environmental microbiology

Environmental Microbiology (2015) doi:10.1111/1462-2920.13039

Global patterns of pelagic dinoflagellate diversity across protist size classes unveiled by metabarcoding

Noan Le Bescot,[1,2] Frédéric Mahé,[1,2,3] Stéphane Audic,[1,2] Céline Dimier,[1,2] Marie-José Garet,[1,2] Julie Poulain,[4] Patrick Wincker,[4,5,6] Colomban de Vargas,[1,2] and Raffaele Siano[7]*

[1]*EPEP – Evolution of Protists and Pelagic Ecosystems, Centre National de la Recherche Scientifique (CNRS) UMR 7144 Station Biologique de Roscoff, 29680 Roscoff, France.*

[2]*Station Biologique de Roscoff, Sorbonne Universités, UPMC Univ Paris 06, UMR7144, 29680 Roscoff, France.*

[3]*Department of Ecology, Technische Universität Kaiserslautern, Erwin-Schrödinger Str. 14, D-67663 Kaiserslautern, Germany.*

[4]*Centre National de Séquençage, CEA-Institut de Génomique, GENOSCOPE, Evry Cedex, France.*

[5]*Université d'Evry, UMR 8030, rue Gaston Crémieux CP5706, 91057 Evry Cedex, France.*

[6]*Centre National de la Recherche Scientifique (CNRS), UMR 8030, rue Gaston Crémieux CP5706, 91057 Evry Cedex, France.*

[7]*Ifremer, Dyneco PELAGOS, BP 70, 29280 Plouzané, France.*

Summary

Dinoflagellates (Alveolata) are one of the ecologically most important groups of modern phytoplankton. Their biological complexity makes assessment of their global diversity and community structure difficult. We used massive V9 18S rDNA sequencing from 106 size-fractionated plankton communities collected across the world's surface oceans during the Tara Oceans expedition (2009–2012) to assess patterns of pelagic dinoflagellate diversity and community structuring over global taxonomic and ecological scales. Our data and analyses suggest that dinoflagellate diversity has been largely underestimated, representing overall ~1/2 of protistan rDNA metabarcode richness assigned at ≥ 90% to a reference sequence in the world's surface oceans. Dinoflagellate metabarcode diversity and abundance display regular patterns across the global scale, with different order-level taxonomic compositions across organismal size fractions. While the pico to nano-planktonic communities are composed of an extreme diversity of metabarcodes assigned to Gymnodiniales or are simply undetermined, most micro-dinoflagellate metabarcodes relate to the well-referenced Gonyaulacales and Peridiniales orders, and a lower abundance and diversity of essentially symbiotic Peridiniales is unveiled in the meso-plankton. Our analyses could help future development of biogeochemical models of pelagic systems integrating the separation of dinoflagellates into functional groups according to plankton size classes.

Received 16 October, 2014; accepted 25 August, 2015. *For correspondence. E-mail raffaele.siano@ifremer.fr; Tel. 00 33 2 98 22 42 04; Fax 0033 2 98 22 45 48.

Introduction

Dinoflagellates are flagellated protists belonging to the eukaryotic super-group Alveolata, and form one of the most diverse lineages of modern phytoplankton (Adl *et al.*, 2005; Guiry, 2012). They display outstanding complexity in terms of life mode (autotrophic, mixotrophic, heterotrophic, parasitic, mutualistic), life cycle (e.g. cyst formation), ecology (pelagic and/or benthic in both marine and freshwater habitats) and capacity for adaptation to environmental changes (Schnepf and Elbrächter, 1992; Jeong, 1999; Gómez, 2012a,b). They have developed significant variability in morphology, pigment composition and photosynthetic activity over a large spectrum of cell size. In global pelagic ecosystems, they are part of the so-called pico-nano-, micro- and meso-planktonic size fractions and are responsible for a significant part of primary production (Gaines and Elbrächter, 1987). Some species produce massive blooms well known for the red, green or brownish water discolorations they induce, and can also produce toxins poisonous to fish and humans once accumulated in shellfish (Faust and Gulledge, 2002). If basic abiotic factors, such as turbulence or nutrient availability, can select dominant life forms (Smayda and Reynolds, 2003), the dominant species or genera within communities of selected taxonomic groups are hardly predictable (Not *et al.*, 2012).

Genetic and morphological data allow the distinction of nine clear-cut major dinoflagellate orders: Gonyaulacales, Peridiniales, Gymnodiniales, Suessiales, Prorocentrales,

Dinophysiales, Phytodiniales, Noctilucales and Pyrocystales (Not *et al*., 2012). The phylogenetic position of Lophodiniales and Thoracosphaerales is uncertain and they may actually belong to Peridiniales (Not *et al*., 2012). The orders Gonyaulacales and Peridiniales are characterized by the presence of five latitudinal, one cingular and one sulcal series of cellulose-like thecal plates within the cortical alveoli of the cell covering (amphiesma). The Prorocentrales and the Dinophysiales instead share the division of the theca into lateral halves joined by a sagittal suture. The Gymnodiniales, or unarmoured dinoflagellates, are clearly a polyphyletic group distinguished by the absence of thecal plates within the cortical alveoli. The Phytodiniales are a group of poorly understood genera characterized by a life-cycle shift from a coccoid cell or continuous-walled colonial stage to a vegetative stage. Similar life cycle shifts have been observed in genera of the order Suessiales that includes essentially symbiotic species. The Noctilucales are an early-diverging order containing aberrant dinoflagellates characterized by a highly motile ventral tentacle, which is missing in typical dinoflagellates and other alveolates (Not *et al*., 2012). At lower taxonomic levels, various species concepts had a strong influence on the number of dinoflagellate species and their biogeography (Lundholm and Moestrup, 2006). Gómez (2012a,b; 2013) recognized 2294 dinokaryotic dinoflagellates (including Noctilucales) belonging to 238 genera, while dinoflagellates *sensu lato* comprised 2377 species belonging to 259 genera with 1555 free-living marine morpho-species. In many cases, classical morpho-species can be further split into cryptic or pseudo-cryptic species detected by DNA sequencing-based methods. One extreme example is the single marine genus *Symbiodinium* that split into order-level genetic clades when analysed with molecular tools (Rowan and Powers, 1992; Coffroth and Santos, 2005). Considering also that the smallest, benthic and/or symbiotic *sensu lato* dinoflagellate diversity is still largely uncharacterized (Moreira and Lopez-Garcia, 2002; Coffroth and Santos, 2005; Worden, 2006), the real number of dinoflagellate species may well be significantly bigger than the number of currently morphologically recognized taxa.

