J. Euk. Microbiol., 45(5), 1998 pp. 561–570

The Phylogenetic Position of Alpha- and Beta-Tubulins from the *Chlorarachnion* Host and *Cercomonas* (Cercozoa)

PATRICK J. KEELING,[1,2] JAMES A. DEANE and GEOFFREY I. McFADDEN

Plant Cell Biology Research Centre, Botany School, University of Melbourne, Parkville VIC 3052, Australia

ABSTRACT. Alpha and beta-tubulin genes from *Chlorarachnion* and an alpha-tubulin gene from *Cercomonas* have been characterised. We found the *Cercomonas* and *Chlorarachnion* alpha-tubulins to be closely related to one another, confirming the proposed relationship of these genera. In addition, the *Chlorarachnion* host and *Cercomonas* also appear to be more distantly related to Heterolobosea, Euglenozoa, chlorophytes, heterokonts, and alveolates. *Chlorarachnion* was also found to have two distinctly different types of both alpha- and beta-tubulin, one type being highly-divergent. *Chlorarachnion* contains a secondary endosymbiont of green algal origin, raising the possibility that one type of *Chlorarachnion* tubulins comes from the host and the other from the endosymbiont. Probing pulsed field-separated chromosomes showed that the highly-divergent genes are encoded by the host genome, and neither alpha- nor beta-tubulin cDNAs were found to include 5′ extensions that might serve as targeting peptides. It appears that *Chlorarachnion* has distinct and divergent tubulin paralogues that are all derived from the host lineage. One *Chlorarachnion* beta-tubulin was also found to be a pseudogene, which is still expressed but aberrantly processed. Numerous unspliced introns and deletions resulting from mis-splicing are contained in the mRNAs from this gene.

Supplementary key words. Cercozoa, Chlorarachniophyta, evolution, sarcomonad.

CHLORARACHNIOPHYTES are unicellular marine algae with a number of unusual characteristics that have attracted the attention of microscopists and molecular biologists alike. Foremost among these is the chloroplast, which has been acquired by a secondary endosymbiosis, that involved the engulfment and retention of a photosynthetic eukaryote by a phagotrophic eukaryote. This process is thought to have occurred in a number of photosynthetic eukaryotes, and is most often revealed only by the presence of more than two membranes around the plastid [13, 14, 41]. However, in two lineages, the chlorarachniophytes and cryptophytes, the reduction of the engulfed endosymbiont has not been so complete. In these organisms the plastid is associated with a small nucleus called the nucleomorph, and a small volume of endosymbiont cytoplasm. A great deal of what is now known about chlorarachniophytes relates specifically to the endosymbiont: its evolutionary origin has been identified as green algal [9, 22, 40], its gene expression and processing have been studied in some detail [16], the general organisation of its tiny genome and structure of its chromosomes has been determined [15, 16], and there is currently a *Chlorarachnion* nucleomorph genome sequencing project underway [30].

In contrast to the nucleomorph, relatively little is known about the host that engulfed this green alga. The host-component of chlorarachniophytes can be quite complex, the best studied having a life cycle including at least three distinct cell types. One is a highly motile flagellate with a true single flagellum that swims in a unique fashion with the flagellum coiled about the body, and can temporarily transform to an amoeboflagellate. A second type is a phagotrophic amoeba that forms a web of interconnected pseudopodia resulting in a large reticulate plasmodium. These cells are also unique in that they are at once photosynthetic and also phagotrophic. Walled cysts are also known, and these appear to be able to give rise to either amoebae or flagellates [20].

The complex of morphological, ultrastructural, and life-history characters that can be assembled for chlorarachniophytes does not immediately suggest an evolutionary relationship with any particular eukaryotic group, and their classification has accordingly been difficult and often revised [5, 8, 19, 20]. Recently small subunit ribosomal RNA (rRNA) phylogenies have been found to group the *Chlorarachnion* host with a diverse group of protists that includes sarcomonads and filose amoebae [4, 6]. Here we have examined genes encoding the cytoskeletal proteins, alpha- and beta-tubulin, from the host genome of *Chlorarachnion,* and alpha-tubulin from *Cercomonas,* a sarcomonad flagellate. These are the first protein-coding genes reported from either cercomonads or chlorarachniophyte-hosts. Phylogenetic trees of alpha-tubulin confirmed the proposed relationship between these organisms based on rRNA. In addition, the phylogeny of both alpha- and beta-tubulins suggests a relationship between the *Chlorarachnion* host, Heterolobosea, Euglenozoa, chlorophytes, heterokonts, and alveolates. We have also found that *Chlorarachnion* contains two highly dissimilar classes of both alpha- and beta-tubulin paralogues, one of which is evolving very rapidly.

MATERIALS AND METHODS

Strains, culture conditions, and DNA extraction. *Chlorarachnion* sp. strain CCMP 621 was cultured and DNA prepared as described [29]. T. Cavalier-Smith and E. E. Chao kindly provided DNA from *Cercomonas* sp. strain RS/23 (ATCC 50319).

Amplification conditions and molecular techniques. Tubulins were amplified either as a single product, or in two overlapping pieces where a single product could not be amplified. In cases where a gene was amplified in two fragments, the 3′ end was amplified first using degenerate primers, and the 5′ end was then amplified using a degenerate 5′ primer and a specific primer based on a region of the 3′ end that was found to be unique to that gene. In all cases the overlap between these fragments was several hundred nucleotides, and throughout this overlap they were found to be identical. In each case multiple products corresponding to each gene were sequenced.

Chlorarachnion alpha-tubulins 1 and 2 were amplified using primers TCCGAATTCARGTNGGNAAYGCNGGYTGGGA and CGCGCCATNCCYTCNCCNACRTACCA. Alpha-tubulin 3 was amplified using primers GATACCGTNGTNGARCCNTAYAA and CGCGCCATNCCYTCNCCNACRTACCA for the 3′ end, and TCCGAATTCARGTNGGNAAYGCNGGYTGGGA and TGCTGTCAGTGATGACAC for the 5′ end. *Chlorarachnion* beta-tubulins 1, 2, 3, and 4 were amplified using primers GCCTGCAGGNCARTGYGGNAAYCA and TCCTCGAGTRAAYTCCATYTCRTCCAT. Beta-tubulin 4 proved to be a truncated product of mispriming approximately 83 codons into the expected product. This missing sequence was amplified using primers GCCTGCAGGNCARTGYGGNAAYCA and AGGTAGCGGCCATTACGCG. *Chlorarachnion* beta-tubulin 5 was am-

[1] To whom correspondence should be addressed. Telephone: 812-855-2549; Fax: 812-855-6705; Email: "pkeeling@sunflower.bio.indiana.edu"

[2] Current Address: Biology Department, Indiana University, Bloomington, IN 47405-6801, USA.

