Molecular Cloning of a cDNA Encoding Cytochrome c of *Stellaria longipes* (Caryophyllaceae)—and the Evolutionary Implications

Xing-Hai Zhang and C. C. Chinnappa
Department of Biological Sciences, University of Calgary

A cDNA clone encoding cytochrome c of *Stellaria longipes* was isolated and characterized. The nucleotide and predicted amino acid sequences were highly similar to those from other plant, fungal, and animal species. No significant polymorphism was observed among different genotypes (ecotypes). The *S. longipes* genome contained a low copy number of nucleotide sequences of cytochrome c. The gene expression of cytochrome c exhibited a certain degree of tissue specificity, with more transcripts in leaf than in stem. A phylogenetic analysis of the cytochrome c amino acid sequences revealed an unusual aspect of plant cytochrome c evolution.

Introduction

The function of mitochondria is coordinated by both mitochondrial and nuclear genomes. While mitochondria produce some of their own proteins and RNAs, most mitochondrial polypeptides are encoded by nuclear genes (Newton 1988; Levings and Brown 1989; Brennicke et al. 1993). Cytochrome c is one such nuclear-encoded mitochondrial protein, at least in fungi and animals. It has been demonstrated that cytochrome c is synthesized as an apoprotein in cytoplasm, imported to the outer surface of the inner mitochondrial membrane where, with the addition of heme, it becomes a functional protein. However, unlike most nuclear-encoded mitochondrial or chloroplast proteins, cytochrome c does not contain an amino-terminal prosequence (leader sequence) (Zimmermann et al. 1979; Matsuura et al. 1981; Stuart et al. 1987; Pfanner and Neupert 1990). Its translocation requires neither a membrane potential nor adenosine triphosphate (Nicholson and Neupert 1989; Hartl and Neupert 1990; Pfanner and Neupert 1990). In this regard, cytochrome c is a unique protein.

Cytochrome c has been extensively studied in fungal and animal species (Montgomery et al. 1978; Boss et al. 1981; Russel and Hall 1982; Wu et al. 1986; Stuart et al. 1987, 1990; Nicholson and Neupert 1989; Dumont et al. 1991). Amino acid sequences of cytochrome c from 26 higher-plant species have been determined (Baba et al. 1981; Moore and Pettigrew 1990, pp. 115–127). In fact, cytochrome c is the only nuclear-encoded mitochondrial protein whose amino acid sequence is available for nearly 100 species of different kingdoms and has been used for broad-scale studies of molecular evolution. Recently, the nucleotide sequences of cytochrome c from unicellular green alga *Chlamydomonas reinhardtii* (Amati et al. 1988), flowering plants, rice, and *Arabidopsis thaliana* (Kemmerer et al. 1991*a*, 1991*b*) have been reported. Here we present the study on cytochrome c from another plant.

Stellaria longipes Goldie (Caryophyllaceae) is a dicotyledonous herbaceous perennial. It has a circumpolar distribution and grows in a variety of habitats (Chinnappa and Morton 1974, 1984). It has been extensively investigated to understand population differentiation, plasticity, and genetic diversity in the species (Chinnappa and Morton 1974, 1984; Macdonald et al. 1988; Cai and Chinnappa 1989; Muhammad et al. 1992). We attempted a molecular approach to study the evolution of *S. longipes*. To obtain sequence information from *S. longipes* (which was not available for any genes of this species), we constructed a cDNA library, from which several clones were isolated and sequenced (Zhang et al. 1993, and in press). One clone encoded cytochrome c. We used this cDNA as a probe to study gene complexity and expression of cytochrome c in *S. longipes*. A comparative phylogenetic analysis that includes the *S. longipes* cytochrome c is presented.

Material and Methods

Plant Material

Plants of *Stellaria longipes* Goldie were collected from different populations representing various ecological habitats. Each individual plant was designated a ge-

Key words: cytochrome c cDNA, *Stellaria longipes*, paralogy, plant cytochrome c evolution.

Address for correspondence and reprints: C. C. Chinnappa, Department of Biological Sciences, University of Calgary, Calgary, Alberta, Canada T2N 1N4.

Mol. Biol. Evol. 11(3):365–375. 1994.

0737-4038/94/1103-0004$02.00

notype and multiplied by vegetative propagation. The plants were maintained in growth chambers, growing in an annual cycle consisting of 5 mo under a long-day photoperiod (16 h) and warm (18°C–22°C) condition (abbreviated as "LDW") followed by a 7-mo treatment under a short-day photoperiod (8 h) and cold (5°C–8°C) condition. The genotypes used in this study were either diploid ($2n = 26$), tetraploid ($2n = 52$), or hexaploid ($2n = 78$). Genotypes S1–S6 and 366 ($2n = 52$) originated from the Plateau Mountain, Kananaskis Valley, Alberta. The altitudinal difference from S1 to S6 is ~1,000 m, while the difference in horizontal distance is 25 km, with S1 and 366 in the highest position (alpine tundra), S6 in the lowest (prairie), and S2–S5 in between. Genotypes 3005 ($2n = 52$) and 2990 ($2n = 26$) were collected from California and Arizona, respectively. Genotype 4566 ($2n = 78$) was from Manitoba. Genotype ASD ($2n = 52$) was from Athabasca sand dunes, Saskatchewan. The plants, free of insects and pathogens, were collected after growing for 2 wk in LDW, were ground in liquid nitrogen, and were stored at −80°C.

RNA Isolation and cDNA Library Construction

Total RNA was extracted from leaves and stems by using the method described by Logemann et al. (1987). Poly(A^+)-RNA was isolated from total RNA by oligo(dT)-cellulose (Sigma) chromatography (Sambrook et al. 1989, pp. 7.26–7.29). Mitochondrial mRNA does not contain poly(A) tail at its 3′ end (Newton 1988; Hanson and Folkerts 1992). Therefore, it is unlikely that our poly(A^+)-RNA extracts would be contaminated with mitochondrial RNAs, although this possibility still cannot be absolutely ruled out. A cDNA library was made based on the protocol of Gubler and Hoffman (1983) and according to the manufacturer's manual (Stratagene), using *Eco*RI as cloning site and λZAP II as vector. Oligo (dT) was used as a primer to initiate the synthesis of the first-stand cDNA. Poly(A^+)-RNA used for cDNA synthesis was from leaves of genotype S5 grown at LDW. About 30 random clones of plasmid DNA were isolated by the lysozyme and NaOH/SDS (rapid lysate) method (Sambrook et al. 1989, pp. 1.21–1.24).

