

## Sequence of the Tomato Chloroplast DNA and Evolutionary Comparison of Solanaceous Plastid Genomes

Sabine Kahlau,<sup>1</sup> Sue Aspinall,<sup>2</sup> John C. Gray,<sup>2</sup> Ralph Bock<sup>1</sup>

<sup>1</sup> Max-Planck-Institut für Molekulare Pflanzenphysiologie, Am Mühlenberg 1, Potsdam-Golm D-14476, Germany

<sup>2</sup> Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK

Received: 25 October 2005 / Accepted: 14 March 2006 [Reviewing Editor: Rüdiger Cerff]

**Abstract.** Tomato, *Solanum lycopersicum* (formerly *Lycopersicon esculentum*), has long been one of the classical model species of plant genetics. More recently, solanaceous species have become a model of evolutionary genomics, with several EST projects and a tomato genome project having been initiated. As a first contribution toward deciphering the genetic information of tomato, we present here the complete sequence of the tomato chloroplast genome (plastome). The size of this circular genome is 155,461 base pairs (bp), with an average AT content of 62.14%. It contains 114 genes and conserved open reading frames (*ycfs*). Comparison with the previously sequenced plastid DNAs of *Nicotiana tabacum* and *Atropa belladonna* reveals patterns of plastid genome evolution in the Solanaceae family and identifies varying degrees of conservation of individual plastid genes. In addition, we discovered several new sites of RNA editing by cytidine-to-uridine conversion. A detailed comparison of editing patterns in the three solanaceous species highlights the dynamics of RNA editing site evolution in chloroplasts. To assess the level of intraspecific plastome variation in tomato, the plastome of a second tomato cultivar was sequenced. Comparison of the two genotypes (IPA-6, bred in South America, and Ailsa Craig, bred in Europe) revealed no nucleotide differences, suggesting that the plastomes of modern tomato cultivars display very little, if any, sequence variation.

**Key words:** *Lycopersicon esculentum* — *Solanum lycopersicum* — Plastid — Plastid genome — RNA editing — Intraspecific variation

### Introduction

The Solanaceae (nightshade) family consists of more than 3000 species and has its center of diversity near the equator in South America. The family comprises many agriculturally important crop species, such as tomato, potato, bell pepper, and eggplant, as well as a number of ornamental and medicinal plants. For several thousand years, solanaceous crops have been subjected to intensive human selection. This has led to an enormous phenotypic diversity within species and the adaptation of individual varieties to widely different habitats. Showing this great interspecific and intraspecific diversity and being the most important plant family of vegetable crops, Solanaceae have recently become a model of comparative and evolutionary genomics research. These efforts are bundled in the international Solanaceae Genomics Network (SOL Genomics Network; SGN: <http://www.sgn.cornell.edu/>), which aims to take a systems approach to genetic diversity and adaptation. While EST projects are under way for a number of solanaceous species and selected members of sister families, the structural genomics efforts are centered on tomato, one of the classical model systems of plant genetics and breeding. The tomato nuclear genome comprises approximately 950 Mb and the gene-rich

Correspondence to: Ralph Bock; email: rbock@mpimp-golm.mpg.de

euchromatic portion of the tomato nuclear genome is currently being sequenced.

The plastid genome (plastome) of higher plants is a circular molecule of double-stranded DNA. Among the three genomes of the plant cell, the plastome is the most gene-dense, with more than 100 genes in a genome of only 120 to 210 kb (for review see, e.g., Sugiura 1989, 1992; Wakasugi et al. 2001). The plastid genome is the evolutionary remnant of a cyanobacterial genome: The endosymbiotic uptake of a cyanobacterium by eukaryotic cells was followed by (i) the loss of dispensable genetic information from the endosymbiont's genome (e.g., genes for bacterial cell wall biosynthesis), (ii) the elimination of redundant genetic information (e.g., genes for biosynthetic pathways present in both the host's and the endosymbiont's genomes), (iii) the acquisition of new (regulatory) gene functions to coordinate gene expression and metabolism between the host cell and the endosymbiont, and (iv) the massive translocation of genetic information from the endosymbiont's genome to the nuclear genome of the host cell (Martin and Herrmann 1998; Race et al. 1999; Timmis et al. 2004). This has resulted in a dramatic reduction in plastid genome size and coding capacity and, thus, contemporary plastid genomes contain only a small proportion of the genes of their free-living cyanobacterial ancestors. Consequently, all cellular functions fulfilled by present-day plastids are strictly dependent on the import of nuclear-encoded proteins which make up the by far largest fraction of the chloroplast proteome (Abdallah et al. 2000; Rujan and Martin 2001; Martin et al. 2002; Hippler and Bock 2004).

Both genome organization and mechanisms of gene expression in present-day plastids resemble those of their cyanobacterial ancestors. Groups of genes are linked together in operons giving rise to polycistronic mRNAs which undergo complex processing and maturation steps. Likewise, the translational apparatus of plastids is highly similar to that of prokaryotes and the plastid-encoded RNA polymerase is highly homologous to eubacterial RNA polymerases. In addition to these conserved prokaryotic traits, several evolutionary inventions of the eukaryotic cell have added to the complexity of gene expression and its regulation in plastids. These include, for example, transcription of some genes by nuclear-encoded bacteriophage-type RNA polymerases (Hajdukiewicz et al. 1997; Hedtke et al. 1997; Hess and Börner 1999) and RNA editing as an additional processing step changing the coding properties of chloroplast transcripts (Hoch et al. 1991; Bock 2000, 2001).

We report here the complete sequence of the plastid genome (plastome) from two cultivars of tomato, *Solanum lycopersicum*. To determine patterns

of plastid genome evolution in Solanaceae, we have compared the tomato plastid genome, its structure, coding capacity, and RNA editing sites, with the two previously sequenced plastomes from tobacco (*Nicotiana tabacum* [Shinozaki et al. 1986; Wakasugi et al. 1998]) and deadly nightshade (*Atropa belladonna* [Schmitz-Linneweber et al. 2002]).

## Materials and Methods

### Plant Material

*Solanum lycopersicum* cv. IPA-6 is a commercially grown South American tomato cultivar. Seeds of cv. Ailsa Craig were obtained from Unwins Seeds (Histon, Cambridge, UK) and germinated and grown in a greenhouse with supplementary lighting of 200  $\mu\text{mol photons m}^{-2} \text{ s}^{-1}$ .

### Purification of Chloroplasts

For large-scale purification of chloroplasts, IPA-6 plants were grown for 6 weeks in the greenhouse (16 h/22°C light, 8 h/20°C dark). For each isolation, the young expanded leaves from 10 plants were pooled. Leaf material (50 g) was homogenized for 2 × 5 s at high speed and 2 × 5 s at low speed in 2 L ice-cold extraction buffer (350 mM sorbitol, 50 mM Tris-HCl, pH 8.0, 5 mM EDTA, 15 mM 2-mercaptoethanol, 0.1% BSA) in a Waring blender. The homogenate was filtered through four layers of gauze (Hartmann) and one layer of Miracloth (Calbiochem). All subsequent steps were performed at 4°C. The chloroplast suspension was centrifuged at 100g for 5 min and the resulting pellet (largely consisting of cell nuclei) was discarded. The supernatant was centrifuged for 10 min at 2000g to pellet the chloroplasts. Following resuspension in 400 ml wash buffer (350 mM sorbitol, 50 mM Tris-HCl pH 8.0, 25 mM EDTA, 0.1% BSA), the chloroplasts were pelleted again by centrifugation for 10 min at 2000g. This washing step was repeated three more times. Subsequently, the chloroplast pellet was resuspended in 30 ml wash buffer and loaded on top of six sucrose step gradients (17.5 ml 60% sucrose, 22.5 ml 37% sucrose, each with the same concentrations of Tris and EDTA as present in the wash buffer). The gradients were centrifuged for 1 h at 7000g. The chloroplast band at the interphase between the 37% and the 60% sucrose layers was collected, mixed with ~3 vol dilution buffer (175 mM sorbitol, 50 mM Tris-HCl, pH 8.0, 25 mM EDTA), and pelleted by centrifugation for 10 min at 2000g and 4°C.

### Isolation of Nucleic Acids

DNA was extracted from purified chloroplasts by lysing the chloroplast pellet in 1–2 vol lysis buffer (50 mM Tris-HCl, pH 8.0, 20 mM EDTA, 2% *N*-lauroylsarcosine sodium salt) for 15 min followed by one extraction each with phenol, phenol/chloroform (1:1), and chloroform. Subsequently, the DNA was precipitated with 2.5 vol ethanol at –20°C overnight. Total cellular RNA was extracted using the peqGOLD TriFast reagent (Peqlab GmbH, Erlangen, Germany). RNA samples for cDNA synthesis were purified by treatment with RNase-free DNase I (Roche, Mannheim, Germany).

### Cloning and DNA Sequencing

Purified IPA-6 plastid DNA (25  $\mu\text{g}$ ) was used to generate fragments for construction of a shotgun library by mechanical shear-

ing. Fragments of an average length of 2 kb were cloned into pUC19 and 768 randomly selected clones were terminally sequenced from both ends (AGOWA GmbH, Berlin). The sequence data from the 1536 sequencing reactions were assembled into contigs and remaining gaps were closed by PCR using primers derived from the sequences flanking the gap. A list of PCR primers is available upon request. Amplified products were purified using the GFX PCR (DNA and Gel Band Purification) kit (Amersham). To exclude mutations introduced by PCR amplification, PCR products were directly sequenced by cycle sequencing.

The nucleotide sequence of the Ailsa Craig plastome was determined on cloned *Pst*I restriction fragments obtained previously (Phillips 1985) using primers designed from the sequence of the tobacco plastid genome (Wakasugi et al. 1998) and from the resulting tomato sequence. Uncloned regions and regions spanning the *Pst*I sites used for the original cloning were amplified by PCR and the PCR products directly sequenced.