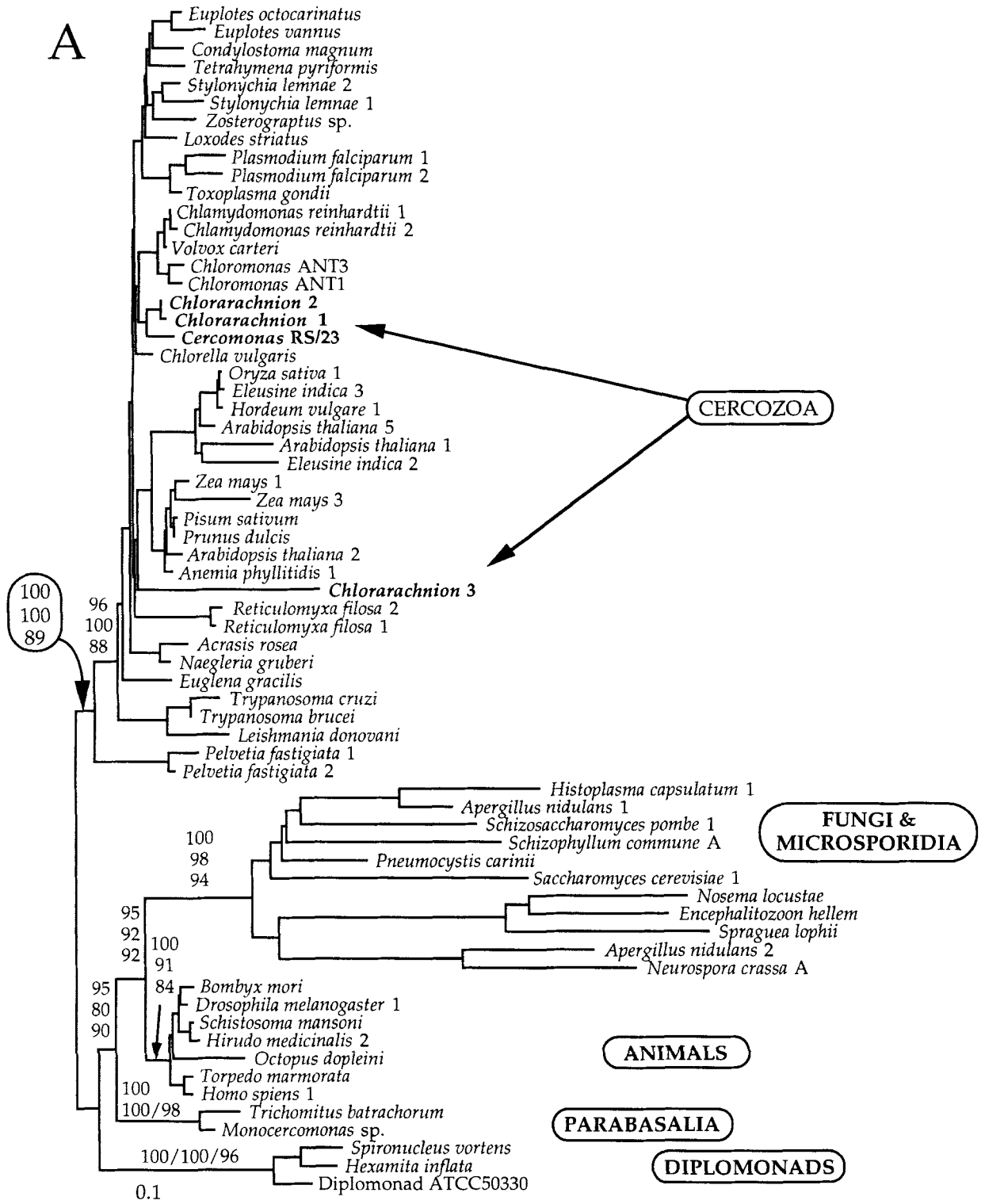

Fig. 1. Fitch-Margoliash trees of (**A**) alpha-tubulin and (**B**) beta-tubulin showing the overall topology of tubulins from diverse eukaryotes, and showing the presence of a strongly supported clade (supporting numbers circled in bold) consisting of Cercozoa, Heterolobosea, Euglenozoa, chlorophytes, heterokonts, alveolates, rhodophytes (sampled only for beta-tubulin), and *Reticulomyxa* (shown only in alpha-tubulin). The branching order within this clade is shown in Fig. 2. Numbers at nodes correspond from top to bottom or left to right to the bootstrap support from Fitch-Margoliash, neighbor-joining, and the percent occurrence in 1,000 puzzling steps. A dash indicates support less than 50%, and supporting numbers have been shown only for the major groups; support for nodes within these groups has been excluded for clarity.

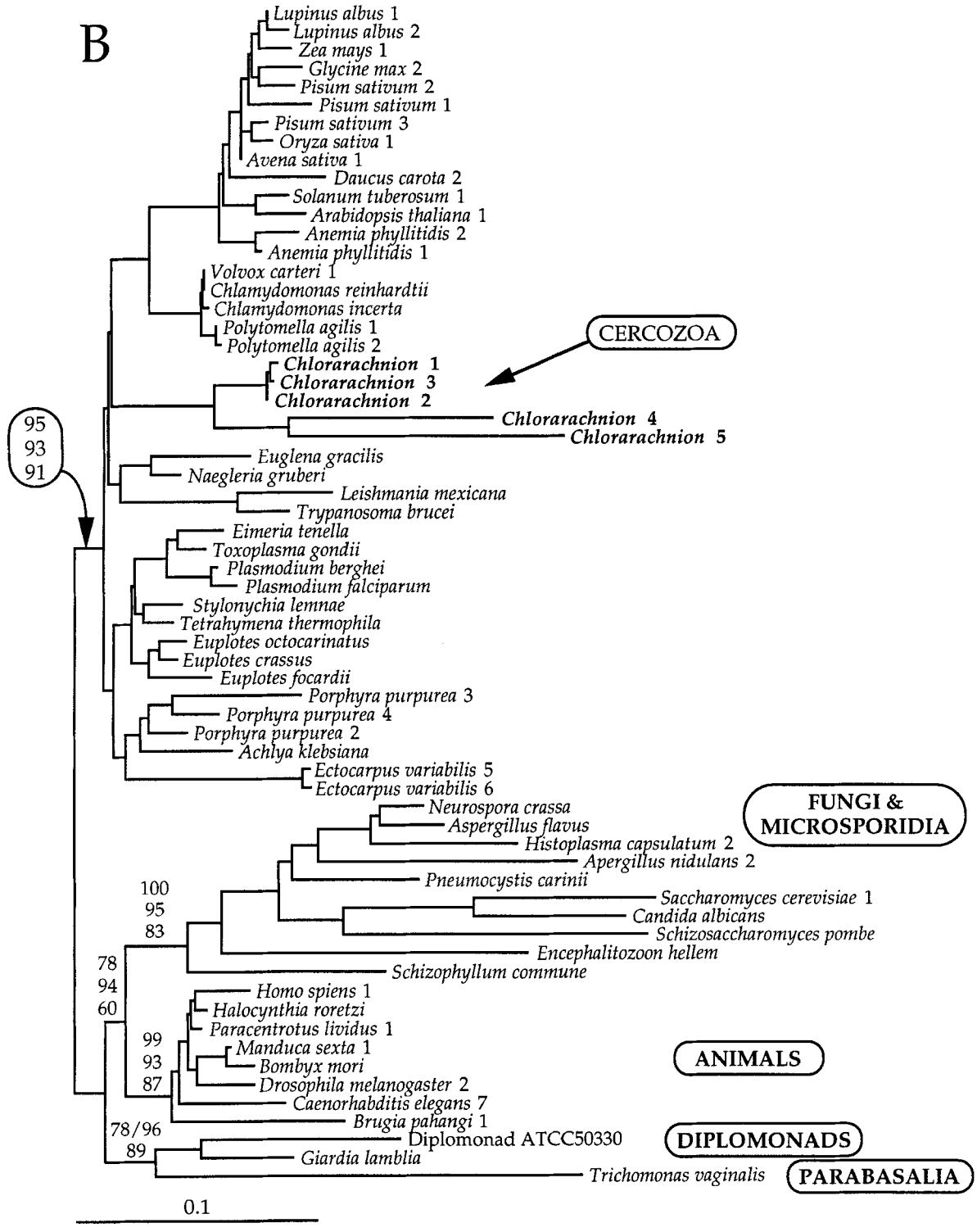

Fig. 1. Continued.

plified with primers GATACCGTNGTNGARCCNTAYAA and TCCTCGAGTRAAYTCCATYTCRTCCAT for the 3′ end, and GCCTGCAGGNCARTGYGGNAAYCA and GTTCAGATCGCCGTAGGTAGG for the 5′ end.

