Analysis of a pair of genes, *DOPPELGANGER1* (*DPL1*) and *DOPPELGANGER2* (*DPL2*) responsible for reproductive isolation between two rice subspecies

Yoko Mizuta

DOCTOR OF PHILOSOPHY

Department of Genetics,

School of Life Science,

The Graduate University for Advanced Studies

2010

Contents

General Introduction

The evolution of species isolation mechanisms is a central topic in the study of speciation. Reproductive isolation is a genetic mechanism to isolate species (Dobzhansky, 1937; Coyne and Orr, 2004). Identification of genes responsible for barriers to gene flow between two species provides insight into molecular mechanisms of reproductive isolation and relationships between evolutions of barrier genes (Clark *et al.*, 2009; Ferris *et al.*, 1997; Herrmann *et al.*, 1999; Masly *et al.*, 2006; Noor and Feder, 2006; Oliver *et al.*, 2009; Orr and Presgraves, 2004; Phadnis and Orr, 2009).

In general, reproductive isolation is divided into two large categories: (I) before fertilization or zygote formation (prezygotic isolation), and (II) after fertilization or zygote formation (postzygotic isolation). Prezygotic isolation would be the strongest impediments to gene exchange, and such isolation reduces the chance of meeting and mating between incompatible species (Orr and Coyne, 2004). Prezygotic isolation is more effective than postzygotic isolation at preventing gene flow. However, postzygotic isolation, which is involved in most speciation events, is also important. If prezygotic isolation is essentially complete, hybrids can never be formed and we will know nothing about postzygotic isolation. Addition to that, postzygotic isolation via complete hybrid sterility, hybrid inviability restricts gene flow between species, just like a prezygotic isolation (Stebbins, 1958).

Several recent analyses documented that genetic factors caused postzygotic isolations and speciation, such as hybrid sterility (Ting *et al.*, 1998; Mihola *et al.*, 2009; Lee *et al.*, 2008), hybrid inviability (Wittbrodt *et al.*, 1989; Weis and Schartl, 1998) and hybrid breakdown (Seidel *et al.*,

2008). Despite more than two decades of studies in this field, only a small number of genes have been identified and the whole view of the molecular mechanisms of reproductive isolation remains unclear. How genes cause postzygotic isolation? To address these fundamental questions in speciation, Asian cultivated rice, *Oryza sativa* L., is an ideal species due to its population structure and genetic diversity. Rice experienced ancient genome duplication (Paterson *et al.,* 2004) and has syntenic relationships with other cereal species (Bolot *et al.,* 2009). *O. sativa* has two generally differentiated subspecies, *indica* and *japonica* (Kato, 1930) that are derived from the ancestral wild rice species, *O. rufipogon* (Oka, 1988). Not only ancestral wild rice, but also cultivated rice comprises a large variety of strains showing a wide-range of genetic diversity and structured populations (Garris *et al.,* 2005). *O. sativa* can produce hybrids with related wild rice species, but various genetic incompatibilities, such as hybrid inviability, hybrid sterility, hybrid weakness and hybrid breakdown are present (Oka, 1988). Even in intraspecific crosses of *O. sativa*, many reproductive barrier loci have been reported in hybrids between *japonica* and *indica* (http://www.shigen.nig.ac.jp/rice/oryzabase/top/top.jsp). Attempts for detection and isolation of the causal genes for reproductive isolation have not much succeeded yet, and the molecular mechanisms of isolation and speciation are still largely unknown.

To investigate the molecular mechanism of postzygotic isolation and the status of evolutionary differentiation within *O. sativa*, Harushima *et al.* (2001, 2002) established a new method to map reproductive barriers on chromosomes using an *indica-japonica* F_2 population. In these results, detected reproductive barriers were caused by the incompatible interactions among genes from *indica* and *japonica*. Thus, it became possible to detect the incompatible interaction

between genes for map-based cloning of the causal genes. From the result of the clarification of interaction, a prominent interactive pair of marker loci was detected in the vicinity of previously mapped reproductive barriers on chromosomes 1 and 6 in an *indica-japonica* F_2 population (Harushima *et al*., in preparation). It was also confirmed that this interaction occurs only in the male gametophyte (pollen) by the analysis of reciprocal backcrosses progenies.

In this study, I identified hybrid incompatible genes on rice chromosomes 1 and 6, responsible for this male gametophytic reproductive isolation in the hybrid between *indica* cultivar (cv.) Kasalath and *japonica* cv. Nipponbare. I focused on the following three questions: (I) How do genes cause this reproductive isolation? (II) What mechanism works on this reproductive isolation? (III) When this reproductive isolation was established? To address these questions, map-based cloning and characterization of the hybrid incompatible genes was carried out, and described in chapter I. In chapter II, I clarified when and how these hybrid incompatible genes have evolved along with species differentiation in plant evolution. My studies of hybrid incompatible genes and analyses of rice reproduction would provide fundamental understanding of molecular functions in plant reproduction and the mechanisms of species diversification.

Chapter I

Isolation and characterization of the reproductive isolation genes

Introduction

Intersubspecific hybrids between *indica* and *japonica* have significant agronomic character, such as hybrid vigor, heterosis, or various beneficial traits that provides great potential for further productivity increases in rice. However, varying degrees of reproductive isolations are commonly seen in cross between *indica* and *japonica* varieties (Stebbins, 1958; Oka, 1988; http://www.shigen.nig.ac.jp/rice/oryzabase/top/top.jsp). The partial or complete sterility and inviability of the hybrids forms a reproductive barrier hindering the utilization of agronomic useful traits for rice breeding.

Recently, the International Rice Genome Sequencing Project (IRGSP) completed the sequencing of the *japonica* rice genome (International Rice Genome Sequencing Project, 2005), and the draft sequences of *indica* rice cultivar, 9311, have been determined (Zhao *et al.,* 2004). The high-density linkage map constructed with 2,275 RFLP markers covering rice whole genome was reported (Harushima *et al.,* 1998). These researches provide useful information to identify and positional cloning of genes that responsible for reproductive isolation in rice. As a result of these advances, a growing number of reproductive barrier loci have been identified using the various cross combinations of rice (Liu *et al.*, 2001). Recent work in rice, Chen *et al.* (2008) cloned a gene encoding an aspartic protease at the locus *S5* conditioning embryo-sac sterility in *indica-japonica* hybrid. Long *et al.* (2008) isolated two adjacent genes encoding a small ubiquitin-like modifier E3 ligase-like protein and an F-box protein that responsible for *indica-japonica* hybrid male sterility. These two independent researches provided insights into the genetics of postzygotic reproductive

isolation in hybrids between *indica* and *japonica*, although a whole view of the molecular mechanism of reproductive isolation in a rice intersubspecific cross remains unknown.

To investigate the molecular mechanisms of reproductive isolation and the status of evolutionary differentiation within *O. sativa*, Harushima *et al.* developed a method to map reproductive barriers on chromosomes by regression analysis of allele frequencies using an *indica-japonica* F_2 population (Harushima *et al.,* 2001). The cross between *indica* cultivar (cv.) Kasalath and *japonica* cv. Nipponbare showed almost all fertile F_1 and F_2 population. Nevertheless, 33 reproductive barriers were mapped along whole chromosomes as a deviation from Mendelian segregation in an *indica-japonica* F_2 population. This reproductive isolation is one of hybrid breakdown. Hybrid breakdown is defined as a decrease of hybrid viability in a F_2 population or later generations compared to the F_1 or parents (Stebbins, 1958).

Two major genetic models for hybrid breakdown were supposed. One is an allelic interaction at a single locus (Kitamura, 1962; Ikehashi and Araki, 1986) and another is epistatic interaction at two independent loci (Oka, 1974; Sato and Morishima, 1987). In the former case, the interaction between alleles on one locus that derived from different parents causes genetic incompatibility in hybrid (Chen, *et al.* 2008; Koide *et al.* 2008). In the later case, unfavorable allelic combinations between two loci that derived from different parents, result in those alleles being eliminated from the progeny of a hybrid. Our laboratory have been investigated the reproductive isolation by the epistatic interaction between two loci. Two-way interactions were sought in a cross between *japonica* cultivar Nipponbare and *indica* cultivar Kasalath by testing independent segregation between all pair-wise marker combinations covering the whole genome in an F_2

population (Harushima *et al*., in preparation). As a result, a prominent interactive pair of marker loci was detected in the vicinity of previously mapped reproductive barriers on chromosomes 1 at round 45.4 cM and 6 at around 15.8 cM (Fig. I-1). The barriers on chromosome 1 and 6 reduced Kasalath homozygous and Nipponbare homozygous alleles, respectively. It was also confirmed that this interaction occurs only in the male gametophyte by analysis of the progeny of reciprocal backcrosses between Nipponbare and the Nipponbare/Kasalath F_1 (the F_1 plants were made from Nipponbare fertilized by Kasalath pollen, Fig. I-2). The non-independent segregation was caused by the reduction of transmission of pollen that has the Kasalath allele of marker S11214 at around 45.4 cM on chromosome 1 and the Nipponbare allele of marker S1520 at around 15.8 cM on chromosome 6 (Fig. I-3, Mizuta *et al*., 2007).

To identify the causal genes at each locus, I carried out high-resolution mapping based on the linkage analysis using selected self-pollinated recombinant plants at around the barrier loci from large number of progeny of the hybrid between Kasalath and Nipponbare. I also generated near isogenic lines (NILs) and used for a functional assay of hybrid pollen. My results should provide evidence to how mechanism causes this male gametophytic reproductive isolation.

Materials and Methods

Production of the molecular markers

Japonica cv. Nipponbare genome sequence was completed (International Rice Genome Sequencing Project, 2005) and *indica* cv. 93-11 draft sequence was also determined (Yu *et al.,* 2005); however Kasalath genome sequence was not determined. Thus, polymorphic PCR markers between Nipponbare and Kasalath were designed based on the polymorphisms between Nipponbare and 93-11 genome sequence information. Primer sequences of polymorphic PCR markers for selection and genotyping recombinants are shown in Table I-1.

High-resolution mapping with *DPL1* and *DPL2*

O. sativa ssp. *japonica* cultivar cv. Nipponbare and *O. sativa* ssp. *indica* cv. Kasalath were used as parents for mapping. The F_1 used for this study was produced by hybridizing Nipponbare as female parent to Kasalath as male parent (Kurata *et al.,* 1994). Self-pollinated F_2 and F_3 populations and backcrossed BC_1F_2, BC_1F_3, BC_3F_2 populations were used in this study. BC_1F_2, BC_1F_3, BC_3F_2 backcrossed populations were made by crossing Nipponbare pollen onto F_1, F_2 and F_3 plants. For high-resolution mapping of the reproductive barrier gene on chromosome 1, recombinant plants that had recombination at around the RFLP marker S11214 and that were homozygous for Nipponbare sequence at the interactive locus on chromosome 6 were selected among about 6,000 individuals in BC_1F_2, BC_1F_3, BC_3F_2 progenies of backcross plants. Similarly, for high-resolution mapping of the reproductive barrier gene on chromosome 6, recombinant plants that had

recombination at around the RFLP marker S1520 and that were Kasalath homozygous at the interactive locus on chromosome 1 were selected among about 4,000 individuals in F_2 and F_3 progenies of backcross and F_2 plants.

Genotypes at the barrier loci of the selected recombinants were determined by their progeny test as follows (see also Fig. I-5): when heterozygous genotype of a marker in the vicinity of the barrier segregates as a normal Mendelian ratio in the self-pollinated progenies, the genotype of the barrier is homozygous; when segregation of the marker genotypes in the progeny showed abnormality, the genotype of the barrier is heterozygous.

Total crude genomic DNA was extracted from each seedling using the method as described in Neff *et al.,* (1998) or NaOH-Tris method (Collard *et al.,* 2007). PCR for selecting recombinants was performed with plant crude DNA and Ex Taq polymerase (Takara, Otus, Japan) following manufacturer's protocol. Rice genome annotation and sequence information were referred to in order to identify the causal genes within each candidate region (http://rice.plantbiology.msu.edu/cgi-bin/gbrowse/rice/). The Kasalath genome sequence for the candidate region was kindly provided from T. Matsumoto, the National Institute of Agrobiological Sciences, Tsukuba, Japan (NIAS).

DPL1-N, DPL2-N **and** ***DPL2-K*** **full-length cDNA cloning**

To determine the full-length cDNA sequences of the *DPL*s in Nipponbare and Kasalath, total RNA was extracted from Nipponbare and Kasalath anthers with the Concert Plant RNA Reagent (Invitrogen, Carlsbad, CA). Each 1 μg intact polyA-mRNA was isolated from total RNA of

Nipponbare and Kasalath using Dynabeads Oligo $(dT)_{25}$ (Invitrogen, Carlsbad, CA). The 5' and 3' rapid amplification of cDNA ends (RACE) experiments were performed by the SMART™ RACE cDNA Amplification Kit (Clontech, Mountain View, CA) with the mRNA. PCR products were cloned into a pCR2.1 plasmid with TOPO TA Cloning Kit (Invitrogen, Carlsbad, CA). Their sequences were determined using an ABI 3130xl auto sequencer (Applied Biosystems, Foster City, CA) according to the instruction protocol. The determined sequences are available from DDBJ (accession numbers AB535626 to AB535628).

Complementation test of *DPL1* and *DPL2* genes

Backcrossing Nipponbare up to BC_4 generation and near-isogenic lines (NILs) that carried homozygous *DPL1-K* (*DPL1-K*/*DPL1-K*) together with heterozygous *DPL2* *(DPL2-K/DPL2-N)* were chosen to test for complementation by the *DPL1-N* allele (Fig. I-13A). Similarly, the NILs carrying heterozygous *DPL1* (*DPL1-K/DPL1-N)* together with homozygous *DPL2-N (DPL2-N/DPL2-N)* were selected to test for complementation with the *DPL2-K* allele (Fig. I-13B). Kasalath and Nipponbare BAC clones including *DPL1* and *DPL2* genes were obtained from NIAS DNA Bank (http://www.dna.affrc.go.jp/) and BAC/EST Resource Center at the Clemson University Genomics Institute (http://www.genome.clemson.edu/), respectively. A XhoI-EcoRI genomic DNA fragment of *DPL1-N* containing a region extending from 4.4 kb upstream of the transcription start site to 1.6 kb downstream of the polyadenylation site and a SacII-SpeI *DPL2-K* containing region extending from 2.5 kb upstream to 1.7 kb downstream were isolated from the BAC clones. These fragments were subcloned into pBluescript SK- (Stratagene, La Jolla, CA) and

inserted into the XhoI-SmaI or SpeI-SmaI site of the binary vector (pBGH1) described by Ito *et al.*, (2008), respectively.

The resultant binary vector plasmids were introduced into *Agrobacterium tumefaciens* EHA101 by electroporation. *Agrobacterium*-mediated transformation was referred as described by Nishimura *et al*. (2007) with following modified procedures. The composition of following culture media (N6D, AAM, 2N6-AS, N6D-S and MS-NK) referred to Nishimura *et al*. (2007). For induction of calli, mature seed without rice husker was sterilized in 70% ethanol for 1 min, then in 2.5 % sodium hypochlorite (Wako, Osaka, Japan) for 15 min by two times with vigorous shaking, followed by ten rinses in sterile water. Sterilized seed was placed on the surface of N6D medium and incubated at 32 ℃ for 3-4 weeks. Genomic DNA of all calli was isolated from each of the partial growing calli (about 1-3 mm) by the Automated DNA isolation system (NA-2000, Kurabo, Osaka, Japan). The genotype of each callus was checked by PCR using gene specific primers (Table I-1). After incubation, cultured calli ware collected and suspended in AAM medium containing pre-suspended *Agrobacterium* with gently mixed for 1.5 min. After removal of excess AAM medium using sterilized stainless-steel sieves and papers, infected calli were transferred onto 2N6-AS medium that covered a filter paper containing 500 μl AAM medium. Plate dish was wrapped with aluminum foil and incubate at 25 ℃ in the dark condition for three days. Infected calli was collected in a 50 ml sterile tube and wash with sterile water containing 500 mg/l cefotaxime by five times and sterile water by five times. After remove excess water with sterilized paper, calli were transferred onto N6D-S medium and incubated at 32 ℃ for 3-4 weeks. Growing calli were transferred to MS-NK medium and incubated at 32℃ for 3-4 weeks. After shoot

regeneration, shoots of 0.5-2 cm in length to MS-HF medium in plant boxes. The transgenic plant was transferred to soil pots and grown in an incubation chamber. Genomic DNA of all transgenic plants was isolated from adult leaves by the CTAB method (Murray and Thompson, 1980). The genotype of each transgenic plant and the self-pollinated progeny was checked by PCR using gene specific primers (Table I-1).

To analyze the genotype segregation of the self-pollinated progenies of transgenic plants, genomic DNA was extracted from an endosperm of the mature seed of the self-pollinated progeny. Half of the endosperm of a mature seed was soaked overnight in 400 μl extraction buffer (Tris-HCl 200 mM (pH 7.5), EDTA 25 mM, NaCl 250 mM, SDS 0.5 %), then crushed and well mixed. The samples were then centrifuged for 10 min at 17,400 g. Supernatant (100 μl) was transferred into 1.5 ml tubes and 100 μl TE saturated phenol (Wako, Osaka, Japan) was added. The samples were then centrifuged for 10 min at 17,400 g. After that, 80 μl of the aqueous phase was transferred into 1.5 ml tube and 60 μl isopropanol was added. The samples were centrifuged for 10 min at 17,400 g. The supernatant was then discarded and the pellet was washed with 200 μl 75% ethanol and dried. Finally, the DNA was resuspended in 50 μl of TE Buffer (10 mM Tris HCl, 1 mM EDTA, pH 8.0). The genotype of each transgenic plant and the self-pollinated progeny were checked by PCR using GoTaq Hot Start Polymerase (Promega, Madison, WI) and gene specific primers (Table I-1).

Gene expression analysis

Total RNA was extracted from various rice tissues with the Concert Plant RNA Reagent (Invitrogen, Carlsbad, CA, USA). Tissue samples of shoot and root were prepared from seedlings of

rice at ten days after germination. Adult leaf, mature panicle before anthesis, un-pollinated mature pistil, lemma and palea of un-pollinated mature flowers, and three different three stages of anther were also used. Intact polyA-mRNA was isolated from 1 μg total RNA using Dynabeads Oligo $(dT)_{25}$ (Invitrogen, Carlsbad, CA). cDNA was synthesized using SuperScript III reverse transcriptase (Invitrogen, Carlsbad, CA) according to the manufacturer's protocol. *DPL*s expression was assayed by using 0.5 μl of cDNA for reverse transcription polymerase chain reaction (RT-PCR) analysis with gene specific primers. *Actin* was used as internal control. PCR was carried out in a 20 μl reaction mixture containing 0.5 μl of cDNA, 0. 5U Ex Taq polymerase (TaKaRa, Otsu, Japan) and 0.2 μM primers. PCR was carried out with 30 cycles of denaturation at 94°C for 30 s, annealing at 60°C for 30 s, and extension at 72°C for 30 s followed by 5 min incubation at 72°C. 2 μl of the PCR products were electrophoresed on a 2.0% agarose gel.

For quantitative expression analysis of *DPL*s, 20 ng of cDNA was used with SYBR Premix Ex Taq (Takara, Otsu, Japan). Data were collected using the TaKaRa Thermal Cycler Dice Real Time System in accordance with the instruction manual. Three independently extracted RNA samples were used for quantification. The expression value of ubiquitin was used for normalization of gene expression levels. Gene specific primers are listed in Table I-1.

Microarray experiments

The definition of each stage was described in Notes on Appendix. Collected biological samples were frozen in liquid nitrogen and kept -80 degrees until RNA extraction. Total RNA was extracted with RNeasy plant mini kit (Qiagen, Hilden, Germany) according to manufacturers

protocol. 500 ng total RNA was used for cRNA probe labeling and 3.3 μg labeled cRNA was used for microarray hybridization. Affymetrix rice genome arrays constructed with 57,381 genes were used for transcriptome analysis. The probe sets of Affymetrix Rice Genome Array were designed on 2004, based on NCBI UniGene Build #52, (May, 2004), GenBank mRNAs (July, 2004) and gene predictions from TIGR's osa1 version 2.0 release (April, 2004). More than triplicate experiments with independently collected samples for all stages were performed according to instruction protocol with GeneChip Hybridization Oven 640, Fluidics Station 450 and Scanner 3000 7G.

***In situ* hybridization experiment**

Nipponbare and Kasalath anthers at bi-cellular stage were fixed in formalin-acetic acid-alcohol fixative solution (formalin : glacial acetic acid: 50% ethanol = 5 : 5 : 90) for 24 h at 4°C, and then dehydrated in an ethanol series from 70% to 100%. The ethanol was replaced with xylene and the samples were embedded in Paraplast Plus (McCormick Scientific, St. Louis, MO). Paraffin sections (8 μm thick) were applied to slide glasses Matsunami Adhesive Slide (MAS; Matsunami Glass, Kishiwada, Japan). A truncated *DPL1-N* cDNA fragment was amplified by PCR using the gene specific primers (RT1-1f and GSP1-4, Table I-1) and cloned into pCR II-TOPO plasmid (Invitrogen, Carlsbad, CA). Digoxigenin-labeled antisense and sense RNA probes were produced by *in vitro* transcription with Sp6 and T7 RNA polymerases and DIG-RNA labeling mix following the manufacturer's protocol (Roche, Indianapolis, IN). Probe hybridization, post-hybridization washing, and immunological detection were performed following the method of Kouchi *et al*. (Kouchi and Hara, 1993).

Examination of pollen viability and pollen germinability

NILs at BC_3 generation were produced by repeatedly backcrossing to the parent, Nipponbare. NILs showing heterozygous *DPL1* (*DPL1-K/DPL1-N)* and homozygous *DPL2-N* (*DPL2-N/DPL2-N)* with Nipponbare background were selected by PCR with specific primers in Table I-1. Pollen viability was investigated with spikelets collected from panicles before flowering. They were fixed in an EAA (100% ethanol : acetic acid = 3 : 1) solution overnight at room temperature. To assess pollen maturation stage, pollen grains were spread on a slide glass and stained with a solution of 4', 6'-diamidin-2 phenylindole (DAPI). Their morphology was observed under a fluorescence microscope. The pollen fertility was examined with I_2KI staining (0.2% iodine and 2% potassium iodine) under bright field microscope (OLYMPUS, Tokyo, Japan). To analyze pollen germination, pollen grains from dehisced anthers of Nipponbare, Kasalath and NILs were immediately placed on a germination medium (1 mg H_3BO_3, 3 mg $Ca(NO_3)_2$, 1.7 g sucrose and 0.07 g agarose in 10 ml sterilized water) and were incubated at room temperature in a humid chamber. After 2-3 hours, pollen grains were stained in a 0.005 % aniline blue solution (0.005 % aniline blue in 0.15 M K_2HPO_4, pH 8.6) and counted under a fluorescence microscope (Zeiss, Jena, Germany). Aniline blue is used to stain callose, a component of pollen tubes. Six and four independent analyses of Nipponbare and Kasalath pollen were conducted, while each NIL was examined once. More than 300 pollen grains were observed in each examination.

