Generation and evolutionary fate of insertions of organelle DNA in the nuclear genomes of flowering plants

Christos Noutsos, Erik Richly and Dario Leister

Genome Res. 2005 15: 616-628
Access the most recent version at doi:10.1101/gr.3788705

Article

Generation and evolutionary fate of insertions of organelle DNA in the nuclear genomes of flowering plants

Christos Noutsos,[1] Erik Richly,[1] and Dario Leister[2]

Abteilung für Pflanzenzüchtung und Genetik; Max-Planck-Institut für Züchtungsforschung, D-50829 Köln, Germany

Nuclear genomes are exposed to a continuous influx of DNA from mitochondria and plastids. We have characterized the structure of ~750 kb of organelle DNA, distributed among 13 loci, in the nuclear genomes of *Arabidopsis* and rice. These segments are large and migrated to the nucleus quite recently, allowing us to reconstruct their evolution. Two general types of nuclear insertions coexist; one is characterized by long sequence stretches that are colinear with organelle DNA, the other type consists of mosaics of organelle DNA, often derived from both plastids and mitochondria. The levels of sequence divergence of the two types exclude their common descent, implying that at least two independent modes of DNA transfer from organelle to nucleus operate. The post-integration fate of organelle DNA is characterized by a predominance of transition mutations, associated with the gradual amelioration of the integrated sequence to the nucleotide composition of the host chromosome. Deletion of organelle DNA at these loci is essentially balanced by insertions of nonorganelle DNA. Deletions are associated with the removal of DNA between perfect repeats, indicating that they originate by replication slippage.

[Supplemental material is available online at www.genome.org.]

Transfer of DNA from mitochondria and/or plastids to the nucleus is a ubiquitous, ongoing evolutionary process, which has markedly influenced the evolution of eukaryotic genomes. During the early phase of organelle evolution, the process resulted in a massive relocation of organellar genes to the nucleus. In yeast, as many as 75% of all nuclear genes may derive from proto-mitochondria (Esser et al. 2004), while ~4500 genes in the nucleus of *Arabidopsis* are of plastid descent (Martin et al. 2002). In many eukaryotes, the transfer of functional genes is now rare or has ceased altogether (Boore 1999; Martin et al. 2002; Adams and Palmer 2003), but frequent transfer of mitochondrial genes for succinate dehydrogenase and ribosomal proteins has occurred in angiosperms relatively recently in evolutionary terms (Adams et al. 2000, 2001; Adams and Palmer 2003).

Almost all present-day transfers of mitochondrial (mtDNA) or plastid (ptDNA) DNA to the nucleus give rise to noncoding sequences, the so-called NUMTs (nuclear mtDNA) and NUPTs (nuclear ptDNA). NUPTs and NUMTs vary inter- and intraspecifically in size and copy number (Ayliffe et al. 1998; Pereira and Baker 2004; Richly and Leister 2004a,b), and they contribute to genetic variation via a mutation-inducing mechanism (Blanchard and Schmidt 1995; Bensasson et al. 2003; Ricchetti et al. 2004; Richly and Leister 2004b; Timmis et al. 2004). Experimental data (Ricchetti et al. 1999; Huang et al. 2003; Stegemann et al. 2003) and phylogenetic analyses (Mourier et al. 2001; Woischnik and Moraes 2002; Bensasson et al. 2003; Ricchetti et al. 2004) reveal that organelle-to-nucleus DNA transfer occurs frequently, yet the number and nature of molecular mechanisms underlying these integration events are still unclear.

[1]These authors contributed equally to this work.
[2]Corresponding author.
E-mail leister@mpiz-koeln.mpg.de; fax 49-221-5062413.
Article and publication are at http://www.genome.org/cgi/doi/10.1101/gr.3788705.

In flowering plants, large continuous stretches of organelle DNA have been identified in nuclear genomes. These are highly similar to their counterparts in mitochondria or plastids, and include a 620-kb mtDNA insertion in *Arabidopsis thaliana* (Lin et al. 1999; Stupar et al. 2001), as well as two large ptDNA fragments (33 kb; Yuan et al. 2002; and 131 kb; The Rice Chromosome 10 Sequencing Consortium 2003) in rice. Their existence indicates that such DNA segments can arise by nonhomologous recombination of nuclear sequences with DNA fragments that leak out of mitochondria or chloroplasts (Henze and Martin 2001; Timmis et al. 2004), and this notion is supported by systematic analyses of human NUMTs (Mourier et al. 2001; Woischnik and Moraes 2002). Since the level of sequence conservation between NUPTs/NUMTs and pt/mtDNA correlates with the size of the integrated fragment, the primary nuclear insertions of organelle DNA are assumed to have been large, decaying over evolutionary time into smaller blocks with an increasing level of sequence divergence from the organellar genome of origin (Richly and Leister 2004b). However, a second mode of generation of transfer has also been suggested, based on the finding of a large set of relatively small NUPTs and NUMTs with high homology to organellar DNA in land plants (Richly and Leister 2004b). In addition, some organelle DNA in the nuclear genome has been found to be organized in tandem arrays of fragments derived from disparate regions of the genome(s) of the same or different organelles (Blanchard and Schmidt 1995; Richly and Leister 2004b), suggesting that the generation and evolution of NUPTs and NUMTs involves their rearrangement before or after their integration into the nuclear genome.

Previous intracellular genome comparisons provided inventories of the number, size, and sequence divergence of nuclear organelle DNA in sequenced eukaryotes (Pereira and Baker 2004; Ricchetti et al. 2004, and references therein; Richly and Leister 2004a,b), but a systematic analysis of the structure and complexity of organelle DNA fragments has not been performed yet. In

15:616–628 ©2005 by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/05; www.genome.org

this study, we characterize the structure of 13 nuclear inserts of organelle DNA, which are sufficiently large and recent to allow the systematic study of their generation and evolution. This analysis has been carried out in *Arabidopsis* and rice, whose nuclear genomes are rich in organelle DNA (Pereira and Baker 2004; Richly and Leister 2004a,b), and allows us to propose modes of generation and evolution for nuclear organelle DNA.

Results

Identification of 13 large segments of organelle DNA in the nuclear genomes of *Arabidopsis* and rice

To identify large nuclear insertions of organelle DNA that were integrated sufficiently recently to allow the systematic study of their generation and evolution, the five *A. thaliana* and three completely sequenced chromosomes of rice were analyzed. Thirteen large segments of organelle DNA were identified, comprising a total ~750 kb of DNA (Table 1). The insertions were designated according to their size, origin, and chromosomal location; e.g., "92^{Os1}_{pm}" refers to the complex 92-kb locus containing both NUPTs and NUMTs on chromosome 1 of *Oryza sativa*. The insertions varied in size from ~3 to ~271 kb, and comprised up to 96 NUPTs and/or NUMTs. They could be classified into three groups as follows: (1) long continuous integrants containing only a few distinct NUPTs or NUMTs; (2) integrants containing multiple, rearranged NUPTs or NUMTs from disparate regions of either the mitochondrial or plastid chromosome; and (3) complex loci made up of DNA from both organelles (Table 1). While 33^{Os10}_{p} has been reported to be flanked by 2.1-kb direct repeats (Yuan et al. 2002), no such flanking repeats were detected at any of the other loci (data not shown).

The rice NUPTs 131^{Os10}_{p}, 33^{Os10}_{p}, 6.1^{Os1}_{p}, and 5^{Os4}_{p} are long continuous insertions

Four nuclear insertions consisting of only one or two long sequences that are colinear with ptDNA were identified (Fig. 1; Table 1). The largest, 131^{Os10}_{p} and 33^{Os10}_{p}, originally identified by Yuan et al. (2002) and the Rice Chromosome 10 Sequencing Consortium (2003), consist of one (131 kb) and two (11 and 22 kb) continuous NUPTs, respectively. Almost the entire ptDNA of rice is present at 131^{Os10}_{p} and the locus originated by linearization of the circular plastid chromosome in one of the inverted repeats (IR_A) (Fig. 1). The inversion of the small single-copy region (SSC) most probably arose by recombination between the IRs before insertion into the nucleus. The two NUPTs contained in 33^{Os10}_{p} originated from disparate regions of the plastid chromosome, whereas the two smaller ptDNA integrants 6.1^{Os1}_{p} and 5^{Os4}_{p} each consist of one NUPT (Fig. 1, Table 1).

Of these four NUPTs, 33^{Os10}_{p} is most similar to the plastid chromosome (99.85%; Table 2), whereas 6.1^{Os1}_{p} (98.61%) shows the least similarity. In both 131^{Os10}_{p} and 6.1^{Os1}_{p}, several InDels were found: seven large insertions of nuclear DNA and four deletions (Fig. 1; Supplemental Table 1).

The NUPTs 121^{Os4}_{p} and 24^{Os4}_{p} and NUMTs 271^{At2}_{m} and 27^{Os4}_{m}: Complex insertions containing rearranged organelle DNA

The rice NUPTs 121^{Os4}_{p} and 24^{Os4}_{p}

The insertions 121^{Os4}_{p} and 24^{Os4}_{p} consist of 19 and 13 clustered NUPTs, respectively (Fig. 2; Table 1). While 121^{Os4}_{p} includes five segments of ptDNA that are larger than 10 kb and represents most of the plastid chromosome, 24^{Os4}_{p} is made up of smaller NUPTs (<3.3 kb) derived from disparate regions of the plastome. Unlike the long continuous insertions described in the previous section, no large-scale colinearity between 121^{Os4}_{p} or 24^{Os4}_{p} and the ptDNA was evident; in both loci, the order and orientation of the NUPTs were random. Despite its highly fragmented state, the overall similarity of 121^{Os4}_{p} to ptDNA was higher than that of the long continuous insertion 131^{Os10}_{p} (Table 2). Sequences derived from the IRs were overrepresented in 121^{Os4}_{p} and 24^{Os4}_{p} (37.9%, 41.5%, and 30.9% for 121^{Os4}_{p}, 24^{Os4}_{p}, and ptDNA, respectively). InDels were also found in 121^{Os4}_{p} and 24^{Os4}_{p}; the largest, 24^{Os4}_{p}i2, is ~3.1 kb long and disrupts a NUPT derived from the IR region (Fig. 2; Table 2; Supplemental Table 1).