Cercomonas alpha-tubulin was amplified using primers TCCGAATTCARGTNGGNAAYGCNGGYTGGGA and CGCGCCATNCCYTCNCCNACRTACCA. Beta-tubulin amplification reactions were also carried out with *Cercomonas*. Several products were cloned and sequenced, however, none proved to resemble tubulin.

All amplification reactions consisted of 35 cycles with an annealing temperature of 50° C and an extension time of 1.5 or 2 min. Polymerase chain reaction (PCR) products were purified from agarose gels as described [25], and cloned into pGEM-T

vector (Promega, Madison, WI). Sequencing was carried out on an ABI 373 with dye-terminator chemistry. New sequences have been deposited in GenBank as accession numbers AF043927-AF043936 and AF046802-AF046805.

Pulsed field electrophoresis and southern blotting. *Chlorarachnion* cultures were harvested and resuspended in plug buffer (10 mM Tris-HCl, pH 7.5, 100 mM Na_2EDTA, and 100 mM NaCl). Resuspended cells were mixed with an equal volume of 1.2% low-gelling-temperature agarose in plug buffer at 37° C and pipetted into chilled plug molds. Solid plugs were digested for 48 h at 50° C in two changes of 10 mM Tris-HCl, pH 7.5, 400 mM Na_2EDTA, and 10 mg/ml Pronase E (Sigma P-5147), and stored in 10 mM Tris-HCl, pH 7.5, 10 mM Tris at 4° C. Plugs were electrophoresed in 1% agarose, 0.5× TBE (Tris Boric Acid EDTA) buffer using a BioRad CHEF DRIII apparatus. Electrophoresis conditions were: 14° C, 175V, 16 h with a pulse change interval of 20 s and 16 h at a pulse change interval of 10 s.

DNA from CHEF gels was transferred to Zeta-Probe membrane (BioRad, Richmond, CA) using alkaline transfer according to the manufacturer's instructions. Membranes were cut into strips corresponding to lanes on the gel and probed with fragments of alpha-tubulin 3 and beta-tubulin 4 to determine where these divergent genes were encoded. Probes were amplified from clones using the primers used to originally amplify the fragments from genomic DNA. Probes were gel-isolated and labeled by random priming (Promega).

cDNA library screening and cDNA amplification. *Chlorarachnion* beta-tubulin 4 was used as a probe to screen a *Chlorarachnion* Lambda ZapII cDNA library (Stratagene). Phage infections and screening were carried out according to the manufacturer's protocols (Stratagene). Probes were prepared as described above.

The 5′ end of *Chlorarachnion* alpha-tubulin 3 mRNA was amplified from the cDNA library using PCR. The M13 reverse primer was used with the specific primer TGCTGTCAGTGATGACAC and approximately 100 ng of DNA prepared from the cDNA library by chloroform extraction and ethanol precipitation. A single product was amplified, which was cloned and sequenced as described above. Four clones were sequenced, and all were found to be unique as they differed slightly in the length of the 5′ untranslated region.

Phylogenetic analysis. New alpha- and beta-tubulins were added to an alignment containing a large representation of tubulins from diverse organisms. Alpha- and beta-tubulins were analysed separately in all cases since rooting one subfamily with the other has previously been shown to spuriously root the two halves of the tree [25]. For all analyses 436 characters were used for alpha-tubulin and 429 characters for beta tubulin. Alignments are available from the corresponding author upon request.

Corrected distance measurements were calculated by PROTDIST using the Dayhoff PAM substitution matrix. Trees based on these distances were constructed by neighbor-joining with random addition using the NEIGHBOR program, and using Fitch-Margoliash with 10 random additions and global rearrangements using the FITCH program [11]. One hundred bootstrap replicates were performed on all datasets using SEQBOOT [11], and bootstrap trees constructed as above, with the exception that only a single random addition was used for Fitch-Margoliash trees.

Maximum likelihood quartet puzzling was carried out with 1,000 puzzling steps using PUZZLE 3.1 [39]. Quartets were calculated with rate heterogeneity modeled on a gamma distribution with four rate categories and the rate heterogeneity parameter estimated from the data. Substitution frequency corrections were based on the JTT substitution matrix with the overall frequency of occurrence estimated from the data. Protein maximum likelihood trees were also inferred using constrained trees and the exhaustive search option of ProtML [1] with correction by the JTT substitution probability matrix adjusted for the frequency of amino acid occurrence in the data. Bootstrap probabilities were calculated by the resampling of estimated log-likelihood (RELL) method as implemented by ProtML [1], and the bootstraps collated using mol2con, a perl script by A. Stoltzfus (pers. commun.). The relative likelihood support for each node tested with ProtML was also calculated using TreeCons [23] using a class V weighting scheme, and alpha values of 0.01, 0.05, 0.1 and 0.5. The alpha value was not found to have a great effect on any of the data analysed here, at most 1–2%, and we have therefore reported only the relative likelihood support obtained with a cutoff of 0.1. Constraints were based on the distance and quartet puzzling analyses and were as follows: in both sets, land plants, green algae, Euglenozoa, and alveolates were constrained; in alpha-tubulin the two *Reticulomyxa* genes, the two Heterolobosea, and two *Chlorarachnion* genes were also constrained; and in beta-tubulin trees, the five *Chlorarachnion* sequences were constrained as were the three *Porphyra* sequences. Alpha-tubulins were analysed with *Chlorella* alternatively constrained with green-algae or left free, and in the RELL bootstrap consensus tree of the latter it branched with other green algae. Similarly, the *Chlorarachnion* alpha-tubulin 3 was either constrained with other cercozoan genes, left free, or excluded altogether.

RESULTS AND DISCUSSION

Tubulin genes from *Chlorarachnion* and *Cercomonas*. Amplification reactions for alpha-tubulin from both *Chlorarachnion* and *Cercomonas* yielded single products corresponding to the size expected of a gene that contained no introns. Only a single gene was found to be represented from *Cercomonas*, but three distinct genes were found in *Chlorarachnion*, two very similar (95% identical at the nucleotide level) and a third that differed considerably (alpha-1 and alpha-3, only 74% identical at the nucleotide level). None contained introns.

Beta-tubulin amplification reactions from *Cercomonas* and *Chlorarachnion* both yielded numerous products. Several of these were sequenced from *Cercomonas*, but unfortunately none was found to encode tubulin. Conversely, from *Chlorarachnion* a total of five distinct genes, beta-1, 2, 3, 4, and 5, were sequenced. Three of these, beta-1, 2, and 3, were similar to one another (99 to 93% identical at the nucleotide level). The remaining two, beta-4 and 5, were highly divergent for tubulins: beta-4 and beta-5 were respectively 73% and 76% identical to beta-1, and 74% identical to one another at the nucleotide level. Beta-4 and beta-5 also share a number of otherwise unique substitutions.

Chlorarachnion beta-tubulins 1–3 did not contain introns, but beta-4 contained an extraordinary 10 introns and beta-5 shared 8 of these (it lacked introns 8 and 9). These introns are the first to be characterised from the nucleus of *Chlorarachnion*. They are all comparatively rich in AT base-pairs and range between 57 and 112 nucleotides in length. The introns in beta-5 showed an obvious tendency to be longer than the corresponding introns of beta-4: the average length in beta-4 was 61 nucleotides while that of beta-5 was 91; and 6 of the 8 shared introns were longer in beta-5.