### *cDNA Synthesis and Polymerase Chain Reactions (PCRs)*

Approximately 5 µg purified DNA-free RNA was used in a 50-µl cDNA synthesis reaction. Reverse transcription of RNA samples was primed with a random hexanucleotide mixture (2.5 µg per reaction). Elongation reactions were performed with SuperScript III RNase H-free reverse transcriptase (Invitrogen) according to the manufacturer's instructions. Total cellular DNA or first-strand cDNAs were amplified in an Eppendorf thermal cycler using GoTaq Flexi DNA polymerase (Promega) and gene-specific primer pairs. The standard PCR program was 30 to 40 cycles of 1 min at 94°C, 40 s at 58°C, and 1 to 2.5 min at 72°C, with a 10-min extension of the first cycle at 94°C and a 5-min final extension at 72°C.

### *Bioinformatics Analyses*

The Lasergene software package (DNASTAR; GATC Biotech, Konstanz, Germany) was used to assemble the final genome sequence and to align the genome with other plastid DNAs. Whole-genome alignments were produced with the Martinez/Needleman-Wunsch method, applying a gap penalty of 1.10 and a gap length penalty of 0.33. DNA sequences for individual genes were compared with sequences available in the databases using the NCBI BLAST program. Nucleotide sequences for species other than tomato were extracted from public databases and aligned with the corresponding tomato sequences. The physical map of the tomato plastome was drawn using the Adobe Illustrator software.

## **Results and Discussion**

### *Properties of the Tomato Plastid Genome*

In order to construct a shotgun clone library for sequencing of the tomato IPA-6 plastid genome, chloroplasts were purified from young tomato leaves at large scale. Purified chloroplast DNA was sheared mechanically and used for shotgun cloning followed by terminal sequencing of individual clones. Altogether 768 clones were sequenced in order to obtain a roughly eightfold coverage of the genome. Assembly of the sequence data from 1536 sequencing reactions yielded five sequence contigs. Alignment of these

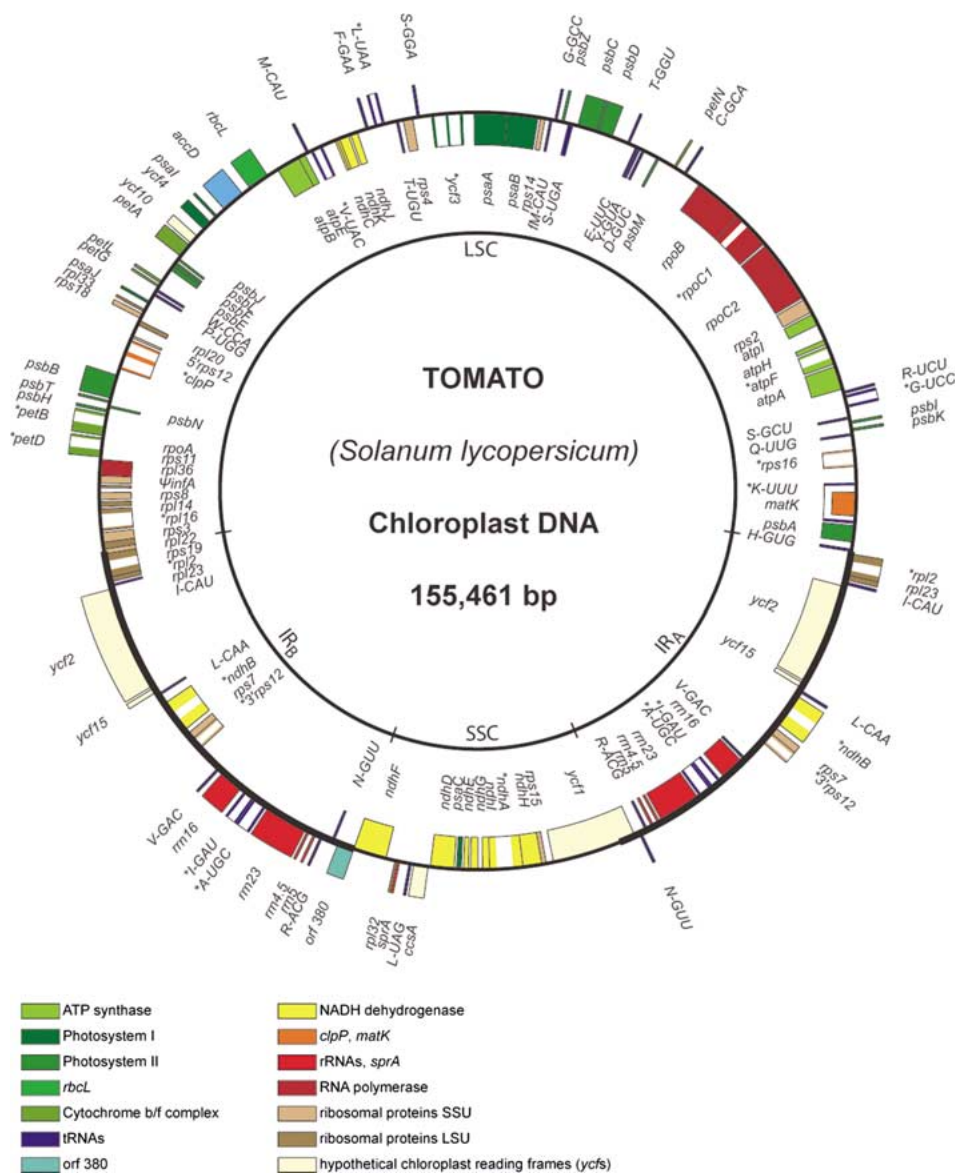
contigs with the tobacco plastid genome confirmed the presence of four gaps and suggested that all four gaps were small enough to be closed by PCR (gap sizes in the tobacco ptDNA: 1901, 1798, 218, and 48 bp, respectively). Primers derived from the termini of the contigs yielded PCR products for all four gaps. Direct sequencing of purified PCR products provided the missing sequence information and thus allowed gap closure and assembly of a single contig with a circular map (Fig. 1), indicating that the complete sequence of the tomato plastid genome had been obtained (accession number AM087200).

In order to identify possible causes for gaps in the sequence despite the high-coverage sequencing, the sequences at the gap termini were inspected. In two cases, the DNA sequences had become ambiguous at oligo(T) tracts (of 10 and 13 Ts, respectively), which apparently caused DNA polymerase stuttering in standard cycle sequencing reactions.

The size of the tomato plastid DNA was found to be 155,461 bp (Fig. 1, Table 1) This deviates only slightly from previous estimates based on gel electrophoretic separations of restriction fragments which suggested a plastome size of 156.6 to 159.4 kb (Phillips 1985). The tomato plastome shows the typical tetrapartite genome organization found in most higher plants with a large single copy region (LSC) and a small single copy region (SSC) separating two inverted repeat regions ( $IR_A$  and  $IR_B$ ; Fig. 1). As expected, the tomato plastome harbors the conserved set of genes present in the plastid genomes of dicotyledonous plants. With the exception of several open reading frames (ORFs; which are discussed below), the gene content is identical to the previously analyzed plastid DNAs from *Nicotiana* (Shinozaki et al. 1986; Wakasugi et al. 1998; accession number of an updated release from September 2005, Z00044.2) and *Atropa* (Schmitz-Linneweber et al. 2002). The tomato plastome harbors 114 genes and conserved ORFs (*ycfs*: hypothetical chloroplast reading frames). Incorporating the data from recent reverse genetics studies, these genes can be grouped as follows (Shimada and Sugiura 1991) (Fig. 1).

#### Photosynthesis-Related Genes.

- 7 genes for photosystem I proteins, including *ycf3* and *ycf4*, 2 genes for proteins involved in photosystem I assembly (Ruf et al. 1997; Boudreau et al. 1997)
- 15 genes for subunits of photosystem II, including *psbZ* (formerly *ycf9*), encoding a protein that couples the light-harvesting complex protein CP26 to photosystem II (Ruf et al. 2000; Swiatek et al. 2001)
- 6 genes for subunits of the cytochrome  $b_6$  complex, including the most recently discovered *petN* (formerly *ycf6* (Hager et al. 1999))



**Fig. 1.** Physical map of the tomato plastid genome. Genes inside the circle are transcribed clockwise; genes outside the circle are transcribed counterclockwise. Asterisks indicate intron-containing genes; introns are depicted as open boxes.

- 6 genes for subunits of the chloroplast ATP synthase
- 11 genes for subunits of a chloroplast NAD(P)H dehydrogenase suggested to be involved in cyclic electron flow around photosystem I (Burrows et al. 1998; Shikanai et al. 1998; Joet et al. 2001; Munekage et al. 2004)
- *rbcL*, encoding the large subunit of Rubisco
- *yef10*, a conserved ORF encoding a chloroplast inner envelope protein reportedly involved in inorganic carbon uptake (Sasaki et al. 1993b; Rolland et al. 1997)

#### Genetic System Genes

- 30 tRNA genes believed to constitute a complete set for decoding all codons in protein-coding genes
- 4 rRNA genes

- 21 genes for ribosomal proteins (9 proteins of the large subunit and 12 proteins of the small subunit of the plastid 70S ribosome)
- 4 genes for subunits of the *E. coli*-like plastid RNA polymerase (PEP)
- *matK*, a gene suggested to encode RNA maturase involved in the removal of a subset of chloroplast group II introns (Hess et al. 1994; Liere and Link 1995; Mohr et al. 1993; Jenkins et al. 1997)
- *clpP*, encoding a subunit of a chloroplast protease (Maurizi et al. 1990; Gray et al. 1990; Shanklin et al. 1995; Majeran et al. 2000)

#### Other Genes and Conserved Open Reading Frames

- *accD*, encoding a subunit of acetyl-CoA carboxylase (Sasaki et al. 1993a, 1995)

