Intracellular Infection of Diverse Diatoms by an Evolutionary Distinct Relative of the Fungi

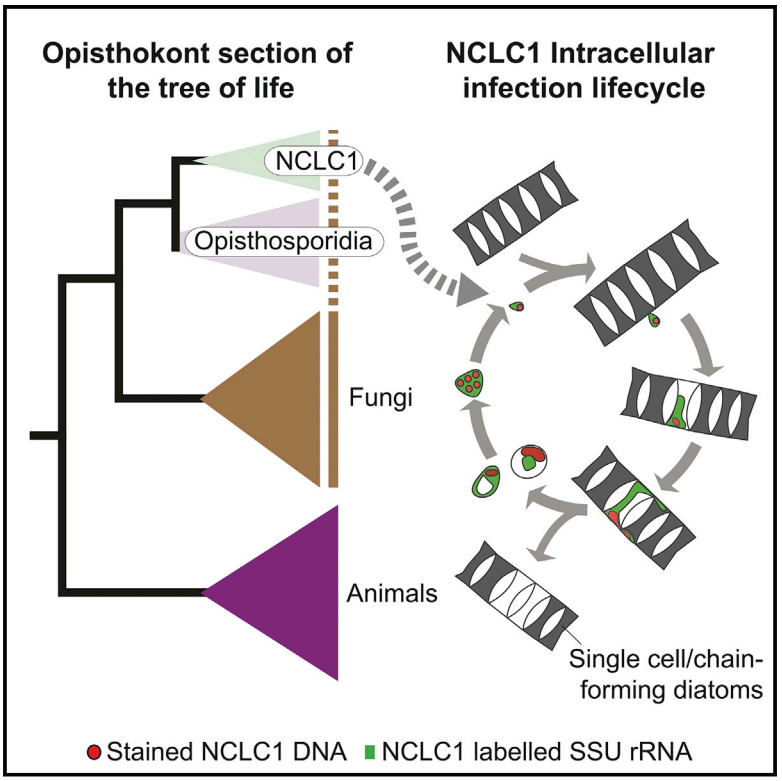
Aurélie Chambouvet, Adam Monier, Finlay Maguire, Sarah Itoïz, Javier del Campo, Philippe Elies, Bente Edvardsen, Wenche Eikreim, Thomas Richards

Report

Current Biology

Intracellular Infection of Diverse Diatoms by an Evolutionary Distinct Relative of the Fungi

Graphical Abstract



Authors

Aurélie Chambouvet, Adam Monier, Finlay Maguire, ..., Bente Edvardsen, Wenche Eikreim, Thomas A. Richards

Correspondence

aurelie.chambouvet@univ-brest.fr (A.C.), t.a.richards@exeter.ac.uk (T.A.R.)

In Brief

Most microbial life remains uncultured and unstudied. Sequencing has shown a diversity of forms branching close to the Fungi. Chambouvet et al., using microscopy of labeled marine samples, show that a newly identified relative of the fungi forms intracellular infections of diatoms, potentially determining the fate of important phytoplankton blooms.

Highlights

- Environmental DNA phylogenies identify an addition to the Opisthosporidia, close to the Fungi
- This group forms intracellular infections within key marine Diatom phytoplankton
- Intracellular infection is a clade-wide lifestyle of the Opisthosporidia
- This putative parasite is found throughout the North-East Atlantic and North Sea

Chambouvet et al., 2019, Current Biology *29*, 4093–4101
December 2, 2019 Crown Copyright © 2019 Published by Elsevier Ltd.
https://doi.org/10.1016/j.cub.2019.09.074

CellPress

Current Biology
Report

Intracellular Infection of Diverse Diatoms by an Evolutionary Distinct Relative of the Fungi

Aurélie Chambouvet,[1,7,*] Adam Monier,[2,7] Finlay Maguire,[2,3] Sarah Itoïz,[1] Javier del Campo,[4] Philippe Elies,[5] Bente Edvardsen,[6] Wenche Eikreim,[6] and Thomas A. Richards[2,8,*]
[1]CNRS, Univ Brest, IRD, Ifremer, LEMAR, F-29280 Plouzane, France
[2]Living Systems Institute, School of Biosciences, University of Exeter, Exeter, EX4 4QD, UK
[3]Faculty of Computer Science, Dalhousie University, Halifax, Nova Scotia, B3H 4R2, Canada
[4]Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, FL, 33149 USA
[5]Plateforme d'Imagerie et de Mesures en Microscopie, Université de Bretagne Occidentale, 29200, Brest, France
[6]Section for Aquatic Biology and Toxicology, Department of Biosciences, University of Oslo, PO Box 1066 Blindern, 0316 Oslo, Norway
[7]These authors contributed equally
[8]Lead Contact
*Correspondence: aurelie.chambouvet@univ-brest.fr (A.C.), t.a.richards@exeter.ac.uk (T.A.R.)
https://doi.org/10.1016/j.cub.2019.09.074

SUMMARY

The Fungi are a diverse kingdom, dominating terrestrial environments and driving important ecologies. Although fungi, and the related Opisthosporidia, interact with photosynthetic organisms on land and in freshwater as parasites, symbionts, and/or saprotrophic degraders [1, 2], such interactions in the marine environment are poorly understood [3–8]. One newly identified uncultured marine lineage has been named novel chytrid-like-clade-1 (NCLC1) [4] or basal-clone-group-I [5, 6]. We use ribosomal RNA (rRNA) encoding gene phylogenies to demonstrate that NCLC1 is a distinct branch within the Opisthosporidia (Holomycota) [7]. Opisthosporidia are a diverse and largely uncultured group that form a sister branch to the Fungi or, alternatively, the deepest branch within the Fungi, depending on how the boundary to this kingdom is inferred [9]. Using culture-free lineage-specific rRNA-targeted fluorescent *in situ* hybridization (FISH) microscopy, we demonstrate that NCLC1 cells form intracellular infection of key diatom species, establishing that intracellular colonization of a eukaryotic host is a consistent lifestyle across the Opisthosporidia [8–11]. NCLC1 infection-associated loss and/or envelopment of the diatom nuclei infers a necrotrophic-pathogenic interaction. Diatoms are one of the most diverse and ecologically important phytoplankton groups, acting as dominant primary producers and driving carbon fixation and storage in many aquatic environments [12–14]. Our results provide insight into the diversity of microbial eukaryotes that interact with diatoms. We suggest that such interactions can play a key role in diatom associated ecosystem functions, such as the marine carbon pump through necrotrophic-parasitism, facilitating the export of diatoms to the sediment [15, 16].

RESULTS AND DISCUSSION

An Addition to the Opisthokont Phylogeny

Symbiotic interactions, from parasitism through to mutualism, influence global biogeochemical processes by shaping microbial community composition and phenology [15, 17]. In marine environments, parasites have been shown to terminate algal blooms and drive phytoplankton succession [18, 19] and carbon sequestration into the deep oceans by killing algae and facilitating transit of algal carcasses down the water column [15, 16]. Diatoms can be one of the most abundant eukaryotic algae in marine environments [20], but little is known about the top-down control of these algae, particularly the role of parasitism. In freshwater and some marine ecosystems, diatoms are parasitized by zoosporic fungi, chytrids (fungi that produce spore cells with a swimming tail) [21–24]. Oomycete protist parasites i have also been shown to act as a significant source of top-down control of toxic diatom bloom species [25]. However, in many marine ecosystems, the abundance and diversity of fungi, and their influence on food webs, remain poorly understood [3, 26]. The phylogenetic analyses of small subunit rRNA encoding gene (SSU rDNA) data demonstrate a diversity of sequences branching with chytrids [3, 5]. One environmental SSU rDNA group with weak phylogenetic affinities to known chytrids and which has been detected in both the sunlit water column and deep-sea sediments is the NCLC1 group [4, 5].

