Mutational Decay and Age of Chloroplast and Mitochondrial Genomes Transferred Recently to Angiosperm Nuclear Chromosomes[1][w]

Chun Y. Huang*, Nicole Grünheit, Nahal Ahmadinejad, Jeremy N. Timmis, and William Martin

Australian Centre for Plant Functional Genomics, School of Agriculture and Wine, University of Adelaide, Waite Campus, Glen Osmond, South Australia 5064, Australia (C.Y.H.); Institute of Botany III, University of Duesseldorf, 40225 Duesseldorf, Germany (N.G., N.A., W.M.); and School of Molecular and Biomedical Sciences, University of Adelaide, South Australia 5005, Australia (J.N.T.)

Transfers of organelle DNA to the nucleus established several thousand functional genes in eukaryotic chromosomes over evolutionary time. Recent transfers have also contributed nonfunctional plastid (pt)- and mitochondrion (mt)-derived DNA (termed nupts and numts, respectively) to plant nuclear genomes. The two largest transferred organelle genome copies are 131-kb nuptDNA in rice (*Oryza sativa*) and 262-kb numtDNA in Arabidopsis (*Arabidopsis thaliana*). These transferred copies were compared in detail with their bona fide organelle counterparts, to which they are 99.77% and 99.91% identical, respectively. No evidence for purifying selection was found in either nuclear integrant, indicating that they are nonfunctional. Mutations attributable to 5-methylcytosine hypermutation have occurred at a 6- to 10-fold higher rate than other point mutations in Arabidopsis numtDNA and rice nuptDNA, respectively, revealing this as a major mechanism of mutational decay for these transferred organelle sequences. Short indels occurred preferentially within homopolymeric stretches but were less frequent than point mutations. The 131-kb nuptDNA is absent in the *O. sativa* subsp. *indica* or *Oryza rufipogon* nuclear genome, suggesting that it was transferred within the *O. sativa* subsp. *japonica* lineage and, as revealed by sequence comparisons, after its divergence from the *indica* chloroplast lineage. The time of the transfer for the rice nupt was estimated as 148,000 (74,000–296,000) years ago and that for the Arabidopsis numtDNA as 88,000 (44,000–176,000) years ago. The results reveal transfer and integration of entire organelle genomes into the nucleus as an ongoing evolutionary process and uncover mutational mechanisms affecting organelle genomes recently transferred into a new mutational environment.

Mitochondria (mt) and plastids (pt) are the descendants of once free-living prokaryotes, a proteobacterium and a cyanobacterium, respectively. During evolution, the bulk of their nuclear genomes has either been transferred to the eukaryotic host genome or lost, such that only remnants of the prokaryotic genomes are retained in the extant organelles (Timmis et al., 2004). As a consequence of this intracellular DNA transfer, several thousand functional nuclear genes have been acquired by plants during the evolution of chloroplasts (Martin et al., 2002), and similarly large numbers of successfully transferred mt-derived genes have been inferred (Gabaldón and Huynen, 2003; Esser et al., 2004). However, organelles not only donated functional genes; nonfunctional organelle DNA fragments are also found in nearly all eukaryote nuclear genomes (Ricchetti et al., 1999; Arabidopsis Genome Initiative, 2000; Mourier et al., 2001; Yuan et al., 2002; Rice Chromosome 10 Sequencing Consortium, 2003; Richly and Leister, 2004a). This continuous influx of organelle DNA from the mt and pt genomes still occurs today at very rapid rates, as is revealed both by genome comparisons (Mourier et al., 2001; Bensasson et al., 2003; Hazkani-Covo et al., 2003; Richly and Leister, 2004a, 2004b) and by recent direct laboratory measurements of organelle-to-nucleus DNA transfer (Ricchetti et al., 1999; Huang et al., 2003; Stegemann et al., 2003).

Most of the nuclear-integrated DNA segments that have been transferred from mitochondria (numts) and plastids (nupts) are currently less than 1 kb in length (Ricchetti et al., 1999; Mourier et al., 2001; Richly and Leister, 2004a), but some are large, ranging from several to more than 100 kb (Ayliffe and Timmis, 1992; Lin et al., 1999; Yuan et al., 2002; Rice Chromosome 10 Sequencing Consortium, 2003). The two largest examples of organelle DNA identified in nuclear genomes so far are the 620-kb copy of mtDNA on chromosome 2 of Arabidopsis (*Arabidopsis thaliana*; Lin et al., 1999; Stupar et al., 2001) and the 131-kb copy of ptDNA on chromosome 10 of rice (*Oryza sativa*; Rice Chromosome 10 Sequencing Consortium, 2003). These numtDNA and nuptDNA sequences duplicate their organelle-located copies, but it is not known whether they are on their way to becoming functional genes or pseudogenes and their mutational spectrum has not been examined in detail. Here, we report the

[1] This work was supported by the Grains Research and Development Corporation (C.Y.H.), the Australian Research Council (C.Y.H., J.N.T.), and the Deutsche Forschungsgemeinschaft (W.M.).

* Corresponding author; e-mail chunyuan.huang@adelaide.edu.au; fax 61–8–8303–7102.

[w] The online version of this article contains Web-only data.

Article, publication date, and citation information can be found at www.plantphysiol.org/cgi/doi/10.1104/pp.105.060327.

Downloaded from on October 25, 2020 - Published by www.plantphysiol.org
Copyright © 2005 American Society of Plant Biologists. All rights reserved.

mutations that have accumulated in these large numt and nupt DNA fragments relative to their coexisting organelle sequences and derive estimates for the age of the transfer events, taking into account the biases among the types of substitutions observed.

RESULTS

Mutations in Oryza nuptDNA

Three rice plastid genome sequences are currently available for comparison with the 131-kb nuptDNA on chromosome 10 of rice: those of *O. sativa* subsp. *japonica* (Hiratsuka et al. [1989]; updated in Tang et al. [2004]), *O. sativa* subsp. *indica* (Tang et al., 2004), and wild rice *Oryza nivara* (Masood et al., 2004). For simplicity, we refer to these sequences as *japonica*, *indica*, and *nivara*, respectively. The sequence organization of the 131-kb nuptDNA reflects the integration of a single, nearly contiguous molecule of rice pt DNA representing approximately 97% of the 134-kb organelle genome (Fig. 1). A 12.4-kb inversion in the nuptDNA corresponding to nucleotides 101,238 to 113,698 of *japonica* ptDNA (Fig. 1) probably results from homologous recombination between the inverted-repeat regions in the pt genome (Oldenburg and Bendich, 2004).

The *japonica* rice (cv Nipponbare) nuptDNA bears 39, 43, and 47 indels relative to the *indica*, *nivara*, and *japonica* chloroplast genomes, respectively. The majority of these indels entail single nucleotides (Fig. 2A). Indels in ptDNA-nupt comparisons were approximately 2- to 3-fold more frequent than in comparisons of the *indica-nivara* (13 indels), *indica-japonica* (16), and *nivara-japonica* (21) chloroplast genomes (Fig. 2B).

There were 271, 292, and 297 isolated single-nucleotide substitutions (i.e. those flanked neither by indels nor other substitutions) in the 131-kb nuptDNA relative to the *indica*, *nivara*, and *japonica* chloroplast genomes, respectively (Fig. 2C). Their distribution (Supplemental Table I) along the nupt in comparison to the *japonica* plastome was not significantly different from random using a chi-square test. Among the 12 possible kinds of point mutations, C → T and G → A transitions were by far the most prevalent substitution types in ptDNA-nupt comparisons (Fig. 2C), accounting for over one-half of all substitutions observed in each case. They are far more frequent than any other types of substitution, including the reverse transitions T → C and A → G (Fig. 2C). C → T transitions (and G → A transitions on the opposite strand) are the hallmark of spontaneous deamination of 5-methylcytosine (5^mC), which produces a G-T mismatch at the deaminated site. The mismatch can be restored to the G-C pair in one direction, but creates an A-T pair in the other direction (Holliday and Grigg, 1993; Finnegan et al., 1998), resulting in T → C transitions (and A → G on the opposite strand). Far fewer single-nucleotide substitutions and no predominance of C → T and G → A transitions were observed in comparisons of the three rice chloroplast genomes (Fig. 2D), indicating that the methylation-derived substitutions have occurred in the nucleus.

