

Phylogenetic analysis of novel alveolates and distribution in different marine systems

**Agnès Groisillier¹, Laure Guillou¹, Ramon Massana², Klaus Valentin³, Daniel
Vaulot¹**

¹Station Biologique, UMR 7127, CNRS and University Pierre & Marie Curie, BP74,
29682 Roscoff Cedex, France

²Departament de Biologia Marina i Oceanografia, Institut de Ciències del Mar,
CMIMA, CSIC. Passeig Maritim de la Barceloneta 37-49, 08004 Barcelona,
Catalonia, Spain.

ABSTRACT

Alveolate phylum is one of the main abundant lineages in genetic libraries. Novel alveolate sequences have been found in oceanic and sediments area and more recently in coastal environments. Novel alveolates are shared in two distinct groups I and II. Group I sequences dominate in sediment environment (75 to 95%) whereas group II sequences are more present in oceanic systems (50 to 100%). In three coastal sites of four, group II represents 75 to 100% and in the last site, group I sequences dominate at 72%. Maximum-Likelihood phylogenetic trees of full and partial 18s rDNA sequences of Alveolates have showed that group I is composed of 5 clades and the group II, more complex, included at least 18 different clades. Certain clades include specifically sediment or coastal sequences but never only oceanic clones. This analyse help us to understand the phylogenetic relationships and the distribution in oceans of novel alveolates what should contribute to the study of their physiology and their ecological role.

INTRODUCTION

These last years, different molecular studies have permitted to reveal the extraordinary diversity of picoplanktonic world. This world composed of prokaryotic and eukaryotic cells with a diameter of less than 3 μm is the dominant contributor to the biomass and primary productivity in all aquatic environments (Li 1994, Campbell et al. 1994). The ubiquitous presence of completely novel lineages, with no representatives previously obtained in cultures, was revealed among the three domains, bacteria (Giovanonni et al., 1990), archaea (DeLong, 1992, Fuhrman et al., 1992), and more recently for small eukaryotes in oceanic environments (Diez et al., 2001, López-García et al., 2001, Moon-van der Staay et al, 2001) and from marine sediments (Dawson & Pace, 2002, Edgcomb et al., 2002, López-García et al., 2003, Stoeck et al., 2003). Alveolates and stramenopiles formed the most abundant lineages in all clones libraries. Novel alveolates are mainly composed of two groups I and II (Moon-van der Staay et al, 2001, Moreira & López-García, 2002). All recent studies do not permit to elucidate the physiology and the ecological role of these novel alveolate groups.

In this study, we have analysed all available genetic libraries from oceanic (Diez et al., 2001, López-García et al., 2001, Moon-van der Staay et al, 2001) and sediments (Dawson & Pace, 2002, Edgcomb et al., 2002, López-García et al., 2003, Stoeck et al., 2003) systems, but also from recent coastal libraries (Romari & Vaultot, submitted, Valentin et al., in prep, Massana et al., in prep). These data have helped us to define different clades in novel alveolate group I and II and to analyse the

distribution of these organisms in the three different systems: oceanic, sediment and coastal.

MATERIALS AND METHODS

Genbank accession numbers

(AF069516) *Amoebophrya* sp., () *Amoebophrya* sp. ex *Dinophysis*, (AF472554) *Amoebophrya* sp. ex *Gymnodinium instriatum*, (AF472553) *Amoebophrya* sp. ex *Karlodinium micrum*, (AF472555) *Amoebophrya* sp. ex *Scrippsiella* sp., (MUCC99) *Alexandrium tamarense*, (L13719) *Amphidinium belauense*, (Z15105) *Babesia equi*, (AJOO9663) *Caenomorpha* sp., (AF300283) *Chilidonella uncinata*, (AY142075) *Colpodella* sp., (AY078092) *Colpodella pontica*, (AF151376) *Cryptosporidium serpentis*, (AF111186) *Cyclospora colobi*, (AF239261) *Dinophysis norvegica*, (AF164134) *Engelmanniella mobilis*, (X65150) *Furgasonia blochmanni*, (AF172714) *Gymnodinium breve*, (AF286023) *Hematodinium* sp., (Z29516) *Metopus contortus*, (X53486) *Oxytricha granulifera*, (X03772) *Paramecium tetraurelia*, (AF133909) *Parvilucifera infectans*, (L07375) *Perkinsus* sp., (X75762) *Perkinsus marinus*, (Y16234) *Prorocentrum arenarium*, (AF022156) *Pyrocystis noctiluca*, (U97112) *Strombidium purpureum*, (M88521) *Symbodinium microadriaticum*, (X75453) *Toxoplasma gondii*. () DH147-EKD6, () DH147-EKD16, () DH148-EKD, () DH147-EKD3, () DH148-EKD27, () DH147-EKD18, () DH148-EKD22, () DH145-EKD10, () DH145-EKD20, () DH144-EKD3, () DH147-EKD21, () OI011010, () OI011023, () OI011055, () OI011261, () OI011115, () OI011009, () OI011012, () OI011011, () OI011029, () OI011033, () OI011038, () OI011511, () OI011255, () OI011027, () OI011005, () OLI02001, () A1_E004, () A1_E005, () A1_E007, () A1_E008, () A1_E010, () A1_E011, () A1_E045

