Promoter Shuffling at a Nuclear Gene for Mitochondrial RPL27. Involvement of Interchromosome and Subsequent Intrachromosome Recombinations[1]

Minoru Ueda, Shin-ichi Arimura, Masayuki P. Yamamoto, Fumio Takaiwa, Nobuhiro Tsutsumi, and Koh-ichi Kadowaki*

Genetic Diversity Department (M.U., K.K.) and Department of Plant Biotechnology (M.P.Y., F.T.), National Institute of Agrobiological Sciences, Tsukuba, Ibaraki 305–8602, Japan; and Laboratory of Plant Molecular Genetics, Graduate School of Agricultural and Life Sciences, University of Tokyo, Bunkyo-ku, Tokyo 113–8657, Japan (M.U., S.A., N.T.)

The *Reclinomonas americana* mitochondrial genome contains a mitochondrial ribosomal protein L27 (*rpl27*) gene, whereas the *rpl27* gene is absent from all plant mitochondrial genomes examined to date. This suggests that plant mitochondrial *rpl27* genes have been transferred previously from the mitochondrial genome to the nuclear genome. A nuclear cDNA encoding mitochondrial RPL27 was identified in rice (*Oryza sativa*). Three similar sequences were identified: *rpl27-1* and *rpl27-2* on chromosome 8 and *rpl27-3* on chromosome 4. Harr plot analysis suggests that they were generated by inter- and intrachromosomal duplications. Interestingly, the transcribed *rpl27* gene (*rpl27-1*) acquired a promoter sequence that was derived from the rice *spt16* (*Osspt16*) gene, the homolog of a global transcription factor in yeast (*Saccharomyces cerevisiae*) located downstream from the *rpl27-3* sequence on chromosome 4, after inter- and intrachromosomal recombination. Reverse transcription-PCR and promoter assay revealed that the *rpl27* mRNAs were mainly transcribed from *rpl27-1*. A repeat of seven nucleotides (AATAGTT) was identified at the junction of *rpl27-1* and *rpl27-2* on chromosome 8, and the same repeat was also identified at the 5′ end of *rpl27-2* and the 3′ end of *rpl27-1*. This repeat (AATAGTT) contains the hot-spot sequence AGTT, which is preferentially recognized by topoisomerase I in wheat (*Triticum aestivum*) germ, suggesting the involvement of topoisomerase I in this recombination. We here report the example of promoter shuffling and show that this promoter shuffling resulted from a recent segmental duplication through inter- and intrachromosomal recombination events.

The endosymbiont hypothesis of the origin of the mitochondria is generally accepted (Gray, 1992). It is thought that most of the genes of the ancestral endosymbiont have been either transferred to the nuclear genome or lost during evolution.

The complete mitochondrial genome sequences of various species have been determined. A limited number of genes are encoded, and their relative positions are largely conserved among vertebrate mitochondrial genomes (Gray, 1992). This is in marked contrast to the plant mitochondrial genome sequences of liverwort (*Marchantia polymorpha*), Arabidopsis (*Arabidopsis thaliana*), sugar beet (*Beta vulgaris*), rapeseed (*Brassica napus*), and rice (*Oryza sativa*), in which gene order and gene content are highly variable (Bullerwell and Gray, 2004). The liverwort mitochondrial genome encodes 16 kinds of ribosomal protein genes that have already been lost from vertebrate mitochondrial genomes. In higher plant mitochondrial genomes, the situation is more complex because the numbers of ribosomal protein genes in mitochondrial genomes vary among higher plants (Adams and Palmer, 2003). A mitochondrial ribosomal protein gene missing from one plant species but encoded by another plant species is likely to be encoded by the nuclear genome in the former species because ribosomal proteins are essential for protein synthesis. These discrepancies in gene content strongly suggest that gene transfer from the mitochondrial genome to the nuclear genome is an ongoing process in higher plants. Several transfer events from the mitochondrial to the nuclear genome have been identified, and they further our understanding of the mechanism of gene transfer from the mitochondrial to the nuclear genome. These events include the RNA-mediated transfer of the cytochrome *c* oxidase subunit 2 gene in cowpea (*Vigna unguiculata*; Nugent and Palmer, 1991) and the alternative splicing involved in the expression of the ribosomal protein S14 gene in rice and maize (*Zea mays*; Figueroa et al., 1999; Kubo et al., 1999). Sequence duplication/recombination was involved in the genomic acquisition of the presequence of ribosomal protein S11 (Kadowaki et al., 1996). In addition to these findings, evidence was also presented on exchange for loci sharing RNA

[1] This work was supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science, Sports and Culture of Japan (grant no. 15208001 to N.T. and K.K.).

* Corresponding author; e-mail kadowaki@affrc.go.jp; fax 81–29–838–7408.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Koh-ichi Kadowaki (kadowaki@affrc.go.jp).

Article, publication date, and citation information can be found at www.plantphysiol.org/cgi/doi/10.1104/pp.105.075564.

helicase-like target presequences, as well as loci sharing RNA binding protein target presequences, during the evolution of mitochondrial genes in Arabidopsis (Elo et al., 2003).

Although there have been several significant findings in this context, they are still limited, and the mechanism by which the promoter sequence of a newly translocated gene is acquired is unknown. Recently, whole-genome sequencing projects have been undertaken in many species, including humans (*Homo sapiens*), yeast (*Saccharomyces cerevisiae*), and Arabidopsis. These projects have revealed that inter- and intrachromosomal duplications occur in a complex manner (Arabidopsis Genome Initiative, 2000; Bailey et al., 2002; Dujon et al., 2004). During evolution, inter- and intrachromosomal duplications have occurred incidentally and have sometimes conferred a new function on a gene product or abolished its function, thus increasing genetic diversity. In higher plants, whole-genome sequencing projects have been conducted for Arabidopsis and rice. These genome sequences, together with full-length cDNA sequences, provide useful information about gene transfer events from organelles to the nucleus.

The mitochondrial genome of the heterotrophic flagellate *Reclinomonas americana* contains an *rpl27* gene (Lang et al., 1997), whereas the *rpl27* gene is absent from all plant mitochondrial genomes. We have identified a mitochondrial *rpl27* gene in the rice nuclear genome. Detailed analysis shows that the *rpl27* gene acquired a promoter sequence via inter- and intrachromosomal duplications from the rice *spt16*-related (*Osspt16*) gene, which is a homolog of the yeast *spt16* gene. The flexibility of the rice nuclear genome structure and the process of promoter shuffling are also discussed.