About 90% of single-nucleotide indels in ptDNA-nupt comparisons occurred within homopolymeric stretches of two to 11 nucleotides (data not shown), suggesting a role for DNA replication slippage. Indels >10 bp were usually flanked by direct repeats involving the terminal of 3 to 4 bp of the inserted/deleted sequence, as exemplified by five nupt-*japonica* comparisons (Fig. 3A).

Relationship of the Rice nupt to Three Rice Plastomes

With the exception of 12 sites, all polymorphisms in ptDNA-nupt and ptDNA-ptDNA comparisons were autapomorphic (i.e. only one sequence differed). Seven of the 12 nonautapomorphic sites were informative, suggesting relationships between the sequences. They

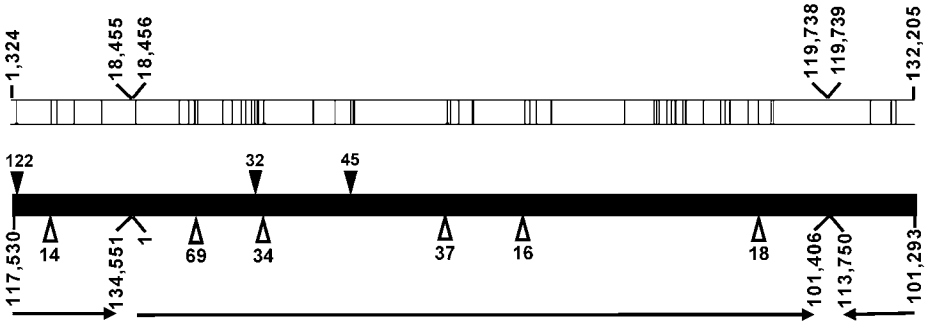

Figure 1. Alignments of *japonica* pt genome sequences with the 131-kb nuptDNA insert on *japonica* chromosome 10. The white and black boxes represent nuptDNA and ptDNA, respectively. Vertical numbers are the coordinates of nuptDNA on chromosome 10 of *japonica* and ptDNA of *japonica*, respectively. The ptDNA sequences were rearranged to make them colinear with the nuptDNA. Arrows give the 5′-3′ orientation of ptDNA sequences. Short indels (<10 bp) in the nuptDNA, relative to the ptDNA sequences, are represented by vertical lines in the white box. White triangles indicate the position of nucleotide deletions (absent in nuptDNA but present in the pt copy), and black triangles indicate the position and number of nucleotide insertions present in nuptDNA, but absent in ptDNA.

Downloaded from on October 25, 2020 - Published by www.plantphysiol.org
Copyright © 2005 American Society of Plant Biologists. All rights reserved.

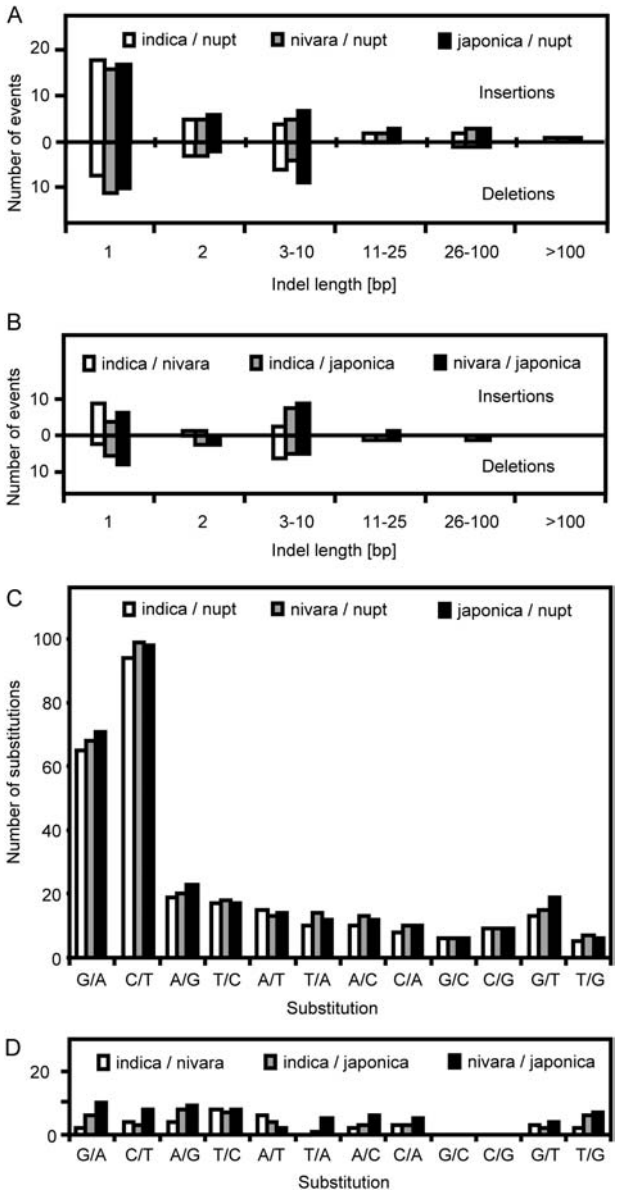

Figure 2. Indel length and nucleotide substitution classes in rice chloroplast and nupt sequences. A, Indel length classes in ptDNA-nuptDNA comparisons. Indels identified in the nuptDNA with specified length were pooled into six length groups shown on the *x* axis. The histograms indicate the number of insertions or deletions in each size class. White, gray, and black bars designate the comparisons shown at the top. For example, there are 18 single-nucleotide insertions and seven single-nucleotide deletions in the *nivara* ptDNA relative to the nupt. B, Indels in comparisons of rice ptDNA. Labels and conventions as in A. C, Nucleotide substitution frequencies in the nuptDNA relative to three rice plastomes. G/A indicates that G is present in the plastome, where A is present in the nuclear copy. D, Nucleotide substitution frequencies in comparisons of three rice plastomes. Labels and conventions as in C.

comprise three substitutions, one short inversion, and three indels (Table I). Five polymorphisms (loci 1–5) were shared between the *japonica* plastome and the *japonica* nupt, suggesting that nuclear integration of the *japonica* nuptDNA occurred from a *japonica* plastome progenitor. However, two additional sites linked the nupt to the other ptDNAs: one to the *nivara* plastome (a G → A transition; locus 6) and one to the *indica* plastome (a 32-bp indel; locus 7). The remaining five polymorphisms (loci 8–12) required more than one mutation and were uninformative.

Absence of the 131-kb nuptDNA in the Nuclear Genome of *indica*

To determine whether the 131-kb nuptDNA is present in *indica*, PCR analyses were conducted. *Oryza rufipogon* was included in the analyses because it is considered an outgroup to both *indica* and *japonica* (Khush, 1997), and its position is currently unresolved within the rice group (Nishikawa et al., 2005). As shown in Figure 4, A and B, primers F1 and R1, designed to detect the presence of the 131-kb nuptDNA, amplified a 677-bp product from total DNA of *japonica* (lane 3) as expected, but no products were obtained from DNA samples of *indica* (lane 2) or *O. rufipogon* (lane 4). Primers F1 and R2, designed to detect the absence of the nuptDNA, yielded a PCR product from *indica* (Fig. 4C, lane 2) and *O. rufipogon* (Fig. 4C, lane 4) templates, which was similar in size to that of 949 bp calculated from the nupt locus on chromosome 10 of *japonica*. As expected, this PCR product was not amplified from the DNA template of *japonica* (Fig. 4C, lane 3).