Coastal genetic libraries

Several genetic libraries were constructed, following conventional methods (Diez et al, 2001, Moon-van der Staay et al. 2001) from four coastal sites all along the year, the Roscoff Coast in the English Channel at the North Sea (RA000412, RA000609, RA000907, RA001219, RA010412, RA010516, RA010613 and RD010517, A= coastal site Astan, D= estuarine site Dourduff), the Blanes Bay in the Catalan Coast at the Mediterranean Sea (BL000921, BL001221, BL010320 and BL010625), the Helgoland Coast at the North Sea (He000327, He000427, He000803, He001005, He001206 and He010218), and the Orkney Coast in the Atlantic Ocean (Or000415). Each system was submitted to detailed analyse: Massana et al. in prep for Blanes, Valentin et al. in prep for Helgoland and Orkney and Romari & Vaultot, submitted for Roscoff. Sequences from Roscoff have been deposited in the Genbank database under the accession number AY295352 to AY295760.

Sequencing and Phylogenetic analyses

We have completely sequenced 14 sequences from Pacific and 13 sequences from Roscoff. Sequencing reactions were performed with a BD kit () and an ABI PRISM model XX automated sequencer. Two sequences from Helgoland and one from Blanes have been sequenced by Qiagen.

New complete and selected sequences for phylogenetic analyses were exported from ARB software (www2.mikro.biologie.tu-muenchen.de/arb/) database and aligned, first automatically using CLUSTALW (Thompson et al., 1994) then by hand using BIOEDIT. Divergent regions were eliminated using Gblocks software (Castresana, 2000). Model of DNA substitution and associated parameters were estimated using Modeltest 3.0 (Posada & Crandall, 1998). Maximum likelihood distance, neighbour joining distance and parsimony analyses were performed using

PAUP 4.0b10 software (Swofford, 2002). Bootstrap values were assessed from 1000 replicates. For analysis of group I and II partial sequences, the *Perkinsus sp.* (L07375) sequence has been used as outgroup.

All full sequences obtained in this study have been deposited in Genbank (Accession numbers XX to XXX).

RESULTS AND DISCUSSION

Different studies have showed that Alveolate phylum represented the largest group of picoeukaryote clone types from all clone libraries (Diez et al. 2001, Moon-van der Staay et al. 2001, Lopez-Garcia et al. 2001, Edgcomb et al. 2002, Dawson and Pace, 2002, Stoeck et al. 2003, Lopez-Garcia et al. 2003). In this study we have used 288 full and partial SSU rDNA gene sequences of novel alveolates. 171 sequences arise from four different coastal sites: 56 from Roscoff (Romari et Vaultot, submitted), 59 from Helgoland (Valentin et al. in prep), 7 from Blanes (Massana et al. in prep) and 3 from Orkney (Valentin et al. in prep). Another part of sequences correspond to oceanic sites (32 from Pacific (Moon-van der Staay et al. 2001), 10 from Antarctica (Lopez-Garcia et al. 2001), 8 from Mediterranean and 3 from North Atlantic (Diez et al. 2001)). We have also used alveolate sequences retrieved in hydrothermal vents (49) (Edgcomb et al. 2002), in Oxygen-depleted marine environment (20) (Stoeck et al. 2003) and in Mid-Atlantic Ridge (4) (Lopez-Garcia et al. 2003). A preliminary phylogenetic analysis of all these sequences was done by distance methods (neighbour-joining, NJ) by using the ARB software (data not showed) allowing the selection of some clones for complete sequencing.

