

Widespread occurrence and genetic diversity of marine parasitoids belonging to *Syndiniales* (*Alveolata*)

L. Guillou,^{1*} M. Viprey,¹ A. Chambouvet,¹
R. M. Welsh,² A. R. Kirkham,³ R. Massana,⁴
D. J. Scanlan³ and A. Z. Worden²

¹Station Biologique de Roscoff, UMR7144, CNRS et Université Pierre et Marie Curie, BP74, 29682 Roscoff Cedex, France.

²Monterey Bay Aquarium Research Institute, 7700 Sandholdt Road, Moss Landing, CA 95039, USA.

³Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK.

⁴Institut de Ciències del Mar, CMIMA, Passeig Marítim de la Barceloneta 37-49, 08003 Barcelona, Catalonia, Spain.

Summary

***Syndiniales* are a parasitic order within the eukaryotic lineage *Dinophyceae* (*Alveolata*). Here, we analysed the taxonomy of this group using 43655 18S rRNA gene sequences obtained either from environmental data sets or cultures, including 6874 environmental sequences from this study derived from Atlantic and Mediterranean waters. A total of 5571 out of the 43655 sequences analysed fell within the *Dinophyceae*. Both bayesian and maximum likelihood phylogenies placed *Syndiniales* in five main groups (I–V), as a monophyletic lineage at the base of ‘core’ dinoflagellates (all *Dinophyceae* except *Syndiniales*), although the latter placement was not bootstrap supported. Thus, the two uncultured novel marine alveolate groups I and II, which have been highlighted previously, are confirmed to belong to the *Syndiniales*. These groups were the most diverse and highly represented in environmental studies. Within each, 8 and 44 clades were identified respectively. Co-evolutionary trends between parasitic *Syndiniales* and their putative hosts were not clear, suggesting they may be relatively ‘general’ parasitoids. Based on the overall distribution patterns of the *Syndiniales*-affiliated sequences, we propose that *Syndiniales* are exclusively marine. Interestingly, sequences belonging to groups II, III and V were largely retrieved from**

the photic zone, while Group I dominated samples from anoxic and suboxic ecosystems. Nevertheless, both groups I and II contained specific clades preferentially, or exclusively, retrieved from these latter ecosystems. Given the broad distribution of *Syndiniales*, our work indicates that parasitism may be a major force in ocean food webs, a force that is neglected in current conceptualizations of the marine carbon cycle.

Introduction

The *Alveolata*, one of the major eukaryotic lineages, is composed of four protist classes: the *Ciliophora*, the *Apicomplexa*, the *Perkinsea* and the *Dinoflagellata*. *Alveolata* have adopted a large range of trophic modes and habitats. They can be important marine primary producers. For instance, about half of dinoflagellates are photosynthetic (Lessard and Swift, 1986), with some being responsible for toxic algal blooms. Ciliates and dinoflagellates can also be active predators; others are coral symbionts (e.g. the dinoflagellate *Symbiodinium* spp.) and some ciliates even reside in mammalian guts (Williams and Coleman, 1992). Finally, a large number of species are parasites, broadly distributed throughout these *Alveolata* classes. For instance, *Apicomplexa* is composed solely of obligate parasites. In addition to phylogenetic relatedness based on gene sequence comparisons, *Alveolata* are unified by specific morphological characters. These include the presence of membrane-bound flattened vesicles, termed alveoli (Cavalier-Smith, 1993; Patterson, 1999), distinct pores called micropores that pierce the outer membrane and the presence of a more or less developed apical complex apparatus used by alveolate parasites to enter their host (reviewed by Leander and Keeling, 2003).

The discovery of novel marine alveolate (MALV) lineages in marine planktonic communities by culture-independent techniques, specifically MALV groups I and II (Díez *et al.*, 2001a; López-García *et al.*, 2001; Moon-van der Staay *et al.*, 2001), has raised questions regarding functional roles of these diverse populations. Phylogenetic analyses showed MALV Group II belongs to the *Syndiniales*, a dinoflagellate order exclusively composed of marine parasites. This assignment was based on the close phylogenetic relatedness of the environmental 18S

Received 17 April, 2008; accepted 29 June, 2008. *For correspondence. E-mail lguillou@sb-roscoff.fr; Tel. (+33) 2 98 29 23 79; Fax (+33) 2 98 29 23 24.

rRNA gene sequences (Moon-van der Staay *et al.*, 2001; Skovgaard *et al.*, 2005) with three previously described genera: *Amoebophrya* spp., *Hematodinium* spp. and *Syndinium* spp. More recently, 18S rRNA gene sequences derived from *Ichthyodinium chaberladi* (Gestal *et al.*, 2006) and *Duboscquella* sp. (Harada *et al.*, 2007), two parasitic *Syndiniales* genera, showed an affiliation with MALV Group I.

MALV groups I and II have been retrieved from various marine habitats, mainly from the picoplankton fraction (< 2 or < 3 µm size fractionated samples). These groups frequently form the majority of sequences within marine environmental clone libraries (see López-García *et al.*, 2001; Moon-van der Staay *et al.*, 2001; Massana *et al.*, 2004; Romari and Vaulot, 2004; Not *et al.*, 2007). This has led to a large increase in the number of MALV sequences deposited in GenBank over the last few years. In a previous study, Groisillier and colleagues (2006) detected five distinct clades within MALV Group I, and 16 clades within MALV Group II. However, in that study, many sequences could not be clearly assigned to a specific clade, suggesting further discrete lineages might exist. Interestingly, some clades comprised sequences retrieved only from very specific habitats, such as anoxic environments or deep sea hydrothermal vents, while other clades contained sequences obtained from widely varying habitats.

In the present work our aims were (i) to broaden representation of the different environments in which members of the *Alveolata* are potentially encountered by constructing and sequencing 18S rRNA gene libraries; (ii) to clarify the phylogeny of the *Alveolata* using recently published and the new 18S rRNA gene sequences for this group; (iii) to undertake a rigorous review of clade organization within the above mentioned MALV groups I and II; (iv) to compare the genetic diversity of environmental sequences derived from culture-independent PCR surveys (targeting the 18S rRNA gene) to our present knowledge of the taxonomy of *Syndiniales*; and (v) to extract general information on the preferred habitats and distribution of specific members of the *Syndiniales*.

Results

Sample sites and environmental conditions

Most of the novel environmental sequences obtained in this study were retrieved from coastal and oceanic waters in the Atlantic Ocean and Mediterranean Sea (Table 1). The collection sites varied in terms of season, depth and level of oligotrophy. For example, the Bermuda Atlantic Time-series Station (BATS) is relatively oligotrophic, and at the time of sampling was already strongly stratified. In contrast, the northern Sargasso Sea Station, although relatively close to BATS, was likely still influenced by deep

winter mixing that occurs in this area, bringing nutrients to surface waters (see Cuvelier *et al.*, 2008 for further discussion). The Florida Straits sites represent three distinct water types. Station 1, on the western side of the Florida Straits, was relatively coastal. Station 4 represents the core of the Gulf Stream Current-forming waters, which are highly oligotrophic, while Station 14 is also oligotrophic but a shallow setting on the eastern side of the Florida Straits (see Cuvelier *et al.*, 2008). Other environmental sequences from the Atlantic Ocean were obtained during the Atlantic Meridional Transect (AMT 15, see also Zwirgmaier *et al.*, 2007), which extended from 48°N [south-west (SW) of the UK] to 40°S (SW of Cape Town, South Africa). Environmental sequences from the Mediterranean Sea were collected in late summer along a Mediterranean transect sampled during the PROSOPE cruise in 1999 (Garczarek *et al.*, 2007), from high nutrient levels in the Morocco upwelling to strong phosphorus limitation in the eastern most basin. Although half of the genetic libraries were built using a PCR approach biased towards green algae, dinoflagellate sequences were retrieved in both data sets (the majority of them were retrieved using general eukaryotic primers). We also used a PCR approach biased towards MALV Group II (using the specific primer ALV01) that we compared with the use of general eukaryotic primers from a coastal site (English Channel, France). Most of genetic libraries were built on very small size fractions (less than 2–3 µm), although some were processed on larger size fractions and even after incubations (see genetic libraries from coastal sites from the Mediterranean Sea). This heterogeneous data set offered us a very large range of environmental sequence origins. In total, 6874 new environmental sequences were generated and screened for dinoflagellate sequences.

Alveolate 18S rRNA gene sequence data set

The completed data set comprised 43655 eukaryotic 18S rRNA gene sequences obtained either from GenBank or from the environmental clone libraries described above. From marine environments, 351 environmental clone libraries were analysed with major contribution from sites (Table 2) in the Atlantic (Díez *et al.*, 2001a; Countway *et al.*, 2007; Not *et al.*, 2007), Indian (Not *et al.*, 2008), Arctic (Lovejoy *et al.*, 2006; Stoeck *et al.*, 2007) and Pacific (Moon-van der Staay *et al.*, 2001) Oceans, Mediterranean Sea (Viprey *et al.*, 2008), Antarctic (López-García *et al.*, 2001), as well as several coastal sites (Massana *et al.*, 2004; Romari and Vaulot, 2004; Medlin *et al.*, 2006; Worden, 2006). These encompassed a range of marine habitats including the photic zone, sediments, hydrothermal vents and anoxic ecosystems (Table 2). Terrestrial and continental ecosystems were also included

Table 1. Description of the sampling sites and characteristics of clone libraries constructed for this study.