RESULTS

Nuclear Genes Encoding Mitochondrial RPL27 in Rice and Arabidopsis

The numbers of genes in the mitochondrial genomes of higher plants vary. Comparative analysis of the genes lost in different species makes it possible to determine the mechanisms and processes of gene transfer from the mitochondrial genome to the nuclear genome (Adams and Palmer, 2003). The *R. americana* mitochondrial genome has the largest number of genes of any mitochondrial genome that has been completely sequenced to date (Lang et al., 1997). Investigation of the protein-coding genes in the database revealed that the *rpl27* gene is encoded only in the *R. americana* mitochondrial genome; no other mitochondrial genome contains this gene. This suggests that the mitochondrial *rpl27* gene has already been transferred from the mitochondrial genome to the nuclear genome in most species. To identify the mitochondrial *rpl27* gene that was transferred to the nucleus in plants, a TBLASTN search of the National Center for Biotechnology Information (NCBI) database was conducted using *R. americana* mitochondrial RPL27 as the query. Two full-length cDNAs were found in both the rice and Arabidopsis databases. The rice clones show 49% (GenBank accession no. AK061690) and 50% (GenBank accession no. AK063072) amino acid identity relative to *R. americana* RPL27, and the Arabidopsis clones show 51% (GenBank accession no. AY046039) and 54% identity (GenBank accession no. AK118574) to *R. americana* RPL27. The deduced amino acid sequences of RPL27 from rice, Arabidopsis, *R. americana*, *Escherichia coli*, and Cyanobacterium were compared (Fig. 1). The alignment strongly suggests that the four clones found in rice and Arabidopsis are RPL27 and that these four clones have several conserved domains and extended amino acid sequences at their N and C termini.

The localization of these four clones from rice and Arabidopsis was predicted using three computer programs: TargetP (Emanuelsson et al., 2000), Predotar version 0.5 (http://www.inra.fr/predotar/), and Mitopred (Guda et al., 2004). Almost all programs concurred in predicting that AK063072 and AK118574 are mitochondrial RPL27 proteins and AK061690 and AY046039 are plastid RPL27 proteins; the only exception was the target organelle predicted for AY046039 by Predotar version 0.5.

Evaluation of the Targeting Signal of RPL27

In addition to the amino acid sequences conserved during evolution (AK063072, amino acids 40–120; AK061690, amino acids 61–138), the two rice clones have additional peptide sequences at both their N termini (AK063072, 1–39; AK061690, 1–60) and C termini (AK063072, 121–145; AK061690, 139–195). Mitochondrial proteins encoded by the nuclear genome carry a targeting signal that allows the protein to be sorted from the cytoplasm to the mitochondria. Most mitochondrial matrix proteins, such as ribosomal proteins, contain a targeting signal at the N terminus called the presequence (Pfanner and Geissler, 2001). It was assumed that the N-terminal sequence of rice RPL27 (AK063072) might function as a presequence. To monitor the subcellular localization of predicted mitochondrial RPL27 (AK063072), a reporter experiment was performed using the targeting signal described above and green fluorescent protein (GFP).

DNA encoding the amino acid sequence of the N-terminal extension (residues 1–40 of AK063072, corresponding to Fig. 1) of the predicted mitochondrial RPL27 was ligated to the *gfp* gene, and the fused protein was expressed under the control of the cauliflower mosaic virus (CaMV) 35S promoter. The chimeric gene was introduced into tobacco (*Nicotiana tabacum*) Bright-Yellow 2 (BY-2) suspension-cultured cells, and its expression was monitored 6 h after its introduction using confocal laser-scanning microscopy. The locations of the GFP protein and the mitochondria were visualized as the green color of GFP

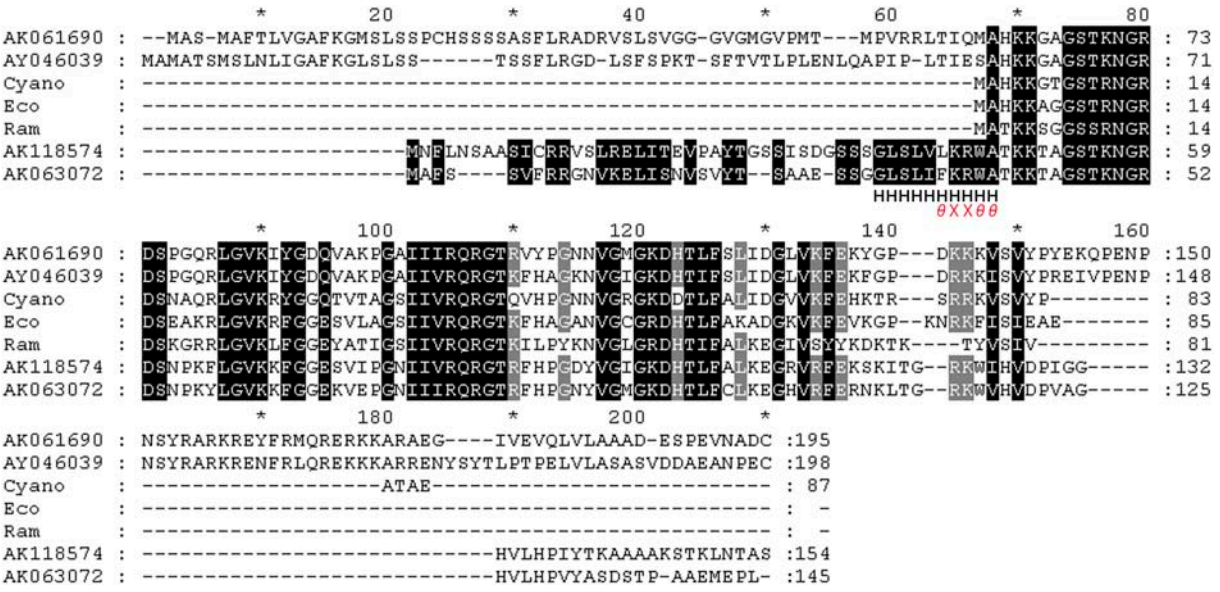

Figure 1. Alignment of the deduced amino acid sequences of RPL27. Black and gray backgrounds indicate amino acids conserved among the seven and six sequences, respectively. AK118574 and AK063072 of the extended N-terminal region were compared. Secondary structure prediction was carried out for the extended N-terminal region. H shows the deduced α-helix structure in rice. Tom20 binding segments ($\theta\chi\chi\theta\theta$) were found in the α-helix structure. AK061690 and AK063072 are GenBank accession numbers defined in the rice full-length cDNA database. AY046039 and AK118574 are GenBank accession numbers in the Arabidopsis full-length cDNA database. The other GenBank accession numbers for proteins used in this alignment are as follows: Cyano, cyanobacterium *Synechocystis* sp. RPL27 (NP_441681); Eco, *E. coli* RPL27 (P02427); and Ram, *R. americana* (NP_044787).

(Fig. 2A) and the red color of MitoTracker Red (Fig. 2C), respectively. If the GFP protein was targeted to the mitochondria, a yellow color should be observed when the green and red colors merge. After the fluorescent signals from GFP and MitoTracker Red were merged, almost all the GFP spots became yellow (Fig. 2B). These results strongly suggest that the GFP fusion protein containing the N-terminal portion of AK063072 were localized to the mitochondria, and that the mitochondrial-targeting signal occurs within the 40 amino acids of the N-terminal extension. These results strongly suggest that clone AK063072 encodes a gene for mitochondrial RPL27 protein.