Dinoflagellate ecology and diversity are still largely based on light microscopy observations. Besides being partly subjective, light microscopy does not allow species discrimination in groups that lack clear morphological features, especially in the pico- (< 2 μm) and nano- (< 20 μm) plankton. In the last decade, the sequencing of environmental clone libraries of polymerase chain reaction (PCR)-amplified ribosomal genes highlighted the presence of many novel dinoflagellates within pico-nanoplankton assemblages (Moon-van der Staay *et al*., 2001). Nevertheless, the taxonomic and ecological significance of small-sized dinoflagellates remains hidden behind a lack of data and taxonomic information (Siano *et al*., 2009; 2010; Siokou-Frangou *et al*., 2010).

The overarching goal of this study was to re-explore, using an objective and semi-quantitative metabarcoding approach (Taberlet *et al*., 2012; de Vargas *et al*., 2015), the overall biodiversity and community structure of pelagic dinoflagellates on global taxonomical and biogeographical scales. Our molecular ecology protocol was thus constrained by the need to target all dinoflagellate lineages of the surface oceans. In dinoflagellates, the classical DNA metabarcodes COI and *Cob* (Cytochrome Oxidase I and b) were tested on a wide range of taxa; however, none of these mitochondrial markers could be PCR amplified from all of the dinoflagellate strains or could resolve common ambiguous genera at the species level (Lin *et al*., 2009; Stern *et al*., 2010). Interestingly, COI was successful in identifying ~70% of cultured species, but when applied to environmental samples it revealed a much higher diversity of uncharacterized species (Stern *et al*., 2010). Although *Cob*-based metabarcoding suffers from a significant lack of reference sequences from well-identified strains (Lin *et al*., 2009), this genetic marker has been used recently to assess the richness of benthic dinoflagellates, providing the largest set of dinoflagellate *Cob* gene sequences (Kohli *et al*., 2013). Generally, a lack of reference sequences hampers the use of mitochondrial genes for accurate appreciation of dinoflagellate communities. Relatively fast-evolving nuclear ribosomal DNA loci (Large SubUnit (LSU), Internal Transcribed Spacers (ITS) rDNA) (Orr *et al*., 2012) are generally preferred to the commonly used protistan 18S rDNA marker or to mitochondrial genes to assess dinoflagellate biodiversity (Edvardsen *et al*., 2003; Litaker *et al*., 2007). ITS-1 and ITS-2 rDNA metabarcodes were successful in identifying cultured species, and also revealed hidden diversity in strains from culture collections (Stern *et al*., 2012). However, their use for global metabarcoding is biased essentially by the difficulty of PCR priming equally across all dinoflagellate lineages, and to a lesser degree by the presence of paralogues and the potential for unidentifiable chimeras (Stern *et al*., 2012). Despite its relatively low variability, 18S rDNA remains advantageous for a comprehensive, first-order assessment of dinoflagellate biodiversity within a broader context of pan-eukaryotes diversity (Pawlowski *et al*., 2012), thanks to its ease in targeting all dinoflagellate together with most other protistan lineages, and its relatively large representation in reference databases (Guillou *et al*., 2013) that allows taxonomic annotation of most eukaryotic environmental metabarcodes. Here, we used the V9 of the 18S rDNA to assess overall dinoflagellate rDNA diversity and abundance at a global scale because of its relatively short length of 130bp, which accommodates the requirements of massive high-throughput HiSeq Illumina sequencing of the 106 protistan communities analysed

herein close to saturation. Our objective was not to characterize environmental dinoflagellate communities at species level, but rather to seek for general patterns of pelagic dinoflagellate biodiversity and community structure across global spatial and time scales, and in particular within relatively underexplored eukaryotic size classes.

Results

Dinoflagellates versus other protists

A total number of 83 860 V9 rDNA dinoflagellate metabarcodes that were present in > 10 copies and with a genetic identity to a dinoflagellate reference sequence of ≥ 90% were retrieved from the world surface oceans (33 stations analysed, four samples per station, one per size fraction), with an average sequence length of 130 ± 1 bp (range: 79–143 bp; mode = 130 bp), allowing taxonomic discrimination between the species and genus/family levels depending on the dinoflagellate lineages. According to our data set (compiled using stringent abundance and %ID thresholds), overall dinoflagellate metabarcode richness accounted for ~49% (range: 38–61%) of total protistan metabarcode richness, dinoflagellate metabarcodes representing > 50% of the total protistan metabarcode diversity in Tara Oceans stations: 18, 45, 76, 98 and 100 (Fig. 1A). In terms of abundance, dinoflagellate metabarcodes accounted on average for ~40% (range: 18–67%) of the total protistan metabarcode abundance, sometimes representing most of the protistan richness even if they only represented < 25% of total abundance (stations: 25, 45, 98 and 102) (Fig. 1B). In terms of global biogeography, the relative richness of dinoflagellate metabarcodes was strikingly stable across the oceans, despite obvious variations in dinoflagellate metabarcode relative abundance. While metabarcode richness ranged from 45–61% and 38–57% in the Mediterranean Sea and Pacific Ocean stations, respectively, their abundance ranged from 20% to 67% and 20% to 49% (Fig. 1).

Dinoflagellate rDNA abundance and richness over the whole protistan community varied among size fractions (Fig. 2). Generally higher absolute metabarcode richness and abundances were observed in the pico-nano-plankton (0.8–5 μm) and then in the nano-plankton (5–20 μm) size fractions, with the exception of stations 64, 70, 78, 82, 100 and 109 where abundances were higher in the nano-plankton. Higher relative richness values were observed in the intermediate sizes fractions (~64% and ~59% for the 5–20 μm and 20–180 μm size fractions) compared with the smallest (~47% – 0.8–5 μm) and largest (~24% – 180–2000 μm) ones, with notable exceptions (stations 18, 45, 67, 70). Interestingly, the abundance of metabarcodes present exclusively in a single size fraction was higher in the smallest size fraction (average: 22 ± 3%). Station 45 was the only one characterized by a higher proportion of size fraction unique metabarcodes in the microplankton (25%) (Fig. 2).

Dinoflagellate community characterization

The abundance and richness of dinoflagellate metabarcodes assignable to orders varied significantly across size fractions. Some orders were represented at most stations and typically dominate in a given size fraction (Gymnodiniales in pico-nano-plankton, Gonyaulacales in micro-plankton and Peridiniales in meso-plankton), while others (Lophodiniales, Phytodiniales, Pyrocystales, Thoracospaerales) contributed much less to overall dinoflagellate communities. Undetermined metabarcodes (sequences that cannot be unambiguously assigned to any dinoflagellate order on the basis of V9 rDNA annotation) spanned from 7% to 49% of the dinoflagellate communities, with an increasing ratio in the smaller size fractions (Fig. S1).