The opisthokonts include a huge diversity of eukaryotic forms but are composed of two major clades: (1) the animals and their protist relatives (the Holozoa) and (2) the fungi and their protist relatives (the Holomycota). Understanding the biodiversity of these groups is important for interpreting the evolutionary ancestry of these major clades (Figure 1A). The branching position of the NCLC1 group remains unresolved, with some analyses suggesting a weak association with holozoan taxa [6], while other analysis suggest that this

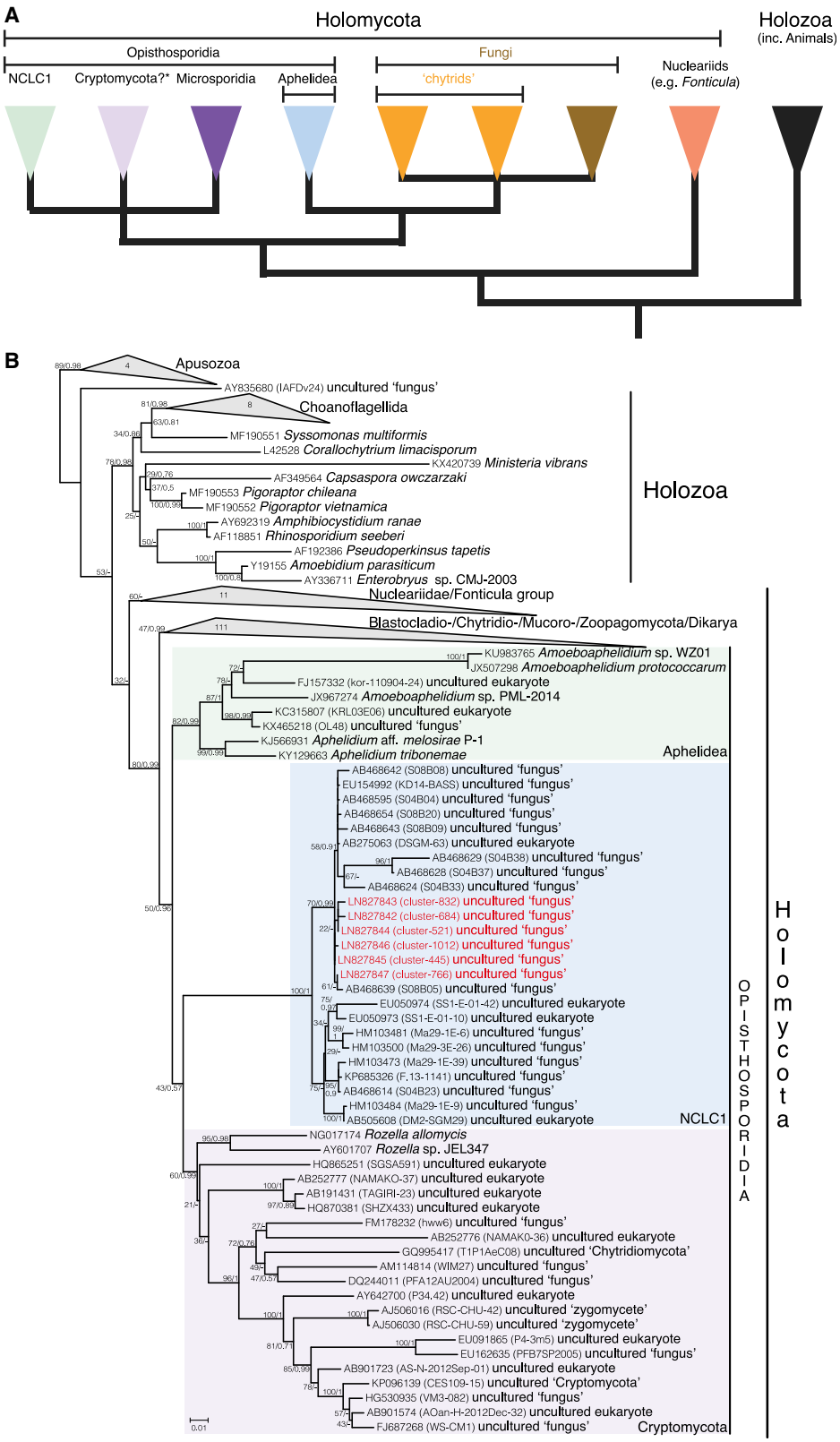


Figure 1. Phylogeny Assessing the Placement of the NCLC1 SSU rDNA Sequences Relative to the Fungi and Other Opithokonts

(A) Summary of the taxonomy and current best understood evolutionary relationships of the opisthokonts. Major phylogenetic relationships are based upon the phylogenetic data reported in [27].

(legend continued on next page)

group branches within the Holomycota [32] close to the Fungi [5]. Here, we investigated the phylogenetic position of the NCLC1 cluster by reconstructing maximum-likelihood (ML) and Bayesian phylogenetic trees including additional SSU rRNA gene sequences from a range of environmental DNA studies [5, 6, 33, 34]. NCLC1 sequences were found to branch sister to the Cryptomycota [35] (known variously as Rozellomycota [28], Rozellida [36], Rozellosporidia [7], or short-branch microsporidia [31]), which includes the genus *Rozella*. The NCLC1-Cryptomycota clade branches sister to the Aphelidea, with both relationships weakly supported by bootstrap analysis (Figure 1B). We recovered moderate bootstrap support (80%) for the separation of the Holomycota and the Holozoa, with the NCLC1 phylotypes clustered with the Holomycota. These results suggest that the NCLC1 represents a newly identified major group that branches with the Opisthosporidia and within the Holomycota, consistent with another analysis [31].

Fluorescent *In Situ* Hybridization Identification of NCLC1 Cells

A previous study on fungal molecular diversity in European coastal waters identified NCLC1-like sequence tags at high relative abundance compared to true fungal sequence tags at an Oslofjord (Norway) sampling site [4, 37]. NCLC1-like sequences were recovered from both large (3–20 μm and 20–1,000 μm) and small (0.6–3 μm) filtration fractions, suggesting that this group has a multifaceted life cycle, either coupled to the infection of larger cells and/or consisting of a larger, possibly multicellular life-cycle stage [4]. To explore the NCLC1 life cycle in marine environments, we used fluorescent *in situ* hybridization (FISH) microscopy to target cells from fixed filtrates sampled from the Oslofjord coastal station shown previously to harbor NCLC1 DNA/RNA diversity. Sampling was conducted to recover water from the sub-surface (1-meter depth) and deep chlorophyll maximum (DCM; 20-meter depth) fractions. From both depths, cells were recovered in two ways; water was sequentially filtered onto two different size-selective filters (0.6–3 μm and 3–20 μm) and recovered from a plankton net haul with a 1,000 μm pre-filtration sieve allowing for the recovery of cells in the range of 20–1,000 μm.

We designed three different FISH probes from the SSU V4 rRNA gene region: probe 1 (CHY-NCLC1-01), which is predicted to target the wider NCLC1 group, including OTU groups 445, 832, 521, 684, 766, and 1012 ([4]; Figure S1) and two probes (CHY-445-01 and CHY-445-02), which specifically target the NCLC1 OTU cluster 445, shown to be highly represented at Oslofjord ([4]; Figure S1). To test each probe, we used two alternative negative controls for comparison; these consisted of either the hybridization buffer without a DNA probe or with the reverse complement of each probe. In each case, the negative controls failed to detect candidate cells.

Using the true probes, the FISH approach identified a series of candidate cells. For all sample types assayed, we observed the same cell types with the three probes in independent hybridization experiments. The FISH probes identified four variant cell forms or "cellular types" (Figures 2, 3, and S2), indicative of detection of either a heterogenous population of microbes or a target group with multiple life-cycle phases. These cell types included an extracellular diatom association, intracellular diatom association, an un-associated, apparently free-living stage, and a multinucleated structure (also not associated with diatoms). The different FISH probe types recovered a similar percentage detection of each cell type and a similar detection profile (Figures 4A and 4B) across all filters, suggesting that the probes are independently, and consistently, detecting the same target population of cells. None of these or any other FISH-labeled cell types were identified in the negative controls.

Identification of NCLC1-Diatom Interactions

One of the four cell types observed using FISH microscopy was an irregular cellular form found inside a range of putative frustules (exoskeletons) of diatom species (Figures 2A–2F). This cell-cell association was only recovered in the 20–1,000 μm plankton net samples. Using bright-field microscopy, we discriminated the diatom' frustules from other phytoplankton species (e.g., dinoflagellates). Using calcofluor white (CFW) staining, which preferentially labels cellulose and/or chitin cell-wall structures, e.g., on the surface of thecate dinoflagellates [18], we further excluded the possibility that NCLC1 was associating with dinoflagellates or any other cells with chitin-cellulose cell walls.

For a separate parallel water mass sample taken at the same time as the FISH samples, the diversity and abundance of the most abundant planktonic species were identified and counted using microscopy of samples fixed with Lugol's solution [40] (Table S1; Figure S3). The combination of the taxonomic identifications obtained using diatom' frustule analysis from the FISH microscopy and species identifications from the fixed-sample analysis allowed us to identify the taxonomy of the host groups as *Chaetoceros*, *Skeletonema, Pseudonitzschia*, and *Leptocylindrus* diatoms (Figures 2A–2F).

We counted the number of diatom intracellular associations observed across the two sample depths (20–1,000 μm plankton net samples taken from both the surface and DCM water samples) using the three different probes. In each case, the individual FISH hybridization experiments were replicated three times each

(B) Maximum-likelihood phylogenetic tree inferred from an SSU alignment of 200 sequences (1,221 parsimony informative sites) under the GTR+F+R6 substitution model. For collapsed groups (gray triangles), the values indicate the number of sequences present in a given group. Aphelidea, NCLC1, and Cryptomycota* (also named Rozellomycota [28], Rozellida [29], Rozellosporidia [30], and/or short-branch microsporidia [31]) are represented by green, blue, and purple rectangles, respectively. The six representative sequences recovered from the Oslo coastal station [4] are colored in red. Numbers on branches are shown in the order of bootstrap support values (percentages; computed from 100 non-paramteric ML bootstrap replicates) and then posterior probabilities (inferred from two converged PhyloBayes chains). Code numbers in front of species names are NCBI - GenBank identifiers. Each collapsed branch is detailed in Table S4. The branch leading to *M. vibrans* was truncated for display purpose. SSU sequences of classically defined microsporidia are excluded from this analysis because they form excessively long branches in SSU rRNA gene trees. The phylogenetic tree is rooted on an Apusozoa outgroup; the scale bar represents the number of estimated substitutions per site. A variant reproduction of this tree is shown in Figure S1 with information about probe specificity annotated on the tree.