Table 1. Size, location, and composition of large nuclear insertions of organelle DNA in *Arabidopsis* and rice

Integrant	Size (bp)	Pseudo-chromosomal location	Type and number of distinct fragments contained
Long continuous integrants			
131^{Os10}_{p}	130,669	10,286,223–10,416,891	1 NUPT
33^{Os10}_{p}	32,974	19,948,780–19,981,753	2 NUPTs
6.1^{Os1}_{p}	6280	20,216,340–20,222,619	1 NUPT
5^{Os4}_{p}	5137	18,622,839–18,627,975	1 NUPT
Complex integrants containing organelle DNA from one organelle			
121^{Os4}_{p}	120,625	8,902,464–9,023,088	19 NUPTs
24^{Os4}_{p}	23,869	23,134,492–23,158,360	13 NUPTs
271^{At2}_{m}	270,815	3,238,037–3,508,851	16 NUMTs
27^{Os4}_{m}	26,730	32,368,959–32,395,688	5 NUMTs
Complex integrants with both types of organelle DNA			
92^{Os1}_{pm}	91,974	33,276,310–33,368,283	36 NUPTs; 11 NUMTs
18^{Os1}_{pm}	18,463	38,837,181–38,855,643	8 NUPTs; 87 NUMTs; 1 ambiguous fragment
12^{Os1}_{pm}	12,393	18,487,740–18,500,132	2 NUPTs; 27 NUMTs
6.2^{Os1}_{pm}	6304	34,148,238–34,154,541	3 NUPTs; 81 NUMTs
3^{Os10}_{pm}	2959	7,678,963–7,681,921	2 NUPTs; 29 NUMTs
Total	749,192	—	88 NUPTs, 256 NUMTs, 1 ambiguous fragment

(NUPT) Nuclear plastid DNA; (NUMT) nuclear mitochondrial DNA.

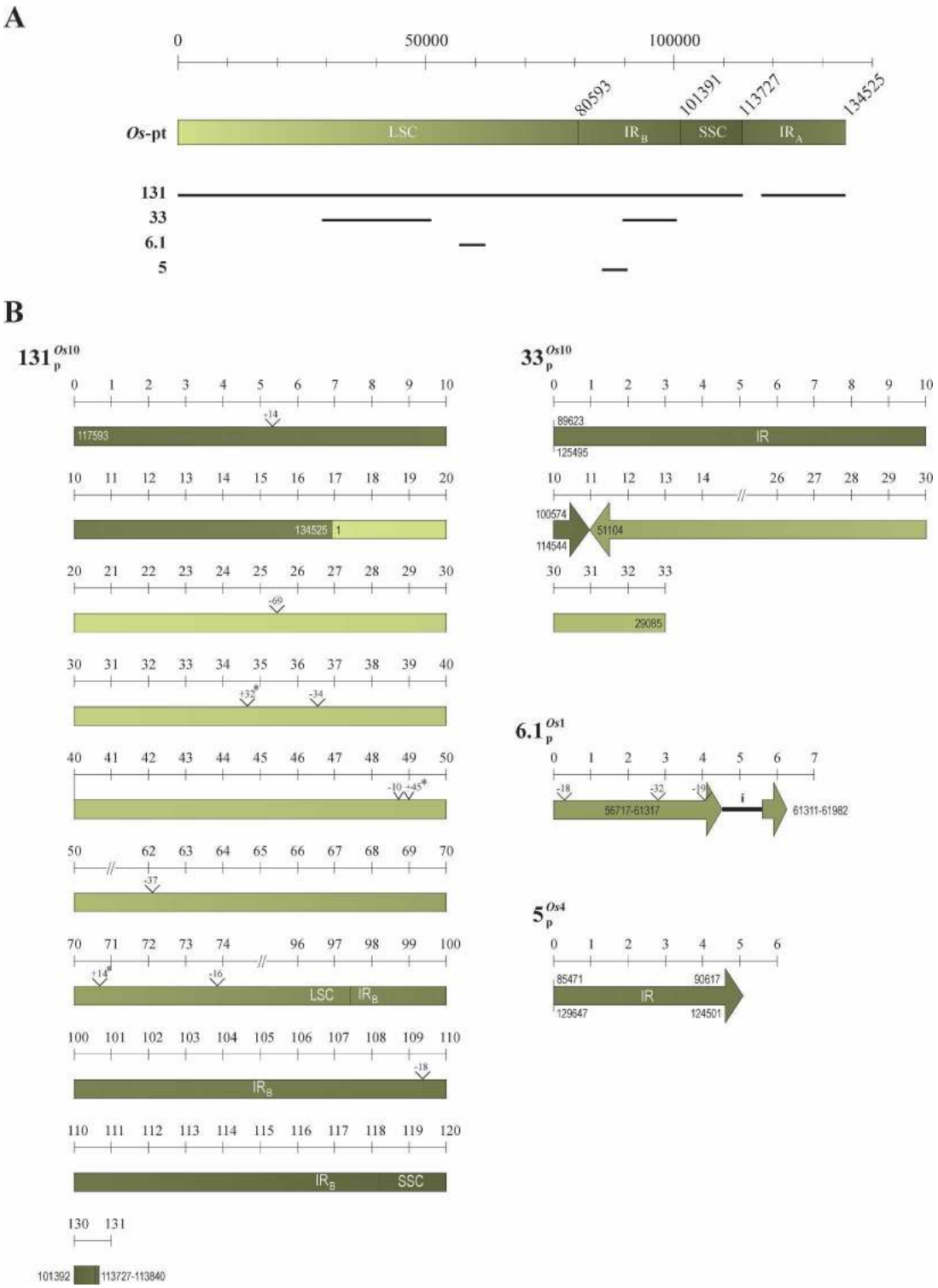

Figure 1. Structure of long continuous insertions of ptDNA in the nuclear genome of *O. sativa*. (*A*) Structure of rice ptDNA (*Os*-pt). The position of the long (LSC) and short (SSC) single-copy regions, as well as of the two inverted repeats (IR_A and IR_B) are indicated. The four nuclear insertions (see *bottom*) are depicted as black lines below their regions of origin in the rice plastid chromosome. When the origin of a fragment of nuclear ptDNA was ambiguous, i.e., when derived from repetitive regions, it was assigned to the repeat located most 5′ of the organellar DNA. For instance, NUPTs solely homologous to IR-specific sequences were always assigned to IR_B. (*B*) Structure of the four nuclear insertions of rice ptDNA 131_p^{Os10}, 33_p^{Os10}, 6.1_p^{Os1} and 5_p^{Os4}. Coloring and numbers indicate the position of the homologous sequence in the plastid chromosome based on the structure reported in *A*. The orientation of NUPTs relative to ptDNA is indicated by arrowheads; arrowheads pointing to the *left* indicate a reverse orientation. In the case of the IR regions of the ptDNA, the position of both homologous ptDNA sequences is provided. Fragments of ptDNA deleted from the nuclear integration are indicated by triangles with the size of the deletion indicated (e.g. "−14") (see Table 3); short insertions or duplications are indicated accordingly (e.g. "+45"; duplications are highlighted by an asterisk; Supplemental Table 2). For 6.1_p^{Os1}, the long insertion of nonorganelle DNA is indicated by a black line and the letter "i" (Supplemental Table 1).

Table 2. Overview of nucleotide composition and mutations of nuclear integrants of organelle DNA

Integrant	Similarity to pt/mtDNA (%)	GC content (%) [integrant/$\Delta_{(integrant\text{-}pt/mtDNA)}$]	Substitutions		InDels (number/total bp)	
			Transitions	Transversions	Deletions	Insertions
131_{p}^{Os10}	99.68	38.83/−0.08	232	168	82/220	66/169
33_{p}^{Os10}	99.85	41.22/−0.01	20	27	21/21	20/27
6.1_{p}^{Os1}	98.61	38.96/+3.05	34	28	12/81	10/1118
5_{p}^{Os4}	99.03	38.78/−0.29	30	14	9/14	4/4
121_{p}^{Os4}	99.82	40.35/+0.02	92	110	174/216	79/715
24_{p}^{Os4}	99.18	44.10/+5.27	98	35	8/376	23/6369
271_{m}^{At2}	99.85	44.63/−0.01	176	157	104/293	133/467
27_{m}^{Os4}	99.05	43.48/−0.19	131	64	27/431	13/4211
92_{pm}^{Os1}	99.84	41.29/+0.45	72	59	45/141	72/10,015
18_{pm}^{Os1}	99.29	42.81/−0.21	63	55	5/11	33/864
12_{pm}^{Os1}	99.27	42.40/−0.15	41	25	3/3	10/3451
6.2_{pm}^{Os1}	99.28	45.61/−0.2	22	9	1/1	14/2017
3_{pm}^{Os10}	99.34	49.17/−1.33	15	1	2/2	15/249
Total	99.48	42.17/+0.2	1026	757	493/1810	492/29,676

The 271/620–kb mtDNA insertion in A. thaliana

A 270-kb stretch of mtDNA sequence has previously been identified on chromosome 2 of *A. thaliana* (Lin et al. 1999). Portions of the mitochondrial chromosome initially believed to be absent from this insert were subsequently identified, and its actual size is now estimated to be ~620 kb (Stupar et al. 2001). Because the sequenced 270-kb stretch was markedly rearranged with respect to the published mtDNA sequence (Unseld et al. 1997), it was thought to have originated from an alternative, recombination-induced, configuration of the mitochondrial chromosome (Lin et al. 1999; Stupar et al. 2001).

