Nucleus-Encoded Genes for Plastid-Targeted Proteins in *Helicosporidium*: Functional Diversity of a Cryptic Plastid in a Parasitic Alga†

Audrey P. de Koning* and Patrick J. Keeling

Program in Evolutionary Biology, Canadian Institute for Advanced Research, and Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada

Plastids are the organelles of plants and algae that house photosynthesis and many other biochemical pathways. Plastids contain a small genome, but most of their proteins are encoded in the nucleus and posttranslationally targeted to the organelle. When plants and algae lose photosynthesis, they virtually always retain a highly reduced "cryptic" plastid. Cryptic plastids are known to exist in many organisms, although their metabolic functions are seldom understood. The best-studied example of a cryptic plastid is from the intracellular malaria parasite, *Plasmodium*, which has retained a plastid for the biosynthesis of fatty acids, isoprenoids, and heme by the use of plastid-targeted enzymes. To study a completely independent transformation of a photosynthetic plastid to a cryptic plastid in another alga-turned-parasite, we conducted an expressed sequence tag (EST) survey of *Helicosporidium*. This parasite has recently been recognized as a highly derived green alga. Based on phylogenetic relationships to other plastid homologues and the presence of N-terminal transit peptides, we have identified 20 putatively plastid-targeted enzymes that are involved in a wide variety of metabolic pathways. Overall, the metabolic diversity of the *Helicosporidium* cryptic plastid exceeds that of the *Plasmodium* plastid, as it includes representatives of most of the pathways known to operate in the *Plasmodium* plastid as well as many others. In particular, several amino acid biosynthetic pathways have been retained, including the leucine biosynthesis pathway, which was only recently recognized in plant plastids. These two parasites represent different evolutionary trajectories in plastid metabolic adaptation.

Plastids, or chloroplasts, are photosynthetic organelles of plants and algae that are derived from an endosymbiotic cyanobacterium (24). All plastids that have been examined have small genomes that encode a handful of proteins involved in housekeeping functions, photosynthesis, and some anabolic functions. These proteins represent only a small fraction of the proteins that are needed for plastid maintenance and activity: most genes for plastid proteins have been transferred to the nuclear genome, and their products are posttranslationally targeted to the plastid. While these genes may be encoded in the nucleus, they are derived from the cyanobacterium, and they can therefore be recognized by their phylogenetic history and by the presence of an N-terminal transit peptide that directs their targeting to the organelle (9).

While plastids are most often associated with photosynthesis, they fulfill many other important metabolic roles in both plants and algae. The best characterized of these is the biosynthesis of fatty acids, isopentyl diphosphate (for isoprenoid biosynthesis), various amino acids, and tetrapyrroles. With such metabolic complexity, it is not surprising that the many plants and algae that have lost photosynthesis have retained cryptic plastids for some other purpose. Cryptic plastids are highly derived nonphotosynthetic organelles that lack thylakoid structures and pigments, which makes them difficult to identify. It is

only by investigations of their metabolism or molecular biology that these plastids are recognized and that the photosynthetic ancestry of the organism that harbors them is exposed. A good example of this is the recently discovered plastid of apicomplexan parasites such as the malaria parasite, *Plasmodium*. This organelle, called the apicoplast, was first identified by the discovery of plastid-encoded genes (13). Not surprisingly, however, the apicoplast genome contains few genes, and most are involved in housekeeping activities that reveal little about why this plastid was retained. The function of the organelle was only discovered through the sequencing of nucleus-encoded genes for plastid-targeted enzymes, which have shown that the apicoplast's role in the malaria parasite is the biosynthesis of fatty acids, isoprenoids, and heme (31). The conversion of a fully functional photosynthetic plastid to the highly derived relict presently seen in apicomplexa is a remarkable transformation that raises a number of questions on the changing role of the organelle as various metabolic pathways become obsolete. Here we have examined this process in an unrelated parasite whose plastid has a very different evolutionary history to compare the effects of metabolic reduction.

Helicosporidia are a group of parasites that infect various invertebrates (5, 38). They were first described in 1921 (22) and since then have been assigned to various types of protists or fungi, but their extremely sophisticated and derived infection mechanism is unlike that of any other eukaryote and they accordingly remained something of a mystery. Recently, however, the first molecular data from this group were characterized and, surprisingly, showed *Helicosporidium* to be a trebouxiophyte green alga that is closely related to another enigmatic

* Corresponding author. Mailing address: 3529-6270 University Blvd., Vancouver, BC V6T 1Z4, Canada. Phone: (604) 822-2845. Fax: (604) 822-6089. E-mail: apdk@interchange.ubc.ca.

Pathway & Protein	Related to Plastid Homologue	Support
Fatty acid biosynthesis ▪ Acyl carrier protein (ACP)	Yes	Fig. 2A
Fatty acid modification ▫ Stearoyl ACP desaturase	Yes	Fig. 2D
Pyruvate dehydrogenation ▫ Dihydrolipoamide dehydrogenase	Yes	89-100%
Tetrapyrrole biosynthesis ▫ Glutamate-1-semialdehyde aminomutase	Yes	94-98%[1]
Isoprenoid biosynthesis (DOXP) ▫ LytB/ispH	Yes	95-98%[2]
Leucine biosynthesis ▪ Isopropylmalate synthase ▪ Isopropylmalate isomerase SSU	Yes Yes	Fig. 3A Fig. 3B
Serine biosynthesis ◪ Phosphoserine aminotransferase	Yes	<50-60%
Amino acid metabolism ◪ Aminotransferase	Yes	96-99%
Reducing Potential ▫ SufB ▪ Ferredoxin ◪ Ferredoxin-Thioredoxin reductase (subunit A)	Yes Yes Yes	Fig. 2B strong[3] Fig. 2C
Translation ▫ Ribosomal protein L15	Yes	<50-53%
Plastid Translation regulation ▪ Protein disulfide isomerase ▫ Poly A binding protein	Yes No	98-100% <50%
Protein folding ◪ GroES/Cpn10 ▫ ClpB/Hsp100	Yes Yes	85-87% 93-95%
Betaine biosynthesis ▫ Betaine aldehyde dehydrogenase	Yes*	99-100%
Sulfur/Cysteine metabolism ▪ Adenylylsulfate kinase ▫ O-acetylserine (thiol) lyases	Yes* Yes*	80-93%[2] 59-76%