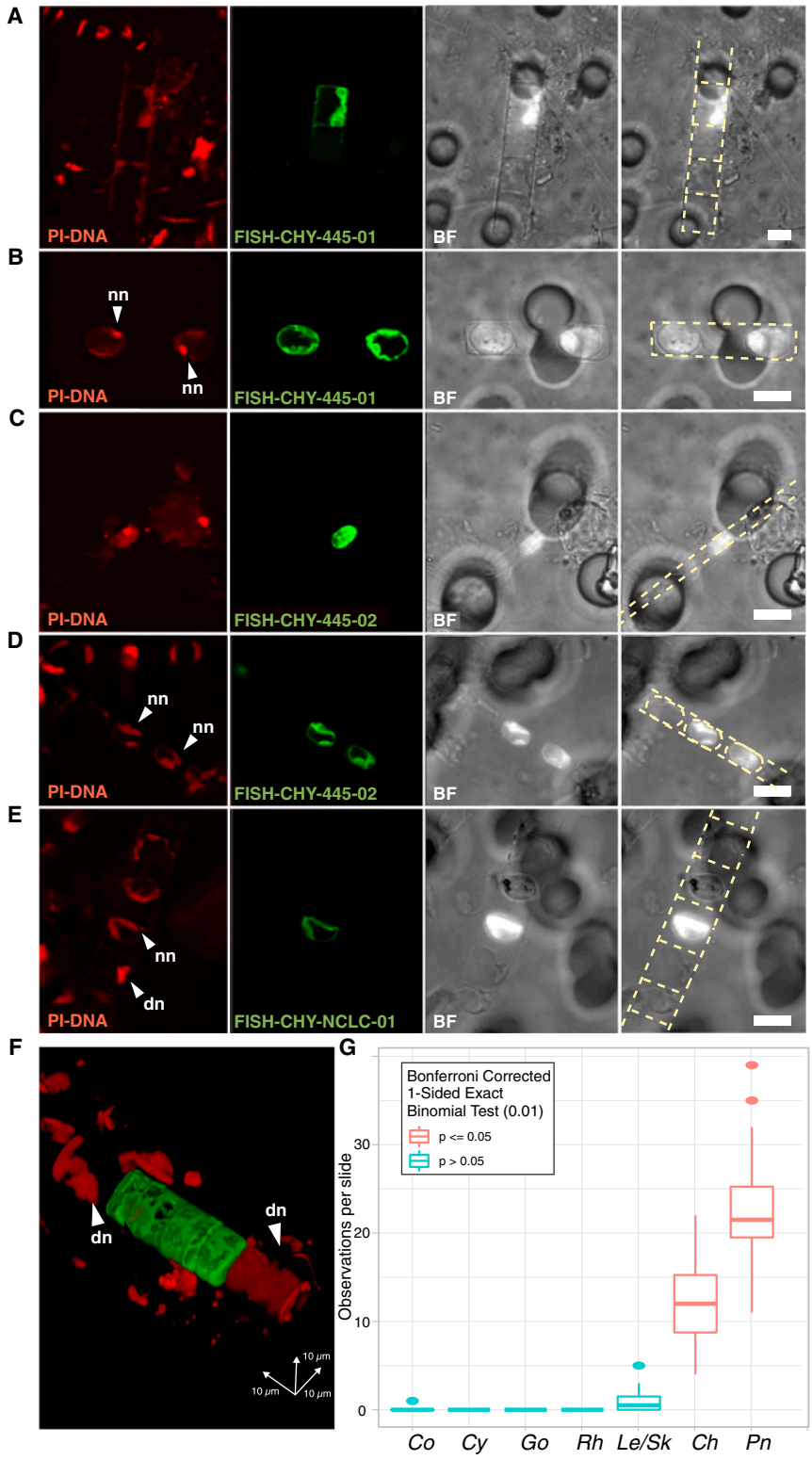


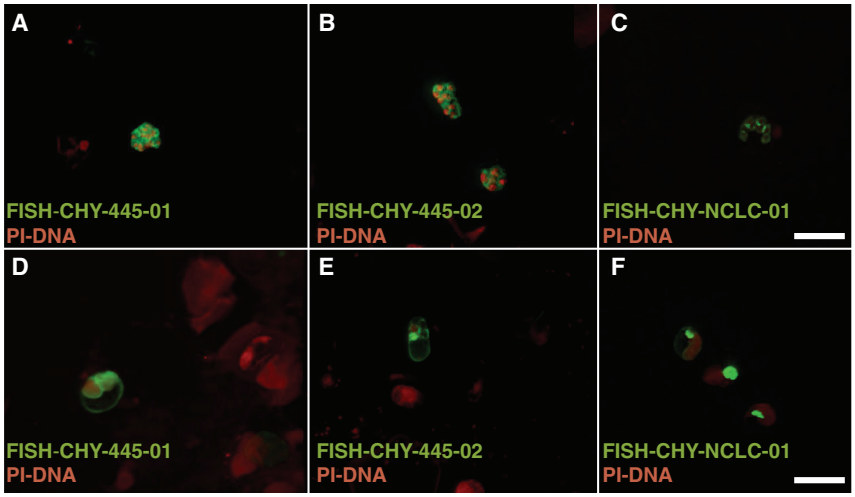
Figure 2. FISH Microscopy Evidence for NCLC1 Intracellular Associations with Diatom Phytoplankton

(A) Intracellular infection of *Chaetoceros*-like diatoms, (B) infection of *Leptocylindrus*-like diatoms, (C) infection of *Pseudonitzschia*-like diatoms, (D) infection of *Skeletonema*-like diatoms, and (E) infection of *Chaetoceros*-like diatoms. Scale bars, 10μm. PI corresponds to nuclear DNA staining with propidium iodide; green displays cells with a positive signal for the horseradish peroxidase (HRP) FISH-labeled probes, with the specific name of the probe included on each image. BF (bright-field) corresponds to the transmitted light with differential interference phase contrast.

(F) 3D confocal reconstruction micrograph displaying an intracellular infection of a *Chaetoceros* spp. diatom by NCLC1. Figure S2 contains more details of this image, including section images showing the presence of PI staining within the parasite conglomeration but absent from the diatom carcass. We used the 3D reconstruction here and in Figure S2 to investigate the precise intracellular geography of the NCLC1 cells; these micrographs show that all DNA-containing compartments within the infected cells are surrounded by the FISH probe, suggesting that they are NCLC1 nuclei. Furthermore, these DNA structures appear more condensed compared to the nuclei of the parallel uninfected diatoms present in the filament. Based on these observations, we hypothesize that the NCLC1 has consumed the host diatom, including the nucleus; however, we cannot exclude the possibility that the NCLC1 cell(s) has encapsulated the diatom nucleus. Either interaction would suggest a parasitic association. Interestingly, these 3D reconstructions suggest that the intracellular NCLC1 cell is showing properties of amoeba-like growth; such properties have been shown for *Rozella* [38] and *Aphelids* (e.g., [7]), and some bona fide fungal chytrids also show amoeba-like crawling [39]. Some of the micrographs show internal illumination of the diatom frustule. We note that (1) we did not see this in any of the negative controls and (2) the 3D reconstruction demonstrates that some of this is associated with cellular amoeba-like extensions of the NCLC1 cell, and (3) we hypothesize that some of this signal is derived from the reflection of the strong FISH light signal off the internal glass structure of diatom frustule. Importantly, this signal is absent from uninfected diatom cells next to the infected cells on the same filament (see Figure S2). Micrographs were obtained using a Confocal Zeiss LSM780 microscope. Putative diatom nuclei are marked with an arrowhead and are labeled "dn," and putative NCLC1 nuclei are marked with an arrowhead labeled "nn."

(G) Distribution of different observed extracellular NCLC1-diatom associations from FISH counts of the two water samples. Counts were summed per slide and probe type and then ordered by median values. Significance testing was performed per interaction under a corrected binomial test with blue indicating an adjusted p value >0.05 and red ≤0.05. *Co* is abbreviated from *Cocconeis* sp.; *Cy*, *Cylindrotheca* sp.; *Go*, *Gophonema* sp.; *Rh*, *Rhizosolenia* sp.; *Le/Sk*, *Leptocynlindrus/Skeletonema* sp.; *Ch*, *Chaetoceros* sp.; *Pn*, *Pseudonitzschia* sp. Box and whisker plots are shown with the outliers excluded (values more than 1.5x the inter-quartile range from the 1st or 3rd quartile) and are shown as dots. The bottom of the whisker line indicates the minimum, the start of box is the first quartile, the median line is shown, the 3rd quartile is the top of box, and maximum is the top of the whisker line.

See also Figure S3 and Table S1.