In Figure 3, A and B, the structure of the sequenced 270-kb stretch (271_{m}^{At2}) is shown. As proposed earlier (Stupar et al. 2001), a D′–A′–C–B structure with a deletion at the D′–A′ junction can be recognized. Such an arrangement is likely to derive from the normal A–B–C–D structure modified by two recombination events (Fig. 3C). In addition to the almost complete nuclear copy of mtDNA, 13 smaller NUMTs were detected within the first 76 kb of 271_{m}^{At2}, which were mainly derived from the D and C regions of the mtDNA (Fig. 3D; Table 1). At least three major rearrangements have taken place in the first 76 kb of the insertion, as well as one event downstream of this region (see legend to Fig. 3).

The rice NUMT 27_{m}^{Os4}

The 27_{m}^{Os4} consists of five clustered NUMTs derived from noncontiguous regions of the mitochondrial chromosome (Fig. 3A; Table 1). This insertion shows less similarity (99.05%) to organellar DNA than most of the others (Table 2). Eight small and two large insertions of nuclear DNA are interspersed among the mtDNA sequences, and one of them (i2) disrupts a repeat IV sequence (Fig. 3A; Table 2; Supplemental Table 1).

Rice loci 92_{pm}^{Os1}, 18_{pm}^{Os1}, 12_{pm}^{Os1}, 6.2_{pm}^{Os1}, and 3_{pm}^{Os10} are complex insertions containing rearranged DNA derived from mitochondria and chloroplast

Five insertions were identified that included both NUPTs and NUMTs (Fig. 4; Table 1). The level of their fragmentation was much higher than that seen in the other insertions, the number of distinct fragments of organelle DNA ranging from 29 (12_{pm}^{Os1}) to 95 (18_{pm}^{Os1}). As in the case of 121_{p}^{Os4}, 24_{p}^{Os4}, 271_{m}^{At2}, and 27_{m}^{Os4} described above, these insertions are characterized by a lack of extended colinearity with organellar DNA, and a random order and orientation of the individual fragments.

The largest of these five insertions, 92_{pm}^{Os1}, is made up of 57,983 bp of ptDNA organized into 36 NUPTs and 24,144 bp of mtDNA organized in 11 NUMTs. In most cases, NUPTs and NUMTs are adjacent to one another; no nuclear "stuffer" sequence is detectable. A number of InDels were found in 92_{pm}^{Os1} (Table 2; Supplemental Table 1).

18_{pm}^{Os1} and 12_{pm}^{Os1} consist of 95 and 29 distinct organelle DNA fragments, respectively, and on average, these are markedly smaller than the ones in 92_{pm}^{Os1} (1579, 184, and 298 bp in 92_{pm}^{Os1}, 18_{pm}^{Os1}, and 12_{pm}^{Os1}, respectively). Both loci consist mainly of NUMTs with only a few NUPTs. In 12_{pm}^{Os1}, a relatively large NUPT (1891 bp) is disrupted by a 3421-bp insertion of nuclear DNA.

The two small insertions 6.2_{pm}^{Os1} and 3_{pm}^{Os10} are made up of 84 and 31 distinct NUMTs or NUPTs, respectively, with an extraordinarily small average size (72 and 87 bp for 6.2_{pm}^{Os1} and 3_{pm}^{Os10}, respectively). As in the other members of this group, in almost all cases, no nuclear "stuffer" DNA is found between adjacent fragments.

Mutations in nuclear organelle DNA: Predominance of nucleotide transitions and insertions

Spectrum of nucleotide substitution and composition

With few exceptions (Cho et al. 2004), nucleotide substitutions occur much less frequently in the organellar DNAs of plants than in the nuclear genome (Wolfe et al. 1987; Palmer and Herbon 1988; Gaut 1998). The incidence of substitutions in our 13 inserts was computed by comparing their sequences with published mtDNA and ptDNA sequences (Hiratsuka et al. 1989; Unseld et al. 1997; Notsu et al. 2002). The rate of substitution varied between 1.23 (271_{m}^{At2}) and 12.01 (6.1_{p}^{Os1}) nt/kb and, except in the cases of 33_{p}^{Os10} and 121_{p}^{Os4}, the rate of transitions exceeded the rate of transversion mutations (Table 2). G→A and C→T transitions formed the highest-frequency substitution class (each of

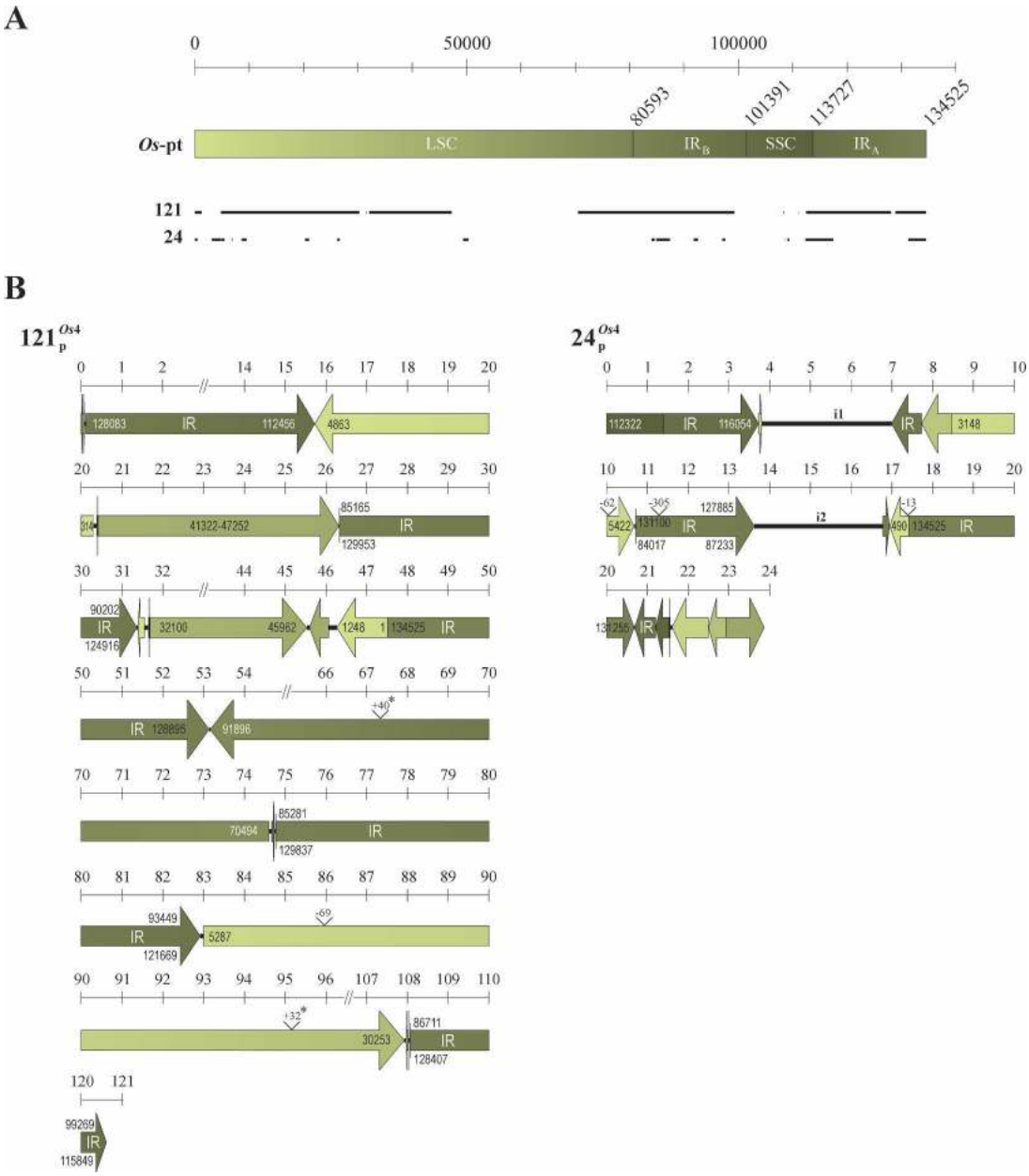

Figure 2. Structures of the complex integrants 121_p^{Os4} and 24_p^{Os4}, which contain rearranged DNA from the rice plastid chromosome. *A* and *B* are depicted as in Figure 1. For 24_p^{Os4}, the two large insertions of nonorganelle DNA are designated as "i1" and "i2" (Supplemental Table 1).

them representing 21% of all substitutions; data not shown), similar to the substitution pattern observed in prokaryotes (Kowalczuk et al. 2001), nematodes (Blouin et al. 1998), *Drosophila* and mammals (Petrov and Hartl 1999), and Cucurbitaceae (Decker-Walters et al. 2004).

The G/C content of the inserts ranged between 38.83% and 49.17%, and differed from that of the corresponding organellar sequence by between −1.33% and +5.27% (Table 2). In nine inserts, the G/C content was changed toward that of the nuclear chromosome, albeit without reaching the same level. Major differences in G/C content with respect to organellar homologs (3_{pm}^{Os10}, 24_p^{Os4}, and 6.1_p^{Os1}) were noted for insertions in nuclear chromosomal regions that have a markedly different G/C content (Fig. 5A). Interestingly, the highest transition/transversion ratios were found in 3_{pm}^{Os10} and 24_p^{Os4}, which also show the highest degree of amelioration to the G/C content of the host chromosome (Table 2; Fig. 5A). This implies that the high rate of transitions was responsible for their amelioration to the nucleotide composition of the nuclear DNA.