Dinoflagellates in the pico-nano-plankton (0.8–5 μm)

In the pico-nano-plankton (30 samples corresponding to the 0.8–5 μm size fraction) undetermined metabarcodes represented on average 45% (range: 26–59%) of the total dinoflagellate metabarcode abundance and 41% (range: 36–49%) of their richness (Fig. 3A). Assignation conflicts leading to undetermined dinoflagellates occurred mainly among two super clusters, the Gymnodiniales/Peridiniales/Prorocentrales and Gymnodiniales/Peridiniales groups (Fig. S2). Based on metabarcodes taxonomically assigned to orders, we observed that Gymnodiniales were more important than other orders in terms of both abundance (average 30%, range: 8–41%) and richness (average: 31%, range: 19–36%). Across stations, dominant Gymnodiniales metabarcodes were related to the genera *Karlodinium* (average: 15%) and *Gyrodinium* (average: 11%). The great majority of the hundred most abundant dinoflagellate metabarcodes (accounting for 70% of total dinoflagellate metabarcodes abundance in the global data set) were significantly more abundant in the 0.8–5 μm than in other size fractions, only 11 of them being, on the contrary, more abundant in the largest size fraction (Fig. S3).

At stations sampled around South Africa within a relatively short time frame (2 weeks between stations 65 and 66, and stations 67 and 68), the metabarcode richness levels of Gymnodiniales had progressively increased from the Indian Ocean (station 65) to the southern Atlantic (stations 66, 68), probably reflecting a spatial community composition change. Within this area, station 67 stood as an exception with only 8% of metabarcodes representing Gymnodiniales, while the Prorocentrales and Peridiniales

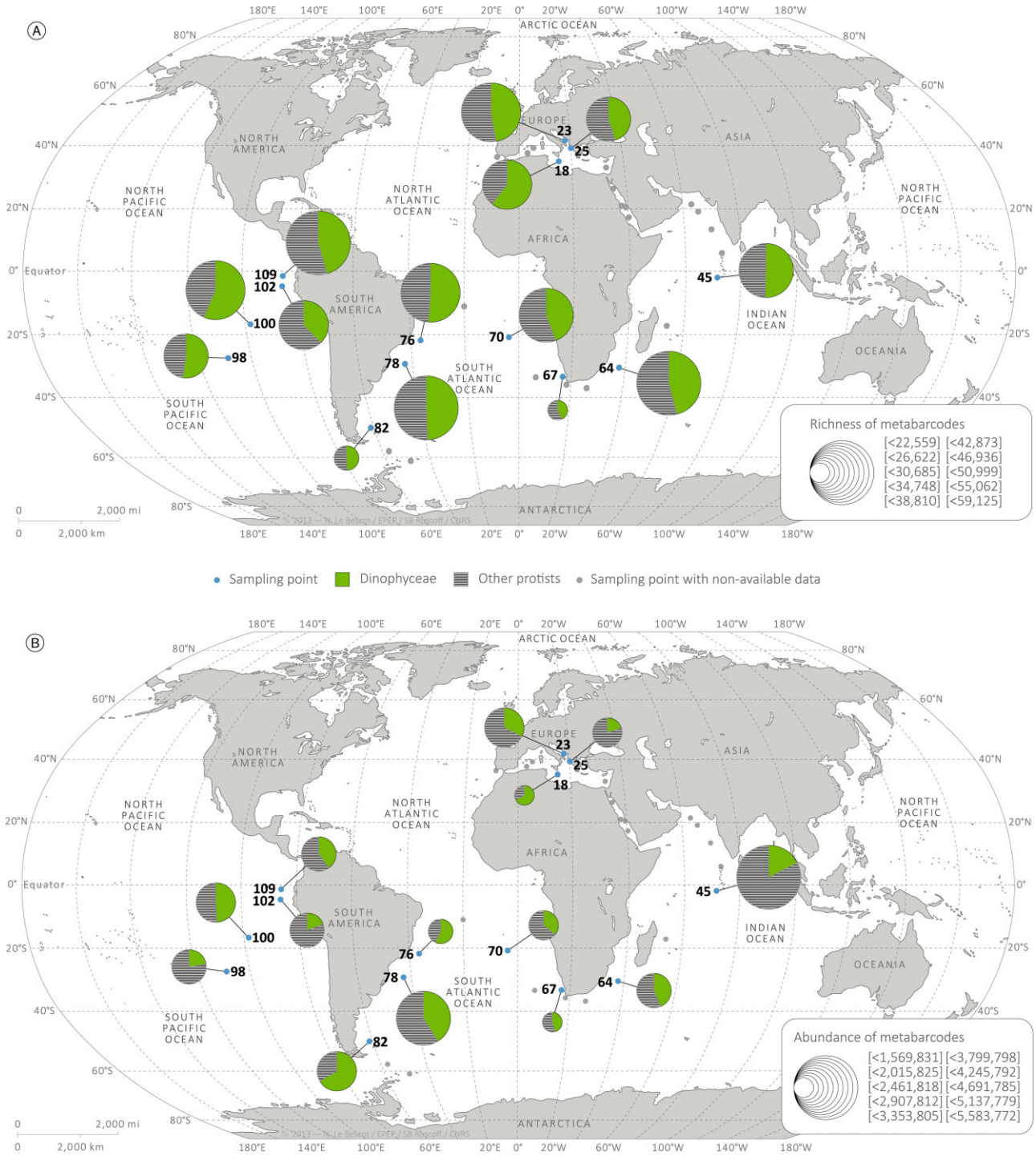

Fig. 1. Metabarcode dinoflagellate richness (A) and abundance (B) (green) over total protist community (grey with stripes). Pie chart sizes are proportional to total metabarcode numbers analysed per sample.

accounted for 21% and 14% of total dinoflagellate diversity. The Peridiniales were the second most important groups among pico-nano-dinoflagellates in terms of both abundance (average 11%, range: 8–19%) and richness (average 14%, range: 12–20%). Uncultured metabarcodes (environmental metabarcodes that do not match any specific order but clearly classified as dinoflagellate) represented > 15% of the total dinoflagellate metabarcode abundance at stations 65, 68 and 78. The richness of uncultured metabarcodes was very similar among these stations, but uncultured metabarcodes abundance almost double at station 78 (Fig. 3A).

