Mitochondrial DNA of *Vitis vinifera* and the Issue of Rampant Horizontal Gene Transfer

Vadim V. Goremykin, Francesco Salamini,† Riccardo Velasco,* and Roberto Viola**

*Istituto Agrario San Michele all'Adige Research Center, San Michele all'Adige (TN), Italy; and †Agricultural Faculty, University of Milan, Milan, Italy

The mitochondrial genome of grape (*Vitis vinifera*), the largest organelle genome sequenced so far, is presented. The genome is 773,279 nt long and has the highest coding capacity among known angiosperm mitochondrial DNAs (mtDNAs). The proportion of promiscuous DNA of plastid origin in the genome is also the largest ever reported for an angiosperm mtDNA, both in absolute and relative terms. In all, 42.4% of chloroplast genome of *Vitis* has been incorporated into its mitochondrial genome. In order to test if horizontal gene transfer (HGT) has also contributed to the gene content of the grape mtDNA, we built phylogenetic trees with the coding sequences of mitochondrial genes of grape and their homologs from plant mitochondrial genomes. Many incongruent gene tree topologies were obtained. However, the extent of incongruence between these gene trees is not significantly greater than that observed among optimal trees for chloroplast genes, the common ancestry of which has never been in doubt. In both cases, we attribute this incongruence to artifacts of tree reconstruction, insufficient numbers of characters, and gene paralogy. This finding leads us to question the recent phylogenetic interpretation of Bergthorsson et al. (2003, 2004) and Richardson and Palmer (2007) that rampant HGT into the mtDNA of *Amborella* best explains phylogenetic incongruence between mitochondrial gene trees for angiosperms. The only evidence for HGT into the *Vitis* mtDNA found involves fragments of two coding sequences stemming from two closteroviruses that cause the leaf roll disease of this plant. We also report that analysis of sequences shared by both chloroplast and mitochondrial genomes provides evidence for a previously unknown gene transfer route from the mitochondrion to the chloroplast.

Introduction

Mitochondrial genomes of higher plants originated from free-living prokaryotes, which, as judged by the phylogenetic evidence, were most closely related to the α-proteobacteria (Dyall et al. 2004). The extant representatives of this group have genomes ranging from 1 to 9 MB in size and encoding from 761 to 5,666 genes (Boussau et al. 2004). During the transformation of the α-proteobacterial–like progenitor of mitochondria into the semiautonomous organelle, its DNA replication system was gradually replaced by new machinery, recruited from bacteriophage (Shutt and Gray 2006). In time, plant mitochondrial DNAs (mtDNAs) lost most of their original gene content, retaining only a few genes necessary for cell respiration and DNA duplication. The end result of this reduction process can still be observed in extant chlorophytes, the mitochondrial genomes of which are less than 60 kb in size and encode at maximum 70 genes (Nedelcu et al. 2000). During the evolution of plant lineages, as charophycean algae evolved into bryophytes and began to colonize land, their mitochondrial genomes started to increase in size. They did not acquire more genes at this stage, but their intergenic spacers underwent frequent duplications (Turmel et al. 2002). As land plants continued to evolve, their mtDNAs at some point began to acquire DNA sequences from the chloroplasts and the nucleus (Brennicke et al, 1993, Cummings et al. 2003). This process, combined with continuing duplication of noncoding DNA, resulted in large mitochondrial genomes of angiosperms, which today range in size from 221 (*Brassica napus*, Handa 2003) to 2,400 kb (*Cucumis melo*, early estimation by Ward et al. [1981]). Small repetitive sequences in noncoding regions appear to have been particularly important for this type of genome expansion in angiosperms, occurring via polymerase slippage or recombination (Lilly and Harvey 2001). However, genome size increase did not extend the coding capacity as the only functional genes recruited were transfer RNA (tRNA) genes of chloroplast origin (Marrchal et al. 1985; Joyce and Gray 1989; Izuchi et al. 1990) and the loss of the protein-coding genes continued (Adams et al. 2002). Due to frequent intra- or intermolecular homologous recombination via repeated sequences in the spacer regions in mtDNA of angiosperms, great variation in both genome organization and size can be observed even within one species (Fauron et al. 1995). This constant recombination has lead to the appearance of incomplete subgenomic and/or isomeric mtDNA forms in some mitochondrial populations. Of all heterogeneous forms of mtDNA present in the organelle, only one, the so-called "master circle," which contains all genes, is thought to be replicated (for discussion, see Ogihara et al. 2005). The large genome size of mitochondrial genomes of angiosperms is contrasted by their sparse gene content. They contain 50–60 household genes encoding respiratory complexes I, II, III, IV, and V, ribosomal subunits, ribosomal and transport RNAs, a few genes of cytochrome c biogenesis, and one maturase-like intron open reading frame (ORF) (Kubo and Mikami 2007). Most published angiosperm mitochondrial genomes have completely lost their polymerase genes, and none of these mtDNA molecules contains the full set of tRNA genes necessary to recognize all codons. All missing components of the gene expression machinery must be imported from the nucleus. The most peculiar feature of mtDNA of angiosperms is uptake of foreign sequences, termed promiscuous DNA (Timmis et al. 2004). Due to absence of such sequences in the mitochondrial genome of *Marchantia*, it was initially thought that they should play a significant role in angiosperm genome size expansion. Subsequent publication of complete mitochondrial genomes has revealed that the contribution of promiscuous DNA is smaller than expected. The proportion of incorporated sequences from the nucleus in published

Key words: grape mtDNA, horizontal gene transfer, molecular evolution.

E-mail: vadim.goremykin@iasma.it.

Mol. Biol. Evol. 26(1):99–110. 2009
doi:10.1093/molbev/msn226
Advance Access publication October 14, 2008

angiosperm mtDNAs ranges from 0.1% to 13.4% of the genome size, and the proportion of sequences from chloroplast DNA constitutes 1.1–6.3% of the genome size. Of all the reported uptakes of promiscuous DNA by mitochondrial genome, perhaps, the most spectacular is the one suggested by Bergthorsson et al. (2003, 2004) and Richardson and Palmer (2007). They argue that unknown factors, perhaps even meteorite bombardment (Bergthorsson et al. 2003, p. 200), have caused the invasion from a myriad of unidentified plant donors of as many as 26 (Bergthorsson et al. 2004) or even hundreds (Richardson and Palmer 2007) of foreign mitochondrial genes into the mtDNA of the angiosperm *Amborella trichopoda*. As the mitochondrial genome of *Vitis* is the largest among those sequenced, we initially thought that its large size might also be due to horizontal gene transfer (HGT) as suggested for other cases by Bergthorsson et al. (2003, 2004). The issue of rampant HGT is, in our opinion, important and has much broader practical implications, far exceeding evolution of *Amborella*. If genes travel from species to species on the scope suggested by Bergthorsson et al. (2003, 2004) and Richardson and Palmer (2007), then we should reevaluate the potential danger of genetically modified crops to other natural populations of plants. This issue has motivated us to reexamine the extent of events of HGT into the genome of *Vitis*.

Materials and Methods

Genomic Sequencing

The mtDNA sequence was produced as a part of the whole genomic sequencing (WGS) project of *Vitis vinifera* (for detailed descriptions of methodology, see Velasco et al. 2007). In brief, shotgun sequencing libraries with average insert sizes of 2, 3, 6, 10, and 12 kb were prepared from the total DNA extracted by a cetyltrimethylammonium bromide method (Doyle JJ and Doyle JL 1987) from young shoots of Pinot Noir, clone ENTAV115. Fosmid and bacterial artificial chromosome libraries were constructed using the same grapevine DNA and end sequenced to build up contig scaffolds. The genomic assembly was carried out as described by Zharkikh et al. (2008). The coverage of the mitochondrial sequence presented in this paper reached 60× (Sanger reads). The variant of the genomic assembly supported by the overwhelming majority of reads is presented.

Sequence Annotation

A preliminary annotation was carried out by mapping BLAST hits employing known mitochondrial genes as queries and, subsequently, by testing for consistency of the ORFs. The tRNA genes were also searched for with the help of tRNAscan-SE program (Lowe and Eddy 1997). In the final step, all ORFs longer than 60 bases, not overlapping with previously annotated regions, were BLASTed against the nr, wgs, and genomic GenBank databases. Several full-length hits found in this search were also mapped onto the mtDNA sequence.

Identification of chloroplast insertions in mtDNA.

