Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 2742–2746, April 1992
Evolution

Sequence of *Prochloron didemni atpBE* and the inference of chloroplast origins

(endosymbiosis/Prochlorophyta/phylogeny/ATP synthase/cyanobacteria)

P. J. LOCKHART[*], T. J. BEANLAND[†], C. J. HOWE[†‡], AND A. W. D. LARKUM[*]

[*]School of Biological Sciences, Macleay Building A12, University of Sydney, NSW 2006, Sydney, Australia; and [†]Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW U.K.

Communicated by Hans Kornberg, December 2, 1991

ABSTRACT **The prochlorophytes, oxygenic photosynthetic prokaryotes containing chlorophylls *a* and *b*, have been put forward as descended from the organisms that gave rise to chloroplasts of green plants and algae by endosymbiosis, although this has always been controversial. To assess the phylogenetic position of the prochlorophyte *Prochloron didemni*, we have cloned and sequenced its *atpBE* genes. Phylogenetic inference under a range of models gives moderate to strong support for a cyanobacterial grouping rather than a chloroplast one. Possible systematic errors in this and previous analyses of prochlorophyte sequences are discussed.**

A range of evidence suggests that plastids originated by the uptake of oxygenic photosynthetic bacteria by nonphotosynthetic hosts (1, 2), although which extant bacteria are the descendants of the protoendosymbionts for the various plastid groups remains unclear (3). For green chloroplasts, which contain chlorophylls *a* and *b* and lack phycobiliproteins, prochlorophytes (oxygenic photosynthetic bacteria with the same pigments) have been suggested (4–6). Ultrastructural and biochemical data from prochlorophytes (4, 7, 8) are inconclusive, the only evidence even that Prochlorales is a monophyletic order coming from the immunochemical cross-reactivity of the chlorophyll-*a*/*b*-binding proteins of *Prochloron didemni* (*P. didemni*) and *Prochlorothrix hollandica* (*Px. hollandica*) (9).

Sequence data are available for only two of the three prochlorophyte species. For *P. didemni*, analyses of a 16S rRNA RNase T1 oligonucleotide catalogue (10) and complete 5S rRNA sequence (11) apparently rule out chloroplast ancestry, although in neither case is such a conclusion statistically excluded (12–14). In contrast, analysis of partial 16S rRNA (15) and *rbcLS* (16) sequences from *Px. hollandica* leads to robust trees placing it among the cyanobacteria. The original analysis of *Px. hollandica* D1 (*psbA*) sequences (17, 18) proposed a common ancestry with chloroplasts to the exclusion of cyanobacteria, but it lacked any estimation of tree robustness and did not utilize D2 (*psbD*) as an outgroup. Use of D2 in this way (19) suggests that the C-terminal 7-amino acid gap common to green chloroplast and *Px. hollandica* D1 proteins that was claimed to support a relationship between these groups (17, 18) may in fact be a primitive feature. Maximum likelihood analysis of *Px. hollandica psbA* data (20) by using an explicit model of amino acid substitution indicates a robust cyanobacterial affinity.

We report here the sequence of *P. didemni atpB* and *atpE* genes.[§] Adopting a hypothesis-testing approach to inference (20, 21), we find some strong evidence ($P = 0.01$) for a cyanobacterial affinity. In common with the analyses discussed, this conclusion is, however, still critically dependent on the inference models employed.