Expansion of NUMT/NUPT loci by further incorporation of DNA

Overall, the numbers of insertion and deletion events that occurred after integration of the 13 segments of organelle DNA are almost identical (492 vs. 493 events). The total length of DNA inserted post-integration, however, is at least 16 times larger than the amount of organelle sequences deleted (29,676 vs. 1810 bp).

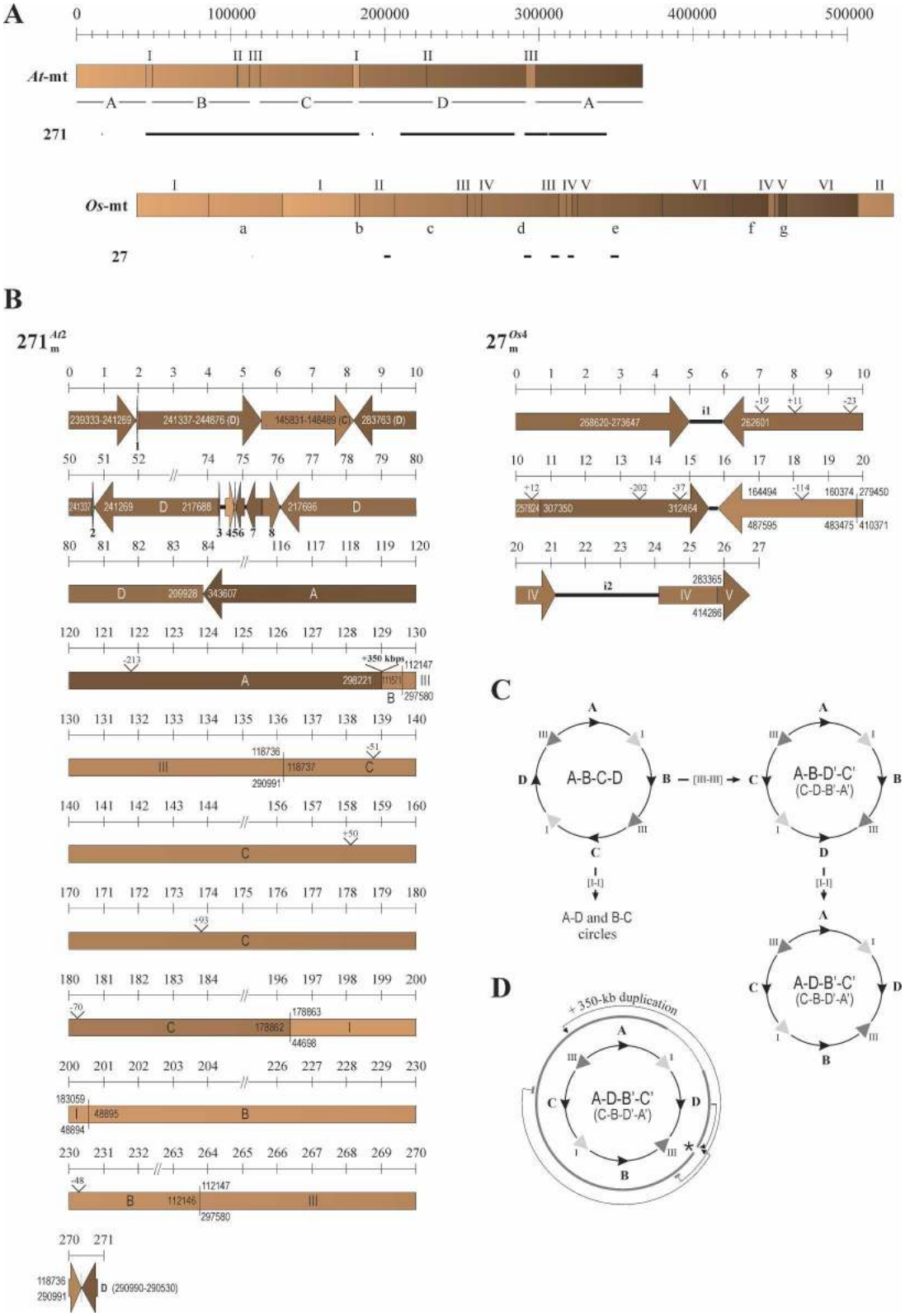

Figure 3. (Legend on next page)

Tandem duplications of organelle DNA already integrated in the nucleus also contributed to the expansion of nuclear loci of organellar origin. However, the frequency of such events was relatively low, and only seven resulted in duplication of stretches larger than 10 bp (Supplemental Table 2). By comparison, the 10 largest insertions of nonorganelle DNA into nuclear DNA of organellar origin amount to 23,805 bp (Supplemental Table 1). These insertions were frequently associated with duplication of short regions at the target site, and two of them were similar to transposons, including the presence of terminal inverted repeats (Supplemental Table 1). The insertion 24_{p}^{Os4}i1 contains a complete nuclear gene encoding a protein kinase, while 92_{pm}^{Os1}i3 is homologous to mtDNA of maize (Supplemental Table 1). The latter sequence is not present in rice mtDNA, indicating that it represents a part of the mitochondrial chromosome that was lost after the insertion of 92_{pm}^{Os1}.

The data thus indicate that large NUMT/NUPT loci were expanded by subsequent insertions of nonorganelle DNA. This also explains the decay of colinearity with organellar DNA and the formation of the so-called "loose clusters" (Richly and Leister 2004b).

Deletion of organelle DNA is associated with direct repeats and palindromic sequences

The presence of deletions of up to several hundred base pairs relative to organellar DNA is characteristic of the nuclear insertions. Twenty-five deletions involved stretches longer than 10 bp (18 are listed in Table 3). Parallel losses were observed for two regions derived from rice ptDNA; a 69-bp stretch corresponding to positions 8535–8603 of ptDNA was absent in both 131_{p}^{Os10} and 121_{p}^{Os4}. In contrast, in the region corresponding to positions 57010–57028 of the plastid chromosome, three independent—and slightly different—deletions were observed in 131_{p}^{Os10}, 6.1_{p}^{Os1}, and 92_{pm}^{Os1} (Table 3). Both repeatedly deleted regions contained two perfect direct 13-bp repeats, one of which was always deleted, together with the inter-repeat region. This, together with the presence of a palindromic sequence in the inter-repeat regions (Table 3), is characteristic for RecA-independent deletion due to replication slippage (for review, see Bzymek and Lovett 2001; Lovett 2004). In the case of the three deletions in the region homologous to positions 57010–57028 of ptDNA, the slight differences in the regions affected imply that the repeat-removing mechanism is somewhat imprecise.

In addition to such parallel losses, 13 unique events were noted (Table 3). These deletions were also associated with the loss of repetitive sequences from nuclear organelle DNA. In such cases, both direct repeats and the inter-repeat palindromic sequences were shorter than those just discussed and/or imperfect. This is compatible with the observation that the frequency of replication slippage depends on the repeat length (Bzymek and Lovett 2001). An additional set of seven deletions was detected, in which neither or both direct repeats were deleted (Supplemental Table 3), indicating variation in replication misalignment or the action of secondary deletion events.

The presence in the insertions of regions that are particularly susceptible to deletion events allows us to further quantify the frequency of deletions. Parallel losses were indeed detected in all insertions that harbored the respective region of the organellar DNA. In contrast, for many of the unique losses, the undeleted version was also detected at a different locus; for example, the −37 deletion found in 131_{p}^{Os10} did not occur in either 33_{p}^{Os10} or 121_{p}^{Os4} (Figs. 1, 2).

Mode of generation of nuclear insertions of organelle DNA: Long continuous fragments versus mosaics

Relative age of insertions

The large insertions of organelle DNA are organized as either continuous or highly fragmented segments. The obvious conclusion is that the fragmented versions derive from originally con-

