

Update section

Short communication

Ninety extra nucleotide in *ndhF* gene of tobacco chloroplast DNA: a summary of revisions to the 1986 genome sequence

Richard G. Olmstead¹, Jennifer A. Sweere^{1,3} and Kenneth H. Wolfe²

¹Department of E.P.O. Biology, University of Colorado, Boulder, CO 80309, USA; ²Department of Genetics, University of Dublin, Trinity College, Dublin 2, Ireland; ³Present address: Laboratory for Molecular Systematics, National Museum of Natural History, Smithsonian Institution, Washington, DC 20560, USA

Received 19 May 1993; accepted 8 June 1993

Key words: tobacco chloroplast genome, *ndhF*, cpDNA

Abstract

Corrections to the published sequence of the tobacco chloroplast gene *ndhF* are presented, including a 90 bp *Alu* I restriction enzyme fragment internal to the gene that was apparently missed during the original sequencing effort. A summary of the corrections to the published tobacco chloroplast DNA that have come to light since its original publication is included.

In the course of sequencing the chloroplast *ndhF* gene from 17 species of the family Solanaceae for phylogenetic purposes (R.G. Olmstead and J.A. Sweere, in preparation) we discovered that all species examined contain a 90 bp (30 amino acid) insert near the 3' end of the gene, relative to the published sequence from tobacco [1]. This prompted us to resequence almost the entire *Nicotiana tabacum ndhF* gene (between coordinates 112180 and 114175 [1]), using independently isolated DNA. We find that the 90 bp insert is also present in tobacco cpDNA (Table 1). Examination of the new sequence suggests that Shinozaki *et al.* [1] overlooked a 90 bp *Alu* I restriction fragment during shotgun sequencing, and instead reported a single *Alu* I site at position 112465–112468. We also found a second minor

error in the published tobacco sequence, at position 112211–112214 (Table 1). Curiously, our revised sequence at this point is identical to that originally published in 1984 by Sugita *et al.* [2] but later changed in [1]. Our revised tobacco *ndhF* sequence has been submitted to GenBank.

The publication of the tobacco chloroplast genome sequence in 1986 was a landmark in cpDNA research. Since then, as homologous genes from other species were sequenced, a number of other possible errors in the tobacco sequence became apparent and most of these regions have now been resequenced. These are summarized in Table 1, together with three positions at which errors are suspected (because they result in frameshifts as compared to other species) but have not been examined by rese-

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number L14953.

Table 1. Revisions to the tobacco chloroplast genome sequence as published by Shinozaki *et al.* [1]. The original sequence is written in lower-case letters. Upper-case letters show bases that were omitted in the original sequence and should be inserted; bases that should be deleted from the original sequence are marked by brackets.

Gene	Sequence ¹	Size	Reference
<i>Revisions identified by resequencing</i>			
<i>rpoC2</i>	19741 cagaaaaatc C gaactaaata	19760	[3]
<i>rpoC2</i>	19781 ttggttact [c] gagaggtag	19800	[3]
<i>rpoC2</i>	19881 gacct [cga] GCAC ggccttctaata	19890	[3]
<i>rpoC2</i>	20301 ata C t C ctccgg	20310	[3]
<i>rpoC2</i>	20381 ttccc c aatt	20390	[3]
<i>rpoC2</i>	20661 cggtgc C acaa	20670	[3]
<i>rpoC2</i>	20761 a C tc C ttta C gggct C cgtagcaa	20780	[3]
<i>rpoC1</i>	23611 tttacga [t] gg	23620	[3]
<i>rpoC1</i>	24081 ccaaaa A ttctt T cacaaaata	24100	[3]
<i>rpoC1</i>	24101 atccatctttt T ccggtttgt	24120	[3]
<i>infA</i>	82451 aa [g] gcgattc	82460	[4]
<i>ndhF</i>	112201 aaccccgtta G tca [g] tcccat	112220	This paper
<i>ndhF</i>	112461 aataAGCTATACTGACTGAAAAGAAC TGCATCCTTTAAAAAATTCATTCCAA TCCATCGAATTATTCGACTTTTGAT GCAAAAAGATTTATAGATGGagctaa	112470	This paper
<i>psaC</i>	119221 tacatcggtggg G caggctcg	119240	[5]
<i>ndhA</i>	123321 tattt [t] cata	123330	[6]
ORF1901	126791 tgtttttatt T tgactaacat	126810	[7]
ORF1901	127791 ttgtaaaatac [g] atatgctt	127810	[7]
<i>Suspected errors that have not been resequenced</i>			
ORF2280	90581/151948 ccgaag G atga	90590/151939	[8, 9]
<i>ndhA</i>	121921 aaa [at] caaac	121930	[10]
<i>ndhA</i>	121981 tctgcttct [t] ggtaaatcaa	122000	[10]