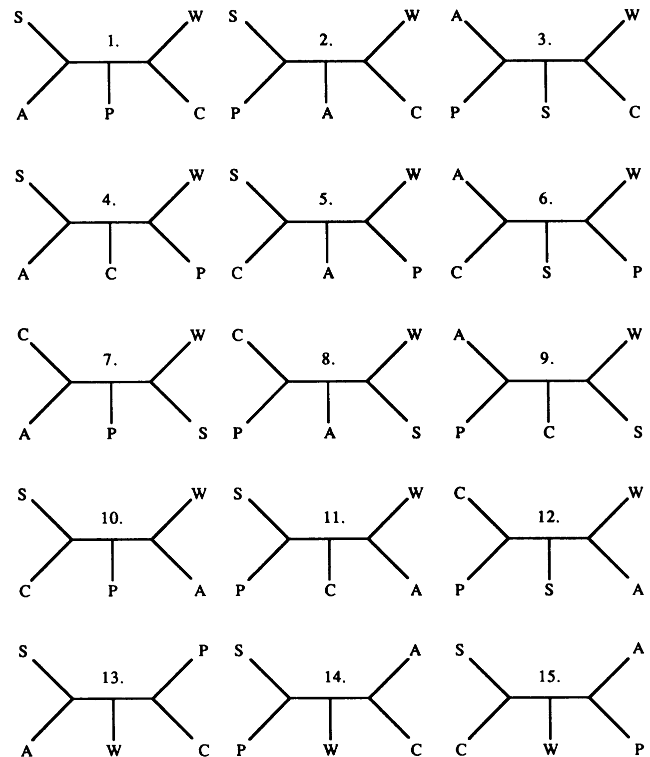

FIG. 1. Unrooted topologies for step 1 testing of the relationships between the oxygenic taxa. A, *Anabaena* sp. PCC7120; S, *Synechococcus* sp. PCC6301; P, *P. didemni*; W, wheat (*Triticum aestivum*) chloroplast; C, *C. reinhardtii* chloroplast. All the trees are unrooted; branch lengths are arbitrary.

MATERIALS AND METHODS

Cloning and Sequencing of *P. didemni atpBE*. *P. didemni* within its host ascidian *Lissoclinum patella* was collected from shallow water around Heron and One Tree Islands on the Great Barrier Reef, north of Gladstone, Australia. High molecular weight genomic DNA was extracted and a genomic library was constructed in λ EMBL3 as described elsewhere (14). Clones hybridizing with a probe containing the *Zea mays* chloroplast *rbcL* and partial *atpB* genes (22) were sequenced, subclones being isolated using a gene-internal probe for the *Synechococcus* sp. PCC6301 *atpB* gene (a gift from A. L. Cozens, Laboratory of Molecular Biology, Cambridge) or synthetic oligonucleotide probes (Oligonucleotide Synthesizing Facilities, Dept. of Biochemistry, University of Cambridge, and School of Biological Sciences, Macquarie