Sequencing revealed that the PCR products contained 913 bp for *indica* rice (EMBL accession no. AJ849475) and were 909 bp in length for *O. rufipogon* (EMBL accession no. AJ849476). There were four single-nucleotide indels and seven substitutions between *indica* rice and *O. rufipogon*, a total of 36 nucleotide insertions and 12 substitutions between *indica* rice and *japonica* rice, and a total of 40 nucleotide insertions and 17 substitutions between *japonica* rice and *O. rufipogon*. Sequence comparison of 160 bp (nucleotides of the *indica* PCR product from 195–320) containing the integration site between *japonica* and *indica* rice (Fig. 4D) revealed a 26-bp insertion (relative to the nuclear sequence of *indica* rice) 16 bp upstream of the integration site, and a 7-bp sequence (CCGAACC) at both sides of the junction in *japonica* rice, which differed from that (CCAAACC) in *indica* rice by one nucleotide. In addition, a 2-bp (CA) insertion was found immediately downstream of CCGAACC at the 5′ end of the junction site of *japonica* rice. There were 11 mutations in this 160-bp region that distinguish *japonica* and *indica* rice ptDNA (Fig. 4D), but there was no difference in the same region between *indica* and *O. rufipogon*. These data suggest that sequence changes, including two insertions (2 and 26 bp) and a 7-bp duplication of neighboring nuclear DNA, may have occurred during the nupt integration, as was observed recently in laboratory transfer experiments (Huang et al., 2004). Database searches of *indica* nuclear genomic sequences identified a single sequence of 841 bp on chromosome 10 of *indica* cv 93-11 (GenBank accession

Downloaded from on October 25, 2020 - Published by www.plantphysiol.org
Copyright © 2005 American Society of Plant Biologists. All rights reserved.

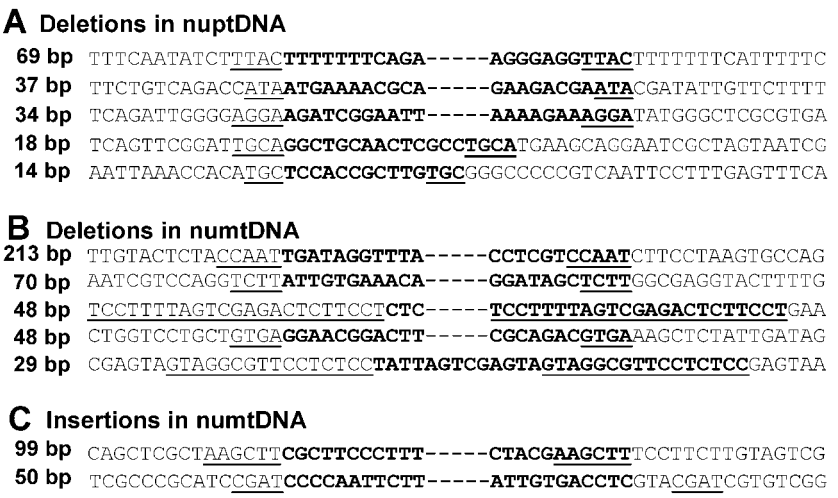

Figure 3. Direct repeats flanking indels in the nuptDNA of rice and the numtDNA of Arabidopsis. Indel length is indicated on the left and direct repeats are underlined. Bold letters indicate nucleotides absent in the nuptDNA (A) or numtDNA (B) but present in its organelle counterpart. Nucleotides present in the numtDNA but absent in its mt counterpart are also shown in bold letters (C).

no. AAAA02029093). This sequence is 99% identical to the 913-bp PCR product from *indica* cv Hsin Tieh Ta (Fig. 4C), and it contains the integration site but lacks the 72-bp sequence at its 5′ end. Like the data in Table I, these findings are compatible with the view that the 131-kb ptDNA fragment was transferred to the nuclear genome of *japonica* after it diverged from *indica*.

Mutational Types in Arabidopsis numtDNA

The 262-kb insert of Arabidopsis numtDNA (GenBank accession no. NC_003071) near the centromere of Arabidopsis chromosome 2, which includes 71.4% of the Arabidopsis mt genome, was aligned with the Arabidopsis organellar genome (Unseld et al., 1997). The numt is colinear with the mt genome, except for a few rearrangements and the complex admixture of 1,790 bp of the mtDNA sequence (Fig. 5). These rearrangements may have occurred prior to transfer in the mt via homologous recombination between repeated sequences (Unseld et al., 1997) or in the nucleus either before or during integration (Lin et al., 1999; Hazkani-Covo et al., 2003; Huang et al., 2004; Richly and Leister, 2004b). The 1,790-bp numtDNA sequence between nucleotides 3,313,242 and 3,315,031 (Fig. 5) could not be aligned with any contiguous stretch of mtDNA, but instead aligned with five short sequences of variable length (68–430 bp) from disparate regions of the Arabidopsis mt genome (data not shown). A 9-bp direct repeat (CTCGTAAAG) was found immediately upstream of the 1,790-bp sequence and at the 3′ end of the sequence (GenBank accession no. NC_003071, coordinates 3,313,233–3,313,240 and 3,315,032–3,315,040, respectively). Approximately 350 kb of duplicated mtDNA sequences were reported to be missing in this large numtDNA insert (Stupar et al., 2001), which are probably located between nucleotides 3,322,802 and 3,322,803 (Fig. 5).

Table I. *Loci with polymorphisms both in the indica-nupt comparison and among sequenced rice pt genomes*

–, Absence of nucleotide(s).

Locus	Coordinates in the *indica* Rice Plastome	*japonica* Plastome	131-kb nuptDNA on Chromosome 10	*O. nivara*	*indica* Plastome
Linking *japonica*-nupt					
1	64,174	C	C	A	A
2	66,410	A	A	G	G
3	62,529–62,536	CTTGGTCT	CTTGGTCT	AGACCAAG	AGACCAAG
4	65,623–65,624	TT	TT	–	–
5	5,015–5,016	–	–	CCTTTAT	CCTTTAT
Linking *nivara*-nupt					
6	8,128	G	A	A	G
Linking *indica*-nupt					
7	17,785–17,786	–	32 nucleotides	–	32 nucleotides
Multiple mutations					
8	75,989–75,990	TT	–	TT	T
9	77,735–77,736	–	G	G	TGG
10	78,434–78,440	TTTTTTT	–	TTT	TTTTTT
11	80,620	–	T	T	TT
12	134,536–134,545	A_{10}	A_9	A_{11}	A_{12}

Downloaded from on October 25, 2020 - Published by www.plantphysiol.org
Copyright © 2005 American Society of Plant Biologists. All rights reserved.