Figure 3. Structures of the complex integrants 271_{m}^{At2} and 27_{m}^{Os4}, which contain rearranged DNA from the mitochondrial chromosome of *Arabidopsis* and rice, respectively. (*A*) Structure of the *Arabidopsis* (*At*-mt) and rice (*Os*-mt) mtDNAs. For the *Arabidopsis* mtDNA, the position of the four single-copy regions A, B, C, and D, as well as the three pairs of specific repeats (I, II, and III), are presented. Repeats I (positions 44,698–48,894 and 178,863–183,059) and II (103,805–104,337 and 227,087–227,619) are directly oriented, while the two repeat III sequences (112,147–118,736 and 297,580–290,991) are inverted. The portion of the mitochondrial chromosome included in each nuclear insertion (*bottom*) is indicated by black lines (as in the previous figures). (*B*) Structure of 271_{m}^{At2} and 27_{m}^{Os4}, depicted according to the scheme used in Figures 1 and 2. For 271_{m}^{At2}, small NUMTs are indicated by Arabic numerals and the positions of their homologs in the *Arabidopsis* mtDNA are as follows: 1 (170,597–170,530); 2 (170,530–170,597); 3 (191,246–191,300); 4 (25,397–25,647); 5 (28,644–28,687); 6 (279,338–279,090); 7 (280,258–279,809); and 8 (191,552–192,070). Four major rearrangements can be recognized (see text). (1) A 5.5-kb region, derived from the D-region, is duplicated (positions 1–5545 and 47,106–52,650) in 271_{m}^{At2}. The two sequences are more similar to each other than to mtDNA, and both harbor a characteristic 68-bp insertion derived from region C (designated as "1" and "2"), suggesting that the insertion of this specific mtDNA segment in the nucleus occurred before the duplication. (2) A 2.6-kb region (positions 5546–8207), derived from region C, is inserted between the D′ terminus of the mtDNA insertion (see above) and the 5.5-kb duplication described above. (3) A 1.8-kb stretch, containing a structure of six short NUMTs, together with very short stretches of nonorganelle DNA, is present at positions 74,295–76,084. The six NUMTs derive from all four major regions of the mtDNA (including those absent at the A/D junction: see text and below). The 1.8-kb insertion is flanked by a duplication of 9 bp (CTTTACGAG) present in the D region, implying that it was inserted after formation of the D′–A′–C–B structure. (4) At positions 129,022–129,597, between A region and III-type repeat, a short stretch of the B region is found. This could be the result of imprecise homologous recombination between repeat III sequences, affecting the B region adjacent to one of the repeats. Alternatively, this short region might represent the beginning of the 350-kb duplication that was previously detected in 271_{m}^{At2}, but not sequenced (Stupar et al. 2001). For 27_{m}^{Os4}, the two large insertions of nonorganelle DNA into nuclear DNA of organellar origin are designated as "i1" and "i2" (Supplemental Table 1). (*C*) Rearrangements of the *Arabidopsis* mtDNA due to recombination across repeat regions. The four single-copy regions A, B, C, and D are separated by two pairs of repeats (I: direct repeats, III: inverted repeats); recombinations involving repeat II sequences are not shown because they are not relevant for the generation of 271_{m}^{At2}. Recombination across repeat II sequences results in A–D and B–C circles, while the A–B–D′–C′ structure derives from III–III recombination (D′ and C′ refer to inverted D and C sequences, respectively). The A–D–B′–C′ arrangement originates from the normal A–B–C–D structure by two rounds of recombination; the first, between the pair of repeat III sequences, inverts the orientation of one of the repeat I sequences, allowing a second recombination between the now inverted repeat I sequences, resulting in an A–D–B′–C′ circle. Another alternative structure, A–C′–B–D′, has been described before (Klein et al. 1994; Unseld et al. 1997). (*D*) Origin of 271_{m}^{At2}. The structure of 271_{m}^{At2} is depicted as a circle around mtDNA of the A–D–B′–C′-type. The 271_{m}^{At2} insertion contains the four major segments of mtDNA in the order D′–A′–C–B, whereby parts of D′ and A′ are either absent in the nuclear mtDNA sequence, or form part of the 350-kb duplication for which no sequence information is available (Stupar et al. 2001). The deletion at the A/D junction is indicated by the dotted segment of the circle and contains mtDNA from position 183,060–209, 928 (region D), an entire I-type repeat, and the region extending from 343,608 to 44,697 (from region A) (see *A*). The breakpoint (indicated by an asterisk) that gave rise to the linear D′–A′–C–B structure should be located in region D, close to the type III repeat; in fact, both ends of 271_{m}^{At2}consist of sequences from the D region. Additional rearrangements of 271_{m}^{At2}, in particular the duplication resulting in the 350-kb region identified before (Stupar et al. 2001), are indicated by arrows.

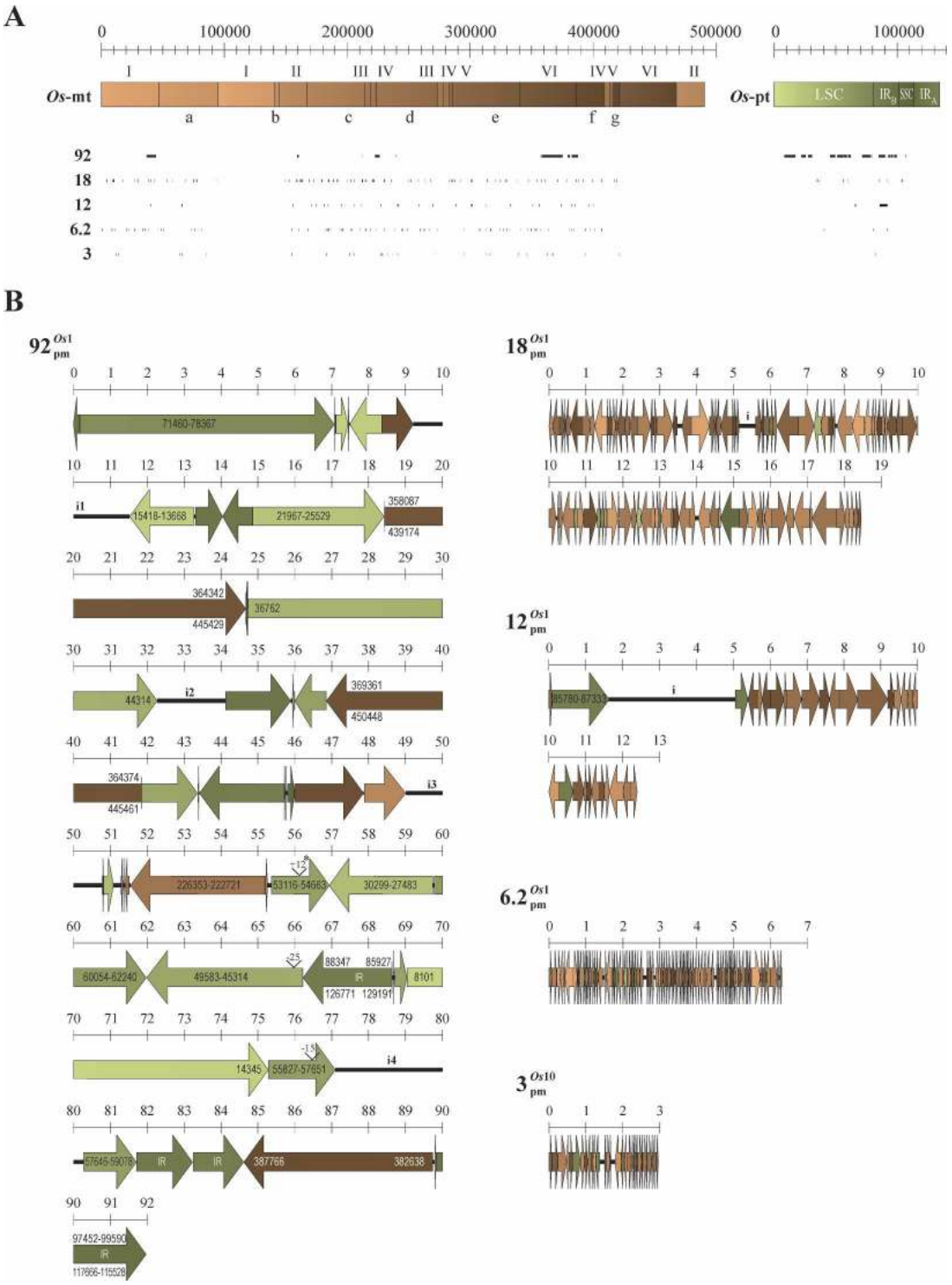

Figure 4. Structure of the complex insertions 92^{Os1}_{pm}, 18^{Os1}_{pm}, 12^{Os1}_{pm}, 6.2^{Os1}_{pm} and 3^{Os10}_{pm}, which contain rearranged DNA from both ptDNA and mtDNA of rice. (*A*) The structures of mtDNA and ptDNA and the portions found in the five insertions (see *B*) are shown, following the scheme used in the previous figures. In *B*, the structures of the five nuclear insertions containing both ptDNA and mtDNA (92^{Os1}_{pm}, 18^{Os1}_{pm}, 12^{Os1}_{pm}, 6.2^{Os1}_{pm}, and 3^{Os10}_{pm}) are shown. For 92^{Os1}_{pm}, the four large inserts of nonorganelle DNA are designated as "i1" to "i4", and for 18^{Os1}_{pm} and 12^{Os1}_{pm} as "i" (Supplemental Table 1).

tinuous ones by rearrangement; i.e., fragmented integrants are older than continuous ones. An alternative explanation is that continuous and highly fragmented integrants might represent two distinct modes of DNA transfer.

To permit one to distinguish between the two scenarios, it is crucial to determine the relative ages of the 13 primary insertions. Two independent measures allow us to assess relative ages as follows: (1) the degree of sequence identity between integrated