After annotation of mitochondrial genes, they were masked in the mtDNA of grape. This was done in order to eliminate the BLAST hits from DNA stretches belonging to the most conserved household genes, which are similar in chloroplast and mitochondrial genomes. The masked genome sequence was BLASTed against a local database containing all published chloroplast genomic sequences and the grape nuclear genomic assembly. The series of adjacent BLAST hits, syntenic in both subject and query sequences and not interrupted by any masked material, were identified, and the sequences of these syntenic regions were mapped onto the mtDNA of *Vitis*.

Phylogenetic Analyses

All sequences of 39 mitochondrial genes from different land plants available from GenBank database and, where possible, of algal outgroups were downloaded and aligned with ClustalW, producing alignments containing from 13 to 553 species. Chloroplast sequences of 61 genes common to the land plant species, used in the control series of experiments, were extracted from 32 published chloroplast genomes and aligned with the help of the same program. BLAST searches of mitochondrial genes against the published nuclear genomes of angiosperms as well as GenBank annotation of single gene entries were used to identify gene copies transferred from mtDNA into the nucleus. These were aligned with mitochondrial genes in order to determine the phylogenetic relationship of such transferred genes to their master copies from mtDNA. As many alignments contained a large number of sequences, phylogenetic analyses by maximum likelihood (ML) method were performed employing the fast PHYML algorithm (Guindon and Gascuel 2003). The ML models utilized during the heuristic search for the best tree were predetermined with the help of Modeltest program (Posada and Crandall 1998). We used the best models chosen employing Akaike information criterion in all our experiments. This should be advantageous to the HGT test procedure applied by Bergthorsson et al. (2004) who used an arbitrarily chosen model (Hasegawa–Kishino–Yano [HKY] + G) in their tree building analyses.

Results

The Mitochondrial Genome Complement of *Vitis*

The predominant mtDNA assembly generated in the WGS project was a 773,279-long base pair circular molecule (fig. 1). This should represent the so-called master circle (Atlan and Couvet 1993) as it contained all genes sequenced. Intergenic spacers constitute the largest part (90.21%, 697591 bp.) of this molecule (where promiscuous DNA is considered as part of the spacer sequences). The protein-coding sequences comprise only 4.98% of the molecule length (38,529 bp). RNA genes constitute 0.91% of the mtDNA of *Vitis* and introns 3.89% (30,100 bp). Gene content in the mitochondrial genome is similar to that of other published angiosperm mtDNAs (e.g., Sugiyama et al. 2005). The large size of the genome being due to the expansion of the spacer regions. These regions contain 1,338 repeated sequences ranging in size from 30 to 651 bases (reaching 52,861 bp in total, which corresponds to 6.84% of the genome length), of which 645 are direct repeats (25,325 bp in total length, 3.28% of the genome length). Most of the genome sequences have no

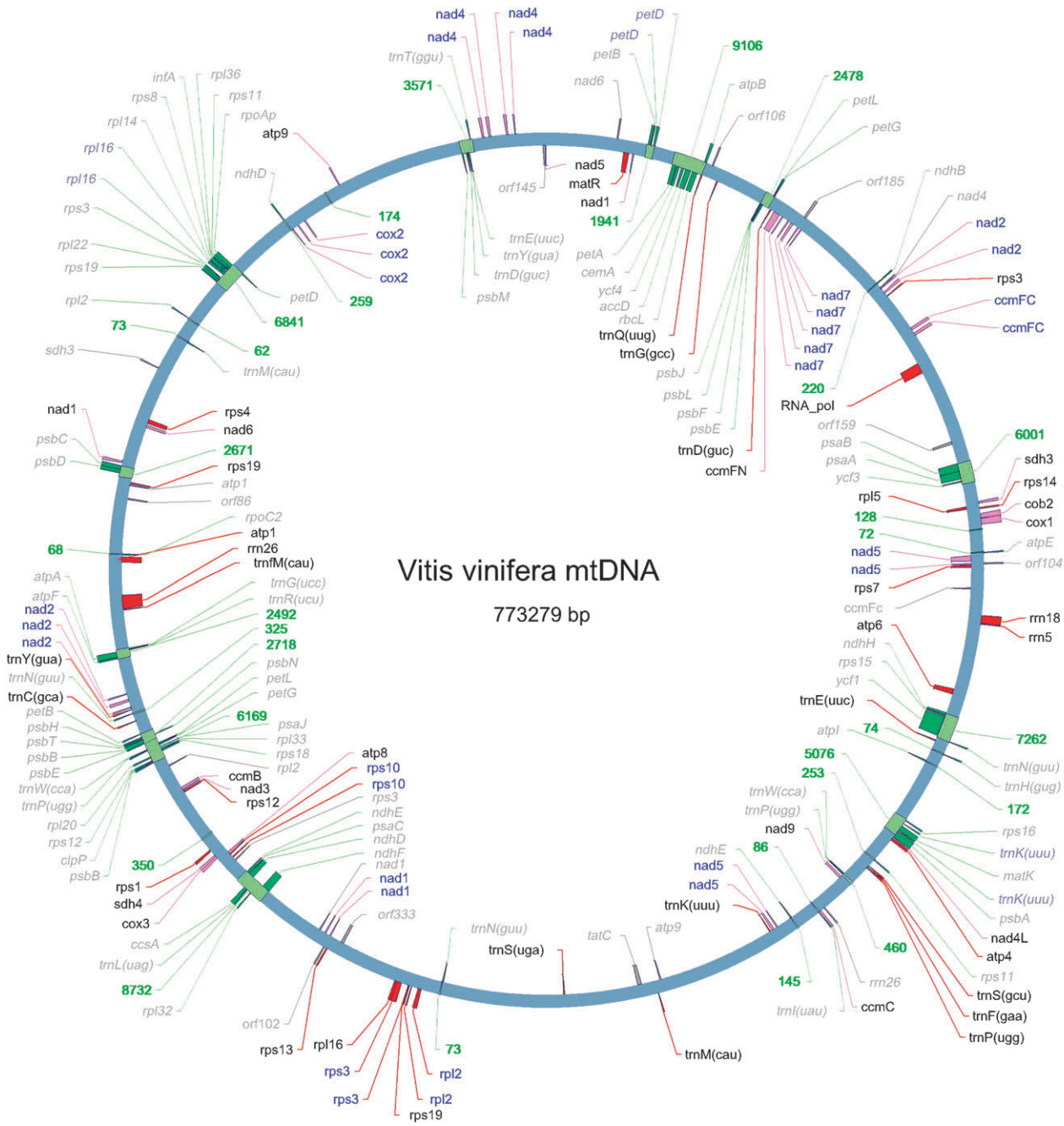

FIG. 1.—Mitochondrial genome of *Vitis vinifera*. The topmost part of the map corresponds to the start and the end of the EMBL sequence entry FM179380. The relative positions and lengths of chloroplast DNA insertions (also listed in table 4) have been indicated in green. The names of the functional copies of the mitochondrial genes are indicated in regular script, and all other gene names have been italicized. The genes shown inside the circle are transcribed clockwise, those outside the circle are transcribed counterclockwise. The genes of genetic apparatus are shown in red, the genes supporting cell respiration are shown in pink, and genes of chloroplast origin are shown in green. ORFs, *ycf*s, and genes of unknown function are highlighted in gray. Intron-containing genes are represented by their exons. Their names are given in blue.

similarity to the sequences of other mitochondrial genomes of angiosperms. For example, the proportion of sequences in mtDNA of *Vitis* exhibiting 70 or more percent similarity to the mitochondrial genomes of *Nicotiana*, *Arabidopsis*, and *Zea* constitutes, respectively, 19.02%, 12.97%, and 11.62% of the total genome size. In contrast to this finding, we observed that both rps3-rpl16 and nad3-rps12 gene clusters were conserved as in the mitochondrial genomes of other angiosperms (Nakazono et al. 1995).