A maximum-likelihood (ML) phylogenetic tree was constructed for the Alveolate phylum (figure 1). Two sequences of primitives eukaryotes were used as outgroup, *Leichmania brasiliensis* (X53913) and *Euglena gracilis* (M12677). Neighbor-joining (NJ) and maximum parsimony (MP) bootstrap values were carried back. The three main groups were represented: Dinoflagellates, heterotrophic group of Ciliates and parasitic group of Apicomplexans. The *Perkinsus* species usually at

the basis of Dinoflagellate (Saldarriaga et al, 2003) was found at the basis of the alveolate phylum. However, the bootstrap value inferior to 50 does not allow confirming this result. On the other hand, this tree suggests also that Colpodellid is the free-living sister group to Apicomplexans (Kuvardina et al, 2002). Most of the novel Alveolate sequences forms two distinct clusters named group I and II (Moreira and Lopez-garcia, 2002). The group I seem to be more closed to Dinoflagellate than group II. The group I is composed of two sub-clusters whereas the group II is more complex composed of a lot of clades who will be detailed further in the analysis. *Amoebophrya* sp. has been long considered a dinoflagellate (Gunderson, 1999, Janson et al, 2000), more exactly at the base of this group although its morphological, structural and life cycle characteristics are completely uncommon among dinoflagellate (Moreira and Lopez-garcia, 2002). It from now on is established that *Amoebophrya* sp. belonging to novel alveolate group II (Moon 2000, Moreira and Lopez-garcia, 2002) and that independently of infected dinoflagellate cell (figure 1).

ML phylogenetic trees were constructed (figures 2 and 3) based upon the analyses of 124 partial sequences for group I and 170 for group II. Existence of a clade was detected based upon 3 criteria: 1) a clade comportes at least 3 sequences, 2) this clade is validated by bootstrap value superior to 70 in NJ and/or in MP 3) there is a nucleotide signature common to all the sequences of a same clade (18 to 20 nucleotides). About a third 18S rDNA gene sequence is available in partial sequences, so, this third criterion is not necessarily found. The nucleotide signature could be in the two other parts of the gene.

The novel alveolate group I is composed by at least of 5 different clades (figure 2). The percentage of clones in each environment: sediments (S), oceanic (O) and coasting are indicated on the right of the tree. The group II, more complex, include at least 18 clades (figure 3). The sequences of Mediterranean (Diez et al., 2001), North-Atlantic (Diez et al., 2001) and Hydrothermal sediment at the Mid-Atlantic ridge (López-García et al. 2003) were not including in the trees. Their 18S rDNA gene sequenced part were not the same that all the other partial sequences. Their position in different clades has been done by addition in the full sequences tree (data not showed). The table 1 show the number of novel alveolate clones among group I and II clades from different marine systems. For coastal genetic libraries (except Orkney), constructed from environmental samples collected at different period of the year, we have separated each sample date. Roscoff, Helgoland and Orkney include a majority of group II sequences (87,5%, 75 % and 100% respectively) whereas group I sequences dominate in Blanes (72%). Concerning oceanic systems, there is a majority of group II sequences (69% for Pacific, 62,5% for Mediterranean and 100% for North Atlantic), except for Antarctic where group I is equivalent in clone number to group II. In marine or hydrothermal sediment, the group I predominate (77,5% in hydrothermal vents, 95% in oxygen-depleted marine environment and 75% in hydrothermal sediment at the Mid-Atlantic ridge). The repartition of the 3 marine systems is also interesting in the different clades of the novel alveolate groups. In group I half of sequences id composed by the clade 1. Clade 4 includes clones of the 3 different environments. The majority of coastal clones (86%) belongs to clade 1 (figure 2). We could also note that Helgoland group I sequences are specific of this clade 1 and these clones have been found exclusively in winter (table 1). The clade 5 is composed of sequences that come from coastal and oceanic environments. Clade

2 includes sequences from the 3 different genetic libraries from sediment environments and one sequence from the Antarctic. The last clade, number 3, is composed exclusively of hydrothermal vents clones. More specifically, the first cm of the core A obtained from the Everest Mound area in the southern Guaymas vent. The difference between cores A and C was the temperature gradient for the top 3 cm, ranged from 3°C to 65°C for core A and from 3°C to 45°C for core C (Edgcomb et al. 2002). For the clades 1, 2, 3 and 4, we have found nucleotide signatures that help us to design probes to be used during *in situ* hybridization (FISH) (data unpublished). There is no nucleotide signature in the partial sequences of clade 5 but the high NJ bootstrap value (93) enable to foresee the presence of a signature on the remainder of the 18S rDNA gene.