Polyclonal antibody production and immunoblot analysis

Two peptides antigens of the residues 1-17 (MGSEDTKDMLKNVDWKTC) and 24-40

(TDPSQPVVKKRLPKKIRC) of DPL1-N common to all other DPLs of Nipponbare and Kasalath were synthesized and used for immunizing rabbits (Medical & Biological Laboratories Co., Ltd, Nagoya, Japan). The immune-serum was affinity purified using these peptide antigens to obtain monospecific immune-serum. Total protein preparation and Western blot detection of DPLs were conducted following the method (Harlow and Lane, 1988). To prepare recombinant DPL1-N protein, the 95 amino acid coding region was fused to a hexahistidine tag at the N-terminal. The construct was cloned into the expression vector pDEST17 (Invitrogen, Carlsbad, CA) and expressed in the BL21-AI *Escherichia coli* strain (Invitrogen, Carlsbad, CA). This recombinant protein was used as a positive control for immunoblot analysis. Recombinant protein and crude extracts of anther were electrophoresed on a 16% SDS-polyacrylamide gel following the Tris-Tricine-SDS method (Schagger, 2006). After electroblotting the gel onto a nitrocellulose membrane, the filter was immuno-stained according the methods (Harlow and Lane, 1988).

Search of the similar predicted TE in the Nipponbare genome

Based on the sequence information, a 518 bp inserted sequence in the Kasalath allele *DPL1* was prediceted as transporsable element (TE). To verify this prediction, the number of candidate TE was sought in the Nipponbare genome. Candidate TE having TIRs (each 12 bp) at its both ends were sought with Electronic PCR program (e-PCR, http://www.ncbi.nlm.nih.gov/projects/e-pcr/). One-base mismatch in each both TIR was allowed for detection the candidate TE. The result was kindaly supported by T. Mochizuki, Plant Genetics, National Institute of Genetics, Japan.

Results

High resolution mapping of the *DPL*s

Firstly, I developed PCR markers, i27556 (36.9 cM) and i18100 (50.8cM) on chromosome 1 for the high-resolution mapping of the causal gene on chromosome 1. Twenty individuals that have recombination between these two markers with Nipponbare homozygous allele at the interactive barrier locus on chromosome 6 were selected from 474 plants of BC_1F_1 and 294 plants of F_2 populations. The genotypes of the twenty recombinant lines at the barrier locus were determined by their progeny analysis of chromosome 1 (see Materials and Methods). The genotypes of the 20 recombinant plants determined by newly developed internal PCR markers (Table I-1) showed that the barrier locus was mapped at 45.4 cM between markers i3068_2 and i7872 spanning 220 kb on chromosome 1 (Fig. I-4). Markers i12296, i6342_2 and i267_2 were co-segregated with the barrier locus on chromosome 1.

For fine mapping of the causal gene on the interactive locus of chromosome 6, 16 individuals that have recombination between PCR markers, i2121 (10.4 cM) and i4790 (32.7 cM) on chromosome 6 with Kasalath homozygous allele of chromosome 1 were selected from 294 F_2 population. Genotypes of the 16 lines with internal 9 PCR markers and progeny (F_3) analysis showed that the barrier locus was mapped at 15.8 cM located between markers i2121 (10.4 cM) and i4099 (18.0 cM) spanning 2.5Mb on chromosome 6 (see Materials and Methods). Markers i3635, i316, i1404_2 and i11601_2 were co-segregated with the male gametophytic barrier locus on chromosome 6 (Fig. 1-4).

Finally, fine mapping using more than 10,000 self-pollinated and backcrossed progeny indicated that candidate genes were located within regions of 59 kb on Nipponbare chromosome 1 and 11 kb on Nipponbare chromosomes 6, including nine and two predicted genes, respectively (Fig. I-5 and Fig. I-6). A pair of genes, one from each region shared a high degree of homology with each other (98% identity in the amino acid level, Fig. I-6). Additionally, both genes have different sequences between Nipponbare and Kasalath. I regarded these homologous genes as primary candidates, and these were designated *LOC_Os01g15448* on chromosome 1 as *DOPPELGANGER1* (*DPL1*) and *LOC_Os06g08510* on chromosome 6 as *DOPPELGANGER2* (*DPL2*), respectively.

Gene structure and expression of the *DPL*s

To determine the full-length cDNA sequences of the *DPL*s in Nipponbare and Kasalath, 5' and 3' rapid amplification of cDNA ends experiments were performed. Sequence analysis of the Nipponbare and Kasalath genomes and their transcripts suggested that alleles on Nipponbare chromosome 1 (*DPL1-N*) and Kasalath chromosome 6 (*DPL2-K*) had the same coding sequence structure (Fig. I-7). In contrast, alleles on Kasalath chromosome 1 (*DPL1-K*) and Nipponbare chromosome 6 (*DPL2-N*) had structural differences from the above two alleles. *DPL1-K* had a 518 bp insertion in the predicted coding sequence and the transcript could not be detected in any tissues (Fig. I-8A). The inserted fragment had diagnostic terminal inverted repeats (TIR) and was widely dispersed in Nipponbare genome (Fig. I-9). This terminal inverted repeats has homology to the transposable elements (TE) in other plants,

Ips-r in pea and *Ac*/*Ds* in maize (Bhattacharyya, *et al*., 1990). These findings suggested that this inserted fragment is a TE.

The transcript of *DPL2-N* was a read through product of the second intron generating a premature stop codon (Fig. I-7 and I-8A). Transcripts of *DPL1-N, DPL2-K* and *DPL2-N* were detected in all tissues examined in both quantitative RT-PCR and microarray experiments (Fig. I-10 and Fig. I-11). Relatively higher expression was observed in anther especially at the bi- cellular (Fig. I-11) and tri-cellular pollen stages (Fig. I-10). Higher expression of *DPL*s in pollen was also observed in *in situ* hybridization experiments (Fig. I-8B). Anti-DPL antibodies recognized small proteins of 10.3 and 10.4 kDa for DPL1-N and DPL2-K in extracts from Nipponbare and Kasalath mature anthers, respectively (Fig. I-8C). However, DPL2-N protein (12.6 kDa) was not detected in extracts from Nipponbare. The lack of *DPL1-K* transcript and the absence of DPL2-N protein suggested that both *DPL1-K* and *DPL2-N* were loss of function alleles.

Pollen maturation and germinability of the hybrid pollen

Near isogenic lines (NILs) having heterozygous *DPL1* (*DPL1-K*/*DPL1-N*) in Nipponbare background (*DPL2-N*/*DPL2-N*) were generated and used for a functional assay of pollen. By examining the transmission rate of pollen to the progeny of self-pollinated plants, it could be inferred that hybrid pollen possessing *DPL1-K* and *DPL2-N* together was not transmitted to the next generation. Therefore, half of the pollen grains from the NILs possessing both *DPL1-K*/*DPL1-N* and *DPL2-N*/*DPL2-N* should have some defect in function.

However, mature pollen of the NILs did not show any defect in pollen shape, number of nuclei, or starch accumulation (Fig. I-12, A and B).

Examination of pollen germination activity was then carried out *in vitro* revealing that about half of the NILs pollen grains could not germinate, whereas almost all pollen grains of Nipponbare and Kasalath germinated and pollen tubes were elongated (Fig. I-12, C and D). This suggests that the presence of the two non-functional alleles, *DPL1-K* and *DPL2-N* in the hybrid pollen causes a defect in pollen germination due to the loss of *DPL* function.

Complementation test of the *DPL* genes

Assuming that both *DPL1-K* and *DPL2-N* are loss of function alleles, pollen having these two alleles together should be rescued by transformation with a functional allele, *DPL1-N* or *DPL2-K*. When *DPL1-N* was introduced into NILs possessing both *DPL1-K*/*DPL1-K* and *DPL2-K*/*DPL2-N* (Fig. I-13A), progeny of self pollinations were obtained carrying homozygous *DPL1-K* and homozygous *DPL2-N,* generated by rescued pollens possessing both *DPL1-K* and *DPL2-N* alleles (Table I-2). All of these *DPL1-K*/*DPL1-K*; *DPL2-N*/*DPL2-N* plants contained the *DPL1-N* transgene with no exception, while no progeny having the same allele combination were produced by the NILs transformed with vector only. Similarly, when NILs possessing both *DPL1-K*/*DPL1-N* and *DPL2-N*/*DPL2-N* were transformed with *DPL2-K* (Fig. I-13B), those progeny that had homozygous *DPL1-K* and homozygous *DPL2-N* always contained transgenic *DPL2-K* (Table I-2). These results clearly showed that a functional *DPL1-N* or *DPL2-K* allele is essential for pollen transmission. In conclusion, *DPL1-K* and *DPL2-N* are loss of function alleles that act as a pair of reproductive barrier genes by their combination in hybrid pollen.

Discussion

In this study, I identified and characterized hybrid incompatible genes, *DOPPELGANGER1* (*DPL1*) and *DOPPELGANGER2* (*DPL2*) in an intersubspecific cross of rice, which were disrupted in each subspecies independently. The overall structural similarity and functional conservation of *DPL1* and *DPL2* indicated that they are paralogous genes.

Two pioneers of fly genetics, Dobzhansky and Muller proposed how various reproductive isolations occurred on the basis of the epistatic interaction of two or more genes that have evolved independently in two related species (Dobzhansky, 1937; Orr and Coyne, 2004; López-Fernández, 2007). Dobzhansky-Muller model results from deleterious epistatic interactions between two loci (*A* and *B*) from different parental genomes. When a species of genotype *AABB* is subdivided into two different populations, new mutations independently arise in each population (*e. g.*, *AaBB* or *AABb*), and may eventually fix yielding genetically divergent groups (*e. g.*, *aaBB* and *AAbb*). If the populations subsequently hybridize (*AaBb*), interaction between derived alleles *a* and *b* in the hybrids could be deleterious because they evolved in different genetic backgrounds and their compatibility has not been tested by natural selection.

Despite its simplicity, only a few genes explaining about Dobzhansky-Muller model have been identified due to the following reasons (Coyne and Orr, 2004); (I) the model does not require that the *A* and *B* alleles have a devastating effect on hybrids. A hybrid incompatibility might be caused only resulting from the accumulative effects of many ‘weak’ genetic incompatibility (II) hybrid incompatibility might be caused by the inappropriate alleles at not only two alleles but three

or more loci. Multi-locus interactions cannot generally be detected via single and two-locus analyses because statistical limitation and genetic complexity (III) both substitutions might occur, for example, in one population and none in the other (for example, one population evolved *AABB* → *AAbb* → *aabb*, while the other population remains *AABB*).

In plants, recent two researches, hybrid necrosis and hybrid embryo lethality in the intraspecific crosses of *Arabidopsis* were identified as Dobzhansky-Muller type hybrid incompatibility (Bomblies *et al.*, 2007; Bikard *et al.*, 2009). Likewise, after gene duplication of *DPL*, an ancestral population seems to have diverged forming the Kasalath ancestral population, which subsequently lost the function of *DPL1* by TE insertion, and the Nipponbare ancestral population, which lost the function of *DPL2* by means of a mutation generating a splicing defect. When they met again by crossing, hybrid pollen having the loss of function alleles together became non-functional and failed to transmit themselves to the next generation (Fig. I-14). This is a typical case of the Dobzhansky-Muller model for barrier formation by genetic incompatibility between species, like Bomblies *et al.* (2007) and Bikard *et al.* (2009).

Hybrid pollen carrying two defective alleles of *DPL*s together became non-functional, and did not germinate (Fig. I-12). Complementation tests of *DPL*s using NILs also indicated that functional *DPL*s are essential for pollen transmission in rice (Table I-2). Plant fertilization has a unique mechanism that has evolved in angiosperms (Cresti *et al.*, 1999). This process begins when a pollen grain adheres to the stigma of the pistil (female reproductive structure), germinates, and grows a long pollen tube. While this pollen tube is growing, a haploid generative cell travels down the tube behind the tube nucleus. Thus, proper pollen maturation and pollen germination are

necessary to succeed plant reproduction. By the study of mutant analysis and by the gene tagging strategies, several male gametophytic mutants that affect pollen germination have been isolated (Han *et al.*, 2006; Cole *et al.*, 2005; Golovkin and Reddy, 2003). However little is known about plant male gamete formation and transmission.

The expressions of *DPL1*s in anther are up-regulated during the late stage of pollen development in bi-cellular stage (Fig. I-11) or tri-cellular stage (Fig. I-10). This difference might derive from a dispersion of expression of *DPL* genes in tri-cellular stage, or cross-hybridization of some probes with both *DPL1-N* and *DPL-2-N* for microarray analysis. It is interest that *DPL*s are expressed in all examined rice tissues, but their defects impair only male gametophyte, pollen. The expression of *DPL1*s in anther at late pollen developmental stage is higher than other tissues and stages suggesting that *DPL*s have important role for male gamete transmission.

It seems worthwhile to elucidate that the role of *DPL*s during pollen development and germination contributes to the molecular basis of plant reproduction. Further studies about molecular function of *DPL*s may provide new information for plant reproduction.

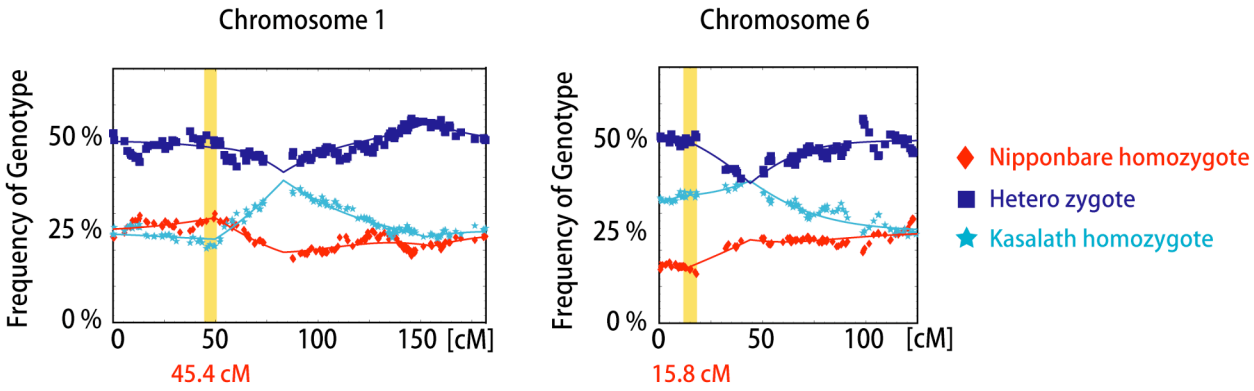


Fig. I-1. Frequencies of each allele of RFLP markers and the best-fitted regression curves in the F_2 population of Nipponbare and Kasalath that modified from Harushima *et al.* (2001). Each allele frequency is plotted along the genetic linkage map. For each chromosome, the left and right ends correspond to the short and long arm ends of the genetic linkage map (Harushima *et al.* 1998). The frequency of Nipponbare homozygous genotypes (red diamonds), Kasalath homozygous genotypes (sky blue stars), and heterozygous genotypes (blue squares) of the individual markers that mapped at different locations are plotted. The best-fitted regression curves for each allele frequency on the chromosome are also presented in the respective colors. The yellow regions indicate a pair of interactive barriers.

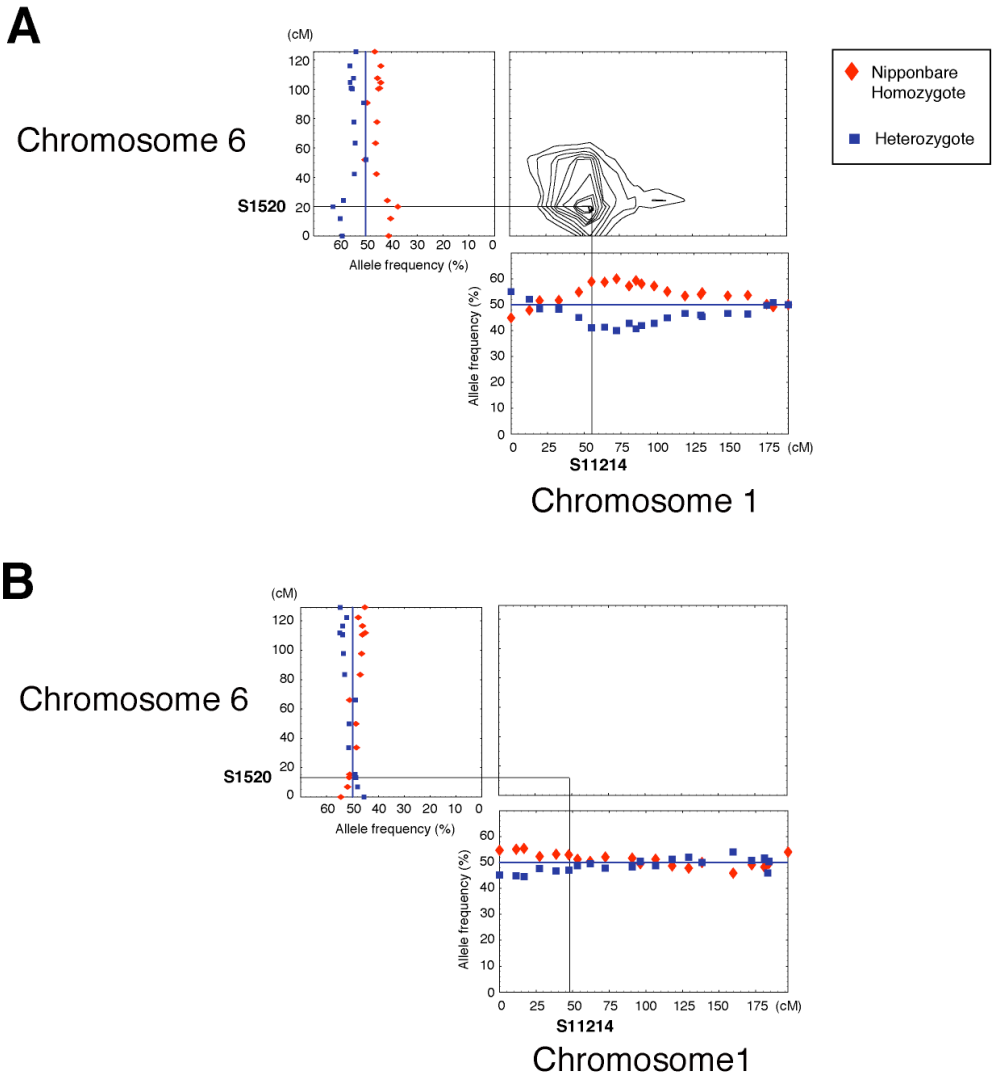


Fig. I-2. Chi-square tests for independence of markers between chromosome 1 and 6, and marker allele frequencies in reciprocal backcross populations. The population (Nipponbare//Nipponbare/Kasalath, 236 plants) was made by crossing F_1 plants pollen onto Nipponbare (A), and 234 plants of Nipponbare/Kasalath//Nipponbare population that was made by reciprocal way (B). Chi-square values were represented as contours from 7.5 to 45 by 2.5. The maximum chi-square value, 39.6, was observed between RFLP markers S11214 on chromosome 1 and S1520 on chromosome 6 in Nipponbare//Nipponbare/Kasalath (A), whereas no interaction between S11214 and S1520 was observed in Nipponbare/Kasalath//Nipponbare (B). These two RFLP markers referred to Harushima *et al.* (*S2*). Based on the Mendel's law, the allele frequency of each marker was expected as Nipponbare homozygote : Heterozygote =1 : 1 (Blue lines), whereas the reduction of heterozygote at S11214 on chromosome 1 and Nipponbare allele S1520 on chromosome 6 was observed (A). These results indicated the reduction of pollen transmission that has Kasalath allele S11214 on chromosome 1 and Nipponbare allele S1520 on chromosome 6.

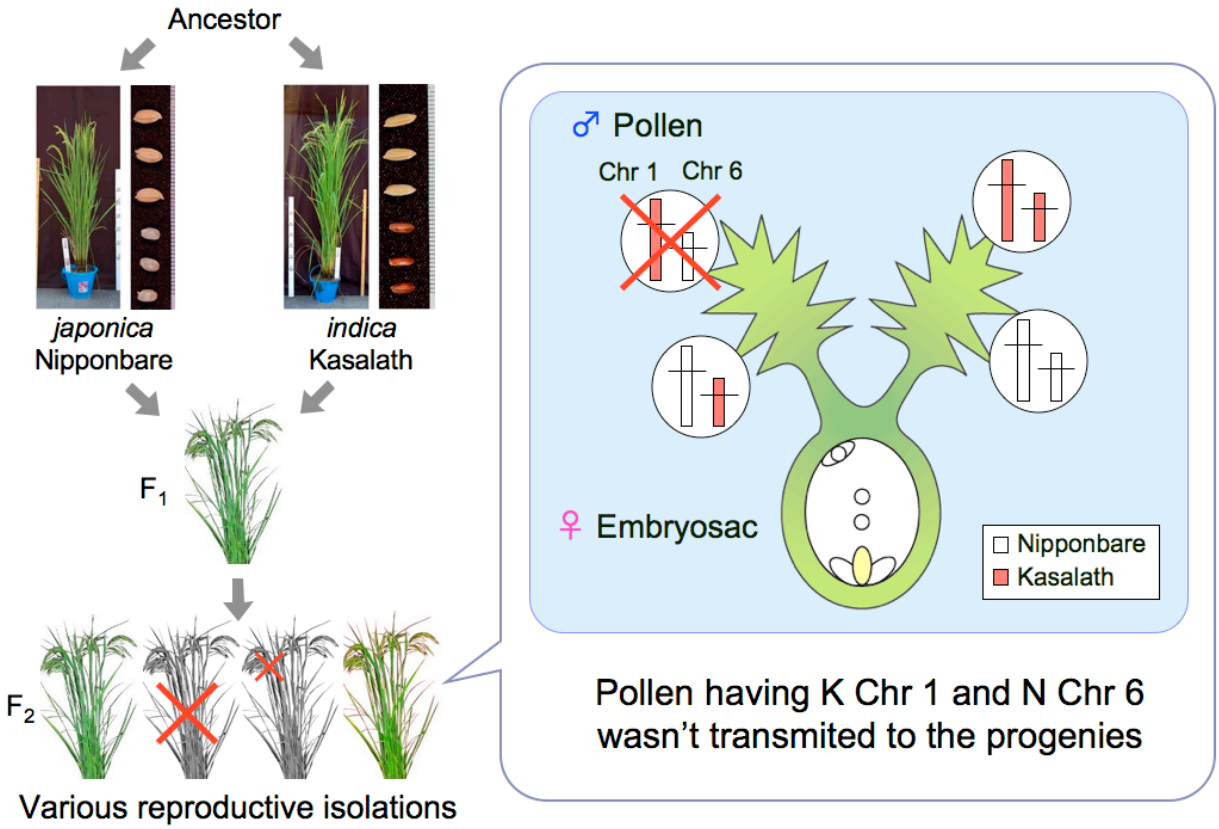


Fig. I-3. The schematic representation of the male gametophytic reproductive isolations caused by the epistatic interaction between Kasalath chromosome 1 and Nipponbare chromosome 6 in the hybrid pollen.