| Environment and cruise | Station and Site | Coordinates | Date | Code library | Depth (m) | Primer | Sample details (μm) | Total number of clones with insert screened |
|---|---|-------------|--------------|--------------|-----------------|-----------------|----------------------------------|---|
| Mediterranean Sea PROSOPE Oceanic euphotic zone Accession numbers from EU792921 to EU793941 + EU848494 to EU848495 See also Viprey <i>et al.</i> (2008) | St 1 Strait of Gibraltar | 38°08'N | 14-Sep-99 | C1-30 | 30 | Euk328f-CHLO02 | < 3 | 246 |
| | | 05°18'W | | E1-30 | 30 | Euk328f-Euk329r | < 3 | 135 |
| | St 3 Algerian Basin | 37°08'N | 16-Sep-99 | E1-80 | 80 | Euk328f-CHLO02 | < 3 | 195 |
| | | | | E1-80 | 80 | Euk328f-Euk329r | < 3 | 69 |
| | | 03°03'E | | C3-5 | 5 | Euk328f-CHLO02 | < 3 | 165 |
| | | | | E3-5 | 5 | Euk328f-Euk329r | < 3 | 46 |
| | St 5 Strait of Sicily | 36°47'N | 18-Sep-99 | C3-25 | 25 | Euk328f-CHLO02 | < 3 | 23 |
| | | | | E3-25 | 25 | Euk328f-Euk329r | < 3 | 56 |
| | | 13°32'E | | C3-95 | 95 | Euk328f-CHLO02 | < 3 | 173 |
| | | | | E3-95 | 95 | Euk328f-Euk329r | < 3 | 27 |
| 33°08'N | | 20-Sep-99 | C5-25 | 25 | Euk328f-CHLO02 | < 3 | 151 | |
| | | | E5-25 | 25 | Euk328f-Euk329r | < 3 | 168 | |
| 22°02'E | | | C5-55 | 55 | Euk328f-CHLO02 | < 3 | 185 | |
| | | | E5-55 | 55 | Euk328f-Euk329r | < 3 | 131 | |
| Ionian Basin | | 33°08'N | 22°02'E | CM-5 | 5 | Euk328f-CHLO02 | < 3 | 173 |
| | | | | EM-5 | 5 | Euk328f-Euk329r | < 3 | 62 |
| St 9 Tyrrhenian Basin | 41°08'N | 28-Sep-99 | CM-50 | 50 | Euk328f-CHLO02 | < 3 | 171 | |
| | | | EM-50 | 50 | Euk328f-Euk329r | < 3 | 102 | |
| | 10°43'E | | CM-110 | 110 | Euk328f-CHLO02 | < 3 | 134 | |
| | | | EM-110 | 110 | Euk328f-Euk329r | < 3 | 115 | |
| | English Channel, France | 41°08'N | 28-Sep-99 | C9-5 | 5 | Euk328f-CHLO02 | < 3 | 101 |
| | | | | E9-5 | 5 | Euk328f-Euk329r | < 3 | 144 |
| | Coastal euphotic zone Accession numbers from EU785054 to EU785255 | 10°43'E | | C9-65 | 65 | Euk328f-CHLO02 | < 3 | 223 |
| | | | | E9-65 | 65 | Euk328f-Euk329r | < 3 | 119 |
| | Catalan, Mediterranean Sea, Spain | 43°38'N | 30-Sep-99 | CD-15 | 15 | Euk328f-CHLO02 | < 3 | 172 |
| | | | | ED-15 | 15 | Euk328f-Euk329r | < 3 | 118 |
| Coastal euphotic zone Accession numbers from EU785256 to EU785301 | 07°02'E | | CD-50 | 50 | Euk328f-CHLO02 | < 3 | 201 | |
| | | | ED-50 | 50 | Euk328f-Euk329r | < 3 | 196 | |
| Catalan, Mediterranean Sea, Spain | 41°40'N | 21-Aug-01 | RA1 | Surface | Euk528f-Euk329r | > 12 | 183 | |
| | | | | RA2 | Surface | ALV01- Euk329r | 0.2-3 | 96 |
| Coastal euphotic zone Accession numbers from EU785302 to EU785340 See also Massana <i>et al.</i> (2006) | 04°00'W | | RA3 | Surface | ALV01- Euk329r | > 12 | 96 | |
| | | | BAFRAC Tnico | Surface | EUKA-EukB | 0.6-1 | 38 | |
| Atlantic Ocean PROSOPE | Barcelona Harbour | 41°23'N | 28-Feb-01 | BAFRAC Tnano | Surface | EUKA-EukB | 1-10 | 29 |
| | | | | | BAFRAC Tmicro | Surface | EUKA-EukB | 10-50 |
| Coastal euphotic zone Accession numbers from EU785302 to EU785340 See also Massana <i>et al.</i> (2006) | Blanes Bay | 41°40'N | 06-Nov-01 | BLENRimf2 | Surface | EK-1A-516-GC | < 2 | 15 |
| | | | | | BLENRD2T4 | Surface | EK-1A-516-GC | Incubation after filtration < 2 |
| Atlantic Ocean PROSOPE | St UPW Morocco upwelling | 31°02'N | 09-Sep-99 | BLENRL2T4 | Surface | EK-1A-516-GC | Incubation after filtration < 2 | 21 |
| | | | | | CU-30 | Surface | Euk328f-CHLO02 | < 3 |
| Coastal euphotic zone Accession numbers from EU793942 to EU793987 See also Viprey <i>et al.</i> (2008) | Northern coastal Station 1 | 48°74'N | 19-Sept-04 | AMT15-1 | 10 | Euk328f-Euk329r | < 3 | 217 |
| | | | | | EU-30 | 30 | Euk328f-Euk329r | < 3 |
| Atlantic Meridional Transect (AMT 15) | Station 15 | 7°08'W | 04-Oct-04 | AMT15-15 | 10 | Euk328f-Euk329r | < 3 | 207 |
| | | | | | AMT15-27 | 30 | Euk328f-Euk329r | < 3 |
| Oceanic euphotic zone Accession numbers from EU780589 to EU780636 See also Zwirgmaier <i>et al.</i> (2007) | Southern gyre Station 27 | 17°83'N | 18-Oct-04 | AMT15-27 | 30 | Euk328f-Euk329r | < 3 | 119 |
| | | | | | AMT15-33 | 5 | Euk328f-Euk329r | < 3 |
| Oceanic euphotic zone Accession numbers from EU780589 to EU780636 See also Zwirgmaier <i>et al.</i> (2007) | Southern temperate Station 33 | 23°56'S | 24-Oct-04 | AMT15-33 | 5 | Euk328f-Euk329r | < 3 | 234 |
| | | | | | AMT15-33 | 5 | Euk328f-Euk329r | < 3 |

Table 1. cont.

| Environment and cruise | Station and Site | Coordinates | Date | Code library | Depth (m) | Primer | Sample details (μm) | Total number of clones with insert screened |
|---|---|-------------|-----------|--------------|-----------|-----------------|----------------------------------|---|
| Florida Straits, Atlantic Ocean Coastal euphotic zone Accession numbers from EU818567 to EU818691 and EU836584 to EU836597 | Station 1 | 25°30'N | 31-Mar-05 | FS01B | 5 | Euk328f-Euk329r | < 2 | 90 |
| | | 80°04'W | | | | | | |
| | | 25°30'N | 31-Mar-05 | FS01C | 70 | Euk328f-Euk329r | < 2 | 85 |
| | | 80°04'W | | | | | | |
| | | 25°30'N | 01-Aug-05 | FS01A | 5 | Euk328f-Euk329r | < 2 | 85 |
| | | 80°03'W | | | | | | |
| | | 25°30'N | 01-Aug-05 | FS01D | 65 | Euk328f-Euk329r | < 2 | 85 |
| | | 80°03'W | | | | | | |
| | | 25°30'N | 31-Mar-05 | FS04E | 5 | Euk328f-Euk329r | < 2 | 81 |
| | | 79°57'W | | | | | | |
| Florida Straits, Atlantic Ocean Oceanic euphotic zone Accession numbers from EU818078 to EU818566 and EU836598 to EU836612 | Station 4 | 25°30'N | 31-Mar-05 | FS04F | 85 | Euk328f-Euk329r | < 2 | 88 |
| | | 79°57'W | | | | | | |
| | | 25°30'N | 01-Aug-05 | FS04G | 5 | Euk328f-Euk329r | < 2 | 90 |
| | | 79°57'W | | | | | | |
| | | 25°30'N | 01-Aug-05 | FS04H | 89 | Euk328f-Euk329r | < 2 | 169 |
| | | 79°57'W | | | | | | |
| | | 25°29'N | 30-Mar-05 | FS14I | 70 | Euk328f-Euk329r | < 2 | 74 |
| | | 79°20'W | | | | | | |
| | | 25°29'N | 30-Mar-05 | FS14J | 5 | Euk328f-Euk329r | < 2 | 90 |
| | | 79°20'W | | | | | | |
| Sargasso Sea Atlantic Ocean Oceanic euphotic zone Accession numbers from EU817966 to EU818077 and EU836613 to EU836615 | Station 14 | 25°29'N | 31-Jul-05 | FS14K | 5 | Euk328f-Euk329r | < 2 | 69 |
| | | 79°20'W | | | | | | |
| | | 25°29'N | 31-Jul-05 | FS14L | 80 | Euk328f-Euk329r | < 2 | 83 |
| | | 79°20'W | | | | | | |
| | | 25°29'N | Dec-08 | FS14M | 58 | Euk328f-Euk329r | < 2 | 83 |
| | | 79°20'W | | | | | | |
| | | 25°29'N | Dec-08 | FS14N | 58 | Euk328f-Euk329r | \geq 2 | 81 |
| | | 79°20'W | | | | | | |
| | | 31°39'N | 29-May-05 | OC413BATS_O | 75 | Euk328f-Euk329r | < 2 | 90 |
| | | 64°37'W | | | | | | |
| Bermuda Atlantic Time-Series Station | Bermuda Atlantic Time-Series Station | | 01-Jun-05 | OC413BATS_P | 15 | Euk328f-Euk329r | < 2 | 87 |
| | | | 05-Jun-05 | OC413NSS_Q | 15 | Euk328f-Euk329r | < 2 | 84 |
| | | | 08-Jun-05 | OC413NSS_R | 70 | Euk328f-Euk329r | < 2 | 83 |
| | | | | | | | | 6874 |

For further information on the primers used, see also Table S2. Accession numbers are provided for dinoflagellate sequences only.

Table 2. Environment types represented by 18S rRNA gene sequences considered herein, and the proportion falling within the *Dinoflagellata*.

| Origin of the sequences (number of clone libraries considered) | Number of 18S rRNA gene sequences | Number of sequences belonging to dinoflagellates | Number of sequences per dinoflagellate group | | | | 'Core' dinoflagellates |
|--|---|--|--|-------------------------|-----------------------|---------------------------|---------------------------|
| | | | Syndiniales Group I | Syndiniales Group II | Syndiniales Others | 'Core' dinoflagellates | |
| Cultures and isolates ^a | 21 795 | 727 (3.3%) | 13 (1.8%) | 24 (3.3%) | 9 (1.2%) | 681 (93.7%) | |
| Env. Continental (150) | 5 711 | 35 (0.6%) | 0 | 0 | 35 (100%) | | |
| Plankton from the water column (42) | 1 213 | 35 (2.9%) | 0 | 0 | 0 | 35 (100%) | |
| Biofilm/microbial mats (16) | 392 | 0 | 0 | 0 | 0 | 0 | |
| Soil (39) | 3 361 | 0 | 0 | 0 | 0 | 0 | |
| Sediment (30) | 438 | 0 | 0 | 0 | 0 | 0 | |
| Others: aerosol, fermentor, gut communities . . . (23) | 307 | 0 | 0 | 0 | 0 | 0 | |
| Env. Marine (347) | 16 149 | 4809 (29.8%) | 1151 (23.9%) | 2411 (50.1%) | 156 (3.2%) | 1091 (22.7%) | |
| Plankton from the water column (262) | 13 528 | 4494 (33.2%) | 1073 (23.9%) | 2315 (51.5%) | 154 (3.4%) | 952 (21.2%) | |
| Sediment (46) | 1 624 | 139 (8.6%) | 54 (38.8%) | 18 (12.9%) | 0 | 67 (48.2%) | |
| Others: biofilm, incubation, gut communities . . . (39) | 997 | 176 (17.7%) | 24 (13.6%) | 78 (44.3%) | 2 (1.1%) | 72 (40.9%) | |
| Total | 43 655 | 5571 (12.8%) | 1164 (20.9%) | 2435 (43.7%) | 165 (3.0%) | 1807 (31.5%) | |

a. Including analysis of single cells.

Env. means environmental sequences. The number of clone libraries screened for each environment is given in brackets in the first column. In columns 4–7, brackets indicate, as a percentage, the number of sequences obtained for a particular dinoflagellate group compared with the corresponding environment (column 3). 'Core' dinoflagellates means all dinoflagellates except *Syndiniales*.