Inter- and Intrachromosomal Duplications around the Mitochondrial *rpl27* Gene

The rice nuclear genome has been completely sequenced, and the genome sequence is available (http://rgp.dna.affrc.go.jp/). To determine the locus of the mitochondrial *rpl27* gene in the nuclear genome (AK063072), a BLASTN search of the rice genome database (http://riceblast.dna.affrc.go.jp/) was conducted. The mitochondrial *rpl27* gene was identified on chromosome 8. Interestingly, a sequence homologous but not identical to that of the *rpl27* gene was also found on chromosome 4. This suggests that the *rpl27* gene and the homologous region on a different chromosome have an evolutionary relationship.

To understand the relationship between these sequences, Harr plot analysis of the *rpl27* gene on chromosome 8 and the *rpl27*-related sequence on chromosome 4 was performed (Fig. 3). This indicated that 29 kb of chromosome 4 and 32 kb of chromosome 8 share a common DNA sequence around the *rpl27* gene. This strongly suggests that an interchromosomal duplication occurred between chromosome 4 and chromosome 8 in the past. The sequence similarity of the overall region duplicated between chromosome 4 and chromosome 8 is 97%, suggesting a relatively recent event.

Detailed analysis identified a tandem duplication of 5.2 kb within chromosome 8 (Fig. 3). The intrachromosomal duplication created a new 5.2-kb sequence containing an additional *rpl27* sequence. There is no base substitution between the two 5.2-kb regions, which suggests that the intrachromosomal duplication

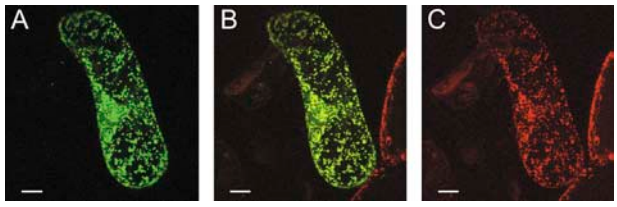

Figure 2. Transient expression in tobacco BY-2 cells of GFP fused to the N-terminal peptide (amino acids 1–40) of rice RPL27. The N-terminal peptide of rice RPL27 corresponds to Figure 1. A, GFP fluorescence. B, Merging of GFP and MitoTracker Red fluorescence. C, Mitotracker Red fluorescence. Scale bar = 10 μm.

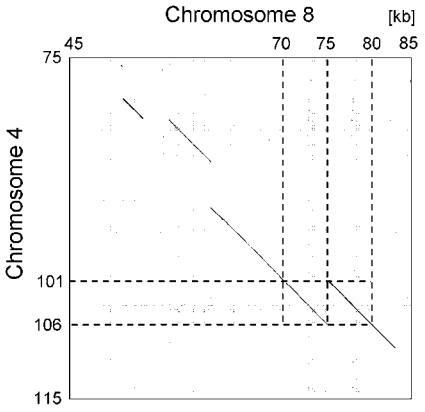

Figure 3. Harr plot analysis of the region of sequence encoding mitochondrial *rpl27* and its related sequences on chromosome 4 and chromosome 8. The horizontal sequence represents 40 kb of the bacterial artificial chromosome clone AP003872 from chromosome 8. The vertical sequence represents 40 kb of a contig of two bacterial artificial chromosome clones, AL663005 and AL663021, from chromosome 4. AP003872, AL663005, and AL663021 are GenBank accession numbers. Dots (duplicated regions across chromosomes) are placed at locations at which more than six nucleotides are continuously identical. Lines formed show duplicated regions. Two lines corresponding to 70 to 75 and 75 to 80 on chromosome 8 represent a sequence duplication event on the same chromosome.

occurred very recently. These results suggest that the intrachromosomal duplication occurred after the interchromosomal duplication.

BLASTX searches of the NCBI database for the homologous region between chromosome 4 and chromosome 8 identified unique sequences in the duplicated region. These include a 5.4-kb retroelement sequence on chromosome 4 and a 3-kb retroelement sequence on chromosome 8, so the results of Harr plot analysis show discontinuous lines and indicate that the duplication occurred in the past.

Whole-genome sequencing projects show in detail that there are huge numbers of genomic rearrangements in the human, yeast, and Arabidopsis genomes (Arabidopsis Genome Initiative, 2000; Bailey et al., 2002; Dujon et al., 2004). Furthermore, large-scale interchromosomal duplications have been found in rice (Paterson et al., 2004; Salse et al., 2004). Duplications can allow new functions to be conferred on the duplicate gene; they can also generate gene families or disrupt functional genes. Because duplication may have some effect on a gene, the effects of the inter- and intrachromosomal duplications around the *rpl27* gene were carefully analyzed.

Promoter Shuffling in the Mitochondrial *rpl27* Gene

The *rpl27* gene translocation event from the mitochondrial genome to the nuclear genome and the subsequent inter- and intrachromosomal duplication events ultimately resulted in the presence of three copies of the *rpl27* sequence on chromosome 4 and chromosome 8 (Fig. 4A). We designated the AK063072 sequence on chromosome 8 that was identified as a full-length cDNA, *rpl27-1*. The homologous sequence upstream from AK063072 on chromosome 8 was designated *rpl27-2*. The sequence on chromosome 4 homologous to that of AK063072 was designated *rpl27-3*. It is interesting that a sequence similar to the first exon (5′ untranslated region [UTR], 5′ UTR) of *rpl27-1* on chromosome 8 was identified approximately 1.6 kb downstream from *rpl27-3* on chromosome 4. Detailed analysis showed that this sequence was submitted previously as a full-length cDNA (GenBank accession no. AK121570) and that the cDNA was deduced to encode a homolog of the yeast *spt16* gene (Belotserkovskaya et al., 2003). We designated the AK121570 sequence *Osspt16*.

An interchromosomal duplication event that occurred between the sequences on chromosome 4 and chromosome 8 resulted in the translocation and duplication of the whole *rpl27* gene and part of the *Osspt16* gene (Fig. 4A). There are two *Osspt16*-related sequences on chromosome 4 and chromosome 8 (Fig. 4A). The deduced open reading frame (ORF) of *Osspt16* on chromosome 8 is 161 amino acids shorter than that on chromosome 4. There are two possibilities regarding the origin of the *Osspt16* sequence. One is that the sequence on chromosome 8 is a truncated form that originated on chromosome 4. The other possibility is that the sequence on chromosome 4 is an extended form that originated on chromosome 8. A BLASTX search of the NCBI database was conducted using the *Osspt16* sequence as the query, and two expressed sequence tags were identified in maize (GenBank accession no. AF545812) and Arabidopsis (GenBank accession no. NM_117139). When the expressed sequence tags from maize and Arabidopsis were aligned with *Osspt16*, they were seen to be highly conserved in terms of the size of the ORF (data not shown). This strongly suggests that the *Osspt16* gene on chromosome 4 is the original form.