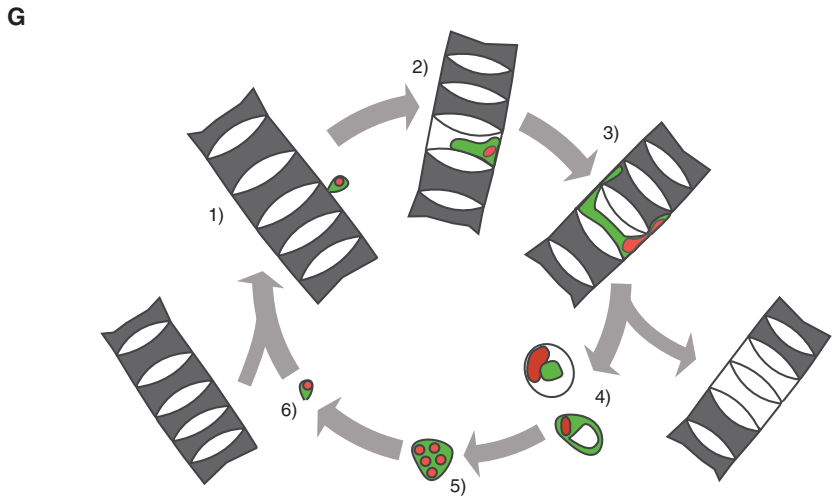


Figure 3. Merged Epifluorescence Micrographs of Additional NCLC1 Putative Life-Cycle Stages Detected and a Cartoon Showing the Inferred Life Cycle of the NCLC1 Group Identified

(A–F) FISH microscopy evidence for additional putative life-cycle phases of the NCLC1 cells identified is shown. The red and green colors correspond with the nuclear staining with propidium iodide and the FISH positive signal of the labeled probes, respectively (probes are named on the bottom corner of each image). (A–C) putative multinucleated structures and (D–F) putative spore or cyst life stage or potentially secondary host associations. The cells identified in (D), (E), and (F) are marked by sub-compartment localization of the FISH probe, suggesting either that the cyst cells have an extensive vacuole or organelle systems or that these cells represent infections of additional secondary hosts. Scale bars, 10 μm.

(G) Carton illustration of the putative partial life cycle of the NCLC1 cluster 445 organisms; these are surmised from the FISH data presented and demonstrate (1) attachment to host diatoms, (2) intracellular invasion (Figures 2 and S2), (3) spread of NCLC1 between cells in a diatom filament (Figure S2), (4) release of NCLC1 cells or a secondary infection (Figures 3D–3F), and (5) NCLC1 multinucleate phase (Figures 3A–3C).

with three independant counts. These analyses consistently demonstrated that ~2%–8% of the FISH identified cells were indicative of intracellular infections of diatoms (Figures 4A and 4B; Table S2).

Counterstaining of DNA with propidium iodide demonstrated that the FISH probe staining often surrounded a DNA structure. In several of the images, the diatom nucleus was not visible within the infected cell as a separate entity (Figure 2A, 2B, 2E, and S2), yet in some cases the diatom nucleus was observed in uninfected diatom cells residing next to the NCLC1-infected cell within the diatom filament (Figures 2E and S2). These microscopy results suggest that the DNA structure identified is either the NCLC1 nucleus/nuclei and that the diatom nucleus is absent, suggesting, in some cases, that the NCLC1 association is with a diatom carcass, or, alternatively, that the host diatom nucleus is actually surrounded by the infecting NCLC1 cell(s) (see Figures 2B, 2D, 2E, and S2, with further rationale outlined in the Figure 2 legend). Either characteristic implies a parasitic interaction; however, we note that such analyses are complicated by the FISH process of sampling, which can damage cells, and the limitations of microscopy, which cannot completely account for cellular structures throughout the z axis of the microscopic field. These limitations also prevent quantitative comparisons. Nonetheless, these results are consistent with the hypothesis that NCLC1 cells are present within dead diatom cells or diatom cells with nuclei smothered by the NCLC1 infection.

In addition to the intracellular associations identified, the FISH analysis demonstrated a large proportion of candidate NCLC1 cells positioned proximate to a diatom cell surface, suggesting an epibiotic (surface-to-surface) association. Three diatom genera were detected in high cell concentrations: *Chaetoceros, Skeletonema*, and *Pseudonitzschia* (Table S1; Figure S3). From our FISH micrographs, we inferred the taxonomic affiliation of the NCLC1-bearing diatoms based on the bright-field silhouette present on the filter. These results indicate that the NCLC1 association was present across a range of diatom hosts, including members from the most abundant diatom genera identified (i.e., *Chaetoceros*, *Skeletonema*, and *Pseudonitzschia*). We note that the pattern of epibiotic associations was similar for both surface and DCM water-column samples (~30%; see Table S2; Figures 4A and 4B), consistent with the detection of similar patterns of phytoplankton biodiversity present in both samples (Figure S3; Table S1), a result that suggests that the surface and DCM zones were highly homogeneous in terms of diatom species community composition.

Although this putative extracellular association and, indeed, the intracellular associations discussed above suggest a symbiotic-parasitic interaction between NCLC1 and diatoms, these

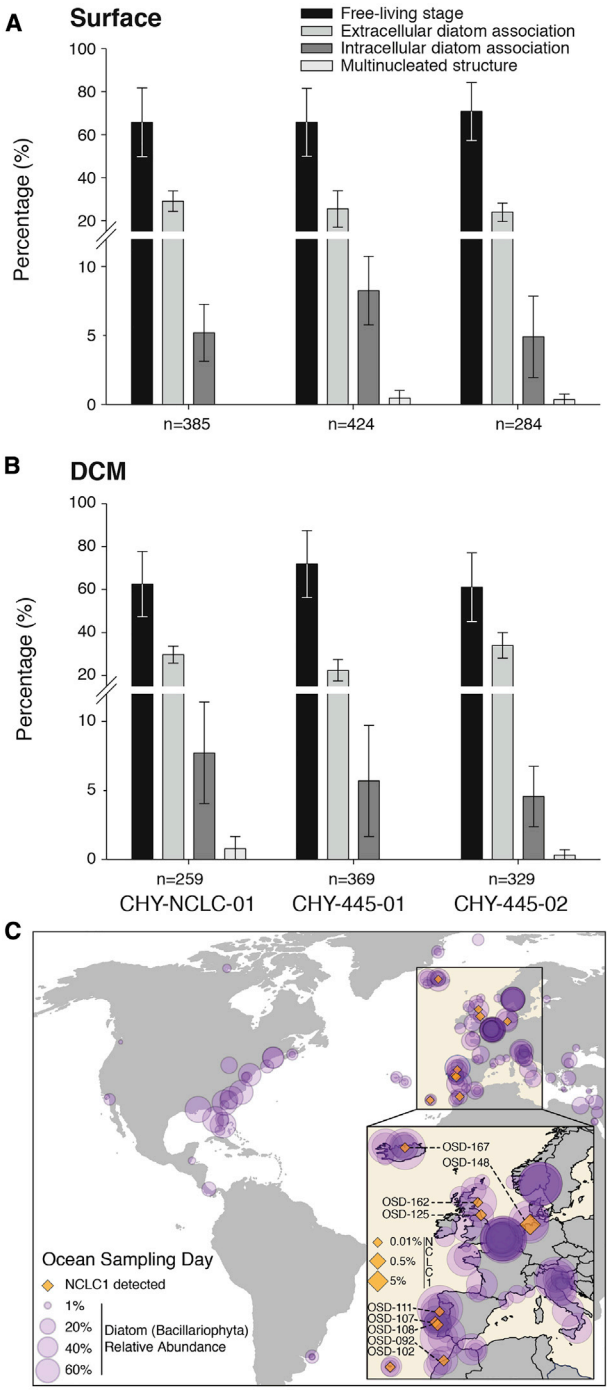


Figure 4. Detection Provenance of NCLC1

(A and B) Percentage of different cell types recovered in the FISH analysis from (A) Sub-surface and (B) DCM depths. The different probes used are listed across the x axis of (B). The total number of NCLC1 cells identified using each probe is also listed below the x axis. Mean % shown is derived from independent hybridization experiments conducted on three separate filter pieces per probe. In each case, three independent counts were conducted per filter piece with a minimum of 200 FISH positive cells observed. Error bars indicate the standard deviation. See Table S2 for data.

(C) Geographical distribution of NCLC1 and diatoms across the Ocean Sampling Day (OSD) data. Samples where SSU-V4 amplicon sequence variants (ASV) classified as NCLC1 and diatoms (Bacillariophyta) were detected are indicated by orange diamonds and purple circles, NCLC1 and diatoms, respectively. Diamond and circle marker sizes are scaled according to NCLC1 and diatom relative abundance recovered (see key).

See also Table S3.

observations could also be the product of filtration, which artifactually, but consistently, suggests a physical contact between abundant diatom species and NCLC1 cells. To test if the observed NCLC1-diatom extracellular associations were a filtration artifact, we used conservative Bonferroni-corrected exact binomial tests with a minimal hypothesized association parameter (0.01). This resulted in rejection of a null hypothesis of minimal chance interaction between NCLC1 and *Chaetoceros* spp. and between NCLC1 and *Pseudonitzschia* spp. (both intracellular or extracellular diatom associations were counted), regardless of probe used (Figure 2G). Collectively, these data suggest that the NCLC1 group detected is forming a bona fide cellular association with multiple diatom species.