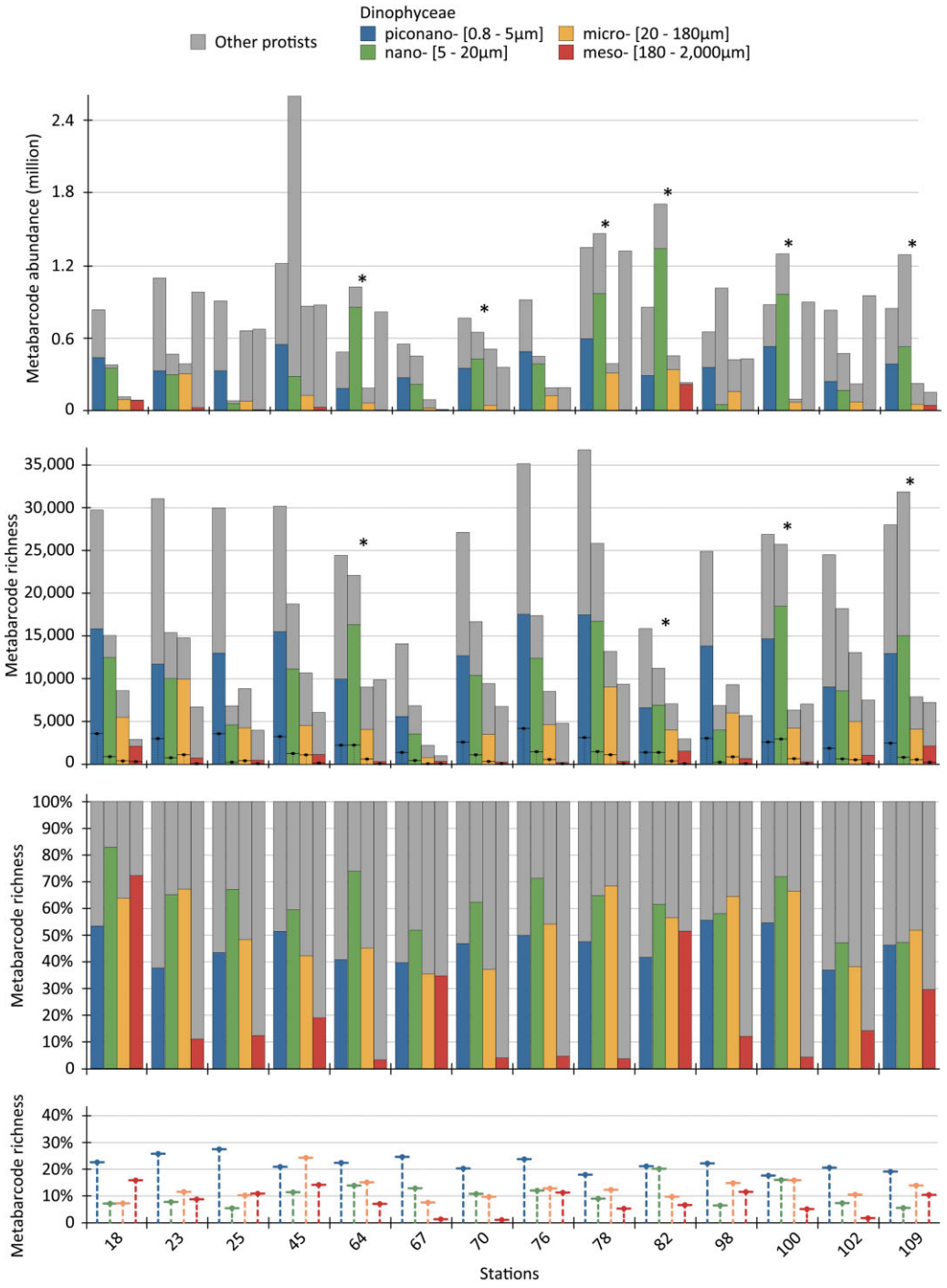

Fig. 2. Abundances and richness of dinoflagellate metabarcodes in the pico-nano- (0.8–5 μm), nano- (5–20 μm), micro- (20–180 μm) and meso- (180–2000 μm) plankton size classes. The asterisk indicates exceptional stations where metabarcode abundance is higher in the nanoplankton. Dashed lines represent the number of metabarcodes unique to their size class.

Dinoflagellates in the nano-plankton (5–20 μm)

In nano-planktonic samples, the relative proportion of dinoflagellate undetermined metabarcodes was still very high (average of 35%). On the other hand, order-level assigned metabarcodes were relatively variable in terms of both richness and abundance, and a unique biodiversity pattern does not emerge across the 15 analysed stations. A large number of Gymnodiniales metabarcodes was present in 10 stations over the 15 analysed. At station

