DNA Research **10**, 59–65 (2003)

Complete Nucleotide Sequence of the Chloroplast Genome from a Leptosporangiate Fern, *Adiantum capillus-veneris* L.

Paul G. Wolf,[1,*] Carol A. Rowe,[1] Robert B. Sinclair,[2] and Mitsuyasu Hasebe[3]

Department of Biology, Utah State University, 5305 Old Main Hill, Logan, UT 84322, USA,[1] *The Jackson Laboratory, 600 Main Street, Bar Harbor, Maine 04609, USA,*[2] *and National Institute for Basic Biology, 38 Nishigounaka, Myo-daiji-cho, Okazaki 444-8585, Japan*[3]

(Received 3 December 2002; revised 10 February 2003)

Abstract

We determined the complete nucleotide sequence of the chloroplast genome of the leptosporangiate fern, *Adiantum capillus-veneris* L. (Pteridaceae). The circular genome is 150,568 bp, with a large single-copy region (LSC) of 82,282 bp, a small-single copy region (SSC) of 21,392 bp and inverted repeats (IR) of 23,447 bp each. We compared the sequence to other published chloroplast genomes to infer the location of putative genes. When the IR is considered only once, we assigned 118 genes, of which 85 encode proteins, 29 encode tRNAs and 4 encode rRNAs. Four protein-coding genes, all four rRNA genes and six tRNA genes occur in the IR. Most (57) putative protein-coding genes appear to start with an ATG codon, but we also detected five other possible start codons, some of which suggest tRNA editing. We also found 26 apparent stop codons in 18 putative genes, also suggestive of RNA editing. We found all but one of the tRNA genes necessary to encode the complete repertoire required for translation. The missing *trnK* gene appears to have been disrupted by a large inversion, relative to other published chloroplast genomes. We detected several structural rearrangements that may provide useful information for phylogenetic studies.
Key words: annotation; genome structure; inversion

1. Introduction

In the past decade biologists have witnessed a significant improvement in understanding phylogenetic relationships in most groups of organisms. This progress has resulted from an increase in the use of nucleotide sequence data from one or a few genes, and also from improvements in methods of phylogenetic analysis. However, in some ways, the use of gene sequence data has had diminishing returns especially at deeper (ancient) phylogenetic levels, such that many aspects of relationships among major clades remain unresolved. For example, it is not clear which group is the most basal clade of land plants.[1–3] The problems associated with inferring deep phylogeny using nucleotide data probably relate to the weak historical signal contained in data with a limited number of character states (four) such that homoplasy is likely. These limitations can be overcome to some degree by increasing the number of genes or by incorporating other types of data. One emerging approach is to use complete genome sequences. Such data can be used for traditional phylogenetic analyses of aligned nucleotide regions. However, the data can also be used to infer changes in genome structure, such as inversions, translocations, nucleotide insertions or deletions, gene losses, or expansion and contraction of repeat units. These are complex characters with a large number of potential states and therefore less prone to homoplasy than the four-state nucleotide characters. In fact, structural rearrangements in chloroplast genomes have been used since the 1980s to infer phylogeny.[4–6] For example, a 60-kb inversion in bryophytes and lycopods relative to other vascular plants is strong evidence that lycopods are the basal lineage of vascular plants.[7] These early studies were done by hybridizing labeled heterologous probes to restriction-digested DNA. Current technology makes it easier to determine complete genome sequences, which leads to simpler data storage and comparison, and eliminates the need for additional cross-taxon probing. Furthermore, structural analysis can be made at any scale, and can therefore detect small rearrangements that would be missed by a large heterologous probe. In addition to phylogenetic studies, complete genome sequence data can be used for many applications, including studies of cellular function, gene function, post-transcriptional modification, and genetic manipulation. This paper is part of a larger effort to sequence and analyze complete

Communicated by Satoshi Tabata
* To whom correspondence should be addressed. Tel. +435-797-4034, Fax. +435-797-1575, E-mail: wolf@biology.usu.edu

organellar genomes from all major clades of green plants. Here we describe and compare the first complete nucleotide sequence of a leptosporangiate fern chloroplast genome.