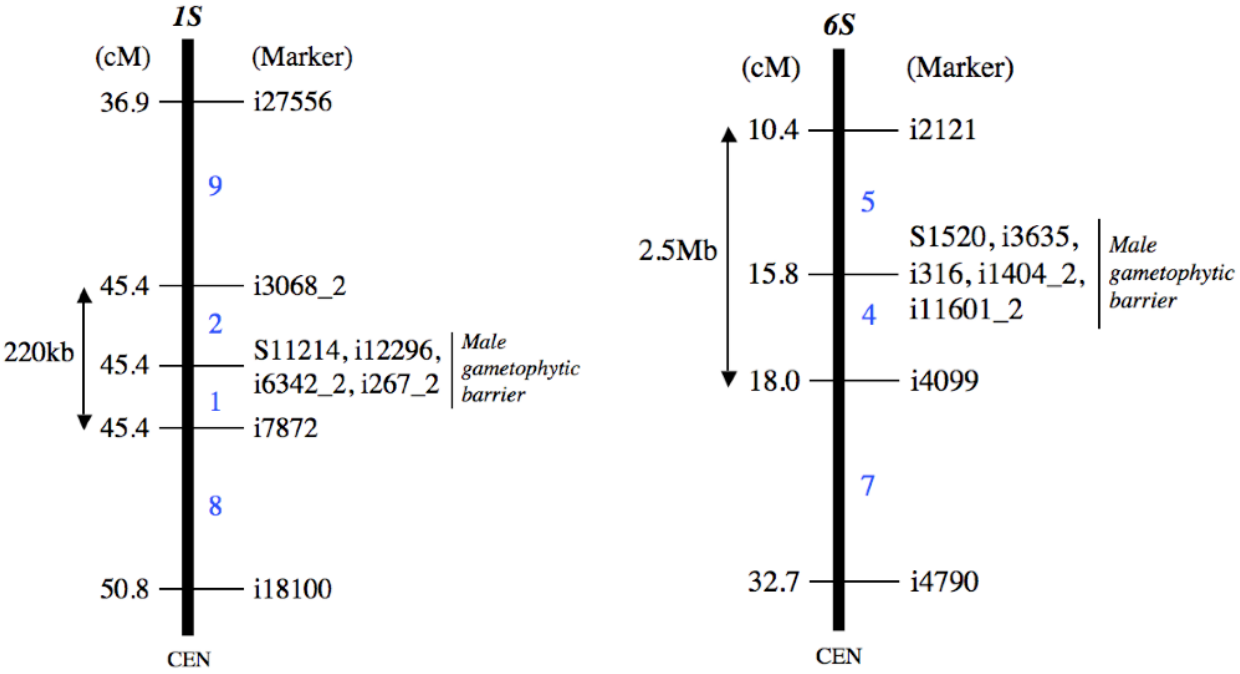


Fig. I-4. Mapping of the male reproductive barrier loci on chromosomes 1 and 6. The blue figures are the number of recombinant individuals. Twenty recombinants that have recombination between PCR markers, i27556 (36.9 cM) and i18100 (50.8 cM) on chromosome 1 with Nipponbare homozygous allele of chromosome 6 were selected from 474 plants of BC_1F_1 and 294 plants of F_2 populations. Sixteen individuals that have recombination between PCR markers, i2121 (10.4 cM) and i4790 (32.7 cM) on chromosome 6 with Kasalath homozygous allele of chromosome 1 were selected from 294 F_2 populations. The markers, S11214 and S1520 referred to Harushima, *et al.* (1998). Primer sequences of each marker are described in Table I-1.

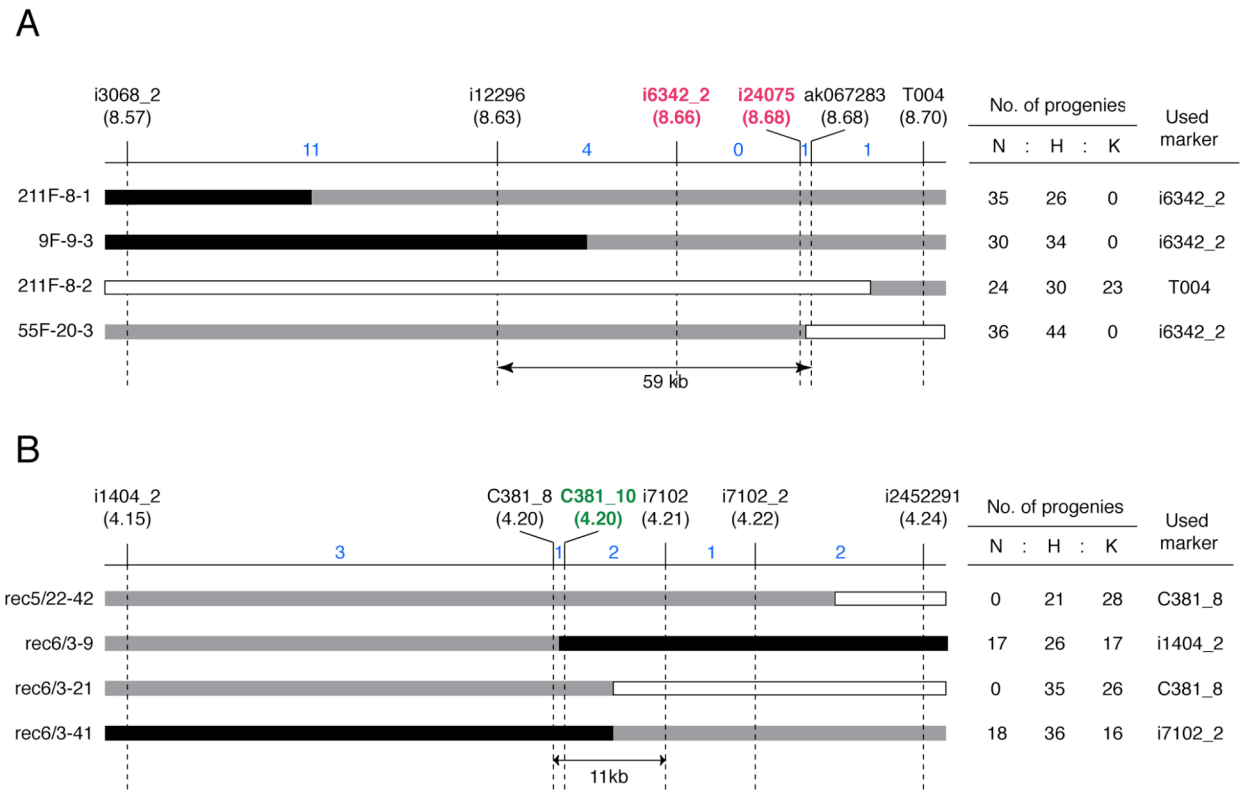


Fig. I-5. The graphical genotypes of the recombinants and segregation of their self-pollinated progenies used for fine mapping. (A) The four plants carrying genetic recombination in the vicinity of the barrier on chromosome 1 are shown. All plants are Nipponbare homozygous chromosome 6. (B) The four plants carrying recombination in the vicinity of the barrier on chromosome 6 were shown. Four recombinants have Kasalath homozygous chromosome 1. Black, gray and white lines show the genotypes of Kasalath homozygous (K), heterozygous (H) and Nipponbare homozygous (N) regions. The segregation of heterozygous markers in their self-progenies is shown in right side. Co-segregated markers with the barrier loci on chromosome 1 and 6 were indicated in magenta and green, respectively. The candidate regions were shown as double-headed arrows below the graphical genotypes. Numbers in blue between dashed lines indicate number of recombinant plants between the selection markers.

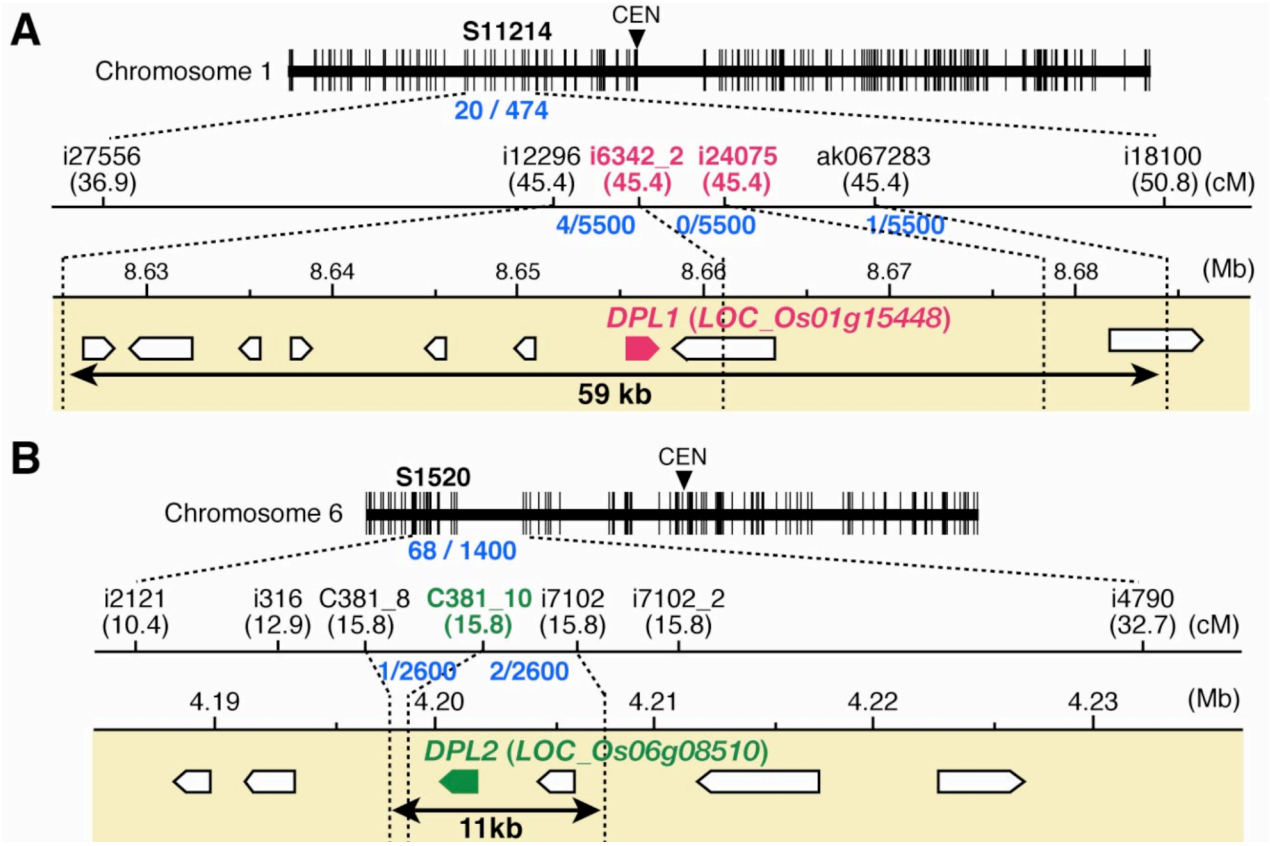


Fig. I-6. Fine mapping of the reproductive barrier genes. The candidate regions were delimited within 59 kb on chromosome 1 (A) and within 11 kb on chromosome 6 (B) of Nipponbare. Numbers in blue between dashed lines indicate number of recombinant plants between the selection markers, and that of the analyzed segregants. Co-segregated markers with the barrier loci and two highly similar genes on chromosome 1 and 6 were indicated in magenta and green, respectively.

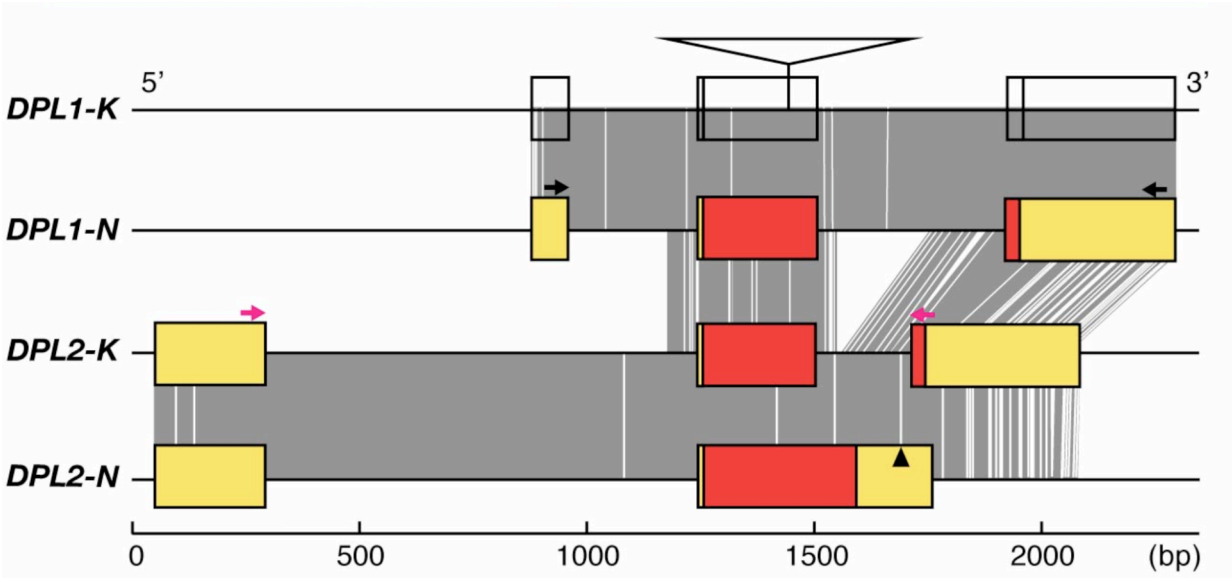


Fig. I-7. Schematic gene structure and alignment of *DPL*s of Nipponbare (N) and Kasalath (K). Each transcript sequence was determined by 5' and 3' RACE experiments. Red and Yellow boxes indicate the coding and untranslated regions, respectively. Gray lines between gene structures show identical bases. The inserted TE to the *DPL1-K* is indicated by triangle. The base change from A to G in the second intron of *DPL2-N* was marked with an arrowhead and was predicted as a loss of branch site for splicing (see Table II-2). Arrows indicate positions of gene specific primers for *DPL1* and *DPL2* used for RT-PCR analysis. The empty boxes of *DPL1-K* were predicted gene structure based on the *DPL1-N*, due to the absence of *DPL1-K* transcript in any tissues.

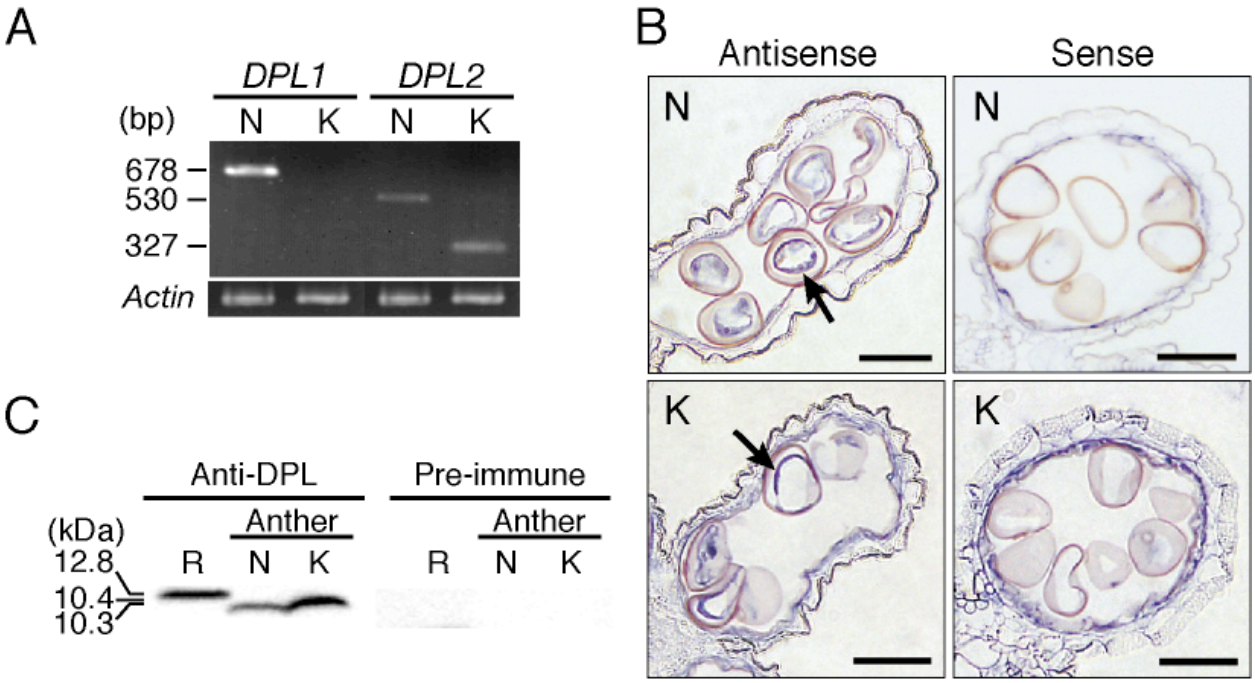


Fig. I-8. Expression analysis of *DPL*s. (A) RT-PCR of *DPL* expression in mature anther of Nipponbare and Kasalath. Gene specific primers are shown in Fig. I-7. (B) RNA *in situ* hybridization analysis of *DPL*s in anther at bi-cellular stage. Transverse sections of Nipponbare (N) and Kasalath (K) anthers were hybridized with *DPL1-N* antisense or sense probes. Arrows showed clear expression signal of *DPL*s in the cytoplasm of pollen grains. Bars = 40 μm. (C) Immunoblot analysis of DPL proteins. Predicted molecular weights of DPL1-N, DPL2-K and an N-terminal hexahistidine tag recombinant protein are 10.3, 10.4 and 12.8 kDa, respectively. The lane indicated by R, N and K are recombinant protein, extracts from mature anther of Nipponbare, and extracts from mature anther of Kasalath, respectively. The crude extracts were extracted from 30 anthers of Nipponbare and Kasalath, respectively.

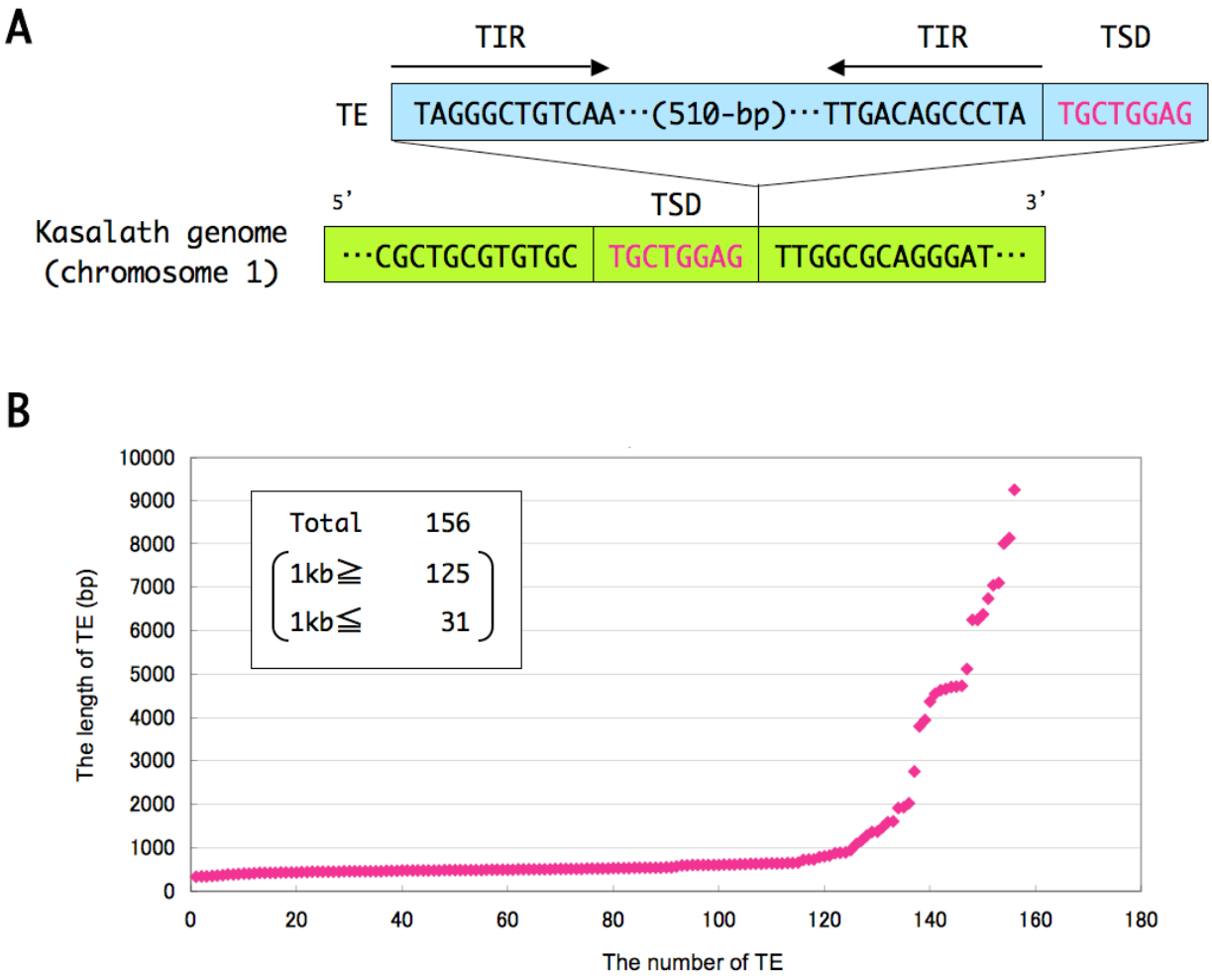


Fig. I-9. The predicted transposable element (TE) of Kasalath allele *DPL1* in Nipponbare genome. (A) The struture of the predicted TE in the *DPL1* gene on Kasalath genome. Arrows indicate the predicted terminal inverted repeats (TIR, 12 bp). Magenta sequences are predicted as target site duplication (TSD, 8 bp). (B) The number and length of candidate TE in the Nipponbare genome. Each candidate TE is aligned from left to right according to the length of sequence. Candidate TE having TIRs at its both ends were sought by e-PCR (see Materials and Methods). In total, 156 candidate TEs were detected.