DNA Sequencing and Computer Searching

DNA sequencing was performed according to the chain-termination method of Sanger et al. (1977) and manufacturers' manuals (Pharmacia and USB). Both strands of the cDNA clones were sequenced with ^{35}S-dATP (Amersham). Sequence searching was carried out with MacVector release 3.0 (IBI) and GenBank release 59.0. Search results showed that the sequence of a cDNA clone closely resembled the cytochrome c sequence from *Neurospora crassa*. Subsequently, this cDNA was completely sequenced and designated "SlC."

Isolation of Genomic DNA and Southern Blot Hybridization

Genomic DNA was extracted according to the method of Doyle and Dickson (1987) and purified by CsCl–ethidium bromide centrifugation (Beckman) at 170,000 *g* for 17 h. The DNA (10 μg/lane) was digested to completion by restriction endonucleases (15–20 units/lane; Pharmacia), separated on 1% agarose gel, and blotted to GeneScreen Plus nylon membranes (DuPont). The membranes were prehybridized overnight at 65°C in 5 × saline sodium citrate (SSC), 2 × Denhardt's solution, 2% sodium dodecyl sulfate (SDS). The cDNA S1C was labeled with [α-^{32}P]dCTP (Amersham) by primer extension using the Klenow fragment of DNA polymerase I (Pharmacia). The primer was a 17-mer oligonucleotide—5′-CGGCAGACACACTGGCT-3′ (nt 166–182)—of the SlC corresponding to amino acids 42–47 of *S. longipes* cytochrome c; fig. 1). Gel electrophoresis and autoradiogram of labeled samples showed that the labeled fragments obtained under our conditions were no larger than 120 bp (corresponding to amino acids 42–83 in fig. 1; data not shown) and therefore within the coding region. After overnight hybridization with the probe at 55°C, the membranes were washed once with 4 × 1% SSC at room temperature for 15 min and one or two times with 4 × SSC, 1% SDS at 55°C for 15 min and were exposed to Kodak XAR-5 film at −80°C.

Northern Blot Analysis

Total RNA (30 μg/lane) was denatured by 50% deionized formamide and 2.2 M formaldehyde at 85°C for 15 min and chilled on ice. The RNA samples were separated by 1.2% agarose and 2.2 M formaldehyde gel electrophoresis in 40 mM morpholinopropane sulfonic acid, 10 mM sodium acetate, 1 mM ethylenediamine-tetraacetate pH 7.0 and were blotted onto a GeneScreen Plus nylon membrane (DuPont). The filter was both prehybridized and hybridized in 5 × SSC, 2 × Denhardt's solution, 2% SDS with 5 mg of denatured yeast tRNA at 60°C. SlC was labeled with [α-^{32}P]dCTP by the random priming (Klenow fragment) method (Feinberg and Vogelstein 1983). After one 15-min wash in 2 × SSC at room temperature and one or two 15-min washes in 2 × SSC, 1.5% SDS at 60°C, the filter was exposed to Kodak XAR-5 film with an intensifying screen at −80°C. The amount of RNA loaded to the gels was calibrated by ethidium bromide staining and hybridization to an 18S rDNA clone of soybean.

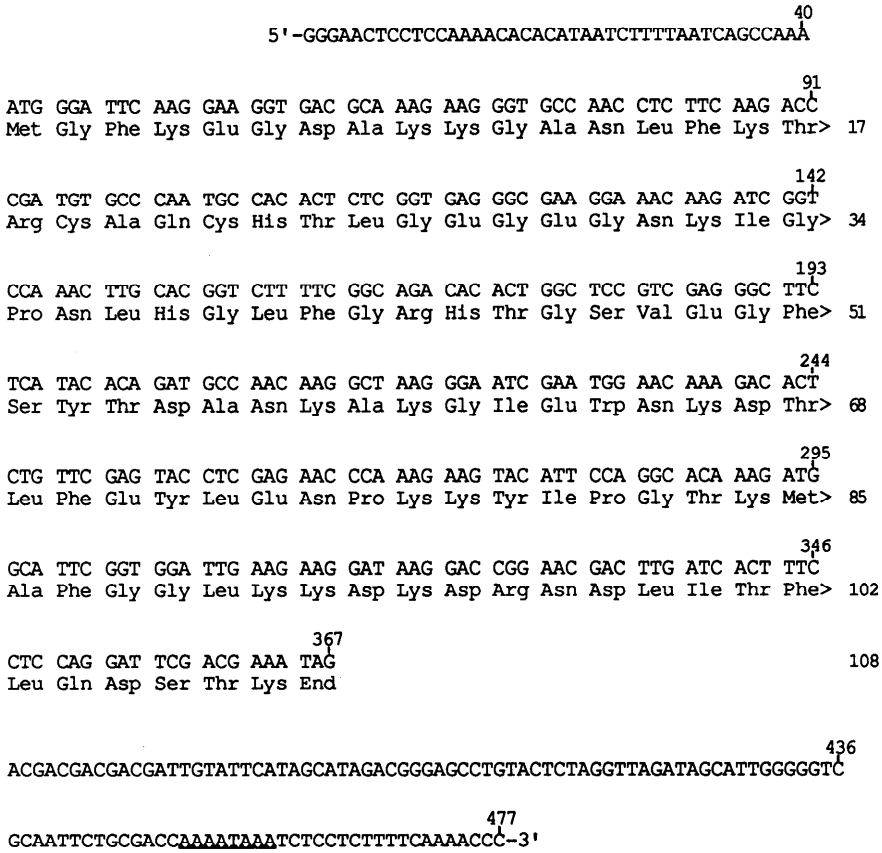

FIG. 1.—Nucleotide and predicted amino acid sequences of cytochrome c cDNA (i.e., SlC) from *Stellaria longipes*. The nucleotides and the amino acids are numbered above and on the right side of the sequence, respectively. The putative polyadenylation signal sequence is underlined.