The alveolate group II including 170 partial 18S rDNA sequences is a very complex group (figure 3). On the base of NJ and MP bootstrap values, we have found at least 18 different clades. Nucleotide signatures specific of certain clades are also used in FISH analysis (data unpublished). Few sediment sequences are represented in the group II and these sequences are grouped in only 3 different clades: 6, 17 and 18. The clade 17 contains only clones from sediment whereas clade 6 includes also 2 environmental clones from Blanes and clade 18 includes also one clone from Antarctic and one from hydrothermal sediment at the Mid-Atlantic ridge. As in the group I, clones from cores A and C of hydrothermal vents are spleat into different clades. There are a lot of coastal sequences in the group II, mainly from Roscoff and Helgoland. We retrieve sequences from group II throughout the year in the coastal site of Roscoff. In Helgoland sequences from group II are also present all the year except in February. Moreover predominance of Group II is particularly

evident from August to December; where group I is represented by very few sequences (table 1). Some clades include only coastal sequences, as clades 3, 9, 10 and 12, or only coastal sequences plus *Amoebophrya* species, as clades 4 and 5. Half of clades contain sequences from Oceanic ecosystem mainly from the equatorial Pacific. But the low number of sequences obtained for Antarctic, Mediterranean Sea and North-Atlantic clones libraries does not mean that they miss of this group II. *Amoebophrya* sequences, the unique organisms described among this new alveolate lineage group II (Maranda, 2001, Gunderson et al., 2002) are localised in clades 1, 2, 4 and 5. These 4 clades form a sub-clade with clade 3.

The alveolate group I and group II include also some sequences belonging to any clade (9 in group I and 22 in group II). These clones are present in all marine systems. This result show that novel alveolates could be still more complex. The complete sequencing of these clones and the discovery of others novel alveolate sequences could be stabilised their position in the phylogenetic trees.

This study highlights the great need for further molecular methods and novel alveolate cultures to understand the physiological role and the evolution of these eukaryotic organisms. As soon as more eukaryotic sequences from a variety of environment become available, it will become possible to design oligonucleotide probes that will be enable us to determine the morphology and ecological role of these unknown protists. Fluorescent in situ hybridization (FISH) method have permit to visualized and enumerated in field samples and enrichment two novel stramenopiles lineages (Massana et al., 2002). In the same way, Rappé et al (2002) have achieved the cultivation of the ubiquitous SAR11 marine bacterioplankton.

In summary, alveolates and stramenopiles are the main lineages retrieved in all clones libraries. In novel alveolates, group I and group II are very distinct groups. They are very ubiquitous, majoritary uncultivated, organisms. But, we have found in each group many different clades. Certain ones contain all marine systems, and others are more specific for example of sediments sequences (clade 17 group II, clade 3 group I). The group I clade 1 includes all the novel alveolate group I Helgoland sequences. These results do not permit to conclude on the phototrophic or heterotrophic function of these organisms. The only sign is the presence of Dinoflagellate parasites *Amoebophrya* species in group II.

ACKNOWLEDGMENTS

This work was supported by EU project PICODIV (EVK3-CT1999-00021).

REFERENCES

- Campbell, L., Nolla, H. A. & Vaulot, D. (1994) *Limnol. Oceanogr.* **39**, 954-961.
- Castresana, J. (2000) *Mol. Biol. Evol.* **17**, 540-552.
- Dawson, S. C. & Pace, N. R. (2002) *Proc. Natl. Acad. Sci.* **99**, 8324-8329.
- DeLong, E. F. (1992) *Proc. Natl. Acad. Sci.* **89**, 5685-5689.
- Diez, B., Pedrós-Aliós, C. & Massana, R. (2001) *Appl. Environ. Microbiol.* **67**, 2932-2941.
- Edgcomb, V. P., Kysela, D. T., Teske, A., De Vera Gomez, A. & Sogin, M. L. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 7658-7662.
- Fuhrman, A. J., McCallum, K. & Davis, A. A. (1992) *Nature* **356**, 148-149.