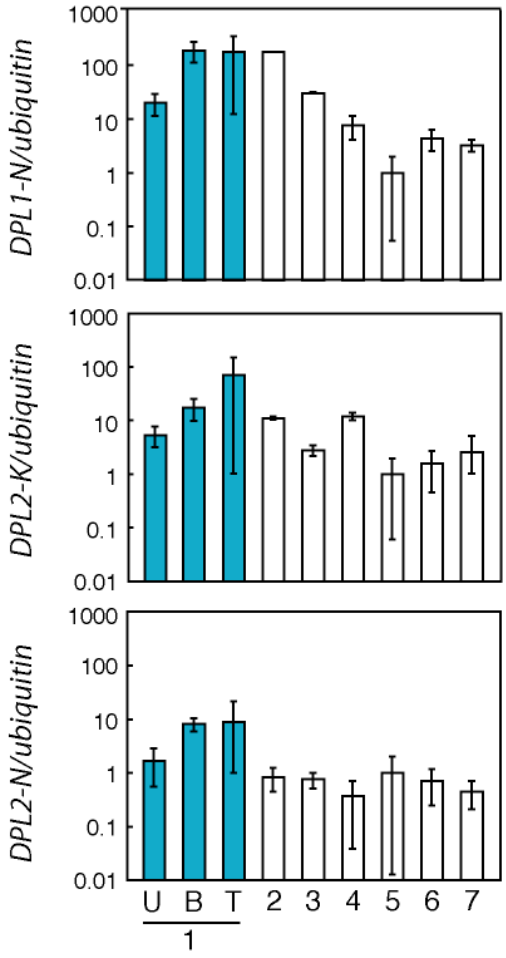


Fig. I-10. Expression profiles of *DPL1-N, DPL2-K* and *DPL2-N* in various tissues of Nipponbare and Kasalath by quantitative RT-PCR. Number indicates tissues: 1, anther; 2, un-pollinated mature pistil; 3, lemma and palea of un-pollinated mature flowers; 4, mature panicle before anthesis; 5, leaf; 6, shoot; 7, root. Blue bars represent expression level in anthers at three different stages of pollen: U, uni-nucleus stage; B, bi-cellular stage; T, tri-cellular stage. Error bars indicate standard deviations.

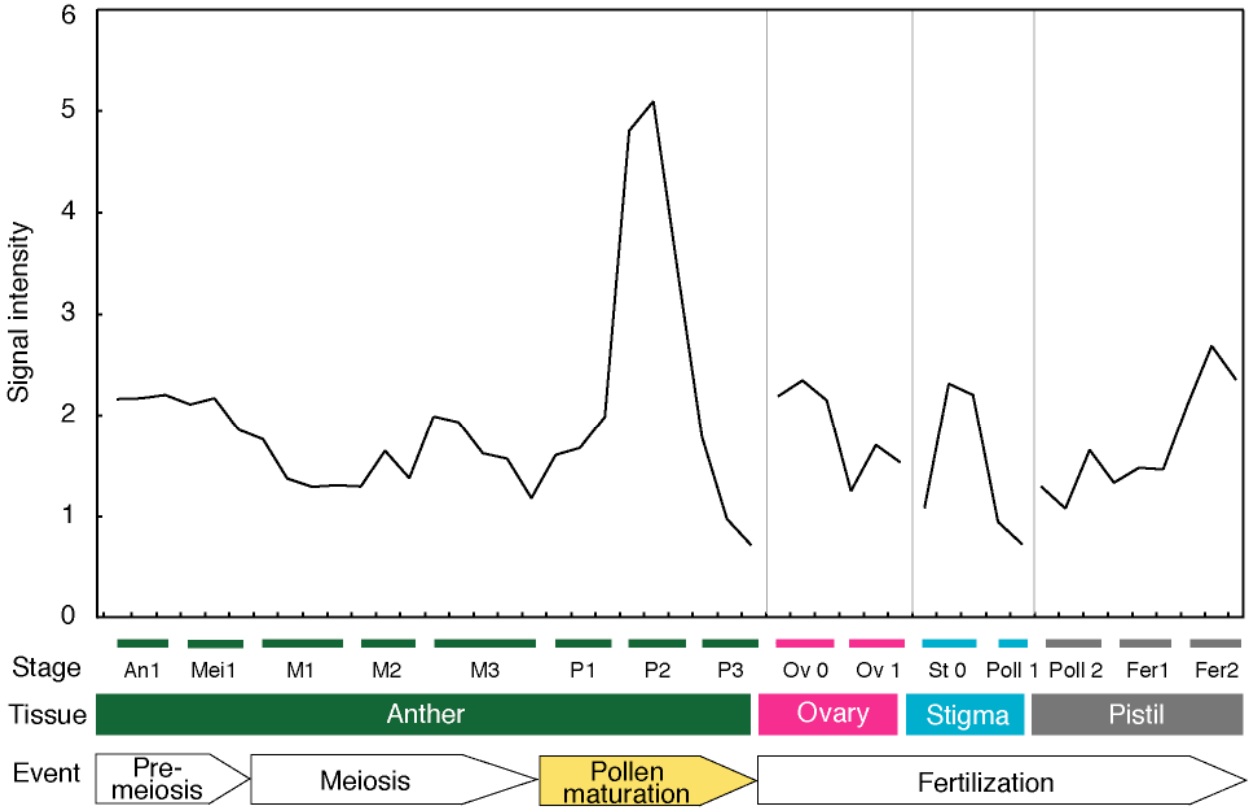


Fig. I-11. The expression pattern of *DPL1-N* and/or *DPL-2-N*. Some probes in a both probe set of *DPL1-N* and *DPL-2-N* cross-hybridized with both *DPL1-N* and *DPL-2-N*. Thus, the signal intensity indicates *DPL1-N* and/or *DPL-2-N* normalized signal intensities. Expression levels remained low during An1 and P1 in anther and ovary, stigma and pistil. The high expression was observed during the pollen maturation in the bi-cellular (P2) stage. Stage symbols and definitions referred to Notes and Appendix.

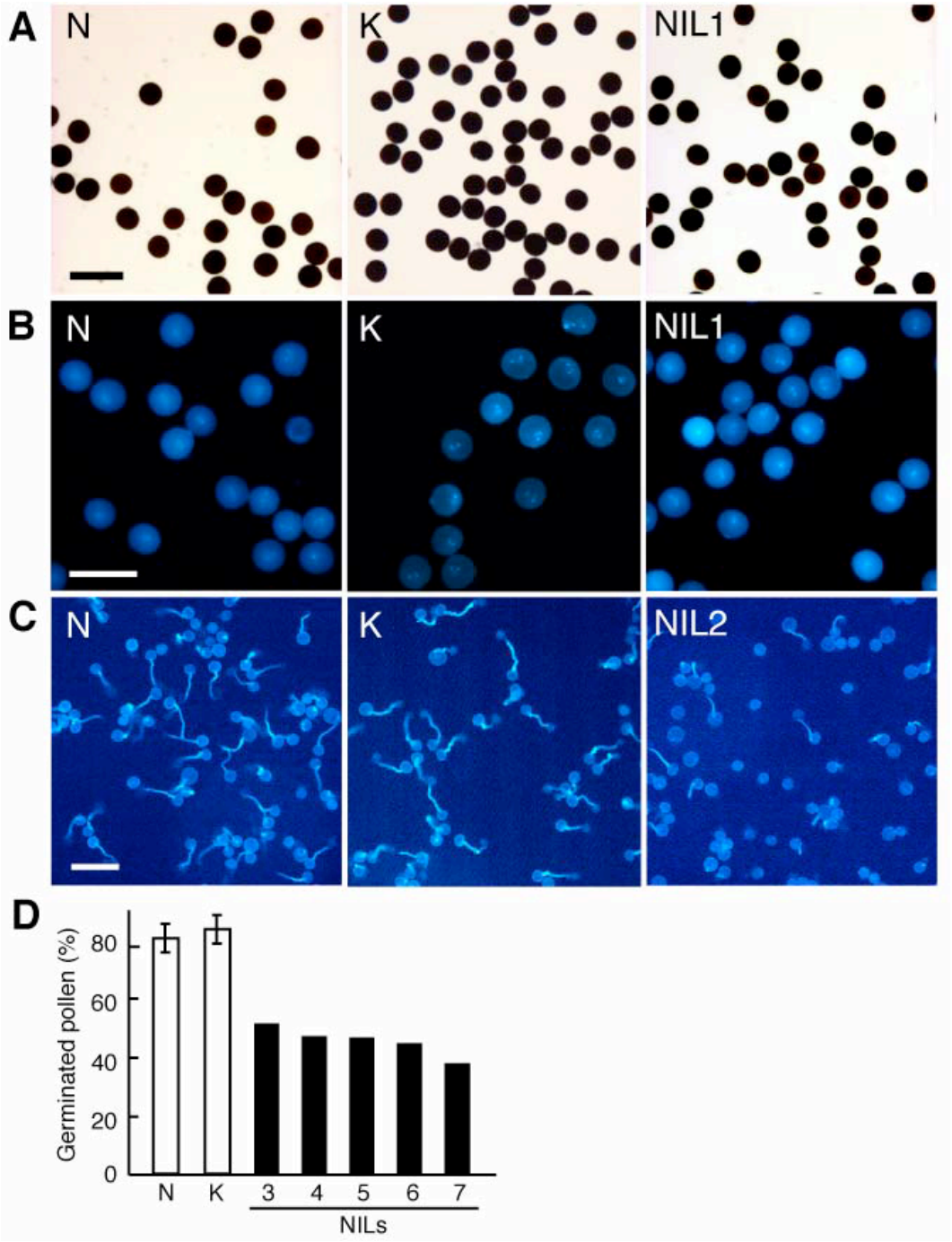


Fig. I-12. Pollen phenotypes and pollen tube germination of Nipponbare (N), Kasalath (K) and near-isogenic lines (NILs). All NILs has *DPL1-K*/*DPL1-N* in Nipponbare background (*DPL2-N*/*DPL2-N*). Mature pollen was stained with I_2KI solution (A) and DAPI (B). (C) Pollen germination of Nipponbare, Kasalath and NILs *in vitro*. Germinated pollen tubes were visible by callose fluorescence stained with aniline blue. Bars = 80 μm (A and B) and 160 μm (C). (D) Pollen germination rate of Nipponbare and Kasalath (white bars) and five NILs (black bar) on the pollen germination solid medium. Error bars indicate standard deviations.

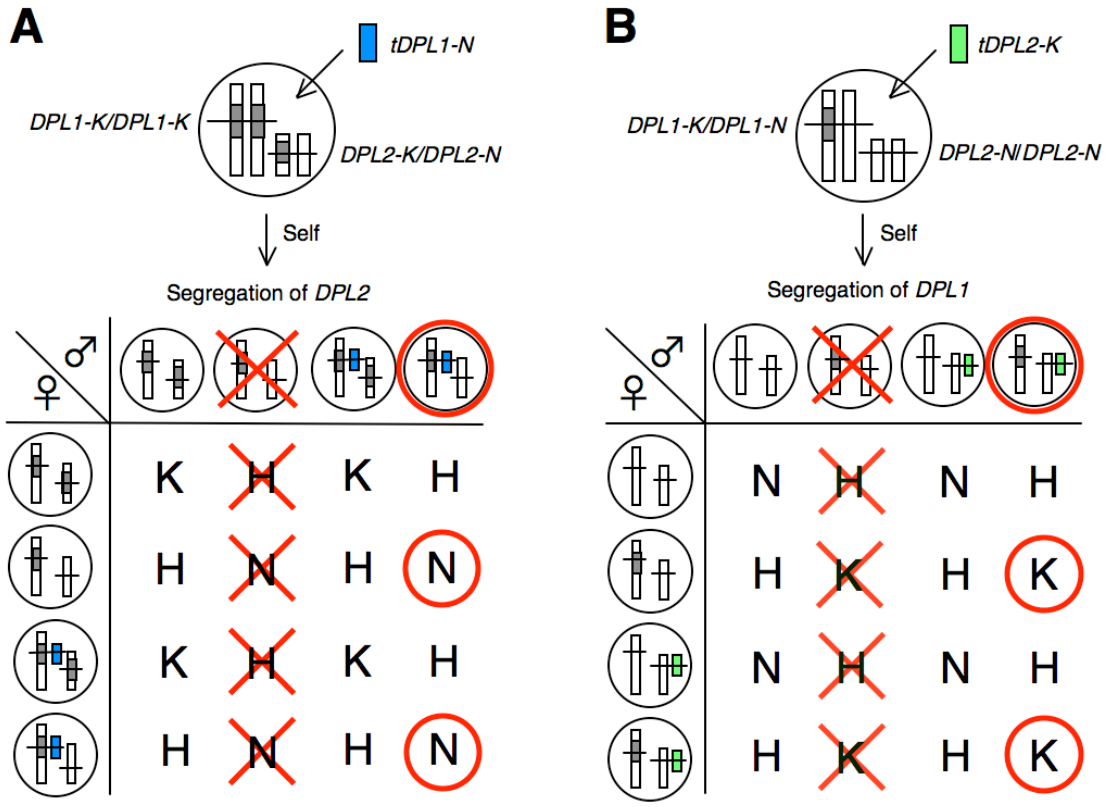


Fig. I-13. Schematic representation showing the concept of the complementation test. White and gray segments indicate Nipponbare and Kasalath fragments, respectively. Assuming that both of the *DPL1-K* and *DPL2-N* are loss of function alleles, the pollen (♂) having these two alleles together should be rescued by transforming with functional alleles, *DPL1-N* (*tDPL1-N*) or *DPL2-K* (*tDPL2-K*) that is shown as red circles. (A) When *tDPL1-N* was introduced into NILs possessing *DPL1-K* homozygote (*DPL1-K*/*DPL1-K*) and *DPL2* heterozygote (*DPL2-K*/*DPL2-N*), *DPL2-N* homozygous plants should be found in their self-pollinated progeny. (B) When *tDPL2-K* was transformed to NILs having *DPL1* heterozygous (*DPL1-K*/*DPL1-N*) and *DPL2-N* homozygous (*DPL2-N*/*DPL2-N*), *DPL1-K* homozygous plants should be found in self-pollinated progeny.

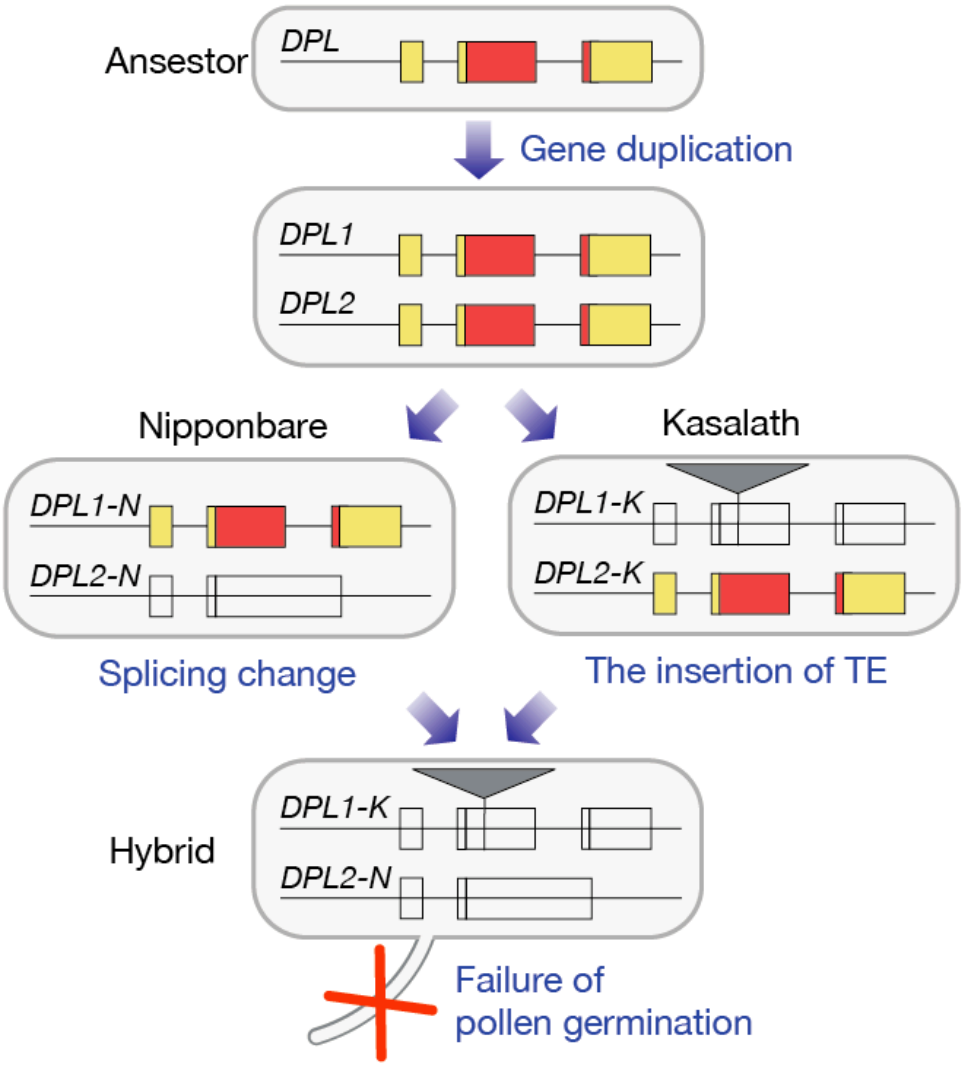


Fig. I-14. Model of the male gametophitic reproductive isolation by the combination of disrupted *DPL*s between Nipponbare and Kasalath. Red and Yellow boxes indicate the coding and untranslated regions, respectively. The inserted TE to the *DPL1-K* is indicated by triangle.

Table I-1 Primer sequences.

Primer name	Sequence (forward: 5' to 3')	Sequence (reverse: 5' to 3')	Gene for	Amplified size (bp)	
				Nipponbare	Kasalath
High-resolution mapping					
i27556	CCATGCATATGCAAACGAGC	ATCGCAGTACCAAGCTATCC	*DPL1*	421	356
i12296	AGGAAGAGGGAGCTCGGTAG	TGAGCCAATCATAACCGTCA	*DPL1*	491	518
i6342_2	GAATATTTGGGAGCCCCTGT	TGTTAAAAGGAGGCGCACTT	*DPL1*	523	478
i24075	GTTCGTGCTTCATCGACTCA	GAGGTTTATGTTGGGCGAAA	*DPL1*	700	700
ak067283	GTGACAACTCGTGTCATGGG	TTCTCCTGCAGACACATTGC	*DPL1*	901	901
i18100	GCTATCCCCTCACTCTCCCT	TGTTGCATTTGCACAGTCCC	*DPL1*	563	598
i276_2	GTTTAAAACGGGATGGGGAT	AAGACGAAGGAGTGGCTCAA	*DPL1*	585	560
i3068_2	GAGAAAAAGTGCCTCGTTGC	TTGACCTGTTCATCGCAGAG	*DPL1*	677	643
i7872	GGTCATTACCAGTCACATGC	ACTACGTCAGCTATCACTGC	*DPL1*	576	531
T004	GCTCAGCGTCGATCATTGTA	CCCAAGTCCTCCAACAAGAA	*DPL1*	677, 303*	960
i2121	AATTGATGCAGTGAGCTGCC	CAACCAACGAAAATCACGCC	*DPL2*	578	639
i316	CCTGCGAAGAATCGAATAGG	ATGGTGGTGTGTTCGTCTCA	*DPL2*	701	741
C381_8	GGCATCCGATGAAAGTGTAAA	AAGGCATAGGATCCTCGTTG	*DPL2*	249	156, 93*
C381_10	CAGTGAGCTGATGCAGAAACA	CGGCACTAGTTGAAAGTTGGA	*DPL2*	338	185, 155*
i7102	CGCACATATGCCGTACACTC	GTATTAAAGCCCCCTCCCAC	*DPL2*	423	391
i7102_2	TCGATGAAAATGTGCGGTTA	TTTGCCAGGACAAGAGGAAT	*DPL2*	499	422
i3635	CGCGACGTAAATAGACGTGA	GTTGCTGACCACTCCTCTCC	*DPL2*	795	747
i11601_2	AAGGTTGTGGCATTTCGAGC	GCCTATGACTCAGACAAACC	*DPL2*	599	504
i4099	TCTCTCATCCGATTCGTACG	TGTGTCGAAGATCGTAGCTG	*DPL2*	809	586
i4790	ATACGGCTGATTAGACGGAG	TGTTTCGTGGGCTATAGCTC	*DPL2*	828	1511
i1404_2	CCAACGGTTTGGAGAATACG	GGTGGCTGCTAGCTTCTGAC	*DPL2*	661	640
i2452291	CACACAAAACACCAACTCGG	TGGCCTTGTAGGCTGTATCC	*DPL2*	477	389
RT-PCR and *in situ* hybridization					
RT1_1f	AACAACGCTGGCGACCTT		*DPL1*		
GSP1_4		GATATTTCCATCAGAGCATG	*DPL1*		
RT6_1f	TGTCGGATTCACCGATTGTA		*DPL2*		
pENTR_N-His_N1r		TCATTTGCCCATTTCCCAGACA	*DPL2*		
RAc1	AACTGGGATGATATGGAGAA		*Actin*		
RAc2		CCTCCAATCCAGACACTGTA	*Actin*		
Real-time RT-PCR					
QRT1_1	AACAACGCTGGCGACCTTC	GTCCTCACTGCCCATCCTGA	*DPL1*		
QRT6_1	CGGATTCACCGATTGTAGCC	AACTGGTTGGCTTGGGTCAG	*DPL2*		
Ubq	AACCAGCTGAGGCCCAAGA	ACGATTGATTTAACCAGTCCATGA	*Ubiquitin*†		
Confirmation of transformant					
35S_2	CTACAAATGCCATCATTGCG	CGGTGAGTTCAGGCTTTTTC	*35S* (in vector)		
RT1-1f/RT1-2r	AACAACGCTGGCGACCTT	ATAGATTGCCAATGCAGAAG	*DPL1*	388, 154*	542
C381_5	TCTGGGCCAAATTTGACTGT	TTTGTTCTGCATTCCCTCCT	*DPL2*	298	272

*After the cleavage of PCR products (T004, AciI; C381_8, Hae II; C381_10, AciI; RT1-1f/RT1-2r, BglI).

†Primer sequences referred to Ishikawa *et al.* (2005).

Table I-2. Transformed *DPL1-N* and *DPL2-K* rescued pollen transmission possessing *DPL1-K* and *DPL2-N*. N, Nipponbare homozygous; H, heterozygous; K, Kasalath homozyogus.

