Theor Appl Genet (2011) 122:373–383
DOI 10.1007/s00122-010-1453-z

ORIGINAL PAPER

Two complementary recessive genes in duplicated segments control etiolation in rice

Donghai Mao · Huihui Yu · Touming Liu · Gaiyu Yang · Yongzhong Xing

Received: 5 June 2010 / Accepted: 8 September 2010 / Published online: 26 September 2010

Abstract The main objective of this study was to identify the genes causing etiolation in a rice mutant, the thylakoids of which were scattered. Three populations were employed to map the genes for etiolation using bulked segregant analysis. Genetic analysis confirmed that etiolation was controlled by two recessive genes, *et11* and *et12*, which were fine mapped to an approximately 147-kb region and an approximately 209-kb region on the short arms of chromosomes 11 and 12, respectively. Both regions were within the duplicated segments on chromosomes 11 and 12. They possessed a highly similar sequence of 38 kb at the locations of a pair of duplicated genes with protein sequences very similar to that of HCF152 in *Arabidopsis* that are required for the processing of chloroplast RNA. These genes are likely the candidates for *et11* and *et12*. Expression profiling was used to compare the expression patterns of paralogs in the duplicated segments. Expression profiling indicated that the duplicated segments had been undergone concerted evolution, and a large number of the paralogs within the duplicated segments were functionally redundant like *et11* and *et12*.

Communicated by Y. Xue.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-010-1453-z) contains supplementary material, which is available to authorized users.

D. Mao · H. Yu · T. Liu · G. Yang · Y. Xing (✉)
National Key Laboratory of Crop Genetic Improvement and National Center of Plant Gene Research (Wuhan), Huazhong Agricultural University, Wuhan 430070, China
e-mail: yzxing@mail.hzau.edu.cn

Introduction

Since it began as a cyanobacterial symbiont, approximately 1–1.5 billion years ago, the chloroplast has become an indivisible portion of its host eukaryotes, and its functions underpin the global ecosystem (Douzery et al. 2004). Photosynthesis takes place within the chloroplast, and the photosynthetic process comprises two distinctive steps. One is the light-dependent reaction, which occurs on the thylakoid membrane. Light energy excites the chlorophyll pigments, which are bound by protein complexes, such as photosystem I and photosystem II, initiates electron flow, and generates ATP. The other step takes place in the chloroplast stroma, where ATP is used to fix carbon dioxide by Rubisco to generate sugar, the food source depended on by all beings either directly or indirectly (Blankenship 2002).

Nearly, every aspect of the chloroplast has been the subject of biological research (for review, Jung and Chory 2010). The chlorophyll biosynthetic pathway from glutamyl-tRNA to chlorophylls *a* and *b* requires 15 steps in the higher plants. Genes encoding enzymes for all steps in *Arabidopsis* have been identified, such as *HEMA1* to *HEMA3* encoding glutamyl-tRNA reductase; *GSA1* and *GSA2* for glutamate 1-semialdehyde aminotransferase; *HEMB1* and *HEMB2* for porphobilinogen synthase; *CHLD*, *CHLL*, *CHLI1*, and *CHLI2* encoding subunits of Mg chelatase, and *CAO* for the final enzyme chlorophyllide, an oxygenase (Beale 2005). The basic mechanism of light signaling for photomorphogenesis has been clarified in *Arabidopsis*. The structures and functional states of dimers of phytochrome, such as CRY1, are changed by the transition between day and night. The dimers then cooperate with the COP10/DET1/DDB1 complexes to determine the fates of transcription factors, such as positive regulator

HY5 or negative regulator PIF3, through ubiquitination and subsequent degradation. The transcription factors, which are not degraded, bind onto promoters of photomorphogenesis genes to promote or repress the expression of these genes. Most of the genes involved, such as *CRY1*, *COP1*, *DET1*, *DDB1*, *COP10*, and *PIFs*, have been cloned in *Arabidopsis thaliana* (Waters and Langdale 2009). Reflecting its complexity, chloroplast gene expression requires four major post-transcriptional processes: RNA processing, editing, splicing, and turnover. For chloroplast RNA editing alone, 17 proteins have been identified as involved in the editing efficiency of *Arabidopsis* (Stern et al. 2010).

Although most genes in these pathways have been cloned and their molecular mechanisms have been characterized in the model dicot plant, *A. thaliana*, few studies have examined the counterparts in monocot plants (Zhang et al. 2006; Takano et al. 2009). *Oryza sativa* is a model plant for monocots, especially for the grass families, and this species has large genomic and phenotypic differences from *Arabidopsis*. For example, the size of rice genome is about 2.7 times larger than the *Arabidopsis* genome, and rice has higher copy numbers of many gene families than *Arabidopsis* (AGI 2000; IRGSP 2002; Yu et al. 2002; Bennetzen et al. 2004). The phytochrome gene family is small, but important for making the chloroplast. The *Arabidopsis* phytochrome family contains five members, *phyA* to *phyE*, and rice possesses only three, *phyA* to *phyC* (Takano et al. 2009; Franklin and Quail 2010). The difference in the number of phytochrome genes between *Arabidopsis* and rice has been caused by duplication within their genomes and subsequent natural selection (Clack et al. 1994). Further studies of chloroplast development are needed in monocot plants.

Mutants play an important role in the study of genetics and molecular biology. Since the late 1950s, *Arabidopsis* mutation databases have been built by the EMS method (Muller 1961). Along with highly efficient transformation, random insertion libraries have been generated by the method of T-DNA insertion or other transposon systems (McElver et al. 2001; Marsch-Martinez et al. 2002). Because *Arabidopsis* enjoys a huge range of genetic and genomic resources, genetic and molecular studies have progressed much further in *Arabidopsis* than in other species (Koornneef and Meinke 2010). In contrast, rice mutant libraries have been constructed only for the past several years (Jeon et al. 2000; Chen et al. 2003; Wu et al. 2003).

Besides artificial mutants, natural mutants provide novel resources and play important roles in gene isolation. In this study, an etiolating mutant in chloroplast biosynthesis was found within a rice recombinant inbred line RI184 from the cross between 02428 and Teqing; the mutant grew pale yellow leaves and subsequently died at the seedling stage. We constructed F_2 and F_3 populations by crossing the RI184 plant the progenies of which segregated with wild types and etiolation mutants with Zhenshan 97 to fine map the genes accounting for the etiolation. Two genes controlling etiolation were located in a pair of duplicated segments on chromosomes 11 and 12, respectively.

Materials and methods

Plant materials

RI184 was one of 191 recombinant inbred lines in an F_6 population derived by single-seed descendent from a cross between an *indica* variety, Teqing (TQ), and a widely compatible *japonica* variety, 02428. The seedlings of RI184 were segregated by color: normal green and pale yellow (etiolated). Six normal green plants were selfed to get six families. Four of the six families were segregated by normal and etiolated leaves. The four plants each produced segregated families in color were crossed with ZS97 to obtain F_1 plants. Their F_1 plants were self-crossed to attain four potential F_2 segregation populations for gene mapping (Fig. 1).