(150 clone libraries, mostly from freshwater, sediments and soils). Taken together, this resulted in 4844 dinoflagellate sequences from these environmental data sets, with 1164 and 2435 belonging to MALV groups I and II respectively (Table 2). An additional 727 sequences were derived from dinoflagellate cultures including a few from the amplification of single cells (see Table 2), most of these belonging to marine photosynthetic dinoflagellates (a bias already pointed out by Murray *et al.*, 2005). The mean length of the dinoflagellate sequences in the database was 950 nucleotides (range 108–1841 bp).

Phylogenetic analyses

Dinoflagellate sequences longer than 1600 bp (1018 sequences in total, 291 retained for the final tree, see the complete list of sequences in the Table S1) were used to perform phylogenetic analyses. Sequences belonging to the major *Alveolata* lineages were also included (Fig. 1). In order to avoid long-branch attraction artefacts, highly divergent alveolate groups were removed after preliminary phylogenetic analyses by neighbour joining (NJ). These divergent groups included *Haemosporida* (*Apicomplexa* including the human parasite *Plasmodium*), two ciliate groups, Mesodiniidae (*Myrionecta* and *Mesodinium*) and *Ellobiopsidae* (Silberman *et al.*, 2004). Highly divergent dinoflagellates, such as members of *Noctilucales* and *Oxyrrhis marina*, were also excluded. Bayesian phylogeny, using 1137 positions in the 18S rRNA gene sequence alignments, delineated four primary lineages within the *Alveolata* (Fig. 1), the ciliates, the *Apicomplexa*, the *Perkinsea* and the *Dinophyceae*. As previously observed (Leander and Keeling, 2003; Groisillier *et al.*, 2006), ciliates fell in the basal region of the tree, followed by *Apicomplexa* (Fig. 1). *Perkinsea* are the closest relatives of dinoflagellates. As frequently found in 18S rRNA gene phylogenies, many of these backbone nodes did not retain bootstrap support. At the basal part of *Dinophyceae*, several distinct taxa were placed in the bayesian analysis, here termed *Syndiniales* groups I–V, as a monophyletic lineage (Fig. 1). The general tree topology obtained with maximum likelihood (ML) was similar (data not shown). The existence of environmental MALV Group I and II was previously described (López-García *et al.*, 2001; Moon-van der Staay *et al.*, 2001), but here renamed (*Syndiniales* groups I and II) given more definitive placement within the *Syndiniales*. The genetic diversity and clade nomenclature of *Syndiniales* groups I and II are described in detail in separate analyses using partial sequences (Figs 2 and 3). For the definition of clades, we modified the general criteria chosen by Groisillier and colleagues (2006) so that a clade must (i) contain environmental sequences from at least 2 different clone libraries, and (ii) be bootstrap supported, at the defining nodes,

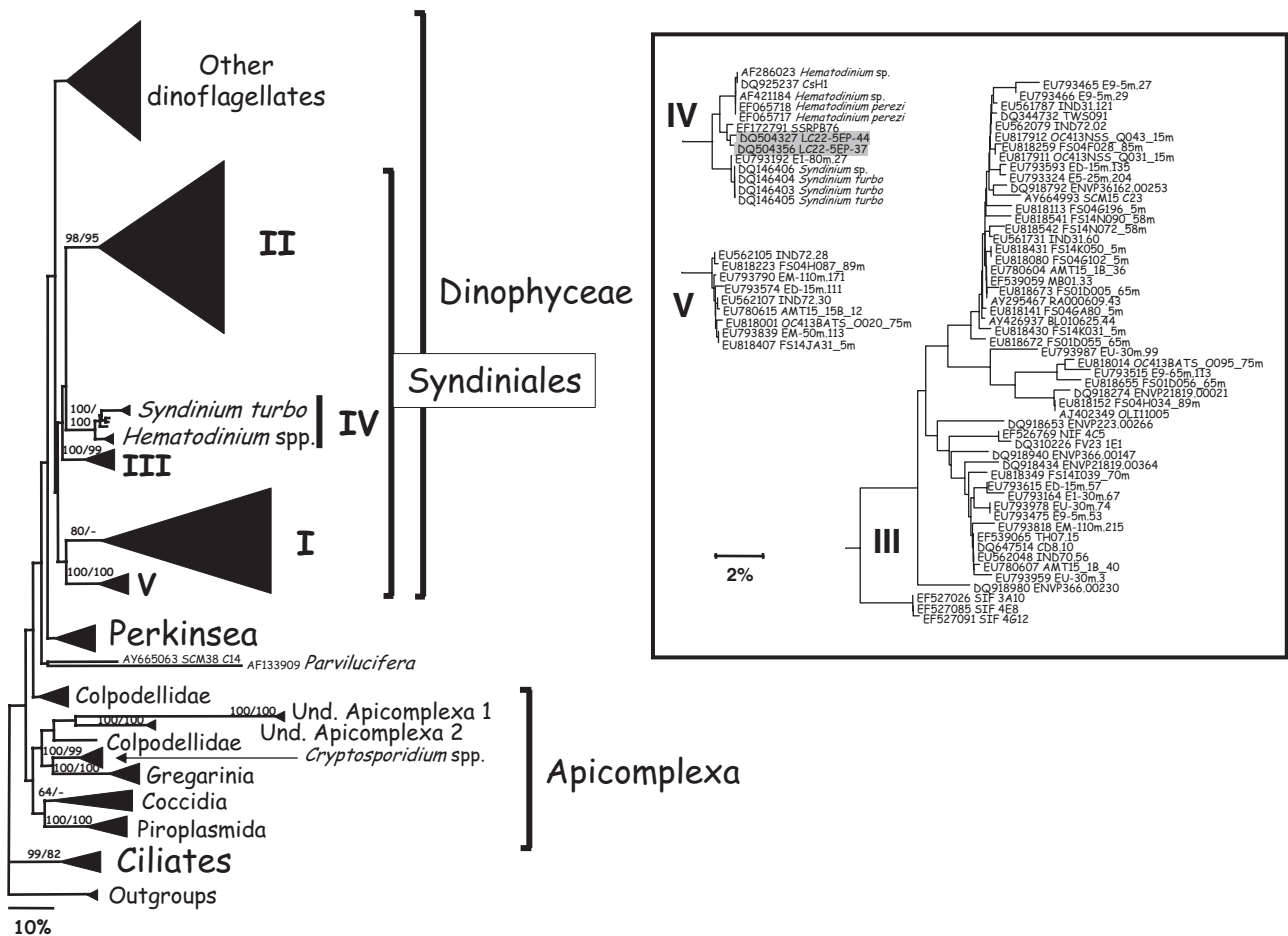


Fig. 1. Phylogeny of dinoflagellates using near-complete 18S rRNA gene sequences. Left: Bayesian phylogeny of alveolates based on analysis of 291 near full-length 18S rRNA gene sequences. Five sequences of *Bolidophyceae* (stramenopiles) were used as an outgroup. Gblock retained 1137 positions for the phylogenetic analyses. Bootstrap values, given at the principal nodes of the tree, correspond to neighbour-joining and maximum parsimony analyses respectively (1000 replicates, values > 60% shown). For neighbour-joining bootstrapping, a GTR + G + I model was selected with the following parameters: Lset Base = (0.2757 0. 1787 0.2494), Nst = 6, Rmat = (1.0746 2.9600 1.2514 1.0820 4.7079), Rates = gamma, Shape = 0.7091, Pinvar = 0.2330. The scale bar corresponds to 10% sequence divergence. Inset (at the right): Details of groups III to VI, based upon analysis of partial (500 bp) sequences (59 sequences including outgroups). The scale bar corresponds to 2% sequence divergence. The two sequences in grey belonging to Group IV are from hydrothermal vents.

by values greater than 60% in NJ and maximum parsimony (MP) analysis. Some exceptions to these rules were made as detailed below. The mean sequence identity within a clade was 87% for *Syndiniales* Group I (ranging from 76.6% for clade 5 to 91.9% for clade 7) and 93.5% for *Syndiniales* Group II (ranging from 80.8% for clade 8 to 99.4% for clades 35 and 43, see also Fig. S1). *Syndiniales* Group I contained eight different clades (Fig. 2). Clades 1–5 were described previously by Groisillier and colleagues (2006), while three new clades emerged from this study. All of these eight clades are supported by bootstrap values > 60%, except for Clade 3, which is only supported by MP, bootstrap analyses (74%). Nevertheless, the tree topology is identical with the two different methods (NJ and MP) and minimal sequence identities within this clade are inside the range of other clades

belonging to this group (see Fig. S1). *Ichthyodinium chabreladi*, a parasitoid of fish eggs, belongs to Group I Clade 3, while *Duboscquella* spp., a parasitoid of tintinnides, belongs to Group I Clade 4 (Harada *et al.*, 2007). *Syndiniales* sequences retrieved by single-cell PCR on radiolarian isolates (Dolven *et al.*, 2007) belong to Clade 1 (DQ916408, though not in the tree because this sequence contained several nucleotide ambiguities) and Clade 2 (DQ916404–DQ916407 and DQ916410). Clades 5–8 are only composed of environmental sequences. Some sequences within clades 1–4 (highlighted in grey in Fig. 2) were retrieved exclusively from suboxic and anoxic ecosystems. Together, clades 1 and 5 were the most commonly retrieved from environmental clone libraries, representing ≥ 75% of sequences belonging to *Syndiniales* Group I.