The first exon of *rpl27-1* shares 94.7% sequence similarity with the first exon of *Osspt16* on chromosome 4 (Fig. 5, hatched boxes). The first exon of *rpl27-1* is located in the UTR and the translational ATG codon is present in the second exon. Because spliced full-length cDNAs of *rpl27-1* and *Osspt16* were identified in the rice full-length cDNA database (knowledge-based Oryza molecular biological encyclopedia), both sequences should have promoter sequences in their 5′ regions. The sequences around the 5′ ends of the cDNAs (possible transcriptional initiation sites) of *rpl27-1* and *Osspt16* were compared (Fig. 5B) and showed a high degree (97.3%) of similarity. In short, the entire first exon, UTR, and promoter sequences of the *rpl27-1* locus shared sequence similarity with the *Osspt16* locus. These results strongly suggest that the 5′ UTR of *rpl27-1*, including the promoter sequence, is derived from the *Osspt16* gene.

Assay for Transcription Activity of the Duplicated *rpl27* Gene with the *Osspt16* Promoter

Interchromosomal duplication followed by intrachromosomal duplication created two additional

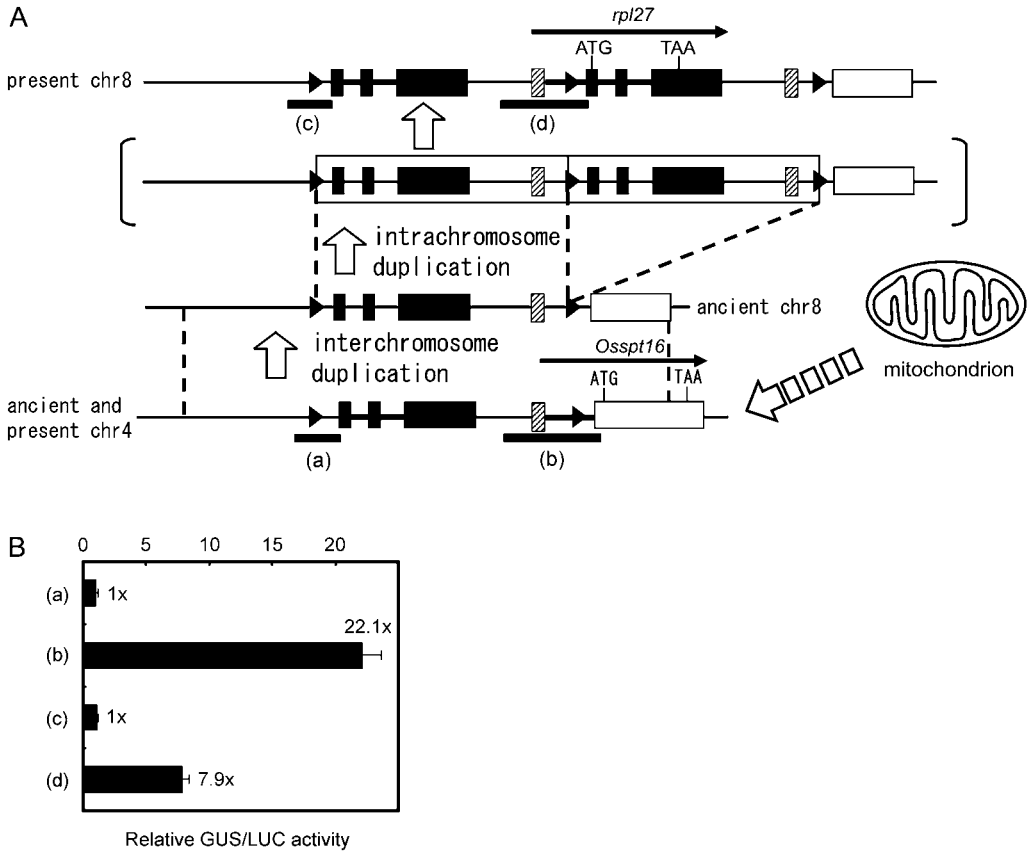

Figure 4. Schematic representation of promoter acquisition by the *rpl27* gene mediated by inter- and subsequent intrachromosomal duplications and the relative promoter activities of each *rpl27* sequence. A, Schematic representation of promoter acquisition by the *rpl27* gene mediated by inter- and subsequent intrachromosomal duplications. The white arrow with white rectangles shows the past *rpl27* gene transfer event from the mitochondrial genome to the nuclear genome. Horizontal arrows indicate transcribed regions identified from the full-length cDNA sequence. Exons and introns are indicated by boxes and thick horizontal lines, respectively. Black dotted lines indicate the regions duplicated. Black triangles show tandem repeats (AATAGTT) that are likely to have been involved in the intrachromosomal duplication. Black lines below *rpl27-3* (a), *rpl27-2* (c), *rpl27-1* (d), and *Osspt16* (b) indicate the cloned regions for transient promoter assays. B, Relative GUS/LUC activities. Effects of the inter- and intrachromosomal duplications on the promoter of the *rpl27* gene. Values show the activity of each construct relative to that of the *rpl27-3* construct (a). The characters a, b, c, and d to the left of the graph correspond to those in A. The experiment was repeated three times per plasmid, and the means are shown by black bars. Standard errors are indicated by lines extending from the bars. The graph represents an experiment performed with a single preparation of protoplasts. Numbers in the graph show the extent of activation (-fold) for each construct relative to that of the *rpl27-3* construct (a).

copies of *rpl27*-related sequences on chromosome 8 (*rpl27-1* and *rpl27-2*) from the original copy on chromosome 4 (*rpl27-3*). However, questions still remain regarding when and how the *rpl27*-related sequences acquired transcriptional apparatus. First, to measure the transcriptional activity of the three *rpl27* sequences, the promoter regions of the three *rpl27*-related sequences and that of *Osspt16* (used as a control) were fused to a β-glucuronidase (GUS) reporter gene and these constructs were transfected into rice protoplasts. The transcriptional activities of *rpl27-2* and *rpl27-3* were less than one-seventh that of *rpl27-1* (Fig. 4B). In short, this result indicates that the transcriptional activity of the *rpl27-1* gene increased more than 7-fold after acquiring the *Osspt16* promoter by promoter shuffling.

Nucleotide substitutions at three positions were identified in the ORFs of *rpl27-1* and *rpl27-3*. This nucleotide dissimilarity allowed us to distinguish the RNAs transcribed from *rpl27-1* and *rpl27-3*. Reverse transcription (RT)-PCR was performed using a primer set designed to amplify an internal part of the *rpl27* ORF, and the RT-PCR products obtained were sequenced. The sequences of 30 independent clones revealed that all of them were transcribed from *rpl27-1* and no clone was transcribed from *rpl27-3*. In short, the *rpl27* gene is transcribed from chromosome 8 but negligibly from chromosome 4.