These 10 intron positions were compared to all other known tubulin introns to see if any positions were shared with other taxa. Five of the intron positions were unique to these genes, and a number was also found in a single isolated instance in a distantly related gene, suggesting that the introns were inserted

twice independently. Only two introns were found to be shared broadly within another lineage, and they were intron 5, which is shared with a vast diversity of green algae and land plants, and intron 3, which is shared with diverse animals.

Phylogenetic position of *Chlorarachnion* and *Cercomonas* tubulins. Alpha- and beta-tubulin phylogenetic trees are both characterised by a major divergence splitting plants, green and red algae, heterokonts, Euglenozoa, Heterolobosea, and alveolates from a well-supported grouping of animals and the fungi-microsporidia clade. A few other taxa, notably diplomonads and trichomonads, do not definitively fall with either of these groups. Still other beta-tubulins from taxa, such as slime moulds, *Entamoeba,* and *Reticulomyxa* have very divergent sequences that cannot be reliably placed in the tree [25]. To determine to which major group the *Cercomonas* and *Chlorarachnion* tubulins belong, trees including a comprehensive representation of tubulins (but excluding the highly divergent types) were inferred. Alpha-tubulin trees comprising 66 sequences were inferred using a variety of methods, and in all cases the *Cercomonas* and *Chlorarachnion* genes branched within the group containing plants, green algae, heterokonts, Euglenozoa, Heterolobosea, and alveolates (Fig. 1A). Similarly, in all trees of 64 beta-tubulins, those of *Chlorarachnion* also branched within the clade comprising plants, green and red algae, heterokonts, Euglenozoa, Heterolobosea, and alveolates (Fig. 1B). The rate heterogeneity of *Chlorarachnion* tubulins can be seen from these trees: alpha-tubulin 3 is extremely divergent, as are beta-tubulins 4 and 5.

Having identified their general position in the tree, the relationships between *Cercomonas, Chlorarachnion,* and other members of this group were examined more closely. The branch separating the two main divisions in tubulin trees was relatively long and unbroken. To avoid possible artifacts arising from this long branch, trees were inferred from only the sequences belonging to the large and heterogeneous clade including *Chlorarachnion* and *Cercomonas.* For both genes, trees were inferred including and excluding some of the relatively divergent sequences, most importantly, the divergent *Chlorarachnion* beta-4 and -5 sequences, and also the highly divergent *Chlorarachnion* alpha-3.

An alpha-tubulin tree (Fig. 2A) does not include alpha-tubulin 3, which was found to branch very erratically in different analyses. In distance trees, alpha-3 sometimes branched with *Cercomonas* and *Chlorarachnion* but more often did not. In quartet puzzling trees, the *Cercomonas* and three *Chlorarachnion* genes formed a weakly supported clade (found in 60% of the puzzling steps), and in protein maximum likelihood the position of *Chlorarachnion* alpha-3 was not resolved. It appears that *Chlorarachnion* alpha-3 is related to other *Chlorarachnion* alpha-tubulins. However, its relationship is very tenuous, and by branching erratically between individual bootstrap trees it affected the overall support for the tree. The sequence was accordingly removed to determine the branching position of the remaining *Chlorarachnion* and *Cercomonas* sequences.

The first noteworthy relationship is between *Chlorarachnion* alpha-1 and alpha-2 and *Cercomonas* alpha-tubulin (Fig. 2A), which found moderate but consistent support in all analyses. This relationship was also tested in partially constrained protein maximum likelihood trees, which supported this relationship with RELL bootstrap value of 76% and a relative likelihood support of 82%. These are the first protein-coding genes to be sequenced from the *Chlorarachnion* host and a cercomonad, so this is not only an important confirmation of the relationship originally proposed on the basis on small subunit rRNA, but also provides support for the taxonomic validity of the Phylum Cercozoa [7].

The cercozoan sequences in turn branched with an interesting group consisting of the Heterolobosea and Euglenozoa. Heterolobosea and Euglenozoa do not generally branch together in trees based on small subunit rRNA [37], but such a relationship is nevertheless favoured by both alpha- and beta-tubulin phylogeny (Fig. 2A, B and [2, 25]), and also by the presence of discoid mitochondrial cristae in both Heterolobosea and Euglenozoa [7, 36]. The bootstrap support for this cluster in alpha-tubulin was modest at best, and support from quartet puzzling only slightly better. However, in partly constrained protein maximum likelihood trees, the union of Heterolobosea and Euglenozoa was highly favoured, having both RELL bootstrap support and relative likelihood support of 100%. The clustering of the Cercozoa with this Euglenozoa-Heterolobosea group was not well supported by alpha-tubulin, as the Cercozoa were also found to branch with green algae in some distance analyses. Moreover, in quartet puzzling and protein maximum likelihood analyses, the branching order of the Heterolobosea-Euglenozoa clade, the Cercozoa (*Chlorarachnion* and *Cercomonas*), and the green algae was not resolvable with any confidence.

The beta-tubulin tree is more or less congruent with the well-supported features of the alpha- tubulin tree (cf. Fig. 2A, B). Once again, there is strong support from beta-tubulin for the clustering of Heterolobosea and Euglenozoa despite the weak support from distance bootstrap analyses. In protein maximum likelihood the RELL bootstrap support for a Heterolobosea-Euglenozoa clade was 96%, and the relative likelihood support was also 96%. The weak relationship between *Chlorarachnion* and chlorophyte beta-tubulins only appeared in analyses including *Chlorarachnion* beta-4 and beta-5; when these were excluded the remaining *Chlorarachnion* cluster tended to branch with the Heterolobosea-Euglenozoa clade. Furthermore, quartet puzzling weakly supported the relationship between chlorophytes, *Chlorarachnion,* and the Euglenozoa-Heterolobosea group, but failed to resolve the three.

Altogether, with both alpha- and beta-tubulin there is a consistent, and strong support for a large and heterogeneous clade consisting of Cercozoa, Heterolobosea, Euglenozoa, chlorophytes, alveolates, and heterokonts (and rhodopytes in beta-tubulin). There is also strong support for specific relationships between *Chlorarachnion* and *Cercomonas,* and between Heterolobosea and Euglenozoa. There may also be a specific relationship between Cercozoa, Heterolobosea and Euglenozoa, and chlorophytes, but this is much less clear.

Evidence that divergent *Chlorarachnion* tubulins are not targeted to the endosymbiont. Since *Chlorarachnion* contains two eukaryotic genomes, that of the amoeboflagellate host and that of the green algal symbiont, the provenance of the *Chlorarachnion* tubulins is open to question. Nucleomorph genes tend to be highly divergent [16], so we have considered the possibility that the highly divergent alpha-3 and beta-4 and 5 tubulins may be derived from the nucleomorph genome. The alpha-tubulins are particularly suspicious since alpha-1 and alpha-2 are specifically related to *Cercomonas,* but the highly divergent alpha-3 is only tenuously related to any of these. There is not such a hint of polyphyly in *Chlorarachnion* beta-tubulins. However, the highly divergent genes do share two intron positions and a generally high intron density with green algae: intron 1 is shared with *Polytomella* and intron 5 is shared with diverse green algae and land plants. These characteristics provide sufficient reason to examine the possibility that *Chlorarachnion* tubulins may be derived from the green algal endosymbiont.