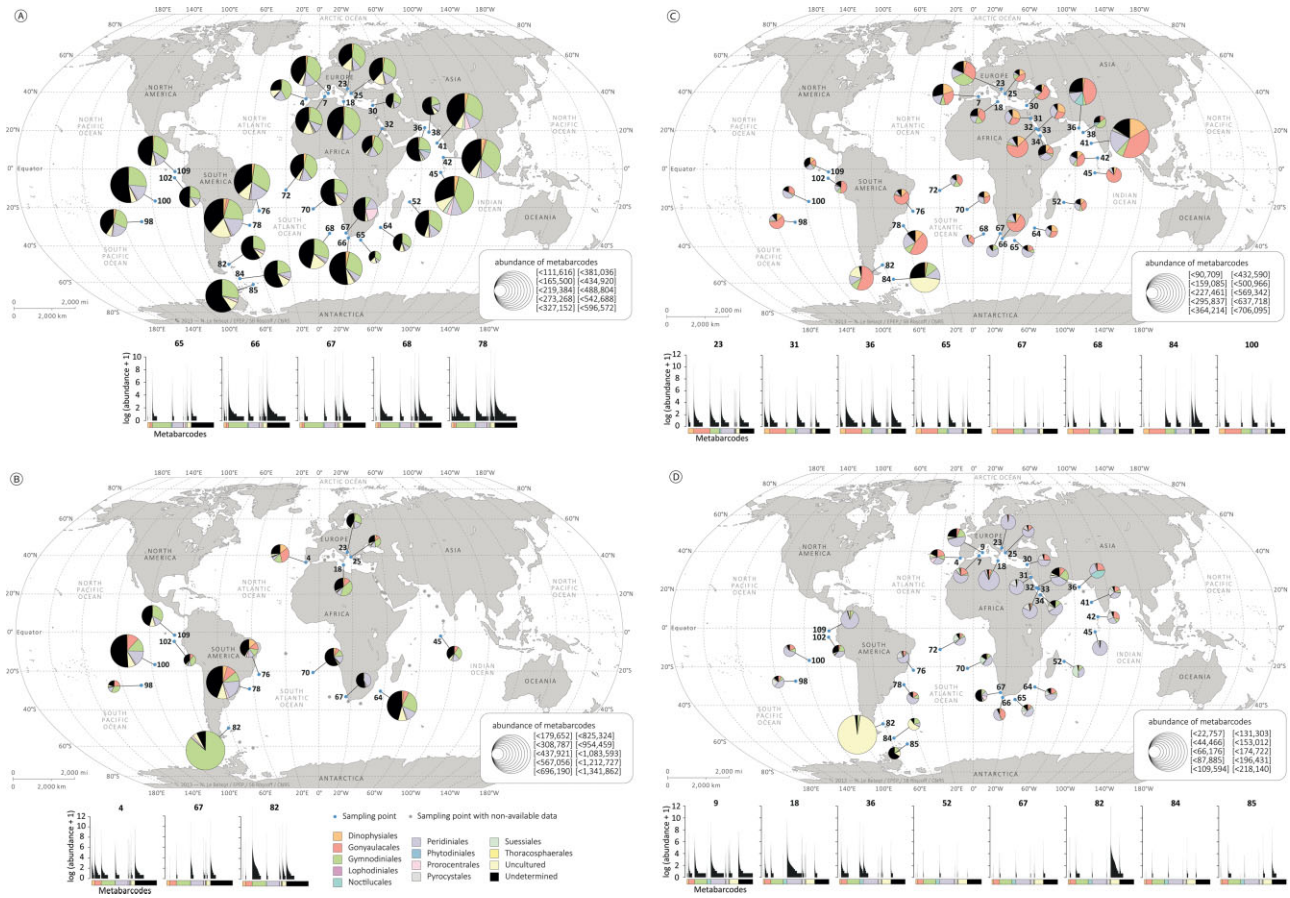

Fig. 3. Dinoflagellate order-based community structure assessed on metabarcode abundance (maps) and log +1 metabarcode abundances (selected stations, graphs) in the pico-nano- (0.8–5 μm) (A), nano- (5–20 μm) (B), micro (20–180 μm) (C) and meso- (180–2,000 μm) (D) plankton. In maps, pie chart sizes are proportional to total metabarcode numbers analysed per sample.

4, metabarcodes from the Gymnodiniales dominated in terms of richness (23%) but Gonyaulacales metabarcodes were more abundant (32%). Station 67 was characterized by a high Peridiniales diversity (24%) and very high relative abundance (39%). Station 82 was largely dominated by Gymnodiniales metabarcodes (86% and 38% of the metabarcode abundance and richness, respectively), and it contained a relatively low number of undetermined metabarcodes (Fig. 3B).

Dinoflagellates in the micro-plankton (20–180 μm)

In the micro-plankton, the relative proportion of undetermined metabarcodes (average of 21%) was significantly less than in smaller size fractions. Micro-dinoflagellate communities were dominated by Gonyaulacales metabarcode abundance and richness in respectively 23 and 19 out of the 30 analysed stations. Gonyaulacales metabarcodes represented on average 37% (range: 18–80%) and 26% (range: 22–48%) of total dinoflagellate metabarcode abundance and richness respectively. Peridiniales metabarcodes were the second most important dinoflagellate order in micro-plankton (abundance average: 23%, range: 21–69%, richness average: 25%, range: 21–48%). At stations 31, 33 and 64, Dinophysiales metabarcodes accounted for up to 25%, and at stations 18, 23, 38, 82 and 84, Gymnodiniales were also important, contributing to up to 20%. In the Red Sea and Indian Ocean, Gonyaulacales, Peridiniales and Dinophysiales accounted for equal parts of the total metabarcodes abundance, except for station 32 and 45. Stations around South Africa (65, 67, 68) were clearly characterized by Peridiniales metabarcodes, except for station 66 which was instead over dominated by Gonyaulacales metabarcodes (67%). Note that at station 84 in the Southern Ocean, undetermined metabarcodes represented 47% of the total metabarcodes abundance (Fig. 3C). At finer taxonomic resolution, > 50% of the Gonyaucales metabarcodes were assigned to the genus *Ceratium* in 20 out of the 30 stations analysed. On average > 75% of the Peridiniales metabarcodes were related to *Protoperidinium*, and Dinophysiales metabarcodes were dominated (> 60%) by *Ornithocercus* at stations 36, 45, 52 and 109 (Table 1).

Table 1. Frequency (%) of major dinoflagellate genera metabarcodes in fractions 20–180 μm and 180–2000 μm at all sampled station (4–109).

	Size fraction 20–180 μm (%)				Size fraction 180–2000 μm (%)		
Order	Gonyaulacales		Peridiniales	Dinophysiales	Peridiniales		
Genus	*Ceratium*	*Gonyaulax*	*Protoperidinium*	*Ornithocercus*	*Blastodinium*	*Brandtodinium*	*Protoperidinium*
4	NA	NA	NA	NA	69	18	2
7	55	22	64	25	93	5	1
9	NA	NA	NA	NA	78	8	0
18	83	11	67	13	85	15	0
23	58	15	61	3	2	98	0
25	57	39	55	14	81	16	0
30	59	9	59	21	81	13	5
31	47	30	91	7	94	1	0
32	91	7	45	8	23	55	3
33	49	30	91	7	93	1	3
34	12	47	83	6	82	1	1
36	7	64	79	63	36	63	1
38	19	65	39	6	NA	NA	NA
41	25	14	55	42	75	21	2
42	57	18	68	37	80	7	4
45	77	7	74	62	99	1	0
52	59	19	59	64	20	74	1
64	67	22	82	11	1	95	3
65	73	14	87	10	13	72	10
66	11	74	89	5	77	2	17
67	34	9	99	0	5	0	16
68	81	11	96	1	NA	NA	NA
70	76	17	87	3	22	63	6
72	73	17	86	40	67	24	6
76	78	12	58	20	32	66	1
78	33	39	79	8	1	96	1
82	99	1	89	0	0	0	82
84	75	0	87	0	3	0	67
85	NA	NA	NA	NA	1	1	55
98	60	14	57	45	21	75	1
100	59	16	88	14	29	67	2
102	89	5	70	10	44	26	10
109	15	58	79	67	95	4	0
Average	56 ± 26	24 ± 20	74 ± 16	20 ± 21	49 ± 36	32 ± 34	10 ± 20

NA = data not available.

Dinoflagellate in the meso-plankton (180–2000 μm)

In the meso-planktonic size fraction, Peridiniales metabarcodes were the most abundant (average 48%, range: 15–95%) and diverse (average 37%, range: 17–66%) in, respectively, 26 and 27 stations out of the 31 analysed. Again the Southern Ocean (stations 82 and 84) stood out, characterized by a majority of uncultured metabarcodes, while station 85 in the Southern Ocean and station 67 in coastal upwelling offshore of South Africa Cape Town were characterized by a higher number of undetermined metabarcodes. At station 67, the taxonomically assigned part of dinoflagellate diversity was made up of relatively equal fractions of Peridiniales (abundance 15% and richness 17%), Prorocentrales (abundance 14% and richness 10%) and Gymnodiniales (abundance 11% and richness 19%). Station 36 was characterized by a high abundance (42%) of Noctilucales metabarcodes, all strictly identical to the *Noctiluca scintillans* reference sequence (Fig. 3D).