There are no nucleotide differences between the two copies of the *rpl27*-related sequences on chromosome 8, *rpl27-1* and *rpl27-2*, except in the first exon of *rpl27-1*. This sequence is unique to *rpl27-1* and is absent from *rpl27-2* and *rpl27-3* (Fig. 4A). Therefore, it is difficult to distinguish the transcription of *rpl27-2* from that of *rpl27-1* using their internal sequences, as described for

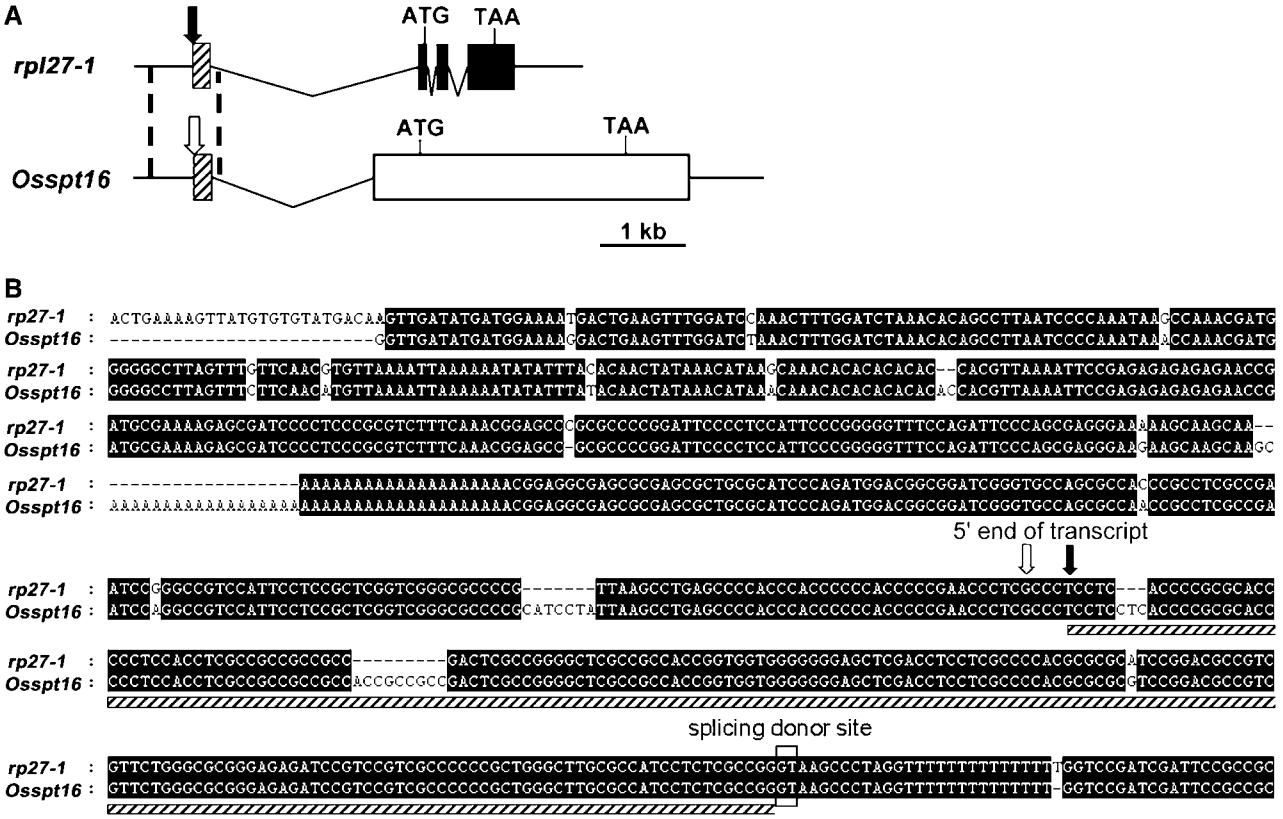

Figure 5. Schematic representations and alignment of genomic DNA sequences of the *rpl27-1* and the *Osspt16* genes. A, Schematic representations of exons and introns of the *rpl27-1* and the *Osspt16* genes determined from full-length cDNAs. Exons and introns are indicated by boxes and oblique lines, respectively. Horizontal lines indicate 5′ and 3′ flanking regions. Hatched boxes corresponding to hatched lines in B show exons of the *rpl27-1* and *Osspt16* genes that share sequence similarity. Black and white arrows indicate the transcription initiation sites of *rpl27-1* and *Osspt16*, respectively. Scale bar is 1 kb. ATG and TAA show translation initiation and stop codons, respectively. The nucleotide sequences between the two dotted lines, from *rpl27-1* and *Osspt16*, are aligned in B. B, Alignment of genomic DNA sequences around the transcription initiation sites of the *rpl27-1* and the *Osspt16* genes. About 500 bp upstream from the transcription initiation site, the first exon, and a part of the first intron of each gene are aligned. Black background indicates nucleotides identical between the *rpl27-1* and the *Osspt16* genes. A splicing donor site is shown by a box.

rpl27-3. Alternatively, RT-PCR was performed using a 5′ primer derived from the unique sequence upstream from *rpl27-2* that is not found in *rpl27-1* and a 3′ primer derived from the evolutionarily conserved *rpl27* sequence. However, no PCR product was detected (data not shown). This strongly suggests that there is no transcription from *rpl27-2*. The results described above demonstrate that the majority of *rpl27* transcripts are derived from *rpl27-1* on chromosome 8.

These results suggest that the *rpl27* gene was translocated from chromosome 4 to chromosome 8 by interchromosomal duplication, and that a subsequent intrachromosomal duplication on chromosome 8 created a new gene with a foreign promoter that originated from the *Osspt16* gene.

DISCUSSION

The mitochondrial *rpl27* genes are absent from mitochondrial genomes of plants, e.g. liverwort, Arabidopsis, and rice (Bullerwell and Gray, 2004). This suggests that translocation of mitochondrial *rpl27* genes occurred before the divergence of monocots and dicots. Genes for mitochondrial RPL27 have been identified in the nuclear genome of Arabidopsis and rice (Fig. 1). Thanks to mass genome-sequencing projects, a large number of segmental duplications have been found in the genomes of many species (Eichler and Sankoff, 2003). Segmental duplications have contributed to an increase in gene-copy number and genome size. In addition, in this study we have shown that segmental duplication also bestowed an already existing promoter sequence to a new gene, which was transferred from the mitochondrial genome to the nuclear genome during evolution. The *rpl27-1* gene acquired a promoter sequence derived from the rice *Osspt16* gene, the homolog of a global transcription factor in yeast, located downstream from the *rpl27-3* sequence on chromosome 4, after inter- and intrachromosomal recombination. RT-PCR and promoter assays revealed that *rpl27* mRNAs were mainly transcribed