Measurement of chlorophyll

Contents of chlorophyll *a* and *b* (Chl *a* and Chl *b*) were dynamically measured according to the method of Arnon (1949) with some modifications. Chlorophyll contents were measured every 2 days from 8 to 18 days after germination, because the etiolated seedlings were too small to remove leaves before 8 days after germination and became withered by 18 days after germination.

Fresh leaves were cut into pieces and 0.015 g of the material was placed into 1.5 ml of the chlorophyll extraction solution (volume ratio of 95% ethanol: pure acetone: water = 4.5:4.5:1) and placed in the dark at 4°C for 15 h. Using the chlorophyll extraction solution as the control, the chlorophyll solution absorbance values were recorded at 663- and 645-nm wavelengths using a spectrophotometer (DU640 manufacturer, city, country). The chlorophyll contents of each genotype were measured using three replicates. Contents of Chl *a* and Chl *b* were calculated by the following equations:

$$\text{Chl } a = [12.72(\text{A663}) - 2.59(\text{A645})] \times V/(1000 \times W) \quad (1)$$

$$\text{Chl } b = [22.88(\text{A645}) - 4.67(\text{A663})] \times V/(1000 \times W) \quad (2)$$

where Chl *a* and Chl *b* represent the contents of Chl *a* and Chl *b* in leaves (mg g^{-1}), A663 is the absorbance at 663 nm and A645 at 645 nm, *V* is the volume of the

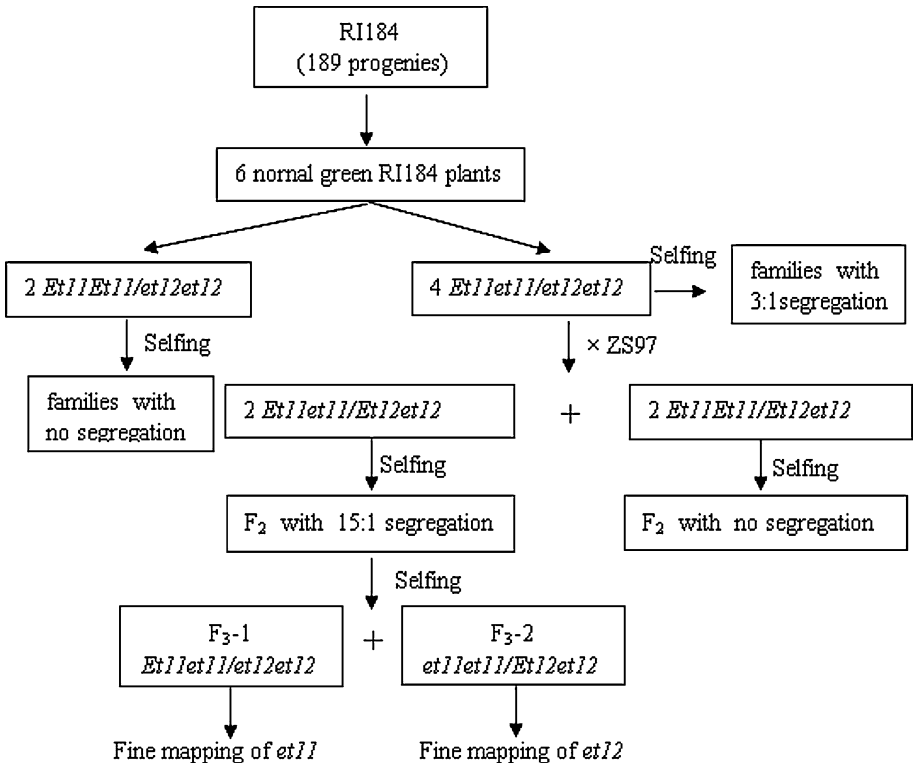

Fig. 1 The flow diagram of population development

chlorophyll extraction solution (ml), and *W* is the weight of leaves (g).

Transmission electron microscopy

Unhusked seeds were put into a 75% ethanol solution for 1 min and then a 0.15% mercuric chloride solution for 15 min for sterilization, and then seeds were washed four to six times with distilled water. The seeds were sown on Murashige and Skoog growth medium agar with 0.8% sucrose, growing under continuous light conditions (50–100 microeinsteins m^{-1} s^{-1}) at 25°C. The leaves of 12-day-old wild-type and mutant seedlings were removed for observation of the chloroplast structures by transmission electron microscopy following the method of Gothandam et al. (2005).

Genetic analysis of etiolation

To perform a primary genetic analysis of etiolation, after germinating at 37°C for 2–3 days, F_2 populations were separately sown in a seeding bed in the field during the rice growing season. Bulked segregant analysis (Michelmore et al. 1991) was carried out in one segregated F_2 population of 178 plants. Five seedlings of each phenotype were separately sampled to extract their DNA, and an equal amount of DNA from the five seedlings was mixed together as a DNA pool. Thus, in the F_2 population two DNA pools were attained for the wild type and etiolating phenotype.

Bulked segregant analysis showed that two unlinked loci were associated with the etiolating phenotype in the F_2 population, indicating that two genes, *et11* and *et12*, controlled etiolation. Plants segregated at one locus, but fixed with a mutant allele at the other locus were selected to produce F_3 populations. Two F_3 populations (F_3-1 and F_3-2) were obtained for fine mapping of the individual genes (Fig. 1).

DNA isolation and polymorphic marker screening

DNA was extracted from fresh leaves of plants from the F_2 and F_3 populations using the cetyl-trimethyl ammonium bromide method (Murray and Thompson 1980) with a few modifications. Markers of RM1–RM600 were designed according to Temnykh et al. (2000, 2001) and other RM series designed according to the rice genome sequences of the Monsanto Company (McCouch et al. 2002) were used for bulked segregant analysis. The SSR assay was conducted by polyacrylamide gel electrophoresis, as described by Wu and Tanksley (1993).

Development of new markers and gene mapping

To fine map the two genes, new markers including single nucleotide polymorphism (SNP) and insertion and deletion

(InDel) were developed in the two targeted regions. Their primers from 19 to 22 bases were designed with Tm from 60 to 74°C (GC% method) using Oligo 5.0 (National Biosciences, USA).

After comparison of sequences in the target regions between subspecies *japonica* (http://rice.plantbiology.msu.edu) and *indica* (http://rice.genomics.org.cn/rice/index2.jsp), the fragments with long polymorphisms were used for developing markers, and those with single nucleotide differences between subspecies were designed as SNP markers. SNP genotyping was done by sequencing PCR products using the Big Dye Terminator mix (PE-Applied Biosystems, USA).