Dominant Peridiniales metabarcodes were assigned to the genera *Blastodinium* and *Brandtodinium*, which include, respectively, parasitic and symbiotic species. At station 52, Suessiales metabarcodes (46%) were all assigned (100% identity) to the *Pelagodinium béii* reference sequences, a well-known symbiotic taxon of planktonic foraminifera. Within the community of symbiotic dinoflagellate genera, *Blastodinium* metabarcodes represented > 70% of total metabarcode abundance within the stations located north of latitude 10°S, while *Brandtodinium* metabarcodes were more important south of 10°S (Fig. 4, Table 1). In Southern Atlantic Ocean stations 82, 84 and 85, Peridiniales contained essentially Peridiniales contained essentially metabarcodes assigned to the genus *Protoperidinium* (Table 1).

Discussion

Initial barcoding studies on dinoflagellates aimed at assessing the best genetic marker to characterize

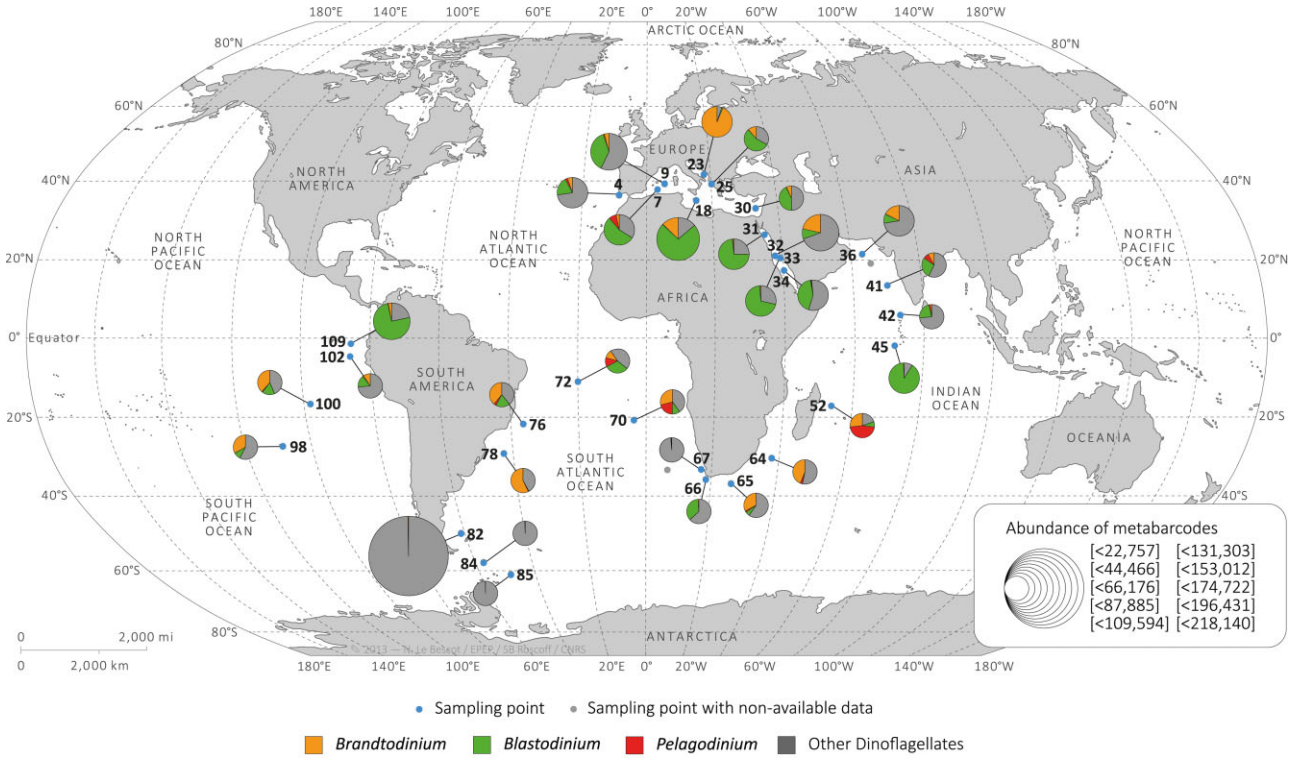

Fig. 4. Dinoflagellate genus-based community structures assessed on metabarcode abundances in fraction 180–2000 μm. Represented genera are characterized by mutualistic (*Brandtodinium*, *Pelagodinium*) or parasitic (*Blastodinium*) associations with other organisms. Pie chart sizes are proportional to total metabarcode numbers analysed per sample.

species and intraspecific diversity (Lin *et al.*, 2009; Stern *et al.*, 2010; 2012). Here, we used a broad metabarcoding approach to assess dinoflagellate overall biodiversity in environmental samples among protistan communities. We continued the exploration of this approach for the study of marine protist biodiversity so far used to evaluate either the diversity of a specific group, such as the uncultured MAST (MArine STramenopiles) (Logares *et al.*, 2012), cercozoan amoebae (Berney *et al.*, 2013), diatoms (Nanjappa *et al.*, 2014) and radiolarians (Decelle *et al.*, 2014), or specific community patters of the rare (Logares *et al.*, 2014) and benthic biosphere (Kohli *et al.*, 2013; Massana *et al.*, 2015). We used a taxonomic-based approach to link classical knowledge of dinoflagellate biodiversity to next-generation sequencing-based analyses. Our ultimate goal was to identify broad patterns of dinoflagellate genetic diversity and community structure based on rDNA metabarcode abundance data in the world ocean across protist size classes. We showed the importance of dinoflagellate rDNA diversity and abundance within total protist communities, and we demonstrated that dinoflagellate communities display strikingly stable but distinct order-level biodiversity patterns across size fractions, independently of space and time.

Overall, we analysed ~84 000 distinct dinoflagellate metabarcodes obtained from 106 total protistan communities from four size fractions and 33 world surface oceans stations. Our objective to compare dinoflagellate diversity to total protistan genetic diversity and to target all dinoflagellates from a relatively large number of samples constrained the choice of molecular marker and sequencing technology. At the time of samples processing, the Illumina GAIIx technology and the V9 rDNA metabarcoding system was the best possible methodological combination to provide both the necessary sequencing power and a valuable taxonomic assessment of total eukaryote diversity based on relatively comprehensive reference database (Guillou *et al.*, 2013; de Vargas *et al.*, 2015). The V9 rDNA barcode presents a combination of advantages for addressing general questions of eukaryotic biodiversity over extensive taxonomic and ecological scales: (i) it is universally conserved in length (130 ± 4bp) and simple in secondary structure, thus allowing relatively unbiased PCR amplification across eukaryotic lineages followed by Illumina sequencing, (ii) it includes both stable and highly variable nucleotide positions over evolutionary time frames, allowing discrimination of taxa over a significant phylogenetic depth, (iii) it is extensively represented in public reference