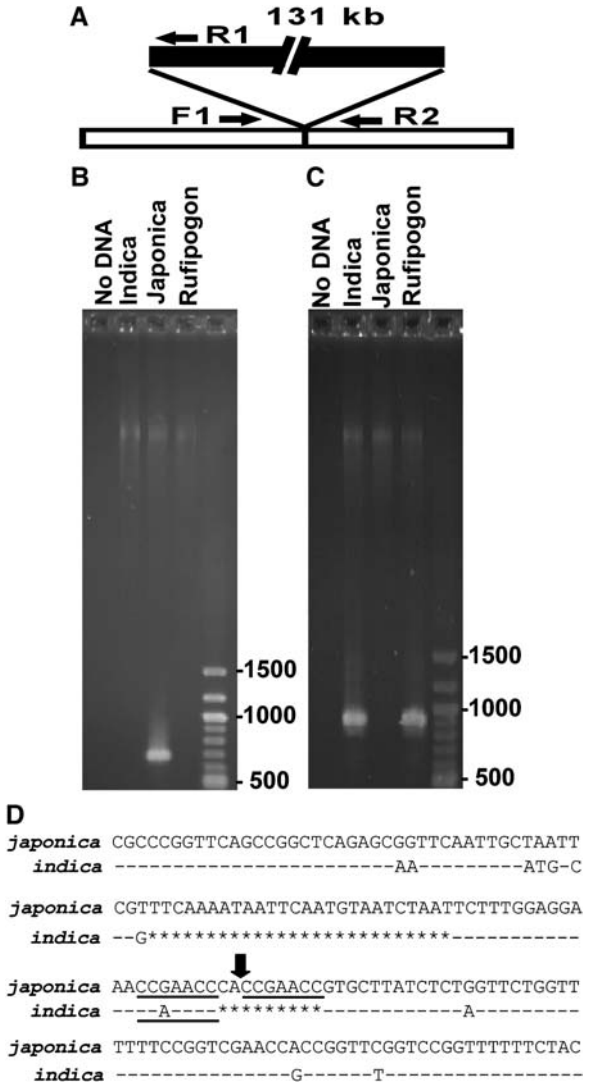

Figure 4. Sequence analysis of the integration site of the 131-kb nuptDNA from three rice taxa. A, Schematic representation of the integration site of the 131-kb nuptDNA on chromosome 10 of *japonica*. Black boxes indicate chloroplast sequences; white boxes are nuclear sequences. Arrows indicate PCR primer positions, and the integration site is shown as a vertical line. B, PCR amplification of the integration site using primers of F1 and R1. C, PCR amplification of the sequence of the integration site using primers of F1 and R2. Total genomic DNA samples (200 ng) from *O. rufipogon*, *indica*, and *japonica* were used in PCR amplification. The lane labeled "No DNA" indicates the negative control. Gels were stained with ethidium bromide. Fragment sizes are indicated in base pairs. D, Sequence comparison at the integration site of the 131-kb ptDNA between *japonica* and *indica*. Dashes represent identical nucleotides and asterisks indicate missing nucleotides. The black arrow points to the insertion site. Repeat motifs are underlined.

Six deletions larger than 10 bp (nucleotides present in mtDNA but missing in the numtDNA) were identified (Fig. 5), five of which were flanked by direct repeats of 4 to 23 bp (Fig. 3B). Two insertions larger than 10 bp were found, one 99 and the other 50 bp long, which also were flanked by direct repeats (Fig. 3C). The 6-bp repeats associated with the 99-bp insertion are *Hin*dIII sites. A database search revealed that this 99-bp integrant (nucleotides 3,412,715–3,412,813 of GenBank accession no. NC_003071) was similar to a 170-bp fragment in *Brassica napus* mtDNA (GenBank accession no. AP006444; nucleotides 141,851–142,020) but with a deletion of 71 bp, an insertion of 5 bp, and three nucleotide substitutions. The 99-bp fragment was present neither in the Arabidopsis mtDNA nor in the nuclear genome of Arabidopsis outside the 262-kb numtDNA. The 50-bp insertion (nucleotides 3,397,037–3,397,086 of GenBank accession no. NC_003071) was identical to a sequence in *B. napus* mtDNA (GenBank accession no. AP006444, nucleotides 65,464–65,415), which, like the 99-bp integrant, was absent from the mt and nuclear genome of Arabidopsis. These two regions were probably deleted through replication slippage or rearrangement from the Arabidopsis mt genome after the origin of the numtDNA.

In the 262-kb numtDNA, as in the rice nupt, single-nucleotide indels predominated over larger ones (Fig. 6A), with 144 single-nucleotide insertions and 123 single-nucleotide deletions found among a total 611 deleted and 320 inserted nucleotides encompassing 287 individual events. The nucleotides flanking inserted or deleted single nucleotides were examined for common patterns, revealing that 79% and 73%, respectively, involved homopolymeric stretches of 2 to 10 nucleotides. Homopolymeric stretches and simple tandem repeats of two to six nucleotides also flanked most of the 2- to 10-bp indels (data not shown).

In this Arabidopsis numtDNA, 241 single-nucleotide substitutions (i.e. those flanked to neither indels nor other substitutions) were observed relative to the mt copy (Fig. 6B). Their distribution (Supplemental Table II) along the numt in comparison to the Arabidopsis mt genome was not significantly different from random using a chi-square test. As in the case of rice nuptDNA, C → T and G → A transitions were far more frequent (accounting for 46% of the total) than other substitutions, including the reverse transitions T → C and A → G (Fig. 6B).

C → T and G → A Transitions at CG Dinucleotides and CNG Trinucleotides

The observed bias toward C → T and G → A transitions prompted us to examine the fate of cytosine residues in CG dinucleotides and CNG trinucleotides, which are both primary targets for DNA methylation in plants (Finnegan et al., 1998). In the rice nuptDNA, 48% of the C → T (and G → A) transitions occurred at CG dinucleotides and 19% occurred at CNG trinucleotides, in total accounting for 67% of the observed 5^{m}C-derived mutations (Table II). Similarly, in the numtDNA of Arabidopsis, 32% of the C → T transitions occurred at CG dinucleotides and 38% occurred at CNG trinucleotides, accounting for 70% of this mutation type (Table II).

Downloaded from on October 25, 2020 - Published by www.plantphysiol.org
Copyright © 2005 American Society of Plant Biologists. All rights reserved.

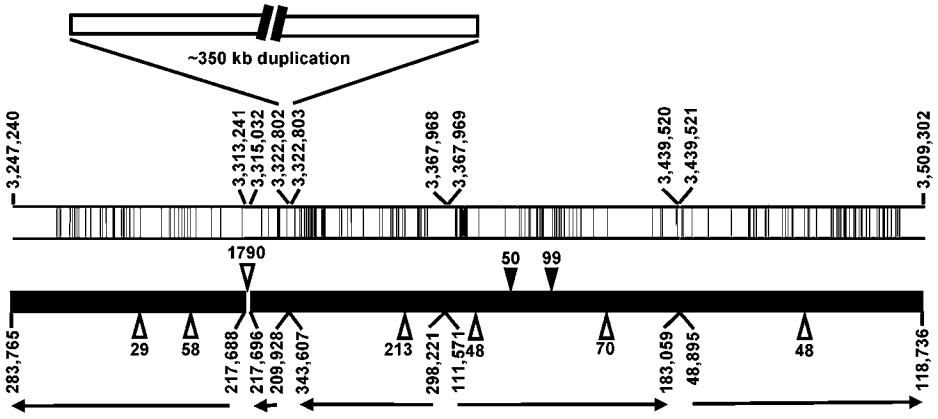

Figure 5. Alignments of Arabidopsis mt genome sequences with the 262-kb numtDNA insert on Arabidopsis chromosome 2. The white box represents numtDNA and the black box shows mtDNA. Vertical numbers indicate coordinates of numtDNA on Arabidopsis chromosome 2 and Arabidopsis mtDNA, respectively. The mtDNA sequences were rearranged to produce colinearity with the 262-kb numtDNA. Arrows indicate the 5′-3′ orientation of mtDNA sequences. Short indels (<10 bp) in numtDNA, relative to the mtDNA, are represented by vertical lines in the white box. White triangles point to the position of deletions of nucleotides absent in numtDNA but present in the mitochondrial copy. Black triangles indicate the position of insertions of nucleotides present in numtDNA but absent in mtDNA. Numbers at triangles indicate the length of indels in base pairs.

Are the 131-kb nuptDNA and 262-kb numtDNA under Selection?