Evolutionary Analysis

Since the amino acid sequences of cytochrome c from 26 higher-plant species are very similar to each other, only the sequences of 4 taxa (maize, tomato, buckwheat, and ginkgo) from divergent evolutionary lines were chosen for comparison with those deduced from the *S. longipes* and *Arabidopsis thaliana* cytochrome c nucleotide sequences. Six species from other kingdoms were included as outgroups in the preliminary analysis. In the final analysis, green flagellate *Euglena gracilis* was used to root the tree, because of its presumed close relationship to green plants (Baba et al. 1981). The same tree topologies were obtained when *Chlamydomonas reinhardtii* and human cytochrome c amino acid sequences were included in the analysis. To unify the sequence length, one and two amino acids at the carboxyl-termini for rat and ginkgo, respectively, were omitted (Swofford and Olsen 1990), and three to eight gaps were introduced into the amino-termini in the alignment. The amino acid sequences were aligned according to the method of Moore and Pettigrew (1990, pp. 118–125) and Kemmerer et al. (1991*b*). The data set was analyzed using the Phylogenetic Analysis Using Parsimony (PAUP) software package (version 3.0; Swofford 1989). All amino acid characters were considered as multistate and unordered. Gaps in the sequence were treated as the 21st amino acid (Swofford 1989) and were weighted as one-half the value of that for substitutions (Swofford and Olsen 1990; Mindell 1991). Both a heuristic search with tree bisection-reconnection (TBR) and a branch-and-bound search were used. Bootstrap values were calculated from 100 replicate Wagner parsimony analyses using heuristic option and stepwise addition sequence of the taxa in PAUP.

Results

Nucleotide Sequence of *Stellaria longipes* Cytochrome c

We constructed a cDNA library from leaves of a genotype (S5) of *Stellaria longipes*. About 30 random clones were isolated and partially sequenced. After searching through the GenBank database, a cDNA clone showed high similarity to the cytochrome c cDNA of *Neurospora crassa*. The cDNA clone (i.e., SlC) was

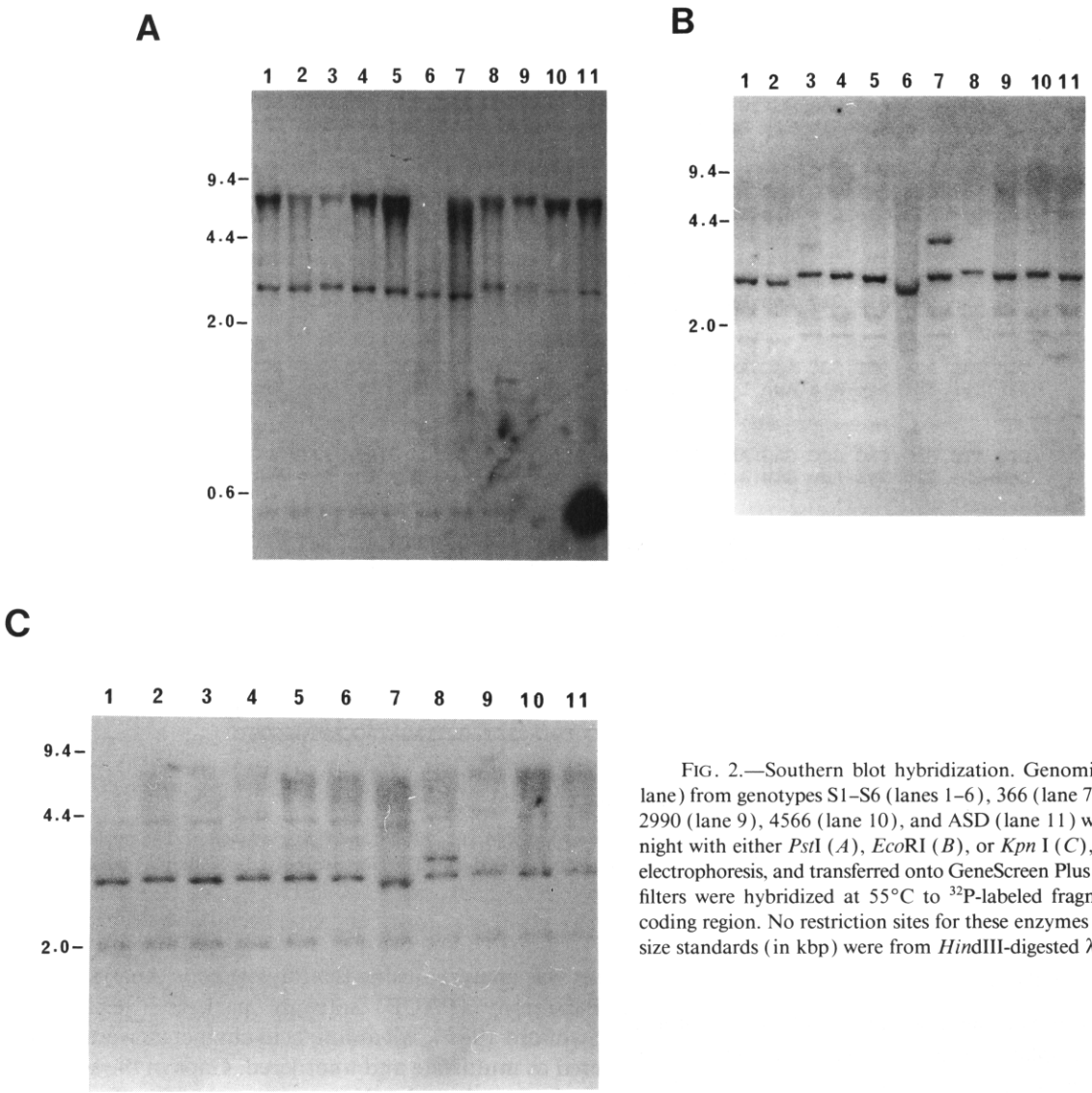

FIG. 2.—Southern blot hybridization. Genomic DNA (10 μg/lane) from genotypes S1–S6 (lanes 1–6), 366 (lane 7), 3005 (lane 8), 2990 (lane 9), 4566 (lane 10), and ASD (lane 11) was digested overnight with either *Pst*I (*A*), *Eco*RI (*B*), or *Kpn* I (*C*), separated by gel electrophoresis, and transferred onto GeneScreen Plus membranes. The filters were hybridized at 55°C to ^{32}P-labeled fragments of the SlC coding region. No restriction sites for these enzymes exist in SlC. The size standards (in kbp) were from *Hin*dIII-digested λ DNA.

completely sequenced. In order to rule out the possibility of fungal contamination in preparation of the cDNA library, a northern blot of total RNA from *S. longipes* was hybridized to the 3′ noncoding region of the cDNA SlC. The signals were similar to those from the hybridization of the SlC coding region, suggesting that the SlC is unlikely a fungal contaminant.