**Table 1.** Properties of the sequenced solanaceous plastid genomes

	<i>Solanum lycopersicum</i>	<i>Nicotiana tabacum</i>	<i>Atropa belladonna</i>
Genome size (bp)	155,461	155,943	156,688
LSC (bp)	85,882	86,686	86,868
SSC (bp)	18,363	18,573	18,008
IR (bp)	25,608	25,342	25,906
Coding regions (%)	58.79	58.50	58.03
Introns (%)	11.73	11.70	11.54
Intergenic regions (%)	29.48	29.80	30.43
AT content (%)			
Overall	62.14	62.15	62.44
Coding regions	59.62	59.69	60.07
Noncoding regions	65.73	65.63	67.40
tRNAs	47.01	47.06	47.40
rRNAs	44.66	44.64	44.64
Protein-coding genes	61.75	61.82	61.96
LSC	64.01	64.05	64.37
SSC	67.97	67.94	68.35
IR	56.91	56.78	57.15

- *ccsA* (*yef5*), the protein product of which is required for heme attachment to chloroplast c-type cytochromes (Orsat et al. 1992; Xie et al. 1998; Xie and Merchant 1996)
- *yef1* and *yef2*, two genes encoding essential proteins of unknown function (Drescher et al. 2000)
- *yef15*, an ORF of unknown function
- *sprA*, a stable noncoding RNA of unknown function (Vera and Sugiura 1994; Sugita et al. 1997)

#### Comparison of Solanaceous Plastid Genomes

Completion of the nucleotide sequence of the tomato plastid genome offered the unique opportunity to conduct an in-depth comparison of the plastomes of three closely related species belonging to the same family of dicotyledonous plants: *Nicotiana tabacum* (tobacco [Shinozaki et al. 1986; Wakasugi et al. 1998]), *Atropa belladonna* (Schmitz-Linneweber et al. 2002), and *Solanum lycopersicum* (tomato).

Plastid genome sizes, structural properties, and AT content in different genome regions and gene classes are compared in Table 1. Compared to the tobacco plastome, the inverted repeat region (IR) in tomato is slightly expanded on both ends (into *rps19* in the LSC and into *yef1* in the SSC; Fig. 1, Table 1). Nonetheless, the tomato genome is smaller than the tobacco plastome, which can be chiefly ascribed to deletions in noncoding intergenic spacer regions (Supplementary Table 1). The tomato plastome is also smaller than that of *Atropa*, which is mainly due to a slightly larger IR in *Atropa* (Table 1). While the overall AT content is nearly identical in tobacco and tomato, it is significantly higher in *Atropa* (Table 1). Whether this can be explained by a somewhat stronger selection pressure toward AT richness operating in *Atropa* or,

**Table 2.** Conservation of open reading frames in the tobacco and tomato plastid genomes

ORF in tobacco	Location	Homologous sequence in tomato: mutations; longest ORF
70A	LSC	1-bp insertion; ORF74
74	LSC	2-bp deletion, 7-bp insertion; ORF25
99	LSC	Nonsense mutation; ORF18
103	LSC	1-bp insertion, 77-bp deletion (including initiation codon)
105	LSC	9-bp deletion, 5-bp insertion; ORF74
70B	IR	Conserved
75	IR	Conserved
79	IR	Conserved
92	IR	53-bp deletion, 77-bp deletion; ORF54
115	IR	53-bp deletion, 77-bp deletion; ORF18
131	IR	Conserved
350	IR	Truncated <i>yef1</i> ; IR in tomato extended, resulting in ORF380

alternatively, by differences in the mutation rate and/or spectrum remains to be investigated. Comparison of the AT content in different regions and gene classes reveals striking differences: In all three genomes, AT content is highest in the SSC region and lowest in the IR. Noncoding regions have a dramatically higher AT content than coding regions and protein coding genes are much more rich in AT than RNA genes (Table 1) (Shimada and Sugiura 1991). The latter may be explicable by a high demand for stable GC base pairs to ensure proper folding of the highly structured rRNAs and tRNAs.

While the gene content of the tomato plastome is identical to that of the previously sequenced solanaceous plastid genomes, it was of interest to assess the conservation of ORFs. ORFs that are not conserved between closely related solanaceous species are unlikely to be genuine genes. Table 2 shows a comparison of ORFs in the tobacco and tomato plastomes. This data set excludes highly conserved ORFs (*yef5*), for most of which good experimental support has been gained that they constitute genuine genes (see list above). Many of the tobacco ORFs are not conserved and have suffered frameshift mutations and/or larger deletions in tomato, suggesting that these reading frames are fortuitously present in tobacco and are unlikely to encode functional gene products. A few short ORFs in the IR are conserved, which, however, is not necessarily indicative of a possible functional significance: It is well established that the mutation rate in the IR is much lower than in the single-copy regions of the plastome (Wolfe et al. 1987; Maier et al. 1995). This phenomenon is generally explained by the operation of gene conversion between the two IR copies and may well be responsible for the conservation of some of the short ORFs in the IR (Table 2). In the absence of experimental evidence supporting a possible function, these ORFs

**Table 3.** InDels in coding regions of the tomato plastid genome

Gene	Number of inDels	InDel length (bp)	Location (region)	Frameshift (yes/no)
<b>Insertions</b>				
<i>rrn16</i>	1	9	IR	
<i>yef2</i>	1	3	IR	-
<i>accD</i>	1	9	LSC	-
<i>trnS-UGA</i>	1	1	LSC	
<i>yef1</i>	5	30	SSC	-
<b>Deletions</b>				
<i>yef2</i>	1	9	IR	-
<i>accD</i>	1	24	LSC	-
<i>rps16</i>	1	10	LSC	+
<i>yef1</i>	3	60	SSC	-

were therefore not considered in the physical map of the tomato plastome (Fig. 1). Tomato ORF380 (equivalent to ORF350 in tobacco) represents a special case in that it encodes the N-terminal portion of the Ycf1 protein, the function of which is unknown but has been shown to be essential for cell survival (Drescher et al. 2000). Being a partial duplication of *yef1*, it is not surprising that this ORF is conserved. As also the expression signals upstream of the reading frame (promoter, 5' UTR) are identical to those of *yef1* (Fig. 1), it is reasonable to assume that ORF380 is expressed. However, whether or not the corresponding protein product serves some function remains to be investigated.

When the insertions and deletions (InDels) in the tomato plastome were analyzed using the tobacco sequence as a reference, the vast majority of InDels were located in noncoding spacer regions and introns (Supplementary Table 1). However, a few InDels affect coding regions (Table 3), which prompted us to analyze the consequences for the encoded gene products. With a single exception, all InDels in protein-coding regions do not alter the reading frame and thus just change the lengths of the encoded protein by one or a few amino acids (Table 3). The only exception is *rps16*, where the deletion of 10 nucleotides causes a frameshift mutation (Table 3, Fig. 2). This frameshift mutation, however, has occurred very close to the termination codon so that the resulting changes at the amino acid level are limited to the very C-terminus of the protein (Fig. 2). The C-terminus of Rps16 is not very well conserved among higher plant species (Fig. 2), suggesting that the frameshift mutation is functionally neutral. Two other InDels affect RNA genes (16S rRNA and tRNA-Ser-UGA). Alignment of the corresponding sequences from a number of species revealed that both InDels are in variable regions of the two genes, again suggesting that they are unlikely to negatively impact on gene product function (Fig. 2).

Having three solanaceous plastome sequences available enabled pairwise comparisons of the three

**16S rRNA**

```
TCGGACGGGAAGT-----GGTGTTCAGT      Nt
TCGGACGGGAAACACGGGAAACGGTGTTCAGT    S1
TCGGACGGGAAACACGGGAAACCGTGTTCAGT    Ab
TCGGACGGGAAGT-----GGTGTTCAGT      At
TCGGACGGGAAGCATCCTAG-TGGTGTTCAGT     Pc
```

***trnS-UGA***

```
AACCGGTATAGTTTT-AACAAAGAACTAT      Nt
AACCGGTATAGTTTTGAACAAAGAAGTAT      S1
AACCGGTATAGTTTTGAACAAAGAAGTAT      Ab
AACCGGTATAGTTCGAAACAAAGAAGTAT      Ps
AACCGGTATAGTTTTGAACAAAGAAGTAT      Va
AACCGGTATAGTTCCTTAAACAAGAACTAT      Ca
AACCGGTATAGTTCT-----AGGAACTAT      Os
AACCGGTATAGAT---A-----AAAAACTAT    Pt
AACCGGTATAGTTCT-----AGGAACTAT      Zm
```

***rps16***

```
GAGGTTTTTAAGGAACTTCGTCTTAATCAATCC-----TAA      Nt
GAGGTTTTTAAGGAACTTCGTCTTAATCAACCGAAATCAATTA      S1
GAGGTTTTTAAGGAACTTCGTCTTAATCAACCTAAATCAATTA      Ab
GGGATTTTTACGGAA-----TGA              At
GAATTCITTAAGAT-----AAAGAAAGAACTTTGAGTTAA      Zm
GAATTCITTAAGAA-----AAAGAAAGAACTTTGAGTTAA      Os
```

**Rps16**

```
AQPTGTVDILKKAIEVFKELRPNQS          Nt
AQPTETVDILKKAIEVFKELRLNQPKN       S1
AQPTGTVDILKKAIEVFKELRPNQPKN       Ab
AQPTRTAHDISKAGIFTE                  At
AQPTRTVYDILRKAIEFFKD---KERTLS      Zm
AQPTRTVSDILRKAIEFFKE---KERTLS      Os
```

Ab - *Atropa belladonna*  
 At - *Arabidopsis thaliana*  
 Ca - *Cypripedium acaule*  
 Nt - *Nicotiana tabacum*  
 Os - *Oryza sativa*  
 Pc - *Polytrichum commune*  
 Ps - *Pisum sativum*  
 Pt - *Pinus thunbergii*  
 S1 - *Solanum lycopersicum*  
 Va - *Vigna angularis*  
 Zm - *Zea mays*