When transferred organelle DNA arrives in the nucleus, its expression and mutation are subject to the regulation of the nuclear compartment. In order to become fixed as a functional gene, the transferred DNA must acquire a promoter to become transcribed and expression must lead to a product upon which selection can act, otherwise the sequence will, sooner or later, undergo mutational decay (Martin and Herrmann, 1998). We examined the protein-coding regions of the nupt and numtDNA for evidence of purifying selection, but found none. Nonsynonymous substitutions were more common than synonymous substitutions and in-frame stop codons as well as frameshift mutations were prevalent, affecting 47 and 41 open reading frames, respectively, in the nupt and numtDNA sequences (Table III). The proportions of synonymous, nonsynonymous, and missense mutations in numtDNA and nuptDNA reveal an approximately 2-fold excess of nonsynonymous over synonymous substitutions and numerous missense mutations (Table III) as expected for random mutation in the absence of purifying selection (Graur and Li, 2000). By contrast, comparisons of rice ptDNA coding regions clearly reveal that they are under purifying selection (Table III). The lack of evidence for purifying selection in coding regions of the nupt and numt sequences indicates that they are simply pseudogenes that are undergoing mutational decay. Notwithstanding the sequencing accuracy of expressed sequence tag data, independent work (D. Leister, personal communication) using BLAST searches uncovered no rice or Arabidopsis expressed sequence tags that are (1) distinct from the respective organelle genome sequences and (2) identical to either the nupt or the numt under investigation here, suggesting that very little, if any, transcription currently occurs from either of these two nuclear loci.

Rate of $5^{m}C$ Hypermutation Relative to Other Substitution Types

In the Arabidopsis numt, there were 54 C → T plus 57 G → A transitions and 19 T → C plus 15 A → G transitions relative to Arabidopsis mtDNA, corresponding to an excess of about 77 mutations attributable to $5^{m}C$ deamination. This indicates that approximately 32% (77/241) of all observed numt substitutions are derived from $5^{m}C$ deamination and that the rate of mutation due to this mechanism is about 5.6-fold faster than that of other point mutations. In the rice nupt comparisons to three ptDNAs, there were, on average, 97 C → T plus 68 G → A transitions as compared to 17.3 T → C plus 20.7 A → G transitions, corresponding to an excess of about 127 mutations attributable to $5^{m}C$ deamination, indicating that about 44% (127/287, average of three comparisons) of all observed nupt substitutions are so derived. In the rice nupt, the rate of mutations due to $5^{m}C$ hypermutation is 9.8-, 9.5-, and 9.2-fold faster than that of other point mutations in the *indica*, *nivara*, and *japonica* comparisons, respectively, or 9.5-fold faster on average.

Estimating the Age of These Organelle-to-Nucleus Transfer Events

The *japonica* nuptDNA is 99.77% identical (297/130,625 single-nucleotide differences) to the organelle-localized copy of the *japonica* plastome (Fig. 2C;

Downloaded from on October 25, 2020 - Published by www.plantphysiol.org
Copyright © 2005 American Society of Plant Biologists. All rights reserved.

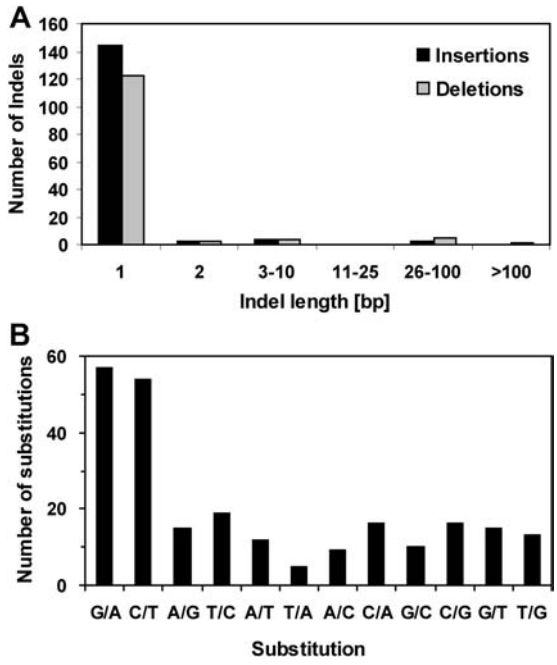

Figure 6. Indel length and nucleotide substitution classes in the 262-kb numtDNA of Arabidopsis. A, Indel length classes in the numtDNA. Indels of specified length were pooled into the six groups shown on the *x* axis. Histograms indicate the number of insertions or deletions. Insertions are nucleotides present in the numtDNA but absent in the Arabidopsis mtDNA. Deletions are nucleotides absent in the numtDNA but present in the Arabidopsis mtDNA. B, Frequency of 12 types of nucleotide substitution in the numtDNA relative to mtDNA (e.g. G/A indicates that G is present in mtDNA and A is present in the nuclear copy).

Supplemental Table I), and the Arabidopsis numtDNA is 99.91% identical (241/259,944 differences) to the DNA from Arabidopsis mt (Unseld et al., 1997). These two large organelle DNA integrants in the nuclear genomes clearly represent evolutionarily recent transfer events, but how recent? Molecular clock estimates can provide some ranges for orientation. Of course, the nuclear substitution rate used in molecular clock calculations is a critical parameter in estimating the timing of evolutionary events (Bromham and Penny, 2003). Estimated values for the nuclear substitution rate in plants vary dramatically, partly because plant fossil calibration points for determining rates are problematic (Koch et al., 2000) and partly due to lineage- and gene-specific rate variation (Zhang et al., 2002). Reported estimates for the nuclear rate in plants, expressed as substitutions per site per year, vary substantially: $>1.1 \times 10^{-9}$ among Boraginaceae (Böhle et al., 1996), 2.6×10^{-9} among palms (Morton et al., 1996), 5.1 to 7.1×10^{-9} among grasses (Wolfe et al., 1989), 6.5×10^{-9} among grasses (Gaut et al., 1996), and 1.5×10^{-8} estimated for Arabidopsis (Koch et al., 2000). To provide a conservative range of error, here we assume a 2-fold uncertainty attached to a mean estimate of 6.5×10^{-9} for the nuclear rate estimated by Gaut et al. (1996). Our timing estimates thereby accommodate a range of plant nuclear rates from 3.3×10^{-9} to 1.3×10^{-8}, approaching the slowest ($>1.1 \times 10^{-9}$) and fastest (1.5×10^{-8}) of a representative range of values from the literature.

Table I suggests that the 131-kb rice nuptDNA is more closely related to the *japonica* plastome among the three sampled here, and Figure 4 suggests that the transfer occurred within the *japonica* rice lineage. We scored 297 single-nucleotide substitutions among 130,625 sites of the pseudogene region between the *japonica* nupt and the *japonica* plastome sequences (Table VI). The 6.5×10^{-9} per site per year rate as applied to 297 substitutions would correspond to insertion of this nuptDNA about 350,000 years ago. However, the predominance of C → T and G → A transitions observed in our data is atypical of the sequences used to estimate the 6.5×10^{-9} per site per year rate (Gaut et al., 1996) and is furthermore not observed in comparison of rice plastomes (Tang et al., 2004; Fig. 2). Correcting for that by excluding the 129 C → T and G → A transitions attributable to 5^{m}C deamination yields about 168 substitutions among 130,625 sites sampled (0.13%) that are attributable to standard point mutations. One more correction is needed because the chloroplast rate is about 3- to 4-fold lower than the nuclear rate (Wolfe et al., 1987;

Table II. *Correlation of CG and CNG sites in the mtDNA and ptDNA with the nuclear C → T and G → A transitions found in Arabidopsis numtDNA and rice nuptDNA*

Trinucleotides	C/G → T/A (nuptDNA)	%	C/G → T/A (numtDNA)	%
CAA	3	1.8	9	8.1
CAC	5	3.0	2	1.8
CAG	20	11.8	20	18.0
CAT	8	4.7	1	0.9
Total CA		21.3	32	28.8
CCA	3	1.8	1	0.9
CCC	10	5.9	6	5.4
CCG	4	2.4	9	8.1
CCT	9	5.3	3	2.7
Total CC		15.4	19	17.1
CGA	23	13.6	12	10.8
CGC	16	9.5	6	5.4
CGG	25	14.8	7	6.3
CGT	17	10.1	10	9.0
Total CG		47.9	35	31.5
CTA	4	2.4	5	4.5
CTC	4	2.4	2	1.8
CTG	8	4.7	13	11.7
CTT	10	5.9	5	4.5
Total CT		15.4	25	22.5
CNG[a]	32	18.9	42	37.8

[a]N refers to A, C, or T in the trinucleotide. CGG is included in the CG dinucleotide count. The trinucleotides (C and two downstream nucleotides) from the mtDNA and ptDNA corresponding to T mutations in the trinucleotides of the numtDNA and nuptDNA were used in the analysis. In the case of the organelle G nucleotide to the nuclear A substitution, the C and its two downstream nucleotides in its complementary strand of the mtDNA and ptDNA were used in the analysis.