Molecular linkage maps for the two genes were constructed using MAPMAKER/EXP version 3.0 (Lander et al. 1987). The linkage groups and the orders of markers on maps were determined by using the "group" and "order" commands of the MAPMAKER program. The recombination values were converted to genetic distances in centimorgans (cM) using the Kosambi function (Kosambi 1944).

To distinguish the paralog pairs on chromosomes 11 and 12, some primer pairs in the "de" series were designed to specifically amplify the duplicated regions of chromosome 12, but with no amplicon on chromosome 11. In addition, other primer pairs in the "de" series simultaneously had amplicons on chromosomes 11 and 12, but due to SNP their amplicons were distinguished by presenting two peaks at SNP sites in the sequencing electropherogram.

Sequence analysis

Gene sequences and their annotations used in this study were from the Institute for Genomic Research database (http://rice.plantbiology.msu.edu) for rice and from the NCBI database (http://www.ncbi.nlm.nih.gov) for *Arabidopsis*. For comparison, the sequences containing *et11* and *et12* on chromosomes 11 and 12 were downloaded from TIGR (http://rice.plantbiology.msu.edu/segmental_dup/100kb/segdup_100kb.shtml), and WU-blast (Washington University BLAST) was used to search the rice genome to identify highly homologous duplicate blocks (http://blast.wustl.edu). The protein sequences of candidate genes and their homologous gene in *Arabidopsis* were compared using the online software ClustalW (http://www.ebi.ac.uk/Tools/clustalw2/index.html; Thompson et al. 1994).

Microarray analysis of the duplicated genes

Sequence comparison revealed a 4.2-Mb fragment containing *et11* on chromosome 11 that was duplicated as the *et12*-containing fragment on chromosome 12. To ascertain whether the duplicated genes have the same expression patterns, the expression profile data of the duplicated genes was downloaded from our laboratory website, Collection of Rice Expression Profiles (http://crep.ncpgr.cn). The profiles were analyzed by collecting 36 tissues of 3 *indica* rice varieties, Minghui 63 (MH63), Zhenshan 97 (ZS97), and their hybrid Shanyou 63 (SY63) (Wang et al. 2010). Because nearly all rice genes have probe sets with 11 probes on the Affymetrix Gene Chip Rice Genome Array, those genes with fewer than three probes having sites targeting other genes within their probe sets were chosen to compare the expression of paralogs. The duplicate regions of 4.2 Mb had 313 pairs of paralogs. Among them, 83 pairs were distinguishable as they had probe sets with low or no sequence similarity. Pearson correlation analysis was performed to compare the expression levels of each pair of paralogous genes among the 36 tissues of the three rice varieties.

Results

Phenotype of the etiolation mutant

The mutant had a pale yellow shoot and leaves after germination. The endosperm of the mutant was nearly depleted after 18 days of growth. Finally, the etiolated seedlings withered and died. The contents of Chl *a* and *b* of the etiolated seedlings gradually decreased along with the growing stage, whereas they increased in the wild type (Fig. 2). In addition, in the wild-type chloroplast the thylakoids were stacked, whereas in the mutant thylakoids were scattered at 12 days of growth. Thus, the

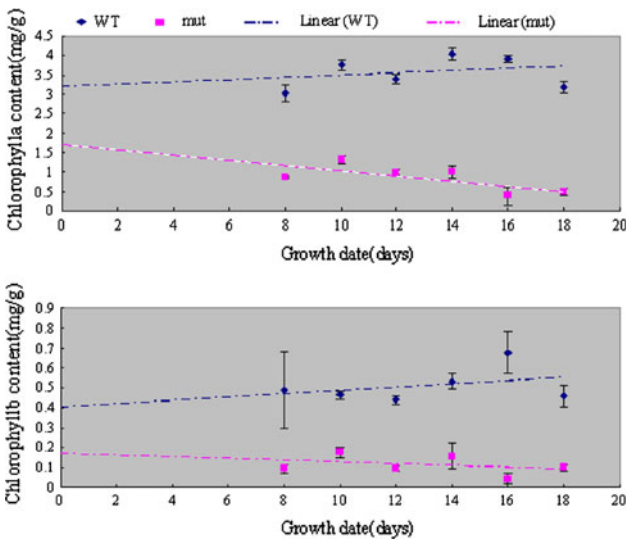

Fig. 2 Dynamic change of chlorophyll *a* and *b* contents in wild-type and etiolated seedlings

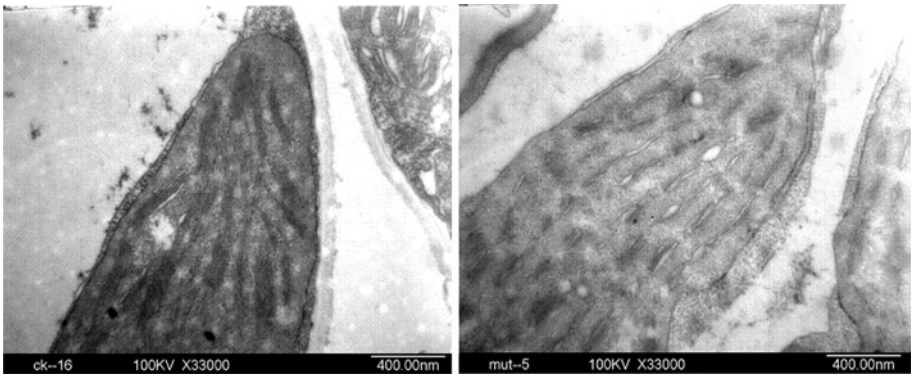

Fig. 3 Ultrastructures of chloroplast thylakoids in the wild type (*left*) and etiolated type (*right*) at the three-leaf stage

thylakoid structure was abnormal in the mutant plants (Fig. 3).

Gene mapping for etiolation

Among the 189 progenies of RI184, there were 146 green and 43 etiolated seedlings. The segregation ratio was in agreement with the expected ratio of 3:1 in the single Mendelian factor model ($\chi^2 = 0.40$, $P = 0.53$). Two of four F_2 populations segregated in seedling color. We randomly investigated one F_2 population of 178 seedlings. There were 167 normal plants and 11 mutants, which fit well the expected ratio (15:1) of a two-gene model including duplicate gene action ($\chi^2 = 0.01$, $P = 0.91$). Clearly, the etiolation was controlled by double-recessive genes. RI184 carried homozygous recessive alleles for one gene and heterozygous alleles for the other. The wild-type variety ZS97, the parent of the F_2 population, carried homozygous dominant alleles for the two genes.

A total of 500 SSR markers evenly distributed on 12 chromosomes were used to screen the two DNA pools and the varieties 02428, Teqing (TQ), and ZS97 to identify polymorphic markers. RM167 on chromosome 11 and RM19 on chromosome 12 were found to have polymorphisms between the two DNA pools. This indicated that two genes controlling the etiolation were likely linked to RM167 and RM19 (Fig. 4). All the RI184 double homozygotes (equal to double mutants) at RM167 and RM19 were etiolated. In contrast, all the wild-type plants carried at least one ZS97 allele at the two marker loci. Thus, our results confirmed that the etiolated seedling phenotype was controlled by two recessive genes linked to RM167 and RM19. The genes were named *et11* on chromosome 11 and *et12* on chromosome 12.