To further explore this association, we used the publicly available 2014 Ocean Sampling Day citizen science [41] projects' SSU V4 rRNA gene sequences to examine the co-occurrences of NCLC1 amplicon sequence variants (ASVs) with other eukaryotic ASVs (see Table S3). These data demonstrate that the detection of NCLC1 was geographically restricted to the North-East Atlantic and the North Sea (Figure 4C). This constitutes an incomplete and biased sampling profile, but it suggests that NCLC1 is resident in these environments but is absent and/or undetected from water samples on the western side of the Atlantic on the date of sampling. This analysis identifies a relatively low abundance for NCLC1, except for one sample (OSD-148), which demonstrated an ~4% relative abundance of an NCLC1 SSU rRNA gene sequence (Figure 4C and Table S3). We identified three significant NCLC1 co-occurrences within the Ocean Sampling Day data; one association was between an NCLC1 ASV and a rhizarian ASV, although we note that sampling of this ASV within the Ocean Sampling Day dataset had limited reads and so may be an artifact, even though a significant pseudo p value was recovered (Table S3). More convincingly, and consistent with the FISH identification of a NCLC1-diatom association reported here, we identified two significant co-occurrence patterns between an NCLC1 and a diatom ASV, providing further support for this interaction across geographically distributed sites (Figure 4C; Table S3).

Detection of Additional NCLC1 Life-Cycle Stages

The FISH experiments also detected two additional variant cell types that were absent from the negative controls and were not associated with diatoms or an identifiable second-party cell, suggesting detection of free-living forms of the target NCLC1 group. This alternative cell type was detected in the 3–20 μm filtrate and the 20–1,000 μm plankton net samples. Using both specific and general NCLC1 probes, we observed a multi-nucleated structure (Figures 3A–3C). This structure was rarely seen in our samples but was detected independently using all three probes and from samples recovered from both sub-surface and DCM water fractions. Specifically, we identified 0%, 0.47% (±SD 0.55), and 0.35% (±0.41) of all the FISH-detected cells from the sub-surface water samples and 0.77% (±0.89),

0%, and 0.3% (±0.41) of all the FISH-detected cells from the DCM water samples. These data are consistent with the hypothesis that the putative NCLC1 cells detected form a multinucleate sporangium-like reproductive life-cycle stage. The low detection rate suggests that this form is rare and/or short-lived; alternatively, the process of fixation or filtration may have disrupted these multicellular life-cycle stages, making them difficult to recover.

The second cell type observed was ovoid (length = 7.70 ± 1.08 μm, width = 3.72 ± 0.55 μm, n = 20) or round (diameter = 6.73 ± 0.58 μm, n = 20). These cellular structures are likely to correspond to either a spore or a cyst life-cycle stage (Figures 3D–3F) or alternatively they may represent an association with a yet- unidentifiable second host. We identified a high proportion of these putative life-cycle-stage cells (60%–70%; see Figures 4A and 4B) from both sample depths using the general and specific NCLC1 probes. CFW staining of the filters coupled with FISH microscopy demonstrated that all cell types identified did not possess a detectable cellulose and/or chitin cell wall; however, we note that the cells sampled possibly represent only a fraction of the NCLC1 life cycle.

Conclusions

These results demonstrate a hitherto-undetected intracellular infection of diatoms, an ecologically important group of marine phytoplankton, including *Chaetoceros*, *Skeletonema*, and *Pseudonitzschia* species, that can form blooms in marine waters. The host range also includes groups responsible for harmful algal blooms (e.g., *Pseudonitzschia* spp.). The infectious agent constitutes a phylogenetically unique branch—most likely a distinct and diverse addition to the Holomycota and possibly the Opisthosporidia [7]—adding an additional branch close to the base of the radiation of the fungal kingdom. The nature of the NCLC1-diatom interaction is unknown but potentially represents a parasitic infection, a mutualistic interaction, a saprotrophic degradation of dead diatoms, or, indeed, an infection that transitions between all three modes of interaction. NCLC1 DNA- and RNA-derived sequences have been detected in the marine water column and sediments, suggesting that NCLC1 is active in both pelagic and benthic environments [4]. As such, the NCLC1 cells may follow diatom carcasses into the marine sediment as saprotrophic degraders of these phytoplankton cells. Consistent with this later hypothesis, we observe NCLC1 as an intracellular infection within diatoms with no identifiable nuclei next to diatoms with identifiable nuclei, consistent with the hypothesis that the host is dead and/or that the diatom nucleus has been smothered by NCLC1 cells. We therefore suggest that this interaction represents a necrotrophic-parasitic interaction followed by a saprotrophic interaction with *Chaetoceros*, *Skeletonema*, and *Pseudonitzschia* diatom carcasses. As such, NCLC1 joins an increasing list of viral (e.g., [42, 43]), protist [2, 24, 25], and fungal pathogens, including putative chytrid associations [44, 45], which are hypothesized to infect diatoms and determine the fate of important phytoplankton blooms.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2019.09.074.

ACKNOWLEDGMENTS

The authors are grateful to Vladyslava Hostyeva for phytoplankton counting and Valentin Foulon for his help with the epifluorescence microscope and the Cytometry core facilities of LEMAR, Brest. We thank Prof. Keith Gull for use of TAT1 antibody. A.C. was supported by the ANR project ACHN 2016 PARASED (ANR-16_ACHN_0003). F.M. is supported by Genome Canada via a Genome Atlantic Postdoctoral Fellowship and a Donald Hill Family Fellowship. A.M. and T.A.R. are funded by the Royal Society through University Research Fellowships. Parts of this project were supported by the *BiodivERsA* ERA-Net project *BioMarKs* and a Gordon and Betty Moore Foundation MMI Grant (GBMF3307).

AUTHOR CONTRIBUTIONS

A.C., A.M., and T.A.R. conceived and designed the study. A.C. conducted FISH Microscopy, A.M. and A.C. conducted phylogenetic and Ocean Day sampling analysis, and F.M. designed and conducted statistical analysis. B.E. and W.E. conducted marine sampling and conducted plankton counts. All authors contributed to drafting the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 20, 2019
Revised: July 12, 2019
Accepted: September 30, 2019
Published: November 14, 2019

REFERENCES

1. James, T.Y., Kauff, F., Schoch, C.L., Matheny, P.B., Hofstetter, V., Cox, C.J., Celio, G., Gueidan, C., Fraker, E., Miadlikowska, J., et al. (2006). Reconstructing the early evolution of Fungi using a six-gene phylogeny. Nature *443*, 818–822.
2. Gleason, F.H., Kagami, M., Lefevre, E., and Sime-Ngando, T. (2008). The ecology of chytrids in aquatic ecosystems: roles in food web dynamics. Fungal Biol. Rev. *22*, 17–25.
3. Richards, T.A., Jones, M.D.M., Leonard, G., and Bass, D. (2012). Marine fungi: their ecology and molecular diversity. Annu. Rev. Mar. Sci. *4*, 495–522.
4. Richards, T.A., Leonard, G., Mahé, F., Del Campo, J., Romac, S., Jones, M.D.M., Maguire, F., Dunthorn, M., De Vargas, C., Massana, R., and

Chambouvet, A. (2015). Molecular diversity and distribution of marine fungi across 130 European environmental samples. Proc. Biol. Sci. *282*, 20152243.