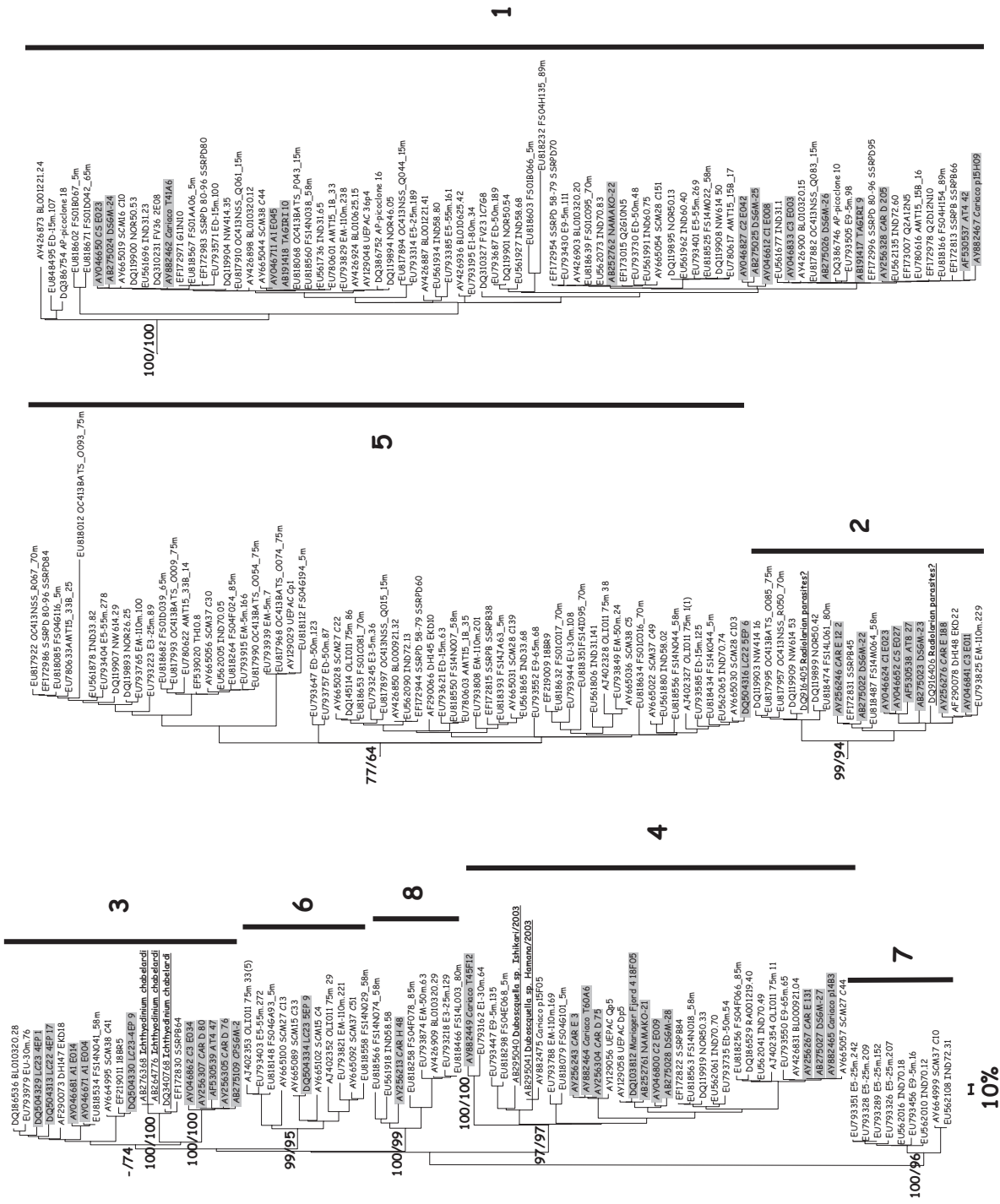
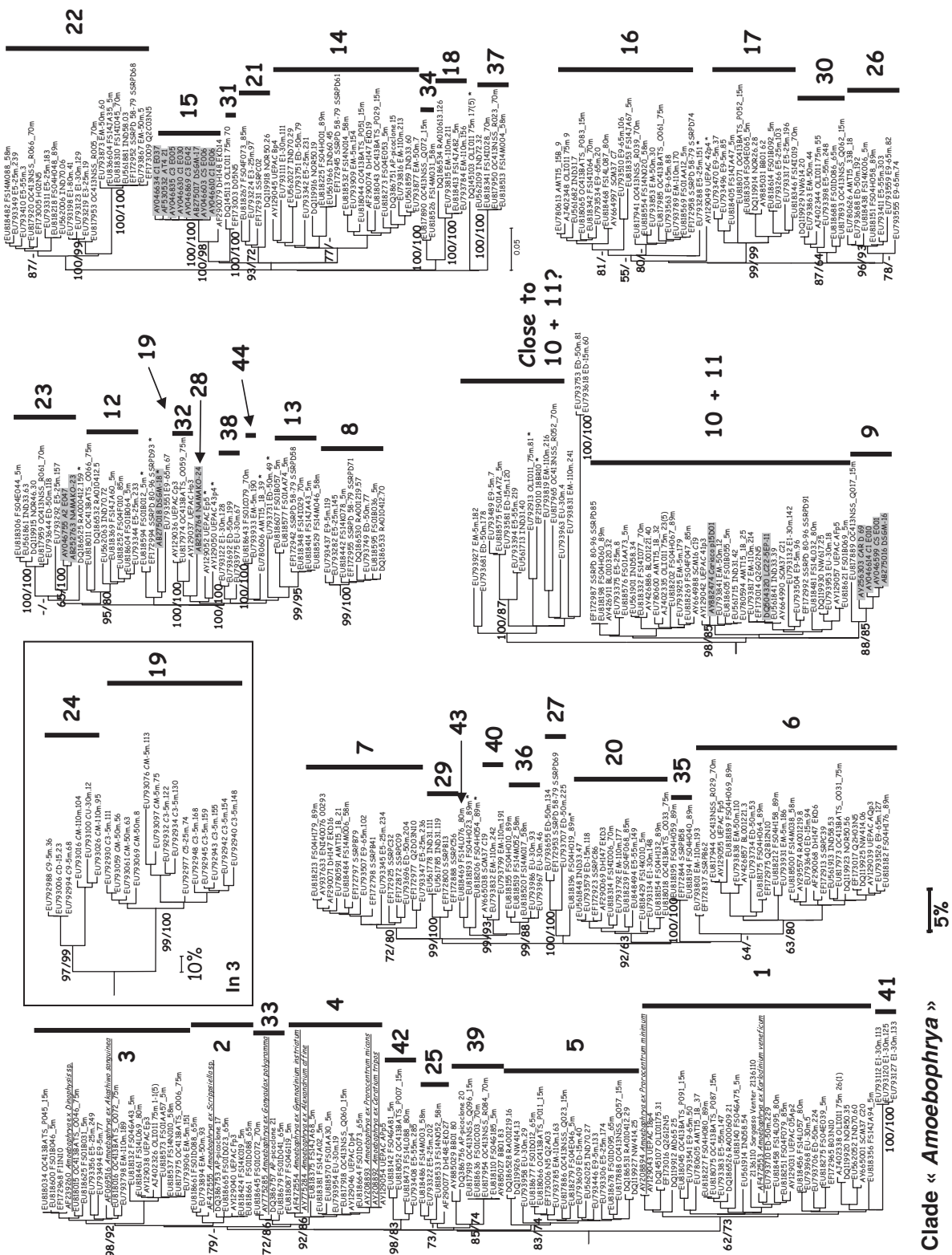


Fig. 2. Phylogeny of *Syndiniales* Group I. Neighbour-joining phylogeny of Alveolate Group I, based upon the analysis of 275 partial sequences, 731 bp in length (including outgroups, not shown). A GTR + G model was selected using the following parameters: Lset Base = (0.2627 0.1810 0.2622), Nst = 6, Rmat = (1.0000 2.8775 1.0000 4.6908), Rates = gamma, Shape = 0.5006, Pinvar = (0). Bootstrap values, given at the principal nodes of the different clades, correspond to neighbour-joining and maximum parsimony analyses respectively (1000 replicates, values > 60% shown). The scale bar corresponds to 10% sequence divergence. The tree was cut in order to better separate the major clades in the figure. Sequences in grey are from hydrothermal and suboxic ecosystems.



Clade « Amoebophrya »

Fig. 3. Phylogeny of *Syncliniales* Group II. Neighbour-joining phylogeny of 423 partial sequences, 738 bp in length and centred at the 5' end (including outgroups, not shown). A GTR + G model was selected using the following parameters: Lset Base = (0.2736 0.1700 0.2545), Nst = 6, Rmat = (1.0000 3.3780 1.2804 1.2804 4.9683), Rates = gamma, Shape = 0.6108, Pinvar = 0.1155). Bootstrap values, given at the principal nodes of the different clades, correspond to neighbour-joining and maximum parsimony analyses respectively (1000 replicates, values > 60% shown). The asterisk ** indicates singletons (sequences that are not included in any clades). The scale bar corresponds to 5% sequence divergence. The tree was cut in order to better separate the major clades in the figure. Inset: a small portion of the neighbour-joining tree obtained using the 3' end of the 18S rRNA gene of *Alveolate* Group II, based upon analysis of 365 partial sequences, 617 bp in length (including outgroups). The scale bar corresponds to 10% sequence divergence. Bootstraps values correspond to neighbour-joining analyses (1000 replicates). Sequences in grey are from hydrothermal and suboxic ecosystems.

By comparison, *Syndiniales* Group II is genetically more diverse than Group I. In addition to clades 1–16 (see Groisillier *et al.*, 2006), 28 new clades emerged from the present study (clades 17–44, Fig. 3). Some of these clades are not supported by bootstrap analyses obtained with one of the two phylogenetic methods tested (clades 2, 6, 10+11, 14, 16, 23, 25 and 26), but in such cases the tree topologies are identical between the two methods. Clade 10 and Clade 11 described by Groisillier and colleagues (2006) are now merged into a single clade (herein named Clade 10+11), which includes clone BL10320.32 (a singleton in the work of Groisillier *et al.*, 2006). However, Clade 10+11 is not supported by bootstrap analyses. Numerous environmental sequences displaying long branches (labelled ‘close to 10+11?’ in Fig. 3) also appear to be closely related to Clade 10+11.

Only a single genus within *Syndiniales* Group II has been formally described, the *Amoebophrya*. All *Amoebophrya* sequences obtained from cultures and single-cell analyses formed a monophyletic group as recognized by Groisillier and colleagues (2006). This monophyletic group now also includes clades 1–5 as well as clades 25, 33, 39, 41 and 42 (Fig. 3). Recently, Kim and colleagues (2008) recognized nine different subgroups within *Amoebophrya*. Subgroup 1 (from the analysis of Kim and colleagues) corresponds primarily with our Clade 1 (except sequence AF290077 which belongs to the novel Clade 25) while subgroups 3, 4, 5 and 6 correspond to clades 4, 3, 5 and 2 respectively. The recognition of subgroup 2 (identified in Kim *et al.*, 2008) is probably skewed by inclusion of a sequence we identified to be a chimera (DQ186527). Thus, sequence AY260468, obtained from the direct amplification of *Ceratium tripos* infected with *Amoebophrya*, likely remains a singleton. Sequence AY295690 (subgroup 8) is also a probable chimera, whereas sequences included within subgroups 7 and 9 belong to Clade 33 in the present study. Sequence DQ916402, obtained from the direct amplification of a cell of spumellarida *Hexacantium giganteum* (Radiolaria), is closely related to Clade 6. This sequence was removed from our global analyses due to the high number of nucleotide ambiguities it contained. Clades 15 and 9 are primarily composed of environmental sequences from anoxic or suboxic ecosystems (highlighted in grey in Fig. 3). Both clades also include sequences retrieved from deep-sea methane cold seeps (DSGM-16 and 17, Takishita *et al.*, 2007) and deep oceanic waters (Countway *et al.*, 2007), but these sequences were not included in this analysis due to their short length. Clades originally containing only sequences from coastal systems (i.e. clades 2, 4, 5, 8, 12 and 13; see Groisillier *et al.*, 2006) are shown here to also include clones from other oceanic regions, including oligotrophic waters. This highlights the importance of gaining sequence representation from a

broader array of geographical locations. Finally, in terms of the number of sequences, the most commonly found clades in rank order were clades 10+11, 7, 6, 1, 19, 3, 22 and 16.