Fine mapping of *et11*

To fine map the two genes, two F_3 populations, each segregated at one gene, were produced (Fig. 1). Each F_3 population showed a segregation ratio of 1:3 etiolating to

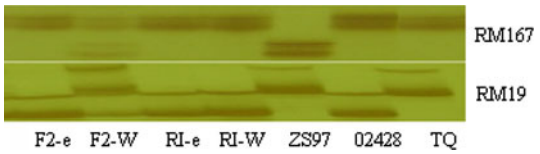

Fig. 4 Polymorphic markers linked to genes controlling etiolation identified by bulk segregant analyses F2-e and F2-W, DNA pools of mutant and wild type F_2 plants from the cross between RI184 and ZS97, respectively; RI-e and RI-W, DNA pools of mutant and wild-type plants in the progenies of RI184; *ZS97* Zhenshan 97, *TQ* Teqing

green plants. Thus, the gene in each F_3 population can be regarded as a dominant marker for genotyping each plant according to its seedling phenotype. In the F_3-1 population containing 124 plants, the region of *Et12* was homozygous for the mutant type, the interval from RM167 to RM3717 on chromosome 11 was homozygous for the mutant type, but the site at RM286 seated on the top of interval between RM167 and RM3717 was heterozygous. Genetic analysis showed that the genetic distance between *Et11* and RM286 was 0.5 cM (Fig. 5). Therefore, the genotype of the F_3-1 population was heterozygous at *Et11*, but homozygous recessive at *Et12*. Then, a large F_3-1 population of 4,100 plants was used to screen etiolated seedlings for the fine mapping of *Et11*. About 1,000 etiolated seedlings were obtained to scan the recombinants between *Et11* and marker RM286, and 11 recombinants were identified. In addition, four recombinants with the green phenotype were obtained from 660 seedlings of RI184 progeny with marker RM286.

The 15 recombinants were further screened with newly developed InDel and SNP markers. One recombinant was left between *Et11* and the SNP marker MSC1, and *Et11* was cosegregated with the SNP marker IBR2. Finally, *et11* was fine mapped into the interval between the telomere and MSC1 containing approximately 147 kb (Table 1).

The deletion of *Et12* in RI184

In the F_3-2 population of 80 plants, RM286 closely linking *Et11* was homozygous, and the interval around *Et12* was

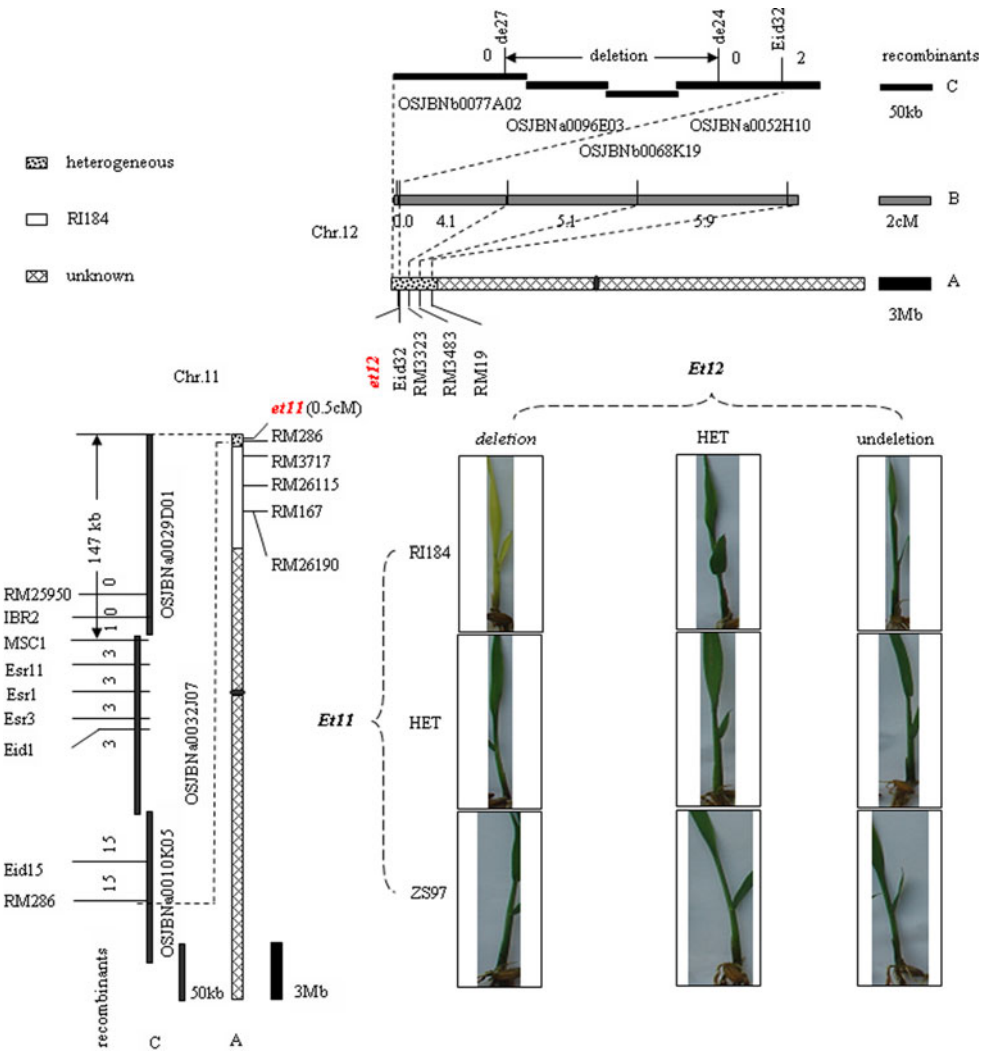

Fig. 5 Genetic analyses of *Et11* and *Et12*. **a** Entire physical maps of chromosomes 11 and 12; **b** local genetic map of chromosome 12; **c** local physical maps of chromosomes 11 and 12. The etiolated seedling phenotype was controlled by the complementary action of two recessive genes, *et11* and *et12*. Only the double-recessive homozygote (*et11et11*/*et12et12*) was an etiolated seedling. The *Et11* gene was fine mapped to approximately 147 kb between the telomere and InDel marker MSC1 on chromosome 11, whereas *Et12* was mapped into an approximately 209-kb region on chromosome 12, which was deleted in RI184