Downloaded from on October 25, 2020 - Published by www.plantphysiol.org
Copyright © 2005 American Society of Plant Biologists. All rights reserved.

Table III. *Mutations in chloroplast, nupt, and numt protein-coding regions*

Types of Mutation	*indica-nivara*	*indica-japonica*	*nivara-japonica*	*indica*-nupt	*nivara*-nupt	*japonica*-nupt	Arabidopsis-numt
Substitutions	13	13	22	107	169	107	64
Synonymous	7	9	16	26	58	36	20
Nonsynonymous	6	4	6	73	102	65	41
Stop	0	0	0	8	9	6	3

Graur and Li, 2000; Muse, 2000; Tang et al., 2004), such that about one-quarter of the observed substitutions probably occurred in the chloroplast, leaving about 126 nuclear mutations that can be assumed to have occurred at a mean rate of 6.5×10^{-9} per site per year. This would correspond to a mean estimate for the insertion date of approximately 148,000 years ago (assuming that 1/10,000 of the substitutions in the nupt sequence might be due to sequencing errors would lead to a 10% more recent age estimate). Assuming a 2-fold uncertainty in the rate estimate (6.5×10^{-9}) as it applies here to rice, we obtain a range for the estimated time of nupt insertion between 74,000 and 296,000 years ago, which is compatible with the estimates for *japonica-indica* divergence at 440,000 years ago based on nuclear gene data (Ma and Bennetzen, 2004) and 86,000 to 200,000 years ago based on chloroplast genome comparisons (Tang et al., 2004).

Repeating the above calculation for the time of insertion of the Arabidopsis numtDNA, where we observed 241 substitutions among 259,944 sites (0.09%), excluding the 77 excess C → T and G → A transitions, correcting for the roughly 10-fold difference in mt versus nuclear substitution rate (Wolfe et al., 1987), and using the 6.5×10^{-9} rate with 2-fold uncertainty, we estimate a mean age of mt-to-nucleus transfer for this copy at 88,000 years with a range of 44,000 to 176,000 years ago.

DISCUSSION

Phylogenetic relationships within the genus Oryza are not simple. Based on recent findings from simple sequence repeat variation in chloroplast and mt genomes among 50 Oryza accessions, Nishikawa et al. (2005) found that the phylogeny of *japonica* (cv Nipponbare), *indica* (cv Kasalath), *nivara*, and *O. rufipogon* was not resolved within the rice complex. In addition, the *indica* plastome sequence stems from a variety with a *japonica* maternal heritage (Tang et al., 2004). Nonetheless, in comparisons of the 131-kb nuptDNA with phylogenetically informative sites among three rice plastomes, the 131-kb nuptDNA on *japonica* chromosome 10 shares five polymorphisms with the *japonica* plastome to the exclusion of the other two rice plastomes (*nivara* and *indica*) currently available (Table I), suggesting that it was transferred subsequent to the divergence of the *japonica* plastome from the *nivara* and *indica* plastome lineages. The absence of the 131-kb nupt in both *indica* (cv Hsin Tieh Ta) and *O. rufipogon* at the integration site and the presence of a single sequence on chromosome 10 of *indica* rice (cv 93-11) containing an empty integration site (Fig. 4) support this view. The schematic history of the rice nupt in the context of sequenced rice plastomes as reconstructed from this work is summarized in Figure 7. Further screening of rice accessions for the presence of nupt and comparisons of further rice chloroplast genomes should provide additional insights.

Dinucleotide and trinucleotide substitutions were rare compared to single-nucleotide substitutions in the chloroplast and nupt comparisons, yet they were just as frequent in comparisons of pt genomes (10 dinucleotide and eight trinucleotide substitutions) as they were in ptDNA-nupt comparisons (11 and seven events, respectively; Table IV). All but two such substitutions occurred outside protein-coding regions. Averof et al. (2000) found that dinucleotide substitutions occur at about one-fortieth the frequency of single-nucleotide substitutions in mammalian nuclear pseudogenes. Table IV indicates a very low rate of dinucleotide and trinucleotide substitutions in the nonfunctional nuptDNA of the rice genome and a roughly 3-fold higher ratio of dinucleotide substitutions versus single-nucleotide substitutions in rice chloroplast noncoding regions over that seen in mammals (Averof et al., 2000).

Chloroplast and mt genomes exhibit very low levels of cytosine methylation (Timmis and Scott, 1983; Ayliffe et al., 1998), but cytosine methylation is extensive in plant nuclear genomes (Finnegan et al., 1998). Hence, relocation to the nucleus places organelle DNA into a fundamentally new mutational environment. We found that $5^{m}C$ hypermutation occurred approximately 5.6-fold faster than other point mutations in the Arabidopsis numt and approximately 9.5-fold faster in the rice nupt. Cytosine methylation and deamination appear to be the predominant mechanisms of mutational decay for organelle DNA after it is integrated into the higher plant nuclear genome. The roughly 2-fold higher rate of $5^{m}C$ hypermutation in the rice nupt versus the Arabidopsis numt might reflects differences in nuclear methyltransferase activities between these genomes (Finnegan et al., 1996; Lindroth et al., 2001), species-specific differences in the DNA repair enzymes that correct G-T mismatches (Wu et al., 2003), a later onset of methylation following integration in the Arabidopsis lineage, or possibly a combination of all these factors.

The vast majority of single-nucleotide insertions and deletions (the most frequent class of indels observed

Downloaded from on October 25, 2020 - Published by www.plantphysiol.org
Copyright © 2005 American Society of Plant Biologists. All rights reserved.

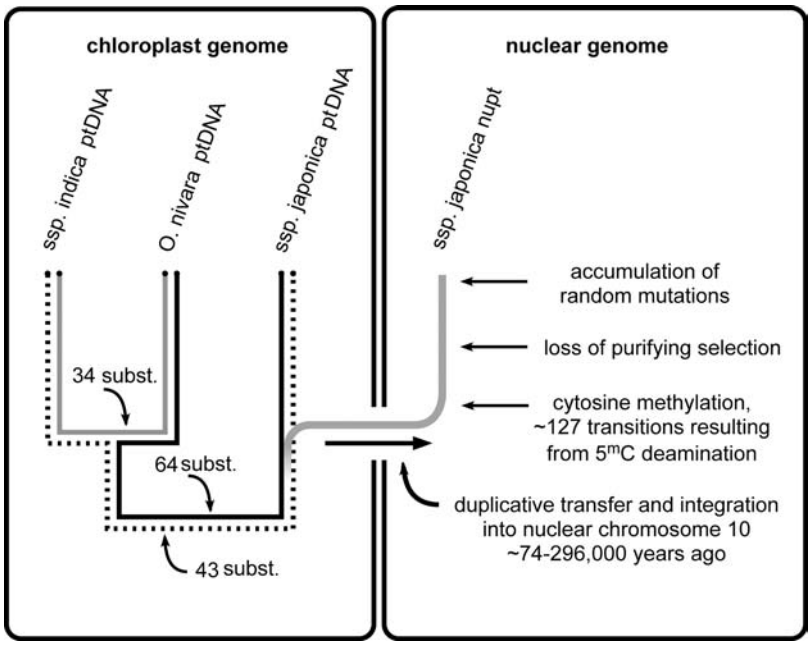

Figure 7. Schematic history of the nupt on rice chromosome 10. The numbers of isolated single-nucleotide substitutions (i.e. those not immediately flanked by other mutations) between the three sequenced rice chloroplast genomes are indicated.