FIG. 1. Summary of cDNAs for putative plastid-targeted proteins from *Helicosporidium* sorted by pathway. At left, filled boxes indicate the presence of an N-terminal extension predicted to be a plastid transit peptide, half-filled boxes indicate a leader of ambiguous function (e.g., the transcript was truncated within the leader or the leader could not unambiguously be predicted to be a transit peptide), and open boxes indicate the transcript was truncated within the mature protein. The second column indicates whether the *Helicosporidium* protein was phylogenetically related to other plastid-targeted proteins, and the right column indicates the statistical support for this. Asterisks indicate proteins that have evolved by lineage-specific duplication in plants, followed by relocation of one copy to a new cellular compartment, thereby obscuring the ancestral location of the plant proteins. Support was standardized to the node uniting *Helicosporidium* specifically with green algal and plant plastid homologues, except where indicated by superscripted numbers as follows: (1) support for plastid

parasite, *Prototheca* (38). Recent descriptions of a plastid-like small subunit rRNA (39) and a fragment containing the elongation factor Tu and ribosomal protein genes (37) in *Helicosporidium* confirmed that, like its green algal relations, this nonphotosynthetic parasite possesses a plastid genome. The metabolic role of the *Helicosporidium* plastid is of considerable comparative interest because it is a primary plastid of green algal ancestry, while the cryptic plastid of *Plasmodium* is a secondary plastid derived from the red algal lineage. For this study, we conducted an expressed sequence tag (EST) survey of *Helicosporidium* and characterized 20 nucleus-encoded genes for plastid-targeted proteins. We show that *Helicosporidium* has retained a considerably higher degree of plastid-derived metabolic diversity than *Plasmodium*.

MATERIALS AND METHODS

Identification and characterization of putative plastid-targeted proteins. Genes encoding putative plastid-targeted proteins were identified from a pool of cDNA sequences produced from an ongoing *Helicosporidium* sp. (strain AT-2000, originally isolated from the black fly, *Simulium jonesii*) EST project. Genes were identified for further characterization based on the following three criteria: (i) their association with pathways known to occur in the plastid, (ii) their phylogenetic relationship to plastid homologues in other organisms, and (iii) the presence of a presequence with hallmarks of plastid-targeting transit peptides.

Clones encoding cDNAs of putative plastid-targeted proteins were bidirectionally sequenced, and conceptual amino acid translations were produced from both unique cDNAs and from reconstructed contiguous sequences for genes with multiple cDNA clones. For genes with truncated cDNAs, some missing sequences were PCR amplified. *Helicosporidium* was cultured in Sabouraud maltose yeast medium at 25°C in the dark. Genomic DNA was isolated from harvested cells by use of a DNeasy Plant mini kit (Qiagen). PCR amplification utilized primer pairs consisting of an "anchor primer" designed from a cDNA sequence and a degenerate primer designed from plant and/or cyanobacterial homologues. PCR products were cloned into the Topo 2.1 vector (Invitrogen) and sequenced on both strands.

Phylogenetic analyses. Inferred amino acid translations of *Helicosporidium* sequences were used as queries for BLAST searching of the nonredundant and EST public databases. In cases in which no chlorophyte homologues were known, the gene of interest was sought in the *Chlamydomonas reinhardtii* unpublished genome sequence (http://www.jgi.doe.gov/). Homologues that minimized the expected scores while maximizing taxonomic sampling were selected from the BLAST output. Amino acid sequences were aligned with CLUSTALX (40) and were manually inspected. Ambiguously aligned regions were not considered in the phylogenetic analyses. The alignments are available upon request.

Phylogenetic trees were produced from amino acid sequence alignments by maximum likelihood (ML) and ML distance methods. ML distances were calculated with Tree-Puzzle 5.1 (35), using the VT model of amino acid substitution, and were corrected for among-site rate variation, which was approximated by a discrete gamma distribution with eight rate categories plus an invariable rate category. The shape parameter (α) and the probability of site-invariability parameter (i) were estimated from the data. ML distance trees were constructed with BIONJ (14) and Weighbor (8) software. Support for ML distance trees was obtained by bootstrapping (100 replicates) with Puzzleboot (A. Roger and M. Holder [www.tree-puzzle.de]). ML trees were constructed with PhyML (15), using the Dayhoff model of amino acid substitution. The among-site rate variation was modeled with a discrete gamma distribution for variable sites (with four substitution rate categories and the alpha parameter estimated from the data) and a proportion of invariable sites (estimated from the data). The significance of the branches was measured with 100 bootstrapped replicates for ML trees.

Presequence characterization. Alignments showed that several *Helicosporidium* proteins contained extensions at the N termini of the polypeptides compared to their bacterium- and organelle-encoded homologues. Moreover, such