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

[‡]To whom reprint requests should be addressed.
[§]The sequence reported in this paper has been deposited in the GenBank data base (accession no. M86384).

```
1                                                          60
ATGGTAGCGA CAACAGAAAC AACAAACATT GGTAAAATTA CCCAAATCAT CGGCCCTGTA
 M  V  A    T  T  E  T    T  N  I    G  K  I    T  Q  I  I    G  P  V

61                                                        120
GTGGATGCGG AGTTTCCATC TGGCAAAATG CCCCGAATCT ACAATGCCTT GAGAGTTGAA
 V  D  A    E  F  P  S    G  K  M    P  R  I    Y  N  A  L    R  V  E

121                                                       180
GGCAAAAATG CCGCCGGACA AGATGTAGCC GTAACCTGCG AGGTGCAGCA GTTGCTGGGA
 G  K  N    A  A  G  Q    D  V  A    V  T  C    E  V  Q  Q    L  L  G

181                                                       240
GACAACCAAG TACGGGCTGT TTCCATGAGC AGCACGGACG GTCTGGTGCG GGGAATGGAA
 D  N  Q    V  R  A  V    S  M  S    S  T  D    G  L  V  R    G  M  E

241                                                       300
ATTACCGATA CTGGCGCACC CATTAACGTT CCTGTGGGCA AGGCTACCCT GGGTCGGATT
 I  T  D    T  G  A  P    I  N  V    P  V  G    K  A  T  L    G  R  I

301                                                       360
TTCAATATCT TGGGGGAACC AGTAGATAAT CAGGGTCCTG TGTATACTGC TGAAACTTCT
 F  N  I    L  G  E  P    V  D  N    Q  G  P    V  Y  T  A    E  T  S

361                                                       420
CCTATTCACC GAGCTGCCCC TAAATTTACC GATTTAGACA CCAAGCCCAC TGTATTTGAG
 P  I  H    R  A  A  P    K  F  T    D  L  D    T  K  P  T    V  F  E

421                                                       480
ACTGGGATCA AGGTTATCGA CTTGCTGACT CCCTATCGTC GCGGCGGTAA AATCGGCCTG
 T  G  I    K  V  I  D    L  L  T    P  Y  R    R  G  G  K    I  G  L

481                                                       540
TTTGGCGGTG CTGGTGTGGG CAAAACCGTT ATCATGATGG AGCTAATCAA CAATATTGCC
 F  G  G    A  G  V  G    K  T  V    I  M  M    E  L  I  N    N  I  A

541                                                       600
ATCAACCACG GTGGAGTCTC CGTCTTCGGC GGTGTGGGAG AGCGCACTCG TGAAGGGAAT
 I  N  H    G  G  V  S    V  F  G    G  V  G    E  R  T  R    E  G  N

601                                                       660
GACCTTTACA ATGAAATGAT TGAATCGAAG GTTATTAACG CTGATAACCT CAACGAGTCT
 D  L  Y    N  E  M  I    E  S  K    V  I  N    A  D  N  L    N  E  S

661                                                       720
AAAATTGCTC TAGTTTACGG TCAGATGAAT GAACCCCCTG GTGCGAGAAT GCGGGTAGGT
 K  I  A    L  V  Y  G    Q  M  N    E  P  P    G  A  R  M    R  V  G

721                                                       780
CTATCTGCTC TGACTATGGC TGAGTATTTC CGGGATGTGA ACAAGCAAGA TGTGTTGCTG
 L  S  A    L  T  M  A    E  Y  F    R  D  V    N  K  Q  D    V  L  L

781                                                       840
TTCATCGACA ATATTTTTCG CTTTGTTCAA GCTGGTTCTG AGGTATCTGC CCTGTTAGGT
 F  I  D    N  I  F  R    F  V  Q    A  G  S    E  V  S  A    L  L  G

841                                                       900
CGCATGCCTT CTGCTGTGGG TTACCAGCCT ACTCTGGGTA CTGACGTGGG AGATTTGCAA
 R  M  P    S  A  V  G    Y  Q  P    T  L  G    T  D  V  G    D  L  Q

901                                                       960
GAGCGGATTA CTTCTACTAA GGAAGGTTCT ATTACCTCTA TTCAAGCGGT TTACGTTCCT
 E  R  I    T  S  T  K    E  G  S    I  T  S    I  Q  A  V    Y  V  P

961                                                      1020
GCGGACGATT TAACCGACCC CGCTCCTGCT ACTACTTTTG CTCACTTAGA CGGTACTACG
 A  D  D    L  T  D  P    A  P  A    T  T  F    A  H  L  D    G  T  T

1021                                                     1080
GTGCTGTCTC GGGGTTTGGC TTCTAAGGGT ATTTATCCTG CTGTAGATCC TTTAGACTCC
 V  L  S    R  G  L  A    S  K  G    I  Y  P    A  V  D  P    L  D  S

1081                                                     1140
ACCAGCACCA TGTTACAGGC GGGAATTGTG GGTGAAGACC ACTACAATAC CGCTCGTGCA
 T  S  T    M  L  Q  A    G  I  V    G  E  D    H  Y  N  T    A  R  A

1141                                                     1200
GTGCAGTCTA CCTTGCAGCG CTATAAAGAA CTGCAAGATA TTATTGCCAT TTTGGGTCTG
 V  Q  S    T  L  Q  R    Y  K  E    L  Q  D    I  I  A  I    L  G  L

1201                                                     1260
GATGAATTGT CGGAAGAAGA CCGCTTGATA GTAGATCGGG CTCGGAAGGT GGAGCGTTTC
 D  E  L    S  E  E  D    R  L  I    V  D  R    A  R  K  V    E  R  F

1261                                                     1320
TTGTCTCAGC CTTTCTTTGT GGCGGAAGTA TTTACTGGCG CACCTGGCAA GTACGTTTCT
 L  S  Q    P  F  F  V    A  E  V    F  T  G    A  P  G  K    Y  V  S

1321                                                     1380
CTGGAAGATA CTATCAAAGG CTTCAAGATG ATTCTGTCTG GGGAATTAGA TGACCTGCCA
 L  E  D    T  I  K  G    F  K  M    I  L  S    G  E  L  D    D  L  P

1381                                                     1440
GAACAGGCAT TCTACTTGGT AGGAGATATT CAGGAAGCTA AGGCTAAAGC TGAAAAACTC
 E  Q  A    F  Y  L  V    G  D  I    Q  E  A    K  A  K  A    E  K  L

1441                                                     1500
AAGCAAGATT AAGATCCCCC TATCCCCCCT TGATCCCCCC TTGGTCCCCC CCAGTGGGGG
 K  Q  D    *

1501                                                     1560
GGAAACAGGG GGGAGAAAAC AACCCCCCTT ATTTAAGGGG GGAGAAGAGG TTGGTTTGTC

1561                                                     1620
AGTTACTGCT TGGTAAAAAC AAACAACAAA CAACCAATAA CAAAAAACAA ACAACAAATA

1621                                                     1680
ACAAACAACC AAAAATGACT TTAACTTTGC GGGTAATTAC CCCAGATAAG ACAGTTTGGG
                M  T    L  T  L    R  V  I  T    P  D  K    T  V  W
1681                                                     1740
ACGATAGTGT AGAAGAAATT GTCCTGCCCA GTACTACGGG ACAGGTAGGG GTTTTGACAG
D  D  S  V    E  E  I    V  L  P    S  T  T  G    Q  V  G    V  L  T

1741                                                     1800
GTCACGCTCC TCTGTTAACG GCTTTGGATA CTGGGGTGAT GCGAGTTCGT CCTGGCAAAG
G  H  A  P    L  L  T    A  L  D    T  G  V  M    R  V  R    P  G  K

1801                                                     1860
ATTGGCAGGC GATCGCCCTC ATGGGTGGAT TTGCTGAAGT AGAGAACAAC GAGGTGAAAG
D  W  Q  A    I  A  L    M  G  G    F  A  E    V  E  N  N    E  V  K

1861                                                     1920
TTCTAGTGAA TGGTGCGGAA GTGGGAGATA GTATCGATAA AGAAACTGCT CGCACTGAGT
V  L  V  N    G  A  E    V  G  D    S  I  D  K    E  T  A    R  T  E

1921                                                     1980
TCCAACAAGC GGAACAAAAT CTCGCTCGAG CCAATCAAGG AGACAACCGC CAAGAGCTAA
F  Q  Q  A    E  Q  N    L  A  R    A  N  Q  G    D  N  R    Q  E  L

1981                                                     2040
TTCAAGCAAC CCAAGAGTTC AAGAAAGCAA GAGCCCGCTT TCAAGCTGCT GGGGGCATGA
I  Q  A  T    Q  E  F    K  K  A    R  A  R  F    Q  A  A    G  G  M

2041  2050
CTTAAGGCAA
T  *
```