Line	Genetic background of NILs	Transgene	Segregation in self-pollinated progenies								
			DPL1					*DPL2*			
			N	:	H	:	K	N	:	H	:	K
TG207	*DPL1-K/DPL1-K; DPL2-K/DPL2-N*	*DPL1-N*						18	:	20	:	12
TG208		*DPL1-N*						27	:	31	:	24
TG216		*DPL1-N*						6	:	42	:	7
TGV6		-						0	:	48	:	32
TGV18		-						0	:	48	:	16
TGV20		-						0	:	42	:	34
TG94	*DPL1-K/DPL1-N; DPL2-N/DPL2-N*	*DPL2-K*	12	:	30	:	19					
TG139		*DPL2-K*	15	:	31	:	14					
TG143		*DPL2-K*	17	:	35	:	28					
TG165		-	12	:	19	:	0					
TG169		-	7	:	9	:	0					
TG170		-	4	:	18	:	0					

Chapter II

Evolutionary analysis of the *DPL1* and *DPL2* genes

Introduction

Gene duplication is one of the major driving forces in the evolution of genomes and genetic systems (Ohno, 1970; Gu *et. al.*, 2003). Duplicate genes are believed to be a major mechanism for the establishment of new gene functions (Long and Lngley, 1993) and the generation of evolutionary novelty (Gilbert *et. al.*, 1997). The most spectacular method of gene duplication is whole genome duplication, which has a major role in the evolutionary history of eukaryotes (Ohno, 1970). Genome duplication is particularly prominent event in plants. *O. sativa* contains duplicated chromosomal regions that are attributable either to ancient segmental duplications (Vandepoele, 2003) or to a paleopolyploid event (Paterson *et. al.*, 2004). The overall structural similarity and functional conservation of *DPL1* and *DPL2* indicated that they are paralogous genes derived from gene duplication event. However, *O. sativa* chromosomes 1 and 6 do not share any homologous region of ancient chromosome, other than the duplicated segment between the *DPL1* and *DPL2* loci spanning only from a site near the end of first intron to the polyadenylation site of these genes. When did the *DPL* gene duplication and disruption occur? Which *DPL* was original? When was this reproductive isolation mechanism established? To investigate these questions, I sought *DPL* orthologs in angiosperm in public databases and carried out comparative analysis for gene duplication events. Next, I examined *DPL1* and *DPL2* genomic and cDNA sequences in various wild and cultivated rice strains to investigate gene disruption events. The aim to examine evolutionary scenario for this reproductive isolation mechanism is to know whether it would works as a positive selection mechanism, or just be a result of random drift.

Materials and Methods

Evolutionary analysis of *DPL*s in rice

In total, 42 wild and cultivated rice varieties in eight species, seven belonging to the *O. sativa* complex (*O. sativa, O. glaberrima, O. rufipogon, O. glumaepatula, O. longistaminata, O. meridionalis and O. barthii*) and one to the *Oryza officinalis* complex (*O. officinalis*) were used to examine their *DPL* orthologs as shown in Table II-2. All examined accessions from *O. glaberrima, O. rufipogon, O. glumaepatula, O. longistaminata, O. meridionalis and O. barthii* were provided by National Bioresource Project (NBPR), National Institute of Genetics (http://www.shigen.nig.ac.jp/rice/oryzabase/top/top.jsp). *Japonica* and *indica* varieties were stocked in author's laboratory that derived from Kyushu University, Japan and National Institute of Agrobiological Sciences (NIAS), Japan. For sequencing, genomic DNA was isolated from seedlings or adult leaves using the CTAB method (Murray and Thompson, 1980) and used as templates for PCR using GoTaq Hot Start Polymerase (Promega, Madison, WI). Purified PCR products were used as templates for sequencing reactions. For transcript analysis, total RNA was extracted from seedlings or mature anthers and used for RT-PCR with gene specific primers listed in Table I-1. Each procedure of RNA extraction and RT-PCR are described in above section. The nucleotide sequences were determined by an ABI 3130xl auto sequencer (Applied Biosystems, Foster City, CA), aligned with ClustalW program in DNA Data Bank of Japan (DDBJ; http://www.ddbj.nig.ac.jp/index-e.html), and used for further analysis by CLC Sequence viewer 6 (http://www.clcbio.com/index.php). Nucleotide variations including single nucleotide substitution

(SNS), nucleotide deletion (DEL), nucleotide insertion (INS) and insertion of the predicted transposable element that are the same as Kasalath allele *DPL1* (TE) but different from the Nipponbare sequence were searched for in the coding region of *DPL* orthologs. The phylogenic tree including *O. sativa* and *O. rufipogon* accessions/varieties was constructed using the coding sequence of *DPL1* and *DPL2* by the NJ method (Saitou and Nei, 1987). The relationship of other rice species referred to the phylogenetic tree constructed on the basis of *p-SINE* insertion patterns (Cheng *et al.*, 2002). DNA sequences are available from DDBJ (accession numbers AB534814 to AB534897).

Genome-to-genome synteny in angiosperms

DPL orthologs were searched for in the two public databases (DDBJ; http://www.ddbj.nig.ac.jp/index-e.html/ and Phytozome database; http://www.phytozome.net/index.php). The deduced amino acid sequences were predicted from deposited mRNA and genomic sequences. The neighboring predicted upstream and downstream genes to the *DPL* ortholog of each species was sought in Phytozome database (Table II-1). The deduced amino acid sequences of predicted genes were compared and classified into clusters based on sequence homology (http://www.phytozome.net/index.php). All-against-all blastp alignments were performed, and the bit score per unit peptide length is chosen as the similarity metric between two peptides in Phytozome 4.0 database.

***DPL* orthologs in angiosperm**

Deduced amino acid sequences of *DPL* orthologs were predicted from deposited mRNA in DDBJ; *Actinidia chinensis* (Accession No. FG527269), *Arabidopsis thaliana* (AV781610), *Arachis hypogaea* (GO330965), *Avena sativa* (GO582432), *Bruguiera gymnorhiza* (BP947345), *Carica papaya* (EX284454), *Citrus clementina* (FC910637), *Citrus limon* (DC891746), *Citrus macrophylla* (FC877729), *Citrus medica* (CX304656), *Citrus reticulata* (EX447729), *Citrus sinensis* (BQ623702), *Coffea arabica* (GT017940), *Coffea canephora* (DV713544), *Eucalyptus gunnii* (CU400557), *Glycine max* (AW706358), *Gossypium hirsutum* (DW496045), *Hevea brasiliensis* (EC606283), *Hordeum vulgare* (AK249681), *Ipomoea batatas* (CB329916), *Linum usitatissimum* (EU829971), *Nicotiana tabacum* (EB684060), *Panicum virgatumc* (FL727324), *Phyllostachys edulis* (FP097812), *Poncirus trifoliate* (CV706376), *Poplus tremuloides* (CA928957), *Prunus persica* (DY638117), *Pseudoroegneria spicata* (FF365103), *Quercus robur* (FP030478), *Saccharum officinarum* (CA066304), *Solanum lycopersicum* (CD002610), *Solanum melongena* (FS017522), *Solanum tuberosum* (BQ117762), *Triticum aestivum* (BQ168308), *Vigna unguiculata* (FG818518) and *Zea mays*_Chr2 (EB676931). The following deduced amino acid sequences of *DPL* orthologs were predicted from deposited genomic sequence in DDBJ; *Lotus japonicus* (AP010463), *Medicago truncatula*_Chr 5 (CU424495), *Medicago truncatula*_Chr 8 (AC157349) and *Vitis vinifera* (AM483775). Deduced amino acid sequences of *DPL* orthologs in *Brachypodium distachyon* and *Sorghum bicolor* were predicted form mRNA sequences in the US Department of Energy Joint Genome Institute (http://www.phytozome.net/index.php; Paterson *et al*., 2009); *Brachypodium distachyon* (Bradi1g47340), *Zea mays*_Chr 9 (GRMZM2G038640_T01), *Sorghum*

*bicolor*_Chr 10 (Sb10g005660.1) and *Sorghum bicolor*_Chr 2 (Sb02g039280.1). Deduced amino acid sequence of *DPL* orthologs in *Populus trichocarpa* (LG_I: 5944319-5944516) in the *Populus* genome project (Tuskan *et al.*, 2006). *Medicago truncatula*_Chr 5 (CU424495) showed a read through product of its second intron. *Zea mays*_Chr 9 (GRMZM2G038640_T01) was predicted to code for a truncated product with a deletion in the coding region. The obtained sequences were aligned by ClustalW software in DDBJ (http://www.ddbj.nig.ac.jp/index-e.html).

Results

Evolutionary conservation of *DPL* genes in angiosperms

The *DPL1* open reading frames are predicted to encode a 94 amino acid protein without any know domain in Nipponbare genome (Fig. II-1). To detect the *DPL* ortholog in other organisms, I performed BLAST search with deduced amino acid sequence of DPL1-N in the public databases. *DPL* is not similar to any proteins with the known functions and has no obvious functional motifs, although many orthologs in flowering plants were detected (Fig. II-1). In total, 46 *DPL* orthologs were detected, and *DPL* genes are highly conserved among angiosperms. Especially, the middle and C-terminus residues of the deduced amino acid sequences of *DPL* ortholog are highly conserved.

The history of gene duplication of *DPL* in angiosperms

Database search indicated that five different species, *Oryza sativa*, *Sorghum bicolor*, *Zea mays*, *Glycine max* and *Medicago truncatula* had two *DPL* orthologs (Fig. II-1). In contrast, *Brachypodium distachyon,* which belongs to a genus close to *Oryza* (Bolot *et al*., 2009) had a single *DPL*. To investigate when gene duplication of *DPL* occurred, and which rice *DPL* is the original, syntenic conservation around *DPL* genes among above five species (*O. sativa*, *S. bicolor*, *Z. mays*, *G. max* and *M. truncatula*) and *B. distachyon* was investigated in the Phytozome 4.0 database (http://www.phytozome.net/index.php, see also Materials and Methods). The regions flanking *DPL2* on rice chromosome 6 had sequence similarity with

the region around the *DPL* ortholog on chromosome 1 of *B. distachyon* (Table II-1 and Fig. II-2). Similarly, the *DPL2* region on rice chromosome 6 exhibited syntenic conservation among all of the grasses examined, having similarity with chromosome 10 of *Sorghum bicolor*, and chromosome 9 of *Zea mays*. These results suggested that *DPL2* on rice chromosome 6 is the most ancestral *DPL* in cereal species.

In contrast to *DPL2*, the sequences flanking *DPL1* on rice chromosome 1 were not similar to sequences flanking the other *DPL* orthologs on sorghum and maize chromosome 2. The similarity of the deduced amino acid sequences of DPL1-N on rice chromosome 1 and DPL2-K on rice chromosome 6 (97.9% identity) was higher than that between DPL2-K and the DPL ortholog on *Brachypodium* chromosome 1 (88.3% identity, Fig. II-3). These data suggested that *DPL1* on rice chromosome 1 was generated by a recent gene duplication of *DPL2* after *Oryza-Brachypodium* differentiation (yellow arrowhead in Fig. II-2).

There was a syntenic relationship between the sorghum *DPL* orthologs on chromosomes 2 and 10 and the maize *DPL* orthologs on chromosomes 2 and 9, respectively (Fig. II-2). The similarity of the deduced amino acid sequences of the DPL ortholog on sorghum chromosome 2 and the *DPL* on maize chromosome 2 (100% identity) was higher than that between the sorghum *DPL*s on chromosomes 2 and 10 (89.4% identity, Fig. II-3). These data suggested that gene duplication of *DPL* occurred before the divergence of sorghum and maize (blue arrowhead in Fig. II-2).

In the Leguminosae, genome structures around the two *DPL* genes in *G. max* and *M. truncatula* showed synteny with each other. The syntenic relationship between the two chromosomes harboring *DPL* genes indicated segmental duplication of the *DPL* containing region

in a common ancestor of *G. max* and *M. truncatula* (pink arrowhead in Fig. II-2).

These results indicated the duplication of *DPL* occurred at least three times independently in the evolution of angiosperms.

The history of gene disruption of *DPL* in rice

To investigate relationships between the disruption of *DPL*s and *Oryza* differentiation, evolutionary analysis was performed with well-diverged rice accessions or varieties belonging to seven closely related species (Vaughan *et al*., 2003). In total, 42 wild and cultivated rice varieties in eight species, seven belonging to the *O. sativa* complex (*O. sativa, O. glaberrima, O. rufipogon, O. glumaepatula, O. longistaminata, O. meridionalis and O. barthii*) and one to the *Oryza officinalis* complex (*O. officinalis*) were used to examine their *DPL* orthologs (Table II-2). Within these eight species, *O. sativa* and *O. glaberrima* are species domesticated from *O. rufipogon* and *O. barthii*, respectively. All 42 accessions and varieties had both *DPL1* and *DPL2* orthologs suggesting that the gene duplication of *DPL* occurred in a common ancestor of them (Fig. II-4).

To discuss when this reproductive isolation mechanism was established, the disruptions of *DPL*s were also investigated along with flowering species differentiation. Using nucleotide variation in the coding region of *DPL*s (Fig. II-4), phylogenic relationships were calculated and illustrated in Fig. II-5 (see also Materials and Methods). Since the nucleotide variations in the coding region of *DPL*s, *O. sativa* and *O. rufipogon* accessions and varieties

could be classified into four groups; group I, tropical and temperate *japonica* and *O. rufipogon*; group II, *indica* and *O. rufipogon*; group III, *indica* and *O. rufipogon* and group IV, *O. rufipogon*

(Fig. II-4).

Gene disruption of *DPL*s occurred at least five times independently, four times in *DPL1* and once in *DPL2*, during rice evolution (Fig. II-5). Gene disruptions of *DPL1* are following four times; (1) *DPL1-K* type disruption. *DPL1-K* type disruption occurred in an ancestor of group III including Kasalath. (2) Different kind of gene disruption of the *DPL1* gene occurred in a common ancestor of *O. glaberrima* and *O. barthii* (Fig. II-4 and Fig. II-5). Because all four species belonging to *O. glaberrima* and *O. barthii* had a 625 bp deletion including second exon of their *DPL1* (Fig. II-4) with no mRNA expressions (Fig. II-6). (3) The *DPL1* genes of W1236 from *O. rufipogon* in group IV had a nucleotide substitution at the 21st polymorphic site generating a premature stop codon (Fig. II-4). (4) The *DPL1* genes of W1169 from *O. glumaepatula* had a nucleotide deletion at the 10th polymorphic site resulting in a frame-shift (Fig. II-4). The mRNA expression of *DPL1* of W1169 was not observed (Fig. II-6).

DPL2-N disruption occurred in a common ancestor of tropical and temperate *japonica* cultivars in group I (Fig. II-5 and Table II-2). The read through product of *DPL2* was only observed in *japonica* cultivars in group I (Fig. II-6); they had a nucleotide substitution at a predicted splicing branch site in the second intron (Table II-2).

Discussion

The purpose of this chapter was to reveal the times of gene duplication of *DPL* and establishment of this reproductive isolation mechanism. This information should clarify the role of this reproductive isolation for plant speciation. I confirmed that the paralogous disrupted genes provide an important source of Dobzhansky-Muller type epistatic interactions after gene duplication. Using syntenic information with recently sequenced angiosperms, it was revealed that multiple duplications of *DPL* in angiosperms, and an original copy in cereal species was *DPL2* (Fig. II-2). This result also suggested the duplication of *DPL* occurred at least three times independently, after *Oryza-Brachypodium* differentiation (yellow arrowhead in Fig. II-2), before *Sorghum*-maize differentiation (blue arrowhead in Fig. II-2), and before *G. max* and *M. truncatula* differentiation (pink arrowhead in Fig. II-2). Recently, multiple comparative analyses revealed that cereal genome had been through the whole genome duplication (Paterson, A. H. *et al.*, 2009; Salse, J. *et al.*, 2009). However, *O. sativa* chromosomes 1 and 6 do not share any homologous region of ancient chromosome, other than the duplicated segment between the *DPL1* and *DPL2* loci spanning only from a site near the end of first intron to the polyadenylation site of these genes (Fig. I-7). A mechanism of gene duplication of *DPL* after *Oryza-Brachypodium* differentiation remains unknown. Similarly, many duplicated genes without genome duplication were found in rice genome (Paterson, A. H. *et al.*, 2009). It is unknown what mechanism made single gene duplicated, like *DPL*.

It is widely recognized that *indica* and *japonica* were polyphyletic, but not completely independently domesticated from different ancestors of *O. rufipogon* at Asian continent (Cheng, *et*

al., 2003; Garris *et al.,* 2005; Gao and Innan, 2008; Khush, 1997; Oka and Chang, 1962). Considering nucleotide variations in the coding region of *DPL*s, *O. sativa* and *O. rufipogon* accessions and varieties could be classified into four groups (Fig. II-4, Table II-2). This classification showed good agreement with phylogenic analyses of *O. rufipogon* and *O. sativa* by p-*SINE* insertion patterns (Cheng *et al.*, 2003; Xu *et al.*, 2007). Based on the results of our phylogenic analysis, independent domestication processes of *japonica* and *indica* were also supported (Fig. II-4, Table II-2). The common ancestor of *O. sativa* and *O. rufipogon* might have had both functional *DLP1* and *DPL2* genes, due to the both *DPL1* and *DPL2* genes in groups II and IV seemed to be functional except in the case of W1921 (Table II-2) and W1236 (Fig. II-4). *DPL2-N* should be originated in the common ancestor of tropical and temperate *japonica* rice after the split from the ancestor of *indica* populations. In contrast, the insertion of the TE in the *DPL1* gene must have occurred in the one population of *indica* including the ancestral varieties of group III. All I to IV groups contained *O. rufipogon* accessions indicated that these four groups must have differentiated before the domestication of *O. sativa*. Our results suggested that *indica* rice is polyphyletically domesticated (at least two times) from the different ancestors of *O. rufipogon*, whereas *japonica* rice is monophyletic population. This result also supported the phylogenic analyses of *O. sativa* by simple sequence repeat (Garris *et al.,* 2005) and single nucleotide polymorphisms (McNally *et al.*, 2009).

Gene disruption of *DPL*s occurred at least five times independently, four times in *DPL1* and once in *DPL2*, during rice evolution (Fig. II-5). Because these disruptions occurred after *japonica-indica* differentiation, reproductive isolation by the combination of disrupted *DPL1* and

DPL2 is not likely to initially act as barriers between *japonica* and *indica*. Although the disruption of the duplicated *DPL*s might have not worked as a positive species isolation mechanism, reinforced the population integrity of *japonica* and *indica* in group III, when the populations came into contact with each other. Independent disruptions of *DPL*s must induce reproductive isolation between *japonica* cultivars in group I and strains in group III, and also between *japonica* cultivars and other *DPL1* disrupted species. Each could indicate relevant timing with the establishment of the reproductive isolation between rice differentiating lineages, and their functional roles in rice reproduction.

These results showed one of the mechanisms of reproductive isolation and a comprehensive story of the evolution of *DPL*s in plant. It remains unknown whether multiple duplications and disruptions of *DPL* genes resulted from adaptive selection or random drift in plant speciation. Further studies of *DPL* function and analyses of reproductive isolation events in *Oryza* will provide fundamental understanding of molecular functions in plant reproduction and the mechanisms of species diversification.

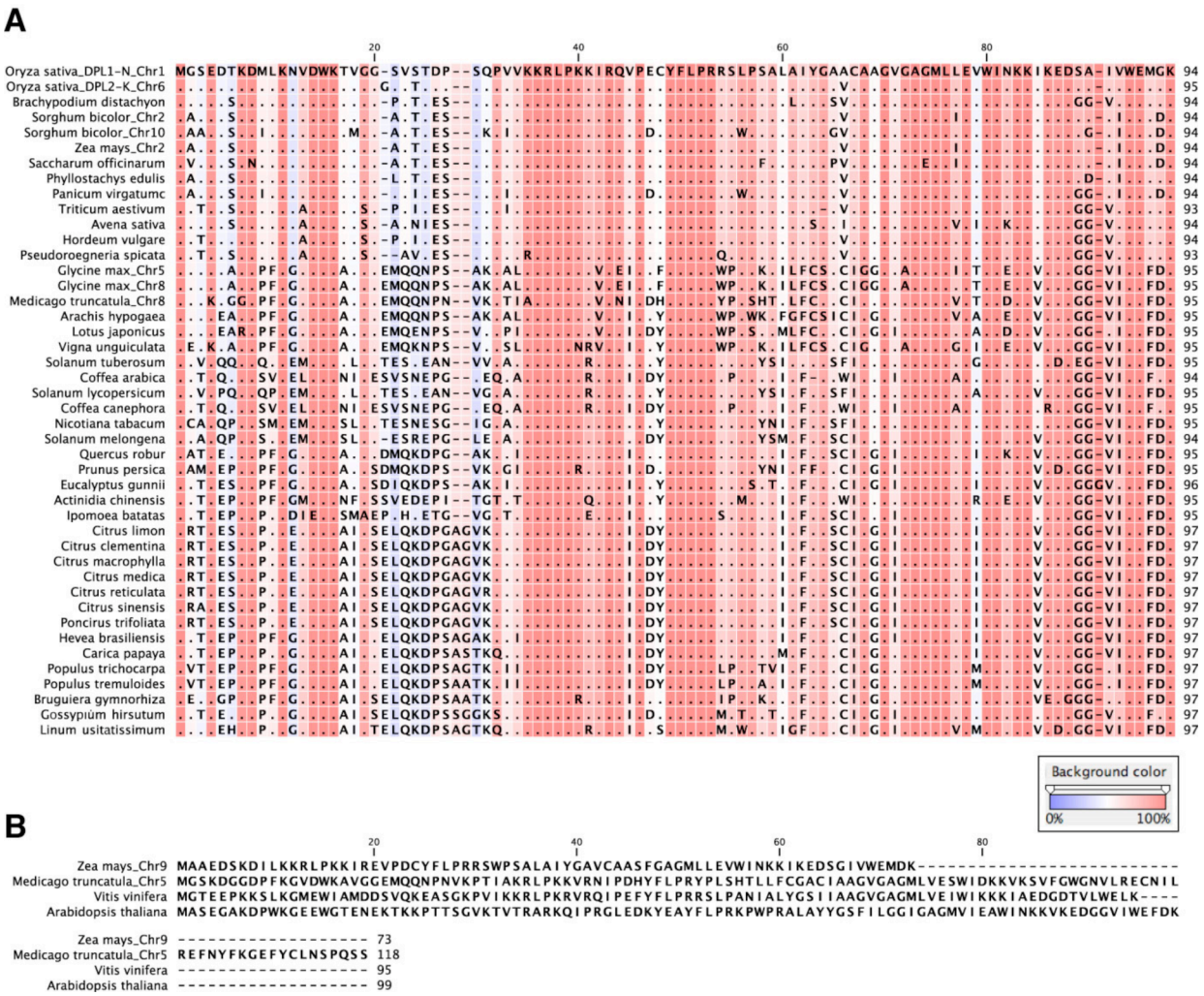


Fig. II-1. (A) Multiple alignments of deduced amino acid sequences of DPL orthologs in angiosperms. Amino acid sequences of DPL orthologs were predicted from deposited mRNA and genomic sequences in public databases (see Materials and Methods) and aligned by ClustalW software in DDBJ. Identical residues to *Oryza sativa*_DPL1-N_Chr1 are shown as dots. Hyphens indicate gaps. Background colors for the residues display the level of conservation at each position in the alignment. The gradient of the color reflects how particular position is conserved in the alignment. (B) The DPL orthologs of *Zea mays*_Chr 9, *Medicago truncatula*_Chr 5, *Vitis vinifera* and *Arabidopsis thaliana* showed low sequence similarity among DPL orthologs in angiosperms. The DPL ortholog of *Medicago truncatula*_Chr 5 showed read through product of its second intron.