**Fig. 2.** InDels in coding regions of the tomato plastid genome. Partial sequence alignments for *rrn16* (16S rRNA), *trnS-UGA*, and *rps16* are shown. InDels are indicated by hyphens. The brace below the 16S rRNA alignment marks a duplicated sequence motif in tomato and *Atropa belladonna* which presumably has arisen from replication slippage. Note that InDels in the highly variable 3' end of the *rps16* reading frame cause substantial interspecific variation in the C-terminus of the Rps16 protein as shown in the amino acid sequence alignment. Abbreviations of species names are given at the bottom.

plastomes in order to deduce phylogenetic relationships and evolutionary trends for individual genes and gene classes. Pairwise homology values were first determined by aligning the entire genomes (Table 4). This analysis revealed that tobacco and *Atropa* may be more closely related to each other than tomato is to either of the two other solanaceous species. The analysis of a subset of genes (or even of the entire SSC) would not have been informative enough to deduce phylogenetic relationships among the three closely related species (Table 4), underscoring the

**Table 4.** Pairwise comparison of solanaceous plastid genomes

	Identity (%)		
	<i>Solanum lycopersicum</i> vs. <i>Nicotiana tabacum</i>	<i>Solanum lycopersicum</i> vs. <i>Atropa belladonna</i>	<i>Nicotiana tabacum</i> vs. <i>Atropa belladonna</i>
ptDNA	95.3	95.2	96.4
LSC	94.0	94.2	95.7
SSC	94.6	94.4	94.3
IR	98.1	97.8	98.9

*Note.* Identity values are given as percentage for the total plastid genome, the two single-copy regions, and the IR (gap penalty, 1.10; gap extension penalty, 0.33).

importance of acquiring large sequence data sets to resolve such relationships. This becomes even more evident when individual genes are compared pairwise and grouped in identity classes (Table 5). While the overall picture seems to confirm that *Atropa* and *Nicotiana* are more closely related than tomato with either tobacco or *Atropa* (Table 5), the analysis of individual genes can tell a different story, illustrating the danger of basing phylogenetic conclusions on the analysis of only one or a few plastid genes. The pairwise comparison of individual genes presented in Table 5 also reveals a set of plastid genes which display a relatively high level of interspecific variation and thus, together with intergenic spacer regions (Kress et al. 2005), may be particularly informative to resolve phylogenetic relationships at the species level. Among them are *clpP*, *yef1*, *yef10*, *accD*, *matK*, and *ccsA*, some of which (e.g., *matK* and *yef1*) are often used in phylogenetic analyses. With the exception of *accD*, the pairwise comparison of all of them supports the closer relationship between *Nicotiana* and *Atropa*. Nonetheless, caution is needed in spite of using entire plastid genome sequences for phylogenetic analyses. When, for example, the three InDels in *trnS-UGA*, *rrn16*, and *rps16* are analyzed (Fig. 2), in all three cases, *Atropa belladonna* and *Solanum lycopersicum* show identical insertions and deletions relative to tobacco, which, other than the total genome sequences, would support a closer phylogenetic association of these two solanaceous species (which would be in congruence with existing phylogenies of the Solanaceae [Olmstead et al. 1999]). Thus, even complete plastome sequences can be insufficient to unambiguously resolve phylogenetic relationships among closely related species.

#### *Evolution of Plastid RNA Editing Patterns in Solanaceous Species*

A hallmark of gene expression in higher plant cell organelles is the requirement for an additional RNA processing step referred to as RNA editing. RNA editing in plastids and mitochondria is a posttranscriptional process changing the identity of single

nucleotides in primary transcripts at highly specific sites. In plastids of seed plants, these changes appear to be restricted to cytidine-to-uridine conversions (Hoch et al. 1991; Kudla et al. 1992; for review see, e.g., Bock 2000, 2001), whereas in chloroplasts of hornworts extensive “reverse” editing by U-to-C transitions has been observed (Kugita et al. 2003). With very few exceptions (Hirose et al. 1996; Kudla and Bock 1999), the vast majority of known plastid editing events alters the coding properties of the affected mRNAs and usually results in the restoration of triplets for phylogenetically conserved amino acid residues (Maier et al. 1992a b). Transgenic experiments creating tobacco plants with a noneditable version of a plastid gene have provided direct evidence for the functional importance of RNA editing in chloroplast gene expression (Bock et al. 1994).

Comparative phylogenetic analyses of RNA editing sites in plastid genomes have revealed that many plastid editing sites are poorly conserved interspecifically (Freyer et al. 1997). For example, of the 26 RNA editing sites located in the chloroplast genome of black pine, *Pinus thunbergii*, only one site is also found in the tobacco plastid genome (Wakasugi et al. 1996). Remarkably, even closely related plant species can differ significantly in their editing patterns (Freyer et al. 1995; Schmitz-Linneweber et al. 2001, 2005). While some plastid RNA editing sites are well conserved, at least within certain taxonomic groups, others appear more or less sporadically in largely divergent taxonomic groups (Freyer et al. 1997; Fiebig et al. 2004). Whether the latter can be explained by several independent acquisition events during evolution or, rather, by several independent losses of ancient editing sites that were present in a common ancestor, is largely unknown.

To assess the evolutionary dynamics of RNA editing in plastid genomes in greater detail, the editing patterns in the three sequenced solanaceous plastomes were compared. In previous work, 34 RNA editing sites had been found in the tobacco plastome and 31 sites in *Atropa* (Hirose et al. 1999; Tsudzuki et al. 2001; Schmitz-Linneweber et al. 2002; Sasaki et al. 2003) (Table 6). In order to identify

**Table 5.** Pairwise comparison of plastid genes in Solanaceae and their grouping in identity classes: Genes in the IR are in boldface