FIG. 2. Sequence of the *atpBE* locus from *P. didemni*. The sequence is shown starting with the *atpB* initiation codon with the predicted amino acid sequence of the β and ε polypeptides (from positions 1–1452 and 1635–2045, respectively).

University). Sequencing was carried out by the dideoxynucleotide chain-termination method (23).

Data Sets for Phylogenetic Inference. The *P. didemni* predicted β and ε polypeptide sequences were aligned with homologues from chloroplasts (wheat and *Chlamydomonas reinhardtii*), cyanobacteria (*Synechococcus* sp. PCC6301 and *Anabaena* sp. PCC7120), and *Escherichia coli* (all sequences from GenBank, release 67) using MSA (24). Regions where gaps greater than three residues long are shared by two or more taxa were omitted to prevent artifacts (e.g., ref. 17). Nucleotide sequences were aligned to correspond to the amino acid alignment. For phylogenetic inference the PHYLIP package (versions 3.3 and 3.4; ref. 25) was used. PROTPARS carries out amino acid parsimony based on the genetic code, and DNAPARS nucleotide parsimony scoring all substitutions equally. DNAML was used for nucleotide maximum likelihood inference. DNADIST estimates a distance matrix from nucleotide data, from which FITCH infers a phenogram using the Fitch–Margoliash least-squares criterion (26). For the models in DNADIST and DNAML, a transition/transversion ratio was estimated from the data, and probabilities in the transition matrix were derived from this and (F option) the observed population base frequences (27). To take account of different rates of substitution at the three codon sites (28), these were assigned weights based on all possible pairwise comparisons of the five oxygenic taxa. For DNAPARS, weights ($W_1/W_2/W_3$) were derived as reciprocals of the pairwise observed substitution rates. For DNADIST and DNAML, transformed rates ($K_1/K_2/K_3$) were estimated using the Kimura two-parameter equation (28).