The nucleotide sequence (477 bp long; fig. 1) of SlC contains a full-length coding sequence of 327 nt (predicting 108 amino acids) and parts of the 5′ and 3′ noncoding regions. The codon usage patterns such as those for CGX, XTA, and G+C in the third base of SlC are typical of dicot plant genes (Murray et al. 1989). The 5′ noncoding region (40 bp) of SlC has a very low G content (5%) and high A+T content (65%). No similarity can be found in the 3′ noncoding region between *S. longipes* and any other species. The sequence, AATAAA, located 19 nt upstream from the 3′ terminus of SlC (fig. 1), may represent the polyadenylation signal of cytochrome c mRNA in *S. longipes*.

Complexity of Cytochrome c Genes among Different Genotypes of *S. longipes*

To investigate the complexity of the cytochrome c gene in the *S. longipes* genome, we used fragments of the cytochrome c coding sequence to probe the *S. longipes* genomic DNA digested by several restriction endonucleases (*Pst*I, *Eco*RI, and *Kpn*I in fig. 2; data of *Dra*I and *Bam*H I blots not shown). Eleven genotypes including diploid, tetraploids, and hexaploid were tested under low-stringency conditions. Three to four hybridizing bands with varied intensities were exhibited, in-

dicating a small gene family (presumably no more than four members) of cytochrome c. It has been reported that a species of rice, *Arabidopsis thaliana* (Kemmerer et al. 1991*a*, 1991*b*), *Chlamydomonas reinhardtii* (Amati et al. 1988), *N. crassa* (Stuart et al. 1987), and chicken (Limbach and Wu 1983) contain only a single cytochrome c gene while several other species and varieties of rice (Kemmerer et al. 1991*a*), *Drosophila melanogaster* (Swanson et al. 1985), rat (Scarpulla 1984), and human (Evans and Scarpulla 1988) bear two or multiple nucleotide sequences of cytochrome c. No significant polymorphism was observed among genotypes of *S. longipes* (fig. 3). The variations in both size and number of hybridizing bands were minimal. This suggests the conservative nature of the cytochrome c genes in the genomes of these different ecotypes or cytotypes.

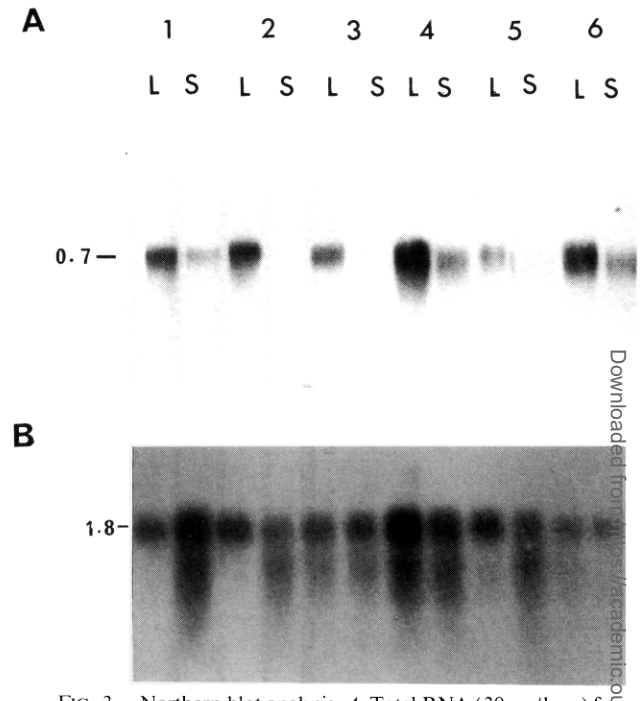

FIG. 3.—Northern blot analysis. *A*, Total RNA (30 μg/lane) from leaf (L) and stem (S) of genotypes S1–S6 (lanes 1–6), separated on 1.2% agarose, 2.2 M formaldehyde gel, blotted, and hybridized to labeled cDNA SlC. *B*, Same blot, stripped and hybridized to ^{32}P-labeled 18S rDNA of soybean. RNA size (in kb) was estimated with the RNA molecular-weight markers (Boehringer-Mannheim).

Expression of Cytochrome c Gene in Genotypes of *S. longipes*

To obtain information on cytochrome c gene expression, we hybridized the ^{32}P-labeled SlC to total RNA from the leaf and stem tissues of different genotypes. The result (fig. 3*A*) showed one species (0.7 kb) of cytochrome c mRNA. There is a certain degree of tissue specificity of cytochrome c gene expression. Leaves show stronger hybridization than stems, in all genotypes tested. In some genotypes, stem hybridization is hardly detectable (fig. 3*A*, lanes 2S, 3S, and 5S). The difference in hybridization intensity was not caused by variation of RNA loadings (fig. 3*B*). This may indicate different steady-state metabolic levels in leaf and stem, in which cytochrome c may play an active role. The tissue preference of cytochrome c gene transcription shown in *S. longipes* has not been reported in other species, although considerable variations in cytochrome c content have been observed in mitochondria from different tissues of rat and chick (Pettigrew and Moore 1987, pp. 37–38). For *S. longipes*, the stem metabolic activity is presumably lower than in leaf cells (such as less total RNA in stems, on a fresh-weight basis; data not shown). We speculate that the tissue specificity of transcription is related to the lower metabolic activity—and therefore lower steady-state level of cytochrome c mRNA—in stem than in leaf tissue. Since no information on gene expression of cytochrome c from other plants is available, it remains to be seen whether tissue specificity and genotypic differences in cytochrome c expression are unique to *S. longipes*.

Evolution of Plant Cytochromes c

There is a considerable sequence similarity of cytochrome c, between *S. longipes* and other taxa. Figure 4 shows a comparison of the amino acid sequences of cytochrome c from 12 species. The 26 amino acids reported to be invariant in all organisms sequenced to date (Moore and Pettigrew 1990, pp. 115–127) are also present in SlC (fig. 4). In fact, seven other residues are also unchanged in those species compared: Gly23, Gly37, Gly45, Asn70, Tyr74, Lys86, and Asp93. The fact that >30% of the amino acids are positionally identical in species from different kingdoms suggests that functional constraints require the chemical stability of vital residues such as those involved in the covalent binding of the heme group and those involved in sustaining the essential biological structure (Salemme 1977).