5. Nagahama, T., Takahashi, E., Nagano, Y., Abdel-Wahab, M.A., and Miyazaki, M. (2011). Molecular evidence that deep-branching fungi are major fungal components in deep-sea methane cold-seep sediments. Environ. Microbiol. *13*, 2359–2370.
6. Bass, D., Howe, A., Brown, N., Barton, H., Demidova, M., Michelle, H., Li, L., Sanders, H., Watkinson, S.C., Willcock, S., and Richards, T.A. (2007). Yeast forms dominate fungal diversity in the deep oceans. Proc. Biol. Sci. *274*, 3069–3077.
7. Karpov, S.A., Mamkaeva, M.A., Aleoshin, V.V., Nassonova, E., Lilje, O., and Gleason, F.H. (2014). Morphology, phylogeny, and ecology of the aphelids (Aphelidea, Opisthokonta) and proposal for the new superphylum Opisthosporidia. Front. Microbiol. *5*, 112.
8. Letcher, P.M., Lopez, S., Schmieder, R., Lee, P.A., Behnke, C., Powell, M.J., and McBride, R.C. (2013). Characterization of *Amoeboaphelidium protococcarum*, an algal parasite new to the cryptomycota isolated from an outdoor algal pond used for the production of biofuel. PLoS ONE *8*, e56232.
9. Richards, T.A., Leonard, G., and Wideman, J.G. (2017). What defines the "kingdom" fungi? Microbiol. Spectr. *5*, https://doi.org/10.1128/microbiolspec.FUNK-0044-2017.
10. Powell, M.J. (1984). Fine structure of the unwalled thallus of *Rozella polyphagi* in its host *Polyphagus euglenae*. Mycologia *76*, 1039–1048.
11. Corradi, N. (2015). Microsporidia: eukaryotic intracellular parasites shaped by gene loss and horizontal gene transfers. Annu. Rev. Microbiol. *69*, 167–183.
12. Falkowski, P.G., Barber, R.T., and Smetacek, V. (1998). Biogeochemical controls and feedbacks on ocean primary production. Science *281*, 200–207.
13. Field, C.B., Behrenfeld, M.J., Randerson, J.T., and Falkowski, P. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. Science *281*, 237–240.
14. Armbrust, E.V. (2009). The life of diatoms in the world's oceans. Nature *459*, 185–192.
15. Worden, A.Z., Follows, M.J., Giovannoni, S.J., Wilken, S., Zimmerman, A.E., and Keeling, P.J. (2015). Environmental science. Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. Science *347*, 1257594.
16. Durkin, C.A., Van Mooy, B.A.S., Dyhrman, S.T., and Buesseler, K.O. (2016). Sinking phytoplankton associated with carbon flux in the Atlantic Ocean. Limnol. Oceanogr. *61*, 1172–1187.
17. Sime-Ngando, T. (2012). Phytoplankton chytridiomycosis: fungal parasites of phytoplankton and their imprints on the food web dynamics. Front. Microbiol. *3*, 361.
18. Chambouvet, A., Morin, P., Marie, D., and Guillou, L. (2008). Control of toxic marine dinoflagellate blooms by serial parasitic killers. Science *322*, 1254–1257.
19. Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., Chaffron, S., Ignacio-Espinosa, J.C., Roux, S., Vincent, F., et al.; Tara Oceans coordinators (2015). Ocean plankton. Determinants of community structure in the global plankton interactome. Science *348*, 1262073.
20. Shubha, S., Louisa, W., Emmanuel, D., Trevor, P., Carla, C., and Heidi, M. (2004). Discrimination of diatoms from other phytoplankton using ocean-colour data. Mar. Ecol. (Berl.) *272*, 59–68.
21. Ibelings, B., de Bruin, A., Kagami, M., Rijkeboer, M., Brehm, M., and Donk, E. (2004). Host parasite interactions between freshwater phytoplankton and chytrid fungi (Chytridiomycota). J. Phycol. *40*, 437–453.
22. Canter, H.M., and Lund, J.W.G. (1951). Studies on plankton parasites: III. examples of the interaction between parasitism and other factors determining the growth of diatoms. Ann. Bot. *15*, 359–371.
23. Sparrow, F.K. (1960). Aquatic Phycomycetes (University of Michigan Press).
24. Scholz, B., Guillou, L., Marano, A.V., Neuhauser, S., Sullivan, B.K., Karsten, U., Küpper, F.C., and Gleason, F.H. (2016). Zoosporic parasites infecting marine diatoms - a black box that needs to be opened. Fungal Ecol. *19*, 59–76.
25. Garvetto, A., Nézan, E., Badis, Y., Bilien, G., Arce, P., Bresnan, E., Gachon, C.M.M., and Siano, R. (2018). Novel widespread marine oomycetes parasitising diatoms, including the toxic genus *Pseudo-nitzschia*: genetic, morphological, and ecological characterisation. Front. Microbiol. *9*, 2918.
26. Lepère, C., Ostrowski, M., Hartmann, M., Zubkov, M.V., and Scanlan, D.J. (2016). In situ associations between marine photosynthetic picoeukaryotes and potential parasites - a role for fungi? Environ. Microbiol. Rep. *8*, 445–451.
27. Torruella, G., Grau-Bové, X., Moreira, D., Karpov, S.A., Burns, J.A., Sebé-Pedrós, A., Völcker, E., and López-García, P. (2018). Global transcriptome analysis of the aphelid *Paraphelidium tribonemae* supports the phagotrophic origin of fungi. Commun. Biol. *1*, 231.
28. James, T.Y., and Berbee, M.L. (2012). No jacket required–new fungal lineage defies dress code: recently described zoosporic fungi lack a cell wall during trophic phase. BioEssays *34*, 94–102.
29. Irino, K., Vaz, T.M., Kato, M.A., Naves, Z.V., Lara, R.R., Marco, M.E., Rocha, M.M., Moreira, T.P., Gomes, T.A., and Guth, B.E. (2002). O157:H7 Shiga toxin-producing Escherichia coli strains associated with sporadic cases of diarrhea in São Paulo, Brazil. Emerg. Infect. Dis. *8*, 446–447.
30. Karpov, S.A., Torruella, G., Moreira, D., Mamkaeva, M.A., and López-García, P. (2017). Molecular Phylogeny of *Paraphelidium letcheri* sp. nov. (Aphelida, Opisthosporidia). J. Eukaryot. Microbiol. *64*, 573–578.
31. Bass, D., Czech, L., Williams, B.A.P., Berney, C., Dunthorn, M., Mahé, F., Torruella, G., Stentiford, G.D., and Williams, T.A. (2018). Clarifying the relationships between microsporidia and cryptomycota. J. Eukaryot. Microbiol. *65*, 773–782.
32. Liu, Y., Steenkamp, E.T., Brinkmann, H., Forget, L., Philippe, H., and Lang, B.F. (2009). Phylogenomic analyses predict sistergroup relationship of nucleariids and fungi and paraphyly of zygomycetes with significant support. BMC Evol. Biol. *9*, 272.
33. 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 raw sediment samples and two enrichment cultures. Extremophiles *11*, 563–576.
34. Tian, F., Yu, Y., Chen, B., Li, H., Yao, Y.-F., and Guo, X.-K. (2009). Bacterial, archaeal and eukaryotic diversity in Arctic sediment as revealed by 16S rRNA and 18S rRNA gene clone libraries analysis. Polar Biol. *32*, 93–103.
35. Jones, M.D.M., Forn, I., Gadelha, C., Egan, M.J., Bass, D., Massana, R., and Richards, T.A. (2011). Discovery of novel intermediate forms redefines the fungal tree of life. Nature *474*, 200–203.
36. Lara, E., Moreira, D., and López-García, P. (2010). The environmental clade LKM11 and *Rozella* form the deepest branching clade of fungi. Protist *161*, 116–121.
37. Logares, R., Audic, S., Bass, D., Bittner, L., Boutte, C., Christen, R., Claverie, J.-M., Decelle, J., Dolan, J.R., Dunthorn, M., et al. (2014). Patterns of rare and abundant marine microbial eukaryotes. Curr. Biol. *24*, 813–821.
38. Powell, M.J., Letcher, P.M., and James, T.Y. (2017). Ultrastructural characterization of the host-parasite interface between *Allomyces anomalus* (Blastocladiomycota) and *Rozella allomycis* (Cryptomycota). Fungal Biol. *121*, 561–572.
39. Fritz-Laylin, L.K., Lord, S.J., and Mullins, R.D. (2017). WASP and SCAR are evolutionarily conserved in actin-filled pseudopod-based motility. J. Cell Biol. *216*, 1673–1688.
40. Dittami, S.M., Hostyeva, V., Egge, E.S., Kegel, J.U., Eikrem, W., and Edvardsen, B. (2013). Seasonal dynamics of harmful algae in outer Oslofjorden monitored by microarray, qPCR, and microscopy. Environ. Sci. Pollut. Res. Int. *20*, 6719–6732.