Using BLAST searches, we evaluated the gene content of the mitochondrial genome. Table 1 gives an overview of these results. Mitochondrial genome of *Vitis* contains at least 37 protein-coding genes (38 genes). Most of the genes (24) encode proteins of the electron transport chain, 9 subunits of nicotinamide adenine dinucleotide dehydrogenase (complex I), 2 subunits of succinate dehydrogenase (complex II), apocytochrome b (complex III), 3 subunits of cytochrome c oxidase (complex IV), 5 subunits of ATP synthase F1 (complex

Table 1
Gene Content of the mtDNA of *Vitis vinifera*

Genes of Mitochondrial Origin	
Complex I	*nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, *nad9*
Complex II	*sdh3*, *sdh4*
Complex III	*cob2*
Complex IV	*cox1*, *cox2*, *cox3*
Complex V	*atp1*, *atp4*, *atp6*, *atp8*, *atp9*
Cytochrome c biogenesis	*ccmB*, *ccmC*, *ccmFC*, *ccmFN*
Ribosome large subunit	*rpl2*, *rpl5*, *rpl16*
Ribosome small subunit	*rps1*, *rps3*, *rps4*, *rps7*, *rps10*, *rps12*, *rps13*, *rps14*, *rps19* (x2)
Intron maturase	*matR*
rRNA genes	*rrn26*, *rrn18*, *rrn5*
tRNA genes	*trnC*(*gca*), *trnD*(*guc*), *trnE*(*uuc*), *trnF*(*gaa*), *trnfM*(*cau*), *trnG*(*gcc*), *trnK*(*uuu*), *trnM*(*cau*), *trnP*(*ugg*), *trnQ*(*uug*), *trnS*(*gcu*), *trnS*(*uga*), *trnY*(*gua*)
Pseudogenes	*atp1*, *atp9*, *ccmFc*, *nad1*, *nad4*, *nad6*, *rpl2*, *rps3* (*x2*), *rrn26*, *sdh3*, *tatC*
Hypothetical genes	8 ORFs
rRNA genes	*rrn26*, *rrn18*, *rrn5*
	trnC(*gca*), *trnD*(*guc*), *trnE*(*uuc*), *trnF*(*gaa*), *trnfM*(*cau*), *trnG*(*gcc*), *trnK*(*uuu*), *trnM*(*cau*), *trnP*(*ugg*), *trnQ*(*uug*), *trnS*(*gcu*), *trnS*(*uga*),
tRNA genes	*trnY*(*gua*)
Genes of Chloroplast Origin	
Genes with intact ORFs	*atp*A, *inf*A, *ndh*E, *pet*A, *pet*G (x2), *pet*L (x2), *psa*C, *psa*J, *psb*A, *psb*J, *psb*M, *psb*N, *psb*T, *rbc*L, *rpl*14, *rpl*16, *rpl*20, *rpl*32, *rpl*33, *rpl*36, *rps*11, *rps*15, *rps*18, *rps*19, *ycf*4
Pseudogenes	*acc*D, *atp*B, *atp*E, *atp*F, *atp*I *ccs*A, *cem*A, *clp*P, *mat*K, *ndh*B, *ndh*D (x2), *ndh*E, *ndh*F, *ndh*H, *pet*B (x2), *pet*D (x2), *psa*A, *psa*B, *psb*B (x2), *psb*C, *psb*D *psb*E (x2), *psb*F, *psb*H, *psb*L, *rpl*2, *rpl*22, *rpo*A, *rpo*C2, *rps*11, *rps*12, *rps*16, *rps*3, *rps*8, *ycf*1, *ycf*3
tRNA genes	*trn*D(guc), *trn*E(uuc), *trn*H(gug), *trn*I(uau), *trn*K(uuu), *trn*L(uag), *trn*M(cau), *trn*N(guu) (x3) *trn*P(ugg) (x2), *trn*R(ucu), *trn*T(ggu), *trn*W(cca) (x2), *trn*Y(gua)
tRNA pseudogenes	*trn*E(uuc), *trn*G(ucc) exon2
Genes of Nuclear Origin	
	Phage-like one-chain Rna polymerase; pfam00940

V) and 4 cytoplasmic membrane proteins required for cytochrome c maturation. Also present is a conserved intron, which contains an ORF for the maturase gene *matR*. The remaining 13 genes (12 gene species) encode ribosomal proteins. *rps*19 has two apparently functional full gene copies. The protein coding capacity of the genome is the highest among completely sequenced mtDNAs of the angiosperms (see Table 2 for an overview of the protein-coding gene content of published mitochondrial genomes).

Some mitochondrial genes have partial pseudogene copies in the mtDNA of *Vitis*. *rps3* has two partial pseudogene copies, whereas *atp1*, *atp9*, *ccmFc*, *nad1*, *nad4*, *nad6*, *rpl2*, and *sdh3* have one copy. All pseudogenes, except *nad6*, *ccmFc*, and one of *rps3* pseudogenes, are identical or near identical to their corresponding functional orthologues in the mitochondrial genome of *Vitis*, suggesting that the primary source of these divergent gene copies is not from HGT but rather from the *Vitis* mtDNA itself.

Translational Dependence of mtDNA of *Vitis*.

BLAST searches against a local chloroplast and mitochondrial tRNA database and tRNA-SE scans identified 32 tRNA sequences in the grape mitochondrial genome (table 1). Of these, at least two are not functional. One is a pseudogene of a chloroplast-like *trn*E(uuc) that has lost its capacity to fold correctly, and the second is an exon of the chloroplast-like *trn*G(ucc) gene. Based on the assumption that U in the third codon/first anticodon wobble position can recognize all bases in the mtDNA, there are 30 tRNAs with intact folding capacity encoded by the mtDNA of *Vitis* that can recognize 43 of the 61 sense codons. The codon–anticodon recognition pattern of the translational machinery of the grape mitochondrial genome is summarized in table 3. The genome lacks at least three functional tRNAs for amino acids Ala, Arg, and Val and lacks at least three tRNAs necessary for recognition of all codons for the amino acids Leu, Gly, and Thr. The missing tRNAs must therefore be imported in the mitochondria of *Vitis* from the cytosol. Most of tRNAs (17 out of 30) found in the mtDNA of *Vitis* are of chloroplast origin (their names are shown in regular script in table 3). Ten of these are single-copy genes, two are duplicated, and one gene occurs in triplicate. With few exceptions, these chloroplast-type mitochondrial tRNAs are included as part of long insertions of chloroplast DNA, wherein the gene order is characteristic of the cpDNA of *Vitis* (fig. 1, table 4). Because expressed tRNAs of chloroplast origin have also previously been reported in the mitochondria of angiosperms (Marrchal et al. 1985; Joyce and Gray 1989; Izuchi et al. 1990), our finding suggests that chloroplast tRNA genes have also been recruited by the grape mitochondrion. Translation of these chloroplast tRNAs might be carried out by the nuclear encoded rpoTm polymerase, imported by the mitochondrion, which utilizing the same promoters as the nuclear encoded polymerase of the chloroplast (Allison et al. 1996; Kühn et al. 2007). The corresponding nuclear gene has been identified in the grape genome (GenBank accession numbers CAO49235 and CAO15344).

RNA Polymerase

The mitochondrial genome of grape contains a 2,928-bp long ORF potentially encoding a 975 amino acids product harboring a conserved domain characteristic of the pfam00949 superfamily of polymerases (Joyce and Steitz 1994). The superfamily includes single-subunit polymerases, presumably of bacteriophage origin, including the enzyme rpoTm which catalyzes mtDNA translation in higher plants (Hedtke et al. 1997; Weihe et al. 1997). However, the product of translation of the ORF has little similarity to the rpoTm enzymes encoded by the nuclear genome of *Vitis* (GenBank accession numbers CAO49235 and CAO15344). The polypeptide resembles the polymerase encoded by the ORF6 of the linear mitochondrial plasmid of *B. napus* (Handa et al. 2002), the RNA polymerase present in the mtDNA of *Daucus carota*