Identity (%)	<i>Solanum lycopersicum</i> – <i>Nicotiana tabacum</i>	<i>Solanum lycopersicum</i> – <i>Atropa belladonna</i>	<i>Nicotiana tabacum</i> – <i>Atropa belladonna</i>
100	<i>petG</i> , <i>psaI</i> , <i>psbF</i> , <i>psbI</i> , <i>psbL</i> , <i>psbN</i> , <b><i>rpl23</i></b> , <i>rps7</i> , <b><i>rnm4.5</i></b> , <b><i>rnm5</i></b> , <b><i>trnA-UGC</i></b> , <i>trnC-GCA</i> , <i>trnD-GUC</i> , <i>trnE-UUC</i> , <i>trnF-M-CAU</i> , <i>trnG-GCC</i> , <i>trnG-UCC</i> , <i>trnH-GUG</i> , <b><i>trnI-CAU</i></b> , <b><i>trnI-GAU</i></b> , <i>trnK-UUU</i> , <i>trnM-CAU</i> , <b><i>trnN-GUU</i></b> , <i>trnQ-UUG</i> , <b><i>trnR-ACG</i></b> , <i>trnR-UCU</i> , <i>trnS-GCU</i> , <i>trnT-GGU</i> , <b><i>trnV-GAC</i></b> , <i>trnV-UAC</i> , <i>trnW-CCA</i> , <i>trnY-GUA</i>	<i>petG</i> , <i>petN</i> , <i>psaI</i> , <i>psbF</i> , <i>psbI</i> , <i>psbJ</i> , <i>psbK</i> , <i>psbN</i> , <b><i>rpl23</i></b> , <i>rps19</i> , <b><i>rps7</i></b> , <b><i>rnm4.5</i></b> , <b><i>rnm5</i></b> , <b><i>trnA-UGC</i></b> , <i>trnC-GCA</i> , <i>trnE-UUC</i> , <i>trnF-M-CAU</i> , <i>trnG-GCC</i> , <i>trnG-UCC</i> , <i>trnH-GUG</i> , <b><i>trnI-CAU</i></b> , <i>trnK-UUU</i> , <b><i>trnL-CAA</i></b> , <i>trnL-UAG</i> , <i>trnM-CAU</i> , <b><i>trnN-GUU</i></b> , <i>trnQ-UUG</i> , <b><i>trnR-ACG</i></b> , <i>trnR-UCU</i> , <i>trnS-GCU</i> , <i>trnT-GGU</i> , <b><i>trnV-GAC</i></b> , <i>trnV-UAC</i> , <i>trnW-CCA</i>	<i>petG</i> , <i>petL</i> , <i>psbF</i> , <i>psbI</i> , <i>psbM</i> , <i>psbN</i> , <b><i>rpl23</i></b> , <b><i>rps7</i></b> , <b><i>rnm4.5</i></b> , <b><i>rnm5</i></b> , <b><i>trnA-UGC</i></b> , <i>trnC-GCA</i> , <i>trnE-UUC</i> , <i>trnF-M-CAU</i> , <i>trnG-GCC</i> , <i>trnG-UCC</i> , <i>trnH-GUG</i> , <b><i>trnI-CAU</i></b> , <i>trnK-UUU</i> , <i>trnL-UAA</i> , <i>trnM-CAU</i> , <b><i>trnN-GUU</i></b> , <i>trnQ-UUG</i> , <b><i>trnR-ACG</i></b> , <i>trnR-UCU</i> , <i>trnS-GCU</i> , <i>trnT-GGU</i> , <i>trnV-UAC</i> , <b><i>trnW-CCA</i></b> , <b><i>ycf15</i></b> , <b><i>ycf3</i></b> , <b><i>ycf9</i></b>
99–99.9	<b><i>ndhB</i></b> , <i>ndhC</i> , <i>ndhI</i> , <i>petD</i> , <i>petL</i> , <i>psaA</i> , <i>psaC</i> , <i>psaI</i> , <i>psbA</i> , <i>psbB</i> , <i>psbD</i> , <i>psbE</i> , <i>psbJ</i> , <i>psbM</i> , <i>rbcl</i> , <b><i>rpl2</i></b> , <i>rpl20</i> , <i>rpl33</i> , <i>rpoB</i> , <i>rps12</i> , <i>rps14</i> , <i>rps18</i> , <i>rps19</i> , <i>rps2</i> , <b><i>rnm16</i></b> , <b><i>rnm23</i></b> , <b><i>ycf15</i></b> , <b><i>ycf2</i></b> , <b><i>ycf3</i></b> , <b><i>ycf9</i></b>	<i>atpB</i> , <b><i>ndhB</i></b> , <i>ndhI</i> , <i>petB</i> , <i>petL</i> , <i>psaA</i> , <i>psaB</i> , <i>psbA</i> , <i>psbB</i> , <i>psbD</i> , <i>psbL</i> , <i>psbM</i> , <b><i>rpl2</i></b> , <i>rpl20</i> , <i>rps12</i> , <i>rps14</i> , <i>rps18</i> , <i>rps2</i> , <b><i>rnm16</i></b> , <b><i>rnm23</i></b> , <b><i>ycf15</i></b> , <b><i>ycf9</i></b>	<i>atpB</i> , <i>atpI</i> , <b><i>ndhB</i></b> , <i>ndhJ</i> , <i>petB</i> , <i>petD</i> , <i>psaA</i> , <i>psaB</i> , <i>psaI</i> , <i>psbA</i> , <i>psbB</i> , <i>psbC</i> , <i>psbD</i> , <i>psbJ</i> , <i>psbL</i> , <i>psbT</i> , <i>rbcl</i> , <b><i>rpl16</i></b> , <b><i>rpl2</i></b> , <i>rpl20</i> , <i>rpl33</i> , <i>rpl36</i> , <i>rpoB</i> , <i>rps11</i> , <i>rps12</i> , <i>rps18</i> , <i>rps19</i> , <i>rps2</i> , <i>rps4</i> , <b><i>rnm16</i></b> , <b><i>rnm23</i></b> , <b><i>ycf10</i></b> , <b><i>ycf2</i></b> , <b><i>ycf4</i></b>
98–98.9	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpH</i> , <i>atpI</i> , <i>atpL</i> , <i>ndhA</i> , <i>ndhD</i> , <i>ndhG</i> , <i>ndhH</i> , <i>ndhJ</i> , <i>ndhK</i> , <i>petA</i> , <i>petB</i> , <i>petN</i> , <i>psbB</i> , <i>psbC</i> , <i>psbK</i> , <i>rpl14</i> , <i>rpoA</i> , <i>rpoC1</i> , <i>rpoC2</i> , <i>rps11</i> , <i>rps15</i> , <i>rps4</i> , <i>rps8</i> , <i>trnF-GAA</i> , <b><i>trnL-CAA</i></b> , <i>trnL-UAA</i> , <i>trnL-UAG</i> , <i>trnP-UGG</i> , <i>trnS-GGA</i>	<i>atpA</i> , <i>atpE</i> , <i>atpF</i> , <i>atpH</i> , <i>atpI</i> , <i>ndhC</i> , <i>ndhG</i> , <i>ndhJ</i> , <i>ndhK</i> , <i>petA</i> , <i>petD</i> , <i>psaC</i> , <i>psbC</i> , <i>psbE</i> , <i>rbcl</i> , <i>rpl14</i> , <i>rpl33</i> , <i>rpoA</i> , <i>rpoB</i> , <i>rpoC1</i> , <i>rps11</i> , <i>rps3</i> , <i>rps4</i> , <i>rps8</i> , <i>trnD-GUC</i> , <i>trnF-GAA</i> , <i>trnL-UAA</i> , <i>trnP-UGG</i> , <i>trnS-GGA</i> , <i>trnY-GUA</i> , <b><i>ycf2</i></b> , <b><i>ycf3</i></b> , <b><i>ycf4</i></b>	<i>atpA</i> , <i>atpE</i> , <i>atpF</i> , <i>atpH</i> , <i>matK</i> , <i>ndhC</i> , <i>ndhE</i> , <i>ndhF</i> , <i>ndhG</i> , <i>ndhH</i> , <i>ndhI</i> , <i>ndhK</i> , <i>petA</i> , <i>petN</i> , <i>psaC</i> , <i>psbE</i> , <i>psbH</i> , <i>psbK</i> , <i>rpl14</i> , <i>rpoA</i> , <i>rpoC1</i> , <i>rpoC2</i> , <i>rps14</i> , <i>rps3</i> , <i>rps8</i> , <i>trnD-GUC</i> , <b><i>trnL-CAA</i></b> , <i>trnL-UAG</i> , <i>trnY-GUA</i>
97–97.9	<i>matK</i> , <i>ndhE</i> , <i>ndhF</i> , <i>psbH</i> , <i>psbT</i> , <i>rpl16</i> , <i>rpl22</i> , <i>rps3</i>	<i>accD</i> , <i>ndhA</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhH</i> , <i>psbH</i> , <i>rps16</i> , <i>rpl16</i> , <i>rpl22</i> , <i>rpl32</i> , <i>rps15</i> , <b><i>trnI-GAU</i></b>	<i>ccsA</i> , <i>clpP</i> , <i>ndhA</i> , <i>ndhD</i> , <i>rpl22</i> , <i>rpl32</i> , <i>rps15</i> , <b><i>trnI-GAU</i></b> , <i>trnP-UGG</i>
96–96.9	<i>ccsA</i> , <i>rpl32</i> , <i>rpl36</i> , <i>trnS-UGA</i> , <b><i>ycf4</i></b>	<i>accD</i> , <i>trnS-UGA</i>	<i>accD</i> , <i>trnS-UGA</i>
95–95.9	<i>accD</i>	<i>rps16</i> , <b><i>ycf1</i></b>	<i>rps16</i> , <b><i>ycf1</i></b>
94–94.9	<i>rps16</i>	<b><i>ycf1</i></b>	<b><i>ycf1</i></b>
93–93.9	<b><i>ycf1</i></b>	<b><i>ycf10</i></b>	<b><i>ycf10</i></b>
90–90.9	<b><i>ycf10</i></b>		
86–86.9	<i>sprA</i>	<i>clpP</i>	<i>sprA</i>
85–85.9	<i>clpP</i>		
84–84.9			



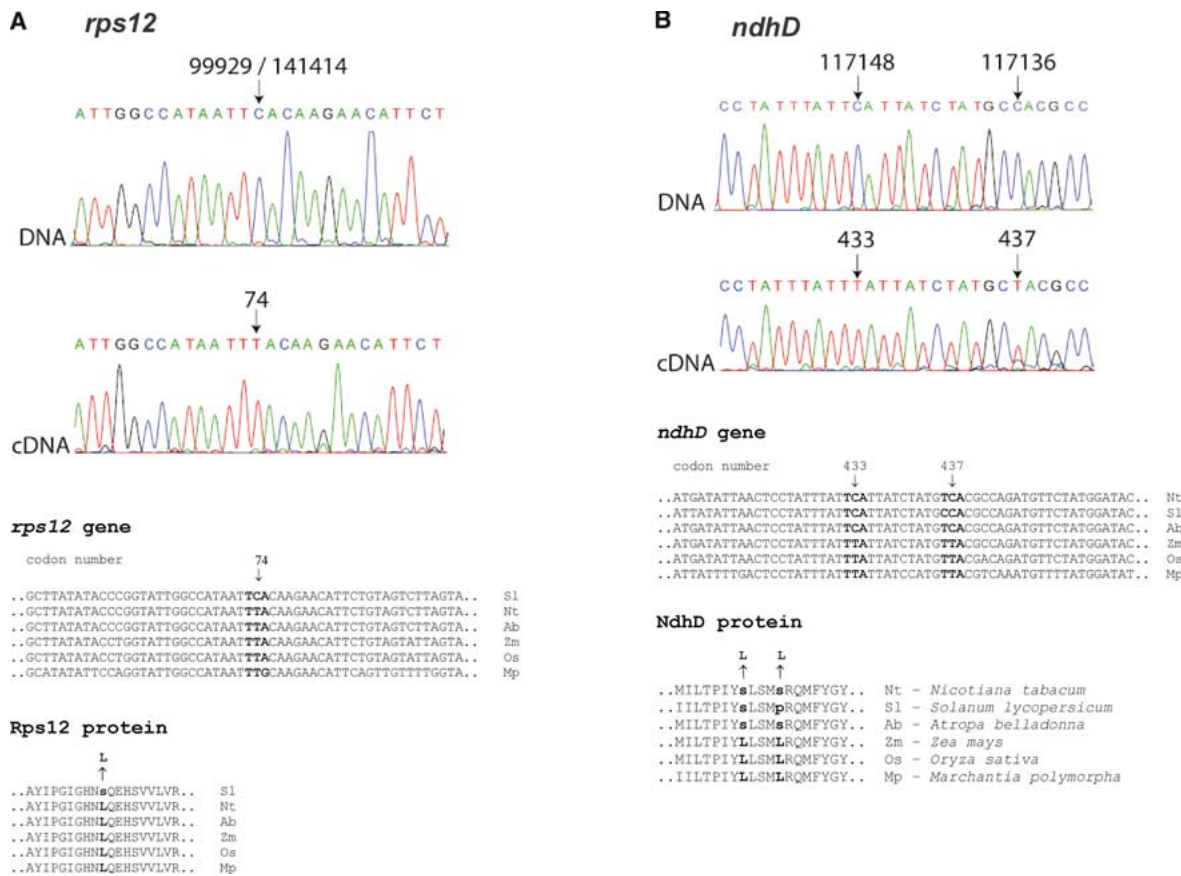
**Table 6.** Comparison of RNA editing patterns in solanaceous plastid genomes