Table 1 summarizes the comparison of cytochrome c amino acid sequences from 28 plants, 2 protists, 2 fungi, and 2 animals. Two flowering plants—*S. longipes* and *A. thaliana*—share a low degree (18.9%) of sequence dissimilarity. Both are similar to fungus *N. crassa*. It is surprising that these two plants exhibit higher sequence divergences from other plants than from species of fungi and animals (table 1). On the other hand, all the higher plants sequenced so far, except *A. thaliana* and *S. longipes*, are very similar, with only 4%–15% sequence dissimilarity, and contain a conserved N-extension motif

FIG. 4.—Alignment of the amino acid sequences of cytochromes c from 12 taxa. A gap is introduced into the sequence of *Euglena gracilis* to increase the similarity. The initiation amino acid methionine predicted by nucleotide sequence is omitted from the alignment. The number above the sequence indicates the residue position according to the vertebrate numeration for cytochrome c. Asterisks represent identical residues. The amino acids different from the *Stellaria longpipes* cytochrome c are presented. The 26 amino acids invariant throughout all known cytochrome c sequences (Moore and Pettigrew 1990, pp. 115–127) are underlined.

Table 1
Comparison of Amino Acid Sequences of Cytochromes c from Different Taxa

Species (common name)[a]	Family[b]	Protein Size[c] (amino acids)	% Sequence Dissimilarity to	
			Stellaria longipes	Maize
Stellaria longipes (chickweed)	Caryophyllaceae	107	...	46.8
Arum maculatum (cuckoopint)	Araceae	111	46.8	4.5
Acer negundo (box elder)	Aceraceae	112	46.8	9.0
Allium porrum (leek)	Liliaceae	111	46.8	8.1
Arabidopsis thaliana	Brassicaceae	111	18.9	44.1
Abutilon theophrasti (velvet leaf)	Malvaceae	111	45.0	7.2
Brassica napus (rape) and *B. oleracea* (cauliflower)	Brassicaceae	111	45.0	5.4
Cucurbita maxima (pumpkin)	Cucurbitaceae	111	46.8	6.3
Cannabis sativa (hemp)	Cannabaceae	111	45.9	9.0
Fagopyrum esculentum (buckwheat)	Polygonaceae	111	43.2	9.0
Guizotia abyssinica (Niger)	Asteraceae	111	45.0	7.2
Gossypium barbadense (cotton)	Malvaceae	111	45.0	7.2
Ginkgo biloba (ginkgo)	Ginkgoceae	113	45.0	10.8
Helianthus annuus (sunflower)	Asteraceae	111	45.9	9.0
Lycopersicum esculentum (tomato)	Solanaceae	111	47.7	3.6
Nigella damascena (love-in-a-mist)	Ranunculaceae	111	45.9	11.7
Oryza sativa (rice)	Poaceae	111	47.7	7.2
Phaseolus aureus (mung bean)	Fabaceae	111	45.9	9.9
Pastinaca oleracea (parsnip)	Apiaceae	111	45.9	10.8
Ricinus communis (castor bean)	Euphorbiaceae	111	46.8	6.3
Sesamum indicum (sesame)	Pedaliaceae	111	47.7	7.2
Sambucus nigra (elder)	Caprifoliaceae	111	46.8	8.1
Spinacea oleracea (spinach)	Chenopodiaceae	111	43.2	15.3
Solanum tuberosum (potato)	Solanaceae	111	47.7	6.3
Triticum aestivum (wheat)	Poaceae	112	44.1	8.1
Tropaeolum majus (nasturtium)	Tropaeolaceae	111	45.0	9.9
Zea mays (maize)	Poaceae	111	46.8	...
Chlamydomonas reinhardtii	...	111	49.5	25.2
Euglena gracilis	...	102	49.5	52.3
Neurospora crassa	...	107	18.9	43.2
Saccharomyces cerevisiae (yeast)	...	108	34.2	41.4
Drosophila melanogaster (fruit fly)	...	107	31.5	40.5
Rattus (rat)	...	104	37.8	37.8

[a] Data for *A. thaliana* are from Kemmerer et al. (1991*b*); those for *C. reinhardtii* are from Amati et al. (1988); those for *S. cerevisiae* (iso-1-cytochrome c) are from Boss et al. (1981); and those for other species are from reviews by Moore and Pettigrew (1990, pp. 115–127). All plant and algal sequences are derived from protein sequencing, except for *S. longipes* and *C. reinhardtii* (both from cDNA), *A. thaliana* (from genomic DNA), and rice (from both protein and nuclear DNA sequencing; Kemmerer et al. 1991*a*).

[b] Only the families of higher plants are presented.

[c] Excluding the initiator amino acid methionine.

(table 1 and fig. 4). The perplexingly odd relationship between *S. longipes, A. thaliana,* and many other higher plants studied so far reveals the need for more careful examinations of plant cytochrome c evolution.

PAUP analysis of the cytochrome c sequences, using both heuristic search and branch-and-bound methods, resulted in the two very similar most parsimonious cladograms. One of the trees is presented in figure 5. The phylogenetic tree shows that the *S. longipes* cytochrome c is unexpectedly grouped with *Neurospora* as well as with *Arabidopsis* and separate from other plants and protists. Similar results were also observed in the analysis of the rice and *A. thaliana* cytochrome c (Kemmerer et al. 1991*a*, 1991*b*).

Discussion

Both studies on *Arabidopsis thaliana* (Kemmerer et al. 1991*b*) and studies on *Stellaria longipes* reveal unexpected results regarding plant cytochrome c evo-