Table 2
Comparison of the Coding Capability of Mitochondrial Genomes

	Vitis vinifera	*Beta vulgaris*	*Nicotiana tabacum*	*Arabidopsis thaliana*	*Brassica napus*	*Oryza sativa*	*Zea mays*	*Triticum aestivum*	*Sorghum bicolor*	*Tripsacum dactyloides*	*Marchantia polymorpha*	*Physcomitrella patens*
Complex I												
nad1	+	+	+	+	+	+	+	−	+	+	+	+
nad2	+	+	+	+	+	+	+	+	+	+	+	+
nad3	+	+	+	+	+	+	+	+	+	+	+	+
nad4	+	+	+	+	+	+	+	+	+	+	+	+
nad4L	+	+	+	+	+	+	+	−	+	+	+	+
nad5	+	+	+	+	+	+	+	+	+	+	+	+
nad6	+	+	+	+	+	+	+	+	+	+	+	+
nad7	+	+	+	+	+	+	+	+	+	+	−	+
nad9	+	+	+	+	+	+	+	+	+	+	+	+
Complex II												
sdh3	+	−	+	−	−	−	−	−	−	−	+	+
sdh4	+	−	+	−	−	−	−	−	−	−	+	+
Complex III												
cob	+	+	+	+	+	+	+	+	+	+	+	+
Complex IV												
cox1	+	+	+	+	+	+	+	+	+	+	+	+
cox2	+	+	+	+	+	+	+	+	+	+	+	+
cox3	+	+	+	+	+	+	+	+	+	+	+	+
Complex V												
atp1	+	+	+	+	+	+	+	+	+	+	+	+
atp4	+	+	+	+	+	+	+	+	+	+	+	+
atp6	+	+	+	+	+	+	+	+	+	+	+	+
atp8	+	+	+	+	+	+	+	+	+	+	+	+
atp9	+	+	+	+	+	+	+	+	+	+	+	+
Cytochrome c biogenesis												
ccmB	+	+	+	+	+	+	+	+	+	+	+	+
ccmC	+	−	+	+	+	+	+	+	+	+	+	+
ccmFC	+	+	+	+	+	+	+	+	+	+	+	+
ccmFN	+	+	+	+	+	+	+	+	+	+	+	+
Ribosomal												
rpl2	+	−	+	+	+	+	−	−	−	−	+	+
rpl5	+	+	+	+	+	+	−	+	−	−	+	+
rpl6	−	−	−	−	−	−	−	−	−	−	+	+
rpl16	+	−	+	+	+	+	+	+	+	+	+	+
rps1	+	−	+	−	−	+	+	+	+	+	+	+
rps2	−	−	−	−	−	+	+	+	+	+	+	+
rps3	+	+	+	+	+	+	+	+	+	+	+	+
rps4	+	+	+	+	+	+	+	+	+	+	+	+
rps7	+	+	−	+	+	+	+	+	+	+	+	+
rps8	−	−	−	−	−	−	−	−	−	−	+	−
rps10	+	−	+	−	−	−	−	−	−	−	+	−
rps11	−	−	−	−	−	−	−	−	−	−	+	+
rps12	+	+	+	+	+	+	+	+	+	+	+	+
rps13	+	+	+	−	−	+	+	+	+	+	+	+
rps14	+	−	−	−	+	−	−	−	−	−	+	+
rps19 (x2)	+	−	+	−	−	+	−	−	−	−	+	+
Other ORFs												
matR	+	+	+	+	+	+	+	+	+	+	−	−
tatC	−	+	+	+	+	+	+	+	+	+	+	+
Total number	37	29	35	31	31	35	32	31	32	32	40	39

(Robison and Wolyn 2005), and a number of variously annotated single-subunit RNA polymerases from mtDNA of *Cucumis*, *Triticum*, and other higher plants.

Promiscuous DNA

Chloroplast Insertions in Grape mtDNA

The mtDNA of grape contains 30 fragments of chloroplast DNA ranging in size from 62 to 9,106 nt (table 4). The total extent of chloroplast DNA sequences present in the mtDNA of *V. vinifera* is 68,237 bp, corresponding to 8.8% of the whole mitochondrial genome length and to 42.4% of the grape chloroplast genome. This is the largest proportion of cpDNA sequences accumulated by a plant mitochondrial genome, both in absolute and relative terms. Most of the insertions are unique to the grape mtDNA, as evident from the observation that only 9 out of 30 insertions have full-length homologs in the

Table 3
Recognition of Anticodons by tRNA Genes Found in the mtDNA of *Vitia vinifera*

UUU Phe	***trnF*(gaa)**	UCU Ser	***trnS*(uga)**	UAU Tyr	***trnY*(gua)**	UGU Cys	***trnC*(gca)**
UUC Phe	*	UCC Ser	*	UAC Tyr	*trn*Y(gua)	UGC Cys	*
UUA Leu	No	UCA Ser	*	UAA —		UGA —	
UUG Leu	No	UCG Ser	*	UAG —		UGG Trp	*trn*W(cca)x2
CUU Leu	*trn*L(uag)	CCU Pro	***trnP*(ugg)**, *trn*P(ugg)x2	CAU His	*trn*H(gug)	CGU Arg	no
CUC Leu	*	CCC Pro	**	CAU His	*	CGU Arg	no
CUA Leu	*	CCA Pro	**	CAA Gln	***trnQ*(uug)**	CGA Arg	no
CUG Leu	*	CCG Pro	**	CAG Gln	*	CGG Arg	no
AUU Ile	*trn*I(uau)	ACU Thr	*trn*T(ggu)	AAU Asn	*trn*N(guu)x3	AGU Ser	***trnS*(gcu)**
AUC Ile	*	ACC Thr	*	AAC Asn	*	AGC Ser	*
AUA Ile	*	ACA Thr	No	AAA Lys	***trnK*(uuu)**, *trn*K(uuu)	AGA Arg	*trn*R(ucu)
AUG Met	***trnM*(cau)**, *trn*M(cau), ***trnf*M(cau)**	ACG Thr	No	AAG Lys	**	AGG Arg	*
GUU Val	No	GCU Ala	No	GAU Asp	***trnD*(guc)**, *trn*D(guc)	GGU Gly	***trnG*(gcc)**
GUC Val	No	GCC Ala	No	GAC Asp	**	GGC Gly	*
GUA Val	No	GCA Ala	No	GAA Glu	***trnE*(uuc)**, *trn*E(uuc)	GGA Gly	No
GUG Val	No	GCG Ala	No	GAG Glu	**	GGG Gly	No

NOTE.—The names of the tRNA genes of mitochondrial origin have been highlighted in bold. The names of the tRNA genes of chloroplast origin are shown in regular font. Asterisks in columns below gene names indicate that the tRNA species shown above recognize corresponding codons too.

mtDNA of other plant species. Two are found in only one species: a 1,941-bp insertion (in *Sorghum*) and a 2,478-bp insertion (in *Nicotiana*). Grape mtDNA shares only two small insertions (of 86 and 73 bp) of chloroplast origin that are common with the rosids (as represented by *Arabidopsis* and *Brassica*), to which *Vitis* is a sister (Jansen et al. 2006). Thus, most of the events of DNA transfer from chloroplast genomes to the mitochondrial genome of grape have occurred after the separation of the phylogenetic lineage leading to *Vitis*.

Table 4
Chloroplast Insertions in the mtDNA of *Vitis vinifera*

Length	Position	Genes Contained
9,106	36915–46020	***pet*A–***cemA*–***ycf*4–***acc*D–***rbc*L–***atp*B**
8,732	474436–483167	***ndh*F–*rpl*32–*trn*L(uag)–***ccs*A–*ndh*D–*psa*C–*ndh*E**
7,262	235673–242934	***ndh*H–*rps*15–***ycf*1–*trn*N(guu)**
6,841	668317–675157	***rps*19–***rpl*22–***rps*3–*rpl*16–*rpl*14–***rps*8–*inf*A–*rpl*36–*rps*11–***rpo*A–*pet*D**
6,169	523895–530052	***psb*B–*clp*P–*rps*12–*rpl*20–*rps*18–*rpl*33–*psa*J–*trnP*(ugg)–*trnW*(cca)–*petG*–*petL*–*psb*E**
6,001	162304–168304	***psa*B–*psa*A–*ycf*3**
5,076	268066–273141	***rps*16–*trn*K(uuu)–*mat*K–*psb*A**
3,571	748716–752286	***psb*M–*trn*D(guc)–*trn*Y(gua)–*trn*E(uuc)–*trn*T(ggu)**
2,718	530053–532770	***psb*B–*psb*T–*psb*N–*psb*H–*pet*B**
2,867	606700–609566	***psb*D–*psb*C**
2,492	554939–557430	***atp*F–*atp*A–*trn*R(ucu)–*trn*G(ucc)**
2,478	65057–67534	***psb*J–*psb*L–*psb*F–*psb*E–*pet*L–*pet*G**
1,941	28850–30790	***pet*B–*pet*D**
460	291324–291783	***trn*W(cca)–*trn*P(ugg)**
350	498165–498514	
325	538650–538974	***trn*N(guu)**
259	693978–694236	***ndh*D**
253	282259–282511	***rps*11**
220	105310–105529	***ndh*B**
174	708001–708174	
172	250617–250788	***atp*I**
145	311184–311328	***ndh*E**
128	181810–181937	
86	303835–303920	***trn*I(uau)**
74	246745–246818	***trn*H(gug)**
73	649719–649791	***trn*M(cau)**
73	417117–417189	***trn*N(guu)**
72	188354–188425	***atp*E**
68	584577–584644	***rpo*C2**
62	655445–655506	***rpl*2**

NOTE.—Pseudogenes with ORF interrupted by stop codons and tRNA pseudogenesare shown in regular font. Genes with intact ORFs and genes which encode tRNAs with intact folding ability are highlighted in bold.

mtDNA Insertions in cpDNA?