Hypotheses Tested. A two-step approach was used for parsimony and maximum likelihood methods (19), as exhaus-

Table 1. Estimates of the parameters used in inference programs

Gene	Weights ($W_1/W_2/W_3$)	Rates ($K_1/K_2/K_3$)	Transition/ transversion ratio
atpB	8:17:2	2.16:1.00:18.55	1.2
atpE	4:5:2	1.58:1.00:5.06	1.5

Weights are rounded to whole numbers for input to DNAPARS, and transition/transversion ratios are given to one decimal place only because of variation between different codon positions.

tive pairwise analysis in a single step would have required the testing of too many trees to be practicable. In step 1, 15 unrooted trees (Fig. 1, trees 1–15) were analyzed, an exhaustive test that requires no assumption on the monophyly of the green chloroplasts. Tree robustness for these programs was assessed by a site-by-site pairwise test (27) that determines whether the step (or support) difference between each tree and the best is significant. This requires that the site-by-site step (support) differences are normally distributed, an assumption for which there is empirical support (19). If the step (support) difference for two trees exceeds its standard error by a factor of 1.96 or greater, the trees are considered different (null hypothesis rejected) at $P = 0.05$.

Step 2 of testing takes all trees not different from the best at $P = 0.05$ (or other appropriate level) and analyzes all possible rooted versions using *E. coli* as an outgroup. Use of a "purple" bacterium (as defined by Woese; ref. 29) to root the oxygenic taxa is standard (15) and justified by comparison of the photosynthetic systems of the two groups. Sequences from all six taxa were also used together in a single-step analysis without pairwise testing to verify that the best trees identified in step 2 were the best overall.

For DNADIST/FITCH, robustness was determined empirically by bootstrapping, again using *E. coli* as an outgroup.

RESULTS AND DISCUSSION

Features of the Sequence. The sequence of 2050 base pairs (bp) of the *P. didemni atpBE* locus is shown in Fig. 2. The genes are separated by 182 bp, more than in cyanobacteria (*Synechococcus* sp. PCC6301, 70 bp; *Anabaena* sp. PCC7120, 93 bp) and in contrast to the overlapping genes of many higher plant chloroplasts (30–33). *P. didemni* lacks the $tRNA^{Met}$ gene (data not shown) found downstream of the higher plant chloroplast *atpE* gene (33, 34), although, like the size of the intergenic spacer, this does not suggest a cyanobacterial affinity as lack of the $tRNA^{Met}$ and long intergenic spacers are both primitive features (i.e., characteristic of *E. coli*) (35). Interestingly, the *P. didemni* intergenic region has four direct repeats of a 10-bp element (AACAAACAAC) containing a 7-bp sequence similar to a repetitive element in the *Anabaena* sp. PCC7120 *atpBE* intergenic region (31). The role of such elements is unclear.

Phylogenetic Inference Parameters. Estimates of parameters derived from the data set and used in the inference models are shown in Table 1. All agree with reported values (28, 36, 37) and indicate a higher degree of conservation of *atpB* than *atpE*.

***atpB*.** Step 1 testing using *atpB* sequences for all programs selected trees 1–3 (Fig. 1) for rooting. For PROTPARS, this group of trees was favored over any of trees 4–15 at $P = 0.05$.

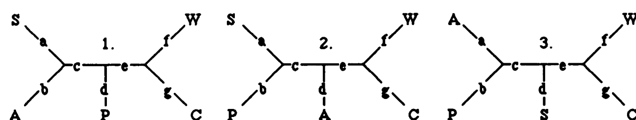

FIG. 3. Topologies for step 2 testing of the relationships between the oxygenic taxa. For trees 1–3 of Fig. 1, 21 trees a–g were produced by introducing *E. coli* as an outgroup on the branch indicated. Thus tree 1a has the outgroup on the *Synechococcus* sp. PCC6301 branch, etc. Branch lengths are arbitrary; abbreviations are as defined in Fig. 1.

