, MT, "*Preliminary studies on taxonomy and biosystematics of the AA-genome. Oryza species (Poaceae).*" , In: Jacobs SWL, Everett J, eds. Grasses, systematics and evolution. Melbourne, Australia:CSIRO, (2000) 51-58.

Lu, BR, Zheng, KL, Qian, HR, Zhuang, JY, *"Genetic differentiation of wild relatives of rice as assessed by RFLP analysis."*, Theor Appl Genet **106** (2002) 101-6.

Mochizuki, K, Ohtsubo, H, Hirano, H, Sano, Y, Ohtsubo, E, *"Classification and relationships of rice strains with AA genome by identification of transposable elements at nine loci."*, Jpn J Genet **68** (1993) 205-17.

Nayer, NM, *"Origin and cytogenetics of rice."*, Advanced in Genetics Incorporating Genetic Medicine **17** (1973) 153-292.

Nishikawa, T, Vaughan, DA, Kadowaki, K, *"Phylogenetic analysis of Oryza species, based on simple*

sequence repeats and their flanking nucleotide sequences from the mitochondrial and chloroplast genomes.", Theor Appl Genet **110** (2005) 696-705.

Oh, SH, Potter, D, *"Phylogenetic utility of the second intron of LEAFY in Neillia and Stephanandra (Rosaceae) and implications for the origin of Stephanandra."*, Mol Phylogenet Evol **29** (2003) 203-15.

Park, KC, Kim, NH, Cho, YS, Kang, KH, Lee, JK, Kim, NS, *"Genetic variations of AA genome Oryza species measured by MITE-AFLP."*, Theor Appl Genet **107** (2003) 203-9.

Reddy, AS, Kiefer-Meyer, MC, Delseny, M, *"Characterization of new variants of a satellite DNA from Oryza officinalis, specific for the CC genome of wild rice."*, Genome **36** (1993) 750-61.

Ren, F, Lu, BR, Li, S, Huang, J, Zhu, Y, *"A comparative study of genetic relationships among the AA-genome Oryza species using RAPD and SSR markers."*, Theor Appl Genet **108** (2003) 113-20.

Saitou, N, Nei, M, (1987) "*The neighbor-joining method: A new method for reconstructing phylogenetic trees.*" Molecular Biology and Evolution **4** (1987) 406-425.

Sang, T, "*Utility of low-copy nuclear gene sequences in plant phylogenetics.*" , Critical Reviews in Biochemistry and Molecular Biology **37** (2002) 121–147.

Second, G, "*Origin of the genetic diversity of cultivated rice: study of the polymorphism scored at 40 isozyme loci.*", Japanese Journal of Genetics , **57** (1982) 25-57.

Soltis, DE, Soltis, PS, "*Choosing an approach and an appropriate gene for phylogenetic analysis.*" In: Soltis, DE, Soltis, PS, Doyle, JJ, eds. Molecular systematics of plants. II. DNA sequencing. Dordrecht, the Netherlands: Kluwer, (1998) 1–42.

Swigonová, Z, Lai, J, Ma, J, Ramakrishna, W, Llaca, V, Bennetzen, JL, Messing, J, "*Close split of sorghum and maize genome progenitors*", Genome Res **14** (2004) 1916-1923.

Tateoka, T, *"Taxonomic studies of Oryza III. Key to the species and their enumeration."*, Bot Mag Tokyo **76** (1963) 165-173.

Vaughan, DA. "*The genus Oryza L. Current status of taxonomy.*" , Manila, The Philippines: IRRI. (1989)

Vaughan, DA, Morishima, H, Kadowaki, K, *"Diversity in the Oryza genus."*, Curr Opin Plant Biol **6** (2003) 139-46.

Wang, ZY, Second, G, Tanksley, SD, "*Polymorphism and phylogenetic relationships among species in the genus Oryza as determined by analysis of nuclear RFLPs.*" , Theor Appl Genet **83** (1992) 565-581.