Gene	Site No.	Codon No.		Codon conversion	<i>Nicotiana tabacum</i>	<i>Atropa belladonna</i>	<i>Solanum lycopersicum</i>
<i>atpA</i>	1	264	Pro	(cCc) → Leu (cUc)	+	T	T
	2	265	Ser	(ucC) → Ser (ucU)	+	+	-
<i>atpF</i>		31	Pro	(cCa) → Leu (cUa)	+	+	⊕
<i>ndhA</i>	1	114	Ser	(uCa) → Leu (uUa)	+	+	⊕
	2	189	Ser	(uCa) → Leu (uUa)	T	+	T
	3	358	Ser	(uCc) → Phe (uUc)	+	+	⊕
<i>ndhB</i>	1	50	Ser	(uCa) → Leu (uUa)	+	+	+
	2	156	Pro	(cCa) → Leu (cUa)	+	+	+
	3	196	His	(Cau) → Tyr (Uau)	+	+	+
	4	204	Ser	(uCa) → Leu (uUa)	+	+	+
	5	246	Pro	(cCa) → Leu (cUa)	+	+	+
	6	249	Ser	(uCu) → Phe (uUu)	+	+	+
	7	277	Ser	(uCa) → Leu (uUa)	+	+	+
	8	279	Ser	(uCa) → Leu (uUa)	+	+	+
<i>ndhD</i>	9	494	Pro	(cCa) → Leu (cUa)	+	+	+
	1	1	Thr	(aCg) → Met (aUg)	+	+	+
	2	128	Ser	(uCa) → Leu (uUa)	+	+	T
	3	200	Ser	(uCa) → Leu (uUa)	+	T	+
	4	225	Ser	(uCg) → Leu (uUg)	+	T	+
	5	293	Ser	(uCa) → Leu (uUa)	T	+	+
	6	433	Ser	(uCa) → Leu (uUa)	±	⊕	±
	7	437	Pro	(cCa) → Leu (cUa)	±	⊕	±
	7	437	Ser	(uCa) → Leu (uUa)	±	⊕	±
	<i>ndhF</i>		97	Ser	(uCa) → Leu (uUa)	+	+
<i>ndhG</i>		17	Ser	(uCg) → Leu (uUg)	+	⊕	+
<i>petB</i>		204	Pro	(cCa) → Leu (cUa)	+	+	⊕
<i>psbE</i>		72	Pro	(Cca) → Ser (Uca)	+	T	⊕
<i>psbL</i>		1	Thr	(aCg) → Met (aUg)	+	+	⊕
<i>rpl20</i>		103	Ser	(uCa) → Leu (uUa)	+	+	⊕
<i>rpoA</i>		277	Ser	(uCa) → Leu (uUa)	+	+	⊕
<i>rpoC1</i>		21	Ser	(uCa) → Leu (uUa)	+	+	⊕
<i>rpoC2</i>		1248	Ser	(uCa) → Leu (uUa)	+	+	⊕
<i>rpoB</i>	1	113	Ser	(uCu) → Phe (uUu)	+	+	+
	2	158	Ser	(uCa) → Leu (uUa)	+	+	+
	3	184	Ser	(uCa) → Leu (uUa)	+	+	+
	4	667	Ser	(uCu) → Phe (uUu)	+	+	⊕
	5	809	Ser	(uCa) → Leu (uUa)	T	+	+
<i>rps2</i>	1	45	Thr	(aCa) → Ile (aUa)	+	⊕	⊕
	2	83	Ser	(uCa) → Leu (uUa)	+	+	⊕
<i>rps12</i>		74	Ser	(uCa) → Leu (uUa)	T	T	±
<i>rps14</i>	1	27	Ser	(uCa) → Leu (uUa)	+	+	⊕
	2	50	Pro	(cCa) → Leu (cUa)	+	T	T

Note. T, T at the DNA level (no editing); +, experimentally determined editing sites; ⊕, editing inferred from editing site conservation in sister species; ±, newly identified RNA editing sites; -, no editing (although genomically encoded C present).

possible tomato-specific sites of C-to-U RNA editing, all protein-coding genes in the tomato plastome were conceptually translated and the resulting amino acid sequences compared with the corresponding sequence from tobacco, which, however, were obtained by translating the edited mRNA sequences. All mismatches were evaluated with respect to the possibility that a C-to-U conversion in tomato potentially could restore a codon for the amino acid residue present in the corresponding position in the tobacco sequence. This led to the identification of altogether nine candidate sites (data not shown). Two of these candidate sites (*ndhD* codon 293 and *rpoB* codon 809) were

previously identified as RNA editing sites in *Atropa belladonna* (Schmitz-Linneweber et al. 2002) and analysis of the corresponding cDNA sequences from tomato confirmed C-to-U RNA editing at these sites also in tomato (Table 6 and data not shown). The remaining seven potential editing sites were also analyzed experimentally by directly sequencing amplified tomato cDNA samples. Whereas in six cases no evidence for RNA editing was detected, a new editing site was discovered in the *rps12* gene (Fig. 3A). The editing event changes a genomic serine codon into a conserved leucine codon by a C-to-U change in second codon position (Fig. 3A, Table 6).



**Fig. 3.** Discovery of novel RNA editing sites. **A** Identification of a tomato-specific RNA editing site in the *rps12* transcript. DNA and corresponding cDNA sequences are shown. The editing position is indicated by a vertical arrow and numbered in the DNA sequence with the nucleotide numbers in the plastid genome (two numbers because the sequence is part of the IR) and in the cDNA sequence with the codon numbers in the *rps12* reading frame. Partial *rps12* nucleotide and amino acid sequences from selected plant species are also shown, with the editing position marked by vertical arrows. Note that the editing event changes a genomic serine codon (TCA) to a leucine codon (UUA) in tomato. In all other species, a leucine codon is already present at the DNA level, and hence, no editing occurs. **B** Identification of two novel RNA editing sites in the *ndhD* transcripts of solanaceous plastids. DNA

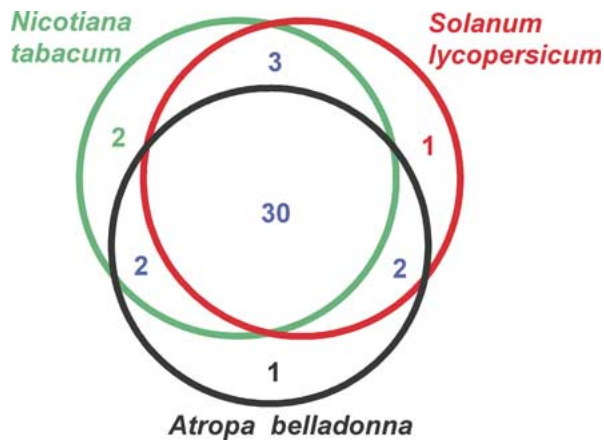
and corresponding cDNA sequences from tomato are shown. The two editing positions are indicated by vertical arrows and numbered in the DNA sequence with the nucleotide numbers in the plastid genome and in the cDNA sequence with the codon numbers in the *ndhD* reading frame. In the bottom panel, partial *ndhD* nucleotide and amino acid sequences from selected plant species are depicted. The editing positions are marked by vertical arrows. Note that the editing event in codon 437 changes a proline codon (CCA) to a leucine codon (CUA) in tomato but alters a UCA serine codon to a UUA leucine codon in the two other solanaceous species. Amino acid residues derived from the unedited sequence are shown in lowercase letters in the amino acid sequence alignments.

The two other solanaceous species, tobacco and *Atropa*, contain the TTA leucine codon at the DNA level and, thus, lack RNA editing at this site. As none of the other candidate sites identified in our computer analyses could be confirmed experimentally, the newly discovered site in *rps12* is the only tomato-specific RNA editing site (Table 6, Fig. 4).

Most of the editing sites present in tobacco and *Atropa* are also conserved in tomato (Table 6, Fig. 4), suggesting that these sites also undergo C-to-U editing in tomato. About two-thirds of the sites were experimentally analyzed in tomato, and as expected, editing was confirmed (Table 6). A notable exception is site 2 in the *atpA* gene. This site is unique in that it is the only site in tobacco (and *Atropa*) plastids where editing is silent, that is, does not

change the coding properties of the affected triplet. This is because editing occurs in the third position of the codon and both the unedited codon UCC and the edited codon UCU specify the amino acid serine. Editing at this site is only partial, with a large fraction of the *atpA* mRNA population remaining unedited (Hirose et al. 1996). Interestingly, analysis of amplified *atpA* cDNAs from tomato revealed no evidence of RNA editing, indicating that the editing at this site in tobacco and *Atropa* is functionally irrelevant.

To obtain a genome-wide picture of plastid RNA editing, we also were interested in discovering additional editing sites shared by the three solanaceous species which might have escaped detection in the earlier work with tobacco and *Atropa* (Hirose et al. 1999; Tsudzuki et al. 2001; Schmitz-Linneweber et al.



**Fig. 4.** Overview of RNA editing site evolution in Solanaceae. All sites present in a given species are enclosed in the respective color-coded circle. Shared sites between two or more species are shown in blue. Unique sites are shown in the color of the respective species. The majority of sites (30) is shared between all three species, whereas only a few sites are unique or shared by only two species.

2002; Sasaki et al. 2003; Tillich et al. 2005) (Table 6). Potential RNA editing sites can be most easily identified by comparing the conceptual translations of plastid genes from higher plant species with those from the liverwort *Marchantia polymorpha*, a species known to lack RNA editing in plastids (Bock 2001). In these analyses, we discovered two closely spaced codons in the *ndhD* gene where conserved leucine residues could be restored by C-to-U editing in all solanaceous species (Fig. 3B). Although the *ndhD* genes from tobacco and *Atropa* had been analyzed in several earlier studies on plastid RNA editing (Neckermann et al. 1994; Hirose et al. 1999; Tsudzuki et al. 2001; Schmitz-Linneweber et al. 2002; Sasaki et al. 2003), the location of the sites within an otherwise highly conserved protein domain (Fig. 3B) prompted us to test for the editing of these two candidate sites experimentally. To this end, we analyzed *ndhD* cDNA sequences from both tobacco and tomato. Tobacco was included because, for the second candidate editing site, the potentially edited codons are different between the solanaceous species: Whereas tobacco (and *Atropa*) has a genomic TCA serine codon, tomato has a CCA proline codon in this position (Fig. 3B). Interestingly, in both cases, C-to-U conversion in the second codon position would restore the conserved leucine residue, because both UUA and CUA triplets specify leucine. Comparison of *ndhD* DNA and cDNA sequences revealed that the two candidate sites undergo RNA editing in both tomato (Fig. 3B) and tobacco (data not shown; Fig. 3B, Table 6). As the two serine codons are also present in *Atropa* (Fig. 3B), there is little doubt that they also undergo editing in the third solanaceous species, although we have not verified this experimentally. In order to confirm that the seven identified

*ndhD* editing sites (Table 6) represent the full set, we amplified and sequenced the complete *ndhD* cDNA from tomato. No other sites of RNA editing were found, which is in line with the absence of additional candidate sites from our bioinformatics analysis.