Other *Syndiniales* groups (III–V) also emerged from the present study (detailed in Fig. 1). *Syndiniales* Group III contains 71 environmental sequences, including clone OLI11005 (AJ402349) that was previously placed outside Group II (see Groisillier *et al.* 2006), as well as numerous environmental sequences retrieved from various oceanic surface waters (Mediterranean Sea, Indian Ocean, Sargasso Sea), coastal waters [Southern Taiwan Strait, Blanes Bay (Spain) and the English Channel] and from a supersulfidic anoxic fjord (DQ310226). *Syndiniales* Group IV contains the genera *Syndinium* and *Hematodinium*, five closely related environmental sequences, a sequence retrieved from the Mediterranean Sea, E1–80m.27, closely related to *Syndinium turbo* and a clade allied with the genus *Hematodinium* and composed of sequences from the Sargasso Sea (EF172791 and DQ918583) and from deep sea hydrothermal vents (DQ504327 and DQ504356). *Syndiniales* Group V contains 67 environmental sequences, all collected within the euphotic zone, but includes clones retrieved from very different oceanic ecosystems (Indian Ocean, Atlantic Ocean, Mediterranean Sea and Sargasso Sea). Sequence BL000921.23 from Blanes also belongs to *Syndiniales* Group V (not included here due to sequence length dissimilarities). Finally, two partial environmental sequences could not be assigned to any specific group: EU793524 (E9–65m.123) and EU818559 (F14DEC+), but phylogenetic analysis placed them closely related, although separated from these *Syndiniales* groups.

Ecological distributions

Dinoflagellate sequences represent less than 1% of environmental sequences from continental ecosystems (aquatic and terrestrial), whereas they represent close to 30% of the sequences obtained from marine ecosystems. With respect to aquatic environments (both marine and freshwater), dinoflagellate sequences as a whole were obtained primarily from the plankton, as opposed to other environments such as sediments (Table 2). Sequences belonging to *Syndiniales* groups I and II were absent in samples collected to date from continental ecosystems, however, they represent the largest portion of dinoflagellate sequences from marine systems. *Syndiniales* Group II alone represented about half of all environmental dinoflagellate sequences (Table 2).

We divided marine ecosystems included in our sequence database into various categories: (i) anoxic and suboxic ecosystems, (ii) ecosystems with hydrothermal activities, (iii) sediments, or (iv) the water column

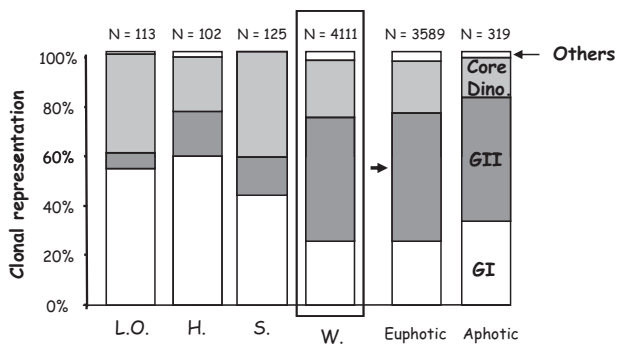


Fig. 4. The relative contribution of dinoflagellate sequences within environmental clone libraries. Left: relative contribution (in percentage) from (i) low oxygenated (L.O.) ecosystems (sediments or deep waters), (ii) Hydrothermal (H) vent ecosystems (collected directly in the hydrothermal chimney, on collectors or on sediments close to the chimney), (iii) sediments (S) that can be oxic or anoxic, deep or more coastal, and (iv) plankton from the water column (W). For this last category, two additional conditions were compared (right side of Fig. 4): whether the water sample was collected from the euphotic or the aphotic layer. N = Total number of environmental sequences considered.

(wherever possible, we also specified whether sequences were derived from the aphotic or euphotic zone; Fig. 4). In cases where these categories overlapped (e.g. anoxic sediments collected from deep hydrothermal vents, or anoxic water columns), environmental sequences were included in both categories. Considering all dinoflagellate sequences, *Syndiniales* Group II sequences were more abundant in clone libraries derived from the water column (planktonic ecosystems) than in other habitats, while the relative importance of the remaining dinoflagellate sequences (including *Syndiniales* Group I) was greater in anoxic environments, hydrothermal vents and sediments (Fig. 4). Within planktonic ecosystems, the relative contribution to clone libraries of *Syndiniales* Group I and II compared with other dinoflagellate sequences was similar between euphotic and aphotic waters (Fig. 4).

Environmental sequences derived from the euphotic zone, both coastal and oceanic waters, have generally been recovered using three different primer sets Euk328/Euk329r, EukA/EukB and EukA/EukB' (for primer references see Table S2). Although the contribution of dinoflagellate sequences to the total number of clones is

very similar using either the EukA/EukB or EukA/EukB' primer sets, the relative contribution of Group II is higher using the Euk328f/Euk329r primer set (Fig. 5). However, the distribution of *Syndiniales* Group I and II clades within the euphotic zone, as evidenced from these clone libraries, is quite similar with the different primer sets (Fig. 6). According to clone library composition, sunlit surface marine waters are dominated by *Syndiniales* Group I clades 1, 4 and 5, and Group II clades 1, 6, 7 and 10+11 (Fig. 6). Comparing sequences obtained using the same primer sets, *Syndiniales* clade distributions within the aphotic zone are quite different from surface waters, mainly dominated by *Syndiniales* Group I clades 1, 2 and 3 and by *Syndiniales* Group II clades 6 and 7 (Fig. 6).

Discussion

Our phylogenetic analysis of available 18S rRNA gene sequences from described parasitic genera (*Amoeboophrya*, *Ichthyodinium*, *Duboscquella*, *Syndinium*, *Hematodinium*) and many environmental sequences, resolved for the first time the *Syndiniales* as a monophyletic lineage at the base of the more classical dinoflagellates. The basal position is not supported by bootstrap analyses, but the same topology was generated by Bayesian and ML phylogenies. This topology agrees with historical descriptions of the *Duboscquodina* as a tribe (a taxonomic category between a genus and a subfamily) by Cachon (1964), and *Syndiniales* as an Order by Loeblich (1976). However, at higher taxonomic levels, *Syndiniales* are still difficult to place. Historically, *Syndiniales* have been retained within dinoflagellates because their short-lived dispersal stages, called dinospores, have a classic naked gymnoid morphology while also their nucleus has the chromatin condensed during part of the life cycle, like the dinokaryon of dinoflagellates. *Syndiniales* have trichocysts, a typical feature for *Dinophyceae*. However, Loeblich (1976) separated them from the rest of *Dinophyceae* (named also 'core' dinoflagellate) and placed *Syndiniales* inside the class *Syndiniophyceae* due to peculiarities of nuclear division in *Syndiniales* (Ris and Kubai, 1974; Soyer, 1974). Nevertheless, for all *Dinophyceae* and *Syndiniales*, chro-

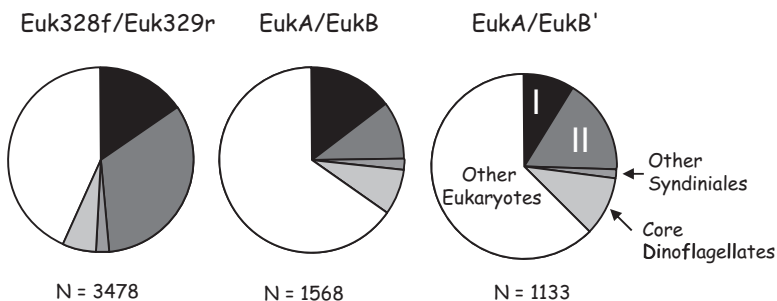


Fig. 5. Proportion of dinoflagellate sequences retrieved using different primer sets. Relative proportion of environmental sequences belonging to *Syndiniales* (I, II and others = III–VI) and the rest of dinoflagellates (= Core Dinoflagellates) compared with other eukaryotes retrieved using different primer sets during PCR amplification. For description of primers, see Table S2. Only clone libraries constructed from the euphotic zone of marine waters have been considered. N = the total number of sequences.

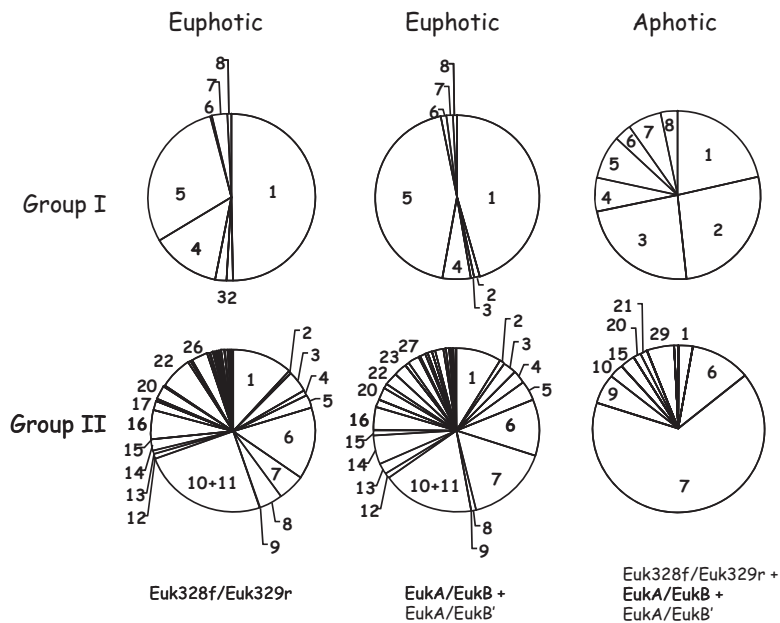


Fig. 6. *Syndiniales* clades retrieved using different primer sets, in the euphotic and aphotic layers of marine waters. The relative proportion of clades belonging to *Syndiniales* (groups I and II) using different sets of primers is indicated. Only clone libraries constructed from the euphotic zone of marine waters have been included (as in Fig. 5). N = the total number of sequences.

mosomes stay anchored to the persistent nuclear envelope during mitotic division, but their segregation is driven by microtubules through the nuclear envelope in *Dinophyceae* while centrioles participate in the formation of the mitotic apparatus in *Syndiniales* (Hollande, 1974; Ris and Kubai, 1974). Another *Syndiniales* peculiarity is their low chromosome number, 4–10 for *Syndiniales* compared with >100 for some *Dinophyceae* members such as *Gonyaulax* and *Ceratium* (see Loeblich, 1976). Taking all this together, it seems reasonable that these last two features (peculiarities in nuclear division and low chromosome number) are derived from an initial dinoflagellate model (Xiao-Ping and Jing-Yan, 1986), being forced by the obligately parasitic lifestyle of *Syndiniales*. Controversially, some recent phylogenies using 28S rRNA gene sequences, have placed the *Syndiniales* *Ichthyodinium chaberladi* (Group I Clade 3) within dinoflagellates (Gestal *et al.*, 2006) and an environmental sequence belonging to Group I clade 1 close to *Perkinsus* (Massana *et al.*, 2008). More sequences are needed to obtain a LSU rRNA phylogeny robust enough to provide accurate affiliations for the *Syndiniales*. Protein coding genes, such as actin and β - and γ -tubulins may also be a better alternative to resolve basal branches of the dinoflagellate lineage (Saldarriaga *et al.*, 2003), including several deep genera such as *Perkinsus*, *Oxyrrhis* and *Noctiluca*, which are almost impossible to place using phylogenies based upon ribosomal genes.