Wang, A, Xia, Q, Xie, W, Datla, R, Selvaraj, G, "*The classical Ubisch bodies carry a sporophytically produced structural protein (RAFTIN) that is essential for pollen development.*", Proc Natl Acad Sci USA **100** (2003) 14487-14492.

Xiong, L, Yang,Y, "*Disease Resistance and Abiotic Stress Tolerance in Rice Are Inversely Modulated by an Abscisic Acid-Inducible Mitogen-Activated Protein Kinase*", Plant Cell **15** (2003) 745-759.

Yoshida, K, Miyashita, NT, Ishii, T, *"Nucleotide polymorphism in the Adh1 locus region of the wild rice Oryza rufipogon."*, Theor Appl Genet **109** (2004) 1406-16.

Zhang, Y, Luo, L, Xu, C, Zhang, Q, Xing, Y, *"Quantitative trait loci for panicle size, heading date and plant height co-segregating in trait-performance derived near-isogenic lines of rice (Oryza sativa)."*, Theor Appl Genet **113** (2006) 361-368.

Zhu, Q, Ge, S, *"Phylogenetic relationships among A-genome species of the genus Oryza revealed by intron sequences of four nuclear genes."*, New Phytol **167** (2005) 249-65.

Chapter III

Ammiraju, JS, Luo, M, Goicoechea, JL, Wang, W, Kudrna, D, Mueller, C, Talag, J, Kim, H, Sisneros, NB, Blackmon, B, Fang, E, Tomkins, JB, Brar, D, MacKill, D, McCouch, S, Kurata, N, Lambert, G, Galbraith, DW, Arumuganathan, K, Rao, K, Walling, JG, Gill, N, Yu, Y, SanMiguel, P, Soderlund, C, Jackson, S, Wing, RA, *"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 **16** (2006) 140-7.

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 **309** (2005) 741-5.

Brar, DS, Khush, GS, *"Alien introgression in rice."*, Plant Mol Biol **35** (1997) 35-47.

(a) Cho, SK, Jeung, JU, Kang, KH, Shim, KS, Jung, KW, You, MK, Ok, SH, Chung, YS, Hwang, HG, Choi, HC, Moon, HP, Shin, JS, *"Identification of genes induced in wound-treated wild rice (Oryza minuta)."*, Mol Cells **17** (2004) 230-6.

Cho, SK, Jung, KW, Jeung, JU, Kang, KH, Shim, KS, You, MK, Yoo, KS, Ok, SH, Shin, JS, *"Analysis of differentially expressed transcripts from planthopper-infested wild rice (Oryza minuta)."*, Plant Cell Rep **24** (2005) 59-67.

Cho, SK, Ok, SH, Jeung, JU, Shim, KS, Jung, KW, You, MK, Kang, KH, Chung, YS, Choi, HC, Moon, HP, Shin, JS, *"Comparative analysis of 5,211 leaf ESTs of wild rice (Oryza minuta)."*, Plant Cell Rep **22** (2004) 839-47.

Cockcroft, CE, den Boer, BG, Healy, JM, Murray, JA, *"Cyclin D control of growth rate in plants."*, Nature **405** (2000) 575-9.

de Oliveira, DE, Seurinck, J, Inze, D, Van Montagu, M, Botterman, J, "*Differential expression of five Arabidopsis genes encoding glycine-rich proteins.*" , Plant Cell **2** (1990) 427-436.

Dewitte, W, Riou-Khamlichi, C, Scofield, S, Healy, JM, Jacqmard, A, Kilby, NJ, Murray, JA, *"Altered cell cycle distribution, hyperplasia, and inhibited differentiation in Arabidopsis caused by the D-type cyclin CYCD3."*, Plant Cell **15** (2003) 79-92.

Goldstein, J, Pollitt, NS, Inouye, M, *"Major cold shock protein of Escherichia coli."*, Proc Natl Acad Sci U S A **87** (1990) 283-7.

