Materials and Methods

Collection of sequences and preparation of alignments

The set of 103 dinoflagellate minicircle sequences with confirmed circularity was collected from GenBank (http://www.ncbi.nlm.nih.gov) (see Table S1). Amino acid sequences of Rpl28, Rpl33, Ycf16, and Ycf24 were obtained via BLAST searches across GenBank (non-redundant protein and EST databases), TbestDB (http://tbestdb.bcm.umontreal.ca/searches/login.php), and *Cyanophora* Genome Project Dragonblast (http://cyanophora.biology.uiowa.edu/dragon).

Amino acid alignments were obtained in MAFFT 6.238b (Katoh et al. 2005) and sites suitable for further phylogenetic analyses were extracted from the alignments with Gblocks 0.91b assuming less stringent criteria (Talavera and Castresana 2007). The alignments also were edited manually in GeneDoc 2.6.002 (Nicholas et al. 2007). We analyzed a total of five sets of aligned sequences including concatenated alignments of Ycf16 and Ycf24 sequences (Table S2).

Phylogenetic analyses

Phylogenetic trees were inferred via the Bayesian approach in PhyloBayes (Lartillot and Philippe 2004) as well as maximum-likelihood in both PhyMl 3.0 (Guindon and Gascuel 2003) and TreeFinder (Jobb et al. 2004).

Amino acid substitution models selected for particular alignments and approaches are shown in Table S2. The models applied in PhyMl were selected in ProtTest 2.2 considering five criteria (-lnL, AIC, AICc, BIC, HQ) and assuming optimization of model, branches, and topology of the tree (Abascal et al. 2005). The models used in TreeFinder were chosen according to the Propose model module in this program considering all criteria (-lnL, AIC, AICc, BIC, HQ) and assuming optimized frequencies of amino acids. In all analyses we applied five discrete categories for gamma distributed rates.

To find a tree close to optimal, and to avoid the trap of local optima in global tree searches, maximum likelihood (ML) trees were carried out in PhyMl and TreeFinder in several stages. At first, an initial ML tree was constructed based on the default start tree determined by the given program. Then, a set of 500 starting tree topologies was generated assuming the initial tree as a center tree. We generated 100 trees for each of five topological distances: 7, 10, 15, 20, and 25 NNI steps. We imposed topological constraints on the generated trees fixing phylogenetic relationships that were supported by bootstrap values

equal to or higher than 75% in a bootstrap tree. The bootstrap tree for this approach was the consensus of 1000 ML trees calculated in PhyMl or TreeFinder. The 500 generated trees and the initial tree were used as starting trees for global tree searches in these programs (now with no constraints). The final tree was selected according to the best maximum-likelihood value.

We assumed search depth = 2 in TreeFinder and the best heuristic search algorithms, i.e. NNI and SPR, in PhyMl. Edge support was assessed by the bootstrap analysis with 1000 replicates in each of these two programs. Additionally, we applied the LR-EWL (Local Rearrangements-Expected Likelihood Weights) method in TreeFinder and the approximate likelihood ratio test (aLRT) based on χ^2 and Shimodaira-Hasegawa-like procedure in PhyMl. The minimum of these two aLRT support values was shown at selected nodes in the trees presented.

We performed two types of analysis in PhyloBayes assuming models: LG+Γ and CAT+Γ with the number of components, weights, and profiles inferred from the data (Table S2). Two independent Markov chains were run for 200,000 and 1,000,000 cycles for the first and the second approach, respectively. After obtaining convergence, the last 100,000 to 10,000 and 500,000 to 250,000 trees from each respective chain were collected to compute a posterior consensus.

The approximately unbiased test (AU) was performed to compare trees presented in figures with alternative topologies that assume different positions of dinoflagellate *Ceratium horridum* and *Pyrocystis lunula* sequences (Table S3). This test was carried out in TreeFinder (Jobb et al. 2004) assuming 10^5 replicates and in Consel v0.1k assuming 10^7 replicates (Shimodaira and Hasegawa 2001). In the latter case site-wise log-likelihoods were calculated in PhyMl 3.0 for the trees analyzed (Guindon and Gascuel 2003). The substitution models applied in these analyses were the same as in the inferences of corresponding trees (Table S2).