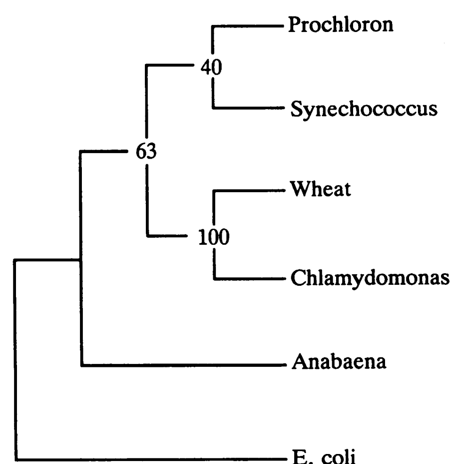

FIG. 4. Phenogram from bootstrap analysis of *atpB* sequences using DNADIST and FITCH. The numbers indicate the number of times out of 100 that the species to the right of each node were placed in a monophyletic group. The tree is rooted using *E. coli* as an outgroup. Branch lengths are arbitrary.

Testing the 21 rooted trees 1a–3g (step 2; Fig. 3) for PROTPARS gave a best tree with *P. didemni* not related to the chloroplasts to the exclusion of *Anabaena* sp. PCC7120 (Fig. 3, tree 2e). Although this tree was poorly resolved from most of the other trees tested in step 2 (suggesting no unique phylogeny is supported), it was favored significantly ($P = 0.05$) over the three topologies containing *P. didemni* ancestral to the chloroplast clade (trees 1a, 1b, and 1c). This is therefore evidence against *P. didemni* being descended from the chloroplast protoendosymbiont.

For DNAPARS, trees 1–3 (Fig. 1) were again favored over trees 4–15 (at $P = 0.15$), as they were for DNAML ($P = 0.05$). At step 2, the best topologies for both DNAPARS and DNAML were the same (Fig. 3, tree 2a) and placed *P. didemni* with the cyanobacteria. However, in neither case was this conclusion significantly favored over a chloroplast affinity. Distance matrix analysis favored tree 2d, although bootstrapping (Fig. 4) suggested that the support for this topology over a chloroplast affinity is not significant ($P = 0.76$; 24 of the 100 trees contained a *P. didemni*–chloroplast clade). Therefore, in contrast to PROTPARS, analysis of the *atpB* nucleotide sequence does not permit a robust conclusion as to the phylogeny of *P. didemni*.

***atpE*.** Step 1 testing using *atpE* sequences for PROTPARS, DNAPARS, and DNAML selected ($P = 0.05$) the same trees 1–3 as *atpB* (Fig. 1) for rooting. At step 2, rooted trees containing a *Prochloron–Anabaena–Synechococcus* clade were favored over any lacking such a grouping by PROTPARS ($P = 0.05$) and DNAPARS ($P = 0.1$) (Table 2). In neither case was the branch order within this bacterial clade determined. The best rooted DNAML tree was 1e, apparently confirming this conclusion, although under likelihood inference this tree was not signif-

Table 2. Results of inference from *atpE* sequences

Program	Best topology	Evidence for a *Prochloron–Synechococcus–Anabaena* clade, *P* value
PROTPARS	1e	0.05
DNAPARS	3e	0.1
DNADIST/FITCH	1e	0.01
DNAML	1e	NS

Best topologies are described in Fig. 3. No branch order within the *Prochloron–Synechococcus–Anabaena* clade is specified. NS, not significant.

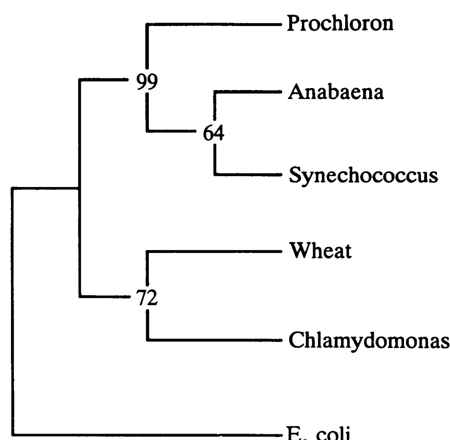

FIG. 5. Phenogram from bootstrap analysis of *atpE* sequences using DNADIST and FITCH. The numbers indicate the number of times out of 100 that the species to the right of each node were placed in a monophyletic group. The tree is rooted using *E. coli* as an outgroup. Branch lengths are arbitrary.

icantly different ($P = 0.05$) from trees 1a, 1b, and 1c containing *P. didemni*–chloroplast clades.