and cyanobacteria as a whole; (2) support for *Helicosporidium* and green algae; (3) in the ferrodoxin protein family, there is a distinct plastid type that includes the *Helicosporidium* sequence, but the position of *Helicosporidium* within the plastid type is not resolved.
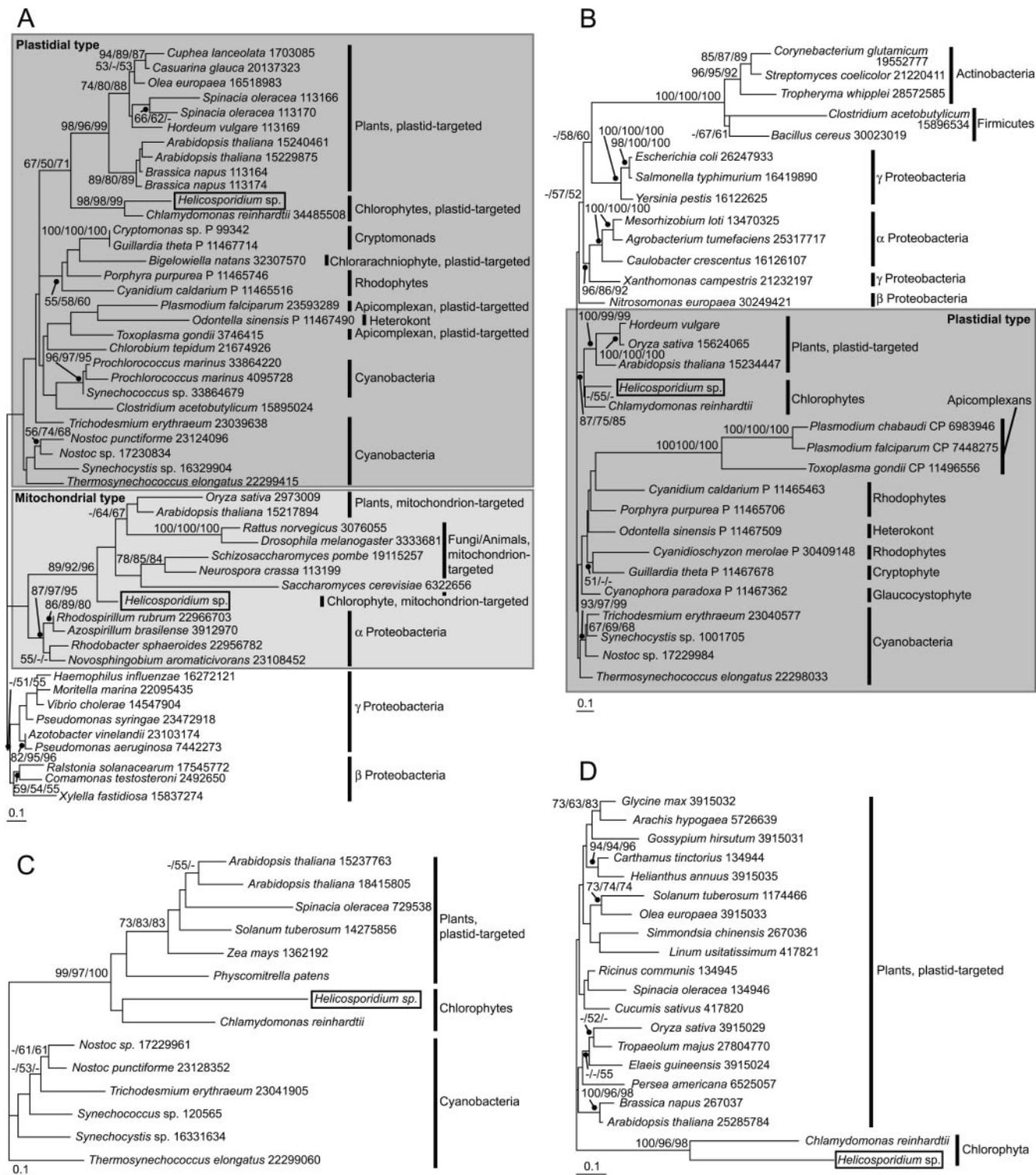
FIG. 2. Protein maximum likelihood phylogenies of four proteins showing different types of support for plastid-targeted proteins in *Helicosporidium*. Major groups are indicated to the right, and numbers at nodes correspond to bootstrap support (%) from (left to right) maximum likelihood, weighted neighbor joining, and BioNJ analyses. GenBank GI numbers for sequences are given after the species names; sequences without GI numbers were assembled from the National Center for Biotechnology Information EST database or the *C. reinhardtii* genome (http://www.jgi.doe.gov). A "P" following a sequence name denotes a gene that is carried in the plastid. (A) ACP phylogeny shows strong support (98 to 99%) for a relationship between the *Helicosporidium* protein and the plastid-targeted homologue from the green alga *Chlamydomonas*. The *Helicosporidium* mitochondrial ACP is also shown. (B) SufB phylogeny, showing the relationship between the *Helicosporidium* protein and plastid-targeted proteins in green algae and plants. Other algal SufBs are plastid encoded, and the protein is not known to exist in eukaryotes outside of this association with the plastid. (C) Ferredoxin-thioredoxin reductase subunit A is only known to exist in plastids and cyanobacteria, and the *Helicosporidium* homologue branches strongly (97 to 100%) with the plastids. (D) Stearoyl-ACP desaturase is only known to exist in plastids, and the *Helicosporidium* homologue is closely related to the *Chlamydomonas* homologue (100%).
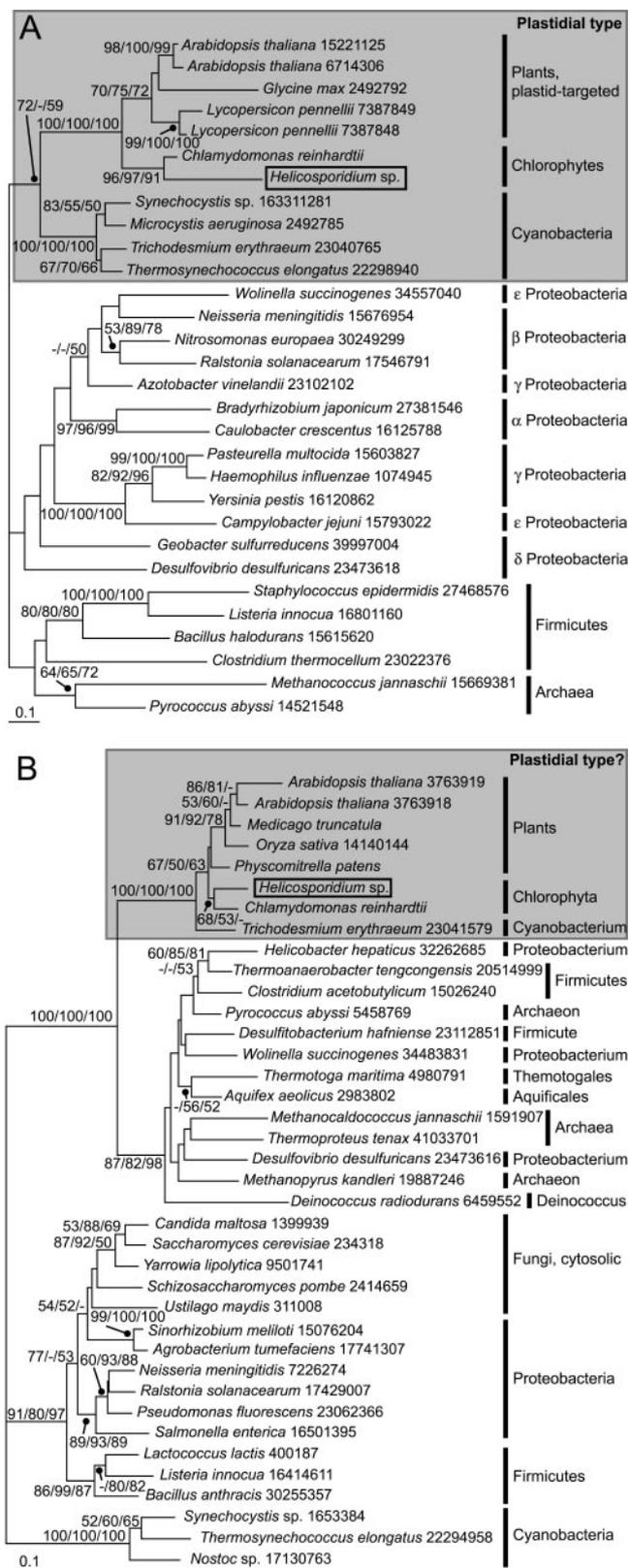
FIG. 3. The leucine biosynthetic pathway is present in the *Helicosporidium* plastid. Protein maximum likelihood phylogenies for the isopropylmalate isomerase small subunit (A) and isopropylmalate synthase (B) show that the *Helicosporidium* proteins are closely related to the green algal homologues, and plastid-targeted proteins as a whole