2. Materials and Methods

Adiantum capillus-veneris L. is in the fern family Pteridaceae within a large clade of recently derived leptosporangiate families; a clade that includes the majority of fern species.[8–10] The chloroplast genome of this species has been previously cloned and mapped,[11,12] providing a starting point for our study. *Pst* I fragments of *Adiantum capillus-veneris* chloroplast DNA were cloned in pUC18 and introduced into competent JM101 cells.[13] Inserts were gel-purified from the larger *Pst* I clones (more than 10 kb), digested with *Sau*3AI and subcloned back into pUC18 to provide additional start sequences for primer walking. All inserts (original *Pst* I and a selection of *Sau*3AI) were then end-sequenced and several rounds of primer-walking were performed to obtain single-stranded coverage of each *Pst* I fragment. Reverse primers were designed to allow sequence reads for the second strand, and additional pairs of primers from near the ends of *Pst* I fragments were used to amplify the intervening regions (bridges) from undigested genomic DNA by PCR. These PCR products were then sequenced, and all *Pst* I fragments and bridges were assembled to obtain the complete chloroplast genome sequence, taking into account the large inverted repeat region. All sequencing was performed with the PRISM™ Big Dye™ 2.0 Terminator Cycle Sequencing Kit (PE Applied Biosystems, Foster City, CA, USA), and reactions were run on an ABI 377 DNA sequencer. We used Sequencher 3.1 (Gene Codes Corporation, Ann Arbor, MI, USA) for sequence analysis and assembly.

To locate putative genes in the *Adiantum* chloroplast genome, we extracted all protein sequences from the published chloroplast genomes of *Psilotum nudum* (NC 003386) and *Nicotiana tabacum* (NC 001879), and used PSI-BLAST 2.2.3[14] to compare each amino acid sequence to all six reading frames in *Adiantum*. We used BLASTN 2.2.3[14] to locate ribosomal RNA genes, and tRNAscan-SE v.1.1[15] to locate putative transfer RNA genes. Any regions in the *Adiantum* sequence larger than 300 bp that did not contain features after this process were compared with the current release of GenBank (26 Aug 2002) using BLASTX[14] to identify additional features. We then compared the gene order in *Adiantum* against its putatively closest relative for which a complete sequence is published (*Psilotum*), as well as to *Nicotiana*. For any discrepancies, we made additional comparisons to *Marchantia polymorpha* (NC 001319), *Zea mays* (NC 001666), and *Pinus thunbergii* (NC 001631). We considered regions to be homologous to annotated regions from other genomes if the e-value of the BLAST hit was less than 10^{-4}.

3. Results and Discussion

The complete chloroplast genome sequence of *Adiantum capillus-veneris* is deposited in GenBank under accession no. AY178864. The genome assembled according to the map of Hasebe and Iwatsuki[12] with few exceptions. We detected one additional (57 bp) *Pst* I fragment between fragments 6 and 12 that were described as adjacent by Hasebe and Iwatsuki.[12] We also found no evidence of *Pst* I fragments 24 (500 bp) and 25 (200 bp) reported by Hasebe and Iwatsuki.[12] This is not too surprising since the latter study used agarose gel electrophoresis which has much lower resolving power to detect fragments.

3.1. Genome structure

In the fully assembled genome sequence, the large single-copy region (LSC) is 82,282 bp, the small single copy region (SSC) is 21,392 bp and the inverted repeats (IR) are 23,447 bp for a total of 150,568 bp (Fig. 1). The overall structure of the genome is typical of vascular plants[16] with especially good synteny in the LSC. However, several rearrangements, not detected by restriction fragment analyses,[12,17] were detected here. For example, the genes *psbM* and *petN* (in the LSC) are included in an $\sim$ 300 bp inversion unique to *Adiantum*, among complete genomes sequenced to date. It should be possible to design PCR primers to screen other ferns to determine if this inversion is unique to a particular clade. Furthermore, *Adiantum* shares an inversion of about 3300 bp (including *psbD*, *psbC*, *psbZ*) with *Psilotum*, relative to *Marchantia*, *Pinus*, *Nicotiana*, and *Zea*. This inversion may unite all members of the moniliform clade,[18] which includes the horsetails, ferns, and Psilotaceae. Gene order data will be required from *Equisetum*, *Osmunda*, and eusporangiate ferns to test this hypothesis. Both the $\sim$ 300 bp inversion and the $\sim$ 3300 bp inversion are in the same region as the end-point of a large $\sim$ 60 kb inversion in all vascular plants except the lycopods.[5] This suggests the possibility of a hotspot for inversions in this region. Stein et al.[4] suggest a similar hotspot associated with *psbA*.