DNADIST/FITCH again favored tree 1e, bootstrap values (Fig. 5) strongly supporting the same *Prochloron–Anabaena–Synechococcus* clade ($P = 0.01$; only 1 of 100 replicate trees grouped the prochlorophyte specifically with the chloroplasts). This is thus strong evidence for a cyanobacterial affinity, suggesting that chlorophyll-*a/b*-binding polypeptides in chloroplasts and *P. didemni* may have arisen independently. The order within the *P. didemni*–cyanobacteria clade is once again not determined.

Overall, *atpE* thus offers evidence ranging from being merely consistent with (DNAML) to very strongly in favor of (FITCH) a cyanobacterial affinity for *P. didemni*. This conclusion is wholly consistent with findings for *atpB* and the same as that drawn from analysis of the *Px. hollandica* 16S rRNA (15), *rbcLS* (16), and *psbA* (19, 20) sequences. Hence analyses of prochlorophyte sequence data at present provide little or no evidence for a phylogeny placing prochlorophytes as most closely related to green chloroplasts.

Systematic Errors. It should be recognized, however, that a systematic error characterizes all these analyses. At present, inference from sequence data assumes explicitly or otherwise that the sequences studied have evolved with a fixed base frequency since divergence from a common ancestor (25, 38–40). Violation of this assumption is very noticeable with oxygenic taxa, whose genomes typically show a wide range of G + C contents, with chloroplast genes usually being A + T-rich and prochlorophyte sequences either intermediate in G + C content between those from plastids and cyanobacteria (e.g., *atpE*) or within the cyanobacterial range (e.g., *psbAI, psbAII, atpB, rbcLS*, and 16S rRNA) (15–18).

When sequences contain such a substitutional bias, inference has been shown to be unreliable (40–42), sequences tending to group because of shared substitution processes rather than necessarily common ancestry. This problem is compounded when long edges are present (40–42) and where the radiation occurs over a relatively short time scale, both of which may apply to the oxygenic taxa diverging after the evolution of the ability to split water (43). When the data analyzed fit the inference models so poorly, tests of tree robustness become less reliable (40–42), hence the lack of robustness for much of the data presented, and even topologies with strong statistical support become suspect (40). Until methods are developed to overcome such problems, the prochlorophyte phylogenies determined to date should be treated with caution.

We are grateful to David Judge, David Penny, Mike Hendy, and "Fast" Eddie Holmes for discussions; to Roger Hiller and Mike Weldon for synthesis of oligonucleotides; and to Cheryl Handford for technical assistance. This work has been supported by the Australian Research Council, the University of Sydney, the Science and Engineering Research Council, the Cambridge Philosophical Society, Corpus Christi College, Cambridge, and the Wellcome Foundation.