here) occurred at homopolymeric regions in the nupt and numtDNAs. Direct repeats (3–6 bp) and homopolymeric stretches (2–10 bp) flank most indels of two nucleotides or more in both numtDNA and nuptDNA. Two direct repeats of 6 bp flanking a deletion were found previously in a tobacco (*Nicotiana tabacum*) nuptDNA (Ayliffe and Timmis, 1992). Ma and Bennetzen (2004) found that 89% of those nuclear deletions in rice not associated with solo long-terminal repeats were bound by short flanking repeats of 2 to 21 bp. This has also been observed within long-terminal repeat retrotransposons in Arabidopsis and rice (Devos et al., 2002; Ma et al., 2004). Slipped-strand mispairing within such repeats during replication appears to be the primary cause of insertion and deletion in the human genome, accounting for 70% insertions of short repeats of 2 to 4 bp in length (Zhu et al., 2000).

In the Arabidopsis numt, 287 indel events were observed as compared to 241 point mutations (164 after correction for 77 5^{m}C-derived transitions), indicating that substitutions and indels occur at roughly similar rates. After correction for 5^{m}C-derived transitions, indels were found to occur at one-third the rate of substitutions in rice ptDNA-nupt comparisons, and at one-half the rate in ptDNA. Replication slippage clearly plays an important role in the early mutational decay of transferred organelle DNA, but less so than 5^{m}C hypermutation in both species.

The dates of integration that we estimate for the 131-kb rice nuptDNA (between 74,000–296,000 years ago) and for the 262-kb Arabidopsis numtDNA (between 44,000–176,000 years ago) indicate that organelle DNA flux to the nucleus is a dynamic, ongoing process in plants. It is noteworthy that the first two higher plant genomes sequenced each contain a copy of a more or less complete organelle genome that was transferred somewhere on the order of 44,000 to 296,000 years ago. This suggests that fixation of complete organelle DNA integrants in the nucleus is rare in rice and Arabidopsis, and it implies transfer frequencies that are considerably lower than those found in recent laboratory experiments (Huang et al., 2003). Among 250,000 tobacco male gametes tested, 16 had a large fragment of chloroplast DNA newly integrated into the nucleus, although it was not necessarily a complete pt genome (Huang et al., 2003). But transfer alone is not sufficient for an organelle sequence to become observable in nuclear DNA through genome sequencing. The integration event must become fixed in the sequenced population, which, in the case of rice and Arabidopsis, were an inbreeding agricultural

Table IV. *Relative frequency of single, dinucleotide, and trinucleotide substitutions in plastome and nupt comparisons*

Types of Mutation	*indica-nivara*	*indica-japonica*	*nivara-japonica*	*indica*-nupt	*nivara*-nupt	*japonica*-nupt
Single-nucleotide substitutions	34	43	64	271	292	297
Dinucleotide substitutions	1	3	6	2	3	6
(in protein-coding regions)	(0)	(0)	(1)	(0)	(1)	(0)
Trinucleotide substitutions	2	2	4	1	3	3
(in protein-coding regions)	(0)	(0)	(0)	(0)	(0)	(0)

Downloaded from on October 25, 2020 - Published by www.plantphysiol.org
Copyright © 2005 American Society of Plant Biologists. All rights reserved.

cultivar and inbred experimental material, respectively. Those genomes that have been sequenced are unique examples that cannot represent the variability present among the species. We would expect high levels of nupt and numt heterogeneity (in both presence and sequence) to exist within the total genome pool of a species (Ayliffe et al., 1998) similar to the differences we have highlighted here between the subspecific *japonica* and *indica* nuclear genomes of rice. Likewise, we would expect wild, normally outbreeding, populations of plants to maintain high levels of polymorphism for nupts and numts. Most large numt and nupt transfers have probably been eliminated in the small rice and Arabidopsis genomes, and larger plant nuclear genomes, including the allotetraploid tobacco, may retain many more. The copies that have been retained provide insights into the mutational fate of recently transferred organelle DNA, a source of genetic novelty that is unique to the eukaryotic lineage.

MATERIALS AND METHODS

Sequence Alignments and Analysis

Sequences of Arabidopsis (*Arabidopsis thaliana*) mtDNA, rice (*Oryza sativa*) ptDNA, and nuclear organelle DNA integrants were retrieved from GenBank (accession no. NC_001284 for the Arabidopsis mt genome, AY522329 for the *indica* rice chloroplast genome, AY522331 for the *japonica* rice chloroplast genome, AP006728 for the *Oryza nivara* chloroplast genome, NC_003071 for the numtDNA sequence on Arabidopsis chromosome 2, and AE017082 for the nuptDNA sequence on rice chromosome 10). Initial nupt *indica* segment alignments were obtained using MegaBlast (http://www.ncbi.nlm.nih.gov/BLAST). Refined segment alignments were made with ContigExpress of Vector NTI version 8 (InforMax, Bethesda, MD). A 130-kb alignment (available upon request) of the three pt genomes and the nupt (after colinearization to ptDNA by rearranging the segments corresponding to nupt positions 1,324–18,455 and 119,739–132,205 in Fig. 1) was prepared with MLAGAN (http://lagan.stanford.edu). Regions corresponding to the recombination junctions (nupt positions 18,455–18,456 and 119,738–119,739 in Fig. 1) were excluded from analysis. Deletions, insertions, and substitutions were identified by shell scripts and by visual inspection of the alignments.

PCR Amplification of Genomic Sequences and Sequencing

Three rice taxa were used in the PCR analysis of the 131-kb nuptDNA integrant to investigate the time of its transfer to the nucleus: *japonica* (cv Nipponbare), *indica* (cv Hsin Tieh Ta), and *Oryza rufipogon* Griff. Total genomic DNA was prepared from young leaves and used in the amplification of the integration site of the 131-kb nuptDNA with primers F1 (5′ TGCTGTCGGATAGTCTGATG) and R1 (5′ CCTGCATCTGGACATAAAGA) or R2 (5′ TTCCGGTTAGCATCACTTTT). Products of PCR were separated by gel electrophoresis, purified by using a QIAquick gel extraction kit (Qiagen, Valencia, CA), and sequenced by using Applied Biosystems (Foster City, CA) sequencing technology.

Sequence data from this article have been deposited with the EMBL/GenBank data libraries under accession numbers AJ849475 and AJ849476.

ACKNOWLEDGMENTS

We thank U. Baumann and S. Gregory for technical advice on initial rice ptDNA alignments, J. Kretschmer for technical assistance in sequence analysis, L. Lewin for providing rice seeds, and Dario Leister both for discussions and for communicating results prior to publication.

Received January 26, 2005; revised March 30, 2005; accepted April 5, 2005; published June 10, 2005.