Within *Syndiniales*, five main families have been described: the *Sphaeriparidae* and the *Coccidinidae* (no sequences available to date for both), the *Amoebophryidae*, the *Duboscquellidae* and the *Syndinidae* (Table S3). As *Ichthyodinium* (*Syndinidae*) and *Duboscquella* (*Duboscquellidae*) are clearly allied within Group I by 18S

rRNA gene analyses, while *Hematodinium* and *Syndinium* (*Syndinidae*) are members of Group IV and quite distinct from Group I, it is likely that the taxonomy of *Syndinidae* will soon be modified. Nevertheless, it is possible that *Amoebophryidae* will remain synonymous with *Syndiniales* Group II. To date, seven different species of *Amoebophrya* have been described (see Table S3). However, this clearly does not reflect the genetic diversity observed in Group II; herein we find at least 44 distinct clades. Unfortunately, only one of the described species, *A. ceratii*, has been genetically characterized to date. *Amoebophrya ceratii* is a parasitoid of many (if not all) dinoflagellates, including heterotrophic species (for a review see Park *et al.*, 2004) and was most recently reported as being a 'complex species' rather than a real species (see Coats *et al.*, 1996). Currently, available SSU rRNA gene sequences from this complex species are from strains able to infect photosynthetic dinoflagellates, in accordance with the fact that this group is largely retrieved from the euphotic zone in environmental studies. Nevertheless, *A. ceratii* was also described as being able to parasitize other *Amoebophrya* species (e.g. it is frequently observed in *A. leptodisci*) or other *Syndiniales* (such as the genus *Keppenodinium*; see Cachon 1964). Consequently, despite recent additions of a substantial number of new sequences, all *Amoebophrya* SSU rRNA gene sequences fall within a monophyletic lineage, which contains at least 10 different clades. However, the taxonomic correspondence of each clade is still not clear. Within *Hematodinium* and *Syndinium*, SSU rRNA gene sequences from different species are almost identical, and species can only be distinguished using more variable genetic regions (ITS, see Hudson and Adlard, 1996;

Skovgaard *et al.*, 2005). In the case of *Amoebophrya*, genetic variability is much more evident, even using conserved markers such as the SSU rRNA gene.

All described *Syndiniales* are parasitoids, and obligately kill their host. A potential exception is *Sphaeripara catenata*, which in some cases only causes castration of its host, but not death (Cachon and Cachon-Enjumet, 1964). Described *Syndiniales* generally complete their life cycle in less than 3 days, leading to the release of hundreds of free-living dispersive dinospores. Environmental sequences obtained from genetic libraries likely result from such dinospores. In general, *Syndiniales* produce two different types of dinospores, the micro and macropores that can both be very small (from 1 to 12 µm in diameter depending on the species). There is a general agreement that dinospores can only survive a few days after their release (e.g. Coats and Park, 2002). Hence, the fact that they systematically dominate clone libraries from coastal and oceanic waters, whatever the season considered, suggests that their production is fairly constant. Alternatively, it could mean that other more resistant planktonic stages are produced. To our knowledge, cyst production has never been reported within *Syndiniales*, with the notable exception of cyst-like cells described in natural populations of *Duboscquella cachoni* infecting the tintinnid *Eutintinnus pectinis* (Coats, 1988). In this particular case, cyst-like formation was completely independent of dinospore production, and the cysts formed were non-motile. Within dinoflagellates, permanent cysts are produced by the fusion of two gametes. Sexuality, issued from different individuals (anisogamous and heterothallic), has only been observed once in *Coccidinidae*, with the production of a planozygote-like body (Chatton and Biecheler, 1936). Nevertheless, cyst production was not observed.

The recent reports of *Amoebophrya* spp. in freshwater systems (Lefèvre *et al.*, 2008) are based on placement of two environmental sequences incorrectly labelled uncultured *Amoebophrya* Clone E and F (Di Giuseppe and Dini, 2004). These sequences are in fact closely related to the genus *Cryptocaryon* (Ciliates, Protosmatea). Thus, to date, environmental sequences of *Syndiniales* have been retrieved solely from marine ecosystems. This supports the 'marine-ness' of this group, together with the fact that all described *Syndiniales* species are marine. In contrast, other unicellular alveolate parasites, such as *Perkinsea*, *Colpodellids* or even other parasitic dinoflagellates, have both marine and continental species. This is interesting given that almost all known *Syndiniales* hosts have species well adapted to continental ecosystems (within ciliates, dinoflagellates, cercozoa and crabs).

Although the distribution of *Syndiniales* is relatively homogeneous at the group level (note their similar contribution to both euphotic and aphotic zones), communi-

ties are quite different at the clade level. Some clades are uniformly distributed vertically through the water column (e.g. Group I clade 1, Group II clade 6), whereas others (e.g. Group I clades 2 and 3, Group II clade 7) are more prevalent in aphotic layers. We speculate that these 'deeper' clades are able to parasitize well-known deep planktonic organisms such as *Phaeodarea* (*Cercozoa*), *Acantharea* (*Radiolaria*) and *Polycystinea* (*Radiolaria*). Recent analyses of environmental sequences belonging to these groups highlight the increasing contribution of radiolarian environmental sequences with depth (Not *et al.*, 2007), in particular sequences belonging to *Spumellarida* (*Polycystinea*). Interestingly, radiolarians are also exclusively marine, and well known to be oceanic, or blue-water, organisms. Hence, both *Syndiniales* and radiolarians may have shared a common evolutionary history inhabiting similar ecosystems as discussed by Not and colleagues (2007). Sequences retrieved from single radiolarian cells (Dolven *et al.*, 2007), likely infected by *Syndiniales*, belong to both Group I and Group II (Fig. 3). Nevertheless, it is difficult to assign an identity to these sequences as *Syndiniales* able to infect radiolarians are described within three of the five described families (the *Amoebophryidae*, the *Duboscquellidae* and the *Syndinidae*). Much more information on host specificity is provided by data from the genus *Amoebophrya*. Phylogenies of *Amoebophrya* clades (Fig. 3) indicate that host-parasite co-evolution is not a central feature driving the observed diversity. Parasites of the same host genus, but of different species, belong to completely separate clades. In fact, within the 'A. ceratii complex', host specificity as assessed in cultures, is highly strain dependent, with some being extremely specific (Coats and Park, 2002), while others have broader host ranges (Kim, 2006). This fact could mask clear co-evolutionary patterns. The same conclusion can be drawn when the entire *Syndiniales* order is considered. For example, all described *Syndiniales* that operate as parasitoids of metazoans are not closely related. The genus *Ichthyodinium*, a parasite of fish eggs, is more closely related to parasites that infect ciliates (i.e. *Duboscquella*) belonging to Group I than to Group IV, which contains *Syndinium* and *Hematodinium* (two genera which infect metazoans). Thus, even between groups, *Syndiniales* and their hosts do not seem to co-evolve. Consequently, we expect that parasites within *Syndiniales* are generalists or opportunists with respect to host range. This is also congruent with the detection of clades widely distributed in all explored marine ecosystems. Nevertheless, the presence of more specialized clades or subclades, related with depth or specifically retrieved from extreme environments, such as Group I clades 3 and 4, and Group II clades 9 and 15, confirms that specific interactions may exist.

Conclusions

Based on the integration of a large number of environmental sequences, we propose a new phylogeny for *Syndiniales*. This order should be considered a monophyletic sister group of the rest of *Dinophyceae* based upon 18S rRNA gene phylogenies. The *Syndiniales* are composed of five main groups, which include environmental sequences formerly referred to as novel marine alveolate groups I and II. All *Syndiniales* species described to date are parasitic marine organisms. The exclusively marine lifestyle of this order as a whole, including sequences from uncultured organisms, is confirmed by the fact that their 18S rRNA gene sequences have not been retrieved from terrestrial or freshwater PCR surveys, but rather come solely from marine ecosystems. *Syndiniales* appear to colonize all the marine habitats investigated thus far, from oceanic surface waters to sediments. Some clades are likely adapted to fairly specific habitats (and/or hosts that reside only in specific habitats). However, these parasites and their hosts do not appear to share co-evolutionary trends. This suggests that at least some *Syndiniales* are highly opportunistic, with a capacity to infect very different marine hosts, from many trophic levels within marine food webs. Even so, we cannot completely exclude that some members of the *Syndiniales* may have retained specific ancestral traits still found among the different alveolate lineages, e.g. phagotrophy and/or the inheritance of photosynthetic genes. In fact, proteins characteristic of secondary endosymbiont plastids have recently been detected from both the non-photosynthetic genera *Oxyrrhis* and *Perkinsus* (Matsuzaki *et al.*, 2008; Slamovits and Keeling, 2008). Hence, the ecological success of the *Syndiniales* may yet have a more surprising and complex explanation. That said, it is certainly still the case that the overall role of parasitism in marine systems requires further investigation and incorporation into food web and carbon flow models.

Experimental procedures

Environmental sampling and library processing

Materials and methods for each of the libraries published herein varied by site. Thus, details are provided separately for each of the regions sampled. See also Table 1 for additional details on the clone libraries constructed.

English Channel, France (RA). Ten litres of surface sea water was sampled weekly at the ASTAN site (SOMLIT station, English Channel, France, see http://www.domino.u-bordeaux.fr/somlit_national/) from the end of June through to the middle of September 2004. Water samples were size-fractionated using a peristaltic pump (with the flow rate fixed

at 100 ml min⁻¹) through sequential 47-mm-diameter polycarbonate filters (12 and 3 µm, Osmonic Poretics Products) in filter housings connected in series, terminated by a 0.2-µm-pore-size Sterivex unit (Millipore). Following sample collection, filters were then submerged in DNA lysis buffer (0.75 M sucrose, 40 mM EDTA, 50 mM Tris-HCl, pH 8), immediately frozen in liquid nitrogen and stored at -80°C. DNA was extracted using a 3% (w/v) CTAB (Cetyltrimethylammonium bromide) extraction procedure (Doyle and Doyle, 1987). PCR conditions for amplifying the 18S rRNA gene were as described in Moon-van der Staay and colleagues (2001). PCR products were cloned using the TOPO-TA cloning kit (Invitrogen). Clones (96 randomly selected) were partially sequenced using the primer Euk528f for RA1 and the primer ALV01 (for primer references see Table S2) for both RA2 and RA3 on an ABI Prism 3100 (Applied Biosystems).

Barcelona Harbour, Mediterranean Sea, Spain (BAFRACT). Fifty litres of sea water was collected on 28 February 2001 from Barcelona harbour (Spain) and pre-filtered through a 50 µm mesh. On return to the laboratory (less than one hour), the sample was then fractionated through seven different pore-sizes (20, 14, 10, 5, 3, 1 and 0.6 µm), using similar filter housings connected in series as described above. Sample storage and DNA extraction were as described above. Genetic diversity in each size fraction was determined by DGGE (not shown), as described by Díez and colleagues (2001b). Clone libraries were performed for size fractions displaying the most 'contrasting' DGGE band patterns. PCR conditions were as in Díez and colleagues (2001a). Cloning was as described above. Restriction fragment length polymorphism (RFLP) analysis was used to detect polymorphisms among the retrieved clones. Briefly, PCR products were digested with 1 U µl⁻¹ of restriction enzyme HaeIII (Gibco BRL) for 6–12 h at 37°C. The digested products were separated by electrophoresis at 80 V for 2–3 h in a 2.5% low-melting-point agarose gel. One clone per unique RFLP pattern was partially sequenced using the internal primer Euk528f. Sequencing was performed by Qiagen Genomics Sequencing Services (Germany).