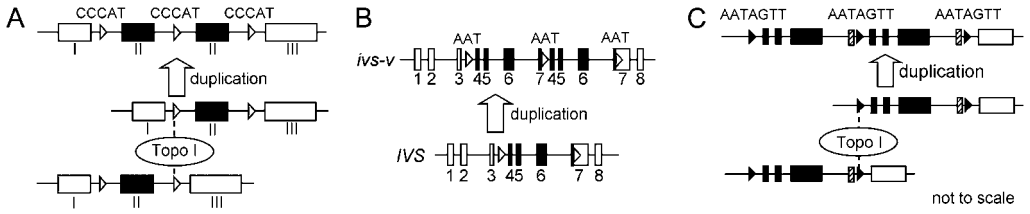

Figure 6. The possible mechanism for duplication via tandem repeats. A, Illegitimate recombination leads to exon duplication of the super αA-crystallin protein gene in hamster, cited from van Rijk et al. (1999). Recombination takes place between two nonhomologous sites with the identical sequence CCCAT in introns I and II of the super αA-crystallin protein gene. Topoisomerase I preferentially nicks the sequence CAT (Zhu and Schiestl, 1996), suggesting that the two CCCATs function in illegitimate recombination. White triangles indicate the sequence CCCAT. White and black boxes represent exons of the αA-crystallin gene. Black boxes also indicate the duplicated exon (exon II; van Rijk et al., 1999; Long, 2001). B, Genomic structures of the *IVS* and *ivs-v* genes in morning glory, cited from Park et al. (2004). Tandem repeats (AAT) were found at the junctions of the tandem duplication in *ivs-v*. White and black boxes indicate exons that are coding regions of the *IVS* and *ivs-v* genes. Black boxes indicate the duplicated region in the *ivs-v* gene. White triangles indicate the tandem repeat (AAT; Park et al., 2004). C, A possible mechanism for intrachromosomal duplication, which is probably triggered by topoisomerase I. Topoisomerase I preferentially nicks the hot-spot sequence (AGTT) in wheat germ (Been et al., 1984). Tandem repeats (AATAGTT) that contain the hot-spot sequence (AGTT) occur at the junction of the sequence duplication and upstream and downstream from the *rpl27* gene on rice chromosome 8. The repeat contains the wheat hot-spot sequence (AGTT) for topoisomerase I and the sequence (AAT) observed in B. Dotted lines indicate possible points of DNA breakage and DNA ligation by topoisomerase I. Other symbols correspond to those in Figure 4A. Topo I, Topoisomerase I (in A and C). Numbers in A and B indicate the order of exons for each gene.

from *rpl27-1*. No such novel findings have been observed in two Arabidopsis *rpl27* genes and one rice-chloroplast *rpl27* gene. Hence, we have focused our experimental analysis on rice mitochondrial *rpl27* only.

We have demonstrated that rice mitochondrial *rpl27-1* contains targeting information within 40 amino acids of its N terminus. To search for sequences homologous with these 40 amino acids, BLASTN and TBLASTN searches against the rice genome database were conducted. However, we did not find any homologous sequences with those in the rice genome. Regarding Arabidopsis mitochondrial *rpl27*, we did not find any sequences in the Arabidopsis genome that were homologous with its N-terminal portion.

Nucleotide sequence similarity around the 5′ portion of mitochondrial *rpl27* was not found when Arabidopsis and rice were compared. However, a comparison of the N-terminal amino acid sequences revealed that 27 amino acids were conserved between the two (Fig. 1). It is well known that mitochondrial targeting signals are loosely conserved and enriched in specific amino acids (Arg, Leu, and Ser), although a characteristic property of presequences is the high tendency to form an amphipathic α helix that presents one positively charged surface and one hydrophobic surface (Pfanner and Geissler, 2001). In rice *rpl27-1*, an amphipathic α-helix structure was observed at positions 31 to 40 (GLSLIFKRWA). It is noteworthy that an amphipathic α-helix structure was also found at positions 38 to 47 (GLSLVLKRWA) in Arabidopsis mitochondrial *rpl27*. Tom20, a 20-kD subunit of TOM (translocases of outer mitochondrial membranes), is an import receptor on the mitochondrial surface that binds to mitochondrial presequences in an early step of importing protein into mitochondria. The common pattern of Tom20 binding segments is described as $\theta\chi\chi\theta\theta$, where θ is a hydrophobic/aromatic residue (Leu, Met, Trp, Cys, Ala, Tyr, Phe, or Ile) and χ is any amino acid residue. Most of the χ residues have a long aliphatic side chain, often with a polar group at the end (Arg, Lys, Leu, or Ile; Muto et al., 2001). A common profile of the Tom20 binding segments was observed at the conserved amphipathic α-helix structure in rice (FKRWA) and Arabidopsis (LKRWA; Fig. 1). This strongly suggests that presequences of the mitochondrial *rpl27* genes in higher plants are of the same origin. However, they are quite divergent while retaining the same biological function during evolution.

There are two *Osspt16* sequences. One is located on chromosome 4 and the other on chromosome 8. It is possible that the *Osspt16* sequence on chromosome 4 is the original form and that the *Osspt16* sequence on chromosome 8 is a duplicated form. It is also possible that these sequences on chromosome 4 and chromosome 8 arose independently from the ancient mitochondrial genome. However, the latter possibility is unlikely for the following reasons. No *spt16* gene has been identified in any plant mitochondrial genome, including lower plants, such as liverwort, suggesting that gene transfer of *spt16* was completed a long time ago during evolution or the *spt16* gene has never been encoded by the mitochondrial genome. Two DNA segments from chromosome 4 and chromosome 8 containing *Osspt16* showed 97% DNA sequence identity, although one of the two sequences seems to be nonfunctional. This evidence strongly suggests that the two DNA segments were duplicated relatively recently, rather than the translocation of the DNA segments into the nuclear genome from the cytoplasmic genome occurring before the divergence of monocots and dicots. In addition, the *spt16* gene is one of two components of the facilitates chromatin transcription (FACT) complex. The FACT complex is widely conserved in eukaryotic cells (Belotserkovskaya et al.,

2003). The regulation of gene transcription in prokaryotes and eukaryotes is fundamentally different. The FACT complex is indispensable for the elongation of polymerase II in yeast; however, we identified no *spt16* gene in the database for bacteria such as *E. coli*. This suggests the *spt16* gene is a eukaryotic gene and not a prokaryotic gene. Mitochondria are widely accepted as a descendant of prokaryotes, such as α proteobacterium. Hence, it is less likely that the two sequences including *spt16* were transferred from mitochondria to the nuclear genome independently.