We detected a large rearrangement (incorporating most of the inverted repeat region) relative to the other sequenced chloroplast genomes. This structure has been identified previously using mapping studies[4,17] and appears to be derived in most ferns, but not in the basal lineage represented by *Osmunda*.[4] Relative mapping of *Adiantum* with *Nicotiana*[17] used gene order to interpret this as an expansion of the LSC in *Nicotiana*, and two overlapping inversions in the ferns. This results in expansion of the IR in most ferns. Examination of sequence data confirms the gene order interpretation of Hasebe and

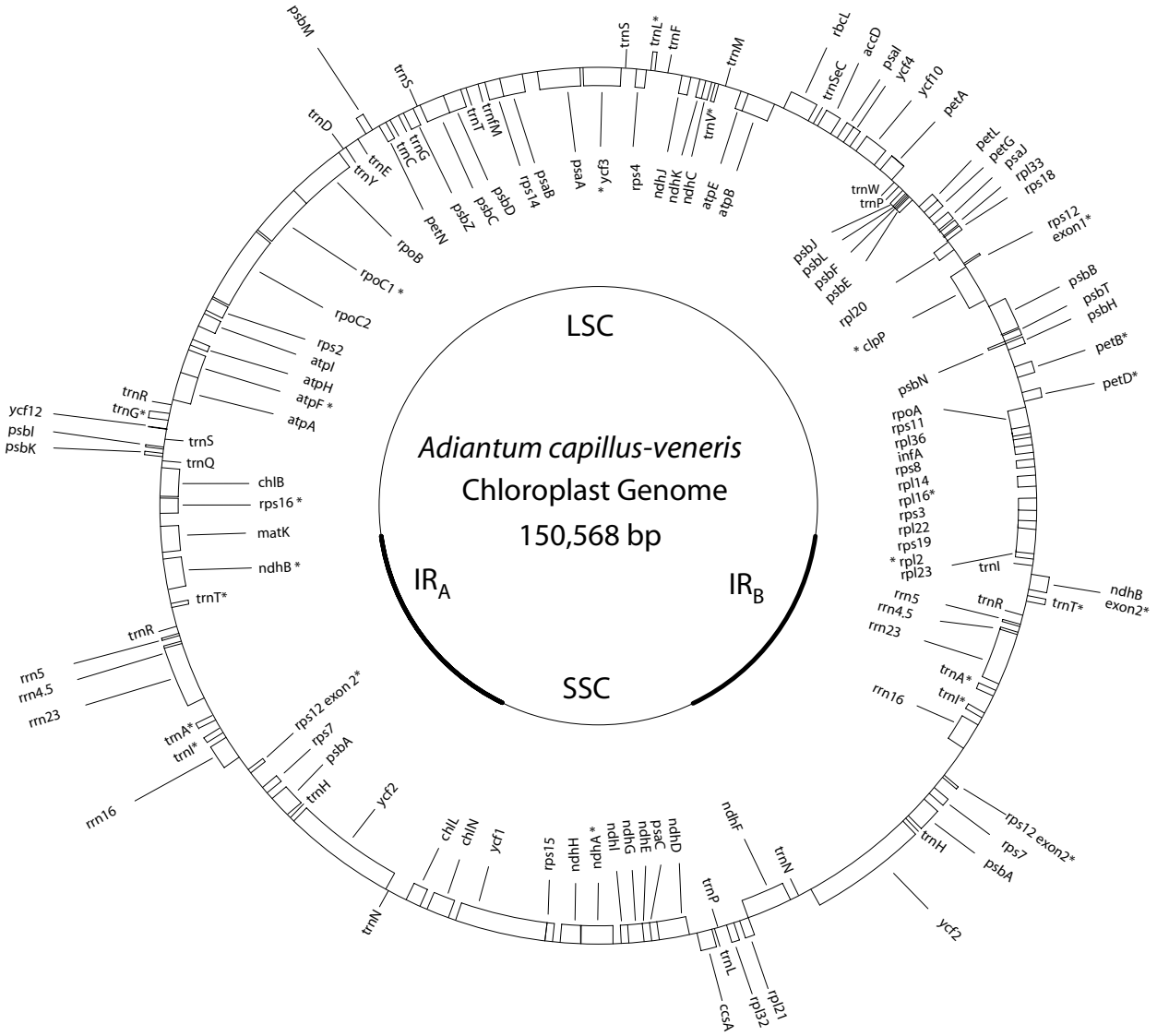

Figure 1. Gene map of *Adiantum capillus-veneris* chloroplast genome. Thick lines represent the inverted repeats (IR; 23,447 bp each) which separate the large single-copy region (LSC; 82,282 bp) from the small single-copy region (SSC; 21,392 bp). Genes on the outside of the map are transcribed clockwise, those on the inside are transcribed counterclockwise. Asterisks denote genes with introns. Nucleotide positions are numbered starting at the boundary of IR_A and LSC, with position 1 in the intron of *ndhB*.