During the search for chloroplast insertions in the mitochondrion genome, we identified a large number of mitochondrial sequences highly similar to a region of cpDNA of *Daucus carota* (Ruhlman et al. 2006). In BLAST analyses, this sequence did not show homology to the chloroplast genomes of other published cpDNA sequences. However, a large number of significant BLAST hits for this region were obtained with other published mitochondrial genomes (95% sequence identity with mtDNA of *Brassica*, *Oenothera*, and *Arabidopsis*; slightly lower values were registered for other mitochondrial genomes). The sequence (positions 99437–99562 and 140717–140592 in the cpDNA of *Daucus*) is a part of the intergenic spacer separating *rps*12 and *trn*V(gac). This sequence is contained in a large 1,439-bp fragment of the cpDNA of *Daucus* (positions 99309–100747 and 139407–140845) that is absent from the chloroplast genomes of other species. BLAST analysis of this longer sequence against the nonredundant GenBank database revealed that the 1,439-bp stretch contains another 74-bp short conserved region (positions 99364–99436 and 140718–140790) with 95–94% identity to the coding sequence of the mitochondrial cytochrome c oxidase subunit 1 (*cox1*).

Assuming correct assembly of the *Daucus* chloroplast genome, this finding suggests a transfer of DNA from mitochondrion into the chloroplast. Whereas the transfer of the chloroplast DNA into the mitochondrial genomes is well-known phenomenon (Stern and Lonsdale 1982), a reverse transfer route has not yet been reported (Lonsdale 1989; Cummings et al. 2003). To our knowledge, this is the first potential case of such transfer.

Viruses

Grape mtDNA contains two DNA fragments with homology to sequences from leaf roll–associated grape closteroviruses 1 and 8. These sequences are a partial copy of the coding sequence of the coat protein of the leaf roll–associated virus 8 (positions 471430–471697 in grape mtDNA), which has a 98% identity to the donor sequence (GenBank accession number AF233936), and a partial copy of the coding sequence of the HSP70 gene of the leaf roll–associated virus 1 (positions 469419–469475), identical to the donor sequence (GenBank accession number AF233935). The grape mitochondrial genome also contains 7 ORFs, ranging in size from 81 to 466 amino acids, which are similar to the RNA polymerases of the genus *Mitovirus*, family Narnaviridae. This genus includes the simplest of all known viruses. These are parasites of fungal mitochondria that do not form true viral particles (Cole et al. 2000). Their genome consists of one gene, the RNA-dependent RNA polymerase necessary to replicate the viral genome. ORFs similar to the mitoviral RNA polymerase were described for the mitochondrial genome of *Arabidopsis thaliana* (Marienfeld et al. 1997; Hong et al. 1998), *Vicia faba* (Marienfeld et al. 1997), and *Brassica napus* (Tuomivirta and Hantula 2005). These findings were interpreted as cases of horizontal DNA transfer of the viral sequences between fungi and plants (Marienfeld et al. 1997; Hong et al. 1998). These authors did not determine though whether such transfer happened once or there were several independent cases of HGT from fungi to plants. Our phylogenetic analysis of the amino acid alignment of conserved domains of known mitoviruses to similar ORFs present in plant mitochondrial genomes supported a monophyletic origin of these plant ORFs but did not identify a specific donor mitovirus (results not shown).

Investigation of HGT

To trace possible events of HGT into the mitochondria, we extracted the coding sequences of genes and sequences of pseudogenes from the completely sequenced mtDNAs and sorted them into FASTA files. In each was included homologues from the nonredundant GenBank database. *rps8* and *rps11* genes were excluded from these analyses because only a few sequences were available for these genes. In total, 39 individual gene alignments were created from these files. ML trees were then reconstructed for these data sets using optimal substitution models. Many of the trees reconstructed from mt sequences had topologies incongruent with that expected for the phylogenetic history of some plant species (supplementary fig. 1s, Supplementary Material online, contains the trees and corresponding alignments). For instance, 30 out of 39 trees did not support the monophyly of eudicots, and 12 trees did not support monophyly of the monocots. In *nad2* tree, grasses were sister to rosids, whereas in *ccmB* tree, an asterid *Archytaea* was embedded into the monocot cluster. In *nad3* tree, a well-supported branch (87% bootstrap support) divided two bryophytes, *Amborella*, *Carthamus* (an asterid), *Magnolia*, and two monocot species (*Allium* and *Pennisetum*) from the large crown group, comprising the bulk of monocots (9 species), eudicots (19 species), magnoliids, and basal angiosperms. Not only the order of branching among the angiosperms was highly variable. Some species were found among totally unrelated plant phyla. For example, in the *atp9* tree, the pine sequence was embedded within angiosperms. In the *cox2* tree, moss *Physcomitrella* appeared as a sister to tracheophyte *Isoetes*. In the *rpl2* tree, maize grouped outside angiosperms and gymnosperms, in a well-supported branch (97% bootstrap proportion support) with bryophytes *Marchantia* and *Physcomitrella*. Most unexpectedly, in the *cob2* tree, *Picea*, a conifer, clustered with *Pedinomonas*, a green alga, with 100% BP support.

A number of observations may be relevant for explanation of these results. We observed no correlation between the taxa included in the above analyses and support for 1) monophyly of monocots, and 2) monophyly of eudicots (table 5). Thus, we conclude that the above abnormalities in the trees

Table 5
Dependence of Results of Phylogeny Reconstruction on Taxon Sampling of Mitochondrial Genes

Gene	OTUs	Eudicots	Monocots
matR	553	–	+
cob2	216	–	–
atp1	197	+	+
nad5	192	–	–
rps13	154	–	–
cox1	150	–	–
ccmB	137	–	–
cox3	137	–	+
nad6	134	–	+
rps3	100	+	+
rps19	95	–	–
rps2	69	–	+
rpl5	55	+	+
rps10	51	–	+[a]
cox2	43	–	–
nad3	40	–	–
nad7	39	–	+
atp6	38	–	+
nad9	38	–	+
atp9	37	–	–
atp8	36	+	+
rpl16	36	–	+
rps14	34	+[a]	–
rps1	33	–	–
rps4	33	–	+
sdh4	31	–	0[b]
rps12	30	–	+
rps7	30	+	+
nad4	28	+	+
ccmFN	27	–	+
atp4	26	–	+
ccmC	26	–	+
rpl2	20	+	–
rpl6	16	–	+
nad2	15	–	+
nad4L	15	+	+
nad1	14	–	+
tatC	14	–	+
ccmFC	13	–	+

NOTE.—First column, gene names; second column, number of species (OTUs) in alignment; third column, does the tree have a monophyletic eudicot cluster? (+, yes; –, no); and fourth column, does the tree have a monophyletic monocot cluster? (+, yes; –, no).

[a] With exclusion of nuclear copies.