extensions were also present in eukaryotic nucleus-encoded homologues that were targeted to organelles. The extensions were analyzed with TargetP (11) and iPSORT (2) software to determine if they had characteristics that were typical of known target peptides. The amino acid compositions of the extensions were also examined. Extensions were classified as being potentially recognized by the mitochondrion or plastid import machinery based on the phylogenetic relationship of the proteins to mitochondrion- or plastid-targeted homologues. The average amino acid compositions of nine putative plastid and eight putative mitochondrion targeting peptides were compared to the average amino acid compositions of the remainder of the respective proteins, and Wilcoxon signed-rank tests were used to detect significant differences.

Nucleotide sequence accession numbers. Forty-one new sequences have been deposited in GenBank under accession numbers AY596480 to AY596508 and AY596510 to AY596521.

RESULTS

Plastid-targeted proteins can be identified by the following three criteria: their phylogenetic relationship to cyanobacterial and other plastid homologues, the presence of an N-terminal transit peptide, and the inferred function of the protein (as part of a common plastid pathway). These criteria are likely very effective at identifying genes derived from the plastid endosymbiont whose products function in other known plastids. An initial survey of 1,360 *Helicosporidium* EST sequences representing 341 unique and identifiable genes revealed 36 potentially plastid-related proteins that fulfilled one or more of these criteria. Preliminary phylogenetic analyses immediately identified 16 of these as mitochondrial, cytosolic, or of an uncertain type, and these were not analyzed further. Some of the remaining 20 cDNAs were truncated, and additional sequence data for 6 of these were acquired by PCRs from genomic DNA.

To determine the likelihood that each of these genes functions in the plastid, we performed comprehensive phylogenetic analyses of all 20 proteins, and the results are summarized in Fig. 1. Many of these phylogenies showed a strongly supported relationship between the *Helicosporidium* gene and a plastid-targeted homologue in green algae or plants. An example of such evidence is the acyl carrier protein (ACP), which is targeted to the plastid in all plants and algae and for which two *Helicosporidium* homologues were sequenced (Fig. 2A). One *Helicosporidium* homologue is closely related to mitochondrial ACP (89 to 96%), while the other forms a very strong group with the green algal plastid ACP (98 to 99%), which is in turn closely related to plant plastid ACPs and more distantly related to those of cyanobacteria. This strongly supports the plastid origin of this protein and, by extension, its plastid location, which is also supported by the presence of a 47-amino-acid N-terminal extension that is predicted to encode a plastid transit peptide.

In other cases, the phylogeny and the known distribution of a protein combined to provide support for a plastid location. For instance, all known eukaryotic homologues of SufB are associated with the plastid (Fig. 2B). SufB is plastid encoded in red algae and organisms containing secondary plastids of red

are consistently supported at 100%. Both *Helicosporidium* proteins are predicted to encode transit peptides, altogether supporting the presence of this pathway in its cryptic plastid. The annotation and numbers at nodes are the same as those in Fig. 2.

algal origin, while green algae and plants have a plastid-targeted SufB, and no other eukaryotic SufB is known. The *Helicosporidium* gene forms a well-supported group with the plastid-targeted proteins of plants and other green algae (75 to 87%), supporting its plastid origin. Other proteins have even narrower distributions. For example, subunit A of ferredoxin-thioredoxin reductase (Fig. 2C) is only known to exist in cyanobacteria and plastids, and the *Helicosporidium* homologue encodes a predicted transit peptide and branches with the green algae within the strongly supported plastid clade (97 to 100%). Stearoyl ACP desaturase (Fig. 2D), on the other hand, is exclusively known from plastids, and the *Helicosporidium* homologue forms a strongly supported group with the other green algal homologue (100%). The green algal homologue is not clearly plastid targeted itself, and the *Helicosporidium* homologue is truncated, so while this gene is derived from the plastid, whether it is plastid targeted is equivocal.