Table 1 Fine mapping of *Et11*

Plants	Phenotype	RM25950	IBR1	MSC1	Esr11	Esr1	Esr3	Eid1	Eid15	RM286
A313	Mutant	R	R	R	R	R	R	R	H	H
A347	Mutant	R	R	R	R	R	R	R	H	H
Y15	Mutant	R	R	R	R	R	R	R	H	H
Y154	Mutant	R	R	R	R	R	R	R	H	H
Y219	Mutant	R	R	R	H	H	H	H	H	H
Y236	Mutant	R	R	R	R	R	R	R	H	H
Y354	Mutant	R	R	R	R	R	R	R	H	H
Y396	Mutant	R	R	R	R	R	R	R	H	H
Y405	Mutant	R	R	R	R	R	R	R	H	H
Y526	Mutant	R	R	H	H	H	H	H	H	H
Y607	Mutant	R	R	R	R	R	R	R	H	H
G12	Green	H	H	H	H	H	H	H	T	T
G22	Green	H	H	H	T	T	T	T	T	T
G444	Green	H	H	H	H	H	H	H	T	T
G470	Green	H	H	H	H	H	H	H	T	T

G series were from R184 selfing progenies, the others were from population F_3-1

R is the genotype of RI184 (*et11et11*) in population F_3-1, T is the genotype of TQ (*et11et11*) in RI184 selfing population, and H is heterozygote (*Et11et11*). Genetic analysis of *Et11* in RI 184 selfing population showed that *et11* came from DNA fragment of the parent TQ

heterozygous. The genetic analysis showed that *Et12* cosegregated with the InDel marker Eid32.

The polymorphisms among the four parents TQ, 02428, ZS97, and RI184 were tested using InDel markers (with prefix "Eid") in the *Et12* region. With the Eid19 primer pair and Eid31-33 primer pairs, all four parents amplified products. However, with markers from Eid20 to Eid30, TQ, 02428, and ZS97 amplified PCR products, but no products were amplified in RI184 (Fig. 6). Then markers with prefix "de" were used to confirm whether RI184 had a deletion in this interval and, if so, to measure the length of the deletion. Ten markers (sequential order from the telomere to centromere: de13, de15, de27, de29, de23, de1, de4, de24, de31, and de19) were designed based on the sequences of the short arm of chromosome 12. Comparison of sequences in the duplicated regions of chromosomes 11 and 12 suggested that de27, de29, and de31 were able to anchor both chromosomes 11 and 12, but their PCR products could be distinguished by an SNP. Other markers had special anchors only on chromosome 12. Markers de13, de15, and de19 had PCR products in the four rice varieties, but de23, de1, de4, and de24 had no PCR products in RI184. The sequencing results showed that de31 had double signal peaks at an SNP position in all the varieties, whereas de27 and de29 had double signal peaks except in RI184 (Fig. 6).

The analysis indicated that RI184 had an approximately 209-kb deletion from de27 to de24 on chromosome 12. The markers at the two ends of the deletion region were used to genotype two recombinants between *et12* and InDel marker Eid32 from 270 plants in the F_3-2 population. No recombinant between the deletion markers and *Et12* was identified, indicating that *Et12* was located in the deleted region (Fig. 5).

Thus, RI184 had a heterozygous interval from the telomere to RM286 on chromosome 11 and an approximately 209-kb deletion from de27 to de24 on chromosome 12. This explains why etiolation in progeny of the RI184 line seemed to be controlled by a single gene.

Paralogs in the regions of *et11* and *et12*

Segmental duplication of an approximately 4.2-Mb fragment was identified between the two regions containing *et11* and *et12* based on a combined WU-blast and DAGchainer search. The fragment contains 313 pairs of paralogs. Almost all the genes in the 147-kb interval containing *Et11* have their paralogs in the corresponding region on chromosome 12. However, the paralogs of most genes in the 209-kb region of chromosome 12 are found in the counterpart on chromosome 11. The 147-kb interval and the 209-kb region share highly similar sequences of 38 kb, which possess genes with the same functions as predicted by the gene annotation of TIGR, such as part of RNA recognition motif-containing protein (LOC_Os11g01190 and LOC_Os12g01190), senescence-induced receptor-like serine/threonine-protein kinase precursor (LOC_Os11g01200 and LOC_Os12g01200), *HCF152* (LOC_Os11g01210 and LOC_Os12g01210), expressed protein (LOC_Os11g01220 and LOC_Os12g01220), IBR domain-containing protein (LOC_Os11g01240 and LOC_Os12g01230), and mitochondrial substrate

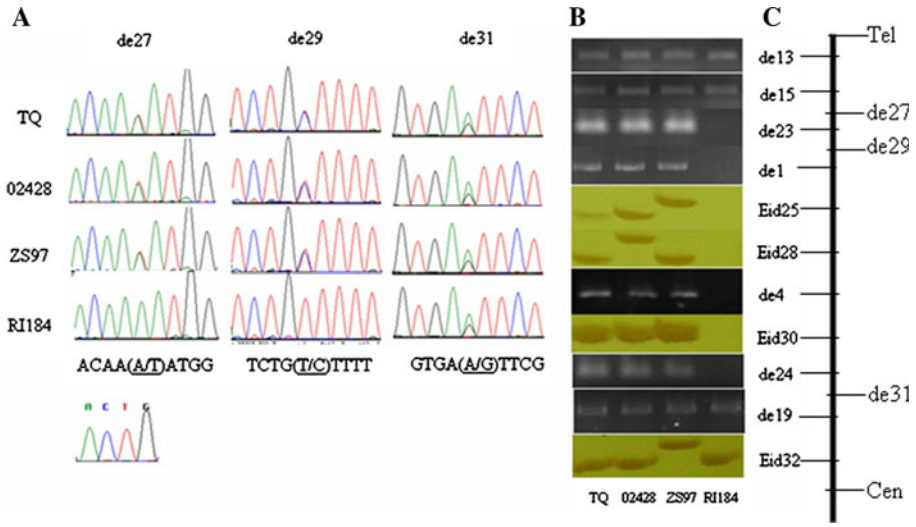

Fig. 6 A deletion of approximately 209 kb on chromosome 12 in RI184. **a** The primer pairs of de27, de29, and de31 can anchor both of the duplicated regions of chromosomes 11 and 12, and the sequenced PCR products from the two chromosomes are presented. At the SNP sites on chromosomes 11 and 12, a double peak means no deletion of de27 and de29 on chromosome 12 in RI184, a single peak means a deletion of de31 on chromosome 12. **b** Pairs of primers that specifically bind onto chromosome 12. PCR product means no deletion on chromosome 12, whereas no PCR product means a deletion of the region. **c** The physical order of primers on the short arm of chromosome 12

Table 2 Distinguishable paralogs in the fine-mapping regions of *et11* and *et12* (the deletion cosegregating with *et12*) using WU-blast