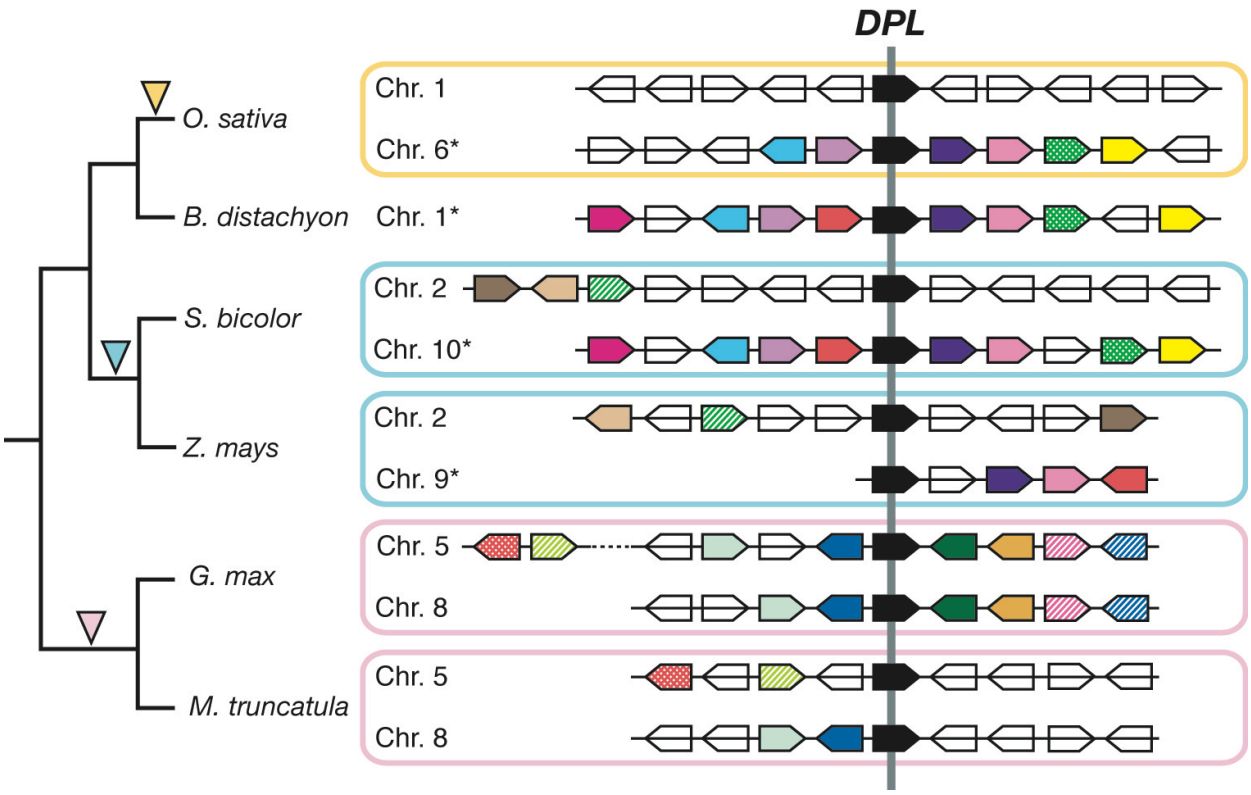


Fig. II-2. Evolutionary history of gene duplication and disruption of *DPL*s in plant. Historical representation of three independent duplications of the *DPL* among the divergence of rice, sorghum and Leguminosae. Predicted upstream and downstream neighboring genes to the *DPL* ortholog in each species are shown. The syntenic genes showing similarity with each other are shown as the same color (see also Table II-1). Asterisks show the predicted ancestral chromosome of *DPL* in Grasses. Each predicted gene duplication point is shown by inverted triangles.

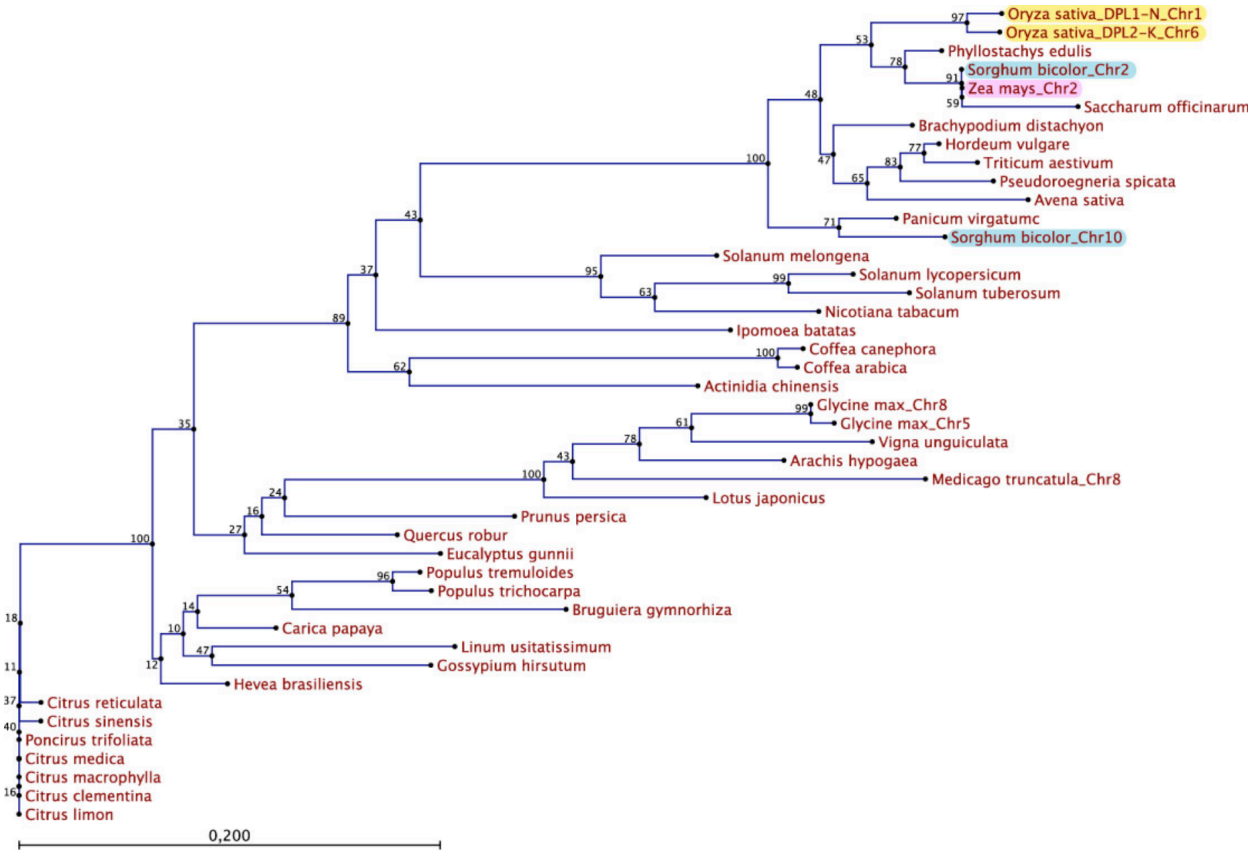
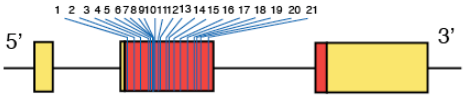


Fig. II-3. A phylogenetic tree of *DPL* genes in angiosperm. The phylogenetic tree was constructed by the NJ method (Saitou and Nei, 1987) based on the deducted amino acid sequences of *DPL* orthologs shown in Fig. II-1. Colored circles coincided with Fig. II-2. Distance bar is shown below the tree.

A

Species	Variety/Accession	Group	Variations in the coding sequence of *DPL1*																					Variations in the coding sequence of *DPL2*																		
			INS	SNS	SNS	SNS	SNS	SNS	SNS	DEL	SNS	DEL	SNS	SNS	SNS	SNS	SNS	SNS	SNS	SNS	TE	SNS	SNS	SNS	SNS	SNS	SNS	DEL	SNS	SNS	SNS	SNS	SNS	SNS	SNS	SNS	SNS	SNS	SNS	SNS	SNS	SNS	SNS
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
O. sativa (*japonica*)	Nipponbare	I	-	T	G	G	G	C	C	-	C	-	G	G	A	T	T	T	C	G	-	G	A	T	T	T	-	T	G	T	A	C	T	T	A	G	T	C	G	C	G	A
O. sativa (*japonica*)	Asominori	I	-	T	G	G	G	C	C	-	C	-	G	G	A	T	T	T	C	G	-	G	A	T	T	T	-	T	G	T	A	C	T	T	A	G	T	C	G	C	G	A
O. sativa (*japonica*)	FL7	I	-	T	G	G	G	C	C	-	C	-	G	G	A	T	T	T	C	G	-	G	A	T	T	T	-	T	G	T	A	C	T	T	A	G	T	C	G	C	G	A
O. sativa (*japonica*)	FL102	I	-	T	G	G	G	C	C	-	C	-	G	G	A	T	T	T	C	G	-	G	A	T	T	T	-	T	G	T	A	C	T	T	A	G	T	C	G	C	G	A
O. sativa (*japonica*)	Taichung 65	I	-	T	G	G	G	C	C	-	C	-	G	G	A	T	T	T	C	G	-	G	A	T	T	T	-	T	G	T	A	C	T	T	A	G	T	C	G	C	G	A
O. sativa (*japonica*)	Koshihikari	I	-	T	G	G	G	C	C	-	C	-	G	G	A	T	T	T	C	G	-	G	A	T	T	T	-	T	G	T	A	C	T	T	A	G	T	C	G	C	G	A
O. sativa (*japonica*)	Shinriki	I	-	T	G	G	G	C	C	-	C	-	G	G	A	T	T	T	C	G	-	G	A	T	T	T	-	T	G	T	A	C	T	T	A	G	T	C	G	C	G	A
O. sativa (*japonica*)*	Dinalaga	I	-	T	G	G	G	C	C	-	C	-	G	G	A	T	T	T	C	G	-	G	A	T	T	T	-	T	G	T	A	C	T	T	A	G	T	C	G	C	G	A
O. sativa (*japonica*)*	Nanjing Xiang Dao	I	-	T	G	G	G	C	C	-	C	-	G	G	A	T	T	T	C	G	-	G	A	T	T	T	-	T	G	T	A	C	T	T	A	G	T	C	G	C	G	A
O. sativa (*japonica*)*	Ketan Nangka	I	-	T	G	G	G	C	C	-	C	-	G	G	A	T	T	T	C	G	-	G	A	T	T	T	-	T	G	T	A	C	T	T	A	G	T	C	G	C	G	A
O. rufipogon	W1958	I	-	T	G	G	G	C	C	-	C	-	G	G	A	T	T	T	C	G	-	G	A	T	T	T	-	T	G	T	A	C	T	T	A	G	T	C	G	C	G	A
O. rufipogon	W1945	I	-	T	G	G	G	C	C	-	C	-	G	G	A	T	T	T	C	G	-	G	A	T	T	T	-	T	G	T	A	C	T	T	A	G	T	C	G	C	G	A
O. sativa (*indica*)	T0421	II	-	T	G	G	G	C	C	-	C	-	G	G	A	T	T	T	C	G	-	G	A	T	T	T	-	T	G	T	A	C	T	T	C	G	T	C	G	C	G	A
O. sativa (*indica*)	T0710	II	-	T	G	G	G	C	C	-	C	-	G	G	A	T	T	T	C	G	-	G	A	T	T	T	-	T	G	T	A	C	T	T	C	G	T	C	G	C	G	A
O. sativa (*indica*)	T0724	II	-	T	G	G	G	C	C	-	C	-	G	G	A	T	T	T	C	G	-	G	A	T	T	T	-	T	G	T	A	C	T	T	C	G	T	C	G	C	G	A
O. sativa (*indica*)	Dao Ren Qiao	II	-	T	G	G	G	C	C	-	C	-	G	G	A	T	T	T	C	G	-	G	A	T	T	T	-	T	G	T	A	C	T	T	C	G	T	C	G	C	G	A
O. sativa (*indica*)	IR24	II	-	T	G	G	G	C	C	-	C	-	G	G	A	T	T	T	C	G	-	G	A	T	T	T	-	T	G	T	A	C	T	T	C	G	T	C	G	C	G	A
O. rufipogon	W0120	II	-	T	G	G	G	C	C	-	C	-	G	G	A	T	T	T	C	G	-	G	A	T	T	T	-	T	G	T	A	C	T	T	C	G	T	C	G	C	G	A
O. rufipogon	W0106	III	-	T	G	G	G	T	C	+	N/A	N/A	G	G	A	T	T	T	C	G	+	G	A	T	T	T	-	T	G	T	A	C	T	T	C	G	T	C	G	C	G	A
O. rufipogon	W1681	III	-	T	G	G	G	T	C	+	N/A	N/A	G	G	A	T	T	T	C	G	+	G	A	T	T	T	-	T	G	T	A	C	T	T	C	G	T	C	G	C	G	A
O. rufipogon	W1866	III	-	T	G	G	G	T	C	+	N/A	N/A	G	G	A	T	T	T	C	G	+	G	A	T	T	T	-	T	G	T	A	C	T	T	C	G	T	C	G	C	G	A
O. rufipogon	W0610	III	-	T	G	G	G	T	C	-	C	-	G	G	A	T	T	T	C	G	+	A	A	T	T	T	-	T	G	T	A	C	T	T	C	G	T	C	G	C	G	A
O. sativa (*indica*)	Kasalath	III	-	T	G	G	G	T	C	-	C	-	G	G	A	T	T	T	C	G	+	G	A	T	T	T	-	T	G	T	A	C	T	T	C	G	T	C	G	C	G	A
O. sativa (*indica*)	Dular	III	-	T	G	G	G	T	C	-	C	-	G	G	A	T	T	T	C	G	+	G	A	T	T	T	-	T	G	T	A	C	T	T	C	G	T	C	G	C	G	A
O. sativa (*indica*)	N22	III	-	T	G	G	G	T	C	-	C	-	G	G	A	T	T	T	C	G	+	G	A	T	T	T	-	T	G	T	A	C	T	T	C	G	T	C	G	C	G	A
O. sativa (*indica*)	9311	III	-	T	G	G	G	T	C	-	C	-	G	G	A	T	T	T	C	G	+	G	A	T	T	T	-	T	G	T	A	C	T	T	C	G	T	C	G	C	G	A
O. sativa (*indica*)	Kinandang puti	III	-	T	G	G	G	T	C	-	C	-	G	G	A	T	T	T	C	G	+	G	A	T	T	T	-	T	G	T	A	C	T	T	C	G	T	C	G	C	G	A
O. rufipogon	W2003	IV	-	T	G	G	G	T	C	-	C	-	G	G	A	T	T	T	C	G	-	G	A	T	T	T	-	T	G	T	A	C	T	T	C	G	T	C	G	C	G	A
O. rufipogon	W1294	IV	-	T	G	G	G	T	C	-	C	-	G	G	A	T	T	T	C	G	-	G	A	T	T	T	-	T	G	T	A	C	T	T	C	G	T	C	G	C	G	A
O. rufipogon	W1921	IV	-	T	G	G	G	T	C	-	C	-	G	G	A	T	T	T	C	G	-	G	A	T	T	T	-	T	G	T	A	C	T	T	C	G	T	C	G	C	G	A
O. rufipogon	W1236	IV	-	T	G	G	G	T	C	-	C	-	G	G	A	T	T	T	C	G	-	G	T	T	T	T	-	T	G	T	A	C	T	T	C	G	T	C	G	C	G	A
O. glaberrima	C0025		N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	T	T	T	-	T	G	T	A	C	C	T	C	G	T	C	G	C	G	A
O. glaberrima	C8863		N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	T	T	T	-	T	G	T	A	C	T	T	C	G	T	C	G	C	G	A
O. barthii	W1588		N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	T	T	T	-	T	G	T	A	C	C	T	C	G	T	C	G	C	G	A
O. barthii	W0698		N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	T	T	T	-	T	G	T	A	C	C	T	C	G	T	C	G	C	G	A
O. glumaepatula	W1169		-	T	G	G	G	T	C	-	C	+	G	G	A	T	T	T	C	C	-	G	A	T	T	T	-	T	G	T	A	C	T	T	C	A	T	C	G	C	G	A
O. glumaepatula	W1187		-	T	G	G	G	T	C	-	C	-	G	G	A	T	T	T	C	G	-	G	A	T	T	T	-	T	G	T	A	C	T	T	C	A	T	C	G	C	G	A
O. longistaminata	W1508		+	T	G	G	G	T	C	-	C	-	G	G	A	T	T	T	C	G	-	G	A	T	T	T	+	T	G	T	A	C	T	T	C	G	T	T	A	C	A	A
O. longistaminata	W1540		+	T	G	G	G	T	C	-	C	-	G	G	A	T	T	T	C	G	-	G	A	T	T	C	+	T	G	T	A	C	T	T	C	G	T	C	A	C	A	A
O. meridionalis	W1625		-	T	A	G	C	T	C	-	G	-	G	A	A	T	T	T	T	G	-	G	A	C	T	T	+	C	G	C	A	T	T	T	C	G	T	T	G	C	G	A
O. meridionalis	W1635		-	T	A	G	C	T	C	-	G	-	G	A	A	G	T	T	T	G	-	G	A	C	T	T	+	C	G	C	A	T	T	T	C	G	T	T	G	T	G	A
O. officinalis	W0002		-	C	G	A	C	T	T	-	C	-	A	A	C	T	C	C	T	G	-	G	A	C	C	T	+	T	A	C	C	C	T	C	C	G	C	C	G	C	G	G

B



C

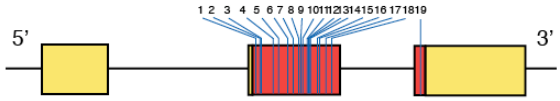


Fig. II-4. Variation in the coding sequences of *DPL*s in the rice 42 accessions/varieties. (A) *O. rufipogon* and *O. sativa* accessions/varieties were classified into four groups (I-IV) based on variations in the coding sequence of *DPL1* and *DPL2*. Each polymorphic site in the coding regions of *DPL1* and *DPL2* is shown in B and C, respectively. Each gene structure refers to Nipponbare *DPL1* and Kasalath *DPL2* that is shown in Fig. I-7. N/A means not applicable sites by following reasons; the polymorphic sites at the 9th and 10th in *DPL1* orthologs of W0106, W1681 and W1866 were ignored because of nucleotide deletions at the 8th site (14 bp). Four accessions from *O. glaberrima* and *O. barthii* showed 652 bp deletions of *DPL1* orthologs including coding region. The polymorphic sites at the 10th and 21st in *DPL1* orthologs of W1169 and W1236 were one nucleotide insertions resulting in a frame-shift and generating premature stop codon, respectively. The polymorphic sites at the first position in *DPL1* orthologs of *O. longistaminata* and 4th in *DPL2* orthologs of *O. longistaminata, O. meridionalis* and *O. officinalis* are three nucleotide insertions. INS, nucleotides insertion; SNS, Single nucleotide substitution; DEL, nucleotide deletion; TE, insertion of predicted transposable element.

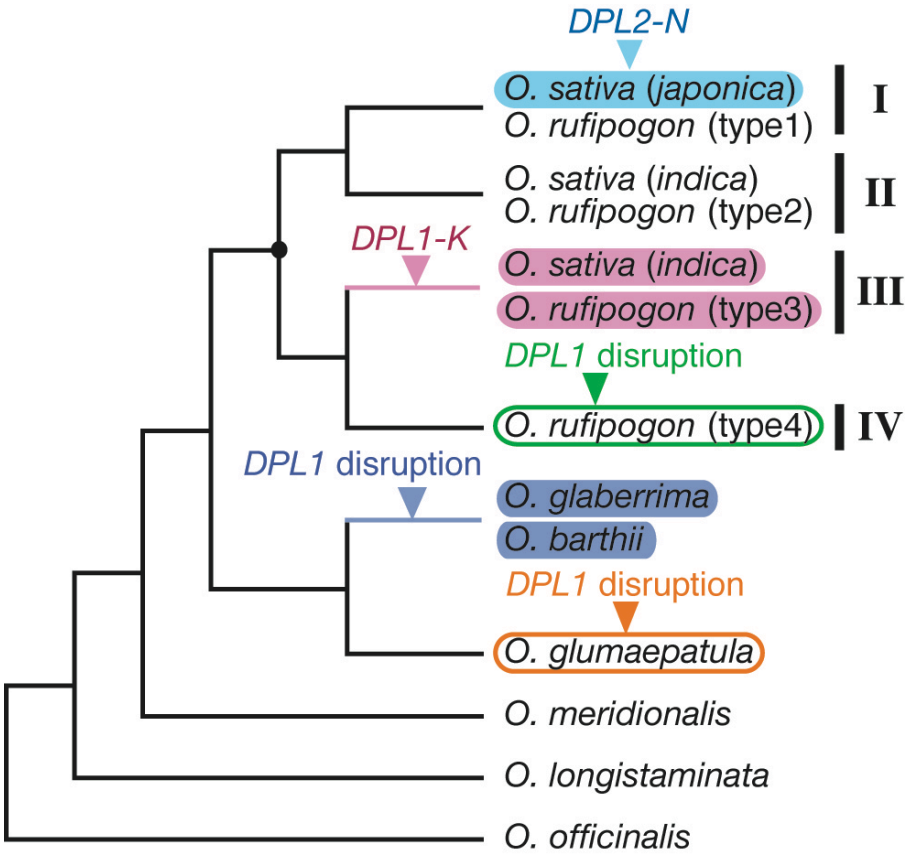


Fig. II-5. Phylogenetic relationships of eight species from *Oryza sativa* complex and *Oryza officinalis* complex. A clade consisting of *O. sativa* and *O. rufipogon* varieties (shown as black node) was divided into four groups according to *DPL* nucleotide variations as shown in Fig. II-4. The species having disrupted *DPL*s are shown as inverted triangles and colored circle. Partial species having disrupted *DPL*s are shown as open circles.transposable element that is the same or similar to that observed in the Kasalath allele of *DPL1*.