- Giovannoni, S. J., Britschgi, T. B., Moyer, C. L. & Field, K. G. (1990) *Nature* **345**, 60-63.
- Gunderson, J. H., Goss, S. H. & Coats D. W. (1999) *J. Eukaryot. Microbiol.* **46**, 194-197.
- Kuvardina, O. N., Leander, B. S., Aleshin, V. V., Myl'nikov, A. P., Keeling, P. J. & Simdyanov, T. G. (2002) *J. Eukaryot. Microbiol.* **49**, 498-504.
- Li, W. K. W. (1994) *Limnol. Oceanogr.* **39**, 169-175.
- López-García, P., Rodríguez-Valera, F., Pedrós-Aliós, C. & Moreira, D. (2001) *Nature* **409**, 603-607.
- López-García, P., Philippe, H., Gail, F. & Moreira, D. (2003) *Proc. Natl. Acad. Sci.* **100**, 697-702.
- Maranda, L. (2001) *J. Phycol.* **37**, 245-248.
- Massana, R., Guillou L., Diez, B. & Pedrós-Aliós, C. (2002) *Appl. Environ. Microbiol.* **68**, 4554-4558.
- Moon-van der Staay, S. Y., De Wachter, R. & Vaulot D. (2001) *Nature* **409**, 607-610.
- Moreira, D. & López-García, P. (2002) *TRENDS in Microbiol.* **10**, 31-38.
- Norén, F., Moestrup, Ø. & Rehnstam-Holm, A.-S. (1999) *Europ. J. Protistol.* **35**, 233-254.
- Posada, D. & Crandall, K. A. (1998) *Bioinformatics* **14**, 817-818.
- Rappé, M. S., Connon, S. A., Vergin, K. L. & Giovannoni, S. J. (2002) *Nature* **418**, 630-633.
- Stoeck, T. & Epstein, S. (2003) *Appl. Environ. Microbiol.* **69**, 2657-2663.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673-4680.
- Swofford, D. L. (1998) Sinauer Associates, Sunderland, Massachusetts.