The current picture of RNA editing in solanaceous species is summarized in Table 6 and Fig. 4. Although it cannot be formally excluded that there are additional sites present in the three genomes which thus far have escaped detection, the reliable prediction of candidate sites by rather simple bioinformatics analyses makes it unlikely that many more sites remain to be discovered. It is noteworthy that the vast majority of sites (31 out of 41) creates leucine codons (Table 6), confirming codon biases of RNA editing observed earlier (Maier et al. 1995; Bock 2000, 2001). Figure 4 illustrates that the total number of sites is highly similar in the three species (37 in tobacco, 35 in *Atropa*, 36 in tomato). Most of the sites (30) are conserved in all three species, suggesting that they were already present in their common ancestor. Only four sites are unique in that they occur only in one of the three species (two in tobacco, one in *Atropa*, one in tomato). Few sites are absent from only one of the three species: tobacco and tomato share three sites that are absent from the *Atropa* plastome, tomato and *Atropa* have two sites that are missing in tobacco, and, finally, tobacco and *Atropa* also have two sites in common that are not present in tomato (Fig. 4, Table 6).

The finding that each species and each pair of species show a few specific sites (Fig. 4) allows some speculation about the evolutionary origin of these sites. The most parsimonious explanation may be that the sites shared by two of the three species were present in the common ancestor of the Solanaceae and were then lost in one of the three species by a genomic C-to-T mutation. The opposite scenario (acquisition in two species) may seem less likely but, at present, cannot be entirely excluded.

With three plastid genomes being sequenced and the editing patterns determined, the Solanaceae family currently offers the most comprehensive data set about the evolutionary flexibility and dynamics of plastid RNA editing among closely related species. In view of the functional importance of mRNA processing by editing (Bock et al. 1994) and the recently discovered role for RNA editing in nucleocytoplasmic incompatibility phenomena in solanaceous plants (Schmitz-Linneweber et al. 2005), the significance of this dynamic of RNA site evolution cannot be underestimated.

#### *Assessing Intraspecific Plastome Variation in Tomato*

The cultivated tomato, *Solanum lycopersicum*, has been subjected to extensive breeding programs in

both Latin America and Europe. The tomato was introduced from Mexico into Europe in the sixteenth century but was initially regarded as poisonous, and the fruits were not widely consumed until the nineteenth century (Simpson and Ogorzaly 2001). To determine whether there were differences in the plastome sequences of European and Latin American tomato cultivars, we compared the plastome sequence of IPA-6, a commercial tomato cultivar bred in Brazil, with that of Ailsa Craig, one of the oldest cultivars bred and grown in Europe. Ailsa Craig was bred in a nursery in Girvan, Ayrshire, Scotland, overlooking the island of Ailsa Craig, and was released in 1910 (Lisman 1961). It is suited to a northern European climate and is unlikely to have been directly involved in the origins of IPA-6. Surprisingly, the nucleotide sequences of the IPA-6 and Ailsa Craig plastomes were absolutely identical and did not show a single nucleotide difference. This indicates a remarkable conservation of sequence over a period of at least several hundred years of separation of the two tomato cultivars and suggests that the plastomes of modern tomato varieties display very little, if any, sequence variation. An earlier comparison of large sequence stretches of the closely related plastid genomes of *Nicotiana sylvestris* and its allopolyploid descendant *Nicotiana tabacum* had revealed a single nucleotide substitution in 4656 bp of plastid DNA sequence (Clarkson et al. 2004), pointing to a very low degree of sequence variation also in tobacco.

**Acknowledgments.** We thank the MPI-MP Green Team for plant care and cultivation, Andy Phillips for gifts of cloned fragments of the Ailsa Craig plastome, Graham Seymour for details on the origin of Ailsa Craig, and Helaine Carrer (University of São Paulo, Brazil) for obtaining information on the origin of IPA-6. This research was supported by a collaborative grant from the European Union (FP6 Plastomics Project LSHG-CT-2003-503238) to J.C.G. and R.B. and by the Max Planck Society.

## References

- Abdallah F, Salamini F, Leister D (2000) A prediction of the size and evolutionary origin of the proteome of chloroplasts of Arabidopsis. *Trends Plant Sci* 5:141–142
- Bock R (2000) Sense from nonsense: how the genetic information of chloroplasts is altered by RNA editing. *Biochimie* 82:549–557
- Bock R (2001) RNA editing in plant mitochondria and chloroplasts. In: Bass B (ed) *Frontiers in molecular biology: RNA editing*. Oxford University Press, New York, pp 38–60
- Bock R, Kössel H, Maliga P (1994) Introduction of a heterologous editing site into the tobacco plastid genome: the lack of RNA editing leads to a mutant phenotype. *EMBO J* 13:4623–4628
- Boudreau E, Takahashi Y, Lemieux C, Turmel M, Rochaix J-D (1997) The chloroplast *ycf3* and *ycf4* open reading frames of *Chlamydomonas reinhardtii* are required for the accumulation of the photosystem I complex. *EMBO J* 16:6095–6104
- Burrows PA, Sazanov LA, Svab Z, Maliga P, Nixon PJ (1998) Identification of a functional respiratory complex in chloroplasts through analysis of tobacco mutants containing disrupted plastid *ndh* genes. *EMBO J* 17:868–876
- Clarkson JJ, Knapp S, Garcia VF, Olmstead RG, Leitch AR, Chase MW (2004) Phylogenetic relationships in Nicotiana (Solanaceae) inferred from multiple plastid DNA regions. *Mol Phylogenet Evol* 33:75–90
- Drescher A, Ruf S, Calsa T Jr, Carrer H, Bock R (2000) The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. *Plant J* 22:97–104
- Fiebig A, Stegemann S, Bock R (2004) Rapid evolution of RNA editing sites in a small non-essential plastid gene. *Nucleic Acids Res* 32:3615–3622
- Freyer R, Lopez C, Maier RM, Martin M, Sabater B, Kössel H (1995) Editing of the chloroplast *ndhB* encoded transcript shows divergence between closely related members of the grass family (Poaceae). *Plant Mol Biol* 29:679–684
- Freyer R, Kiefer-Meyer M-C, Kössel H (1997) Occurrence of plastid RNA editing in all major lineages of land plants. *Proc Natl Acad Sci USA* 94:6285–6290
- Gray JC, Hird SM, Dyer TA (1990) Nucleotide sequence of a wheat chloroplast gene encoding the proteolytic subunit of an ATP-dependent protease. *Plant Mol Biol* 15:947–950
- Hager M, Biehler K, Illerhaus J, Ruf S, Bock R (1999) Targeted inactivation of the smallest plastid genome-encoded open reading frame reveals a novel and essential subunit of the cytochrome *b6f* complex. *EMBO J* 18:5834–5842
- Hajdukiewicz PTJ, Allison LA, Maliga P (1997) The two RNA polymerases encoded by the nuclear and the plastid compartments transcribe distinct groups of genes in tobacco plastids. *EMBO J* 16:4041–4048
- Hedtke B, Börner T, Weihe A (1997) Mitochondrial and chloroplast phage-type RNA polymerases in Arabidopsis. *Science* 277:809–811
- Hess WR, Börner T (1999) Organellar RNA polymerases of higher plants. *Int Rev Cytol* 190:1–59
- Hess WR, Hoch B, Zeltz P, Hübschmann T, Kössel H, Börner T (1994) Inefficient *rpl2* splicing in barley mutants with ribosome-deficient plastids. *Plant Cell* 6:1455–1465
- Hippler M, Bock R (2004) Chloroplast proteomics. *Prog Bot* 65:90–105
- Hirose T, Fan H, Suzuki JY, Wakasugi T, Tsudzuki T, Kössel H, Sugiura M (1996) Occurrence of silent RNA editing in chloroplasts: its species specificity and the influence of environmental and developmental conditions. *Plant Mol Biol* 30:667–672
- Hirose T, Kusumegi T, Tsudzuki T, Sugiura M (1999) RNA editing sites in tobacco chloroplast transcripts: editing as a possible regulator of chloroplast RNA polymerase activity. *Mol Gen Genet* 262:462–467
- Hoch B, Maier RM, Appel K, Igloi GL, Kössel H (1991) Editing of a chloroplast mRNA by creation of an initiation codon. *Nature* 353:178–180
- Jenkins BD, Kulhanek DJ, Barkan A (1997) Nuclear mutations that block group II RNA splicing in maize chloroplasts reveal several intron classes with distinct requirements for splicing factors. *Plant Cell* 9:283–296
- Joet T, Cournac L, Horvath EM, Medgyesy P, Peltier G (2001) Increased sensitivity of photosynthesis to antimycin A induced by inactivation of the chloroplast *ndhB* gene. Evidence for a participation of the NADH-dehydrogenase complex to cyclic electron flow around photosystem I. *Plant Physiol* 125:1919–1929
- Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. *Proc Natl Acad Sci USA* 102:8369–8374
- Kudla J, Bock R (1999) RNA editing in an untranslated region of the Ginkgo chloroplast genome. *Gene* 234:81–86