Inter- and intrachromosomal recombinations are believed to have been involved in the acquisition of the promoter sequence by the *rpl27* gene, which was translocated from the mitochondrial genome to the nuclear genome. Therefore, we undertook to determine the process of this intrachromosomal recombination event. A repeat of seven nucleotides (AATAGTT) was identified at the junction of the duplicated sequences on chromosome 8 and the same repeat was also identified at the 5′ and 3′ ends of the duplicated sequences on chromosome 8 (Fig. 4A). It is possible that this 7-bp repeat was involved in the intrachromosomal recombination event or is a footprint of the sequence duplication. Illegitimate recombination is mediated by topoisomerase I, which recognizes small repeat sequences then nicks the DNA and ligates the nicked DNA (Sherratt and Wigley, 1998). Analysis of illegitimate recombination in yeast has revealed that there is a tendency for topoisomerase I to recognize sequences containing a particular stretch of nucleotides [(G/C)(A/T)T, hot-spot sequences; Zhu and Schiestl, 1996]. In higher plants, topoisomerase I extracted from the wheat (*Triticum aestivum*) germ preferentially recognized the hot-spot sequence AGTT (Been et al., 1984). It is noteworthy that in our study, a microhomology (AATAGTT) of seven nucleotides was identified at both ends of the repeats and at the junction of the repeats, and that this repeat contained the hot-spot sequence (AGTT) observed in wheat germ. In transformation experiments with the αA-crystallin gene in hamsters, exon shuffling (tandem duplication) occurred in a transformed product of the αA-crystallin gene by illegitimate recombination. The sequences of the exon-shuffled regions suggest that the exon shuffling occurred by illegitimate recombination at two CCCAT repeat sequences. The CCCAT repeat contains the hot-spot sequence (CAT) of yeast (Fig. 6A; van Rijk et al., 1999; Long, 2001). Interchromosomal duplication of the αA-crystallin gene in hamsters is thought to have involved topoisomerase I. The recombination observed in this study seems to be analogous to the recombination of the αA-crystallin gene (Fig. 6, A and C).

The wild-type morning glory (*Ipomoea tricolor*) has bright-blue flowers and dark-brown seeds, whereas its spontaneous mutant, Blue Star, carrying the mutable ivory-seed-variegated (*ivs-v*) allele, has pale-blue flowers with a few fine, blue spots and ivory seeds with tiny dark-brown spots. The mutant allele is caused by an intragenic tandem duplication of 3.3 kb. Each of the tandem repeats is flanked by the 3-bp sequence AAT, indicating that a 3-bp microhomology is involved in generating the tandem duplication (Park et al., 2004). It is extremely interesting that the repeat identified in the *rpl27* region (AATAGTT) contains the 3-bp microhomology (AAT) involved in generating the tandem duplication in the *ivs-v* mutant of morning glory (Fig. 6B). We speculate that topoisomerase I and the AATAGTT repeat were implicated in the recombination described above, resulting in the promoter shuffling involving the *rpl27* gene, as well as in the *rpl27* sequence duplication (Fig. 6C).

MATERIALS AND METHODS

Plant Material

Etiolated seedlings of rice (*Oryza sativa* L. var Nipponbare) were used as the plant material.

Sequencing Procedures

Purified DNA was used for cycle sequencing with the Dye Terminator cycle sequencing quick start kit (Beckman Coulter). PCR was performed with an initial denaturation step at 96°C for 1 min followed by 30 cycles of 96°C for 20 s, 50°C for 20 s, and 60°C for 4 min. PCR products were precipitated and subjected immediately to sequence analysis using the CEQ 2000 XL DNA analysis system (Beckman Coulter).

Database Analysis

The available rice genome database (http://rgp.dna.affrc.go.jp/) and the BLAST programs in the RiceBLAST BLAST search service (http://riceblast.dna.affrc.go.jp/) were used. Default values were used for all parameters in the TBLASTN and BLASTN programs. Subcellular localization of the protein was predicted with TargetP (Emanuelsson et al., 2000), Predotar version 0.5 (http://www.inra.fr/predotar/), and Mitopred (Guda et al., 2004).

RT-PCR Analysis of Transcripts

Total RNA was isolated and further purified by incubation with RNase-free DNase I (TaKaRa), according to the manufacturer's instructions. First-strand cDNA synthesis was performed using 1 μg of purified RNA, 0.5 units of Moloney murine leukemia virus reverse transcriptase, and 20 pmol of random hexamer primer (BD Biosciences). The resultant cDNAs were used as templates to amplify the mitochondrial *rpl27* cDNA. Amplification of mitochondrial *rpl27* cDNA was performed with forward (5′-TCATCAGTTTTCAGGAGAGG-3′) and reverse (5′-CAATGGCTCCATTTCAGCTG-3′) primers and LA *Taq* DNA polymerase (TaKaRa). cDNA was denatured at 94°C for 5 min and amplified with 35 cycles of 94°C for 30 s, 54°C for 30 s, and 72°C for 1 min. PCR products were subsequently cloned into the pCR2.1-TOPO plasmid vector (Invitrogen). Independent clones were sequenced using universal primers.

GFP Expression Analysis

A nucleotide sequence thought to encode the targeting signal of rice RPL27, AK063072 (RPL27N), was amplified by PCR. The PCR product and the S65TGFP vector encoding the CaMV 35S promoter and 3′ nopaline synthase (NOS) transcription terminator (Chiu et al., 1996) were doubly digested with *Sal*I and *Nco*I and ligated in frame as follows: CaMV35Spro::RPL27N::sGFP(S65T)::NOSter. The nucleotide sequence of the resultant plasmid was confirmed by DNA sequencing. Plasmid DNA (10 μg) was precipitated onto 1.0 μm spherical gold beads (Bio-Rad), and the beads were bombarded into suspension-cultured tobacco (*Nicotiana tabacum*) BY-2 cells using a PDS-1000 particle delivery system (Bio-Rad). After bombardment, the cells were incubated in the dark at 24°C for 6 h. GFP fluorescence was visualized with a confocal laser-scanning microscope, as described previously (Arimura and Tsutsumi, 2002).

Staining of Mitochondria with MitoTracker Dyes

BY-2 suspension cells were incubated with 500 nM MitoTracker Red CMXRos (Invitrogen) in modified Murashige and Skoog medium enriched with 0.2 mg L^{-1} 2,4-D for 5 min at 24°C and washed three times in the above medium.