In addition to being highly divergent, nucleomorph genes have a relatively high density of 18–20 nucleotide introns and tend to be greater than 70% AT [16]. No nuclear protein-coding

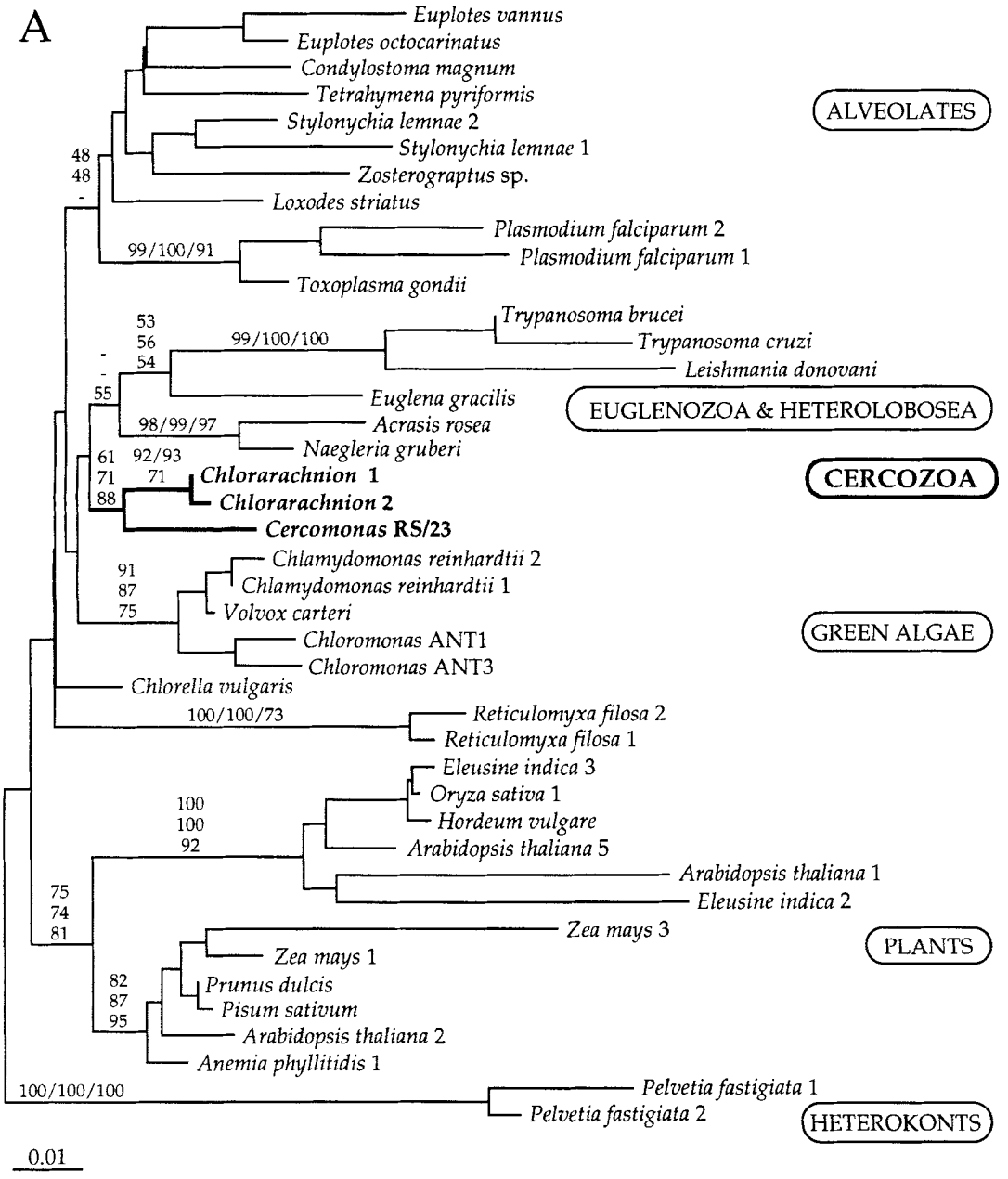

Fig. 2. Fitch-Margoliash trees showing detailed relationships within the clade incuding *Chlorarachnion* and *Cercomonas* for (**A**) alpha-tubulin and (**B**) *Chlorarachnion* beta-tubulin. Numbers at nodes correspond from top to bottom or left to right to the bootstrap support from Fitch-Margoliash, neighbor-joining, and the percent occurrence in 1,000 puzzling steps. A dash indicates support less than 50%, and supporting numbers have been shown only for the major divisions; support for nodes within these groups has been excluded for clarity. The union of *Cercomonas* and *Chlorarachnion*, and Euglenozoa and Heterolobosea were also strongly supported by protein maximum likelihood, as described in text.

genes have previously been characterised from *Chlorarachnion*, but the nuclear (host) DNA has a buoyant density consistent with approximately 50% AT content (G. I. M., unpubl. data). All the tubulins described here were close to the 50% AT expected for the host genome: the most AT-rich was alpha-tubulin 3, which was 55% AT. Moreover, the divergent beta-tubulins 4 and 5 also contained large spliceosomal introns, which are completely uncharacteristic of nucleomorph introns. Furthermore, when hybridised to Southern blots of CHEF gels, alpha-tubulin 3 and beta-tubulin 4 hybridised specifically with host chromosomes and did not hybridise to the three nucleomorph chromosomes (Fig. 3).

Thus, all the tubulin genes described here are apparently encoded by the host genome. However, the host genome must contain a large number of genes whose products are targeted to the plastid [17]. Examples of nuclear-encoded proteins that are targeted to the plastid have been found to encode a signal peptide at their amino-terminus, suggesting that targeting to the symbiont from the cytoplasm is initially via the endomembrane system (J. A. D. & G. I. M., unpubl. data). Any nuclear-encoded proteins that function in the endosymbiont cytoplasm surrounding the plastid (i.e. the periplastid space) would therefore be expected to include a signal peptide as well. To address this possibility, the 5′ ends of alpha-tubulin 3 and beta-tubulin 4