- Kudla J, Igloi GL, Metzclaff M, Hagemann R, Kössel H (1992) RNA editing in tobacco chloroplasts leads to the formation of a translatable psbL mRNA by a C to U substitution within the initiation codon. *EMBO J* 11:1099–1103
- Kugita M, Yamamoto Y, Fujikawa T, Matsumoto T, Yoshinaga K (2003) RNA editing in hornwort chloroplasts makes more than half the genes functional. *Nucleic Acids Res* 31:2417–2423
- Liere K, Link G (1995) RNA-binding activity of the matK protein encoded by the chloroplast trnK intron from mustard (*Sinapis alba* L.). *Nucleic Acids Res* 23:917–921
- Lisman TA (1961) The romance of Ailsa Craig. *Grower* 56:582–583
- Maier RM, Hoch B, Zeltz P, Kössel H (1992a) Internal editing of the maize chloroplast ndhA transcript restores codons for conserved amino acids. *Plant Cell* 4:609–616
- Maier RM, Neckermann K, Hoch B, Akhmedov NB, Kössel H (1992b) Identification of editing positions in the ndhB transcript from maize chloroplasts reveals sequence similarities between editing sites of chloroplasts and plant mitochondria. *Nucleic Acids Res* 20:6189–6194
- Maier RM, Neckermann K, Igloi GL, Kössel H (1995) Complete sequence of the maize chloroplast genome: gene content, hot-spots of divergence and fine tuning of genetic information by transcript editing. *J Mol Biol* 251:614–628
- Majeran W, Wollman F-A, Vallon O (2000) Evidence for a role of ClpP in the degradation of the chloroplast cytochrome b6f complex. *Plant Cell* 12:137–149
- Martin W, Herrmann RG (1998) Gene transfer from organelles to the nucleus: How much, what happens, and why? *Plant Physiol* 118:9–17
- Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, Leister D, Stoebe B, Hasegawa M, Penny D (2002) Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc Natl Acad Sci USA* 99:12246–12251
- Maurizi MR, Clark WP, Kim S-H, Gottesman S (1990) ClpP represents a unique family of serine proteases. *J Biol Chem* 265:12546–12552
- Mohr G, Perlman PS, Lambowitz AM (1993) Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function. *Nucleic Acids Res* 21:4991–4997
- Munekage Y, Hashimoto M, Miyake C, Tomizawa K-i, Endo T, Tasaka M, Shikanai T (2004) Cyclic electron flow around photosystem I is essential for photosynthesis. *Nature* 429:579–582
- Neckermann K, Zeltz P, Igloi GL, Kössel H, Maier RM (1994) The role of RNA editing in conservation of start codons in chloroplast genomes. *Gene* 146:177–182
- Olmstead RG, Sweere JA, Spangler RE, Bohs L, Palmer JD (1999) Phylogeny and provisional classification of the Solanaceae based on chloroplast DNA. In: Nee M, Symon DE, Lester RN, Jessop JP (eds) *Solanaceae IV: Advances in biology and utilization*. Royal Botanic Gardens, Kew, UK, pp 111–137
- Orsat B, Monfort A, Chatellard P, Stutz E (1992) Mapping and sequencing of an actively transcribed *Euglena gracilis* chloroplast gene (ccsA) homologous to the *Arabidopsis thaliana* nuclear gene cs (ch-42). *FEBS Lett* 303:181–184
- Phillips AL (1985) Restriction map and clone bank of tomato plastid DNA. *Curr Genet* 10:147–152
- Race H, Herrmann RG, Martin W (1999) Why have organelles retained genomes? *Trends Genet* 15:364–370
- Rolland N, Dorne A-J, Amoroso G, Sültemeyer DF, Joyard J, Rochaix J-D (1997) Disruption of the plastid ycf10 open reading frame affects uptake of inorganic carbon in the chloroplast of *Chlamydomonas*. *EMBO J* 16:6713–6726
- Ruf S, Kössel H, Bock R (1997) Targeted inactivation of a tobacco intron-containing open reading frame reveals a novel chloroplast-encoded photosystem I-related gene. *J Cell Biol* 139:95–102
- Ruf S, Biehler K, Bock R (2000) A small chloroplast-encoded protein as a novel architectural component of the light-harvesting antenna. *J Cell Biol* 149:369–377
- Rujan T, Martin W (2001) How many genes in Arabidopsis come from cyanobacteria? An estimate from 386 protein phylogenies. *Trends Genet* 17:113–121
- Sasaki T, Yukawa Y, Miyamoto T, Obokata J, Sugiura M (2003) Identification of RNA editing sites in chloroplast transcripts from the maternal and paternal progenitors of tobacco (*Nicotiana tabacum*): Comparative analysis shows the involvement of distinct trans-factors for ndhB editing. *Mol Biol Evol* 20:1028–1035
- Sasaki Y, Hakamada K, Suama Y, Nagano Y, Furusawa I, Matsuno R (1993a) Chloroplast-encoded protein as a subunit of acetyl-CoA carboxylase in pea plant. *J Biol Chem* 268:25118–25123
- Sasaki Y, Sekiguchi K, Nagano Y, Matsuno R (1993b) Chloroplast envelope protein encoded by chloroplast genome. *FEBS Lett* 316:93–98
- Sasaki Y, Konishi T, Nagano Y (1995) The compartmentation of acetyl-coenzyme A carboxylase in plants. *Plant Physiol* 108:445–449
- Schmitz-Linneweber C, Tillich M, Herrmann RG, Maier RM (2001) Heterologous, splicing-dependent RNA editing in chloroplasts: allotetraploidy provides trans-factors. *EMBO J* 20:4874–4883
- Schmitz-Linneweber C, Regel R, Du TG, Hupfer H, Herrmann RG, Maier RM (2002) The plastid chromosome of *Atropa belladonna* and its comparison with that of *Nicotiana tabacum*: the role of RNA editing in generating divergence in the process of plant speciation. *Mol Biol Evol* 19:1602–1612
- Schmitz-Linneweber C, Kushnir S, Babiychuk E, Poltnigg P, Herrmann RG, Maier RM (2005) Pigment deficiency in nightshade/tobacco hybrids is caused by the failure to edit the plastid ATPase  $\alpha$ -subunit mRNA. *Plant Cell* 17:1815–1828
- Shanklin J, DeWitt ND, Flanagan JM (1995) The stroma of higher plant plastids contain CplP and CplC, functional homologs of *Escherichia coli* ClpP and ClpA: an archetypal two-component ATP-dependent protease. *Plant Cell* 7:1713–1722
- Shikanai T, Endo T, Hashimoto T, Yamada Y, Asada K, Yokota A (1998) Directed disruption of the tobacco *ndhB* gene impairs cyclic electron flow around photosystem I. *Proc Natl Acad Sci USA* 95:9705–9709
- Shimada H, Sugiura M (1991) Fine structural features of the chloroplast genome: comparison of the sequenced chloroplast genomes. *Nucleic Acids Res* 19:983–995
- Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J* 5:2043–2049
- Simpson BB, Ogorzaly MC (2001) *Economic botany*, 3rd ed. McGraw-Hill, New York
- Sugita M, Svab Z, Maliga P, Sugiura M (1997) Targeted deletion of sprA from the tobacco plastid genome indicates that the encoded small DNA is not essential for pre-16S rRNA maturation in plastids. *Mol Gen Genet* 257:23–27
- Sugiura M (1989) The chloroplast chromosomes in land plants. *Annu Rev Cell Biol* 5:51–70
- Sugiura M (1992) The chloroplast genome. *Plant Mol Biol* 19:149–168
- Swiatek M, Kuras R, Sokolenko A, Higgs D, Olive J, Cinque G, Müller B, Eichacker LA, Stern DB, Bassi R, Herrmann

- RG, Wollman F-A (2001) The chloroplast gene *ycf9* encodes a photosystem II (PSII) core subunit, *psbZ*, that participates in PSII supramolecular architecture. *Plant Cell* 13:1347–1367
- Tillich M, Funk HT, Schmitz-Linneweber C, Poltnigg P, Sabater B, Martin M, Maier RM (2005) Editing of plastid RNA in *Arabidopsis thaliana* ecotypes. *Plant J* 43:708–715
- Timmis JN, Ayliffe MA, Huang CY, Martin W (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nature Rev Genet* 5:123–136
- Tsudzuki T, Wakasugi T, Sugiura M (2001) Comparative analysis of RNA editing sites in higher plant chloroplasts. *J Mol Evol* 53:327–332
- Vera A, Sugiura M (1994) A novel RNA gene in the tobacco plastid genome: its possible role in the maturation of 16S rRNA. *EMBO J* 13:2211–2217
- Wakasugi T, Hirose T, Horihata M, Tsudzuki T, Kössel H, Sugiura M (1996) Creation of a novel protein-coding region at the RNA level in black pine chloroplasts: The pattern of RNA editing in the gymnosperm chloroplast is different from that in angiosperms. *Proc Natl Acad Sci USA* 93:8766–8770
- Wakasugi T, Sugita M, Tsudzuki T, Sugiura M (1998) Updated gene map of tobacco chloroplast DNA. *Plant Mol Biol Rep* 16:231–241
- Wakasugi T, Tsudzuki T, Sugiura M (2001) The genomics of land plant chloroplasts: gene content and alteration of genomic information by RNA editing. *Photosynthesis Res* 70:107–118
- Wolfe KH, Li W-H, Sharp PM (1987) Rates of nucleotide substitutions vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci USA* 84:9054–9058
- Xie Z, Merchant S (1996) The plastid-encoded *ccsA* gene is required for heme attachment to chloroplast c-type cytochromes. *J Biol Chem* 271:4632–4639
- Xie Z, Culler D, Dreyfuss BW, Kuras R, Wollman F-A, Girard-Bascou J, Merchant S (1998) Genetic analysis of chloroplast c-type cytochrome assembly in *Chlamydomonas reinhardtii*: one chloroplast locus and at least four nuclear loci are required for heme attachment. *Genetics* 148:681–692