Oligonucleotide Primers for GFP and GUS Fusion Constructs

The putative promoter region of each gene and deduced targeting signal of predicted mitochondrial RPL27 (AK063072) were amplified by PCR and fused to the gene for GUS and GFP in reporter constructs, respectively. For GFP-fused protein, forward (5′-GAGACGTCGACCATGGCTTTTTC-3′) and reverse (5′-CTGTCTTT TCCATGGCCAACG-3′) primers, containing *Sal*I (GTCGAC) and *Nco*I (CCATGG) sites, respectively, were used. For GUS-fused protein, F1 (5′-GCAGGTAAGCTTCAGCCTTTCAATC-3′), F2 (5′-TTGCTTAAGCTTATACTTATAAGCC-3′), R1 (5′-TGATTGTCGACAATCACCTGAAATC-3′), and R2 (5′-TTTTAGGATCCAGCTCACATTAACCCAG-3′) primers were used, into which a *Hin*dIII (AAGCTT), *Bam*HI (GGATCC), or *Sal*I (GTCGAC) site was introduced (underlined). F1 and R2 primers were used for *rpl27-2* and *rpl27-3*. Sequences from −966 to −6 and from −1,000 to −6 relative to the translation initiation codons of *rpl27-2* and *rpl27-3*, respectively, were cloned. F2 and R2 primers were used for *Osspt16*. Sequences from −3,213 to −13 relative to the translation initiation codon of *Osspt16* were cloned. F2 and R1 primers were used for *rpl27-1*. Sequences from −3,472 to −3 relative to the translation initiation codon of *rpl27-1* were cloned.

Transient GUS Expression Assay Using Rice Protoplast

The putative promoter region of each gene, amplified by PCR, was ligated upstream from the GUS reporter gene::3′ NOS transcription terminator (Wu et al., 1998). The prepared GUS plasmid (10 μg) was transfected into rice protoplasts using the electroporation method described previously (Wu et al., 1998). Each transfection included 10 μg of the pAHC18 plasmid, which contains a ubiquitin promoter-luciferase (LUC)-encoding construct (Bruce et al., 1989), as an internal standard. GUS activities were normalized to the LUC activity.

ACKNOWLEDGMENT

We thank Dr Y. Niwa for providing the S65TGFP construct.

Received December 14, 2005; revised April 2, 2006; accepted April 3, 2006; published April 7, 2006.

LITERATURE CITED

Adams KL, Palmer JD (2003) Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Mol Phylogenet Evol **29:** 380–395

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

Arimura S, Tsutsumi N (2002) A dynamin-like protein (ADL2b), rather than FtsZ, is involved in Arabidopsis mitochondrial division. Proc Natl Acad Sci USA **99:** 5727–5731

Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE (2002) Recent segmental duplications in the human genome. Science **297:** 1003–1007

Been MD, Burgess RR, Champoux JJ (1984) Nucleotide sequence preference at rat liver and wheat germ type 1 DNA topoisomerase breakage sites in duplex SV40 DNA. Nucleic Acids Res **12:** 3097–3114

Belotserkovskaya R, Oh S, Bondarenko VA, Orphanides G, Studitsky VM, Reinberg D (2003) FACT facilitates transcription-dependent nucleosome alteration. Science **301:** 1090–1093

Bruce WB, Christensen AH, Klein T, Fromm M, Quail PH (1989) Photoregulation of a phytochrome gene promoter from oat transferred into rice by particle bombardment. Proc Natl Acad Sci USA **86:** 9692–9696

Bullerwell CE, Gray MW (2004) Evolution of the mitochondrial genome: protist connections to animals, fungi, and plants. Curr Opin Microbiol **7:** 528–534

Chiu W, Niwa Y, Zeng W, Hirano T, Kobayashi H, Sheen J (1996) Engineered GFP as a vital reporter in plants. Curr Biol **6:** 325–330

Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, De Montigny J, Marck C, Neuveglise C, Talla E, et al (2004) Genome evolution in yeasts. Nature **430:** 35–44

Eichler EE, Sankoff D (2003) Structural dynamics of eukaryotic chromosome evolution. Science **301:** 793–797

Elo A, Lyznik A, Gonzalez DO, Kachman SD, Mackenzie SA (2003) Nuclear genes that encode mitochondrial proteins for DNA and RNA metabolism are clustered in the Arabidopsis genome. Plant Cell **15:** 1619–1631

Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol **300:** 1005–1016

Figueroa P, Gomez I, Holuigue L, Araya A, Jordana X (1999) Transfer of *rps14* from the mitochondrion to the nucleus in maize implied integration within a gene encoding the iron-sulphur subunit of succinate dehydrogenase and expression by alternative splicing. Plant J **18:** 601–609

Gray MW (1992) The endosymbiont hypothesis revisited. Int Rev Cytol **141:** 233–357

Guda C, Fahy E, Subramaniam S (2004) MITOPRED: a genome-scale method for prediction of nucleus-encoded mitochondrial proteins. Bioinformatics **20:** 1785–1794

Kadowaki K, Kubo N, Ozawa K, Hirai A (1996) Targeting presequence acquisition after mitochondrial gene transfer to the nucleus occurs by duplication of existing targeting signals. EMBO J **15:** 6652–6661

Kubo N, Harada K, Hirai A, Kadowaki K (1999) A single nuclear transcript encoding mitochondrial RPS14 and SDHB of rice is processed by alternative splicing: common use of the same mitochondrial targeting signal for different proteins. Proc Natl Acad Sci USA **96:** 9207–9211

Lang BF, Burger G, O'Kelly CJ, Cedergren R, Golding GB, Lemieux C, Sankoff D, Turmel M, Gray MW (1997) An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. Nature **387:** 493–497

Long M (2001) Evolution of novel genes. Curr Opin Genet Dev **11:** 673–680

Muto T, Obita T, Abe Y, Shodai T, Endo T, Kohda D (2001) NMR identification of the Tom20 binding segment in mitochondrial presequences. J Mol Biol **306:** 137–143

Nugent JM, Palmer JD (1991) RNA-mediated transfer of the gene *coxII* from the mitochondrion to the nucleus during flowering plant evolution. Cell **66:** 473–481

Park KI, Choi JD, Hoshino A, Morita Y, Iida S (2004) An intragenic tandem duplication in a transcriptional regulatory gene for anthocyanin biosynthesis confers pale-colored flowers and seeds with fine spots in *Ipomoea tricolor*. Plant J **38:** 840–849

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

Pfanner N, Geissler A (2001) Versatility of the mitochondrial protein import machinery. Nat Rev Mol Cell Biol **2:** 339–349

Salse J, Piegu B, Cooke R, Delseny M (2004) New in silico insight into the synteny between rice (*Oryza sativa*) and maize (*Zea mays*) highlights reshuffling and identifies new duplications in the rice genome. Plant J **38:** 396–409

Sherratt DJ, Wigley DB (1998) Conserved themes but novel activities in recombinases and topoisomerases. Cell **93:** 149–152

van Rijk AA, de Jong WW, Bloemendal H (1999) Exon shuffling mimicked in cell culture. Proc Natl Acad Sci USA **96:** 8074–8079

Wu CY, Suzuki A, Washida H, Takaiwa F (1998) The GCN4 motif in a rice glutelin gene is essential for endosperm-specific gene expression and is activated by Opaque-2 in transgenic rice plants. Plant J **14:** 673–683

Zhu J, Schiestl RH (1996) Topoisomerase I involvement in illegitimate recombination in *Saccharomyces cerevisiae*. Mol Cell Biol **16:** 1805–1812