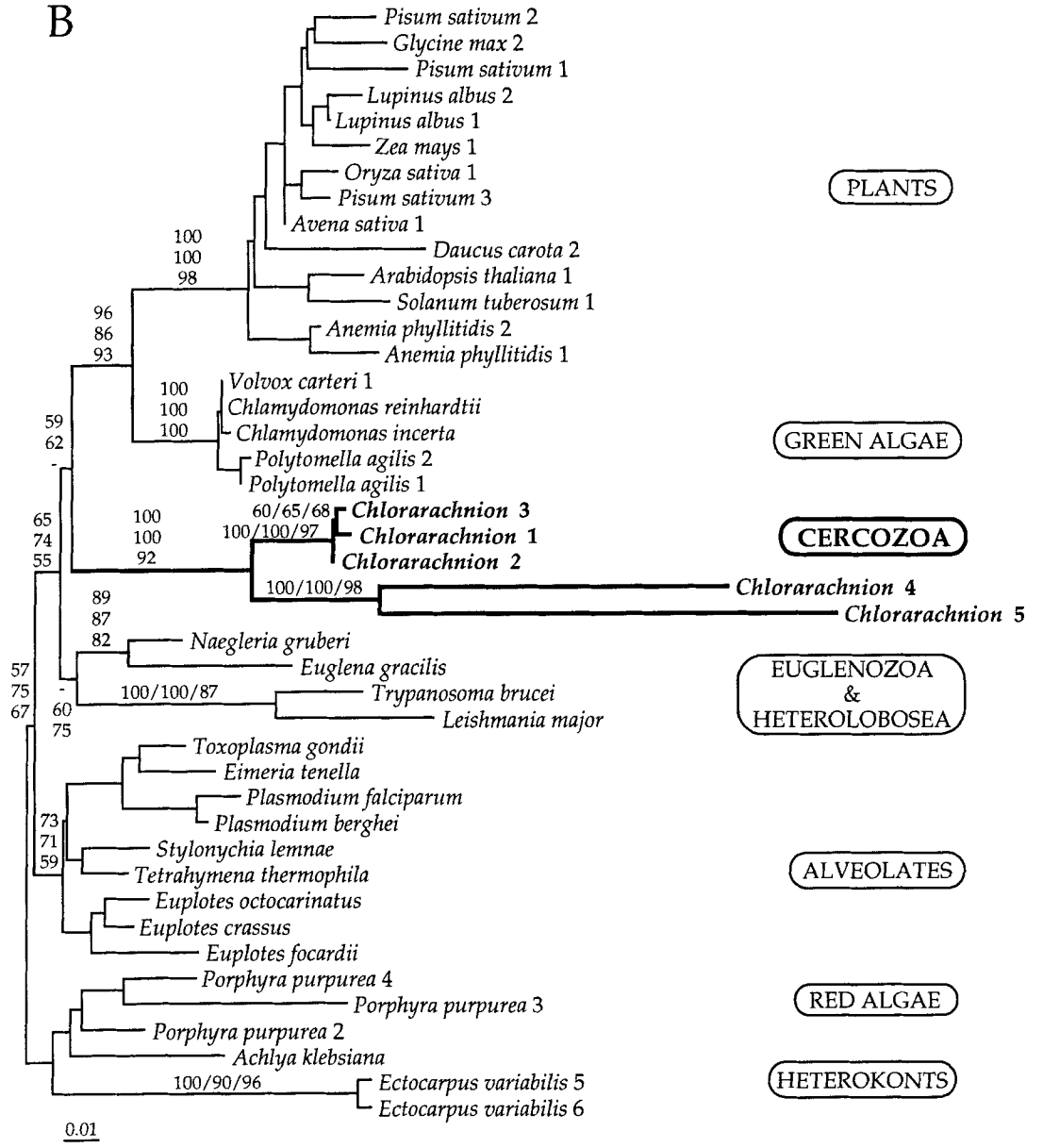

Fig. 2. Continued.

were sequenced from cDNAs to determine if they included any identifiable targeting information that might direct them to the symbiont. In both cases there is clear evidence from the 5′ untranslated region that these genes do not contain any targeting information (Fig. 4). The absence of targeting information, together with the tendency for the long- and short-branch *Chlorarachnion* tubulins to cluster together (a strong tendency for beta-tubulin and a weak tendency for alpha-tubulin), makes it most likely that all of these genes encode cytosolic proteins, and that all were derived from the host lineage.

It is interesting then, that of the eight tubulin genes found in *Chlorarachnion* so far, none are from the nucleomorph. Microtubules have never been observed in the symbionts of either chlorarachniophytes or cryptomonads, and it has been speculated that neither of their nucleomorphs has mitotic division [28, 31]. Nevertheless, tubulins are involved in a host of cellular activities, and the total loss of microtubules from all aspects of the symbiont and from all stages of its life cycle is perhaps unlikely. Ultimately the completed sequence of the nucleomorph genome will address this question [30].

***Chlorarachnion* beta-tubulin 4 is a recently evolved pseudogene.** The coding region of the *Chlorarachnion* beta-tubulin 4 gene and cDNAs were found to contain an in-frame TAG stop codon at a position otherwise highly conserved for tyrosine. This stop is found over 30 codons from the end of even the shortest beta-tubulin, and the sequence it precedes is every bit as conserved as the upstream sequence, arguing that it was not simply a truncated gene. Moreover, the position where termination would be expected to occur has another TAG codon, indicating that TAG likely does not encode tyrosine in *Chlorarachnion.* The most obvious explanation is that *Chlorarachnion* beta-tubulin 4 is a pseudogene. But if this is the case, then *Chlorarachnion* beta-tubulin 4 is unusual since it contains no other obvious pseudogene characteristics such as frame-shifts, a skewed AT content, or insertions and deletions, and it is abundantly represented in the cDNA library. In addition, *Chlorar-*

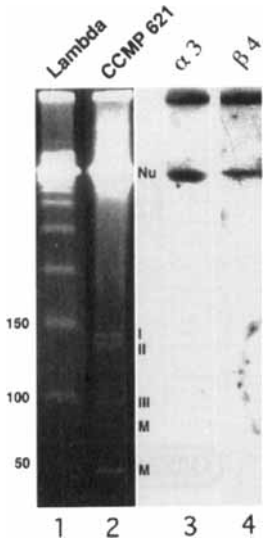

Fig. 3. Localisation of *Chlorarachnion* alpha-tubulin 3 and beta-tubulin 4 genes to the host genome. Lane 1: lambda size markers; lane 2: ethidium stained *Chlorarachnion* sp. CCMP 621 DNA showing the unresolved nuclear chromosomes (Nu), the three nucleomorph chromosomes (I, II, and III), and monomeric and dimeric mitochondrial genomes (M); lane 3: *Chlorarachnion* DNA probed with alpha-tubulin 3 probe; lane 4: *Chlorarachnion* DNA probed with beta-tubulin 4 probe.

achnion beta-tubulin 4 mRNAs are processed, although the processing itself was quite unusual. Of the 4 cDNAs completely sequenced, only one was totally free of introns, and one intron was also mis-spliced resulting in a deletion (Fig. 5).

Although there is presently nothing known about mRNA splicing in the cytosol of *Chlorarachnion*, the processing of these messages does seem unusual. Nevertheless, given that the beta-4 protein is truncated and likely non-functional, it is interesting that beta-4 is expressed and processed at all. It would appear that this gene has only recently lost its function, and is gradually losing efficiency and accuracy of processing due to a lack of selection.

Implications for the evolution of *Chlorarachnion* and *Cercomonas*, and for tubulin phylogeny. Prior to the application of molecular phylogeny there was no clear idea to which evolutionary lineage the chlorarachniophyte host might belong [12, 19, 20, 34]. Even then, original studies using the small subunit rRNA from *Chlorarachnion* revealed little, since no sequences were known from any closely related taxon [5, 8]. Subsequently, a wider sampling of rRNA sequences suggested a relationship between *Chlorarachnion* and *Cercomonas* and a very diverse group of flagellates, amoebae, and amoeboflagellates [4, 6] recently named Cercozoa [7]. Alpha-tubulin trees including genes from *Chlorarachnion* and *Cercomonas* now confirm this relationship, and support the validity of the Cercozoa. Both alpha- and beta-tubulin trees also give a very broad suggestion of the evolutionary position of the Cercozoa, as they strongly and consistently fall within a cluster comprising chlorophytes, rhodophytes, alveolates, heterokonts, and the clade consisting of Heterolobosea and Euglenozoa.