databases across the eukaryotic tree of life, allowing taxonomic assignment among all known eukaryotic lineages (de Vargas *et al*., 2015). The relevance of eight 18 rDNA hypervariable regions for dinoflagellate barcoding was evaluated using 77 cultured strains, confirming that intra-species genetic variation is low compared with interspecies divergence (Ki, 2011). The V9 loop is considered the third most informative 18S rDNA region to infer dinoflagellate diversity (Ki, 2011). Used as a barcode, the V9 rDNA showed a particularly low-resolution power preventing protist biodiversity assessment at species and often-generic levels (de Vargas *et al*., 2015), and it occurs in multiple copies in single cell genome. This last problem challenges the use of rDNA barcodes to assess protistan diversity and relative abundance. Indeed, a broad range of SSU (Small SubUnit) rDNA copy numbers per cell has been estimated in different protist groups, such as diatoms (Galluzzi *et al*., 2004; Godhe *et al*., 2008), dinoflagellates (Godhe *et al*., 2008), ciliates (Gong *et al*., 2013) and a set of microalgal cultured strains (Zhu *et al*., 2005). However, rDNA copy numbers are relatively stable within a given taxon, and they have been shown to correlate positively to cell lengths and even better to bio-volumes (Zhu *et al*., 2005; Godhe *et al*., 2008, and see de Vargas *et al*., 2015 for a synthesis). Godhe and colleagues (2008) demonstrated a linear correlation between SSU rDNA copy numbers assessed by quantitative PCR and light measured bio-volumes for four diatom and nine dinoflagellate species in culture. Thus, the metabarcodes abundance data obtained in this study should not be interpreted as number of individuals of a particular taxon, but as relative numbers of ribosomal genes, independently of the organismal size fraction, are rough proxy for a specific taxon bio-volume (de Vargas *et al*., 2015), a key ecological feature typically overlooked in classical dinoflagellate studies based on cell counting under the microscope.

Global dinoflagellate diversity and abundance

Our study is the first global metabarcoding attempt to characterize pelagic dinoflagellates as a whole across a wide variety of spatial and temporal scales and through different planktonic size fractions. Our V9 rDNA diversity data, analysed with conservative thresholds (> 10 copies and > 90% ID to the nearest reference sequence) indicate that dinoflagellates represent about half of total, accurately assignable protist richness in the world surface oceans, and that proportion is strikingly stable across the > 30 explored stations (range 38–61%). Even when using the entire Tara Oceans V9 rDNA data set without any of our cut-offs, keeping all metabarcodes occurring at least three times and in two different samples (de Vargas *et al*., 2015), the dinoflagellates still make up ~28% of the entire eukaryotic ribosomal diversity and are 1.5 times more diverse than all animals (metazooplankton) together. Thus, our data indicate that the biodiversity of pelagic dinoflagellates has been largely underestimated. This knowledge gap will certainly increase if similar metabarcoding surveys were applied to deep oceans, benthic and/or freshwater ecosystems, and if rare taxa were included in the analysis. In contrast to the remarkable stability of dinoflagellate relative richness over space and time, the relative abundance of dinoflagellate metabarcodes displayed more variability, with values ranging from 18–67% of total protist metabarcode abundance (Fig. 1). This variability is likely due to specific environmental and seasonal parameters characterizing each sampled station, with local selection for the growth or decline of a few dinoflagellate species. Overall, the global consistency of dinoflagellate diversity relative to its fluctuating abundance suggests strong community self-organizational rules at fundamental taxo-functional levels, with putative local selection of specific taxa on ecological time scales.

Dinoflagellate community composition across protist size classes

Significant biodiversity and ecological patterns emerge from our analyses of dinoflagellate metabarcode diversity and abundance across the pico-nano-, nano-, micro- and meso-planktonic size fractions. The fractionation of hundreds of litres of seawater through nets and sieves certainly biases to some extent the relative diversity of particularly fragile plankton across organismal size fractions. However, the ratio of dinoflagellate to all other protist sequences is relatively stable within each size fraction, and displays a remarkable pattern with higher seemingly relative richness in the intermediate (5–20 μm; 20–180 μm) size fractions as compared with the smallest (0.8–5 μm) and largest (180–2000 μm) ones (Fig. 2). In addition, between 2% and 28% of dinoflagellate metabarcodes are strictly specific to a single size fraction, providing further proof that our community fractionation was generally successful. The presence of identical metabarcodes across size fractions is expected due to natural cell debris, ontogenic and life-cycling stages involving different sizes, the presence of symbionts *sensu lato* in larger size fractions and the relatively slow rate of dinoflagellate rDNA evolution, meaning that a single metabarcode could represent several taxa of various sizes.

Dinoflagellates in the pico-nano-plankton (0.8–5 μm)

Previous community analyses of pico-plankton were essentially based on clone libraries of PCR products (Diez *et al*., 2001; Lopez-Garcia *et al*., 2001; Moon-van der

Staay *et al*., 2001; Cheung *et al*., 2008) and culturing methods (Massana *et al*., 2004; Balzano *et al*., 2012). Attempts to assess the ecological role of tiny dinoflagellates on the basis of quantitative estimation methods such as cells counts and carbon contribution are rare (e.g. Siokou-Frangou *et al*., 2010; Zingone *et al*., 2010). Despite being not dominant, small (< 10 μm) dinoflagellates can contribute to up to 76% of total carbon production (Zingone *et al*., 2010). Pico-plankton genetic diversity studies mostly emphasize the importance of pico-green algae (Chlorophyta and Prasinophyta), estimating the relative dinoflagellate contribution on average to be 17% of total pico-eukaryotes (Massana *et al*., 2004). 454-pyrosequencing of total eukaryotic V4 18S rDNA PCR products from tropical waters further estimated that dinoflagellates represent about 19% of pico-eukaryotic communities (Cheung *et al*., 2010). Our results based on massive Illumina sequencing of larger volumes of extracted seawater from > 30 world pico-nano-planktonic communities suggest that dinoflagellates represent in fact a relatively stable and very important part of these communities, accounting on average for ~47% and ~43% of the total pico-nano-eukaryotic community rDNA richness and abundance respectively. Moreover, our data show that pico-nano-dinoflagellates are not only key contributors of pico-nano-planktonic communities but are also significantly more diverse and unexplored as compared with those of larger size fractions (Fig. 2 and 3). The relatively higher proportion of dinoflagellate metabarcodes strictly specific to the pico-nano-sized fraction (18–26%, depending on station) indicates that pico-nano-dinoflagellates do not come from debris of larger cell size fractions. The large majority of the 100 more abundant metabarcodes of the whole data set are more abundant in the pico-nano-plankton. Only 11 of them are in higher proportion in the size fractions > 5 μm (Fig. S3). This is explainable either considering that this high number of metabarcodes really belong to the pico-nano-plankton, or that the cells corresponding to those metabarcodes completely or mostly destroyed during the filtration process in all samples, leaving only few traces (if any) in their appropriate size fraction. This second option is hardly believable; therefore, we consider that the potential accumulation in the smallest size fraction of debris from cells larger than 5 μm is not a major source of bias of our analyses. However, it is not completely excluded that in a low proportion cell breakage have cross contaminated size fractions.