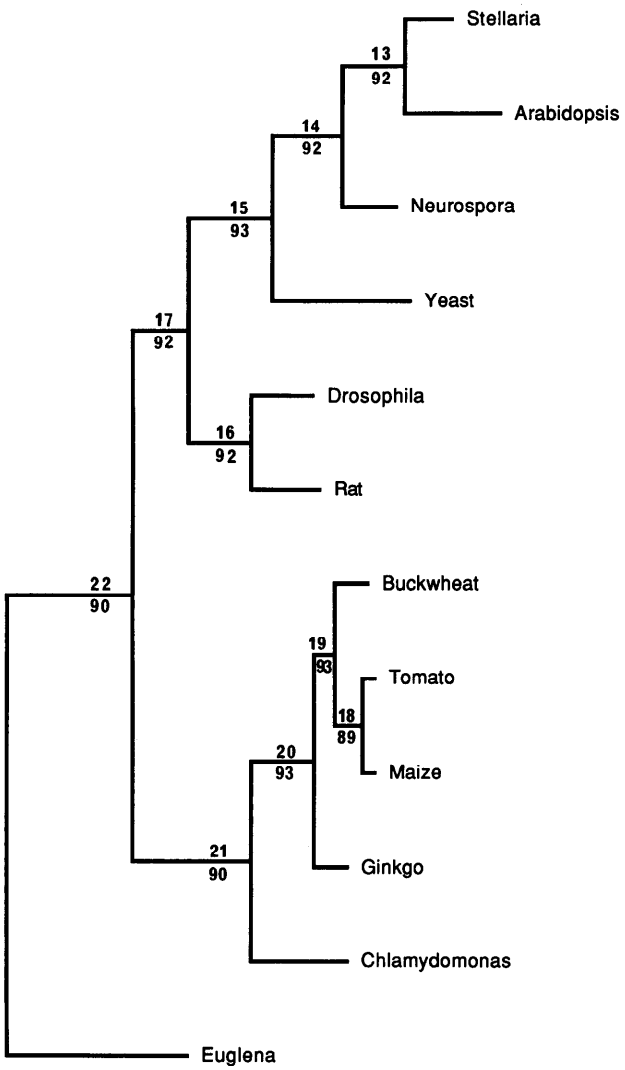

FIG. 5.—One of two most parsimonious trees for the amino acid sequences of cytochromes c from 12 taxa. Consistency index (CI) = 0.814; CI excluding uninformative characters = 0.784; retention index = 0.760; and tree length = 221 steps. Numbers on the branches indicate the number of residue changes. Numbers below branches are bootstrap values.

lution. In an attempt to explain the dissimilarity between *A. thaliana* cytochrome c and those of other plants, Kemmerer et al. (1991*b*) suggested the possibility that certain genes such as the cytochrome c gene in *Arabidopsis* have diverged from those of plants and converged toward those of yeasts and molds. A possible paralogous relationship between cytochrome c genes of *Arabidopsis* and rice was also proposed (Kemmerer et al. 1991*a*). While the possibility of gene convergence remains to be tested, we agree that paralogy may contribute, in large part, to the discrepancy between the molecular phylogeny of cytochrome c and the true species phylogeny.

Gene duplication plays an important role in molecular evolution (Wilson et al. 1977; Li and Graur 1991) and results in the occurrence of two or more homologous genes. Orthologous genes are descendants of an ancestral gene that was present in the last common ancestor of two or more species, whereas paralogous genes are descendants of a duplicated ancestral gene. Orthologous comparison of protein or DNA sequences provides a gene phylogeny from which evolutionary relationship among species can be inferred (Avers 1989, pp. 361–366; Li and Graur 1991, pp. 136–171). Comparison between the products of paralogous genes in the analysis of species phylogeny could lead to anomalous conclusions, which may give false branching order in an inferred species-phylogeny tree (Wilson et al. 1977; Avers 1989, pp. 361–366). This is likely the case for *S. longipes* and *A. thaliana* in the cytochrome c gene tree (fig. 5).

Cytochrome c gene duplication has been observed in many species (Moore and Pettigrew 1990, pp. 280–284). Two cytochrome c genes in *Drosophila melanogaster* differ both in the amino acid sequence and in the rate of evolution (Wu et al. 1986). There are two or more cytochrome c genes in the genomes of *S. longipes* (fig. 2) and rice (Kemmerer et al. 1991*a*). It is possible that at least some of the plants listed in table 1 may also have duplicated genes of cytochrome c. The duplicated genes of different duplication lineages in the same genome may evolve with different tempo and mode (such as fast evolved and slowly evolved genes of cytochrome c) and diverge from each other in their DNA sequences. Some of the duplicated genes, such as processed pseudogenes of cytochrome c found in rat, mouse (Scarpulla and Wu 1983; Scarpulla 1984; Limbach and Wu 1985), and other mammals (Scarpulla et al. 1982), may even lose functions. Because of gene duplication, homologous genes of different species can be either orthologues or paralogues. One of the rice cytochrome c sequences is confirmed by both DNA and protein sequencing (Mori and Morita 1980; Kemmerer et al. 1991*a*). It shows very high similarity with many other flowering plants, except for *A. thaliana* and *S. longipes* (table 1). It is probable that the rice cytochrome c copy is paralogous to those known from *A. thaliana* and *S. longipes* and orthologous to at least some of the cytochromes c from other plants.

It is interesting that a polymerase-chain-reaction product from rice genomic DNA (PCR-1) exhibits 73%–75% of sequence identity to cytochromes c from *A. thaliana, N. crassa* (Kemmerer et al. 1991*a*), and *S. longipes* and 46% to those from the rice gene (OsCc-1) and other plants. If the PCR-1 represents a second rice cytochrome c gene or pseudogene, rather than a fungal contaminant (Kemmerer et al. 1991*a*), it is likely that the cytochromes c from *S. longipes, A. thaliana,* and the PCR-1 rice gene are orthologues, while those from the rice cytochrome c gene (OsCc-1) and from at least some of

the other plants studied (table 1) are paralogues. Availability of more cytochrome c–related nucleotide sequences from *S. longipes,* rice, and *A. thaliana* would be helpful for clarifying the orthologous relationship between different genes.

As another aspect of gene paralogy, there is a possibility that, besides the nuclear genome, plant mitochondria may also encode and synthesize cytochrome c. Some of the plant cytochromes c listed in table 1, except for rice, *A. thaliana,* and *S. longipes,* might be the products of mitochondrial gene expression. Therefore another kind of paralogous relationship could exist. The structural gene of cytochrome c has been proved to be exclusively located in the nuclear genomes of fungi, mammals, and liverwort (Salemme 1977; Gray 1992; Oda et al. 1992). For higher plants, however, no evidence has ruled out the possibility of functional mitochondrial encoded gene(s) of cytochrome c, even though DNA clonings of cytochrome c from rice, *A. thaliana* (Kemmerer et al. 1991*a,* 1991*b*), and *S. longipes* indicate the existence of nuclear cytochrome c–coding sequences. The plant mitochondrial genome is exceptionally large and complex (Newton 1988; Levings and Brown 1989). Its DNA sequence evolves at a much lower rate than nuclear and chloroplast DNAs (Wolfe et al. 1987; Palmer 1992). Plant mitochondria also encode many more proteins than do those of fungi and animals (Gray 1992; Brennicke et al. 1993). It is conceivable that plant mitochondrial genome could possess a cytochrome c gene. Flowering-plant *Oenothera berteriana* mitochondria have been shown to encode a protein involved in assembly and maturation of cytochrome c (Schuster et al. 1993). So far, however, the mitochondrial encoded cytochrome c has yet to be demonstrated. Studies of rice cytochrome c (Mori and Morita 1980; Kemmerer et al. 1991*a*) do not support this speculation.