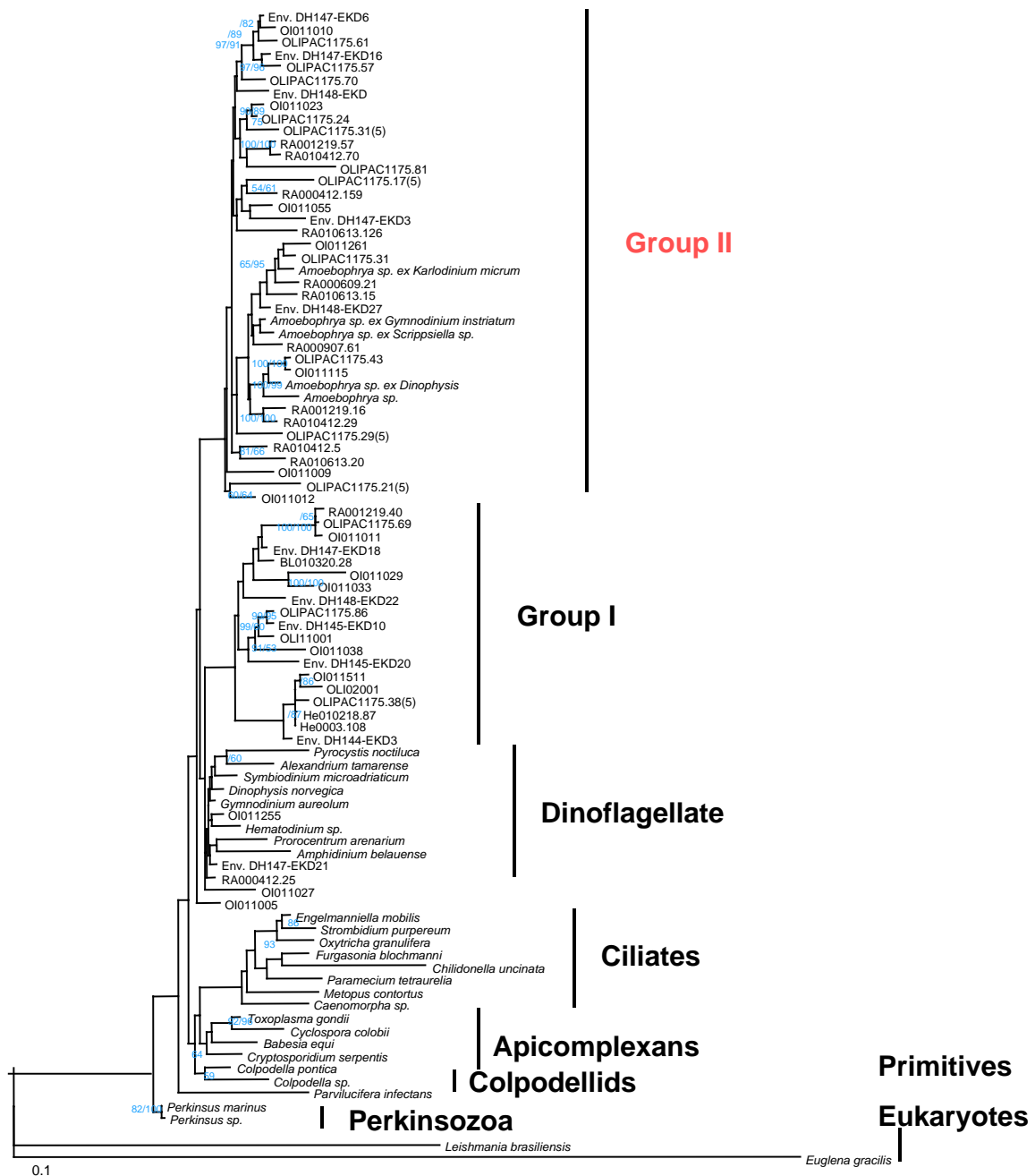


Figure 1. Maximum likelihood phylogenetic tree with full 18S rDNA sequences of novel alveolates. The scale bar indicates 0,1% sequence divergence. Neighbor-joining and maximum-parsimony bootstrap values (left to right, respectively) are shown at the internal branches (1000 replicates , values > 50% shown).



Figure 2. Maximum likelihood phylogenetic tree with partial 18S rDNA sequences of novel group I alveolates. The scale bar indicates 0,1% sequence divergence. Neighbor-joining bootstrap values are show at the internal branches (1000 replicates, values>50 show). The percentage of clones in each environment (S = sediments ; C = coasting ; O = oceanic) belonging to each phylogenetic clade is indicated on the right.