Phylogenetic support for the plastid origin of other proteins ranges from very strong (e.g., glutamate-1-semialdehyde aminotransferase, ClpB, and dihydrolipoamide dehydrogenase) to moderate (e.g., rpL15 and phosphoserine aminotransferase) or weak. The latter category includes proteins for which the phylogeny did not adequately resolve the position of *Helicosporidium* [e.g., poly(A) binding protein] and for which multiple isoforms exist and plastid localization is not certain. These proteins have not been excluded from consideration because some encode transit peptide-like leaders, and all have a functional relationship to known plastid metabolic pathways. Whether these really are targeted to the plastid, however, will require direct localization evidence.

Complete leader sequences were available for 9 of the 20 putative plastid-targeted genes. While there is little sequence conservation among target peptides, some trends have been observed; for example, mitochondrial targeting peptides often have an excess of arginine, alanine, and serine, while the acidic amino acids are rare (11). Plastid targeting peptides have a deficit of acidic amino acids, but arginine and the hydroxylated amino acids are overabundant (7). The amino acid compositions of predicted plastid and mitochondrion targeting peptides from *Helicosporidium* were examined (see the supplemental material), and they have many of the expected features. *Helicosporidium* plastid targeting peptides have an average overabundance of arginine compared to the average amino acid composition of the mature parts of the proteins ($P = 0.01$), and they have an average deficit of acidic amino acids ($0.05 > P > 0.025$). The predicted mitochondrial targeting peptides are deficient in acidic amino acids ($P = 0.01$) and have a surplus of serine ($0.025 > P > 0.01$).

The only definite proof of plastid localization would be immunolocalization to the organelle. Nevertheless, the phylogenetic evidence shows that these proteins are plastid derived, that the full-length genes encode leaders, as expected for plastid-targeted proteins, and that all of these proteins are part of metabolic pathways that are known to operate in plastids, with most being found in other cryptic plastids as well. Moreover, the plastid genomic data showed that *Helicosporidium* does have a plastid. While it remains possible that these proteins function in some other compartment, such a case would be unprecedented in plastid evolution. Plastid-derived enzymes have only once been shown to take up a function in another

compartment, and in that case the original plastid function was retained as well (6). Even in *Plasmodium*, plastid-derived proteins have not been recruited to function in other compartments, so there is no reason to suspect that such an event takes place in *Helicosporidium*.

DISCUSSION

Functional implications of plastid-targeted proteins in *Helicosporidium* sp. As expected for the nonphotosynthetic *Helicosporidium*, no components of the photosynthetic machinery were found. However, some of the major metabolic roles of plastids were represented by the proteins that we characterized. These putative plastid proteins are parts of pathways that have interesting implications for the metabolism of *Helicosporidium* and the role of its cryptic plastid.

Fatty acid metabolism. Evidence for fatty acid metabolism is provided by the presence of ACP, which carries the elongating acyl chain through fatty acid biosynthesis (18). In animals and fungi, ACP is part of a multifunctional, cytosolic, type I fatty acid synthase. In bacteria and plastids, fatty acids are synthesized by a multisubunit type II fatty acid synthase, in which ACP is a small stand-alone protein (18). Such a type II complex also operates in the mitochondria of some eukaryotes (43). *Helicosporidium* encodes both plastid and mitochondrial ACP (Fig. 2A), indicating that some fatty acid metabolism occurs in both compartments. Whether this activity consists of de novo fatty acid biosynthesis or simply a modification of fatty acids that are synthesized elsewhere cannot be concluded until additional components of a type II fatty acid synthase complex are also identified.

The primary products of type II fatty acid biosynthesis are the unsaturated fatty acids palmitoyl-ACP and stearoyl-ACP. In plant plastids, stearoyl-ACP is further modified to oleoyl-ACP (a major precursor of membrane glycerolipids and polyunsaturated fatty acids) by the enzyme stearoyl-ACP desaturase (18). To date, stearoyl-ACP desaturase has only been identified in higher plants (41), but we found homologues in *Helicosporidium* and *Chlamydomonas* (Fig. 2D), suggesting that their plastids play an integral role in fatty acid modification.

In plastids, fatty acids are synthesized from acetyl-coenzyme A produced by the pyruvate dehydrogenase complex (3, 21). The plastid pyruvate dehydrogenase complex is made up of four subunits, which are called E1α, E1β, E2, and E3. *Helicosporidium* encodes plastid-derived dihydrolipoamide dehydrogenase (E3), implying that the *Helicosporidium* plastid metabolizes pyruvate and can provide acetyl-coenzyme A for the synthesis of fatty acids.

Tetrapyrrole biosynthesis. Tetrapyrroles include hemes, which serve as prosthetic groups of respiratory enzymes, and chlorophyll, the major light-harvesting pigment for photosynthesis. All tetrapyrroles are formed from δ-aminolevulinic acid. In animals, fungi, and alpha proteobacteria, δ-aminolevulinic acid is produced through the Shemin pathway, while in most eubacteria, archaea, plants, and algae, it is produced through the C5 pathway (29). In plastid-containing eukaryotes, there is evidence for the Shemin pathway in the mitochondrion and the C5 pathway in the plastid (20, 27, 30, 44), although the plastid C5 pathway seems to be responsible for most tetrapyrrole syn-

thesis (29). *Helicosporidium* encodes a glutamate-1-semialdehyde aminomutase, which catalyzes the third step in the C_5 pathway. In contrast, *Plasmodium* appears to use the mitochondrial Shemin pathway to produce δ-aminolevulinic acid, which is then exported to the apicoplast for further steps in heme biosynthesis (31). *Helicosporidium* tetrapyrrole biosynthesis may be more similar to that of higher plants.