Iwatsuki[17] with one minor difference that does not affect the "two inversion" hypothesis, above. The *trnK* gene was located by Hasebe and Iwatsuki[17] using a *Nicotiana* probe that was probably *matK* (based on the nucleotide position in *Nicotiana*), although not annotated as such at the time. In *Adiantum*, the result of the rearrangement is that the first exon of *ndhB* is on the complementary strand at the beginning of the LSC (close to the junction of LSC-IR_A), the intron spans the LSC-IR_A junction, and exon 2 continues in the inverted repeat. There is, therefore, an orphan exon 2 at the end of the IR_B (Fig. 1). In the other genomes compared here, both *ndhB* exons are in the inverted repeat. Additional gene order data from other basal lineages of ferns (such as Gleicheniaceae and Hymenophyllaceae) could be used to test the "two inversion" hypothesis, by finding a lineage with only one of the two inversions.

Table 1 lists all genes that we detected in the chloroplast genome of *Adiantum*. The genes *rps16* and *chlL* are absent from *Psilotum*[19,20] but are present in *Adiantum*. However, we were unable to locate homologs of *Psilotum orf83* or *orf119* in *Adiantum*. These may be spurious open reading frames (i.e., they are not transcribed) or sequence divergence has lowered the similarity to a level that they are not detectable by BLAST. The gene *psaM* is annotated in *Psilotum*, *Chaetosphaeridium* (NC 004115) and *Pinus*, but not in *Nicotiana*, *Zea*, and *Marchantia*. However, our BLAST analyses located a candidate gene

Table 1. List of genes annotated for *Adiantum capillus-veneris* chloroplast DNA. Asterisk denotes an intron-containing gene.

Gene class				
Ribosomal RNAs	*rrn16* x2	*rrn23* x2	*rrn5* x2	*rrn4.5* x2
Transfer RNAs	*trnQ-UUG*	*trnG-GCC*	*trnM-CAU*	*trnV-UAC**
	trnS-GCU	*trnS-UGA*	*trnSeC-UCA*	*trnC-GCA*
	*trnG-UCC**	*trnT-GGU*	*trnW-CCA*	*trnY-GUA*
	trnR-UCU	*trnfM-CAU*	*trnP-UGG*	*trnP-GGG*
	trnD-GUC	*trnS-GGA*	*trnI-CAU*	*trnL-UAG*
	trnA-TGC x2*	*trnN-GUU x2*	*trnT-UGU* x2*	*trnI-GAU* x2*
	trnE-UUC	*trnF-GAA*	*trnR-ACG x2*	*trnL-CAA**
	trnH-GUG x2			
Photosystem I	*psaA*	*psaB*	*psaC*	*psaI*
	psaJ			
Photosystem II	*psbA*	*psbB*	*psbC*	*psbD*
	psbE	*psbF*	*psbH*	*psbI*
	psbJ	*psbK*	*psbL*	*psbM*
	psbN	*psbT*	*psbZ*	
Cytochrome	*petA*	*PetB**	*PetD**	*petG*
	petL	*petN*		
ATP synthase	*atpA*	*atpB*	*atpE*	*atpF**
	atpH	*atpI*		
Rubisco	*rbcL*			
Chlorophyll biosynthesis	*chlB*	*chlN*	*chlL*	
NADH dehydrogenase	*ndhA**	*ndhB**	*ndhC*	*ndhD*
	ndhE	*ndhF*	*ndhG*	*ndhH*
	ndhI	*ndhJ*	*ndhK*	
Ribosomal proteins	*rpl2**	*rpl14*	*rpl16**	*rpl20*
	rpl21	*rpl22*	*rpl23*	*rpl32*
	rpl33	*rpl36*	*rps2*	*rps3*
	rps4	*rps7 x2*	*rps8*	*rps11*
	rps12 x2	*rps14*	*rps15*	*rps16**
	rps18	*rps19*		
RNA polymerase	*rpoA*	*rpoB*	*rpoC1**	*rpoC2*
Miscellaneous proteins	*infA*	*ccsA*	*matK*	*clpP**
	accD			
Hypothetical proteins	*ycf1*	*ycf2 x2*	*ycf3**	*ycf10*
	ycf12			

in *Marchantia* but not in *Nicotiana*, *Zea*, or in *Adiantum*. The presence of the gene in *Pinus* suggests that it may have been lost independently in fern and angiosperm clades. Based on gene content, the *Adiantum* chloroplast genome is typical for that of other vascular plants, lacking only *psaM* and *trnK* (see below).