This grouping of Euglenozoa and Heterolobosea is of particular interest here, and perhaps a key to the understanding of tubulin phylogeny. The overall structures of tubulin and rRNA phylogenies differ in a fundamental way. Small subunit rRNA trees typically form a series of consecutive, deep, independent branches, including the separate heterolobosean and euglenozoan branches, all "crowned" by an unresolved explosion of diversity that includes *Chlorarachnion* [38, 40]. Tubulin trees, on the other hand, characteristically divide most eukaryotes into one of two large, deeply diverging clades separated by the independent diplomonad and parabasalian lineages [25]. If we ignore for a moment the deep branches of rRNA trees, the small subunit rRNA topology and the tubulin topologies are not so very different: in rRNA trees based on a substitution rate calibration, the eukaryotic "crown." topology is much like the tubulin topologies, with animals and fungi separated from a poorly resolved clade of chlorophytes, heterokonts, alveolates, *Chlorarachnion*, and so forth [40]. The incongruence between rRNA and tubulin phylogenies only becomes evident when one considers certain taxa that branch deeply in small subunit rRNA, such as slime moulds, microsporidia, Heterolobosea, and Euglenozoa. Historically, these deep-branching groups have been widely accepted as representing truly ancient lineages, and the line between them and the "crown" a major demarcation in eukaryotic evolution. Quite significantly, however, there is now growing acceptance that for two of these groups this is not the case: microsporidia appear to be related to fungi (for reviews see [26, 32]), and slime moulds appear to be related to animals and fungi [3, 27]. Tubulin phylogeny was one of the earliest indications that microsporidia and fungi were related [10, 25], and remains one of the strongest arguments for this relationship (Fig. 1). The "crown" position of slime moulds is also in agreement with some tubulin analyses [2, 25]. In light of this, the present assertion that tubulin phylogeny may be correctly representing the evolutionary history of Euglenozoa and Heterolobosea should not be considered overly far fetched. Indeed, a number of other protein phylogenies, including chaperonin 60, V-ATPase, and unrooted trees of actin also support this assertion by suggesting a relationship between either Euglenozoa or Heterolobosea with plants or alveolates [18, 21, P.

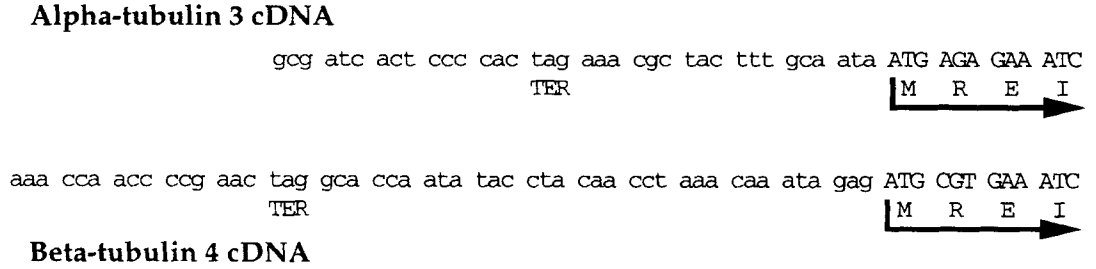

Fig. 4. Untranslated leaders and 5′ ends of *Chlorarachnion* alpha-tubulin 3 and beta-tubulin 4 cDNAs. The amino termini of alpha- and beta-tubulins are both very highly conserved. This position is indicated by the arrow and inferred translation. Upstream of this position there are in-frame termination codons in both genes, indicating that this is almost certainly the start site of translation, and that there are not amino terminal leaders on either protein.

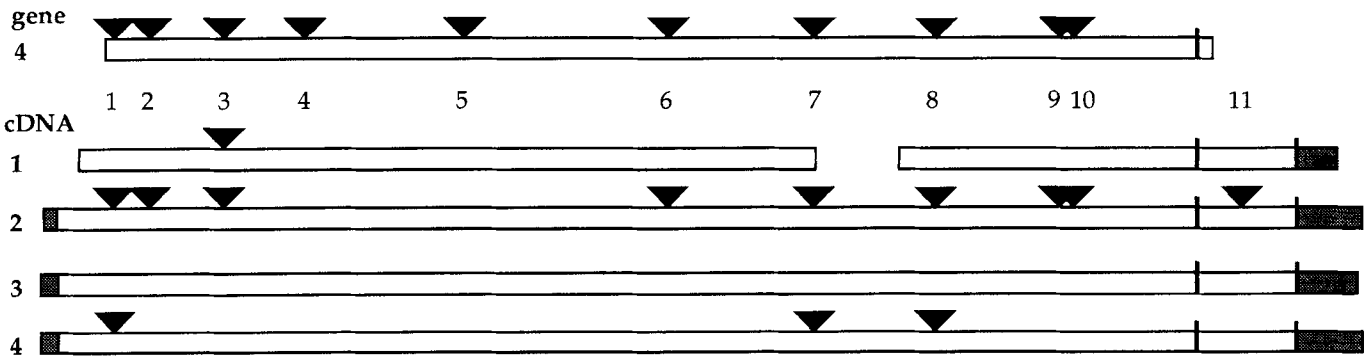

Fig. 5. Structure of the *Chlorarachnion* beta-tubulin 4 gene and processing of its mRNAs. The gene-fragment is shown above (gene 4) with the ten introns marked by triangles and numbered below. The four beta-4 cDNAs are shown below (cDNAs 1–4) with unspliced introns indicated above each cDNA. The 5′ and 3′ untranslated regions are shown as shaded boxes to the left and right of the coding region which is shown as an open box. Also shown are the mis-splicing of intron 7 in cDNA 1 (marked as a gap in the coding region), and the presence of the eleventh intron in cDNA 2. Hatch marks represent TAG codons. The first is located within the coding region (shown in the genomic clone and all four cDNAs), and the second is located at the expected position of termination (downstream of the known genomic sequence, but found in all four cDNAs).

J. K., unpubl. data]. Other genes, such as elongation factor-1 alpha, and elongation factor-2, place the Euglenozoa and Heterolobosea "high" in the tree without much resolution [3, 33], and still others, such as phosphoglucose isomerase and enolase, support the relationship between alveolates and chlorophytes [24, 37, P. J. K., unpubl. data]. These phylogenies are perhaps each hinting that the large, diverse, and well-supported clade of tubulin genes comprising chlorophytes, Heterolobosea, Euglenozoa, heterokonts, alveolates, and Cercozoa may really represent eukaryotic evolutionary history. This interpretation of eukaryotic phylogeny would have interesting implications for our definition of "crown-eukaryotes", or even question whether the concept of a "crown" has any meaning at all.

ACKNOWLEDGMENTS

We would like to thank E. E. Chao and T. Cavalier-Smith for providing *Cercomonas* DNA, J. M. Logsdon Jr. for providing his database of intron positions, and A. Stoltzfus for the use of mol2con. We are also indebted to V. Su and K. Adams for their assistance with sequencing, and to P. R. Gilson for his *Chlorarachnion* DNA preparation and continuing interest. This work was supported by an Australian Research Council grant to G. I. M. P. J. K. was supported by a fellowship from the Medical Research Council of Canada. J. A. D. was supported by an Australian Postgraduate Research Award.

LITERATURE CITED

1. Adachi, J. & Hasegawa, M. 1992. Computer science monographs, No. 27. MOLPHY: programs for molecular phylogenetics, I.-PROTML: maximum likelihood inference of protein phylogeny. Institute of Statistical Mathematics, Tokyo.

2. Baldauf, S. L. & Palmer, J. D. 1993. Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. *Proc. Natl. Acad. Sci. USA*, **90**:11558–11562.

3. Baldauf, S. L. & Doolittle, W. F. 1997. Origin and evolution of the slime molds (Mycetozoa). *Proc. Natl. Acad. Sci. USA*, **94**:12007–12012.

4. Bhattacharya, D., Helmchen, T. & Melkonian, M. 1995. Molecular evolutionary analyses of nuclear-encoded small subunit ribosomal RNA identify an independent rhizopod lineage containing the Euglyphina and the Chlorarachniophyta. *J. Euk. Microbiol.*, **42**:65–69.

5. Cavalier-Smith, T. 1993. Kingdom Protozoa and its 18 phyla. *Microbiol. Rev.*, **57**:953–994.

6. Cavalier-Smith, T. & Chao, E. E. 1997. Sarcomonad ribosomal RNA sequences, rhizopod phylogeny, and the origin of euglyphid amoebae. *Arch. Protistenk.*, **147**:227–236.