Chromosome 11		Chromosome 12		Putative function	BLASTP
Locus	Coordinates	Locus	Coordinates		E-value
LOC_Os11g01180.1	107153-106431	LOC_Os12g01180.1	102956-100009	Superfamily of TFs having WRKY and zinc finger domains	6.00E−126
LOC_Os11g01190.1	109395-115173	*LOC_Os12g01190.1*	105119-110825	RNA recognition motif-containing protein	1.00E−250
LOC_Os11g01200.1	123018-116428	*LOC_Os12g01200.1*	118819-112039	Senescence-induced receptor-like serine/threonine-protein kinase precursor	1.00E−250
LOC_Os11g01210.1	126373-128004	*LOC_Os12g01210.1*	125052-127145	HCF152	1.00E−250
LOC_Os11g01220.1	133454-129261	*LOC_Os12g01220.1*	131889-127617	Expressed protein	1.00E−250
LOC_Os11g01240.1	135610-137959	*LOC_Os12g01230.1*	132654-134280	IBR domain-containing protein	7.20E−201
LOC_Os11g01270.1	147869-143576	*LOC_Os12g01240.1*	139966-135627	Mitochondrial substrate carrier family protein	1.00E−250
LOC_Os11g01300.1	162402-162079	*LOC_Os12g01260.1*	156023-155700	Protein transport protein Sec61 subunit beta, putative, expressed	1.50E−51

Genes noted in bold and italics are located within to the interval of about 147 kb and the deletion of about 209 kb on chromosomes 11 and 12, respectively

carrier family protein (LOC_Os11g01270 and LOC_Os12g01240) (Table 2).

Comparison of gene expression patterns in the duplicated segments

Three crossover events occurred among the duplicated regions in the distal parts of chromosomes 11 and 12 after the segmental duplication (Wang et al. 2007). The total duplicated region can be divided into four interval pairs: the first crossover interval (first COI: 0.13–1.50 Mb), second crossover interval (second COI: 1.50–2.00 Mb), third crossover interval (third COI: from the telomere to 0.13 Mb), and a remnant part with no crossover (NCOI: 2.00–4.20 Mb). The breaking points of each crossover event were referred to by the corresponding positions on chromosome 11 as followed Wang et al. (2007). The first, second, and third COIs and NCOI contain 162, 58, 17, and 76 pairs of paralogs, with densities of 1.2, 1.3, 1.2, and 0.3 pairs per 10 kb, respectively. The younger crossover intervals possess higher paralog density than the older intervals, and NCOI has fewer paralog pairs than COIs, on average (Table 3).

A total of 83 of the 313 pairs of paralogs have special probe sets that could be used to distinguish them between chromosomes 11 and 12 (Table S1). The pairs of special probe sets of NCOI, first COI, second COI, and third COI were 54, 25, 4, and 0. NCOI had most of these probe sets, suggesting that the sequence in the duplicated regions without crossover has diverged much over time (Table 3).

Comparison of the expression profiles of duplicated gene pairs showed that most of the 83 pairs had a similar or

Table 3 Analysis of duplicated genes in the distal regions of chromosomes 11 and 12

Crossing over[a]	Intervals (Mb)[b]		Paralog pairs[c]		SPB[d]		Positive[e]		Irrelevant[e]		Negative[e]		Single silent[f]		Both silent[f]	
	Position	Length	No.	Density	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
NCOI	2.00–4.20	2.20	76	0.3/10 kb	54	71.1	27	50.0	17	31.5	2	3.7	6	11.1	2	3.7
First COI	0.13–1.50	1.37	162	1.2/10 kb	25	15.4	18	72.0	2	8.0	0		5	20.0		
Second COI	1.50–2.00	0.50	58	1.2/10 kb	4	6.9	3	75.0	0	0.0	0		1	25.0		
Third COI	0.00–0.13	0.13	17	1.3/10 kb	0	0.0										

[a] Three crossover events have occurred after the segmental duplication in the distal regions of chromosomes 11 and 12 (Wang et al. 2007). First COI, second COI, and third COI are the intervals from the initial to final physical positions of each crossover event. NCOI is the remnant part of the segmental duplicated region in which no crossover occurred

[b] The physical position and length of each interval from the initial to end position

[c] The number of paralog pairs in each of the four intervals. Paralog numbers divided by the length of each interval equals paralog density

[d] The number and percent of special probe set pairs for duplicated genes in each interval

[e] The number and percent of duplicated gene pairs with positive, irrelevant, and negative expression correlation in each interval

[f] One or both members of the duplicated gene pairs have no expression signal, as the largest expression value is less than 100, as is the background

the same expression pattern at different developmental stages (Table S1). The expression levels of 48 pairs were positively correlated in the 36 tissues examined. For example, LOC_Os11g06020.1 and LOC_Os12g06340.1 have homologous protein sequences, with a BLASTP E value of 1.7E-70, and both encode homeobox domain-containing protein. They have very similar expression patterns, with a positive correlation of 0.672 for expression level (Fig. S2). For different intervals in the duplicated region, at least 72% of paralogs were expressed positively in the first and second COIs, which more than that in the NCOI (50.0%) is. In total, 19 pairs of paralogs expressed irrelevantly. Of them, 17 pairs were located in NCOI and only 2 pairs were located in COIs (Table 3). This result suggests that the paralogs in the COIs tend to retain similar functions. Two pairs of duplicated genes were negatively expressed (Table 3). Both were located into NCOI. For example, LOC_Os11g06880.1 and LOC_Os12g07110.1 have similar protein sequences, with a BLASTP E value of 1.2E-168, and encode putative acyl-CoA synthetase protein, but they show negative expression patterns (Fig. S2). In addition, some genes had expression signals as low as the background in all 36 tissues, suggesting that their expression is silent (Table S1). There were six pairs in the NCOI, five in the first COI, and one in the second COI with expression silence in a single copy of the paralog pairs. Two pairs of duplicated genes were located in the NCOI, and both copies of each pair showed silent expression in all tissues, probably because the ancestral genes of these duplicated genes were silent as well (Table 3).

Discussion

The candidate genes for etiolated seedling

The very large PPR protein family, of which there are 450 members in *Arabidopsis* and 477 in rice, has been well studied. PPR proteins are encoded by nuclear genes and are predicted to be targeted to mitochondria or plastids (O'Toole et al. 2008). Most of the PPR proteins have been shown to play important roles in chloroplast gene expression (Fisk et al. 1999; Koteral et al. 2005). For example, as an RNA-binding protein, HCF152 is required for the processing and expression of psbB-psbT-psbH-petB-petD transcripts and subsequently affects the accumulation of the plastid cytochrome b6f complex, which is located on the chloroplast thylakoid in *A. thaliana* (Meierhoff et al. 2003; Nakamura et al. 2003).

Our findings indicate that the 147-kb region containing *Et11* is composed of 19 genes and the 209-kb region co-segregating with *Et12* possesses 31 genes. A pair of duplicated genes in the two regions possess high protein sequence identities of 59 and 52% with *HCF152* in *Arabidopsis* (Fig. S1), respectively, and is named *OsHCF152s* (LOC_Os11g01210 or LOC_Os12g01210). These genes are likely the candidate genes accounting for the etiolated seedling phenotype.