[b] Only one monocot species in the tree.

cannot be explained by insufficient taxon sampling. However, the affinity of maize with bryophytes in *rpl2* tree is clearly due to long branch attraction because deletion of bryophytes leads to relocation of the branch subtending *Zea* to the branch *Oryza* + *Eichhornia*. The *Pedinomonas–Picea* relationship is also likely to be an LBA artifact because deletion of *Pedinomonas* leads to relocation of *Picea* across several tree nodes to another very long branch subtending *Pseudendoclonium*. To check if HGT may explain the tree abnormalities, we built ML trees from the coding sequences of the 61 genes common to the chloroplast genomes of the land plants. We took chloroplast genes of land plants as a standard for comparison because their evolution is generally considered unaffected by HGT. We assumed that if the tree building results obtained with the mitochondrial sequences were indeed due to HGT, then we expected that unusual phylogenetic relationships should be significantly more pronounced in trees built from mitochondrial sequences than in trees built from chloroplast sequences. However, many unexpected optimal tree topologies were also observed for chloroplast genes. For instance, 20 optimal trees did not support monophyly of eudicots and 38 trees did not support monophyly of monocots. In the optimal petN tree, *Pinus* (a gymnosperm) was adjacent to *Huperzia* (a fern ally), and in *psb*K and *rpl*36 gene trees, *Pinus* joined with *Psilotum* (a fern) (for all tree building results, see supplementary data 2s, Supplementary Material online). The position of the basal angiosperm *Amborella*, whose varying phylogenetic relationships in individual mtDNA gene trees was interpreted as evidence of HGT by Bergthorsson et al. (2003, 2004), was of particular interest to us. *Amborella* assumed a sister group position to *Phalaenopsis* in the *lhb*A tree, sister to the grasses in the *rps*15 tree, sister to the *Calycanthus* in the *rps*18 tree, sister to the lineage containing ((*Liriodendron*, *Calycanthus*), (*Acorus*, *Nymphaea*)) on a *ycf*3 tree, sister to the lineage ((*Ceratophyllum*), (*Acorus*, *Nymphaea*)) on *psb*D tree, a sister to the branch subtending grasses plus *Nymphaea* in *psb*K tree, and sister to the branch subtending *Phalaenopsis* plus *Nymphaea* on the *psb*J tree. It was embedded within a cluster encompassing *Nymphaea*, *Acorus*, and *Ceratophyllum* on the *rpl*16 tree, contained within a cluster encompassing *Nymphaea*, *Liriodendron*, *Calycanthus*, *Ceratophyllum* plus two monocots *Acorus* and *Phalaenopsis* in the *clp*P tree (this branch received 98% BP support); it joined another branch (81% BP) bearing *Nymphaea*, *Liriodendron*, *Calycanthus* plus *Acorus* in the *atp*A tree, and so on.

As the trees in Bergthorsson et al. (2003, 2004) were based on the sequences of polymerase chain reaction (PCR) products generated from the total DNA, we also wished to see if known paralogous gene copies from nucleus and mitochondrion could be easily recognized as such in reconstructed phylogenetic trees. For these experiments, complete gene sequences were used for one-exon genes, and, in several cases, single exons were used to represented genes with nonuniform exon composition among the land plants. Our partial gene sequences were at least as long as those used by Bergthorsson et al. (2003, 2004). We observed 14 cases in which paralogous sequences from nuclear and mitochondrial genomes did not form monophyletic clusters (supplementary data 3s, Supplementary Material online). In four trees (*atp8*, *cox2*, *nad1*, and *rps14*), separation of the nuclear and mitochondrial gene copies was strongly supported (80% and above BP). It is noteworthy that when analyzed under the optimal TIM + G model, the two copies of *sdh4* gene from *Amborella* remained adjacent (supported by 89% BP), unlike the finding of Bergthorsson et al. (2004), who used the suboptimal HKY + G model.

Discussion

The Mitochondrial Genome Complement of *Vitis*

The mitochondrial genome we have assembled is the largest among published organelle genomes. However, this is not due to major changes in the gene set of grape mtDNA, which for the most part is fairly standard and typical for the mtDNA of angiosperms. Grape mtDNA encodes the household keeping genes related to biogenesis of electron-transporting machinery, structural components of the machinery itself, an intron maturase, and several structural ribosomal proteins. Like all other published mitochondrial genomes, grape mtDNA does not encode the full set of tRNAs necessary to recognize all codons and depends on import of nuclear encoded tRNA species for gene expression. The large genome size of *Vitis* mt DNA is not due to its genes but due to expansion of its intergenic spacers, which constitute ~90% of the genome size. Given the presence of a large number of direct repeats in the spacer regions, the growth of the genome most probably occurred via unequal homologous recombination/duplication of the repetitive sequences, a phenomenon known to contribute to the large mtDNA size (Lilly and Harvey 2001). The observation that the genome is completely rearranged with respect to the other angiosperm mtDNAs (see, for example, supplementary figure 4s, Supplementary Material online) seems to support this interpretation. As 80–90% of *Vitis* mtDNA sequence lacks any similarity to the sequences of other published angiosperm mtDNAs, it is not possible to provide a detailed stepwise analysis of its structural evolution. However, it is possible to trace the origin of most multiple gene copies harbored by grape mtDNA. Sequence comparisons have revealed that *Vitis* mtDNA pseudogenes have sequences identical or nearly identical to the corresponding sequences of functional genes also present in the grape mitochondrial genome. These observations raise doubt as to whether HGT is the only mechanism that plant mitochondria possess to establish divergent gene copies (Bergthorsson et al. 2004) because it seems more likely that the primary source of duplicate genes in the *Vitis* mt genome is from recombinations of the mtDNA itself.

An interesting feature of the gene content of the mtDNA of *Vitis* is the presence of a large ORF-encoding single-subunit RNA polymerase, from the pfam00949 superfamily. Because an intact gene encoding the mitochondrial RNA polymerase has previously been found in the nuclear genome of grape (Velasco et al. 2007), it is likely that its mitochondrial genome is transcribed by the nuclear encoded enzyme. The mtDNA-encoded RNA polymerase of *Vitis* is dissimilar to the nuclear homologue and resembles the polymerase encoded by ORF6 from the linear

mitochondrial plasmid of *Brassica napus* (Handa et al. 2002) and the RNA polymerase present in the mtDNA of *Daucus carota* (Robison and Wolyn 2005). Such linear plasmids of mitochondria resemble T-odd bacteriophages. They have terminal inverted repeats, ORFs encoding RNA and/or DNA polymerases, and proteins covalently attached to their 5′ end. These plasmids, originally discovered in the mitochondria of green plants and fungi (Meinhardt et al. 1990), were later found in a broad spectrum of eukaryotes, including the amoebozoan *Physarum polycephalum* (Sakurai et al. 2004), the heterokont *Pylaiella littoralis* (Oudot-Le Secq et al. 2001), and the flagellate *Jacoba libera* (Nosek and Tomashka 2003). Recently, Shutt and Gray (2006) have hypothesized that these linear plastids are a remnant of an ancient viral line (some T3 or T7 bacteriophage) that served as a progenitor of the mtDNA replication machinery of all modern eukaryotes. The function of these plasmids still remains unknown, but their presence has been associated with metabolic disorders. One such case was reported for maize, where the presence/integration of S1 and S2 linear plastids causes the S type of cytoplasmic male sterility (Schardl et al. 1984, 1985). Integration of a linear Kalilo plasmid into the mtDNA of *Neurospora intermedia* leads to a senescence phenotype and hyphal death (Bertrand and Griffiths 1989).

The proportion of sequences homologous to chloroplast DNA in the grape mtDNA is the largest observed in a mitochondrial genome, both in relative (8.8% of the whole mitochondrial genome length) and absolute terms (68,052 bp). As much as 42.3% of the grape cpDNA has become incorporated into the mitochondrial genome. This finding suggests caution when using PCR to amplify chloroplast markers for phylogeny reconstruction among angiosperms.

Investigation of HGT

We have found evidence of HGT into mtDNA of *Vitis* from other species but on a smaller scale than might have been expected given the findings of Bergthorsson et al. (2003, 2004). The only two convincing cases concern partial sequences of two viruses, which parasitize grape, and have been integrated into its mitochondrial genome. The extent of topological incongruence among the trees built from sequences of chloroplast genes of land plants, which are generally assumed to be vertically inherited, suggests that the argument for HGT among angiosperms, based on topological differences between single gene trees (Bergthorsson et al. 2003, 2004), is not strong.

The topological incongruence observed here and in earlier studies for chloroplast and mitochondrial gene trees might be explained by a variety of reasons. For example, the position of *Amborella* sequences in the trees built from *sdh4* gene sequences in Bergthorsson et al. (2004) (*Amborella*1 sister to *Podophyllum* and *Amborella*2 sister to *Nymphaea*) and here (*Amborella1* sister to *Amborella2*) implicates model selection as important for understanding topological incongruence. All trees providing phylogenetic evidence for HGT in Bergthorsson et al. (2003) were built employing a HKY85 model with transition-to-transversion ratio set to 2 (without gamma distribution of rates and without specifying proportion of invariable sites). Similarly, Bergthorsson et al. (2004) inferred HGT into the mtDNA of *Amborella* assuming a HKY + G model. These models were found to be suboptimal under model selection criteria implemented in the present study. Several *Amborella* genes (one or both amplicons of *atp1*, *ccmC*, *nad4*, *nad6*, *nad7*, *rpl16*, and *atp6*) occur on some of the longest branches in the trees shown in Bergthorsson et al. (2004). To explain the nonadjacent position of these amplified sequences, Bergthorsson et al. inferred gene transfer from unidentified land plant donors without considering the possibility of LBA. However, LBA artifacts may also explain some of the groupings in the mitochondrial gene trees (such as the *Pedinomonas* + *Picea* relationship in our *cob2* tree and *Zea* + *Marchantia* + *Physcomitrella* relationship in our *rpl2* tree).