3.2. Codon usage and transfer RNA genes

Start positions of most *Adiantum* chloroplast protein-coding genes could be inferred from comparisons to previously annotated genes. We examined 83 putative genes for which the start position was based on alignments to *Nicotiana* and *Psilotum*. We inferred that 57 start at AUG, 16 at ACG, 5 at AUU, 3 at AUC, one at AUA, and one at GUG. We were unable to locate canonical starts nearby upstream or downstream from these putative start positions. RNA editing of Thr (ACG) to Met (AUG) can restore start codons in chloroplast encoded genes[21–23] and GUG starts have been reported in other chloroplast genomes.[24] It should be emphasized that inferring start positions based only on genome sequences is merely hypothetical.[24] Future studies include plans to examine cDNA sequences to test inferences of our annotation, including start codons and RNA editing.

We detected 29 putative tRNA genes (Table 2), of which six have second copies in the inverted repeat. Most chloroplast genomes encode all necessary ribosomal RNAs, transfer RNAs, and ribosomal proteins for fully functional translational machinery.[24] We examined this potential in the *Adiantum* chloroplast genome by scoring 26,129 codons in 85 putative genes to provide a codon usage table that is listed with corresponding tRNA genes inferred (Table 3). Most amino acids can be incorporated by allowing nucleotide "wobble" among anticodons of tRNA.[25] However, this would leave three codons without corresponding tRNAs: the heavily used UUA (requiring tRNA-Leu) and both codons requiring tRNA-Lys, for which we found no candidate gene. The UUA codon could be read by the tRNA-Leu (CAA), but only if the anticodon is C → U edited, or the normal constraints on wobble are loosened (Table 3). Our failure to detect any tRNA-Lys is more puzzling. This tRNA is encoded by *trnK* and is usually found in the LSC of most land plant chloroplast genomes. The gene is usually split by a large intron that contains *matK*, which encodes a maturase required for splicing introns throughout the chloroplast genome. However, it appears that one of the large inversions in ferns occurred in the intron. Either *trnK* is trans-spliced and we have failed to locate the exons, or *trnK* is not functional. If the latter, then it is not clear how lysine would be incorporated during protein synthesis. One possibility is that post-transcriptional editing could alter an anticodon of a different tRNA[21] to create tRNA-Lys, or to incorporate arginine in place of lysine. These two amino acids have similar properties and are often (though not always) interchangeable. Alternatively, tRNA-Lys could be supplied from nuclear origin, as has been proposed to explain the lack of *trnR*-ACG in the chloroplast genome of *Lotus japonicus*.[26] Import of nuclear-encoded tRNAs into mitochondria has been documented in several systems[27–29] and import into chloroplasts has been speculated for a

Table 2. List of putative tRNA genes located in *Adiantum* chloroplast genome.

	First exon (or all)		Second exon		gene-anticodon	tRNA-amino acid
	start	end	start	end		
1	6235	6164			*trnQ-UUG*	tRNA-Gln
2	7290	7203			*trnS-GCU*	tRNA-Ser
3	8361	8383	9321	9368	*trnG-UCC*	tRNA-Gly
4	9578	9649			*trnR-UCU*	tRNA-Arg
5	27214	27287			*trnD-GUC*	tRNA-Asp
6	27483	27402			*trnY-GUA*	tRNA-Tyr
7	27717	27645			*trnE-UUC*	tRNA-Glu
8	29765	29694			*trnC-GCA*	tRNA-Cys
9	30146	30076			*trnG-GCC*	tRNA-Gly
10	30979	31067			*trnS-UGA*	tRNA-Ser
11	35016	34945			*trnT-GGU*	tRNA-Thr
12	35274	35201			*trnfM-CAU*	tRNA-Met-f
13	42897	42983			*trnS-GGA*	tRNA-ser
14	44775	44808	45389	45435	*trnL-CAA*	tRNA-Leu
15	45736	45808			*trnF-GAA*	tRNA-Phe
16	48841	48808	48177	48138	*trnV-UAC*	tRNA-Val
17	49115	49187			*trnM-CAU*	tRNA-Met-elong
18	53449	53522			*trnSeC-UCA*	tRNA-selenocysteine
19	61937	61864			*trnW-CCA*	tRNA-Trp
20	62197	62124			*trnP-UGG*	tRNA-Pro
21	82234	82161			*trnI-CAU*	tRNA-Ile
22	83676	83709	84218	84260	*trnT-UGU*	tRNA-Thr
23	85246	85173			*trnR-ACG*	tRNA-Arg
24	89988	89952	89147	89112	*trnA-TGC*	tRNA-Ala
25	91123	91088	90091	90056	*trnI-GAU*	tRNA-Ile
26	97859	97932			*trnH-GUG*	tRNA-His
27	105232	105171			*trnN-GUU*	tRNA-Asn
28	109673	109600			*trnP-GGG*	tRNA-Pro
29	109898	109977			*trnL-UAG*	tRNA-Leu
30	127619	127681			*trnN-GUU*	tRNA-Asn
31	134992	134919			*trnH-GUG*	tRNA-His
32	141728	141763	142760	142795	*trnI-GAU*	tRNA-Ile
33	142863	142899	143704	143739	*trnA-UGC*	tRNA-Ala
34	147605	147678			*trnR-ACG*	tRNA-Arg
35	149175	149142	148633	148591	*trnT-UGU*	tRNA-Thr