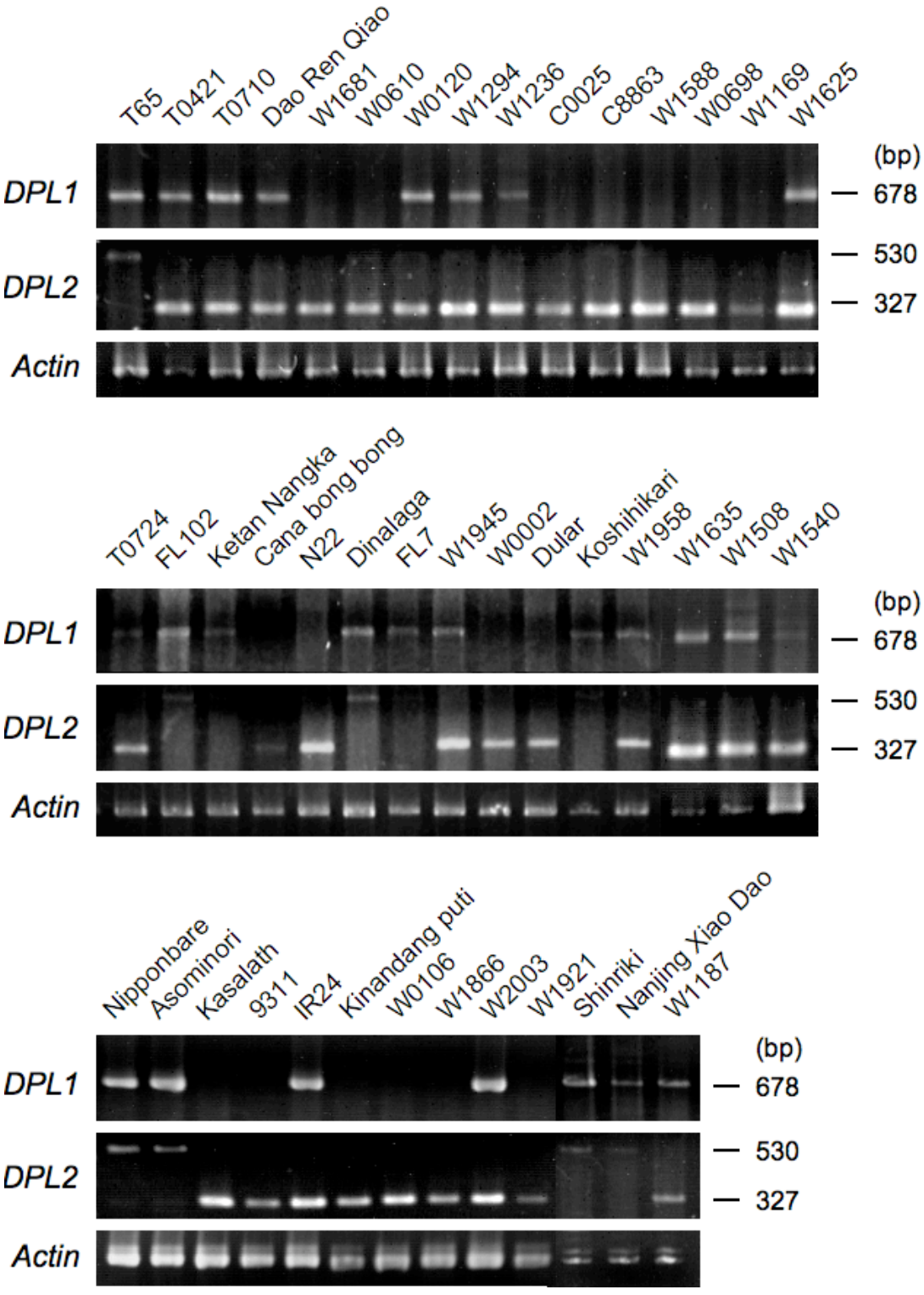


Fig. II-6. RT-PCR analysis of *DPL*s in the other wild and cultivated rice. The PCR products of *DPL2* (530 and 327 bp) associated with the products size of *DPL2-N* (530 bp) and *DPL2-K* (327 bp), respectively. Total RNA was extracted from mature anther or seedling (see Table II-2). Gene specific primers for *DPL*1 and *DPL2* are shown as arrowheads in Fig. I-7. *Actin* was used as control.

Table II-1 The accession and cluster numbers of predicted genes neighbouring *DPL*.

Species	Chr.	Gene order neighbouring *DPL**															
				-7	-6	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Oryza sativa	1	-	-	-	-	12001.m150929	12001.m08139	12001.m08140	12001.m08141	12001.m08142	12001.m150928	12001.m150542	12001.m08145	12001.m08146	12001.m08147	12001.m08148
Oryza sativa	6	-	-	-	-	12006.m05576	12006.m91824	12006.m05575	12006.m05573	12006.m31847	12006.m05570	12006.m05569	12006.m05568	**12006.m05567**	12006.m05566	12006.m05565
Brachypodium distachyon	1	-	-	-	-	Bradi1g47290.1	Bradi1g47300.1	Bradi1g47310.1	Bradi1g47320.1	Bradi1g47330.1	Bradi1g47340.1	Bradi1g47350.1	Bradi1g47360.1	**Bradi1g47370.1**	Bradi1g47380.1	Bradi1g47390.1
Sorghum bicolor	2	-	-	Sb02g039220.1	Sb02g039230.1	Sb02g039240.1	Sb02g039250.1	Sb02g039255.1	Sb02g039260.1	Sb02g039270.1	Sb02g039280.1	Sb02g039290.1	Sb02g039300.1	Sb02g039310.1	Sb02g039315.1	Sb02g039320.1
Sorghum bicolor	10	-	-	-	-	Sb10g005710.1	Sb10g005700.1	Sb10g005690.1	Sb10g005680.1	Sb10g005670.1	Sb10g005660.1	Sb10g005650.1	Sb10g005640.1	Sb10g005635.1	**Sb10g005630.1**	Sb10g005620.1
Zea mays	2	-	-	-	-	GRMZM2G124335_T02	AC202384.4_FGT040	GRMZM2G148194_T05	AC202384.4_FGT050	GRMZM2G159237_T01	GRMZM2G159218_T01	GRMZM2G159210_T01	GRMZM2G159205_T01	AC202384.4_FGT046	GRMZM2G109120_T01	-
Zea mays	9	-	-	-	-	-	-	-	-	-	GRMZM2G038640_T01	GRMZM2G067666_T01	GRMZM2G067654_T01	GRMZM2G067646_T01	GRMZM2G116491_T01	-
Glycine max	5	**Glyma05g37060.1**	Glyma05g37070.1			Glyma05g26040.1	Glyma05g26030.1	Glyma05g26020.1	Glyma05g26010.1	Glyma05g26000.1	Glyma05g25990.1	Glyma05g25980.1	Glyma05g25970.1	Glyma05g25960.1	Glyma05g25950.1	Glyma05g25940.1
Glycine max	8	-	-	-	-	Glyma08g08980.1	Glyma08g08970.1	Glyma08g08960.1	Glyma08g08950.1	Glyma08g08940.1	Glyma08g08930.1	Glyma08g08920.1	Glyma08g08910.1	Glyma08g08900.1	Glyma08g08890.1	Glyma08g08880.1
Medicago truncatula	5	-	-	-	-	CT033771_13	**CT033771_31**	CU424495_8	CU424495_9	CU424495_10	CU424495_11	CU424495_12	CU424495_13	CU424495_14	CU424495_15	CU424495_16
Medicago truncatula	8	-	-	-	-	AC157349_29	AC157349_38	AC157349_35	AC157349_49	AC157349_26	AC157349_48	AC157349_23	AC157349_21	AC157349_20	AC157349_47	AC157349_15

*The sequences and location of each predicted genes are available in Phytozome database 4.0 (genome browse version 1.69, see also Materials and Methods). Each predicted gene in the same cluster is rendered as the same non-white color. These colors of predected genes coincided with Fig. II-2.

Cluster No.	Color	Annotation of each cluster
11671513		Uncharacterized
11679638		Putative RPT2
11687521		Putative (S)-2-hydroxy-acid oxidase (Os07g0616500 protein)
11704422		Uncharacterized
11706663		Uncharacterized
11711279		Uncharacterized
11719340		Uncharacterized
11720951		Uncharacterized
11724722		Similar to Chromatin-remodeling factor CHD3
11776858		Uncharacterized
11777029		Similar to expressed protein in Arabidopsis thaliana; similar to DUF602; [ortholog of At5g58020]
11789995		Similar to Basic helix-loop-helix (BHLH)-like
11812984		Uncharacterized
11818640		Similar to Beta-ketoacyl-ACP synthase
11818640		Similar to Beta-ketoacyl-ACP synthase
11819772		*DPL*
11824306		Putative ubiquitin-specific protease 23
11825416		Uncharacterized
11128555		Similar to COP1-INTERACTIVE PROTEIN 1; [co-ortholog (2of2) of At1g64330, At5g41790]
11174100		Uncharacterized

Table II-2 The variation of *DPL* orthologs in rice 42 accessions and varieties.

Species	Accession/Variety	Original place	*DPL1*	*DPL2*		Tissue for RT-PCR	Group‡
			mRNA expression*	Nucleotide sequence†	Splicing of second intron*		
O. sativa (Temperate *japonica*)	Nipponbare	Japan	+	G	-	mature anther	I
O. sativa (Temperate *japonica*)	Asominori	Japan	+	G	-	mature anther	I
O. sativa (Temperate *japonica*)	FL7	Japan	+	G	-	seedling	I
O. sativa (Temperate *japonica*)	FL102	Japan	+	G	-	seedling	I
O. sativa (Temperate *japonica*)	Taichung 65	Japan	+	G	-	mature anther	I
O. sativa (Temperate *japonica*)	Koshihikari	Japan	+	G	-	seedling	I
O. sativa (Temperate *japonica*)	Shinriki§	Japan	+	G	-	seedling	I
O. sativa (Tropical *japonica*)	Dinalaga	Philippines	+	G	-	seedling	I
O. sativa (Tropical *japonica*)	Nanjing Xiang Dao	China	+	G	-	seedling	I
O. sativa (Tropical *japonica*)	Ketan Nangka	Indonesia	+	G	-	seedling	I
O. rufipogon	W1958	Philippines	+	A	+	seedling	I
O. rufipogon	W1945	China	+	A	+	seedling	I
O. sativa (*indica*)	T0421	India	+	A	+	seedling	II
O. sativa (*indica*)	T0710	China	+	A	+	seedling	II
O. sativa (*indica*)	T0724	China	+	A	+	seedling	II
O. sativa (*indica*)	Dao Ren Qiao	China	+	A	+	seedling	II
O. sativa (*indica*)	IR24	Philippines	+	A	+	mature anther	II
O. rufipogon	W0120	India	+	A	+	seedling	II
O. rufipogon	W0106	India	-	A	+	mature anther	III
O. rufipogon	W1681	India	-	A	+	seedling	III
O. rufipogon	W1866	Thailand	-	A	+	mature anther	III
O. rufipogon	W0610	Burma	-	A	+	seedling	III
O. sativa (*indica*)	Kasalath	India	-	A	+	mature anther	III
O. sativa (*indica*)	Dular	India	-	A	+	seedling	III
O. sativa (*indica*)	N22	India	-	A	+	seedling	III
O. sativa (*indica*)	9311	China	-	A	+	mature anther	III
O. sativa (*indica*)	Kinandang puti	Philippines	-	A	+	mature anther	III
O. rufipogon	W2003	India	+	A	+	mature anther	IV
O. rufipogon	W1294	Philippines	+	A	+	seedling	IV
O. rufipogon	W1921	Thailand	-	A	+	mature anther	IV
O. rufipogon	W1236	New Guinea	+	A	+	seedling	IV
O. glaberrima	C0025	Guinea	-	A	+	seedling	-
O. glaberrima	C8863	Nigeria	-	A	+	seedling	-
O. barthii	W1588	Cameroon	-	A	+	seedling	-
O. barthii	W0698	Guinea	-	A	+	seedling	-
O. glumaepatula	W1169	Cuba	-	A	+	seedling	-
O. glumaepatula	W1187	Brazil	-	A	+	seedling	-
O. longistaminata	W1508	Madagascar	+	A	+	seedling	-
O. longistaminata	W1540	Congo	+	A	+	seedling	-
O. meridionalis	W1625	Australia	+	A	+	seedling	-
O. meridionalis	W1635	Australia	+	A	+	seedling	-
O. officinalis	W0002	Thailand	-	A	+	seedling	-

*For each accession, RT-PCR as well as genome sequencing of *DPL*s were performed. The positions of gene specific primers and their sequences for RT-PCR analyses are shown as in Fig. I-7 and Table I-1. Each total RNA was extracted from seedling or mature anther.

†The predicted branch site at -23 bp upstream of the 3' splice site in the second intron (shown as arrowhead in Fig. I-7). Within the intron, the splice sites and branch site (mainly adenine) are required for splicing (Lal *et al.* 1999). Compare with *DPL2* orthologs in 42 rice accessions/varieties, this nucleotide change coincided with the splicing of the second intron of *DPL2.* No other coincidence was observed between nucleotide sequences and the splicing of the second intron of *DPL2* in the determined sequences.

‡*O. rufipogon* and *O. sativa* accessions/varieties were divided into four groups (I-IV) based on variations in the coding sequence of *DPL1* and *DPL2* (see Fig. II-4).

§The rice seeds were provided from Kyushu University.

Summary

Reproductive isolation is a genetic mechanism to isolate species. Identification of genes responsible for barriers to gene flow between two species provides insight into molecular mechanisms of reproductive isolation and relationships between evolution of barrier genes and species diversification. To investigate the molecular mechanism of reproductive isolation and the status of evolutionary differentiation, Asian cultivated rice, *Oryza sativa* L., is an ideal species due to its population structure and genetic diversity. *O. sativa* has diverged subspecies, *indica* and *japonica*. Inter-subspecific cross between them exhibits various reproductive isolations, but these mechanisms are still largely unknown. The cross between *indica* cultivar (cv.) Kasalath and *japonica* cv. Nipponbare showed almost all fertile F_1 population, although 33 reproductive barriers were mapped along whole chromosomes in their F_2 population. In these barriers, a prominent interactive barrier locus was detected on rice chromosomes 1 and 6. It was also confirmed that this interaction occurs only in the hybrid pollen by the analysis of reciprocal backcrosses progenies.

To identify the causal genes at each locus, map-based cloning of a pair of reproductive barrier genes has been done. Using more than 10,000 individual plants, causal genes were mapped within regions of 59 kb on Nipponbare chromosome 1 and 11 kb on Nipponbare chromosome 6. A pair of genes, one from each region shared a high degree of homology with each other, and both genes have different sequences between Nipponbare and Kasalath. These homologous genes were regarded as primary candidates, and these were named as *DOPPELGANGER1* (*DPL1*) and *DOPPELGANGER2* (*DPL2*), respectively. Hybrid pollen carrying both alleles on Kasalath

chromosome 1 (*DPL1-K*) and Nipponbare chromosome 6 (*DPL2-N*) together became non-functional, and did not germinate.

DPL genes encode plant specific protein with unknown function, which are highly conserved among angiosperms. Sequence analysis of the Nipponbare and Kasalath genomes and their transcripts suggested that alleles on Nipponbare chromosome 1 (*DPL1-N*) and Kasalath chromosome 6 (*DPL2-K*) had the same coding sequence structure. In contrast, alleles on Kasalath chromosome 1 (*DPL1-K*) and Nipponbare chromosome 6 (*DPL2-N*) had structural differences from the above two alleles. *DPL1-K* had an insertion of a predicted transposable element (TE) in the coding sequence and the transcript could not be detected in any tissues. The transcript of *DPL2-N* was a read through product of the second intron generating a premature stop codon. Higher expression of *DPL*s in pollen was also observed in *in situ* hybridization experiments. Anti-DPL antibodies recognized proteins of DPL1-N and DPL2-K in extracts from Nipponbare and Kasalath mature anthers, respectively. However, DPL2-N protein was not detected in extracts from Nipponbare. The lack of *DPL1-K* transcript and the absence of DPL2-N protein suggested that both *DPL1-K* and *DPL2-N* were loss of function alleles. Phenotype observation also indicated that *DPL1-K* and *DPL2-N* were loss of function alleles, due to the hybrid pollen carrying both of them became non-functional, and did not germinate.

The relatively high expressions of both *DPL1-N* and *DPL2-K* were observed in pollen at the late stage of pollen development. Both *DPL1-N* and *DPL2-K* were thought to have normal functions, because they were normally transmitted to progenies. Complementation tests of *DPL*s using near isogenic lines also indicated that *DPL*s are responsible genes for this reproductive

isolation, and either *DPL1-N* or *DPL2-K* are necessary for pollen transmission. These results clearly showed that a functional *DPL1-N* or *DPL2-K* allele is essential for pollen transmission, whereas *DPL1-K* and *DPL2-N* are loss-of-function alleles that act as a pair of reproductive barrier genes by their combination in hybrid pollen.

In this study, the molecular mechanism of male gametophytic reproductive isolation by disrupted DPLs in rice was elucidated. After gene duplication of *DPL*, an ancestral population seems to have diverged forming the Kasalath ancestral population, which subsequently lost the function of *DPL1* by TE insertion, and the Nipponbare ancestral population, which lost the function of *DPL2* by means of a splicing defect. When they met again by crossing, hybrid pollen having the loss of function alleles together became non-functional and failed to transmit themselves to the next generation. This is a typical case of the Dobzhansky-Muller model for barrier formation by genetic incompatibility between species.

To discuss when this reproductive isolation mechanism was established, the duplications and disruptions of *DPL*s were also investigated along with flowering species differentiation. *DPL* was highly conserved among angiosperms. Database search indicated that five different species, *Oryza sativa*, *Sorghum bicolor*, *Zea mays*, *Glycine max* and *Medicago truncatula* had two *DPL* orthologs. Using syntenic information of them, it was suggested that the duplication of *DPL* occurred at least three times, twice in Poaceae and once in Leguminosae. It was suggested that a multiple duplication of *DPL* in angiosperms, and an original copy in cereal species was *DPL2*.

In the genus of *Oryza*, all examined 42 accessions or varieties belonging to eight closely related species had both *DPL1* and *DPL2*. To investigate relationships between the disruption of

*DPL*s and *Oryza* differentiation, it was investigated when the disruptions of *DPL*s occurred in these species including both *O. sativa* and its ancestral species, *O. rufipogon*. Since the nucleotide variations in the coding region of *DPL*s, *O. sativa* and *O. rufipogon* accessions or varieties could be classified into following four groups; group I, tropical and temperate *japonica* and *O. rufipogon*; group II, *indica* and *O. rufipogon*; group III, *indica* and *O. rufipogon* and group IV, *O. rufipogon*. The insertion of TE in *DPL1* was only observed both *indica* and *O. rufipogon* belonging to group III, whereas the read through product of *DPL2* was only observed in *japonica* cultivars belonging to group I. These analyses indicated the relevant timing with an establishment of this reproductive isolation among rice differentiating lineages, and their functional roles in rice reproduction. *DPL1-K* was only observed in the partial indica varieties, suggesting the loss-of-function of *DPL1* in *indica* and of *DPL2* in *japonica* occurred after *japonica-indica* differentiation. It suggested that reproductive isolation by the combination of disrupted *DPL1-K* and *DPL2-N* is not initially act as barriers between *japonica* and *indica*, but functions in maintaining *japonica-indica* subspecies integrity.

Our results also suggested that *indica* rice is polyphyletically domesticated from the different ancestors of *O. rufipogon*, whereas *japonica* rice is monophyletic population. Other gene disruptions of *DPL*s also occurred at least three times independently in rice, suggesting the evolution of *DPL* genes according to species diversification.

These findings showed the molecular mechanisms of reproductive isolation by the combination of disrupted *DPL*s, and a comprehensive story of the evolution of *DPL*s in plant. It remains unknown whether duplications and disruptions of *DPL* genes resulted from adaptive

selection or random drift in plant speciation. Further studies of *DPL* function and analyses of reproductive isolation events in *Oryza* will provide fundamental understanding of molecular functions in plant reproduction and the mechanisms of species diversification.

Acknowledgements

I am grateful to Dr. Nori Kurata (National Institute of Genetics and School of Life Science, Mishima, Japan and School of Life Science, the Graduate University for Advanced Studies) and Dr. Yoshiyuki Harushima (Transdisciplinary Research Integration Center, Research Organization of Information and Systems, Mishima, Japan) for invaluable advices and encouragements to accomplish this research. I gratefully thank Drs. Akira Kanazawa (Research Faculty of Agriculture, Hokkaido University, Sapporo, Japan), Toshiyuki Takano, Hiroyuki Araki, Yasushi Hiromi, Tetsuji Kakutani, Hiroyuki Sasaki and Toshihiko Shiroishi (National Institute of Genetics and School of Life Science, Mishima, Japan and School of Life Science, the Graduate University for Advanced Studies, Hayama, Japan) and Tetsu Kinoshita (Plant Reproductive Genetics, GCOE Research Group, Graduate School of Biological Science, Nara Institute of Science and Technology, Ikoma, Japan) for suggestions and critically reviewing this thesis.

I thank Dr. Takashi Matsumoto, National Institute of Agrobiological Sciences, Tsukuba, Japan (NIAS) for providing Kasalath genome sequence within the mapped region on chromosome 1. I also thank Dr. Akira Kawabe, Division of Population Genetics, National Institute of Genetics, for kindly suggestions. I express my sincere thanks to Dr. Ryo Ishikawa (Faculty of Agriculture, Kobe University, Kobe, Japan) for their advices and encouragement. I express my sincere thanks to Drs. Masahiro Fujita and Yoko Horiuchi, Plant Genetics Laboratoly, National Institute of Genetics, for kindly providing me the result of microarray analyses. I also thank Takako Mochizuki for supporting the detection of transposable element in the Nipponbare genome. I would like to express

my appreciations to all member of the Plant Genetics Laboratory and Experimental farm, National Institute of Genetics for their invaluable suggestions and encouragements on this work.

Lastly, I would like to appreciate heartily to my family for full support to me.

Notes on Appendix: The definition of pollen development stages

Introduction

A genome-wide expression profile in plant species achieves important information for the study of transcriptional networks (Jiao *et al.,* 2009; Birnbaum *et al*. 2003) and genome analysis (Kikuchi *et al*., 2003). Recently, various technologies for plant transcriptome analyses such as microarray, MPSS and SAGE have been developed (Rensink *et al*., 2005). A large scale transcriptome analysis was well described in *Arabidopsis thaliana*, but little is known about rice (Jiao *et al.,* 2009; Nobuta, *et al*. 2007). In addition to that, a transcriptome analysis has not been compiled to date for reproductive organs and stages in rice.

To investigate transcriptional networks during rice reproduction, in my laboratory, expression atlas by microarray analysis was attempted with dissected developmental stages from Nipponbare tissues including reproductive stages and organs by microarray analysis. In these experiments, it was necessary to define the developmental stages for collection of a flesh sample, due to the short-lived rice reproductive organs.

In this note, the definition of pollen development stages to collect anthers for microarray analysis is described. To characterize a relationship between tissue length and developmental stage of anther, pollen development was investigated. My result would help to collect and use rice anther for microarray analysis and other rice studies.

Materials and Methods

Plant materials and phenotype observation of pollen

The reproductive organs of Nipponbare were collected from plants grown in paddy fields under normal condition. The length of flower and anther was investigated under a stereomicroscope (OLYMPUS, Tokyo, Japan). Pollen grains were spread on a slide glass, and their morphology was observed under a differential interference microscope (OLYMPUS, Tokyo, Japan).