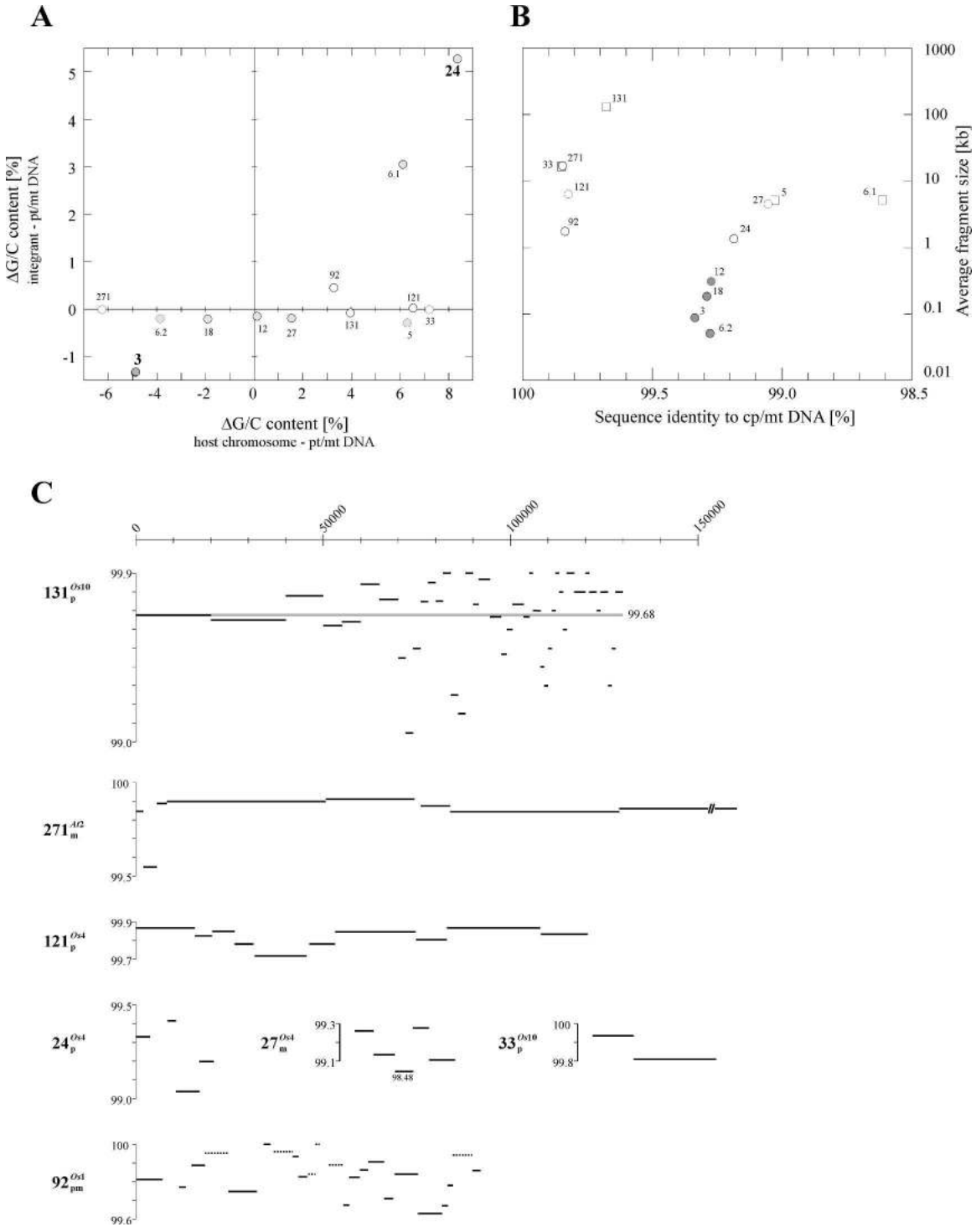

Figure 5. (Legend on next page)

and organellar DNA, taking into account the rate of nucleotide substitutions and of post-integration insertions/deletions; (2) the pattern of distribution of large deletions throughout the 13 insertions.

With respect to sequence identity, the insertions can be ranked by descending levels of identity in the following order: 33_p^{Os10} - 271_m^{At2} - 92_{pm}^{Os1} - 121_p^{Os4} - 131_p^{Os10} - 3_{pm}^{Os10} - 18_{pm}^{Os1} - 6.2_{pm}^{Os1} - 12_{pm}^{Os1} - 24_p^{Os4} - 27_m^{Os4} - 5_p^{Os4} - 6.1_p^{Os1} (Table 2). With respect to the distribution pattern of large deletions, six of the integrated segments contained at least five organelle sequences that include repetitive sequences susceptible to deletion (see above; Supplemental Table 4). Thus, the presence/absence of deletions in these regions allows us to produce an estimate of their relative ages. For nuclear ptDNA, the ranking order in descending level of deletion

Table 3. Deletions in nuclear organelle DNA

Integrant/deletion size[a]	Position of deletion and of direct repeats[b]	Position and structure of inverted repeats[c]
Parallel losses		
131_p^{Os10} (−69)	^{8534}Tttacttttttttcag-----gTTACTTTTTTTCA^{8616}	8587–8619 (13/23)
121_p^{Os4} (−69)		
131_p^{Os10} (−16)	^{57005}ATTctttttttttaga ataCTTTTTTTTTAGA^{57036}	57008–57061 (20/43)
6.1_p^{Os1} (−18)	^{57005}ATTCTTttttttttaga atactttttTTTTTAGA^{57036}	57017–57068 (24/52)
92_{pm}^{Os1} (−15)	^{57005}ATTCTTttttttttaga atacTTTTTTTTTAGA^{57036}	
Unique losses		
131_p^{Os10} (−14)	^{122921}AtgctccaccgcttgTGC^{122936}	122917–122940 (7/13)
		122918–122937 (7/16)
131_p^{Os10} (−34)	^{19671}GaggaagatcggaattagcaattgataaaaaagaaAGGA^{19709}	19661–19683 (7/13)
131_p^{Os10} (−37)	^{45241}CataatgaaaacgcaatttctgtttttcttgaagacgaATA^{45281}	45247–45267 (7/13)
131_p^{Os10} (−18)	^{92552}TtgcaggctgcaactcgccTGCA^{92574}	1692553–92574 (9/20)
24_p^{Os4} (−13)	^{2}CCAatatcttgcttcaGCAAGATAT^{26}	134519–36 (20/47)
271_m^{At2} (−213)	^{305456}Actctaccaattg-----ttccaccCTCGTCCAAT^{305679}	3005546–3005576 (8/16)
		3005593–3005609 (7/14)
271_m^{At2} (−51)	^{121312}Tcttcttcctcttcc-----tcgagaCT.CTTCCT^{121371}	121336–121368 (9/16)
271_m^{At2} (−70)	^{162645}TCTTattgtg-----gatagctcttG^{162719}	162664–162693 (11/25)
		162701–162725 (9/19)
271_m^{At2} (−48)	^{78599}GTGAggaacg-----gcagacgtgaA^{78651}	78588–78609 (8/17)
27_m^{Os4} (−23)	^{258872}ATAAGCTTTaccggttaaaccttataagctttG^{258904}	258874–258902 (6/14)
27_m^{Os4} (−202)	^{310215}GACcaacctcgct-----ctaactCAAC^{310423}	310344–310384 (13/24)
27_m^{Os4} (−37)	^{311575}Aggaaactagcacttt-----aatgatGGAA.CTAG^{311620}	311554–311579 (8/15)
27_m^{Os4} (−114)	^{161943}ATGCccaggg-----gaaaaaatgcA^{162061}	161935–161955 (8/16)
		162009–162034 (8/17)

[a]Name of the locus and size of the deletion (in parentheses).
[b]Nucleotides in lower case are absent in NUMTs or NUPTs, but present in organellar DNA (positions in organellar DNA are indicated in superscript). Direct repeats are underlined.
[c]Positions of inverted repeats are given relative to organelle DNA, together with the length/number of bonds in resulting stem-loop structures (in parentheses).

frequency is as follows: 33_p^{Os10} - 121_p^{Os4} - 131_p^{Os10} - 24_p^{Os4}. When the incidence of direct-repeat-associated deletion of nuclear mtDNA was considered, 27_m^{Os4} appears to be older than 271_m^{At2} (Supplemental Table 4), which supports the ranking obtained on the basis of sequence identity.

Mode of integration: Single events vs. hot spots

The two approaches outlined above yield coherent estimates for the relative ages of the different primary insertions. Strikingly, the highly fragmented insertions, in general, do not appear to be older than the long continuous ones. This is especially evident for 92_{pm}^{Os1} and 121_p^{Os4}, which appear to have entered the nuclear genome at a later time than the largest continuous integrant, 131_p^{Os10}. The most fragmented integrations (3_{pm}^{Os10}, 18_{pm}^{Os1}, 12_{pm}^{Os1}, and 6.2_{pm}^{Os1}) seem to have intermediate relative ages, and should be younger than the less fragmented 24_p^{Os4}, 27_m^{Os4}, 5_p^{Os4}, and 6.1_p^{Os1} (Fig. 5B). Thus, it appears that highly fragmented inserts have not evolved from the long continuous type. Based on these data, we propose that continuous and highly fragmented integrants were initially generated in different ways.

Figure 5. Relationships between base composition, structure, and sequence divergence of large insertions of nuclear organelle DNA. (*A*) The G/C content of the insertion comes to resemble that of its nuclear chromosomal neighborhood. The primary insertion is assumed to have had the same G/C content as the corresponding region of pt/mtDNA. The G/C contents of the chromosomal regions hosting the insertions are based on those of the 300 kb of nuclear DNA immediately flanking the respective insertion. White boxes indicate segments with an overall similarity of >99.5% to pt/mtDNA; the insertions indicated as gray boxes are less similar to organellar DNA. The two integrants with the highest transition/transversion ratios are indicated in bold. (*B*) Relationship between average fragment size and sequence divergence. The level of sequence identity between insertion and organellar DNA was calculated using the BESTFIT algorithm of the GCG package, which considers both nucleotide exchanges and InDels (Devereux et al. 1984). Average fragment sizes were obtained from columns 2 and 4 of Table 1. Squares indicate long continuous integrants; open circles symbolize complex insertions derived from one organelle; and shaded circles stand for complex insertions, including sequences from both organelles. (*C*) Intra-insertion patterns of sequence divergence. Sequence identity between NUPTs and ptDNA and between NUMTs and mtDNA was assessed for the five largest complex loci as in *B*. NUPTs and NUMTs larger than 1 kb were considered. Insertion 131_p^{Os10} served as control. It consists of one large continuous NUPT subdivided into sets of fragments with sizes ranging from 20 kp to 1 kb; the level of its overall sequence identity is indicated by a gray line. In 92_{pm}^{Os1}, NUPTs are indicated by continuous lines and NUMTs by dotted lines. Note that the more diverged fragments in 121_p^{Os4} (26,331–31,363; 31,686–45,533; 46,267–53,135) are mostly derived from the IR region; in 92_{pm}^{Os1} (55,367–56,927; 66,220–68,636; 75,288–81,711; 81,712–83,231) they originate from the IR region, and from positions 55,000–60,000 of ptDNA. The most diverged regions in 24_p^{Os4} (10,721–16,939; 16,940–20,685) derive again from IR sequences.