Blanes Bay. Surface water was collected in Blanes Bay (Catalan coast, north-western Mediterranean Sea). Sea water was rapidly transported to the laboratory (less than 2 h), and gently filtered by gravity through 2- and 5-µm-pore-size polycarbonate filters. One litre of each size fraction was incubated both in the light [around 100 (µmol quanta) m⁻² s⁻¹] and in the dark (four bottles in total). All the incubations were carried out at 20°C, close to *in situ* temperature (18°C). Samples (100 ml) were filtered through 0.2-µm-pore-size Durapore filters for DNA analysis over a course of 5 days. DNA was then extracted following the protocol described in Díez and colleagues (2001b). Abundance and genetic diversity of eukaryotes were monitored by epifluorescence microscopy and DGGE (using the same protocol as described above, data not show). 18S rRNA genes were amplified from the initial day and after 4 days of incubation both in the light and dark. PCR products were cloned using the TOPO-TA cloning kit. Clones were sequenced by the Qiagen Genomics Sequencing Services (Germany).

Atlantic Meridional Transect (AMT15). Both the sampling strategy, DNA extraction procedure and the PCR amplification protocol used have been previously published (Zwirgmaier *et al.*, 2007). Cloning was as described above. Restriction fragment length polymorphism analysis was used to screen for distinct inserts by digesting with HaeIII in a total volume of 20 µl using 5 U of HaeIII at 37°C for 2 h. Fragments were resolved by electrophoresis in a 2.5% agarose gel at 60 V for 3 h. At least two representatives of each RFLP pattern were sequenced in entirety following purification of the PCR product using the QIAquick PCR Purification Kit (Qiagen), on an ABI 3130xl Genetic Analyser (Applied Biosystems).

Florida Straits and Sargasso Sea. Sea water was collected from both the surface and deep chlorophyll maximum on three transect cruises across the Straits of Florida in 2005 as described in Cuvelier and colleagues (2008). Briefly, 1 l samples were first gravity filtered through 47 mm diameter, 2-µm-pore-size filters (GE Osmonics, Minnetonka, Minnesota, USA) and then onto a 0.2-µm-pore-size Supor filter (Pall Gelman, Ann Arbor, Michigan, USA) using approximately 5 psi vacuum. Filters were placed in cryovials and frozen in liquid nitrogen. DNA was extracted using the DNeasy kit according to the manufacturer's recommendations (Qiagen, Germantown, MD, USA). PCR conditions have been published elsewhere (Cuvelier *et al.*, 2008). The PCR product was ligated into vector pCR2.1 (Invitrogen, Carlsbad, CA, USA) and transformed. Clones were then sequenced on an ABI 3730xl sequencer (Applied Biosystems).

Samples were also collected on a cruise from Woods Hole, MA USA to the BATS. Two regions were sampled, BATS and a site north-west of BATS. Samples were collected and processed as for the Florida Straits clone libraries (as reported in Cuvelier *et al.*, 2008).

Development of sequence database

The 18S rRNA gene database was built using all published environmental sequences deposited in GenBank up to December 2007. We also integrated the data sets generated herein and a number of other libraries which are being published independently, but without specific attention to the analysis of *Alveolata* sequences. The complete list of sequences used in this study and their annotation is available upon request (Excel file).

Sequence group assignments

The software package, KeyDNATools, was used to (i) detect sequences belonging to dinoflagellates within a large database including more than 42000 eukaryotic SSU rRNA gene sequences; (ii) remove potential chimeras; and (iii) automatically assign sequences to a specific clade. Details of the software development will be described elsewhere. Briefly, environmental sequences were annotated using small oligonucleotides 'probes' (15 bp length), named 'keys', which were generated *in silico*. The specificity of each key, from the Order to the clade level (where possible) was automatically deduced from a reference database containing as many

eukaryotic lineages as possible as outgroup and at least two sequences from each taxa created within marine dinoflagellates. In this study, more than 110000 keys were generated, each being specific at least to the level of Order. Annotations of 18S rRNA genes were deduced from the specificity of each key when they matched a sequence. As an example, sequence EF172938 was annotated by 175 different keys, providing converging annotation at the order level and covering positions 1–1289 along the sequences (Table 3). Thus, this sequence was considered as a *Syndiniales*, Dino-Group II, and belonging to Dino-Group II-Clade 7 (this last annotation is provided by 47 keys). The 18S rRNA gene sequences belonging to the majority of *Dinophyceae* were not variable enough to allow a clear separation of sequences even at the Order level. Thus, keys specific to most of the *Dinophyceae* are labelled as *Dinophyceae* at the Class and the Order levels (see annotation of sequence EU087283). Wherever possible, more precise information is given (e.g. at the Family, Genus, Species and Clade levels; see annotation of sequence EU087283). New groups/clades were often detected by conflicting annotation (caused by incorrect deduction of specificity by the keys) in a particular taxonomic field. Chimeras were detected when annotations conflicted for separate parts of an individual sequence. As an example (Table 3), sequence AJ965100 was detected as a chimera at the class level, a false gene sequence combining part of a *Ciliophora* (annotation provided by 25 keys covering positions 44–155) gene sequence with part of a *Dinophyceae* (31 keys covering positions 213–474) gene sequence. Automatic annotation of eukaryotic 18S rRNA environmental gene sequences using the final generated keys can be freely tested on the following web site: <http://KeyDNATools.com>.

Phylogenetic analyses

Sequences were aligned using the slow and iterative refinement method FFT-NS-i with Mafft 5.8 software (Katoh *et al.*, 2007). Using the BioEdit sequence editor (Hall, 1999) for visualization of the alignment, we deduced the secondary structures by hand using previously published studies (see for example Lange *et al.*, 1996). For complete sequence alignments only, poorly aligned and highly variable regions of the alignments were automatically removed using Gblocks (Castresana, 2000) with the following parameters: allowing gaps in half positions and the minimum length of a block = 5. Different nested models of DNA substitution and associated parameters were tested using Modeltest v.3.06 (Posada and Crandall, 1998). Settings given by Modeltest were used to perform the NJ and the ML analyses. The NJ, ML and MP analyses were performed using PAUP 4.0b10 (Swofford, 2002). A heuristic search procedure using the tree bisection/reconnection branch swapping algorithm (setting as in MP) was performed to find the optimal ML tree topology (with 70 000 re-arrangements). Bootstrap values for NJ and MP were estimated from 1000 replicates. For MP, the number of re-arrangements was limited to 5000 for each bootstrap replicate. Starting trees were obtained by randomized stepwise addition (number of replicates = 20). Additionally, we used Bayesian reconstruction for the analysis of complete sequences with MrBayes, v.3.0b4 (Huelsenbeck and Ronquist, 2001). The GTR model of substitution was used, taking

Table 3. Examples of three different automatic annotations generated by the KeyDNAtools.

| ID | Length | SuperGroup | Division | Class | Order | Family | Genus | Species | Clade |
|---------------------|--------|----------------------------------|-----------------------------|---|--|---------------------------------|-------------------------------|---------|---|
| Case 1: EF172938 | 1288 | Chromalveolata → 175 (1–1289) | Alveolata → 175 (1–1289) | Dinophyceae → 175 (1–1289) | Syndiniales → 175 (1–1289) | Dino-Group II → 174 (1–1289) | | | Dino-Group II-Clade 7 → 47 (1–1196) |
| Case 2: EU087283 | 886 | Chromalveolata → 53 (38–716) | Alveolata → 53 (38–716) | Dinophyceae → 53 (38–716) | Dinophyceae → 53 (38–716) | Suessiales → 11 (38–656) | Symbiodinium → 7 (142–656) | | Symbiodinium clade B+B'+D → 3 (521–656) |
| Case 3: AJ965100 | 499 | Chromalveolata → 56 (44–474) | Alveolata → 56 (44–474) | Dinophyceae → 31 (213–474)/ Ciliophora → 25 (44–155) | Syndiniales → 31 (213–474)/ Oligotricha → 25 (44–155) | Dino-Group II → 31 (213–474) | | | Strombidium group → 2 (75–99) |

into account a gamma-shaped distribution of the rates of substitution among sites. The chains were run for 1000000 generations. Trees were sampled every 100 generations. The first 5000 sampled trees, corresponding to the initial phase before the chains became stationary, were discarded (burn in).

Acknowledgements

We thank R. Cowen, B.J. Binder and E. Goetze for cruise space as well as the captain and crews of the R/V Oceanus, R/V Walton Smith; R/V Atalante, R/V Melville, RRS Discovery; M.L. Cuvelier, A. Engman, C. Guigand, J.A. Hilton and F. Not for cruise assistance; J. Heidelberg, A. Ortiz, M. Perennou, C. Cormier-Caillault, J. Winkler and A. Gobet for help with clone libraries and sequencing; and P. Zimmermann for help with sequence annotation. This work was supported by the GIS génomique marine and the ANR AQUAPARADOX projects (to L.G.), the Gordon and Betty Moore Foundation and National Science Foundation grant OCE-0836721 (to A.Z.W.) and by the NERC (to A.R.K., D.J.S.).

References

- Cachon, J. (1964) Contribution à l'étude des péridinies parasites. Cytologie, cycles évolutifs. *Ann Sc Nat Zool Paris VI*: 1–158.
- Cachon, J., and Cachon-Enjumet, M. (1964) Cycle évolutif et cytologie de *Neresheimeria catenata* Neresheimer, péridinien parasite d'appendiculaires. *Ann Sc Nat Zool Paris 12 5ème série*: 779–800.
- Castresana, J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17**: 540–552.
- Cavalier-Smith, T. (1993) Kingdom *Protozoa* and its 18 phyla. *Microbiol Rev* **57**: 953–994.
- Chatton, E., and Biecheler, B. (1936) Documents nouveaux relatifs aux *Coccidiniides* (Dinoflagellés parasites). La sexualité du *Coccidinium meslini* n.sp. *C R Acad Sc Paris 203*: 573–576.
- Coats, D.W. (1988) *Duboscquella cachoni* n. sp., a parasitic dinoflagellate lethal to its tintinnine host *Eutintinnus pectinis*. *J Protozool* **35**: 607–617.
- Coats, D.W., and Park, M.G. (2002) Parasitism of photosynthetic dinoflagellates by three strains of *Amoebophrya* (*Dinophyta*): parasite survival, infectivity, generation time, and host specificity. *J Phycol* **38**: 520–528.
- Coats, D.W., Adam, E.J., Gallegos, C.L., and Hedrick, S. (1996) Parasitism of photosynthetic dinoflagellates in a shallow subestuary of Chesapeake Bay, USA. *Aquat Microb Ecol* **11**: 1–9.
- Countway, P.D., Gast, R.J., Dennett, M.R., Savai, P., Rose, J.M., and Caron, D.A. (2007) Distinct protistan assemblages characterize the euphotic zone and deep sea (2500 m) of the Western North Atlantic (Sargasso Sea and Gulf Stream). *Environ Microbiol* **9**: 1219–1232.
- Cuvelier, M.L., Ortiz, A., Kim, E., Moehlig, H., Richardson, D.E., Heidelberg, J.F., et al. (2008) Widespread distribution of a unique marine protistan lineage. *Environ Microbiol* **10**: 1621–1634.