Protein domain and secondary structure predictions

Domain analysis of the potential fragment of FtsY protein encoded on the *Pyrocystis lunula* AF490367 minicircle was carried out with Pfam database searches (http://pfam.sanger.ac.uk/) and secondary structures were predicted with eight algorithms: (1) hhpred (http://toolkit.tuebingen.mpg.de/hhpred#), (2) jufo (http://www.meilerlab.org/web/), (3) phyre (http://www.sbg.bio.ic.ac.uk/phyre/), (4) poter (http://distill.ucd.ie/porter/), (5) pred (http://www.predictprotein.org/), (6) prof (http://www.aber.ac.uk/~phiwww/prof/), (7) psipred (http://bioinf.cs.ucl.ac.uk/psipred/), and (8) sable (http://sable.cchmc.org/).

Discussion

Evidence against bacterial sequence contamination of *Ceratium* and *Pyrocystis* minicircles

The data presented in the main text of our article are most consistent with the transfer of five genes from the Bacteroidetes (*Algoriphagus*-like species) to two dinoflagellate minichromosomes: the *Ceratium* AF490364 minicircle (*ycf16* and *ycf24* genes) and the *Pyrocystis* AF490367 minicircle (*rpl28*, *rpl32*, and *ftsY* genes). However, it also is important to consider potential contamination of dinoflagellate DNA extractions by exogenous bacterial DNA. In the case of contamination, we should expect the following: (i) different types of contaminating sequences, (ii) a relatively high level of similarity between the analyzed *Ceratium* and *Pyrocystis* minicircle genes and Bacteroidetes homologs found in databases, and (iii) different bacterial sources that do not correlate with relationships between targeted dinoflagellate taxa. In contrast to these predictions, all the genes studied clearly are related to plastid functions and the *Ceratium* AF490364 and *Pyrocystis* AF490367 minicircles were isolated along with other molecules encoding typical minicircle- or plastid-specific genes (Laatsch et al. 2004). Moreover, the highest identity at the nucleotide level observed between minicircle genes and Bacteroides homologs did not exceed 78%. Also, the genes from two minichromosomes analyzed belong to related dinoflagellate genera and show relationships to the same or closely related bacterial species; this indicates a shared evolutionary history. Finally, these two *Ceratium* and *Pyrocystis* minicircle sequences are similar in length to typical dinoflagellate minicircles.

Are *Ceratium* minicircles recovered in higher-molecular weight fractions of plastid DNA?

Laatsch et al. (2004) detected DNA in purified *Ceratium* plastids, but in a higher-molecular weight fraction than where minicircles were found; however, this does not preclude the presence of minicircles in both DNA fractions for two reasons. First, recovery of minicircles in the higher-molecular-weight DNA fraction during gel electrophoresis could result from retardation of their movement as open or relaxed circular DNA forms, or replication intermediates (Leung and Wong 2009). Second, the size of minicircles can increase through dimerization (Nelson and Green 2005), and it is possible that they form networks of concatenated circular DNA molecules in the native plastid, similar to structures found in kinetoplastid mitochondria (Lukes et al. 2002). In agreement with this idea, analyses

of the 50-150 kb genome from the peridinin plastid of *Gonyaulax polyedra* (now *Lingulodinium polyedrum*) suggest that it is composed of several distinct molecules (Wang and Morse 2006). The plastid DNA identified by Laatsch et al. (2004) in *C. horridum* could be similar to that of *L. polyedrum*, because these genomes have similar sizes and large proportions of non-coding DNA. The hypothesized *Lingulodinium*-like plastid genome organization in *Ceratium* would explain the *psbB* signal in higher-molecular-weight plastid DNA in the Southern blot analysis of Laatsch et al. (2004).

References

Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21:2104-2105.

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

Jobb G, von Haeseler A, Strimmer K. 2004. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol.* 4:18.

Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform.* 9:286-298.

Leung SK, Wong JT. 2009. The replication of plastid minicircles involves rolling circle intermediates. *Nucleic Acids Res.* 37:1991-2002.

Lukes J, Guilbride DL, Votýpka J, Zíková A, Benne R, Englund PT. 2002. Kinetoplast DNA network: evolution of an improbable structure. *Eukaryot Cell.* 1:495-502.