Isoprenoid biosynthesis. Isoprenoids, such as terpenes, dolichols, cholesterol, ubiquinone, chlorophyll, and carotenoids, are compounds containing isoprene subunits. The isoprene precursors, isopentenyl diphosphate and dimethylallyl diphosphate, are synthesized by either the mevalonate or deoxyxylulose phosphate (DOXP) pathway. Animals, fungi, some eubacteria, and archaea use the mevalonate pathway, while most eubacteria, including the cyanobacteria, use the DOXP pathway (10). Plants use the mevalonate pathway in the cytosol and the DOXP pathway in plastids (10). *Plasmodium falciparum* also uses the DOXP pathway in the apicoplast, but it apparently lacks the mevalonate pathway (31). The final step of the DOXP pathway is catalyzed by IspH (10, 32), and the presence of this enzyme in *Helicosporidium* (Fig. 1) suggests that the DOXP pathway functions in the plastid.

Amino acid biosynthesis. The *Helicosporidium* plastid appears to be involved in the synthesis of several amino acids, of which the leucine pathway is the most interesting. Leucine biosynthesis only takes place in bacteria, archaea, fungi, and plants (23, 42), and for plants, recent evidence suggests that it occurs in the plastid (16, 17). The first committed step of leucine synthesis is catalyzed by isopropylmalate synthase (IPMS). *Helicosporidium* encodes a plastid-derived IPMS homologue (Fig. 3A) with an N-terminal leader that is predicted to be a plastid transit peptide. The second committed step in leucine biosynthesis is catalyzed by isopropylmalate isomerase (IPMI). In prokaryotes and plant plastids, IPMI is comprised of a large and a small subunit (36, 42), whereas in fungi, the two subunits are fused and function in the cytoplasm (23). *Helicosporidium* encodes the small subunit of IPMI, which forms a well-supported clade with plant and *Chlamydomonas* plastid-targeted homologues (Fig. 3B). While the localization of these proteins has not been demonstrated directly in plants or chlorophytes, all of these proteins encode N-terminal leaders that are predicted to be transit peptides, and isolated plant chloroplasts produce leucine when supplied with the substrate for IPMS (17).

There is also evidence for the metabolism of other amino acids. Phosphoserine aminotransferase catalyzes the second step in a pathway from 3-phosphoglycerate to serine. This pathway is cytosolic in fungi and animals, but it is plastidic in plants (19). The *Helicosporidium* phosphoserine aminotransferase has an N-terminal extension and is phylogenetically related to plant homologues. Lastly, one *Helicosporidium* transcript is highly similar to a gene for an amino acid aminotransferase. This gene is found in plants, bacteria, and archaea, but its function is only known for *Arabidopsis* and rice, in which it has been shown to have aminotransferase activity with a high affinity for lysine (34). The *Helicosporidium* sequence is closely related to those of rice and *Arabidopsis* as well as those of other plants and a chlorophyte. The plant and chlorophyte sequences all possess N-terminal extensions that are predicted to be plastid transit peptides, and the enzyme has

been shown to be localized to the plastid of *Arabidopsis* (34). In summary, the presence of genes involved in the synthesis of three different amino acids provides strong evidence that the plastid is an integral part of amino acid metabolism in *Helicosporidium*.

Reducing potential. Ferredoxins are small proteins that are involved in a variety of redox reactions. The best-known role of ferredoxin is in photosynthesis, during which it is oxidized by the electron transport chain and then reduces NADP or thioredoxin. Ferredoxin can also be reduced by NADPH, and the reduced ferredoxin acts as a cofactor for several plastid enzymes, such as those involved in nitrate metabolism (12, 26), sulfite reduction (46), and fatty acid desaturation (33). The oxidation of NADPH and reduction of ferredoxin are dominant in nonphotosynthetic plastids (28). *Helicosporidium* encodes a ferredoxin protein with an N-terminal transit peptide whose closest relatives include plastid-targeted ferredoxins from plants, chlorophytes, and apicomplexa, plastid-encoded ferredoxins from rhodophytes and other algae, and ferredoxins from cyanobacteria. The *Helicosporidium* plastid ferredoxin likely provides reducing power for ferredoxin-dependent enzymes, such as stearoyl ACP desaturase, which is one of the enzymes that has been shown to be ferredoxin dependent in nonphotosynthetic plastids of plants (33).

Ferredoxin can also reduce thioredoxin through the action of ferredoxin-thioredoxin reductase, an enzyme that, up until now, had only been found in photosynthetic eukaryotes and cyanobacteria (4). Plastid thioredoxin is a key cofactor in photosynthesis and light-regulates the redox state of several plastid enzymes. Thioredoxin has been detected in nonphotosynthetic plant plastids, where it may activate enzymes in fatty acid biosynthesis, nitrate metabolism, isoprenoid biosynthesis, tetrapyrrole biosynthesis, and sulfur metabolism (1, 4). The *Helicosporidium* ferredoxin-thioredoxin reductase subunit A encodes a leader and branches with plant enzymes and an uncharacterized *Chlamydomonas* gene (Fig. 2C), but its role in the cryptic plastid is uncertain because of the number of potential targets for a plastid thioredoxin.

Lastly, iron-sulfur clusters are cofactors that bind various enzymes and mediate redox reactions. Two of the putative plastid-targeted proteins in *Helicosporidium* are the Fe-S metalloenzymes IspH and ferredoxin (31). In plastids, iron-sulfur clusters are thought to be synthesized by a pathway that is homologous to one found in bacteria (45) and that involves SufB. Red algae and related plastids encode SufB in the plastid genome, while green algae and plants encode plastid-targeted enzymes, and localization has been demonstrated in *Arabidopsis* (25). The *Helicosporidium* SufB protein is related to plant and chlorophyte homologues (Fig. 2B), as expected of a plastid-targeted enzyme.