- Di Giuseppe, G., and Dini, F. (2004) Identita dei dinoflagellati proliferanti in strutture sperimentali *in situ* nel Lago di Tovel (Nord Italia). *Studi Trentini Sci. Nat. Acta Biol* **81**: 427–438.
- Díez, B., Pedrós-Alió, C., and Massana, R. (2001a) Study of genetic diversity of eukaryotic picoplankton in different oceanic regions by small-subunit rRNA gene cloning and sequencing. *Appl Environ Microbiol* **67**: 2932–2941.
- Díez, B., Pedrós-Alió, C., Marsh, T.L., and Massana, R. (2001b) Application of denaturing gradient gel electrophoresis (DGGE) to study the diversity of marine picoeukaryotic assemblages and comparison of DGGE with other molecular techniques. *Appl Environ Microbiol* **67**: 2942–2951.
- Dolven, J.K., Lindqvist, C., Albert, V.A., Bjørklund, K.R., Yuasa, T., Takahashi, O., and Mayama, S. (2007) Molecular diversity of alveolates associated with neritic North Atlantic radiolarians. *Protist* **158**: 65–76.
- Doyle, J.J., and Doyle, J.L. (1987) A rapid DNA isolation for small quantities of fresh leaf tissue. *Phytochem Bull* **19**: 11–15.
- Garczarek, L., Dufresne, A., Rousvoal, S., West, N.J., Mazard, S., Marie, D., et al. (2007) High vertical and low horizontal diversity of *Prochlorococcus* ecotypes in the Mediterranean Sea in summer. *FEMS Microbiol Ecol* **60**: 189–206.
- Gestal, C., Novoa, B., Posada, D., Figueras, A., and Azevedo, C. (2006) *Perkinsoide chabelardi* n. gen., a protozoan parasite with an intermediate evolutionary position: possible cause of the decrease of sardine fisheries? *Environ Microbiol* **8**: 1105–1114.
- Groissillier, A., Massana, R., Valentin, K., Vaulot, D., and Guillou, L. (2006) Genetic diversity and habitats of two enigmatic marine alveolate lineages. *Aquat Microb Ecol* **42**: 277–291.
- Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symp Ser* **41**: 95–98.
- Harada, A., Ohtsuka, S., and Horiguchi, T. (2007) Species of the parasitic genus *Duboscquella* are members of the enigmatic marine alveolate Group I. *Protist* **158**: 337–347.
- Hollande, A. (1974) Etude comparée de la mitose syndinienne et de celle des përidiniens libres et des hypermastigines. Infrastructure et cycle évolutif des Syndinides parasites de radiolaires. *Protistologica* **X**: 413–451.
- Hudson, D.A., and Adlard, R.D. (1996) Nucleotide sequence determination of the partial SSU rDNA gene and ITS1 region of *Hematodinium cf perezii* and *Hematodinium*-like dinoflagellates. *Dis Aquat Organ* **24**: 55–60.
- Huelsenbeck, J.P., and Ronquist, F. (2001) MrBayes: bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754–755.
- Katoh, K., Kuma, K.-i., Toh, H., and Miyata, T. (2007) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* **33**: 511–518.
- Kim, S. (2006) Patterns in host range for two strains of *Amoebophrya* (*Dinophyta*) infecting thecate dinoflagellates: *Amoebophrya* spp. ex *Alexandrium affine* and ex *Gonyaulax polygramma*. *J Phycol* **42**: 1170–1173.
- Kim, S., Park, M.G., Kim, K.-Y., Kim, C.-H., Yih, W., Park, J.S., and Coats, D.W. (2008) Genetic diversity of parasitic dinoflagellates in the genus *Amoebophrya* and its relationship to parasite biology and biogeography. *J Eukaryot Microbiol* **55**: 1–8.
- Lange, M., Guillou, L., Vaulot, D., Simon, N., Amann, R.I., Ludwig, W., and Medlin, L.K. (1996) Identification of the class *Prymnesiophyceae* and the genus *Phaeocystis* with ribosomal RNA-targeted nucleic acid probes detected by flow cytometry. *J Phycol* **32**: 858–868.
- Leander, B.S., and Keeling, P.J. (2003) Morphostasis in alveolate evolution. *Trends Ecol Evol* **18**: 395–402.
- Lefèvre, E., Roussel, B., Amblard, C., and Sime-Ngando, T. (2008) The molecular diversity of freshwater picoeukaryotes reveals high occurrence of putative parasitoids in the plankton. *PLoS ONE* **3**: e2324.
- Lessard, E.J., and Swift, E. (1986) Dinoflagellates from the North Atlantic classified as phototrophic or heterotrophic by epifluorescence microscopy. *J Plankton Res* **8**: 1209–1215.
- Loeblich, A.R. III (1976) Dinoflagellate evolution: speculation and evidence! *J Protozool* **23**: 13–28.
- López-García, P., Rodríguez-Valera, F., Pedrós-Alió, C., and Moreira, D. (2001) Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature* **409**: 603–607.
- Lovejoy, C., Massana, R., and Pedrós-Alió, C. (2006) Diversity and distribution of marine microbial eukaryotes in the Arctic Ocean and adjacent seas. *Appl Environ Microbiol* **72**: 3085–3095.
- Massana, R., Balagué, V., Guillou, L., and Pedrós-Alió, C. (2004) Picoeukaryotic diversity in an oligotrophic coastal site studied by molecular and culturing approaches. *FEMS Microbiol Ecol* **50**: 231–243.
- Massana, R., Guillou, L., Terrado, R., Forn, I., and Pedrós-Alió, C. (2006) Growth of uncultured heterotrophic flagellates in unamended seawater incubations. *Aquat Microb Ecol* **45**: 171–180.
- Massana, R., Karniol, B., Pommier, T., Bodaker, I., and Béjà, O. (2008) Metagenomic retrieval of a ribosomal DNA repeat array from an uncultured marine alveolate. *Environ Microbiol* **10**: 1335–1343.
- Matsuzaki, M., Kuroiwa, H., Kuroiwa, T., Kita, K., and Nozaki, H. (2008) A cryptic algal group unveiled: a plastid biosynthesis pathway in the oyster parasite *Perkinsus marinus*. *Mol Biol Evol* **25**: 1167–1179.
- Medlin, L.K., Metfies, K., Mehl, H., Wiltshire, K., and Valentin, K. (2006) Picoeukaryotic plankton diversity at the Helgoland time series site as assessed by three molecular methods. *Microb Ecol* **52**: 53–71.
- Moon-van der Staay, S.Y., Watcher, R.D., and Vaulot, D. (2001) Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature* **409**: 607–610.
- Murray, S., Jørgensen, M.F., Ho, S.Y.W., Patterson, D.J., and Jermini, L.S. (2005) Improving the analysis of dinoflagellate phylogeny based on rDNA. *Protist* **156**: 269–286.
- Not, F., Gausling, R., Penn, K., Heidelberg, J., and Worden, A.Z. (2007) Vertical distribution of picoeukaryotic diversity in the Sargasso Sea. *Environ Microbiol* **9**: 1233–1252.
- Not, F., Latasa, M., Scharek, R., Viprey, M., Karleskind, P., Balagué, V., et al. (2008) Protistan assemblages across the Indian Ocean: a specific emphasis on picoeukaryotes. *Deep Sea Res Part I* (in press).

- Park, M.G., Yih, W., and Coats, D.W. (2004) Parasites and phytoplankton, with special emphasis on dinoflagellate infections. *J Eukaryot Microbiol* **51**: 145–155.
- Patterson, D.J. (1999) The diversity of eukaryotes. *Am Nat* **154**: 96–124.
- Posada, D., and Crandall, K.A. (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**: 817–818.
- Ris, H., and Kubai, D.F. (1974) An unusual mitotic mechanism in the parasitic protozoan *Syndinium* sp. *J Cell Biol* **60**: 702–720.
- Romari, K., and Vaulot, D. (2004) Composition and temporal variability of picoeukaryote communities at a coastal site of the English channel from 18S rDNA sequences. *Limnol Oceanogr* **49**: 784–798.
- Saldarriaga, J.F., McEwan, M.L., Fast, N.M., Taylor, F.J.R., and Keeling, P.J. (2003) Multiple protein phylogenies show that *Oxyrrhis marina* and *Perkinsus marinus* are early branches of the dinoflagellate lineage. *Int J Syst Evol Microbiol* **53**: 355–365.
- Silberman, J.D., Collins, A.G., Gershwin, L.-A., Johnson, P.J., and Roger, A.J. (2004) *Ellobiopsidae* of the genus *Thalassomyces* are alveolates. *J Eukaryot Microbiol* **51**: 246–252.
- Skovgaard, A., Massana, R., Balagué, V., and Saiz, E. (2005) Phylogenetic position of the copepod-infesting parasite *Syndinium turbo* (Dinoflagellata, Syndinea). *Protist* **156**: 413–423.
- Slamovits, C.H., and Keeling, P.J. (2008) Plastid-derived genes in the non-photosynthetic alveolate *Oxyrrhis marina*. *Mol Biol Evol* **25**: 1297–1306.
- Soyer, M.-O. (1974) Etude ultrastructurale de *Syndinium* sp. Chatton parasite coelomique de copépodes pélagiques. *Vie Milieu* **XXIV**: 191–212.
- Stoeck, T., Kasper, J., Bunge, J., Leslin, C., Ilyin, V., and Epstein, S. (2007) Protistan diversity in the Arctic: a case of paleoclimate shaping modern biodiversity? *PLoS ONE* **2**: e728.
- Swofford, D.L. (2002) *PAUP*. Phylogenetic Analysis Using Parsimony (*And Others Methods)*. Sunderland, MA, USA: Sinauer Associates.
- Takishita, K., Yubuki, N., Kakizoe, N., Inagaki, Y., and Maruyama, T. (2007) Diversity of microbial eukaryotes in sediment at a deep-sea methane cold seep: surveys of ribosomal DNA libraries from a raw sediment samples and two enrichment cultures. *Extremophiles* **11**: 563–576.
- Viprey, M., Guillou, L., Ferréol, M., and Vaulot, D. (2008) Wide genetic diversity of picoplanktonic green algae (*Chloroplastida*) uncovered in the Mediterranean Sea by a phylum-biased PCR approach. *Environ Microbiol* **10**: 1804–1822.
- Williams, A.G., and Coleman, G.S. (1992) *The Rumen Protozoa*. New York, USA: Springer-Verlag.
- Worden, A.Z. (2006) Picoeukaryote diversity in coastal waters of the Pacific Ocean. *Aquat Microb Ecol* **43**: 165–175.
- Xiao-Ping, G., and Jing-Yan, L. (1986) Nuclear division in the marine dinoflagellate *Oxyrrhis marina*. *J Cell Sci* **85**: 161–175.
- Zwirgmaier, K., Heywood, J.L., Chamberlain, K., Woodward, E.M.S., Zubkov, M.V., and Scanlan, D.J. (2007) Basin-scale distribution patterns of picocyanobacterial lineages in the Atlantic Ocean. *Environ Microbiol* **9**: 1278–1290.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Mean sequence identities within the different clades belonging to *Syndiniales* groups I and II.

Table S1. Details of the sequences used for the phylogenetic analyses described in Fig. 1.

Table S2. Primer sequences used or cited in this study.

Table S3. Taxonomy of *Syndiniales* (= *Syndinida*) and their hosts modified from Loeblich (1976).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.