In the ten gene trees reported for *atp4*, *atp6*, *atp9*, *cox2*, *cox3*, *ccmB*, *nad4*, *nad7*, *rpl16*, and *sdh4* in Bergthorsson et al. (2004), *Amborella* sequences exhibited unexpected phylogenetic relationships which were not strongly supported (below 80% bootstrap proportion). This is due to the relatively small number of informative sites in the data, an important reason why numerous authors, including those who previously advocated use of character-wise small data sets (e.g., Soltis et al. 2004), have turned to studies of large genome-scale data sets in recent times (Martin et al. 2005; Moore et al. 2007). Plant mitochondrial genes show limited sequence variation (Wolfe et al. 1987) and thus are expected to be especially sensitive to this type of stochastic error.

Although factors causing tree incongruence are different from alignment to alignment and, therefore, should not necessarily be exactly the same for chloroplast and mitochondrial genes, their cumulative effect observed in the analyses of 61 vertically inherited genes was so strong that it should warn against overinterpretation of the "strange" branches. The fact that only 28% (17 of 61) of alignments of these genes supported two benchmark clades (monocots and eudicots) illustrates that systematic errors and scarcity of informative characters should not be neglected in explanation of varying topologies of the single gene trees.

There is also a further consideration important for understanding the results observed by Bergthorsson et al. (2003, 2004). PCR amplified sequences of *A. trichopoda* used to infer HGT were obtained from total DNA preparations, and a number of these PCR products were extracted from gels containing multiple bands (Bergthorsson et al. 2004). A mixture of PCR products can be generated when the template contains different sets of pairs of the annealing sites not perfectly complementary to the primers. Because it is known that mitochondrial genes have numerous paralogous gene copies in the nuclear genome (Adams et al. 2002; Adams and Palmer 2003), simultaneous amplification of the paralogous nuclear and mitochondrial gene copies is a possible explanation for the nonadjacent positions of homologous sequences amplified from *Amborella* in the phylogenetic trees of Bergthorsson et al. (2003; 2004). Without studies to discount this possibility, nuclear encoded mt genes remain a concern for confusion in phylogenetic interpretation. This is suggested from our own studies shown here, where paralogous sequences from nuclear and mitochondrial genomes are found on the nonadjacent branches of

optimal ML trees. In these trees, the nuclear sequences are sometimes subtended on the long branches and sometimes on short branches. Therefore, a "conventional mitochondrial like branch length," cited by Bergthorsson et al. (2004) as one of the indirect indications of the mitochondrial nature of their amplified sequences, is not a reliable criterion for a mitochondrial origin of the PCR products. For this reason, we question whether sequences have indeed been transferred into the mtDNA of *Amborella* from different donor species. It is worthwhile to mention that earlier arguments in favor of horizontal transfer of mitochondrial *cox1* introns by Cho and Palmer (1999) were recently reevaluated by Cusimano et al. (2008), who suggested vertical transmission of these sequences. In any event, determination of the mitochondrial genome of *Amborella* will help to support or refute the suggestion of massive HGT into its mtDNA. Finally, although our inference of transfer of mitochondrial sequences into the chloroplast DNA of *Daucus* (Ruhlman et al. 2006) is not based on phylogenetic evidence (and subject to the difficulties of interpretation that we outline), our conclusion nevertheless assumes correct genome assembly. Given the unexpected nature of our finding, it may be important to evaluate the assembly in more detail.

Supplementary Material

Supplementary figures 1s, 2s, and 3s are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported by the "Grapevine Genome" project funded by the Autonomous Province of Trento. Special thanks to Dr Andrey Zahrkikh for the assembly of the mitochondrial sequences.

Literature Cited

Adam-Blondon AF, Bernole A, Faes G, Lamoureux D, Pateyron S, Grando MS, Caboche M, Velasco R, Chalhoub B, et al. 2005. Construction and characterization of BAC libraries from major grapevine cultivars. Theor. Appl. Genet. 110:1363–1371.

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

Adams KL, Qiu Y-L, Stoutemyer M, Palmer JD. 2002. Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. Proc Natl Acad Sci USA. 99:9905–9912.

Allison LA, Simon LD, Maliga P. 1996. Deletion of *rpo*B reveals a second distinct transcription system in plastids of higher plants. EMBO J. 15:2802–2809.

Atlan A, Couvet D. 1993. A model simulating the dynamics of plant mitochondrial genomes. Genetics. 135:213–222.

Bergthorsson U, Adams KL, Thomason B, Palmer JD. 2003. Widespread horizontal transfer of mitochondrial genes in flowering plants. Nature. 424:197–201.

Bergthorsson U, Richardson AO, Young GJ, Goertzen LR, Palmer JD. 2004. Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm *Amborella*. Proc Natl Acad Sci USA. 101:17747–17752.

Bertrand H, Griffiths AJF. 1989. Linear plasmids that integrate into mitochondrial DNA in *Neurospora*. Genome. 31:155–159.

Boussau B, Karlberg EO, Frank AC, Legault BA, Andersson SG. 2004. Computational inference of scenarios for α-proteobacterial genome evolution. Proc Natl Acad Sci USA. 101:9722–9727.

Brennicke A, Grohmann L, Hiesel R, Knoop V, Schuster W. 1993. The mitochondrial genome on its way to the nucleus: Different stages of gene transfer in higher plants. FEBS Lett. 325:140–145.

Buck KW. 1986. Fungal virology-an overview. In: Buck KW, editor. Fungal virology. Boca Raton (FL): CRC Press. p. 1–84.

Buck KW. 1998. Molecular variability of viruses of fungi. In: Bridge PD, Couteaudier Y, Clarkson JM, editors. Molecular variability of fungal pathogens. Wallingford (UK): CAB International. p. 53–72.

Cho Y, Palmer JD. 1999. Multiple acquisitions via horizontal transfer of a group I intron in the mitochondrial *cox1* gene during evolution of the Araceae family. Mol Biol Evol. 16:1155–1165.

Cole TE, Hong Y, Brasier CM, Buck KW. 2000. Detection of an RNA-dependent RNA polymerase in mitochondria from a mitovirus-infected isolate of the Dutch elm disease fungus, *Ophiostoma novo-ulmi*. Virology. 268:239–243.

Cummings MP, Nugent JM, Olmstead RG, Palmer JD. 2003. Phylogenetic analysis reveals five independent transfers of the chloroplast gene *rbc*L to the mitochondrial genome in angiosperms. Curr Genet. 43:131–138.

Cusimano N, Zhang L-B, Renner SS. 2008. Reevaluation of the cox1 group I intron in Araceae and angiosperms indicates a history dominated by loss rather than horizontal transfer. Mol Biol Evol. 25:265–276.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 19:11–15.

Dyall SD, Brown MT, Johnson PJ. 2004. Ancient invasions: from endosymbionts to organelles. Science. 304:253–257.

Fauron C, Casper M, Gao Y, Moor B. 1995. The maize mitochondrial genome: dynamic, yet functional. Trends Genet. 11:228–235.

Ghabrial SA. 1994. New developments in fungal virology. Adv Virus Res. 43:303–388.

Gibbs MJ, Pfeiffer P, Fukuhara T. 2004. Genus endornavirus. In: Fauquet CM, Mayo MA, Maniloff J, Desselberger U, Ball AL, editors. Virus taxonomy. Eighth report of the international committee on taxonomy of viruses. London: Elsevier/Academic Press. p. 603–605.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 52:696–1504.

Handa H. 2003. The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica napus* L.) comparative analysis of the mitochondrial genomes of rapeseed and *Arabidopsis thaliana*. Nucleic Acids Res. 31:5907–5916.