Comparative plastid reduction in obligate parasites. The range of putative plastid-targeted proteins in *Helicosporidium* implies that many biosynthetic pathways may play important roles in the parasite's plastid. Some of these, such as fatty acid, tetrapyrrole, and isoprenoid biosynthesis, are thought to represent the core functions of the cryptic plastid of the malaria parasite, *Plasmodium* (31). Other housekeeping processes also seem to be present in both *Plasmodium* and *Helicosporidium*, such as the generation of reducing power, chaperone activity, and translation. Of the pathways they do share, the two cryptic

plastids seem to do some things differently. For example, while both parasites synthesize tetrapyrroles in the plastid, *Helicosporidium* may do so from endogenously synthesized precursors, while *Plasmodium* has lost this pathway and imports precursors from the mitochondrion (31).

The cryptic *Helicosporidium* plastid is probably also metabolically more diverse and retains several pathways that have been lost by *Plasmodium*, such as amino acid metabolic pathways. Two evolutionary factors may explain this. (i) A selective force may be acting to maintain pathways in the *Helicosporidium* plastid. *Helicosporidium* can survive in a cyst form outside its host and thus might need more metabolic autonomy than *Plasmodium*, which is entirely host associated throughout its life cycle. (ii) The increased complexity of the *Helicosporidium* plastid may reflect its more recent autotrophic ancestor. If plastid reduction is a continuous process, then *Plasmodium* may simply be further along in the process of reducing a full-functioning plastid to the highly specialized relict we see today.

ACKNOWLEDGMENTS

REFERENCES

1. **Balmer, Y., A. Koller, G. del Val, W. Manieri, P. Schurmann, and B. B. Buchanan.** 2003. Proteomics gives insight into the regulatory function of chloroplast thioredoxins. Proc. Natl. Acad. Sci. USA **100:**370–375.

2. **Bannai, H., Y. Tamada, O. Maruyama, K. Nakai, and S. Miyano.** 2002. Extensive feature detection of N-terminal protein sorting signals. Bioinformatics **18:**298–305.

3. **Bao, X., M. Focke, M. Pollard, and J. Ohlrogge.** 2000. Understanding in vivo carbon precursor supply for fatty acid synthesis in leaf tissue. Plant J. **22:**39–50.

4. **Baumann, U., and J. Juttner.** 2002. Plant thioredoxins: the multiplicity conundrum. Cell. Mol. Life Sci. **59:**1042–1057.

5. **Boucias, D. G., J. J. Becnel, S. E. White, and M. Bott.** 2001. In vivo and in vitro development of the protist Helicosporidium sp. J. Eukaryot. Microbiol. **48:**460–470.

6. **Brinkmann, H., and W. Martin.** 1996. Higher-plant chloroplast and cytosolic 3-phosphoglycerate kinases: a case of endosymbiotic gene replacement. Plant Mol. Biol. **30:**65–75.

7. **Bruce, B. D.** 2001. The paradox of plastid transit peptides: conservation of function despite divergence in primary structure. Biochim. Biophys. Acta **1541:**2–21.

8. **Bruno, W. J., N. D. Socci, and A. L. Halpern.** 2000. Weighted neighbor joining: a likelihood-based approach to distance-based phylogeny reconstruction. Mol. Biol. Evol. **17:**189–197.

9. **Cline, K., and R. Henry.** 1996. Import and routing of nucleus-encoded chloroplast proteins. Annu. Rev. Cell Dev. Biol. **12:**1–26.

10. **Cunningham, F. X., T. P. Lafond, and E. Gantt.** 2000. Evidence of a role for LytB in the nonmevalonate pathway of isoprenoid biosynthesis. J. Bacteriol. **182:**5841–5848.

11. **Emanuelsson, O., H. Nielsen, S. Brunak, and G. von Heijne.** 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J. Mol. Biol. **300:**1005–1016.

12. **Emes, M. J., and H. E. Neuhaus.** 1997. Metabolism and transport in non-photosynthetic plastids. J. Exp. Bot. **48:**1995–2005.

13. **Gardner, M. J., D. H. Williamson, and R. J. M. Wilson.** 1991. A circular DNA in malaria parasites encodes an RNA polymerase like that of prokaryotes and chloroplasts. Mol. Biochem. Parasitol. **44:**115–123.

14. **Gascuel, O.** 1997. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Mol. Biol. Evol. **14:**685–695.

15. **Guindon, S., and O. Gascuel.** 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. **52:**696–704.

16. **Hagelstein, P., and G. Schultz.** 1993. Leucine synthesis in spinach chloroplasts: partial characterization of 2-isopropylmalate synthase. Biol. Chem. Hoppe-Seyler **374:**1105–1108.

17. **Hagelstein, P., B. Sieve, M. Klein, H. Jans, and G. Schultz.** 1997. Leucine synthesis in chloroplasts: leucine/isoleucine aminotransferase and valine aminotransferase are different enzymes in spinach chloroplasts. J. Plant Physiol. **150:**23–30.

18. **Harwood, J. L.** 1996. Recent advances in the biosynthesis of plant fatty acids. Biochim. Biophys. Acta **1301:**7–56.

19. **Ho, C. L., M. Noji, M. Saito, M. Yamazaki, and K. Saito.** 1998. Molecular characterization of plastidic phosphoserine aminotransferase in serine biosynthesis from Arabidopsis. Plant J. **16:**443–452.

20. **Iida, K., I. Mimura, and M. Kajiwara.** 2002. Evaluation of two biosynthetic pathways to delta-aminolevulinic acid in Euglena gracilis. Eur. J. Biochem. **269:**291–297.

21. **Ke, J., R. H. Behal, S. L. Back, B. J. Nikolau, E. S. Wurtele, and D. J. Oliver.** 2000. The role of pyruvate dehydrogenase and acetyl-coenzyme A synthetase in fatty acid synthesis in developing arabidopsis seeds. Plant Physiol. **123:**497–508.

22. **Keilin, D.** 1921. On the life history of Helicosporidium parasiticum n. g., n. sp., a new species of protist parasite in the larvae of *Dashelaea obscura* Winn (Diptera:Ceratopogonidae) and in some arthropods. Parasitology **13:**97–113.