Entire organellar chromosomes, or long fragments of organelle DNA, are thought to integrate into the nuclear genome by nonhomologous recombination (Henze and Martin 2001; Timmis et al. 2004). It is more difficult to explain the origin of mosaic insertions derived from different regions of the same organellar chromosome, or even from different organelles. Three mechanisms, in principle, could generate mosaic molecules for integration, i.e., (1) random end-joining of different linear DNA fragments before integration; (2) rapid rearrangement during, or directly after, insertion; (3) ongoing integration of organelle DNA at the same locus, corresponding to a "hotspot" for integration. The data we have presented do not allow us to choose definitively between the three scenarios. However, several of our observations support the relevance of a random end-joining mechanism. (1) When distinct fragments of nuclear organelle DNA are compared with organellar DNA, sequence divergence varies markedly within highly fragmented integrants (121_{p}^{Os4}, 24_{p}^{Os4}, 27_{m}^{Os4}, and 92_{pm}^{Os1}); the degree of this variation, however, does not appear to be markedly different from that observed for the long continuous insertion 131_{p}^{Os10} when the latter is subdivided into smaller segments (Fig. 5C). (2) In highly fragmented integrants, no sequence homology between adjacent fragments, or their template regions in organellar DNA, was evident, excluding the possibility that pre-existing insertions might have facilitated subsequent integrations at the same locus due to homologous recombination. (3) All highly fragmented integrants (3_{pm}^{Os10}, 18_{pm}^{Os1}, 12_{pm}^{Os1}, 6.2_{pm}^{Os1}) show an intermediate level of sequence divergence from the organellar template DNA, which is, however, lower than that associated with the less fragmented complex insertions, such as 121_{p}^{Os4} and 92_{pm}^{Os1} (Fig. 5B). This may suggest that the first type of integrants derives from less fragmented ones.

Discussion

Elucidating the mechanisms by which organelle DNA enters nuclear chromosomes requires detailed insights into the structure of integrated fragments. In plants, moreover, the presence of two distinct types of organelle DNA in the nuclear genome allows one to discriminate between intra- and interchromosomal rearrangements of non-nuclear organelle DNA. In this study, we have characterized the structures of segments of organelle DNA integrated in the chromosomes of *Arabidopsis* and rice. These insertions are large and occurred relatively recently, making it possible to reconstruct their evolution. Thirteen such insertions, totaling ~750 kb, were characterized with respect to their sequence similarity to and colinearity with the organellar genomes from which they originated.

These nuclear loci were selected because of their large size. Almost all of them were more than 99% identical to their organellar genome of origin, allowing an unambiguous assignment (344 of 345 NUPTs or NUMTs; see Table 1) of the contained nuclear DNA to either a plastid or mitochondrial origin and indicating that they were transferred to the nucleus quite recently. The inserts that were most similar to their organellar ancestors were also the largest (Table 1; Fig. 5B), and this supports the idea that primary insertions of organelle DNA are large and then diverge over evolutionary time, fragmenting initially long blocks of homology into shorter stretches (Richly and Leister 2004b). In previous analyses of genomes of flowering plants, also NUPTs and NUMTs with sequence identities to cp/mt DNA lower than 90% have been identified; in all cases, however, these fragments were smaller than 1 kb (Blanchard and Schmidt 1995; Shahmuradov et al. 2003; Richly and Leister 2004b). Our data, however, are not compatible with the view that all primary insertions represent complete, or large fractions of, organellar chromosomes. A subset of the very large inserts, 121_{p}^{Os4} and 92_{pm}^{Os1}, in fact, comprise mosaics of organelle DNA—consisting in the case of 92_{pm}^{Os1} of segments from both plastid and mitochondrial chromosomes. We could not discern any obvious variation in degree of divergence between mtDNA and ptDNA sequences in 92_{pm}^{Os1}. Thus, they probably arose by random end-joining of linear DNA fragments and/or rearrangement of distinct fragments of organellar DNA prior to insertion into the nuclear genome, rather than by multiple insertions of organelle DNA into the same locus over an extended period of time (a hotspot mechanism).

The presence of two types of mosaic arrangements—one with large fragments relatively similar to organellar DNA, and the other with, in part, extremely short fragments that are more diverged in sequence—suggests that the latter type is derived from the former. Because no homologies were detected between adjacent fragments in the complex insertion loci, homologous recombination between different regions of the same organellar chromosomes or between different molecules probably did not play a major role in this process. We conclude that more or less intact organellar chromosomes can give rise to very large continuous integrants (e.g., 271_{m}^{At2} and 131_{p}^{Os10}), or, after degradation, to smaller fragments that are ligated before integration to form the progenitors of complex nuclear loci. Since complex loci containing DNA from both organelles are relatively frequent, one must conclude that the concomitant release of ptDNA and mtDNA is not a rare event, possibly taking place under conditions that affect both organelles, such as cellular stress or gametogenesis-associated organellar degradation (Adams and Palmer 2003; Richly and Leister 2004b; Timmis et al. 2004). Alternatively, transfer of organelle DNA to other cellular compartments might be facilitated by physical interactions between organelles (Yu and Russell 1994; Kwok and Hanson 2004).

Organelle DNA that has been incorporated into the nuclear genome is exposed to the evolutionary influences that act on this compartment. From this point of view, the 13 large insertions can be considered as good probes for the types of mutation that occur in nuclear DNA. In general, NUMT/NUPT loci appear to increase in size due to the incorporation of novel DNA; the amount of sequence lost by deletion is comparatively much less. However, the possibility cannot be excluded that locus expansion is paralleled by a similar, or even greater loss of organelle DNA at its ends—such a process would have remained undetected in our analysis. Nevertheless, it is clear that insertions of nonorganelle DNA result in the fragmentation of initially large blocks of homology, generating loose clusters of NUPTs and NUMTs.

The direct-repeat-associated deletion of nuclear organelle DNA seems to result in rearrangements similar to those that arise due to replication slippage, as described in eukaryotes and prokaryotes (Lovett 2004). It is interesting to note that—in contrast to *Escherichia coli* (Bzymek and Lovett 2001; Lovett 2004)—in nuclear organelle DNA, regions flanked by direct repeats are rarely duplicated, indicating that such regions are more efficiently removed than generated. In contrast, the persistence of repetitive sequences in the organellar genomes might be due to positive selection; about half of the deletion-susceptible regions listed in Table 3 are, in fact, located in coding regions of the respective organellar chromosomes (data not shown).

Nucleotide substitutions occur more often in the nuclear genome than in plastid or mitochondrial chromosomes (Wolfe et al. 1987; Palmer and Herbon 1988; Gaut 1998), and this allows the identification of transitions and transversions in nuclear organelle DNA. In most insertions, nucleotide transitions predominate, changing the sequence composition so that it more closely approximates that of the surrounding nuclear DNA; i.e., the mutagenic processes active in the nucleus result in an amelioration of the organelle DNA to its host chromosome—a process, which is still ongoing in most insertions (see Fig. 5A).

The observed prevalence of C→T and, concomitantly, of G→A transitions on the opposite strand should be due to a high rate of methylation of cytosine residues, followed by deamination (Holliday and Grigg 1993), which in turn could be the result of the nonfunctional nature of nuclear organelle DNA associated with loss of its transcription (Paszkowski and Whitham 2001; Bender 2004). This, however, cannot explain the increase in G/C content observed in four insertions (see Table 2), and indicates that multiple mechanisms operate to ameliorate the organelle DNA to its new environment.

Taking all of the data together, the structure of nuclear inserts of organelle DNA suggests that they originate by two different modes of organelle-to-nucleus DNA transfer and provides insights into some features of the mutational processes affecting nuclear DNA. A possible application of this type of comparative genomics concerns the evolutionary reconstruction of major rearrangements of organellar chromosomes. As shown here, inserts of organelle DNA in nuclear genomes can retain sequences that are absent from the organellar chromosomes of the same species, but highly homologous to organellar DNA from related species (e.g., 92^{Os1}_{pm}i3).

Methods

Identification of large inserts of nuclear organelle DNA

Full-length nucleotide sequences of organellar chromosomes were retrieved from NCBI (http://www.ncbi.nlm.nih.gov/). Nuclear DNA sequences were obtained as pseudo-chromosomes from MIPS (http://mips.gsf.de/proj/thal/db/index.html; *Arabidopsis thaliana*) and TIGR (ftp://ftp.tigr.org/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/; *Oryza sativa* ssp. *japonica*). For the analysis of clusters of NUPTs or NUMTs in rice, only the completely sequenced chromosomes 1 (Sasaki et al. 2002), 4 (Feng et al. 2002), and 10 (Rice Chromosome 10 Sequencing Consortium 2003) were considered.

NCBI-BLASTN (Altschul et al. 1990) was carried out locally with standard settings and without low-complexity filtering as described before (Richly and Leister 2004a,b). NUPTs and NUMTs separated by <5 kb of DNA of nonorganellar origin were considered as clusters according to Richly and Leister (2004b). Individual or clustered NUPTs and NUMTs with a size ≥5 kb, as well as nuclear sequences that yielded more than 50 individual BLAST hits for NUPTs or NUMTs (i.e., 3^{Os10}_{pm}), were selected for further analyses.

Identification of G/C content, InDels, nucleotide substitutions, and repeat regions

Alignments of organellar DNA and nuclear DNA sequences and the subsequent identification of nucleotide exchanges, deletions, duplications, and insertions were performed using the BioEdit sequence alignment editor (Hall 1999) and confirmed by sequence comparisons employing the BESTFIT algorithm of the Wisconsin Package Version 10.0, Genetics Computer Group (GCG) (Devereux et al. 1984). The BESTFIT algorithm was also used to calculate the level of DNA sequence identity between NUPTs or NUMTs and the respective organellar DNA, and their G/C contents. To calculate the local G/C content of the host chromosome, 300-kb stretches of the regions flanking the nuclear organelle DNA were considered.