7. Cavalier-Smith, T. 1998. A revised six kingdom system of life. *Biol Rev.*, in press.

8. Cavalier-Smith, T., Allsopp, M. & Chao, E. 1994. Chimeric conundra: are nucleomorphs and chromists monophyletic or polyphyletic? *Proc. Natl. Acad. Sci. USA*, **91**:11368– 11372.

9. Cavalier-Smith, T., Couch, J. A., Thorstenin, K. E., Gilson, P. R., Deane, J. A., Hill, D. R. A. & McFadden, G. I. 1996. Cryptomonad nuclear and nucleomorph 18S rRNA phylogeny. *Eur. J. Phycol.*, **31**: 315–328.

10. Edlind, T. D., Li, J., Visvesvara, G. S., Vodkin, M. H., McLaughlin, G. L. & Katiyar, S. K. 1996. Phylogenetic analysis of beta-tubulin sequences from amitochondrial protozoa. *Mol. Phylogenet. Evol.*, **5**:359–367.

11. Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package). J. Felsenstein, University of Washington, Seattle.

12. Geitler, L. 1930. Ein grünes Filarplasmodium und andere neue Protisten. *Arch. Protistenk.*, **69**:615–637.

13. Gibbs, S. P. 1978. The chloroplasts of *Euglena* may have evolved from symbiotic green algae. *Can. J. Bot.*, **56**:2882–2889.

14. Gibbs, S. P. 1981. The route of entry of cytoplasmically synthesized proteins into chloroplasts of algae possessing chloroplast ER. *J. Cell Sci.*, **35**:253–266.

15. Gilson, P. & McFadden, G. I. 1995. The chlorarachniophyte: a cell with two different nuclei and two different telomeres. *Chromosoma*, **103**:635–641.

16. Gilson, P. R. & McFadden, G. I. 1996. The miniaturized nuclear genome of a eukaryotic endosymbiont contains genes that overlap, genes that are cotranscribed, and the smallest known spliceosomal introns. *Proc. Natl. Acad. Sci. USA*, **93**:7737–7742.

17. Gilson, P. R. & McFadden, G. I. 1997. Good things in small packages: the tiny genomes of chlorarachniophyte endosymbionts. *Bioessays*, **19**:167–173.

18. Gogarten, J. P., Hilario, E. & Olendzenski, L. 1996. Gene duplications and horizontal gene transfer during early evolution. *Symp. Soc. Gen. Microbiol.*, **54**:267–292.

19. Hibberd, D. J. 1990. Phylum Chlorarachnida *In*: Margulis, L., Corliss, J. O., Melkonian, M. & Chapman, D. J. (ed.), Handbook of Protoctista, Jones & Bartlett Publishers, Boston. Pp. 288–292.

20. Hibberd, D. J. & Norris, R. E. 1984. Cytology and ultrastructure of *Chlorarachnion reptans* (Chlorarachniophyta divisio nova, Chlorarachniophyceae classis nova). *J. Phycol.*, **20**:310–330.

21. Horner, D. S., Hirt, R. P., Kilvington, S., Lloyd, D. & Embley, T. M. 1996. Molecular data suggest an early acquisition of the mitochondrion endosymbiont. *Proc. R. Soc. Lond. B. Biol. Sci.*, **1263**:1053–1059.

22. Ishida, K., Cao, Y., Hasegawa, M., Okada, N. & Hara, Y. 1997. The origin of chlorarachniophyte plastids, as inferred from phylogenetic comparisons of amino acid sequences of EF-Tu. *J. Mol. Evol.*, **45**:682–687.

23. Jermiin, L. S., Olsen, G. J., Mengersen, K. L. & Easteal, S. 1997. Majority-rule consensus of phylogenetic trees obtained by maximum-likelihood analysis. *Mol. Biol. Evol.*, **14**:1296–1302.

24. Katz, L. A. 1996. Transkingdom transfer of the phosphoglucose isomerase gene. *J. Mol. Evol.*, **43**:453–459.

25. Keeling, P. J. & Doolittle, W. F. 1996. Alpha-tubulin from early-diverging eukaryotic lineages and the evolution of the tubulin family. *Mol. Biol. Evol.*, **13**:318–326.

26. Keeling, P. J. & McFadden, G. I. 1998. Origins of microsporidia. *Trends Microbiol.*, **6**:19-23.

27. Kuma, K., Nikoh, N., Iwabe, N. & Miyata, T. 1995. Phylogenetic position of *Dictyostelium* inferred from multiple protein data sets. *J. Mol. Evol.*, 41:238–246.

28. Ludwig, M. & Gibbs, S. 1989. Evidence that nucleomorphs of *Chlorarachnion reptans* (Chlorarachniophyceae) are vestigial nuclei: morphology, division and DNA-DAPI fluorescence. *J. Phycol.*, **25**:385–394.

29. McFadden, G. I., Gilson, P. R., Hofmann, C. J., Adcock, G. J. & Maier, U. G. 1994. Evidence that an amoeba acquired a chloroplast by retaining part of an engulfed eukaryotic alga. *Proc. Natl. Acad. Sci. USA*, **91**:3690–3694.

30. McFadden, G. I., Gilson, P. R., Douglas, S. E., Cavalier-Smith, T., Hofmann, C. J. & Maier, U. G. 1997. Bonsai genomics: sequencing the smallest eukaryotic genomes. *Trends Genet.*, **13**:46–49.

31. Morrall, S. & Greenwood, A. D. 1982. Ultastructure of nucleomorph division in species of Cryptophyceae and its evolutionary implications. *J. Cell Sci.*, **54**:311–328.

32. Müller, M. 1997. What are the microsporidia? *Parasitol. Today*, **13**:455–456.

33. Nakamura, Y., Hashimoto, T., Kamaishi, T., Adachi, J., Nakamura, F., Okamoto, K. & Hasegawa, M. 1996. Phylogenetic position of kinetoplastid protozoa inferred from the protein phylogenies of elongation factors 1a and 2. *J. Biochem.*, **119**:70–79.

34. Norris, R. E. 1967. Micro-algae in enrichment cultures from Puerto Penasco, Sonora, Mexico. *Bull. South. Cal. Acad. Sci.*, **66**:233–250.

35. O'Kelly, C. J. 1993. The jakobid flagellates: structural features of *Jakoba, Reclinomonas* and *Histiona* and implications for the early diversification of eukaryotes. *J. Eukaryot. Microbiol.*, **40**:627–636.

36. Patterson, D. J. 1994. Protozoa: evolution and systematics. *In*: Hausmann, K. & Hülsmann (ed.), Progress in Protozoology, Proceedings of the IX International Congress of Protozoology, Berlin 1993. Gustav Fisher, Stuttgart. Pp. 1–14.

37. Read, M., Hicks, K. E., Sims, P. F. & Hyde, J. E. 1994. Molecular characterisation of the enolase gene from the human malaria parasite *Plasmodium falciparum*. Evidence for ancestry within a photosynthetic lineage. *Eur. J. Biochem.*, **220**:513–520.

38. Sogin, M. L. 1997. History assignment: when was the mitochondrion founded? *Curr. Opin. Genet. Dev.*, **7**:792–799.

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

40. Van de Peer, Y., Rensing, S. A., Maier, U. G. & De Wachter, R. 1996. Substitution rate calibration of small subunit ribosomal RNA identifies chlorarachniophyte endosymbionts as remnants of green algae. *Proc. Natl. Acad. Sci. USA*, **93**:7732–7736.

41. Whatley, J. M., John, P. & Whatley, F. R. 1979. From extracellular to intracellular: the establishment of mitochondria and chloroplasts. *Proc. R. Soc. Lond.* B., **204**:165–187.

Received 3-2-98, 4-8-98; accepted 5-12-98