Another possibility is the lateral transfer of fungal genes into a plant lineage. The *Neurospora* cytochrome c is more similar to that of *S. longipes* and *A. thaliana* than to yeast or animal ones (table 1 and fig. 5). This could be due to an ancient gene duplication in a common ancestor of plants, fungi, and animals, with the known sequences of *Neurospora/Stellaria/Arabidopsis* being true orthologues. All other sequences are either a second branch of orthologues or represent branches of paralogues. This would explain the odd relationship of *S. longipes* and *A. thaliana* to *Neurospora* and other plants in the phylogenetic tree (fig. 5). However, it is hard to explain why 26 of 28 plant cytochrome c amino acid sequences are very similar to each other while very dissimilar to the two other plant sequences deduced from DNA sequences.

In summary, the odd dissimilarity of cytochrome c between *S. longipes* (and *A. thaliana*) and 26 other higher plants could be attributed to (1) comparison of paralogous genes from ancient gene duplications, which occurred within nuclear genomes and possibly within mitochondrial genomes as well, and/or (2) lateral gene transfer resulting in horizontal evolution. To confirm gene paralogy requires evidence of at least two considerably different cytochrome c genes from the same plant species. More sequence data on plant cytochrome c genes, therefore, are needed to resolve the puzzle of plant cytochrome c evolution.

Sequence Availability

The nucleotide sequence data reported will appear in the EMBL, GenBank, and DDBJ nucleotide sequence databases, under accession number Z21499.

Acknowledgments

We are grateful to Dr. Maurice M. Moloney for his advice and for allowing us to use his lab facility, to Janis M. Gall and Dr. Peta C. Bonham-Smith for their help with sequencing and cDNA synthesis, to Elizabeth E. Dickson for discussion of phylogenetic analysis, and particularly to Dr. Jeff D. Palmer of Indiana University for advice and critical comments on the manuscript. We thank the anonymous reviewers for their critical comments of the earlier versions of this article. This research was supported by Natural Sciences and Engineering Research Council (NSERC) of Canada operating grants to C. C. Chinnappa and M. M. Moloney.

LITERATURE CITED

AMATI, B. B., M. GOLDSCHMIDT-CLERMONT, C. J. A. WALLACE, and J.-D. ROCHAIX. 1988. cDNA and deduced amino acid sequences of cytochrome c from *Chlamydomonas reinhardtii:* unexpected functional and phylogenetic implications. J. Mol. Evol. **28**:151–160.

AVERS, C. J. 1989. Process and pattern in evolution. Oxford University Press, New York and Oxford.

BABA, M. L., L. L. DARGA, M. GOODMAN, and J. CZELUSNIAK. 1981. Evolution of cytochrome c investigated by the maximum parsimony method. J. Mol. Evol. **17**:197–213.

BOSS, J. M., S. GILLAM, R. S. ZITOMER, and M. SMITH. 1981. Sequence of the yeast iso-1-cytochrome c mRNA. J. Biol. Chem. **256**:12958–12961.

BRENNICKE, A., L. GROHMANN, R., HIESEL, V. KNOOP, and W. SCHUSTER. 1993. The mitochondrial genome on its way to the nucleus: different stages of gene transfer in higher plants. FEBS Lett. **325**:140–145.

CAI, Q., and C. C. CHINNAPPA. 1989. Gene duplication of isozymes in the tetraploid of *Stellaria longipes* (Caryophyllaceae). J. Hered. **80**:112–117.

CHINNAPPA, C. C., and J. K. MORTON. 1974. The cytology of *Stellaria longipes.* Can. J. Genet. Cytol. **16**:499–514.

———. 1984. Studies on the *Stellaria longipes* complex (Caryophyllaceae)—biosystematics. Syst. Bot. **9**:60–73.

DOYLE, J. J., and E. E. DICKSON. 1987. Preservation of plant samples for DNA restriction endonuclease analysis. Taxon **36**:715–722.

DUMONT, M. E., T. S. CARDILLO, M. K. HAYES, and F. SHERMAN. 1991. Role of cytochrome c heme lyase in mitochondrial import and accumulation of cytochrome c in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **11**:5487–5496.

EVANS, M. J., and R. C. SCARPULLA. 1988. The human somatic cytochrome c gene: two classes of processed pseudogenes demarcate a period of rapid molecular evolution. Proc. Natl. Acad. Sci. USA **85**:9625–9629.

FEINBERG, A. P., and B. VOGELSTEIN. 1983. A technique for radiolabelling DNA restriction fragments to high specific activity. Anal. Biochem. **132**:6–13.

GRAY, M. W. 1992. The endosymbiont hypothesis revisited. Int. Rev. Cytol. **141**:233–357.

GUBLER, U., and B. J. HOFFMAN. 1983. A simple and very efficient method for generating cDNA libraries. Gene **25**: 263–269.

HANSON, M. R., and O. FOLKERTS. 1992. Structure and function of the higher plant mitochondrial genome. Int. Rev. Cytol. **141**:129–172.

HARTL, F.-U., and W. NEUPERT. 1990. Protein sorting to mitochondria: evolutionary conservation of folding and assembly. Science **247**:930–938.

KEMMERER, E. C., M. LEI, and R. WU. 1991*a*. Isolation and molecular evolutionary analysis of a cytochrome c gene from *Oryza sativa* (rice). Mol. Biol. Evol. **8**:212–226.

———. 1991*b*. Structure and molecular evolutionary analysis of a plant cytochrome c gene: surprising implications for *Arabidopsis thaliana*. J. Mol. Evol. **32**:227–237.

LEVINGS, C. S. III, and G. G. BROWN. 1989. Molecular biology of plant mitochondria. Cell **56**:171–179.

LI, W.-H., and D. GRAUR. 1991. Fundamentals of molecular evolution. Sinauer, Sunderland, Mass.

LIMBACH, K. J., and R. WU. 1983. Isolation and characterization of two alleles of the chicken cytochrome c gene. Nucleic Acids Res. **24**:8931–8950.

———. 1985. Characterization of a mouse somatic cytochrome c gene and three cytochrome c pseudogenes. Nucleic Acids Res. **13**:617–644.

LOGEMANN, J., J. SCHELL, and L. WILLMITZER. 1987. Improved method for the isolation of RNA from plant tissues. Anal. Biochem. **163**:16–20.

MACDONALD, S. E., C. C. CHINNAPPA, and D. M. REID. 1988. Evolution and phenotypic plasticity in the *Stellaria longipes* complex: comparisons among cytotypes and habitats. Evolution **42**:1036–1046.