Jeanmougin, F, Thompson, JD, Gouy, M, Higgins, DG, Gibson, TJ, "*Multiple sequence alignment with Clustal x*", Trends Biochem Sci **23** (1998) 403-405.

Jeong, DH, An, S, Park, S, Kang, HG, Park, GG, Kim, SR, Sim, J, Kim, YO, Kim, MK, Kim, SR, Kim, J, Shin, M, Jung, M, An, G, *"Generation of a flanking sequence-tag database for activation-tagging lines in japonica rice."*, Plant J **45** (2006) 123-32.

Karlson, D, Nakaminami, K, Toyomasu, T, Imai, R, *"A cold-regulated nucleic acid-binding protein of winter wheat shares a domain with bacterial cold shock proteins."*, J Biol Chem **277** (2002) 35248-56.

Kikuchi, S, Satoh, K, Nagata, T, Kawagashira, N, Doi, K, Kishimoto, N, Yazaki, J, Ishikawa, M, Yamada, H, Ooka, H, Hotta, I, Kojima, K, Namiki, T, Ohneda, E, Yahagi, W, Suzuki, K, Li, CJ, Ohtsuki, K, Shishiki, T, Otomo, Y, Murakami, K, Iida, Y, Sugano, S, Fujimura, T, Suzuki, Y, Tsunoda, Y, Kurosaki, T, Kodama, T, Masuda, H, Kobayashi, M, Xie, Q, Lu, M, Narikawa, R, Sugiyama, A, Mizuno, K, Yokomizo, S, Niikura, J, Ikeda, R, Ishibiki, J, Kawamata, M, Yoshimura, A, Miura, J, Kusumegi, T, Oka, M, Ryu, R, Ueda, M, Matsubara, K, Kawai, J, Carninci, P, Adachi, J, Aizawa, K, Arakawa, T, Fukuda, S, Hara, A, Hashizume, W, Hayatsu, N, Imotani, K, Ishii, Y, Itoh, M, Kagawa, I, Kondo, S, Konno, H, Miyazaki, A, Osato, N, Ota, Y,

Saito, R, Sasaki, D, Sato, K, Shibata, K, Shinagawa, A, Shiraki, T, Yoshino, M, Hayashizaki, Y, Yasunishi, A, Rice Full-Length cDNA Consortium, "*Collection, Mapping, and Annotation of Over 28,000 cDNA Clones from japonica Rice*" , Science **301** (2003) 376-379.

Kimura, M, (1980), "*A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences*", J. Mol. Evol **16** (1980) 111-120.

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 **312** (2006) 1392-6.

Kumar, S, Tamura, K, Nei, M, *"MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment."*, Brief Bioinform **5** (2004) 150-63.

La, H, Li, J, Ji, Z, Cheng, Y, Li, X, Jiang, S, Venkatesh, PN, Ramachandran, S, *"Genome-wide analysis of cyclin family in rice (Oryza Sativa L.)."*, Mol Genet Genomics **275** (2006) 374-86.

Meijer, M, Murray, JA, *"Cell cycle controls and the development of plant form."*, Curr Opin Plant Biol **4** (2001) 44-9.

Miyao, A, Tanaka, K, Murata, K, Sawaki, H, Takeda, S, Abe, K, Shinozuka, Y, Onosato, K, Hirochika, H, *"Target site specificity of the Tos17 retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome."*, Plant Cell **15** (2003) 1771-80.

Nakaminami, K, Karlson, DT, Imai, R, *"Functional conservation of cold shock domains in bacteria and higher plants."*, Proc Natl Acad Sci USA **103** (2006) 10122-7.

Oakenfull, EA, Riou-Khamlichi, C, Murray, JA, *"Plant D-type cyclins and the control of G1 progression."*, Philos Trans R Soc Lond B Biol Sci **357** (2002) 749-60.