41. Kopf, A., Bicak, M., Kottmann, R., Schnetzer, J., Kostadinov, I., Lehmann, K., Fernandez-Guerra, A., Jeanthon, C., Rahav, E., Ullrich, M., et al. (2015). The ocean sampling day consortium. Gigascience *4*, 27.
42. Kimura, K., and Tomaru, Y. (2015). Discovery of two novel viruses expands the diversity of single-stranded DNA and single-stranded RNA viruses infecting a cosmopolitan marine diatom. Appl. Environ. Microbiol. *81*, 1120–1131.
43. Kim, J., Kim, C.-H., Youn, S.-H., and Choi, T.-J. (2015). Isolation and physiological characterization of a novel algicidal virus infecting the marine Diatom *Skeletonema costatum*. Plant Pathol. J. *31*, 186–191.
44. Gutiérrez, M.H., Jara, A.M., and Pantoja, S. (2016). Fungal parasites infect marine diatoms in the upwelling ecosystem of the Humboldt current system off central Chile. Environ. Microbiol. *18*, 1646–1653.
45. Hanic, L.A., Sekimoto, S., and Bates, S.S. (2009). Oomycete and chytrid infections of the marine diatom *Pseudo-nitzschia* pungens (Bacillariophyceae) from Prince Edward Island, Canada. Botany *87*, 1096–1105.
46. Tragin, M., and Vaulot, D. (2018). Green microalgae in marine coastal waters: the Ocean Sampling Day (OSD) dataset. Sci. Rep. *8*, 14020.
47. Yilmaz, L.S., Parnerkar, S., and Noguera, D.R. (2011). mathFISH, a web tool that uses thermodynamics-based mathematical models for in silico evaluation of oligonucleotide probes for fluorescence in situ hybridization. Appl. Environ. Microbiol. *77*, 1118–1122.
48. Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glöckner, F.O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. *35*, 7188–7196.
49. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.
50. Katoh, K., and Toh, H. (2010). Parallelization of the MAFFT multiple sequence alignment program. Bioinformatics *26*, 1899–1900.
51. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972–1973.
52. Gouy, M., Guindon, S., and Gascuel, O. (2010). SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. *27*, 221–224.
53. Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274.
54. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods *14*, 587–589.
55. Lartillot, N., Lepage, T., and Blanquart, S. (2009). PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics *25*, 2286–2288.
56. Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. Nat. Methods *13*, 581–583.
57. Wickham, H. (2016). ggplot2: elegant graphics for data analysis (Springer-Verlag).
58. Friedman, J., and Alm, E.J. (2012). Inferring correlation networks from genomic survey data. PLoS Comput. Biol. *8*, e1002687.
59. Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., de Vargas, C., Decelle, J., et al. (2013). The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. Nucleic Acids Res. *41*, D597–D604.
60. Hasle, G.R. (1978). The inverted-microscope method. In Phytoplankton Manual, A. Sournia, ed. (Paris: United Nations Educational, Scientific and Cultural Organisation), pp. 88–96.
61. Utermöhl, H. (1958). Zur vervollkommnung der quantitativen phytoplankton methodik. Mitt. Int. Ver. Theor. Agnew. Limnol. *9*, 1–38.
62. Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S., Jobb, G., et al. (2004). ARB: a software environment for sequence data. Nucleic Acids Res. *32*, 1363–1371.
63. Woods, A., Sherwin, T., Sasse, R., MacRae, T.H., Baines, A.J., and Gull, K. (1989). Definition of individual components within the cytoskeleton of *Trypanosoma brucei* by a library of monoclonal antibodies. J. Cell Sci. *93*, 491–500.
64. Soubrier, J., Steel, M., Lee, M.S.Y., Der Sarkissian, C., Guindon, S., Ho, S.Y.W., and Cooper, A. (2012). The influence of rate heterogeneity among sites on the time dependence of molecular rates. Mol. Biol. Evol. *29*, 3345–3358.
65. Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. *73*, 5261–5267.
66. Katoh, K., Kuma, K., Toh, H., and Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. *33*, 511–518.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Tat1 Tubulin Antibody	Gift K. Gull Uni. Oxford, By request	TAT1; RRID: AB_10013740
Fluorescein isothiocyanate (FITC)-conjugated goat anti-mouse immunoglobulins	Jackson ImmunoResearch/ Stratech	https://www.jacksonimmuno.com/catalog/products/115-095-003
Biological Samples		
Filter sections taken as part of the BioMarks project	Biomarks Consortium, by request, but exhaustible.	N/A
Chemicals, Peptides, and Recombinant Proteins		
Calcofluor White	Sigma-Aldrich, USA	https://www.sigmaaldrich.com/catalog/product/sial/18909?lang=en&region=GB
Deposited Data		
All physical and chemical parameters of the samples water column obtained using a CTD	This paper	are available at http://biomarks.eu/ctd007 (and replicated at FigShare https://doi.org/10.6084/m9.figshare.9821936)
tree file, masked and unmasked SSU rDNA alignments	This paper	Zenodo repository: https://doi.org/10.5281/zenodo.2788876.
Ocean Sampling Day 2014 Data	[41, 46]	https://github.com/MicroB3-IS/osd-analysis/wiki/Guide-to-OSD-2014-data see also http://mb3is.megx.net/osd-files?path=/2014/protocols
Oligonucleotides		
FISH Probe (5′-3′) GTCCTAGATTCACTGCTC	This paper, ordered from biomers.net (Germany)	CHY-NCLC-01
FISH Probe (5′-3′) GATTCTAATGCCCCCCAA	This paper, ordered from biomers.net (Germany)	CHY-445-01
FISH Probe (5′-3′) CGATTCTAATGCCCCCCA	This paper, ordered from biomers.net (Germany)	CHY-445-02
FISH Probe (5′-3′) [reverse complement negative control] GAGCTGTGAATCTAGGAC	This paper, ordered from biomers.net (Germany)	CHY-NCLC-01_RC
FISH Probe (5′-3′) [reverse complement negative control] TTGGGGGGCATTAGAATC	This paper, ordered from biomers.net (Germany)	445_01_RC
FISH Probe (5′-3′) [reverse complement negative control] GTTGGGGGGCATTAGAAT	This paper, ordered from biomers.net (Germany)	445_02_RC
Software and Algorithms		
The R code used to test statistical association between NCLC1 and Diatoms in the FISH data	This paper	Zenodo repository: https://doi.org/10.5281/zenodo.2788876.
mathFISH for FISH probe design	[47]	http://mathfish.cee.wisc.edu
ARB software (v.6.0.4) for SSU rRNA probe design	[48]	(https://www.arb-silva.de)
TESTPROBES for FISH probe optimization	[48]	(https://www.arb-silva.de)
R programming language (RCore2013) ‘tidyverse’ set of tools for statisitical analysis	N/A	(https://www.tidyverse.org/)
BLASTN similarity search for sequences with shared sequence identify from NCBI nt database	[49]	https://blast.ncbi.nlm.nih.gov
MAFFT v7.2 sequence alignment for automated sequence alignment	[50]	https://mafft.cbrc.jp/alignment/server/
trimAL v4 for automated sequence alignment refinement and sampling	[51]	http://trimal.cgenomics.org

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
seaview v4 manual sequence alignment program	[52]	http://doua.prabi.fr/software/seaview
IQ-TREE v1.6 for phylogenetic analysis	[53]	http://www.iqtree.org/release/v1.6.7
ModelFinder for finding appropriate model of sequence evolution for phylogenetic analysis	[54]	http://www.iqtree.org/ModelFinder/
PHYLOBAYES v3.3 for Bayesian phylogenetic analysis	[55]	http://megasun.bch.umontreal.ca/People/lartillot/www/download.html
DADA2 for sequence tag analysis	[56]	https://benjjneb.github.io/dada2/dada-installation.html
R package 'ggmap'	[57]	https://github.com/dkahle/ggmap
SparCC	[58]	https://bitbucket.org/yonatanf/sparcc
Other		
PR^2 v4.10 SSU rDNA reference database	[59]	https://github.com/pr2database/pr2database

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact Thomas Richards (T.A.Richards@exeter.ac.uk). This study did not generate any unique reagents other than the FISH oligonucleotide probes. Details of these probes are available in the key resource table.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Sampling

Samples were taken as part of the BioMarKs project (http://www.biomarks.eu) [37] in the outer Oslofjorden station OF (59.253735N, 10.710908E) on the 22nd September 2009. All physical and chemical parameters of the water column obtained using a CTD are available at http://biomarks.eu/ctd007 (and replicated here DOI: 10.6084/m9.figshare.9821936). Water and plankton samples were collected from sub-surface at 1 m depth and the DCM at 20 m depth from the University of Oslo research vessel R/V Trygve Braarud. Sampling was conducted using either: 1) a plankton net with 20 μm mesh-size, for a horizontal net haul where the net sample was then passed through a 1000 μm metallic sieve, or 2) Niskin bottles for collecting water samples. Aliquots from all samples were fixed onboard with neutralized formaldehyde (3.7% final conc.) and kept at 4°C until processed in the lab the day after. In the lab the samples from the plankton net (with a 20 μm 'aperture' size) were collected onto a 47 mm polycarbonate (PC) filters of 12 μm pore-size rendering recover of cells of 20 - 1000 μm diameter. For the Niskin bottle samples, water was pre-filtrated through 20 μm nylon sieve and then successively size-fractionated throughout 3 μm and 0.8 μm PC filters of 25 mm diameter. All filters (including plankton net water samples of 20- 1000 μm and water samples of 3- 20 μm and 0.6- 3 μm size fractions) were then dehydrated in sequential 50%, 80% and 100% ethanol incubations with 3 min of incubation at each step followed by drying at room temperature before final storage at −80°C.

METHOD DETAILS

Phytoplankton counts

Cell counts was performed on 10 mL water samples collected by the Niskin bottles and fixed immediately with Lugol's solution (1% final. conc.). Cells were counted using the Utermöhl method [60, 61].

Probe design

FISH probes were designed based on an alignment of 136 sequences and 316 alignment positions of the V4 region of the SSU rRNA encoding gene (see [4]) using the probes-design tool available through ARB v6.0.4 [62]. We designed three oligonucleotide probes, a general probe targeting the wider NCLC1 clade named CHY-NCLC-01 and two probes, CHY-445-01 and CHY-445-02, which specifically target the cluster 445 (see Key Resources Table). The thermodynamic parameters for all three probes were evaluated using mathFISH [47]. For the two specific probes of the cluster 445, only one specific region of the V4 SSU rDNA contained enough nucleotide polymorphism to allow design of highly specific probes with the optimal themodynamic properties. We therefore decided to use two different probes with a single shift in the nucleotide sequence position (i.e., CHY-445-01 and CHY-445-02).