Handa H, Itani K, Sato H. 2002. Structural features and expression analysis of a linear mitochondrial plasmid in rapeseed (*Brassica napus* L.). Mol Genet Genomics. 267:797–805.

Hedtke B, Borner T, Weihe A. 1997. Mitochondrial and chloroplast phage-type RNA polymerases in *Arabidopsis*. Science. 277:809–811.

Hong Y, Cole TE, Brasier CM, Buck KW. 1998. Evolutionary relationships among putative RNA-dependent RNA polymerases encoded by a mitochondrial virus-like RNA in the

Dutch elm disease fungus, *Ophiostoma novo-ulmi*, by other viruses and virus-like RNAs and by the *Arabidopsis* mitochondrial genome. Virology. 246:158–169.
Izuchi S, Terachi T, Sakamoto M, Mikami T, Sugita M. 1990. Structure and expression of tomato mitochondrial genes coding for $tRNA^{Cys}$(GcA), $tRNA^{Asn}$(Guu) and $tRNA^{Tyr}$ (GUA): a native $tRNA^{Cys}$ gene is present in dicot plants but absent in monocot plants. Curr Genet. 18:239–243.
Jansen RK, Kaittanis C, Saski C, Lee S-B, Tompkins J, Alverson AJ, Daniell H. 2006. Phylogenetic analyses of *Vitis* (Vitaceae) based on complete chloroplast genome sequences: effects of taxon sampling and phylogenetic methods on resolving relationships among rosids. BMC Evol Biol. 6:32.
Joyce C, Steitz T. 1994. Function and structure relationships in DNA polymerases. Annu Rev Biochem. 63:777–822.
Joyce PBM, Gray MW. 1989. Chloroplast-like transfer RNA genes expressed in wheat mitochondria. Nucleic Acids Res. 17:5461–5476.
Kubo T, Mikami M. 2007. Organisation and variation of angiosperm mitochondrial genome. Physiol Plant. 129:6–13.
Kühn K, Bohne A-V, Liere K, Weihe A, Börner T. 2007. Arabidopsis phage-type RNA polymerases: accurate in vitro transcription of organellar genes. Plant Cell. 19:959–971.
Lilly JW, Harvey MJ. 2001. Small, repetitive DNAs contribute significantly to the expanded mitochondrial genome of cucumber. Genetics. 159:317–328.
Lonsdale DM. 1989. The plant mitochondrial genome. In: Marcus A, editor. The biochemistry of plants, vol. 15. New York: Academic Press. p. 229–295.
Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25:955–964.
Marienfeld JR, Unseld M, Brandt P, Brennicke A. 1997. Viral nucleic acid sequence transfer between fungi and plants. Trends Genet. 13:260–261.
Marrchal L, Guillemaut P, Grienenberger JM, Jeannin G, Weil JH. 1985. Sequence and codon recognition of bean mitochondria and chloroplast tRNAsTrp: evidence for a high degree of homology. Nucleic Acids Res. 13:4411–4416.
Martin W, Deusch O, Stawski N, Grünheit N, Goremykin V. 2005. Chloroplast genome phylogenetics: why we need independent approaches to plant molecular evolution. Trends Plant Sci. 10:203–209.
Meinhardt F, Kempken F, Kaemper J, Esser K. 1990. Linear plasmids among eukaryotes: fundamentals and application. Curr Genet. 17:89–95.
Moore MJ, Bell CD, Soltis PS, Soltis DE. 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. Proc Natl Acad Sci USA. 104: 19363–19368.
Nakazono M, Itadani H, Wakasugi T, Tsutsumi N, Sugiura M, Hirai A. 1995. The rps3-rpl16-nad3-rps12 gene cluster in rice mitochondrial DNA is transcribed from alternative promoters. Curr Genet. 27:184–189.
Nedelcu AM, Lee RW, Lemieux C, Gray MW, Burger G. 2000. The complete mitochondrial DNA sequence of *Scenedesmus obliquus* reflects an intermediate stage in the evolution of the green algal mitochondrial genome. Genome Res. 10:819–831.
Nosek J, Tomashka L. 2003. Mitochondrial genome diversity: evolution of the molecular architecture and replication strategy. Curr Genet. 44:73–84.
Ogihara Y, Yamazaki Y, Murai K, et al. (14 co-authors). 2005. Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome. Nucleic Acids Res. 33:6235–6250.
Oudot-Le Secq M-P, Fontaine JM, Rousvoal S, Kloareg B, Loiseaux-De Goeir S. 2001. The complete sequence of a brown algal mitochondrial genome, the ectocarpale *Pylaiella littoralis* (L.) Kjellm. J Mol Evol. 53:80–88.
Posada D, Crandall KA. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics. 14:817–818.
Richardson AO, Palmer JD. 2007. Horizontal gene transfer in plants. J Exp Bot. 58:1–9.
Robison MM, Wolyn DJ. 2005. A mitochondrial plasmid and plasmid-like RNA and DNA polymerases encoded within the mitochondrial genome of carrot (*Daucus carota* L.). Curr Genet. 47:57–66.
Ruhlman T, Lee SB, Jansen RK, Hostetler JB, Tallon LJ, Town CD, Daniell H. 2006. Complete plastid genome sequence of *Daucus carota*: implications for biotechnology and phylogeny of angiosperms. BMC Genomics. 7:224.
Sakurai R, Nomura H, Moriyam Y, Kawano S. 2004. The mitochondrial plasmid of the true slime mold *Physarum polycephalum* bypasses uniparental inheritance by promoting mitochondrial fusion. Curr Genet. 46:103–114.
Schardl CL, Lonsdale DM, Pring DR, Rose KR. 1984. Linearization of maize mitochondrial chromosomes by recombination with linear episomes. Nature. 310:292–296.
Schardl CL, Pring DR, Lonsdale DM. 1985. Mitochondrial DNA rearrangements associated with fertile revertants of S-type male sterile maize. Cell. 43:361–368.
Sederoff RR. 1987. Molecular mechanisms of mitochondrial genome evolution in higher plants. Am Nat. 130:30–45.
Shutt TE, Gray MW. 2006. Bacteriophage origins of mitochondrial replication and transcription proteins. Trends Genet. 22:90–95.
Soltis DE, Albert VA, Savolainen V, et al. (11 co-authors). 2004. Genome-scale data, angiosperm relationships, and "ending incongruence": a cautionary tale in phylogenetics. Trends Plant Sci. 9:477–483.
Stern DB, Lonsdale DM. 1982. Mitochondrial and chloroplast genomes of maize have a 12-kilobase DNA sequence in common. Nature. 299:698–702.
Sugiyama Y, Watase Y, Nagase M, Makita N, Yagura S, Hirai A, Sugiura M. 2005. The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants. Mol Genet Genomics. 272:603–615.
Swofford DL. 2002. PAUP*: phylogenetic analysis using parsimony (* and other methods). Version 4. Sunderland (MA): Sinauer associates.
Timmis J, Ayliffe MA, Hyang CY, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet. 5:123–135.
Tuomivirta TT, Hantula J. 2005. Three unrelated viruses occur in a single isolate of *Gremmeniella abietina* var. *abietina* type A. Virus Res. 110:31–39.
Turmel M, Otis C, Lemieux C. 2002. The chloroplast and mitochondrial genome sequences of the charophyte *Chaetosphaeridium globosum*: insights into the timing of the events that restructured organelle DNAs within the green algal lineage that led to land plants. Proc Natl Acad Sci USA. 99:11275–11280.
Unseld M, Mariefield JR, Brandt P, Brennicke A. 1997. The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. Nat Genet. 15:57–61.
Velasco R, Zharkikh A, Troggio M, et al. (57 co-authors). 2007. A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. PLoS ONE. 2:e1326.
Ward BL, Anderson RS, Bendich AJ. 1981. The size of the mitochondrial genome is large and variable in a family of plants (Cucurbitaceae). Cell. 25:793–803.
Weihe A, Hedtke B, Borner T. 1997. Cloning and characterization of a cDNA encoding a bacteriophage-type RNA

polymerase from the higher plant *Chenopodium album*. Nucleic Acids Res. 25:2319–2325.
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.
Zharkikh A, Troggio M, Pruss D, et al. (15 co-authors). 2008. Sequencing and assembly of highly heterozygous genome of *Vitis vinifera* L. cv Pinot Noir: Problems and solutions. J Biotechnol. 136:38–43.

Peter Lockhart, Associate Editor

Accepted October 4, 2008