¹ The numbers indicate the coordinates of the first and last bases shown. To avoid confusion, we have not adjusted the numbering system. ORF2280 is located in the inverted repeat, so the sequence shown has two sets of coordinates. All other sequences shown are single-copy.

quencing. The cumulative effect of the changes listed in Table 1 is the insertion of 119 nucleotides (one of which is in the inverted repeat) and the deletion of 12 nucleotides, making the 1993 estimate of the genome size 155 952 bp.

References

- Shinozaki K, Ohme M, Tanaka T, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chungwongse 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: The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J* 5: 2041–2049 (1986).
- Sugita M, Kato A, Shimada H, Sugiura M: Sequence analysis of the junctions between a large inverted repeat and single-copy regions in tobacco chloroplast DNA. *Mol Gen Genet* 194: 200–205 (1984).
- Shimada H, Fukuta M, Ishikawa M, Sugiura M: Rice chloroplast RNA polymerase genes: the absence of an intron in *rpoC1* and the presence of an extra sequence in *rpoC2*. *Mol Gen Genet* 221: 395–402 (1990).
- Wolfe KH, Morden CW, Ems SC, Palmer JD: Rapid evolution of the plastid translational apparatus in a non-photosynthetic plant: loss or accelerated sequence evolution of tRNA and ribosomal protein genes. *J Mol Evol* 35: 304–317 (1992).
- Hayashida N, Matsubayashi T, Shinozaki K, Sugiura M, Inoue K, Hiyama T: The gene for the 9 kd polypeptide, a possible apoprotein for the iron-sulfur centers A and B

- of the photosystem I complex, in tobacco chloroplast DNA. *Curr Genet* 12: 247–250 (1987).
6. Matsubayashi T, Wakasugi T, Shinozaki K, Yamaguchi-Shinozaki K, Zaita N, Hidaka T, Meng BY, Ohto C, Tanaka M, Kato A, Maruyama T, Sugiura M: Six chloroplast genes (*ndhA-F*) homologous to human mitochondrial genes encoding components of the respiratory chain NADH dehydrogenase are actively expressed: determination of the splice sites in *ndhA* and *ndhB* pre-mRNAs. *Mol Gen Genet* 210: 365–393 (1987).
 7. Wolfe KH, Morden CW, Palmer JD: Small single-copy region of plastid DNA in the non-photosynthetic angiosperm *Epifagus virginiana* contains only two genes: differences among dicots, monocots and bryophytes in gene organization at a non-bioenergetic locus. *J Mol Biol* 223: 95–104 (1992).
 8. Shimada H, Sugiura M: Fine structural features of the chloroplast genome: comparison of the sequenced chloroplast genomes. *Nucl Acids Res* 19: 983–995 (1991).
 9. Zhou DX, Massenet Q, Quigley F, Marion MJ, Moneger F, Huber P, Mache R: Characterisation of a large inversion in the spinach chloroplast genome relative to *Marchantia*: a possible transposon-mediated origin. *Curr Genet* 13: 433–439 (1988).
 10. Wolfe KH, Sharp PM: Identification of functional open reading frames in chloroplast genomes. *Gene* 66: 215–222 (1988).