Shim, KS, Cho, SK, Jeung, JU, Jung, KW, You, MK, Ok, SH, Chung, YS, Kang, KH, Hwang, HG, Choi, HC,

Moon, HP, Shin, JS, *"Identification of fungal (Magnaporthe grisea) stress-induced genes in wild rice (Oryza minuta)."*, Plant Cell Rep **22** (2004) 599-607.

Thompson, JD, Higgins, DG, Gibson, TJ, (1994), *"Clustal-W – improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice"*, Nucleic Acids Res **22** (1994) 4673 – 4680.

Vaughan, DA, *"The Wild relatives of Rice: A genetic Resources Handbook."*, Manila, Philippines: International Rice Research Institute (1994).

Wang, F, Huo, SN, Guo, J, Zhang, XS, *"Wheat D-type cyclin Triae;CYCD2;1 regulate development of transgenic Arabidopsis plants."*, Planta **224** (2006) 1129-40.

Wing, RA, Ammiraju, JS, Luo, M, Kim, H, Yu, Y, Kudrna, D, Goicoechea, JL, Wang, W, Nelson, W, Rao, K, Brar, D, Mackill, DJ, Han, B, Soderlund, C, Stein, L, SanMiguel, P, Jackson, S, *"The oryza map alignment project: the golden path to unlocking the genetic potential of wild rice species."*, Plant Mol Biol **59** (2005) 53-62.

Xiao, J, Grandillo, S, Ahn, SN, McCouch, SR, Tanksley, SD, Li, J, Yuan, L., *"Genes from wild rice improve yield"*, Nature **384** (1996) 223-224.

Xiao, J, Li, J, Grandillo, S, Ahn, SN, Yuan, L, Tanksley, SD, McCouch, SR, *"Identification of trait-improving quantitative trait loci alleles from a wild rice relative, Oryza rufipogon."*, Genetics **150** (1998) 899-909.

Yu, J, Wang, J, Lin, W, Li, S, Li, H, Zhou, J, Ni, P, Dong, W, Hu, S, Zeng, C, Zhang, J, Zhang, Y, Li, R, Xu, Z, Li, S, Li, X, Zheng, H, Cong, L, Lin, L, Yin, J, Geng, J, Li, G, Shi, J, Liu, J, Lv, H, Li, J, Wang, J, Deng, Y, Ran, L, Shi, X, Wang, X, Wu, Q, Li, C, Ren, X, Wang, J, Wang, X, Li, D, Liu, D, Zhang, X, Ji, Z, Zhao, W, Sun, Y, Zhang, Z, Bao, J, Han, Y, Dong, L, Ji, J, Chen, P, Wu, S, Liu, J, Xiao, Y, Bu, D, Tan, J, Yang, L,

Ye, C, Zhang, J, Xu, J, Zhou, Y, Yu, Y, Zhang, B, Zhuang, S, Wei, H, Liu, B, Lei, M, Yu, H, Li, Y, Xu, H, Wei, S, He, X, Fang, L, Zhang, Z, Zhang, Y, Huang, X, Su, Z, Tong, W, Li, J, Tong, Z, Li, S, Ye, J, Wang, L, Fang, L, Lei, T, Chen, C, Chen, H, Xu, Z, Li, H, Huang, H, Zhang, F, Xu, H, Li, N, Zhao, C, Li, S, Dong, L, Huang, Y, Li, L, Xi, Y, Qi, Q, Li, W, Zhang, B, Hu, W, Zhang, Y, Tian, X, Jiao, Y, Liang, X, Jin, J, Gao, L, Zheng, W, Hao, B, Liu, S, Wang, W, Yuan, L, Cao, M, McDermott, J, Samudrala, R, Wang, J, Wong, GK, Yang, H, "*The Genomes of Oryza sativa: a history of duplications.*", PLoS Biol **3** (2005) e38.

Zhang, Y, Luo, L, Xu, C, Zhang, Q, Xing, Y, *"Quantitative trait loci for panicle size, heading date and plant height co-segregating in trait-performance derived near-isogenic lines of rice (Oryza sativa).",* Theor Appl Genet **113** (2006) 361-368.