For negative controls for the specific and general probe we used the reverse complement of each probe named CHY-445-01-RC, CHY-445-02-RC and CHY-NCLC-01-RC (see Key Resources Table) and the hybridization buffer without any probe. All probes were

tested *in silico* using both ARB software (v.6.0.4) and TESTPROBES available on the Silva website (https://www.arb-silva.de) [48]. The six oligonucleotide probes were purchased from biomers.net (Germany) and were labeled at 5′ end with horseradish peroxidase (HRP).

Fluorescent *in situ* hybridization

For *in situ* hybridization, we followed the protocol described by Chambouvet et al. 2008 [18]. This method is outlined as follows: filter samples prepared for FISH were incubated with 3 μL of probes (10 pmol L^{-1}) and 27 μL of hybridization buffer (HB) that include 35% (v/v) formamide, 0.9 M NaCl, 20 mM TrisBase pH = 7.5, 0.01% SDS (sodium dodecyl sulfate, Sigma-Aldrich, UK) and 2% blocking reagent. Samples were hybridized for 12 h at 35°C or 42°C depending of the probe used (see Key Resources Table) before washing twice at 46°C during 20 min in a washing buffer (56 mM NaCl, 5 mM EDTA, 0.01% SDS, 20 mM Tris HCl pH = 7.5). Filter samples were then equilibrated for 15 min at room temperature in the dark in TNT buffer (100 mM Tris-HCl pH = 7.5, 150 mM NaCl, 0.05% (v/v) Tween 20 (Sigma Aldrich, UK)). Each filter was then transferred onto a new slide before adding 10 μl of TSA mix (TSA™ Fluorescein System, Perkin Elmer, UK) per filter piece (2 μl of FT, 50 μl of amplification diluent and 50μl of 40% dextran sulfate) and incubated for 30 min at room temperature in the dark. To remove excess TSA amplification, samples were incubated twice at 55°C for 20 min in TNT buffer. Filters were then washed twice in sterile water and left dry at room temperature. Finally, filter samples were mounted between a slide and a cover glass using an anti-fade mounting solution AF1 (Citifluor™, Electron Microscopy Science, USA) previously mixed with DNA counterstaining, propidium iodide (final concentration of 10 μg/ ml^{-1}). Counts were performed with a Zeiss Observer Z1. Each picture was obtained from a single image extract from a Z stack using Zeiss Observer Z1 epifluorescence microscope equipped with a 3D module VivaTome, a laser excitation light and a camera AxioCam MR. All FISH experiments were conducted in triplicate for each sample type and each experimental condition.

To detect flagellum structures, we used antibodies and the protocol reported in [35] to identify major tubulin cytoskeleton of flagella. Briefly to ascertain presence of a flagellum, sections of filter that were subjected to TSA-FISH hybridization were re-permeabilized with 0.1% v/v nonidet P-40 in PBS (10 mM Na_2HPO_4, 2 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.2), blocked with 1% w/v bovine serum albumin in PBS then incubated for 1 h with the TAT1 monoclonal antibody [63] against α-tubulin, followed by fluorescein isothiocyanate (FITC)-conjugated goat anti-mouse immunoglobulins (Jackson ImmunoResearch/Stratech). The antibody was a gift from Professor Keith Gull's lab at the university of Oxford. Across all samples, candidate flagella seemed dissociated from cells, this was specifically apparent in the dinoflagellates, where the flagella seemed sheared off, indicating that the fixation and dehydration steps were too rough to perform this cellular structure assay. Therefore we could not reliably assess NCLC1 cells for candidate flagella.

To detect cellulosic and/or chitin cell wall structures, we stained with Calcofluor White (1% final concentration, Sigma-Aldrich, USA) using the protocol reported in [35]. We also used this protocol to check for NCLC1 associations with additional cellulosic and/or chitin containing host cells such as dinoflagellates, which were also present in the environmental samples (Table S1). This check was conducted to rule out additional host associations among cells damaged during the process of cell sampling and FISH microscopy preparation, which could potentially limit our ability to identify host cell morphology.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical testing of NCLC1-diatom associations

In order to test whether each observed NCLC1-diatom interaction was significantly different from an incidental interaction due to filtering artifacts we used a series of 1-sided binomial tests (alpha = 0.05) using a binomial distribution with a minimal theta (0.01). In other words, each observed interaction was treated as a 'success' with the total number of NCLC1 interactions observed (i.e., the row sum) as the total number of trials. In order to control for the multiple comparisons a Bonferroni correction was applied to the results. The boxplots were generated to show the distribution of different interactions by summing the replicates per slide. Plots were ordered by their median values and interactions where the null hypothesis of a minimal interaction was rejected were highlighted in blue. This analysis was conducted in the R programming language (RCore2013) using the 'tidyverse' (https://www.tidyverse.org/) set of tools. The code used to perform these calculations (plotting_and_testing_association.r) can be found in the supplemental data, see Zenodo repository: DOI 10.5281/zenodo.2788876.

SSU sequence alignment and phylogenetic tree reconstruction

SSU sequences from 200 taxa were retrieved from previous publications [4, 6]; additional sequence homologs were identified through BLASTN similarity searches (http://blast.ncbi.nlm.nih.gov//blast.ncbi.nlm.nih.gov/). Sequences were then aligned using MAFFT v7.2 iterative refinement method Q-INS-i (https://mafft.cbrc.jp/alignment/server/, [50]). The alignment was subsequently edited with trimAL v4 [51] using the 'gappyout' parameter (''distribution0based algorithm'') and manually masked with seaview v4 [52], resulting in a final alignment of 1542 nucleotide sites (of which 1221 were parsimony informative). The ML phylogenetic tree was reconstructed with IQ-TREE v1.6 [53] using a thorough nearest neighbor interchange search ('-allnni') under the GTR+F+R6 model that is, the general time reversible model with empirical base frequencies and FreeRate model [64] with 6 categories; this substitution model was determined as best fitting the data by ModelFinder [54], as implemented in IQ-TREE, and based on the Akaike Information Criteria. To evaluate node supports, 100 nonparametric bootstrap trees were reconstructed using the same methodology. In addition, to the ML reconstructions, Bayesian inferences were conducted with PHYLOBAYES v3.3 [55] under the GTR-CAT-Γ model

(with 4 discrete categories). Two independent Markov chain Monte Carlo chains (MCMC) were run for 20k generations and sampled every 10 generations with the first 3k discarded as the burnin. Resulting posterior probabilities were mapped onto the ML phylogeny presented.

Ocean Sampling Day 2014 and sequence co-occurrence analysis

SSU rDNA amplicon sequence and contextual data from the Ocean Sampling Day 2014 initiative [41] were retrieved from the Micro B3 project repository (https://github.com/MicroB3-IS/osd-analysis/wiki/Guide-to-OSD-2014-data); microbial community filtering, DNA extraction and Illumina MiSeq sequencing protocols are described in the OSD handbook (http://mb3is.megx.net/osd-files?path=/2014/protocols; see also [46] for an overview of the sampling and sequencing protocol). OSD amplicon sequences were retrieved as pre-processed sequences (technical sequences removed; non-merged paired end reads) and only SSU-V4 sequences obtained following the NPL022 protocol (according to the OSD nomenclature) were analyzed as part of this study. To allow comparison between depth samples, only samples collected from surface waters were kept for analysis. All OSD amplicon were then processed with DADA2 [56]; SSU-V4 sequences were first error corrected and dereplicated, and paired end reads were merged to produce an amplicon sequence variant (ASV) dataset. ASV sequences were then checked for potential chimeras. ASVs were taxonomically classified using the PR^2 v4.10 reference database ([59]; https://github.com/pr2database/pr2database) and the RDP naive Bayesian classifier [65] using a minimum bootstrap confidence of 50, as implemented in DADA2. To produce an SSU-V4 dataset representing only protists sampled during OSD-2014, ASVs classified as multicellular eukaryotes were discarded from the study. Furthermore, to detect spurious SSU-V4 sequences, ASVs were aligned to PR2 representatives using MAFFT v7 [66] and ASVs with poor sequence overlap thresholds, identified with trimAl v1.4 ([51]; '-seqoverlap' lower than 0.8), were discarded. The final SSU-V4 sequence dataset was comprised of 7,766 ASVs totalling 2,552,000 sequences across 145 samples. ASV geographical distributions were plotted using the R package 'ggmap' [57]. ASV sequence correlations were determined for ASV represented by at least 20 sequences (to limit computation time and potential spurious correlations) using the 'sparse correlations for compositional data' algorithm, SparCC [58], with 20 iterations; to identify significant co-occurrences, pseudo p values were computed from resampled correlation matrices (i.e., 100 bootstrap replicates).

DATA AND CODE AVAILABILITY

All data and code are available with DOI's given in the methods section. Specifically, physical and chemical parameters of the water column obtained using CTD ocean water sampling are available at http://biomarks.eu/ctd007 (and replicated here at figshare DOI: 10.6084/m9.figshare.9821936). The phylogenetic tree file, masked and unmasked SSU rDNA alignments are available at Zenodo repository: DOI 10.5281/zenodo.2788876. All sequence data used were derived from the NCBI 'GenBank' database and accession numbers are provided in Figure. 1B. The R code used to test statistical association in the FISH data are available at Zenodo repository: DOI 10.5281/zenodo.2788876.