Genes involved in transcription or signal transduction tends to retain the primary function because of purifying selection (Blanc and Wolfe, 2004). In *Arabidopsis*, *HCF152* encodes a PPR protein that is imported into the chloroplast, where it is involved in the processing and expression of psbB-psbT-psbH-petB-petD transcript. The mutant *hcf152* results in a seedling lethal phenotype (Meierhoff et al. 2003; Nakamura et al. 2003). The rice mutant RI184 in this study presents a similar phenotype as that of the mutant *hcf152* in *Arabidopsis*. In addition, comparative sequencing of *Os11HCF152*, the candidate of *et1l*, showed that there were many polymorphisms in promoter region and coding regions between the wild type and mutant (Table S3). Of them, 3 SNPs resulted 3 amino acid changes. *Et12* was completely deleted in the mutant. These results provide evidence that *OsHCF152s* are the candidate genes controlling etiolation. *OsHCF152s* likely have retained a similar function as that of *HCF152* under purifying selection in the evolution of rice.

OsHCF152s are the candidate genes for the phenotype of etiolated seedling in rice, it is likely that a single mutation at either *Et11* or *Et12* does not affect the processing and expression of the psbB-psbT-psbH-petB-petD transcripts. However, double mutants of these duplicated genes would result in blocking of biosynthesis of the chloroplast thylakoids, as the mutant phenotype suggests. The light-dependent reaction cannot occur normally within the abnormal thylakoids, and ultimately biological energy cannot be produced to support other biosynthesis, such as that of Chl *a* and *b*. Thus, the mutation results in an etiolated seedling that soon withers and dies.

Functional redundancy of duplicated genes on distal parts of rice chromosomes 11 and 12

During the evolution of the rice genome, a whole-genome duplication occurred about 70 million years ago (Paterson et al. 2004; Wang et al. 2005), and single gene duplication is ongoing at a massive scale in the genome (Yu et al. 2005). It was suggested that a recent segmental duplication had occurred on the initial parts of chromosomes 11 and 12 about 5–7 million years ago (Rice Chromosomes 11 and 12 Sequencing Consortia 2005; Wang et al. 2005). However, recently it was proposed that the apparent segmental duplication actually resulted from the whole-genome duplication and became differentiated from the remainder of the chromosome(s) owing to concerted evolution (Paterson et al. 2009).

Since Ohno (1970) argued that gene or genome duplication supplies raw genetic material for biological evolution, several prevalent models for the fate of duplicated genes have been proposed, such as nonfunctionalization (Harrison et al. 2001), subfunctionalization (Giuseppe et al. 2005), neofunctionalization (Hideki and Fyodor 2010), and subneofunctionalization (He and Zhang 2005). In addition, subfunctionalization was thought to be a transition state to neofunctionalization (Rastogi and Liberles 2005). These hypotheses proposed explanations for the functional divergence of the duplicated genes. There is another situation, however, in which duplicated genes maintain the same function. This is referred to as concerted evolution, which is achieved by gene conversion and/or purifying selection (Nei and Rooney 2000). Gene conversion is often involved in homogenization of small tracts of paralogous DNA sequences, usually between several and several hundred base pairs (Petes et al. 1991), whereas the homogenization of larger tracts of DNA is generally believed to involve crossing over (Szostak and Wu 1980).

Since the segmental duplication on rice chromosomes 11 and 12, concerted evolution had occurred through gene conversion. As involved in a large DNA fragment of more than 3.5 Mb, crossover, instead of gene conversion, has taken place three times in the duplicated regions (Wang et al. 2007). COIs have much greater densities of paralogs than NCOI in the duplicated regions. In addition, the sequences of NCOI are more diverse than those of COIs, which was helpful as we developed special probe sets for distinction of the duplicated genes. The COIs have more duplicated genes with the same or similar expression patterns as compared to NCOI, and fewer pairs of duplicated genes showed irrelevant or negative expression. This situation was observed in the first, second, and third COIs. Moreover, the more recently the crossover event occurred, the higher the degree of similarity in sequences and/or expression patterns. In this study, rice etiolation was shown to be controlled by duplicated genes *et11* and *et12*. They were fine mapped into the third COI in the duplicated segments. A single mutant of either *et11* or *et12* can grow normally. An RI184 plant carrying the *Et11* allele could normally complete its life cycle despite the deletion of about 209 kb on chromosome 12. These results suggest that the genes' functions are redundant and buffer each other. In addition, 31 genes in the deletion region of chromosome 12 are located in the intervals where the first and third crossover events took place, suggesting that crossover has homogenized the differences between the duplicated regions, resulting in functional redundancy.

In summary, two recessive genes, *et11* and *et12*, controlling etiolation in rice were fine mapped to duplicated segments in the short arms of chromosomes 11 and 12, respectively. The duplicated genes *et11* and *et12* are functionally redundant. *OsHCF152* is likely the candidate gene of *et11* and *et12*. Both the duplicated segments possess paralogous genes with similar expression patterns, suggesting that concerted evolution has occurred in the segments.

Acknowledgments The authors express their thanks to the farm technician Mr. Wang Jianbo for his excellent field management and to Cao Jiangbo for his help with the transmission electron microscopy. This work was supported in part by grants from the National Program on the Development of Basic Research (2010CB125901) and the National Natural Science Foundation of China (30830064, 30921091).

References

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

Arnon DI (1949) Copper enzymes in isolated chloroplasts: polyphenoloxidase in *Beta vulgaris*. Plant Physiol 24:1–15

Beale SI (2005) Green genes gleaned. Trends Plant Sci 10:309–312

Bennetzen JL, Coleman C, Liu RY, Ma JX, Ramakrishna W (2004) Consistent over-estimation of gene number in complex plant genomes. Curr Opin Plant Biol 7:732–736

Blanc G, Wolfe KH (2004) Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. Plant Cell 16:1679–1691

Blankenship RE (2002) Molecular mechanisms of photosynthesis. Blackwell Science Ltd, London

Chen S, Jin W, Wang M, Zhang F, Zhou J, Jia Q, Wu Y, Liu F, Wu P (2003) Distribution and characterization of over 1000 T-DNA tags in rice genome. Plant J 36:105–113

Clack T, Mathews S, Sharrock RA (1994) The phytochrome apoprotein family in *Arabidopsis* is encoded by five genes: the sequences and expression of *PHYD* and *PHYE*. Plant Mol Biol 25:413–427

Douzery EJ, Snell EA, Bapteste E, Delsuc F, Philippe H (2004) The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? Proc Natl Acad Sci USA 101:15386–15391