LITERATURE CITED

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

Averof M, Rokas A, Wolfe KH, Sharp PM (2000) Evidence for a high frequency of simultaneous double-nucleotide substitutions. Science **287:** 1283–1286

Ayliffe MA, Scott NS, Timmis JN (1998) Analysis of plastid DNA-like sequences within the nuclear genomes of higher plants. Mol Biol Evol **15:** 738–745

Ayliffe MA, Timmis JN (1992) Tobacco nuclear DNA contains long tracts of homology to chloroplast DNA. Theor Appl Genet **85:** 229–238

Bensasson D, Feldman MW, Petrov DA (2003) Rates of DNA duplication and mitochondrial DNA insertion in the human genome. J Mol Evol **57:** 343–354

Böhle U-R, Hilger HH, Martin W (1996) Island colonisation and evolution of the insular woody habit in *Echium* L. (*Boraginaceae*). Proc Natl Acad Sci USA **93:** 11740–11745

Bromham L, Penny D (2003) The modern molecular clock. Nat Rev Genet **4:** 216–224

Devos KM, Brown JK, Bennetzen JL (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in Arabidopsis. Genome Res **12:** 1075–1079

Esser C, Ahmadinejad N, Wiegand C, Rotte C, Sebastaini F, Gelius-Dietrich G, Henze K, Kretschmann E, Richly E, Leister D, et al (2004) A genome phylogeny for mitochondria among α-proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. Mol Biol Evol **21:** 1643–1660

Finnegan EJ, Genger RK, Peacock WJ, Dennis ES (1998) DNA methylation in plants. Annu Rev Plant Physiol Plant Mol Biol **49:** 223–247

Finnegan EJ, Peacock WJ, Dennis ES (1996) Reduced DNA methylation in *Arabidopsis thaliana* results in abnormal plant development. Proc Natl Acad Sci USA **93:** 8449–8454

Gabaldón T, Huynen MA (2003) Reconstruction of the proto-mitochondrial metabolism. Science **301:** 609

Gaut BS, Morton BR, McCaig BC, Clegg MT (1996) Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. Proc Natl Acad Sci USA **93:** 10274–10279

Graur D, Li W-H (2000) Fundamentals of Molecular Evolution. Sinauer Associates, Sunderland, MA

Hazkani-Covo E, Sorek R, Graur D (2003) Evolutionary dynamics of large numts in the human genome: rarity of independent insertions and abundance of post-insertion duplications. J Mol Evol **56:** 169–174

Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, Kondo C, Honji Y, Sun CR, Meng BY, et al (1989) The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Mol Gen Genet **217:** 185–194

Holliday R, Grigg GW (1993) DNA methylation and mutation. Mutat Res **285:** 61–67

Huang CY, Ayliffe MA, Timmis JN (2003) Direct measurement of the transfer rate of chloroplast DNA into the nucleus. Nature **422:** 72–76

Huang CY, Ayliffe MA, Timmis JN (2004) Simple and complex nuclear loci created by newly transferred chloroplast DNA in tobacco. Proc Natl Acad Sci USA **101:** 9710–9715

Khush GS (1997) Origin, dispersal, cultivation and variation of rice. Plant Mol Biol **35:** 25–34

Koch MA, Haubold B, Mitchell-Olds T (2000) Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in Arabidopsis, Arabis, and related genera (*Brassicaceae*). Mol Biol Evol **17:** 1483–1498

Lin X, Kaul S, Rounsley S, Shea TP, Benito MI, Town CD, Fujii CY, Mason T, Bowman CL, Barnstead M, et al (1999) Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. Nature **402:** 761–768

Lindroth AM, Cao X, Jackson JP, Zilberman D, McCallum CM, Henikoff S, Jacobsen SE (2001) Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. Science **292:** 2077–2080

Downloaded from on October 25, 2020 - Published by www.plantphysiol.org
Copyright © 2005 American Society of Plant Biologists. All rights reserved.

Ma J, Bennetzen JL (2004) Rapid recent growth and divergence of rice nuclear genomes. Proc Natl Acad Sci USA **101:** 12404–12410

Ma J, Devos KM, Bennetzen JL (2004) Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. Genome Res **14:** 860–869

Martin W, Herrmann RG (1998) Gene transfer from organelles to the nucleus: how much, what happens, and why? Plant Physiol **118:** 9–17

Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, Leister D, Stoebe B, Hasegawa M, Penny D (2002) Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc Natl Acad Sci USA **99:** 12246–12251

Masood MS, Nishikawa T, Fukuoka S-I, Njenga PK, Tsudzuki T, Kadowaki K-I (2004) The complete nucleotide sequence of wild rice (*Oryza nivara*) chloroplast genome: first genome wide comparative sequence analysis of wild and cultivated rice. Gene **340:** 133–139

Morton BR, Gaut BS, Clegg MT (1996) Evolution of alcohol dehydrogenase genes in the palm and grass families. Proc Natl Acad Sci USA **93:** 11735–11739

Mourier T, Hansen AJ, Willerslev E, Arctander P (2001) The Human Genome Project reveals a continuous transfer of large mitochondrial fragments to the nucleus. Mol Biol Evol **18:** 1833–1837

Muse SV (2000) Examining rates and patterns of nucleotide substitution in plants. Plant Mol Biol **42:** 25–43

Nishikawa T, Vaughan DA, Kadowaki K-I (2005) Phylogenetic analysis of Oryza species, based on simple sequence repeats and their flanking nucleotide sequences from the mitochondrial and chloroplast genomes. Theor Appl Genet **110:** 696–705

Oldenburg DJ, Bendich AJ (2004) Most chloroplast DNA of maize seedlings in linear molecules with defined ends and branched forms. J Mol Biol **335:** 953–970

Ricchetti M, Fairhead C, Dujon B (1999) Mitochondrial DNA repairs double strand breaks in yeast chromosomes. Nature **402:** 96–100

Rice Chromosome 10 Sequencing Consortium (2003) In-depth view of structure, activity, and evolution of rice chromosome 10. Science **300:** 1566–1569

Richly E, Leister D (2004a) NUMTs in sequenced eukaryotic genomes. Mol Biol Evol **21:** 1081–1084

Richly E, Leister D (2004b) NUPTs in sequenced eukaryotes and their genomic organization in relation to NUMTs. Mol Biol Evol **21:** 1972–1980

Stegemann S, Hartmann S, Ruf S, Bock R (2003) High-frequency gene transfer from the chloroplast genome to the nucleus. Proc Natl Acad Sci USA **100:** 8828–8833

Stupar RM, Lilly JW, Town CD, Cheng Z, Kaul S, Buell CR, Jiang J (2001) Complex mtDNA constitutes an approximate 620-kb insertion on *Arabidopsis thaliana* chromosome 2: implication of potential sequencing errors caused by large-unit repeats. Proc Natl Acad Sci USA **98:** 5099–5103

Tang J, Xia H, Cao M, Zhang X, Zeng W, Hu S, Tong W, Wang J, Wang J, Yu J, et al (2004) A comparison of rice chloroplast genomes. Plant Physiol **135:** 412–420

Timmis JN, Ayliffe MA, Huang CY, Martin W (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet **5:** 123–135

Timmis JN, Scott NS (1983) Spinach nuclear and chloroplast DNAs have homologous sequences. Nature **305:** 65–67

Unseld M, Marienfeld JR, Brandt P, Brennicke A (1997) The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. Nat Genet **15:** 57–61

Wolfe KH, Li W-H, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA **84:** 9054–9058

Wolfe KH, Sharp PM, Li W-H (1989) Rates of synonymous substitution in plant nuclear genes. J Mol Evol **29:** 208–211

Wu S-Y, Culligan K, Lamers M, Hays J (2003) Dissimilar mispair-recognition spectra of Arabidopsis DNA-mismatch-repair proteins MSH2·MSH6 (MutSα) and MSH2·MSH7 (MutSγ). Nucleic Acids Res **31:** 6027–6034

Yuan Q, Hill J, Hsiao J, Moffat K, Ouyang S, Cheng Z, Jiang J, Buell CR (2002) Genome sequencing of 239-kb region of rice chromosome 10L reveals a high frequency of gene duplication and a large chloroplast DNA insertion. Mol Genet Genomics **267:** 713–720

Zhang L, Vision TJ, Gaut BS (2002) Patterns of nucleotide substitution among simultaneously duplicated gene pairs in *Arabidopsis thaliana*. Mol Biol Evol **19:** 1464–1473

Zhu Y, Strassmann JE, Queller DC (2000) Insertions, substitutions, and the origin of microsatellites. Genet Res **76:** 227–236

Downloaded from on October 25, 2020 - Published by www.plantphysiol.org
Copyright © 2005 American Society of Plant Biologists. All rights reserved.