Group II

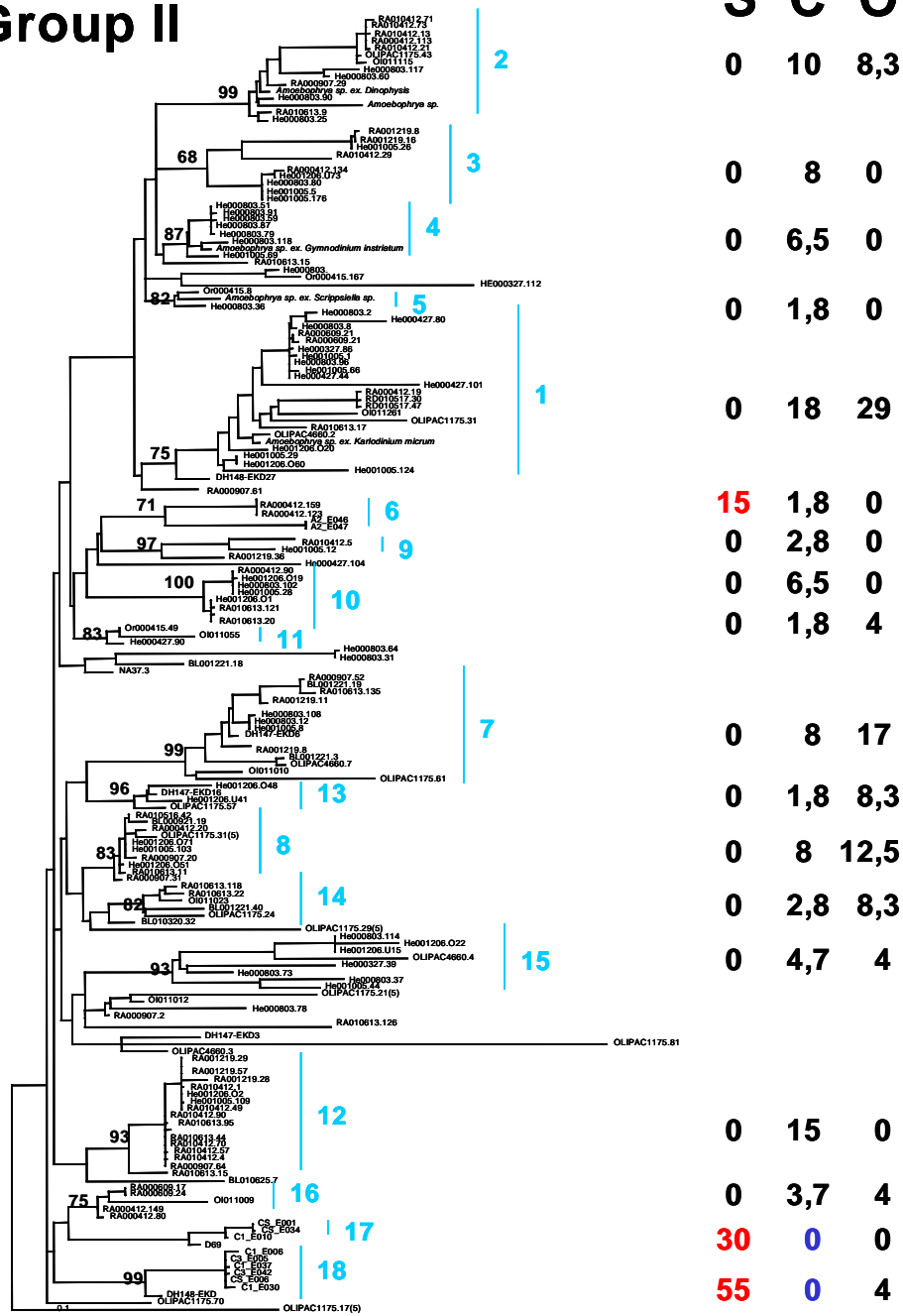


Figure 3. Maximum likelihood phylogenetic tree with partial 18S rDNA sequences of novel group II alveolates. The scale bar indicates 0,1% sequence divergence. Neighbor-joining bootstrap values are show at the internal branches (1000 replicates, values>50 show). The percentage of clones in each environment (S = sediments ; C = coasting ; O = oceanic) belonging to each phylogenetic clade is indicated on the right.

		RA000412	RA000609	RA000907	RA001219	RA010412	RA010516 + RD010517	RA010613	He000327	He000427	He000803	He001005	He001206	He010218	BL000921	BL001221	BL010320	BL010625	Orkney	Pacific	Antarctica	Mediterranean	North Atlantic	Hydrothermal vents	Oxygen-depleted marine env	Ridge
GROUP I	Clade 1				2			2	4				5	11	2	3	5	2		3	1			11	9	1
	Clade 2																				1			3	2	1
	Clade 3																							19		
	Clade 4			1	2										1	1	1			2		3		4	5	
	Clade 5							1							1					3	2					
	others																2			2	1			1	3	1
GROUP II	Clade 1	1	2			2	1	1	3	3	4	2								3	1	3				
	Clade 2	1		1		4	1			4										2						
	Clade 3	1			2	1				1	3	1														
	Clade 4									6	1															
	Clade 5										1								1							
	Clade 6	2																						2		
	Clade 7			1	2		1			2	1					2				3	1					
	Clade 8	1		2			1	1				1	2		1					1			2			
	Clade 9				1	1						1														
	Clade 10	1					2				1	1	2													
	Clade 11									1									1	1						
	Clade 12			1	3	6	3					1	1													
	Clade 13												2							1	1					
	Clade 14						2									1	1			2						
	Clade 15							1		3	1	2								1						
	Clade 16	2	2																	1						
	Clade 17																							3	1	
	Clade 18																					1	2	6		1
	others			2			2	1	1	4						1	1	1	1	7	1	2	1			

Table 1. Number of novel alveolate clones among group I and II clades from different marine systems. RAYMMDD, HeYYMMDD, BLYMMDD and Orkney corresponding to coasting sites; Pacific (Moon-van der Staay et al. 2001), Antarctic (Lopez-Garcia et al. 2001), Mediterranean and North Atlantic (Diez et al. 2001) to oceanic sites; Hydrothermal vents (Edgcomb et al. 2002), Oxygen-depleted marine env. (Stoeck et al. 2003) and Ridge (Lopez-Garcia et al. 2003) to sediments.