1. Raven, P. H. (1970) *Science* **169**, 641–646.
2. Whatley, J. M. & Whatley, F. R. (1981) *New Phytol.* **87**, 233–247.
3. Gray, M. W. (1989) *Trends Genet.* **5**, 294–299.
4. Lewin, R. A. (1984) *Phycologia* **23**, 203–208.
5. Burger-Wiersma, T., Veenhuis, M., Korthals, H. J., van de Wiel, C. C. M. & Mur, L. R. (1986) *Nature (London)* **320**, 262–264.
6. Chisholm, S. W., Olsen, R. J., Zettler, E. R., Goericke, R., Waterbury, J. B. & Welschmeyer, N. A. (1988) *Nature (London)* **334**, 340–342.
7. Cox, G. & Dwarte, D. M. (1981) *New Phytol.* **88**, 427–438.
8. Golecki, J. R. & Jürgens, U. W. (1989) *Arch. Microbiol.* **152**, 77–82.
9. Bullerjahn, G. S., Matthijs, H. C. P., Mur, L. R. & Sherman, L. A. (1990) *FEMS Microbiol. Lett.* **67**, 99–106.
10. Seewaldt, E. & Stackebrandt, E. (1982) *Nature (London)* **295**, 618–620.
11. Mackay, R. M., Salgado, D., Bonen, L., Stackebrandt, E. & Doolittle, W. F. (1982) *Nucleic Acids Res.* **10**, 2963–2970.
12. van Valen, L. M. (1982) *Nature (London)* **298**, 493–494.
13. van den Eynde, H., de Baere, R., de Roeck, E., van de Peer, Y., Vandenberghe, A., Willekens, P. & de Wachter, R. (1988) *J. Mol. Evol.* **27**, 126–132.
14. Lockhart, P. J., Larkum, A. W. D., Beanland, T. J. & Howe, C. J. (1989) *FEBS Lett.* **248**, 127–130.
15. Turner, S., Burger-Wiersma, T., Giovannoni, S. J., Mur, L. R. & Pace, N. R. (1989) *Nature (London)* **337**, 380–382.
16. Morden, C. W. & Golden, S. S. (1991) *J. Mol. Evol.* **32**, 379–395.
17. Morden, C. W. & Golden, S. S. (1989) *Nature (London)* **337**, 382–384.
18. Morden, C. W. & Golden, S. S. (1989) *Nature (London)* **339**, 400.
19. Beanland, T. J. (1990) Ph.D. thesis (Univ. of Cambridge, Cambridge).
20. Kishino, H., Takashi, T. & Hasegawa, M. (1990) *J. Mol. Evol.* **31**, 151–160.
21. Beanland, T. J. (1990) *J. Theor. Biol.* **145**, 535–545.
22. Bedbrook, J. R., Coen, D. M., Beaton, A. R., Bogorad, L. & Rich, A. (1979) *J. Biol. Chem.* **245**, 905–910.
23. Bankier, A. T. & Barrell, B. G. (1989) in *Nucleic Acids Sequencing: A Practical Approach*, eds. Howe, C. J. & Ward, E. S. (IRL, Oxford), pp. 37–78.
24. Lipman, D. J., Altschul, S. F. & Kececiolgu, J. D. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 4412–4415.
25. Felsenstein, J. (1988) *Annu. Rev. Genet.* **22**, 521–565.
26. Fitch, W. M. & Margoliash, E. (1967) *Science* **155**, 279–284.
27. Kishino, H. & Hasegawa, M. (1989) *J. Mol. Evol.* **29**, 170–179.
28. Kimura, M. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 454–458.
29. Woese, C. R. (1987) *Microbiol. Rev.* **51**, 221–271.
30. Cozens, A. L. & Walker, J. E. (1987) *J. Mol. Biol.* **194**, 359–383.
31. Curtis, S. E. (1987) *J. Bacteriol.* **169**, 80–86.
32. Zurawski, G., Bottomley, W. & Whitfeld, P. R. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6260–6264.
33. Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B. Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, A., Shimada, H. & Sugiura, M. (1986) *EMBO J.* **5**, 2043–2049.
34. Howe, C. J., Fearnley, I. M., Walker, J. E., Dyer, T. A. & Gray, J. C. (1985) *Plant Mol. Biol.* **4**, 334–345.
35. Walker, J. E., Saraste, M. & Gay, N. J. (1984) *Biochim. Biophys. Acta* **768**, 164–200.

36. Ritland, K. & Clegg, M. T. (1987) *Am. Nat.* **130,** S74–S100.
37. Zurawski, G. & Clegg, M. T. (1987) *Annu. Rev. Plant Physiol.* **38,** 391–418.
38. Sidow, A. & Wilson, A. C. (1990) *J. Mol. Evol.* **31,** 51–68.
39. Gillespie, J. H. (1986) *Annu. Rev. Ecol. Syst.* **17,** 637–665.
40. Lockhart, P. J., Howe, C. J., Bryant, D. A., Beanland, T. J. & Larkum, A. W. D. (1992) *J. Mol. Evol.* **34,** 153–162.
41. Penny, D., Hendy, M., Zimmer, E. A. & Handby, R. K. (1990) *Aust. Syst. Bot.* **3,** 21–38.
42. Lockhart, P. J. (1990) Ph.D. thesis (Univ. of Sydney, Sydney, Australia).
43. Awramik, S. M. (1990) in *Palaeobiology, A Synthesis*, eds. Briggs, D. E. & Crowther, P. R. (Blackwell, Oxford), pp. 336–341.