Lartillot N, Philippe H. 2004. A Bayesian mixture model for across site heterogeneities in the amino-acid replacement process. *Mol Biol Evol.* 21:1095-1109.

Nicholas KB, Nicholas HBJr, Deerfield DWII. 1997. GeneDoc: Analysis and visualization of genetic variation. *EMBNEW News* 4:14.

Shimodaira H, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17:1246-1247.

Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol.* 56:564-577.

Table S1. List of dinoflagellate minicircle sequences used in this study.

Accession number	Species	Encoded genes (start-end)
AY612430	*Adenoides eludens*	*psbA* (3456-4493)
AY612431		*psbA* (3388-4425)
AY612433		*psbA* (1-435), *psbD* (4090-4599)
AY612435		*psbA* (1-435), *psbD* (3913-4422)
AY612432		*psbA* (1-435), *psbD* (4536-5045)
AY612434		*psbD* (3247-4314)
AY612436		*psbD* (3492-4559)
AY612437		*psbD* (3559-4626)
AY612438		*psbA* (3838-4875)
AF206672	*Amphidinium carterae*	*psbA* (185-1207)
AJ307009		
AJ307010		
AJ307011		
AJ307012		
AJ307013		
AJ307014		
AJ307015		
AJ307016		
AJ311628		*psbD* (386-1453), *psbE* (1574-1807)
AJ311629		*psbB* (286-2220)
AJ311630		*petb* (300-959), *atpA* (1032-2426)
AJ311631		*psbA* (320-2356)
AJ311632		*psbA* (672-1694)
AJ318067		*trnM* (174-237)
AY004258		*23S rRNA* (419-2554)
AY004259		*psbA* (781-1803)
DQ507216		
DQ507217		*petD* (906-1379)
DQ507218		*atpB* (764-2410)
DQ507219		*psbC* (864-2243)
DQ507220		*psbB* (369-1889)
AJ582639	*Amphidinium operculatum*	*psaB* (243-2207)
AJ582640		*23S rRNA* (200-2400)
AF401627		
AF401628		
AF401629		
AF401630		*trnM* (167-230)
AF426172		*psbC* (13-1392)
AJ250262		*psbA* (1-1023)
AJ250263		*psbB* (1-1521)
AJ250264		*psaA* (1-2016)
AJ250265		*petD* (1-474)
AJ250266		*atpB* (1-1647)
AJ582641		
AJ582642		
AJ582643		
AJ582644		
AJ620761		*psbD* (381-1448), *psbE* (1567-1800), *psbI* (2059-2166)
AY048664		*petB* (587-1246), *atpA* (1337-2713)
AF490356	*Ceratium horridum*	*psaA* (1-339)
AF490357		*psaB* (1-638)
AF490358		*psbB* (4397-5883)

AF490359		*psbC* (1-155)
AF490360		*psbD* (735-1062)
AF490361		*psbD* (1-724)
AF490363		*petB* (423-624)
AF490364		*ycf24* (1-1116), *ycf16* (1146-1937)
AF490362		*psbE* (211-439)
AF490368		
DQ318023		*psaB* (4063-6266)
DQ318024		*psbC* (4499-5917)
DQ318025		*psbD* (5650-6696)
DQ318029		
DQ318030		
DQ318031		
DQ318032		
DQ318033		
AY004265	*Heterocapsa niei*	*psbA* (1260-2303)
AY004266		*23S rRNA* (1012-3255)
AY004260	*Heterocapsa pygmaea*	*23S rRNA* (423-2689)
AY004261		*psbA* (826-1872), *trnW* (1937-2008), *trnP* (2048-2120)
AY033400		*psbA* (870-1916), *trnW* (2084-2155), *trnP* (2195-2267)
AY004262	*Heterocapsa rotundata*	*psbA* (744-1790)
AY004263		*23S rRNA* (828-3082)
AF130031	*Heterocapsa triquetra*	*psaA* (737-2935)
AF130032		*psaB* (762-3092)
AF130033		*psbA* (932-1978)
AF130034		*psbB* (716-2233)
AF130035		*psbC* (851-2233)
AF130036		*atpA* (797-2155)
AF130037		*petB* (730-1389)
AF130038		*16S rRNA* (670-2214)
AF130039		*23S rRNA* (682-2958)
AY004267		
AY004268		*trnP* (1980-2050)
AY004269		*trnW* (1840-1909), *trnP* (1955-2025)
AY004270		*trnP* (1753-1823)
AY004271		*trnW* (1624-1685), *trnP* (1731-1801)
DQ168850		
DQ168851		*psbD* (1561-2628)
DQ168852		*trnW* (11-82), *Pro* (137-207), *psbE* (1963-2196)
DQ168853		*trnM* (255-327), *Pro* (467-535), *petD* (1701-2177)
AY004264	*Protoceratium reticulatum*	*23S rRNA* (1475-3772)
AF490367	*Pyrocystis lunula*	*rpl28* (1-246), *rpl33* (274-462)
AF490366		*psbC* (2245-3642)
AJ884897	*Symbiodinium sp.*	*psbA* (1-1029)
AJ884898		*psbA* (1-1029)
AJ884899		*psbA* (1-1029)
AY160085		*psbA* (1-17), *psbA* (749-896)
AY160086		*psbA* (1-17), *psbA* (802-949)
AY160087		*psbA* (1-17), *psbA* (986-1133)
AY160088		*psbA* (1-17), *psbA* (917-1064)
AY160089		*psbA* (1-17), *psbA* (666-813)