Table 3. Total numbers of each codon detected in 85 putative genes in the *Adiantum* chloroplast genome, indicated with tRNAs for which genes have been identified in the chloroplast genome.

UUU	Phe	920		UCU	Ser	526		UAU	Tyr	558	
UUC	Phe	490	trnF-GAA	UCC	Ser	365	trnS-GGA	UAC	Tyr	346	trnY-GUA
UUA	Leu	728		UCA	Ser	434	trnS-UGA	UAA	Ter	57	
UUG	Leu	670	trnL-CAA	UCG	Ser	277		UAG	Ter	23	
CUU	Leu	448		CCU	Pro	332		CAU	His	328	
CUC	Leu	276		CCC	Pro	312	trnP-GGG	CAC	His	214	trnH-GUG (x2)
CUA	Leu	438	trnL-UAG	CCA	Pro	305	trnP-UGG	CAA	Gln	539	trnQ-UUG
CUG	Leu	250		CCG	Pro	180		CAG	Gln	257	
AUU	Ile	971		ACU	Thr	481		AAU	Asn	800	
AUC	Ile	432	trnI-GAU (x2)	ACC	Thr	291	trnT-GGU	AAC	Asn	381	trnN-GUU (x2)
AUA	Ile	521	trnI-CAU	ACA	Thr	376	trnT-UGU	AAA	Lys	919	
AUG	Met	535	trnfM-CAU trnM-CAU	ACG	Thr	208		AAG	Lys	381	
GUU	Val	511		GCU	Ala	623		GAU	Asp	777	
GUC	Val	249		GCC	Ala	358		GAC	Asp	296	trnD-GUC
GUA	Val	541	trnV-UAC	GCA	Ala	460	trnA-UGC (x2)	GAA	Glu	951	trnE-UUC
GUG	Val	236		GCG	Ala	215		GAG	Glu	405	

UGU	Cys	178									
UGC	Cys	146	trnC-GCA								
UGA	Ter	34									
UGG	Trp	431	trnW-CCA								
CGU	Arg	335	trnR-ACG (x2)								
CGC	Arg	181									
CGA	Arg	303									
CGG	Arg	175									
AGU	Ser	436									
AGC	Ser	170	trnS-GCU								
AGA	Arg	459	trnR-UCU								
AGG	Arg	231									
GGU	Gly	599									
GGC	Gly	267	trnG-GCC								
GGA	Gly	660	trnG-UCC								
GGG	Gly	333									

nonphotosynthetic angiosperm.[30] However, we have not found any published evidence for this phenomenon in photosynthetic plants.

We detected an additional tRNA gene that encodes a potential tRNA-selenocysteine (Fig. 2), which has been reported in plants[31] but, as far as we know, not in chloroplasts. Although *trnSeC* was detected by tRNAscan-SE, the sequence does have some problems that brings its function into question. Position 32 is typically a pyrimidine in tRNA molecules but is an 'A' in *Adiantum trnSeC*, and the pairing at the base of the TΨC loop is poor (Fig. 2). Thus, it is possible that *trnSeC* represents a pseudogene of some previously functional tRNA gene. The potential *trnSeC* that we found has a UCA anticodon, corresponding to one of the common stop codons. We detected 26 stop codons within the otherwise open reading frames of 18 putative *Adiantum* genes: 16 of these are UAA, 8 are UGA and 2 are UAG. Some of these could be RNA editing sites,[21] but it is also possible that the UGA codons are read-through by tRNA-SeC.[32] Clearly, transcript sequences are required to distinguish these phenomena.

Although previous DNA hybridization studies identified many evolutionarily significant genome rearrangements, the enhanced resolution of a complete plastid genome sequence can enable researchers to detect more markers, as well as learn more about gene and genome evolution at the nucleotide level. Complete genome sequences also provide public domain data that can be used in ways unanticipated by the original researchers. We believe that as the database of complete plastid genome sequences increases, so will its utility, as patterns and rearrangements can be inferred in more detail.