23. **Kohlhaw, G. B.** 2003. Leucine biosynthesis in fungi: entering metabolism through the back door. Microbiol. Mol. Biol. Rev. **67:**1–15.

24. **McFadden, G. I.** 2001. Primary and secondary endosymbiosis and the origin of plastids. J. Phycol. **37:**951–959.

25. **Møller, S. G., T. Kunkel, and N.-H. Chua.** 2001. A plastidic ABC protein involved in intercompartmental communication of light signaling. Genes Dev. **15:**90–103.

26. **Neuhaus, H. E., and M. J. Emes.** 2000. Nonphotosynthetic metabolism in plastids. Annu. Rev. Plant Physiol. Plant Mol. Biol. **51:**111–140.

27. **Ohhama, T., H. Seto, N. Otake, and S. Miyachi.** 1982. C-13-NMR evidence for the pathway of chlorophyll biosynthesis in green-algae. Biochem. Biophys. Res. Commun. **105:**647–652.

28. **Onda, Y., T. Matsumura, Y. Kimata-Ariga, H. Sakakibara, T. Sugiyama, and T. Hase.** 2000. Differential interaction of maize root ferredoxin:NADP(+) oxidoreductase with photosynthetic and non-photosynthetic ferredoxin isoproteins. Plant Physiol. **123:**1037–1045.

29. **Papenbrock, J., and B. Grimm.** 2001. Regulatory network of tetrapyrrole biosynthesis—studies of intracellular signalling involved in metabolic and developmental control of plastids. Planta **213:**667–681.

30. **Porra, R. J., O. Klein, and P. E. Wright.** 1983. The proof by C-13-NMR spectroscopy of the predominance of the C-5 pathway over the shemin pathway in chlorophyll biosynthesis in higher-plants and of the formation of the methyl-ester group of chlorophyll from glycine. Eur. J. Biochem. **130:**509–516.

31. **Ralph, S. A., G. G. van Dooren, R. F. Waller, M. J. Crawford, M. J. Fraunholz, B. J. Foth, C. J. Tonkin, D. S. Roos, and G. I. McFadden.** 2004. Metabolic maps and functions of the Plasmodium falciparum apicoplast. Nat. Rev. Microbiol. **2:**203–216.

32. **Rohdich, F., S. Hecht, K. Gartner, P. Adam, C. Krieger, S. Amslinger, D. Arigoni, A. Bacher, and W. Eisenreich.** 2002. Studies on the nonmevalonate terpene biosynthetic pathway: metabolic role of IspH (LytB) protein. Proc. Natl. Acad. Sci. USA **99:**1158–1163.

33. **Schultz, D. J., M. C. Suh, and J. B. Ohlrogge.** 2000. Stearoyl-acyl carrier protein and unusual acyl-acyl carrier protein desaturase activities are differentially influenced by ferredoxin. Plant Physiol. **124:**681–692.

34. **Song, J. T., H. Lu, and J. T. Greenberg.** 2004. Divergent roles in Arabidopsis thaliana development and defense of two homologous genes, aberrant growth and death2 and AGD2-like defense response protein 1, encoding novel aminotransferases. Plant Cell **16:**353–366.

35. **Strimmer, K., and A. vonHaeseler.** 1996. Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. Mol. Biol. Evol. **13:**964–969.

36. **Tamakoshi, M., A. Yamagishi, and T. Oshima.** 1998. The organization of the leuC, leuD and leuB genes of the extreme thermophile Thermus thermophilus. Gene **222:**125–132.

37. **Tartar, A., and D. G. Boucias.** 2004. The non-photosynthetic, pathogenic green alga Helicosporidium sp. has retained a modified, functional plastid genome. FEMS Microbiol. Lett. **233:**153–157.

38. **Tartar, A., D. G. Boucias, B. J. Adams, and J. J. Becnel.** 2002. Phylogenetic analysis identifies the invertebrate pathogen Helicosporidium sp. as a green alga (Chlorophyta). Int. J. Syst. Evol. Microbiol. **52:**273–279.

39. **Tartar, A., D. G. Boucias, J. J. Becnel, and B. J. Adams.** 2003. Comparison of plastid 16S rRNA (rrn 16) genes from Helicosporidium spp.: evidence supporting the reclassification of Helicosporidia as green algae (Chlorophyta). Int. J. Syst. Evol. Microbiol. **53:**1719–1723.

40. **Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins.** 1997. The CLUSTAL_X windows interface: flexible strategies for

multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25:**4876–4882.

41. **Tocher, D. R., M. J. Leaver, and P. A. Hodgson.** 1998. Recent advances in the biochemistry and molecular biology of fatty acyl desaturases. Prog. Lipid Res. **37:**73–117.

42. **Velasco, A. M., J. I. Leguina, and A. Lazcano.** 2002. Molecular evolution of the lysine biosynthetic pathways. J. Mol. Evol. **55:**445–459.

43. **Wada, H., D. Shintani, and J. Ohlrogge.** 1997. Why do mitochondria synthesize fatty acids? Evidence for involvement in lipoic acid production. Proc. Natl. Acad. Sci. USA **94:**1591–1596.

44. **Weinstein, J. D., and S. I. Beale.** 1983. Separate physiological roles and subcellular compartments for 2 tetrapyrrole biosynthetic pathways in Euglena-gracilis. J. Biol. Chem. **258:**6799–6807.

45. **Wilson, R. J., K. Rangachari, J. W. Saldanha, L. Rickman, R. S. Buxton, and J. F. Eccleston.** 2003. Parasite plastids: maintenance and functions. Philos. Trans. R. Soc. Lond. B **358:**155–162.

46. **Yonekura-Sakakibara, K., Y. Onda, T. Ashikari, Y. Tanaka, T. Kusumi, and T. Hase.** 2000. Analysis of reductant supply systems for ferredoxin-dependent sulfite reductase in photosynthetic and nonphotosynthetic organs of maize. Plant Physiol. **122:**887–894.