Results

The stage definition during anther development

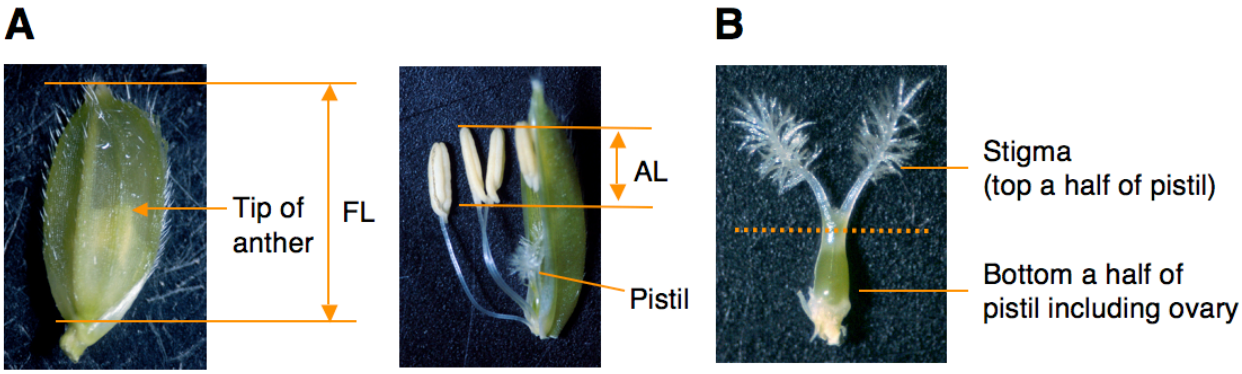
Rice reproductive organs and the length of tissues are shown in Appendix Fig. 1. To characterize a relationship between tissue length and developmental stage of anther, the length of tissues in 27 independent flowers were investigated (Appendix Table 1). In rice, pollen is generated from pollen mother cells in anthers (Appendix Table 2). During the early developmental stage, the young microspore has only one nucleus (Uni-nucleus stage). The first nuclear division produces a large vegetative cell and a generative cell in the cytoplasm along pollen wall (Bi-cellular stage). Then, the generative cell undergoes a second mitosis to form two sperm cells (Tri-cellular stage). Subsequently, the filling of starch in pollen grains is completed and the pollen cytoplasm is full of larger starch granules at tri-cellular stage. Based on these morphological differences, pollen development was observed according to the length of tissues and the position of anther in a flower

(Appendix Table 1). Phenotype observation of pollen indicated that the value of AP (The length of the tip of anther from the bottom per flower length) followed gradual pollen development, and becomes as good indicator to define the developmental stage. When the value of AP was 33-43%, 41-53% and 59-79%, the anther included pollen in P1 (anther including pollen at uni-nucleus stage), P2 (anther including pollen at bi-nuclear stage) and P3 stages, respectively (Appendix Table 1). Meanwhile, the length of flower (FL) and anther (AL) could not be used to determine the stage due to no significance differences among P1-P3 stages. Based on these results, a minimum range of developmental stage of P1, P2 and P3 was defined with the value of AP with 30-40%, 45-55% and 65-100%, respectively (Appendix Tables 2 and 3).

Discussion

In rice, reproductive tissues are short-lived, especially at mature pollen. The rough classification of developmental stage of anther was proposed by Itoh *et al.* (2005), although their tissue length was not appreciable to determine the P1 to P3 stages since an anther and a flower length are almost the same among P1-P3 stages. Morphological analysis of pollen was also investigated by Zhang *et al.* (2005), although investigated cultivar is different (*japonica* cv. Taichung 65) and the tissue length of each stage was not mentioned.

In this study, it was suggested that the value of AP becomes a good indicator to define the developmental stage of pollen, and to collect anthers. This result would help to collect and use anther at pollen developmental stages not only microarray analysis but also other rice studies.



Appendix Fig. 1. Reproductive organs of Nipponbare. (A) The reproductive organs in a flower. FL and AL mean the length of flower and anther, respectively. (B) A rice pistil comprises ovary and stigma. For a microarray experiment, the tissue of un-pollinated and pollinated stigma was defined as the top a half of pistil. Bottom a half of pistil including ovary was also used. Each tissue and stage definition was summarized in Appendix. 4.

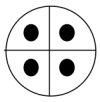
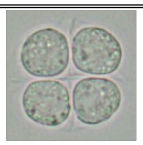
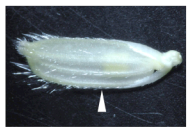
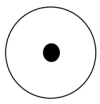
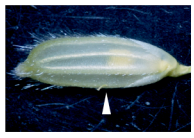
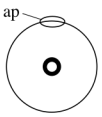
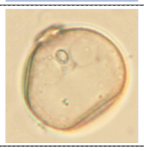
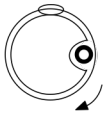
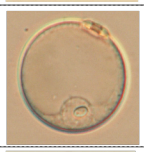
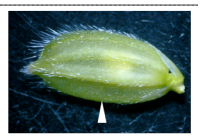
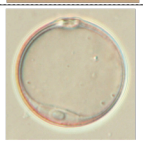
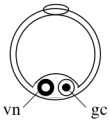
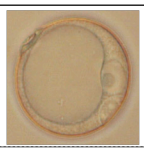
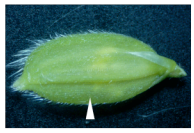
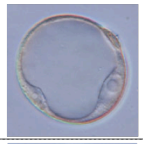
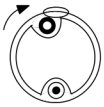
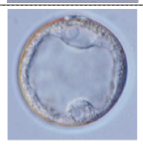
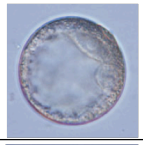
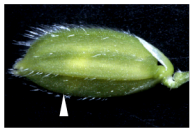
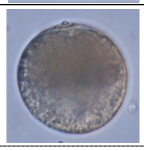
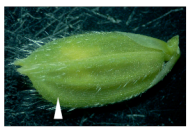
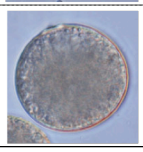
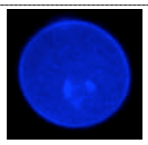
Appendix Table 1 The tissue length and stages of pollne in Nipponbare anther.

Sample	The length of tissues (mm)			Anther position (AP)*		Stage†
	Infrorescent	Flower	Anther	Length (mm)	(%)	
1	140	4.8	0.9	1.6	33	P1
2	140	6.1	1.1	2.2	36	P1
3	118	5.5	1.2	2.0	36	P1
4	140	4.7	0.9	1.8	38	P1
5	152	6.5	1.4	2.5	38	P1
6	153	7.0	1.7	2.7	39	P1
7	118	6.7	1.2	2.6	39	P1
8	118	6.7	1.2	2.6	39	P1
9	140	6.4	1.2	2.5	39	P1
10	140	4.4	1.1	1.8	41	P1
11	153	6.6	1.5	2.7	41	P1
12	140	4.1	0.9	1.7	41	P1
13	153	6.9	1.7	3.0	43	P1
14	152	6.1	1.2	2.5	41	P2
15	162	6.2	1.3	2.7	44	P2
16	177	6.9	1.5	3.1	45	P2
17	162	6.8	1.6	3.1	46	P2
18	162	6.8	1.7	3.1	46	P2
19	177	6.6	1.3	3.2	48	P2
20	162	6.5	1.5	3.3	51	P2
21	175	6.4	1.6	3.4	53	P2
22	162	7.0	1.6	3.4	49	P2
23	177	7.2	1.3	3.5	49	P2
24	162	7.3	1.5	4.3	59	P3
25	162	7.0	1.5	4.3	61	P3
26	175	7.0	1.6	5.2	74	P3
27	162	7.0	1.6	5.5	79	P3

*The position of the tip of anther from the bottom in a flower, which was described in Appendix Fig. 1. The percentage indicates a rate of the length of the position of the tip of anther from the bottom per the tissue length of a flower.

†The developmental stage of pollen was determined by using differential interference microscope.

Appendix Table 2 The developmental stage of pollen with the tissue length and the position of anther in a flower.

Stage		Diagram*	Differential interference / Bright	DAPI	Stereomicroscope§	The length of tissues†			AP‡
						IL (mm)	FL (mm)	AL (mm)	(%)
Tetrad	-			-		70-100	4.0-6.0	0.8-1.1	-25
Uni-nucleus (P1)	P1-1			-		120-160	5.0-7.0	1.1-2.2	30-40
	P1-2			-					
	P1-3			-					
	P1-4			-	-				
Bi-cellular (P2)	P2-1			-		160-170	7.0	2.2	45-55
	P2-2			-	-				
	P2-3			-					
	P2-4			-					
Tri-cellular (P3)	P3-1			-		170-250	7.0	2.2	65-80
	P3-2								

*ap, aparture; vn, vegetative nucleus; gc, generative cell; and sc, sperm cell. Diagram was reffered to Zhang *et al.* (2005).

† IL, inflorescence length; FL, floret length; AL, anther length.

§ White arrowheads indicate the position of the tip of anther from the bottom in a flower.

‡AP means the length of the tip of anther from the bottom per flower length.

Appendix Table 3 Stage and organ definition in rice reproductive phase.

Stage Symbol	Biological Event	Organ	Definition
An1	Formation of hypodermal archesporial cells	Anther	Anther lengthb: 0.1-0.15 mm
Mei1	Pre-meiotic S/G2	Anther	Anther length: 0.2-0.45 mm
M1	Leptotene in Meiosis	Anther	Anther length: 0.4-0.55 mm
M2	Zygotene to Pachytene in Meiosis	Anther	Anther length: 0.45-0.8 mm
M3	Diplotene to Terad in Meiosis	Anther	Anther length: 0.75-1.10 mm
P1	Uni-nuclear pollen	Anther	30 < AP < 40
P2	Bi-cellular pollen	Anther	45 < AP < 55
P3	Tri-cellular mature pollen	Anther	65 < AP < 100
St 0	Stigma Control	Stigma (top a half of pistil)	Before pollination
Ov 0	Ovary Control	Bottom a half of pistil including ovary	Before pollination
Poll 1	Pollination and pollen germination	Stigma (top a half of pistil)	5 min after pollination
Ov 1	Pollination and pollen germination	Bottom a half of pistil including ovary	5 min after pollination
Poll 2	Pollen tube growth	Pistil	15-25 min after pollination
Fer1	Fertilization	Pistil	40-50 min after pollination
Fer2	Zygote formation	Pistil	5-7 hrs after pollination

References

Bhattacharyya, M.K. *et al*. The wrinkled-seed character of pea described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme. *Cell* **60**, 115-122 (1990).

Bikard, D. *et al*. Divergent evolution of duplicate genes leads to genetic incompatibilities within *A. thaliana*. *Science* **323**, 623-626 (2009).

Bolot, S. *et al*. The 'inner circle' of the cereal genomes. *Curr. Opin. in Plant Biol*. **12**, 119-125 (2009).

Birnbaum, K. *et al*. A Gene Expression Map of the *Arabidopsis* Root. *Science* **302**, 1956-1960 (2003).

Bomblies, K. *et al*. Autoimmune response as a mechanism for a Dobzhansky-Muller-type incompatibility syndrome in plants. *PLoS Biol*. **5**, e236 (2007).

Chen, J. *et al*., A triallelic system of *S5* is a major regulator of the reproductive barrier and compatibility of *indica–japonica* hybrids in rice. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 11436-11441 (2008).

Cheng, C. *et al*. Evolutionary relationships among rice species with AA genome based on *SINE* insertion analysis. *Genes Genet. Syst*. **77**, 323-334 (2002).

Cheng, C. *et al*. Polyphyletic origin of cultivated rice: based on the interspersion pattern of *SINE*s. *Mol. Biol. Evol*. **20**, 67-75 (2003).

Clark, N. *et al*., Coevolution of Interacting Fertilization Proteins. *PLoS Genet*. **5**, e1000570 (2009).

Cresti, M., Cai, G., Moscatelli, A. *Fertilization in Higher Plants – Molecular and Cytological Aspects* -. (Springer Pr: 1999)

Cole, R.A. *et al*. *SEC8*, a subunit of the putative *Arabidopsis* exocyst complex, facilitates pollen germination and competitive pollen tube growth. *Plant Physiol* **138**, 2005-2018 (2005).

Collard, B.C.Y. *et al*. Evaluation of 'quick and dirty' DNA extraction methods for marker-assisted selection in rice *Oryza sativa* L., *Plant Breed*. **126**, 47-50 (2007).

Coyne, J.A., Orr, H. A. *Speciation*. (Sinauer Associates Inc., 2004).

Dobzhansky, T. *Genetics and the Origin of Species*. (Columbia Univ Press, 1982).

Ferris, P. J., Pavlovic, C., Fabry, S., Goodenough, U. W. Rapid evolution of sex-related genes in *Chlamydomonas*. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 8634-8639 (1997).

Garris, A. J. *et al*. Genetic structure and diversity in *Oryza sativa* L. *Genetics* **169**, 1631-1638 (2005).

Gao, L., Innan, H. Nonindependent Domestication of the Two Rice Subspecies, *Oryza sativa* ssp. *indica* and ssp. *japonica*, Demonstrated by Multilocus Microsatellites. *Genetics* **179**, 965-976 (2008).

Golovkin, M., Reddy, A.S.N. A calmodulin-binding protein from *Arabidopsis* has an essential role in pollen germination. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 10558-10563 (2003).

Gilbert, W., de Souza, S. J., Long, M. Origins of genes, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 7698-7703 (1997).

Gu, Z. *et al*. Role of duplicate genes in genetic robustness against null mutations. *Nature* **421**, 63-66 (2003).

Han, M. *et al. Rice Immature Pollen 1* (*RIP1*) is a regulator of late pollen development. *Plant Cell Physiol* **47**, 1457-1472 (2006).

Harlow E., Lane D. *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, 1988).

Harushima, Y. *et al*. A high-density rice genetic linkage map with 2275 markers using a single F_2 population. *Genetics* **148**, 479-494 (1998).

Harushima, Y. *et al*. A genome-wide survey of reproductive barriers in an intraspecific hybrid. *Genetics* **159**, 883-892 (2001).

Harushima, Y. *et al*. Diverse variation of reproductive barriers in three intraspecific rice crosses. *Genetics* **160**, 313-322 (2002).

Herrmann, B. G. *et al*. A protein kinase encoded by the *t complex responder* gene causes non-mendelian inheritance. *Nature* **402**, 141-146 (1999).

Ikehashi, H., Araki, H. Genetics of F_1 sterility in remote crosses of rice. *IN: Rice genetics*, International Rice Research Institute, Manila, Philippines. 119-130 (1986).

International Rice Genome Sequencing Project. The map-based sequence of the rice genome International Rice Genome Sequencing Project. *Nature* **436**, 793 (2005).

Ishikawa, R. *et al*. Suppression of the floral activator *Hd3a* is the principal cause of the night break effect in rice. *Plant Cell* **17**, 3326-3336 (2005).

Ito, N., Kurata, N. Disruption of *KNOX* gene suppression in leaf by introducing its cDNA in rice. *Plant Sci*. **174**, 279-289 (2008).

Itoh, J. *et al*. Rice Plant Development: from Zygote to Spikelet. *Plant Cell Physiol*. **46**, 23-47 (2005).

Jiao, Y. *et al*. A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies. *Nat. Genet*. **41**, 258-263 (2009).

Kato, S. On the affinity of the cultivated rice varieties of rice plants, *Oryza sativa* L. *Jour. Dept. Agr. Kyushu Imp. Univ.* **2**, 241-276 (1930).

Khush, G.S. Origin, dispersal, cultivation and variation of rice. *Plant Mol. Biol*. **35**, 25-34 (1997).

Kikuchi, S. *et al*. Collection, Mapping, and Annotation of Over 28,000 cDNA Clones from *japonica* Rice. *Science* **301**, 376-379 (2003).

Kitamura, E. Genetic studies on sterility observed in hybrids between distantly related varieties of rice, *Oryza sativa* L. *Bull. Chugoku Aric. Exp. Station* **A8**, 141-205 (1962).

Koide, Y. *et al*. The evolution of sex-independent transmission ratio distortion involving multiple allelic interactions at a single locus in rice. *Genetics*. 108.090126 (2008). doi:10.1534.

Kouchi, H., Hata, S. Isolation and characterization of novel nodulin cDNAs representing genes expressed at early stages of soybean nodule development. *Mol. Gen. Genet*. **238**, 106-119 (1993).

Kurata, N. *et al*. A 300 kilobase interval genetic map of rice including 883 expressed sequences. *Nat. Genet*. **8**, 365-372 (1994).

Lal, S. *et al*. A splice site mutant of maize activates cryptic splice sites, elicits intron inclusion and exon exclusion, and permits branch point elucidation. *Plant Physiol*. **121**, 411-418 (1999).

Lee, H. *et al*. Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. *Cell* **135**, 1065-1073 (2008).

Liu, Y.S. *et al*. Mapping QTLs for defective female gametophyte development in an inter-subspecific cross in *Oryza sativa* L. *Theor. Appl. Genet*. **102**, 1243-1251 (2001).

Long, M., Langley, C. Natural selection and the origin of jingwei, a chimeric processed functional gene in *Drosophila*. *Science* **260**, 91-95 (1993).

Long, Y. *et al*. Hybrid male sterility in rice controlled by interaction between divergent alleles of two adjacent genes. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 18871-18876 (2008).

López-Fernández, H., Bolnick, D.I. What Causes Partial F_1 Hybrid Viability? Incomplete Penetrance versus Genetic Variation. *PLoS ONE* **2**, e1294 (2007).

Masly, J. P. *et al*. Gene Transposition as a Cause of Hybrid Sterility in *Drosophila*. *Science* **313**, 1448-1450 (2006).

McNally, K. L. *et al*. Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 12273-12278 (2009).

Mihola, O. *et al*. A mouse speciation gene encodes a meiotic Histone H3 methyltransferase. *Science* **323**, 373-375 (2009).

Mizuta, Y. *et al*. Mapping of a pair of reproductive barrier loci observed in a cross between Nipponbare and Kasalath. *Rice genetics newsletter* **23**, 33-35 (2007).

Murray, M. G., Thompson, W.F. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids. Res*. **8**, 4321-4325 (1980).

Neff, M. M. *et al*. dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms:

experimental applications in *Arabidopsis thaliana* genetics. *Plant J*. **14**, 387-392 (1998).

Nishimura, A., Aichi, I., Matsuoka, M. A protocol for *Agrobacterium*-mediated transformation in rice. *Nat. Protocols* **1**, 2796-2802 (2007).

Nobuta, K. *et al*. An expression atlas of rice mRNAs and small RNAs. *Nat. Biotech*. **25**, 473-477 (2007).

Noor, M.A.F., Feder, J.L. Speciation genetics: evolving approaches. *Nat. Rev. Genet*. **7**, 851-861 (2006).

Ohno, S. *Evolution by Gene Duplication* (Springer, New York, 1970).

Oka, H. I. Analysis of genes controlling F_1 sterility in rice by the use of isogenic lines, *Genetics* **77**, 521-534 (1974).

Oka, H. I. *Origin of Cultivated Rice* (Japan Sci. Soc. Press, Tokyo, 1988).

Oka, H. I. The inviability, weakness, and sterility of interspecific hybrids. *Adv. Genet*. **9**, 147-215 (1958).

Oliver, P.L. *et al*. Accelerated evolution of the *Prdm9* speciation gene across diverse metazoan taxa. *PLoS Genet*. **5**, e1000753 (2009).

Orr, H.A., Masly, J.P., Presgraves, D.C. Speciation genes. *Curr. Opin. Genet. Dev*. **14**, 675-679 (2004).

Paterson, A. H., Bowers, J. E., Chapman, B. A. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc. Natl. Acad. Sci. U.S.A*. **101**, 9903-9908 (2004).

Paterson, A. H. *et al*. The *Sorghum bicolor* genome and the diversification of grasses. *Nature* **457**, 551-556 (2009).

Phadnis, N., Orr, H.A. A single gene causes both male sterility and segregation distortion in *Drosophila* hybrids. *Science* **323**, 376-379 (2009).

Rensink, W.A., Buell, C.R. Microarray expression profiling resources for plant genomics. *Trends in Plant Science* **10**, 603-609 (2005).

Saitou, N., Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol*. **4**, 406-425 (1987).

Salse, J. *et al*. Reconstruction of monocotelydoneous proto-chromosomes reveals faster evolution in plants than in animals. *Proc. Natl. Acad. Sci. U.S.A*. **106**, 14908-14913 (2009).

Sato, Y. Morishima, H. Studies on the distribution of complementary genes causing F_1 weakness in common rice and its wild relatives. 2. Distribution of two complementary genes *Hwc-1* and *Hwc-2* genes in native cultivars and wild relatives of tropical Asia. *Euphytica* **36**, 425-431 (1987).

Schagger, H., Tricine-SDS-PAGE. *Nat. Protocols* **1**, 16-22 (2006).

Seidel, H.S., Rockman, M.V., Kruglyak, L. Widespread genetic incompatibility in *C. Elegans* maintained by balancing selection. *Science* **319**, 589-594 (2008).

Stebbins, G.L. The inviability, weakness, and sterility of interspecific hybrids. *Adv. Genet*. **9**, 147-215 (1958).

Ting, C. *et al*. A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* **282**, 1501-1504 (1998).

Tuskan, G.A. *et al*. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**, 1596-1604 (2006).

Vandepoele, K., Simillion, C., Van de Peer, Y. Evidence that rice and other cereals are ancient aneuploids. *Plant Cell* **15**, 2192-2202 (2003).

Vaughan, D. A., Morishima, H., Kadowaki, K. Diversity in the *Oryza* genus. *Curr. Opin. in Plant Biol*. **6**, 139-146 (2003).

Weis, S., Schartl, M. The macromelanophore locus and the melanoma oncogene *Xmrk* are separate genetic entities in the genome of *Xiphophorus*. *Genetics* **149**, 1909-1920 (1998).

Wittbrodt, J. *et al*. Novel putative receptor tyrosine kinase encoded by the melanoma-inducing *Tu* locus in *Xiphophorus*. *Nature* **341**, 415-421 (1989).

Xu, J. *et al*. Phylogenetic analysis of *Oryza rufipogon* strains and their relations to *Oryza sativa* strains by insertion polymorphism of rice *SINE*s. *Genes Genet. Syst*. **82**, 217-229 (2007).

Yu, J. *et al*. The Genomes of *Oryza sativa*: A history of duplications. *PLoS Biol*. **3**, e38 (2005).

Zhao, W. *et al*. BGI-RIS: an integrated information resource and comparative analysis workbench for rice genomics. *Nucleic Acids Res*. **32**, D377-382 (2004).

Zhang, Z. *et al.* Nuclear and cell migration during pollen development in rice (*Oryza sativa* L.). *Sexual Plant Reproduction* **17**, 297-302 (2005).