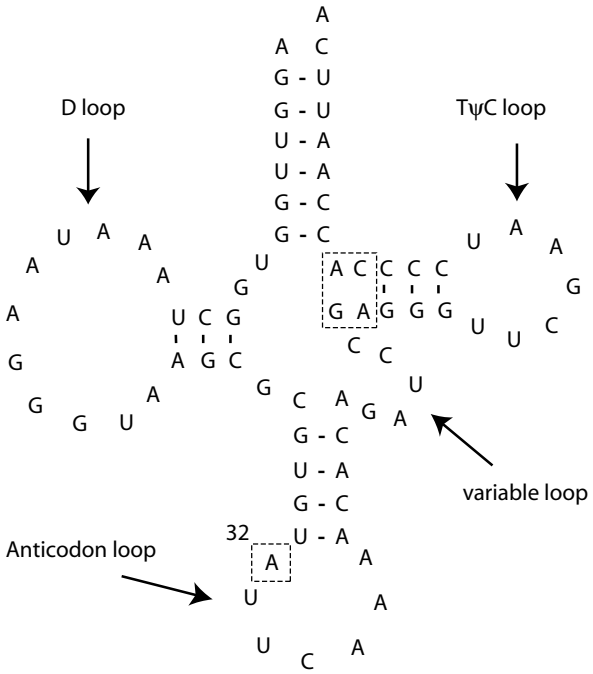

Figure 2. Secondary structure of putative tRNA-selenocysteine encoded in the *Adiantum capillus-veneris* chloroplast genome. Note the 'A' rather than a pyrimidine at position 32 and the poor base pairing in the stem of the TΨC loop.

Acknowledgements: This research was supported by funds from a Utah State Community-University Research Initiative and by National Science Foundation grant DEB-0228432. This is paper no. 1 of the Green Plant Tree of Life Project. Thanks to Gayle Knapp, Bruce Copeland, and Tomoaki Nishiyama for help with data analysis, to Ashwin Manne for drawing figure, and

to Scylla Rottee for critical reading of the manuscript.