MATSUURA, S., M. ARPIN, C. HANNUM, E. MARGOLIASH, D. D. SABATINI, and T. MORIMOTO. 1981. In vitro synthesis and posttranslational uptake of cytochrome c into isolated mitochondria: role of a specific addressing signal in the apocytochrome. Proc. Natl. Acad. Sci. USA **78**:4368–4372.

MINDELL, D. P. 1991. Aligning DNA sequences: homology and phylogenetic weighting. Pp. 73–89 *in* M. M. MIYAMOTO and J. CRACRAFT, eds. Phylogenetic analysis of DNA sequences. Oxford University Press, New York.

MONTGOMERY, D. L., B. D. HALL, S. GILLAM, and M. SMITH. 1978. Identification and isolation of the yeast cytochrome c gene. Cell **14**:673–680.

MOORE, G. R., and G. W. PETTIGREW. 1990. Cytochromes c: evolutionary, structural and physicochemical aspects. Springer, Berlin and Heidelberg.

MORI, E., and MORITA, Y. 1980. Amino acid sequence of cytochrome c from rice. J. Biochem. **87**:249–266.

MUHAMMAD, N., X.-H. ZHANG, N. EMERY, C. C. CHINNAPPA, and M. M. MOLONEY. 1992. A preliminary study on DNA restriction fragment analysis of the *Stellaria longipes* complex (Caryophyllaceae). Beitr. Biol. Pflanzen **67**:119–127.

MURRAY, E. E., J. LOTZER, and M. EBERLE. 1989. Codon usage in plant genes. Nucleic Acids Res. **17**:477–498.

NEWTON, K. J. 1988. Plant mitochondrial genomes: organization, expression and variation. Annu. Rev. Plant Physiol. Mol. Biol. **39**:503–532.

NICHOLSON, D. W., and W. NEUPERT. 1989. Import of cytochrome c into mitochondria: reduction of heme, mediated by NADH and flavin nucleotides, is obligatory for its covalent linkage to apocytochrome c. Proc. Natl. Acad. Sci. USA **86**:4340–4344.

ODA, K., K. YAMATO, E. OHTA, Y. NAKAMURA, M. TAKEMURA, N. NOZATO, K. AKASHI, T. KANEGAE, Y. OGURA, T. KOHCHI, and K. OHYAMA. 1992. Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA: a primitive form of plant mitochondrial genome. J. Mol. Biol. **223**:1–7.

PALMER, J. D. 1992. Mitochondrial DNA in plant systematics: applications and limitations. Pp. 36–49 *in* P. S. SOLTIS, D. E. SOLTIS, and J. D. DOYLE, eds. Molecular systematics of plants. Chapman & Hall, New York.

PETTIGREW, G. W., and G. R. MOORE. 1987. Cytochromes c: biological aspects. Springer, Berlin and Heidelberg.

PFANNER, N., and W. NEUPERT. 1990. The mitochondrial protein import apparatus. Annu. Rev. Biochem. **59**:331–353.

RUSSEL, P. R., and B. D. HALL. 1982. Structure of the *Schizosaccharomyces pombe* cytochrome c gene. Mol. Cell. Biol. **2**:106–116.

SALEMME, F. R. 1977. Structure and function of cytochromes c. Annu. Rev. Biochem. **46**:299–329.

SAMBROOK, J., E. F. FRITSCH, and T. MANIATIS. 1989. Molecular cloning: a laboratory manual, 2d ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

SANGER, F., S. NICKLEN, and A. R. COULSON. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.

SCARPULLA, R. C. 1984. Processed pseudogenes for rat cytochrome c are preferentially derived from one of three alternate mRNAs. Mol. Cell. Biol. **4**:2279–2288.

SCARPULLA, R. C., K. M. AGNE, and R. WU. 1982. Cytochrome c gene-related sequences in mammalian genomes. Proc. Natl. Acad. Sci. USA **79**:739–743.

SCARPULLA, R. C., and R. WU. 1983. Nonallelic members of the cytochrome c multigene family of the rat may arise through different messenger RNAs. Cell **32**:473–482.

SCHUSTER, W., B. COMBETTES, K. FLIEGER, and A. BRENNICKE. 1993. A plant mitochondrial gene encodes a protein involved in cytochrome c biogenesis. Mol. Gen. Genet. **239:** 49–57.

STUART, R. A., W. NEUPERT, and M. TROPSCHUG. 1987. Deficiency in mRNA splicing in a cytochrome c mutant of *Neurospora crassa:* importance of carboxy terminus for import of apocytochrome c into mitochondria. EMBO J. **6:**2131–2137.

STUART, R. A., D. W. NICHOLSON, and W. NEUPERT. 1990. Early steps in mitochondrial protein import: receptor functions can be substituted by the membrane insertion activity of apocytochrome c. Cell **60:**31–43.

SWANSON, M. S., S. M. ZIEMINN, D. D. MILLER, E. A. E. GARBER, and E. MARGOLIASH. 1985. Developmental expression of nuclear genes that encode mitochondrial proteins: insect cytochrome c. Proc. Natl. Acad. Sci. USA **82:** 1964–1968.

SWOFFORD, D. L. 1989. PAUP: phylogenetic analysis using parsimony, version 3.0. Illinois Natural History Survey, Champaign.

SWOFFORD, D. L., and G. J. OLSEN. 1990. Phylogeny reconstruction. Pp. 411–501 *in* D. M. HILLIS and C. MORITZ, eds. Molecular systematics. Sinauer, Sunderland, Mass.

WILSON, A. C., S. S. CARLSON, and T. J. WHITE. 1977. Biochemical evolution. Annu. Rev. Biochem. **46:**573–639.

WOLFE, K. H., W.-H. LI, and P. M. SHARP. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. USA **84:**9054–9058.

WU, C.-I., W.-H. LI, J. J. SHEN, R. C. SCARPULLA, K. J. LIMBACH, and R. WU. 1986. Evolution of cytochrome c genes and pseudogenes. J. Mol. Evol. **23:**61–75.

ZHANG, X.-H., and C. C. CHINNAPPA. Triose phosphate isomerase of *Stellaria longipes* (Caryophyllaceae). Genome (in press).

ZHANG, X.-H., M. M. MOLONEY, and C. C. CHINNAPPA. 1993. Nucleotide sequence of a cDNA clone encoding a dehydrin-like protein from *Stellaria longipes.* Plant Physiol. **103:** 1029–1030.

ZIMMERMANN, R., U. PALUCH, and W. NEUPERT. 1979. Cell-free synthesis of cytochrome c. FEBS Lett. **108:**141–146.

BARBARA A. SCHAAL, reviewing editor

Received August 25, 1993

Accepted December 5, 1993