Table S2. Characteristics of analyzed alignments and applied substitution models.

Alignment set	Number of sequences	Amino acid substitution models used in:			
		PhyMl	TreeFinder	PhyloBayes	
Ycf16	71	LG+I+Γ	LG+Γ+F	LG+Γ	CAT+Γ
Ycf24	71	LG+Γ	LG+Γ+F	LG+Γ	CAT+Γ
Ycf16+Ycf24	62	LG+I+Γ	LG+I+Γ+F/LG+Γ+F	LG+Γ	CAT+Γ
Rpl28	80	LG+I+Γ+F	LG+I+Γ+F	LG+Γ	CAT+Γ
Rpl33	91	LG+I+Γ	MIX+I+Γ+F	LG+Γ	CAT+Γ

In all models we assumed five discrete categories for gamma distributed rates.

Table S3. Result of AU tests obtained by two methods, considering trees presented in figures and alternative topologies with different positions of dinoflagellate *Ceratium horridum* and *Pyrocystis lunula* sequences. Significant p-values less than 0.05 are in **bold**.

Protein	Tree topology with tested position:	TreeFinder		PhyMl + Consel	
		lnL	p-value	lnL	p-value
Ycf16+Ycf24	Tree presented in Fig. 1	-44742	0.234	-44857	1.000
	Ceratium + Apicomplexa	-45048	**0**	-45166	7e-38
Ycf16	Tree presented in Fig. S2	-15188	0.234	-15254	1.000
	Ceratium + Apicomplexa	-15303	**0**	-15371	**2e-14**
	Ceratium + *Amphidinium*	-15312	**0**	-15378	**1e-58**
Ycf24	Tree presented in Fig. S3	-33715	0.234	-33754	1.000
	Ceratium + Haptophyta	-33897	**0**	-33936	**2e-06**
	Ceratium + Apicomplexa	-33906	**0**	-33945	**8e-109**
Rpl33	Tree presented in Fig. S4	-3074	0.980	-3163	0.987
	Pyrocystis + Apicomplexa	-3095	**0.029**	-3186	**0.024**
	Pyrocystis + *Amphidinium*	-3096	**0.017**	-3187	**0.009**
	Pyrocystis + (Haptophyta, *Karlodinium*)	-3117	**0**	-3209	**5e-06**
Rpl28	Tree presented in Fig. S5	-6407	0.976	-6470	0.970
	Pyrocystis + (*Bacteroides*, *Parabacteroides*)	-6414	0.054	-6476	0.058
	Pyrocystis + (*Flavobacterium*, *Kordia*, *Gramella*, *Algoriphagus*)	-6418	0.084	-6482	0.090
	Pyrocystis + *Algoriphagus*	-6419	0.064	-6482	0.052
	Pyrocystis + Haptophyta	-6469	**0**	-6532	**5e-11**
	Pyrocystis + Apicomplexa	-6472	**0**	-6536	**1e-85**