References

1. Duff, R. J. and Nickrent, D. L. 1999, Phylogenetic relationships of land plants using mitochondrial small-subunit rDNA sequences, *Amer. J. Bot.*, **86**, 372–386.
2. Nickrent, D. L., Parkinson, C. L., Palmer, J. D., and Duff, R. J. 2000, Multigene phylogeny of land plants with special reference to bryophytes and the earliest land plants, *Mol. Biol. Evol.*, **17**, 1885–1895.
3. Qiu, Y. L. and Lee, J. 2000, Transition to a land flora: A molecular phylogenetic perspective, *J. Phycol.*, **36**, 799–802.
4. Stein, D. B., Conant, D. S., Ahearn, M. E. et al. 1992, Structural rearrangements of the chloroplast genome provide an important phylogenetic link in ferns, *Proc. Natl. Acad. Sci. USA*, **89**, 1856–1860.
5. Raubeson, L. A. 1992, A rare chloroplast-DNA structural mutation is shared by all conifers, *Biochem. Syst. Ecol.*, **20**, 17–24.
6. Jansen, R. K. and Palmer, J. D. 1987, A Chloroplast DNA inversion marks an ancient evolutionary split in the sunflower family (Asteraceae), *Proc. Natl. Acad. Sci. USA*, **84**, 5818–5822.
7. Raubeson, L. A. and Jansen, R. K. 1992, Chloroplast DNA evidence on the ancient evolutionary split in vascular land plants, *Science*, **255**, 1697–2699.
8. Hasebe, M., Wolf, P. G., Pryer, K. M. et al. 1995, Fern phylogeny based on *rbcL* nucleotide sequences, *Amer. Fern J.*, **85**, 134–181.
9. Hasebe, M., Ito, M., Kofuji, R., Ueda, K., and Iwatsuki, K. 1993, Phylogenetic relationships of ferns deduced from *rbcL* gene sequence, *J. Mol. Evol.*, **37**, 476–482.
10. Wolf, P. G., Soltis, P. S., and Soltis, D. S. 1994, Phylogenetic relationships of dennstaedtioid ferns: evidence from *rbcL* sequences, *Mol. Phyl. Evol.*, **3**, 383–392.
11. Hasebe, M. and Iwatsuki, K. 1990, *Adiantum capillus-veneris* chloroplast DNA clone bank: as useful heterologous probes in the systematics of the leptosporangiate ferns, *Amer. Fern J.*, **80**, 20–25.
12. Hasebe, M. and Iwatsuki, K. 1990, Chloroplast DNA from *Adiantum capillus-veneris* L., a fern species (Adiantaceae); clone bank, physical map and unusual gene localization in comparison with angiosperm chloroplast DNA, *Curr. Genet.*, **17**, 359–364.
13. Sambrook, E., Fritch, F., and Maniatis, T. 1989, Molecular cloning: a laboratory manual, 2nd edition, Cold Spring Harbor Laboratory Press, Cold Spring harbor, New York.
14. Altschul, S. F., Madden, T. L., Schaffer, A. A. et al. 1997, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.*, **25**, 3389–3402.
15. Lowe, T. M. and Eddy, S. R. 1997, tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence, *Nucleic Acids Res.*, **25**, 955–964.
16. Palmer, J. D. and Stein, D. B. 1986, Conservation of chloroplast genome structure among vascular plants., *Curr. Genet.*, **10**, 823–833.
17. Hasebe, M. and Iwatsuki, K. 1992, Gene localization on the chloroplast DNA of the maiden hair fern: *Adiantum capillis-veneris*, *Bot. Mag. Tokyo*, **105**, 413–419.
18. Pryer, K. M., Schneider, H., Smith, A. R. et al. 2001, Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants, *Nature*, **409**, 618–622.
19. Wakasuki, T., Nishikawa, A., Yamada, K., and Sugiura, M. 1998, A complete nucleotide sequence of the plastid genome from a fern, *Psilotum nudum*, *Endocyt. Cell Res.*, **13 (Suppl.)**, 147.
20. Burke, D. H., Raubeson, L. A., Alberti, M. et al. 1993, The *chlL* (*frxC*) gene — phylogenetic distribution in vascular plants and DNA sequence from *Polystichum acrostichoides* (Pteridophyta) and *Synechococcus* sp. 7002 (Cyanobacteria), *Plant Syst. Evol.*, **187**, 89–102.
21. Freyer, R., Kiefer-Mayer, M.-C., and Kössel, H. 1997, Occurrence of plastid RNA editing in all major lineages of land plants, *Proc. Natl. Acad. Sci. USA*, **94**, 6285–6290.
22. Neckerman, K., Zeltz, P., Igloi, G. L., Kössel, H., and Maier, R. M. 1994, The role of RNA editing in conservation of start codons in chloroplast genomes, *Gene*, **146**, 177–182.
23. Hoch, B., Maier, R. M., Appel, K., Igloi, G. L., and Kössel, H. 1991, Editing of a chloroplast mRNA by creation of an initiation codon, *Nature*, **353**, 178–180.
24. Sugiura, M., Hirose, T., and Sugita, M. 1998, Evolution and mechanism of translation in chloroplast, *Ann. Rev. Genet.*, **32**, 437–459.
25. Crick, F. H. 1966, Codon-anticodon pairing: the wobble hypothesis, *J. Mol. Biol.*, **19**, 548–555.
26. Kato, T., Kaneko, T., Sato, S., Nakamura, Y., and Tabata, S. 2000, Complete structure of the chloroplast genome of a legume, *Lotus japonicus*, *DNA Res.*, **7**, 323–330.
27. Akashi, K., Takenaka, M., Yamaoka, S., Suyama, Y., Fukuzawa, H., and Ohyama, K. 1998, Coexistence of nuclear DNA-encoded tRNA(Val)(AAC) and mitochondrial DNA-encoded tRNA(Val)(UAC) in mitochondria of a liverwort *Marchantia polymorpha*, *Nucleic Acids Res.*, **26**, 2168–2172.
28. Dorner, M., Altmann, M., Paabo, S., and Morl, M. 2001, Evidence for import of a lysyl-tRNA into marsupial mitochondria, *Mol. Biol. Cell*, **12**, 2688–2698.
29. Glover, K. E., Spencer, D. F., and Gray, M. W. 2001, Identification and structural characterization of nucleus-encoded transfer RNAs imported into wheat mitochondria, *J. Biol. Chem.*, **276**, 639–648.
30. Morden, C. W., Wolfe, K. H., dePamphilis, C. W., and Palmer, J. D. 1991, Plastid translation and transcription genes in a non-photosynthetic plant: intact, missing and pseudo genes, *EMBO J.*, **10**, 3281–3288.
31. Novoselov, S. V., Rao, M., Onoshko, N. V. et al. 2002, Selenoproteins and selenocysteine insertion system in the model plant cell system, *Chlamydomonas reinhardtii*, *EMBO J.*, **21**, 3681–3693.
32. Beier, H. and Grimm, M. 2001, Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs, *Nucleic Acids Res.*, **29**, 4767–4782.