Fisk DG, Walker MB, Barkan A (1999) Molecular cloning of the maize gene *crp1* reveals similarity between regulators of mitochondrial and chloroplast gene expression. EMBO J 18:2621–2630

Franklin KA, Quail PH (2010) Phytochrome functions in *Arabidopsis* development. J Exp Bot 61:11–24

Giuseppe D, Valentini T, Fruscoloni P, Glauco P (2005) Structure, function, and evolution of the tRNA endonucleases of Archaea: an example of subfunctionalization. Proc Natl Acad Sci USA 102:8933–8938

Gothandam KM, Kim ES, Cho HJ, Chung YY (2005) *OsPPR1*, a pentatricopeptide repeat protein of rice is essential for the chloroplast biogenesis. Plant Mol Biol 58:421–433

Harrison PM, Echols N, Gerstein MB (2001) Digging for dead genes: an analysis of the characteristics of the pseudogene population in the *Caenorhabditis elegans* genome. Nucleic Acids Res 29:818–830

He X, Zhang J (2005) Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. Genetics 169:1157–1164

Hideki I, Fyodor K (2010) The evolution of gene duplications: classifying and distinguishing between models. Nature Rev Genet 11:97–108

Jeon JS, Lee S, Jung KH, Jun SH, Jeong DH, Lee J, Kim C, Jang S, Yang K, Nam J et al (2000) T-DNA insertional mutagenesis genomics in rice. Plant J 22:561–570

Jung HS, Chory J (2010) Signaling between chloroplasts and the nucleus: can a systems biology approach bring clarity to a complex and highly regulated pathway? Plant Physiol 152: 453–459

Koornneef M, Meinke D (2010) The development of Arabidopsis as a model plant. Plant J 61:909–921

Kosambi DD (1944) The estimation of map distances from recombination values. Ann Eugen 12:172–175

Koteral E, Tasakal M, Shikanai T (2005) A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. Nature 433:326–330

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Marsch-Martinez N, Greco R, Van Arkel G, Herrera-Estrella L, Pereira A (2002) Activation tagging using the En-I maize transposon system in *Arabidopsis*. Plant Physiol 129:1544–1556

McCouch SR, Teytelman L, Xu Y, Lobos KB, Clare K, Walton M, Fu B, Maghirang R, Li Z, Xing Y et al (2002) Development and mapping of 2, 240 new SSR markers for rice (*Oryza sativa* L.). DNA Res 9:199–207

McElver J, Tzafrir I, Aux G, Rogers R, Ashby C, Smith K, Thomas C, Schetter A, Zhou Q, Cushman MA et al (2001) Insertional mutagenesis of genes required for seed development in *Arabidopsis thaliana*. Genetics 159:1751–1763

Meierhoff K, Felder S, Nakamura T, Bechtold N, Schusterb G (2003) HCF152, an Arabidopsis RNA binding pentatricopeptide repeat protein involved in the processing of chloroplast psbB-psbT-psbH-petB-petD RNAs. Plant Cell 15:1480–1495

Michelmore RW, Paran L, Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci USA 88:9828–9832

Muller AJ (1961) Mutationen mit Embryonaler Manifestation bei *Arabidopsis thaliana*. Naturwissenschaften 48:579

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321–4326

Nakamural T, Meierhoff K, Westhoff P, Schuster G (2003) RNA-binding properties of HCF152, an *Arabidopsis* PPR protein involved in the processing of chloroplast RNA. Eur J Biochem 270:4070–4081

Nei M, Rooney AP (2000) Purifying selection and birth-and-death evolution in the ubiquitin gene family. Proc Natl Acad Sci USA 97:10866–10871

O'Toole N, Hattori M, Andres C, Iida K, Lurin C, Schmitz-Linneweber C, Sugita M, Small I (2008) On the expansion of the pentatricopeptide repeat gene family in plants. Mol Bio Evol 25:1120–1128

Ohno S (1970) Evolution by gene duplication. Springer, Berlin

Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Natl Acad Sci USA 101:9903–9908

Petes TD, Malone RE, Symington LS (1991) Recombination in yeast. In: Broach J, Jones E, Pringle J (eds) The molecular cellular biology of the yeast saccharomyces. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp 407–421

Rastogi S, Liberles DA (2005) Subfunctionalization of duplicated genes as a transition state to neofunctionalization. BMC Evol Bio 5:28

Rice Chromosomes 11, 12 Sequencing Consortia (2005) The sequence of rice chromosomes 11 and 12, rice in disease resistance genes and recent gene duplications. BMC Biol 3:20

SP IRG (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296:92–100

Stern DB, Goldschmidt-Clermont M, Hanson MR (2010) Chloroplast RNA metabolism. Annu Rev Plant Biol 61:125–155

Szostak JW, Wu R (1980) Unequal crossing over in the ribosomal DNA of *Saccharomyces cerevisiae*. Nature 284:426–430

Takano M, Inagaki N, Xie X, Kiyota S, Baba-Kasai A, Tanabata T, Shinomura T (2009) Phytochromes are the sole photoreceptors for perceiving red/far-red light in rice. Proc Natl Acad Sci USA 106:14705–14710

Temnykh S, Park WD, Ayres N, Cartihour S, Hauck N, Lipovich L, Cho YG, Ishii T, McCouch SR (2000) Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:697–712

Temnykh S, Declerck G, Luashova A, Lipovich L, Cartinhour S, McCouch SR (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res 11:1441–1452

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

Wang X, Shi X, Hao B, Ge S, Luo J (2005) Duplication and DNA segmental loss in the rice genome: implications for diploidization. New Phytol 165:937–946

Wang XY, Tang HB, Bowers JE, Feltus FA, Paterson AH (2007) Extensive concerted evolution of rice paralogs and the road to regaining independence. Genetics 177:1753–1763

Wang L, Xie W, Chen Y, Tang W, Yang J, Ye R, Liu L, Lin Y, Xu C, Xiao J et al (2010) A dynamic gene expression atlas covering the entire life cycle of rice. Plant J 61:752–766

Waters MT, Langdale JA (2009) The making of a chloroplast. EMBO J 28:2861–2873

Wu KS, Tanksley SD (1993) Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol Gen Genet 241:225–235

Wu C, Li X, Yuan W, Chen G, Kilian A, Li J, Xu C, Li X, Zhou DX, Wang S et al (2003) Development of enhancer trap lines for functional analysis of the rice genome. Plant J 35:418–427

Yu J, Hu SN, Wang J, Wong GKS, Li SG, Liu B, Deng YJ, Dai L, Zhou Y, Zhang XQ et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296:79–92

Yu J, Wang J, Lin W, Li S, Li H et al (2005) The genomes of *Oryza sativa*: a history of duplications. PLoS Biology 3(2):e38

Zhang YC, Gong SF, Li QH, Sang Yi, Yan HQ (2006) Functional and signaling mechanism analysis of rice *CRYPTOCHROME 1*. Plant J 46:971–983