Inverted repeats were identified with the StemLoop program in the GCG package. Standard settings were used for the minimum stem length, minimum and maximum loop sizes, and the minimum number of bonds per stem. G-T, A-T, and G-C pairs were scored as 1, 2, and 3 bonds, respectively.

Acknowledgments

We are grateful to Francesco Salamini, Maarten Koornneef, and Paul Hardy for critical reading of the manuscript and to the Deutsche Forschungsgemeinschaft for providing a Heisenberg fellowship (LE 1265/8) to D.L.

References

Adams, K.L. and Palmer, J.D. 2003. Evolution of mitochondrial gene content: Gene loss and transfer to the nucleus. *Mol. Phylogenet. Evol.* **29:** 380–395.

Adams, K.L., Daley, D.O., Qiu, Y.L., Whelan, J., and Palmer, J.D. 2000. Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. *Nature* **408:** 354–357.

Adams, K.L., Rosenblueth, M., Qiu, Y.L., and Palmer, J.D. 2001. Multiple losses and transfers to the nucleus of two mitochondrial succinate dehydrogenase genes during angiosperm evolution. *Genetics* **158:** 1289–1300.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Ayliffe, M.A., Scott, N.S., and Timmis, J.N. 1998. Analysis of plastid DNA-like sequences within the nuclear genomes of higher plants. *Mol. Biol. Evol.* **15:** 738–745.

Bender, J. 2004. Chromatin-based silencing mechanisms. *Curr. Opin. Plant Biol.* **7:** 521–526.

Bensasson, D., Feldman, M.W., and Petrov, D.A. 2003. Rates of DNA duplication and mitochondrial DNA insertion in the human genome. *J. Mol. Evol.* **57:** 343–354.

Blanchard, J.L. and Schmidt, G.W. 1995. Pervasive migration of organellar DNA to the nucleus in plants. *J. Mol. Evol.* **41:** 397–406.

Blouin, M.S., Yowell, C.A., Courtney, C.H., and Dame, J.B. 1998. Substitution bias, rapid saturation, and the use of mtDNA for nematode systematics. *Mol. Biol. Evol.* **15:** 1719–1727.

Boore, J.L. 1999. Animal mitochondrial genomes. *Nucleic Acids Res.* **27:** 1767–1780.

Bzymek, M. and Lovett, S.T. 2001. Instability of repetitive DNA sequences: The role of replication in multiple mechanisms. *Proc. Natl. Acad. Sci.* **98:** 8319–8325.

Cho, Y., Mower, J.P., Qiu, Y.L., and Palmer, J.D. 2004. From the Cover: Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. *Proc. Natl. Acad. Sci.* **101:** 17741–17746.

Decker-Walters, D.S., Chung, S.M., and Staub, J.E. 2004. Plastid sequence evolution: A new pattern of nucleotide substitutions in the Cucurbitaceae. *J. Mol. Evol.* **58:** 606–614.

Devereux, J., Haeberli, P., and Smithies, O. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12:** 387–395.

Esser, C., Ahmadinejad, N., Wiegand, C., Rotte, C., Sebastiani, 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.

Feng, Q., Zhang, Y., Hao, P., Wang, S., Fu, G., Huang, Y., Li, Y., Zhu, J., Liu, Y., Hu, X., et al. 2002. Sequence and analysis of rice chromosome 4. *Nature* **420:** 316–320.

Gaut, B.S. 1998. Molecular clocks and nucleotide substitution rates in higher plants. *Evol. Biol.* **30:** 93–120.

Hall, T.A. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids.*

Symp. Ser. **41:** 95–98.
Henze, K. and Martin, W. 2001. How do mitochondrial genes get into the nucleus? *Trends Genet.* **17:** 383–387.
Hiratsuka, J., Shimada, H., Whittier, R., Ishibashi, T., Sakamoto, M., Mori, M., Kondo, C., Honji, Y., Sun, C.R., Meng, B.Y. 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. and Grigg, G.W. 1993. DNA methylation and mutation. *Mutat. Res.* **285:** 61–67.
Huang, C.Y., Ayliffe, M.A., and Timmis, J.N. 2003. Direct measurement of the transfer rate of chloroplast DNA into the nucleus. *Nature* **422:** 72–76.
Klein, M., Eckert-Ossenkopp, U., Schmiedeberg, I., Brandt, P., Unseld, M., Brennicke, A., and Schuster, W. 1994. Physical mapping of the mitochondrial genome of *Arabidopsis thaliana* by cosmid and YAC clones. *Plant J.* **6:** 447–455.
Kowalczuk, M., Mackiewicz, P., Mackiewicz, D., Nowicka, A., Dudkiewicz, M., Dudek, M.R., and Cebrat, S. 2001. DNA asymmetry and the replicational mutational pressure. *J. Appl. Genet.* **42:** 553–577.
Kwok, E.Y. and Hanson, M.R. 2004. Stromules and the dynamic nature of plastid morphology. *J. Microsc.* **214:** 124–137.
Lin, X., Kaul, S., Rounsley, S., Shea, T.P., Benito, M.I., Town, C.D., Fujii, C.Y., Mason, T., Bowman, C.L., Barnstead, M., et al. 1999. Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* **402:** 761–768.
Lovett, S.T. 2004. Encoded errors: Mutations and rearrangements mediated by misalignment at repetitive DNA sequences. *Mol. Microbiol.* **52:** 1243–1253.
Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., Leister, D., Stoebe, B., Hasegawa, M., and 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.* **99:** 12246–12251.
Mourier, T., Hansen, A.J., Willerslev, E., and Arctander, P. 2001. The Human Genome Project reveals a continuous transfer of large mitochondrial fragments to the nucleus. *Mol. Biol. Evol.* **18:** 1833–1837.
Notsu, Y., Masood, S., Nishikawa, T., Kubo, N., Akiduki, G., Nakazono, M., Hirai, A., and Kadowaki, K. 2002. The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: Frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Mol. Genet. Genomics* **268:** 434–445.
Palmer, J.D. and Herbon, L.A. 1988. Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *J. Mol. Evol.* **28:** 87–97.
Paszkowski, J. and Whitham, S.A. 2001. Gene silencing and DNA methylation processes. *Curr. Opin. Plant Biol.* **4:** 123–129.
Pereira, S.L. and Baker, A.J. 2004. Low number of mitochondrial pseudogenes in the chicken (*Gallus gallus*) nuclear genome: Implications for molecular inference of population history and phylogenetics. *BMC Evol. Biol.* **4:** 17.
Petrov, D.A. and Hartl, D.L. 1999. Patterns of nucleotide substitution in *Drosophila* and mammalian genomes. *Proc. Natl. Acad. Sci.* **96:** 1475–1479.
Ricchetti, M., Fairhead, C., and Dujon, B. 1999. Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. *Nature* **402:** 96–100.
Ricchetti, M., Tekaia, F., and Dujon, B. 2004. Continued colonization of the human genome by mitochondrial DNA. *PLoS Biol.* **2:** E273.
Rice Chromosome 10 Sequencing Consortium. 2003. In-depth view of structure, activity, and evolution of rice chromosome 10. *Science* **300:** 1566–1569.
Richly, E. and Leister, D. 2004a. NUMTs in sequenced eukaryotic genomes. *Mol. Biol. Evol.* **21:** 1081–1084.
———. 2004b. NUPTs in sequenced eukaryotes and their genomic organization in relation to NUMTs. *Mol. Biol. Evol.* **21:** 1972–1980.
Sasaki, T., Matsumoto, T., Yamamoto, K., Sakata, K., Baba, T., Katayose, Y., Wu, J., Niimura, Y., Cheng, Z., Nagamura, Y., et al. 2002. The genome sequence and structure of rice chromosome 1. *Nature* **420:** 312–316.
Shahmuradov, I.A., Akbarova, Y.Y., Solovyev, V.V., and Aliyev, J.A. 2003. Abundance of plastid DNA insertions in nuclear genomes of rice and *Arabidopsis*. *Plant Mol. Biol.* **52:** 923–934.
Stegemann, S., Hartmann, S., Ruf, S., and Bock, R. 2003. High-frequency gene transfer from the chloroplast genome to the nucleus. *Proc. Natl. Acad. Sci.* **100:** 8828–8833.
Stupar, R.M., Lilly, J.W., Town, C.D., Cheng, Z., Kaul, S., Buell, C.R., and 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.* **98:** 5099–5103.
Timmis, J.N., Ayliffe, M.A., Huang, C.Y., and Martin, W. 2004. Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* **5:** 123–135.
Unseld, M., Marienfeld, J.R., Brandt, P., and Brennicke, A. 1997. The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nat. Genet.* **15:** 57–61.
Woischnik, M. and Moraes, C.T. 2002. Pattern of organization of human mitochondrial pseudogenes in the nuclear genome. *Genome Res.* **12:** 885–893.
Wolfe, K.H., Li, W.H., and Sharp, P.M. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl. Acad. Sci.* **84:** 9054–9058.
Yu, H.S. and Russell, S.D. 1994. Occurrence of mitochondria in the nuclei of tobacco sperm cells. *Plant Cell* **6:** 1477–1484.
Yuan, Q., Hill, J., Hsiao, J., Moffat, K., Ouyang, S., Cheng, Z., Jiang, J., and Buell, C.R. 2002. Genome sequencing of a 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.

Web site references

http://www.ncbi.nlm.nih.gov/; NCBI
http://mips.gsf.de/proj/thal/db/index.html; MIPS
ftp://ftp.tigr.org/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/; TIGR

Received February 2, 2005; accepted in revised form March 8, 2005.