The remarkable diversity of pico-nano-dinoflagellates across the world's oceans is characterized by a large proportion of undetermined metabarcodes (average richness and abundance of 41% and 45%, respectively) and Gymnodiniales metabarcodes (average richness and abundance of 31% and 30% respectively). The assignation of most taxonomically determined metabarcodes to the Gymnodiniales should however be interpreted with caution, as this order is paraphyletic and intergeneric relationships are not clear (Orr *et al*., 2012). Peridiniales and Prorocentrales metabarcodes are the next most important contributors to pico-nano-dinoflagellates biodiversity. The fact that assignation conflicts leading to undetermined dinoflagellates metabarcodes concern mostly the Gymnodiniales/Peridiniales/Prorocentrales and Gymnodiniales/Peridiniales super groups (Fig. S2) reinforces our estimation that some of the undetermined barcodes of this size fraction may belong to Gymnodiniales, corroborating our hypothesis that this order is the most abundant and diverse in the pico-nano-plankton size fraction. These conflicts also highlight the limit of our reference database that probably lacks many reference sequences, in particular from small featureless cells, within these three major orders. In fact very few dinoflagellate species < 5 μm have been described to date. The pelagic genera *Karlodinium* and *Gymnodinium* (Gymnodiniales) (Siano *et al*., 2009; Thessen *et al*., 2012), *Azadinium* and *Heterocapsa* (Peridiniales) (Pomroy, 1989; Percopo *et al*., 2013) and *Prorocentrum* (Prorocentrales) (Puigserver and Zingone, 2002) and *Pelagodinium* (Suessiales) (Siano *et al*., 2010) include species of ±10 μm in size. To the best of our knowledge, a single dinoflagellate species, *Prorocentrum nux* Puigserver & Zingone, can be smaller than 3–5 μm (Puigserver and Zingone, 2002), thus truly meeting our definition of pico-nano-plankton. The identification of *Karlodinium* metabarcodes among pico-nano-plankton is somehow coherent with morphological information, whereas the detection of *Gyrodinium* metabarcodes is unexpected since described *Gyrodinium* species are generally >5 μm. Small naked dinoflagellate cells unidentifiable in light microscopy are typically observed in environmental samples but often overlooked in classical phytoplankton surveys (Siokou-Frangou *et al*., 2010), underlying the putative existence of a large biodiversity of truly pico-nano-dinoflagellates taxa and/or small and abundant life stages of taxa belonging to bigger size fractions. As for the great majority of pico-eukaryotic taxa from open oceans, this unveiled diversity, potentially partly heterotrophic, is most likely difficult to cultivate and thus is not available in reference gene database. A similar pattern of extreme and novel pico-planktonic diversity in a classical group of marine phytoplankton that was thought to be morphologically well described was also shown in the haptophytes (Liu *et al*., 2009), and may well be a common features in pelagic protists.

Dinoflagellates in the nano-plankton (5–20 μm)

Metabarcoding data from nano-planktonic communities were available for fewer stations (15) than for other size

fractions (30) due to the challenge in concentrating the relatively dilute biological material within the size range of 5–20 μm from a large volume of seawater. Eukaryotic nano-plankton is arguably the least known component of pelagic ecosystems (Lopez-Garcia *et al*., 2001; de Vargas *et al*., 2015). However, this is the fraction where the relative richness of dinoflagellate to protist metabarcode ratio is the highest (10/15 stations), suggesting a critical and underestimated importance of nano-dinoflagellates. This fraction is also characterized by a relative instability in the abundance of representatives from different orders, when in comparison to the more stable proportions observed in the pico-nano- and micro-/meso-planktonic communities. Nano-dinoflagellate communities are characterized by a mixture of Gymnodiniales, Gonyaulacales, Peridiniales and Dinophysiales as well as by other orders whose relative abundance varies across spatial at temporal scales. The relative lack of analysed samples could partly explain the relatively less homogenous diversity patterns within this size class, and a higher spatio-temporal sampling resolution is needed to understand the ecological and functional role of these highly complex communities.

Dinoflagellates in the micro-plankton (20–180 μm)

The large majority of described, morphologically identifiable (10 μm in size usually represents the lower limit of species detection by light microscopy) and genetically characterized dinoflagellates belong to this size class (Fensome *et al*., 1993; Gómez, 2012a; 2013), making the micro-dinoflagellate communities relatively well known across the world's oceans in terms of both taxonomy and ecology. Our metabarcoding analysis confirms indeed that, except for some stations (7, 23, 36, 84, 9), undetermined and uncultured dinoflagellates metabarcodes account for a relatively low percentage (< 15%) of total micro-planktonic communities. Note that there is a general trend of decreasing frequency of uncultured and/or undetermined dinoflagellate diversity from the smallest to the largest size fractions, matching the fact that species >20 μm are relatively well described, and further supporting the remarkable and novel diversity detected in the pico- and nano-plankton. Among micro-dinoflagellates, the Gonyaulacales and Peridiniales are the major traditional orders including genera that comprise a large number of species (e.g. *Ceratium*, *Protoperidinium*, (Gómez, 2012a; 2013). Our richness and abundance data confirmed that Gonyaulacales is the first, and Peridiniales is the second most important order within world surface oceans micro-dinoflagellate communities. The relative importance of one order over the other can vary across stations, likely depending on local environmental triggers selecting specific taxa.

Dinoflagellate in meso-plankton (180–2000 μm)

The largest described dinoflagellates (< 2 mm) are odd, fragile and relatively rare species belonging to the order Noctilucales, whose phylogenetic position has been recently elucidated (Gómez *et al*., 2010). Apart from *N. scintillans* which is a relatively well-known bloom-forming aberrant dinoflagellate occurring from temperate to subtropical coastal regions (Harrison *et al*., 2011), other Noctilucales (e.g. *Abedinium*, *Spatulodinium*, *Kofoidinium*, *Petalodinium*) are very poorly studied species with aberrant shapes that hampers their identification. The generally low relative abundances of Noctilucales metabarcodes in our metabarcoding data set (average of < 0.1%) confirm the rarity of members of this order in the world's oceans. One exception is Tara Oceans station 36, where a high abundance of *N. scintillans* metabarcodes (42%) was observed, probably corresponding to a bloom of the species.

Other dinoflagellates can be bigger than 180 μm, such as members of both Gonyaulacales (i.e. *Ceratium* spp.) and Pyrocystales (*Pyrocystis* spp.). Although some metabarcodes assignable to these orders were detectable in the meso-plankton (average 10% and 7% for the Gonyaulacales and Pyrocystales, respectively), Peridiniales, metabarcodes were by far the most abundant (average 48%) across the global ocean. Within Peridiniales, some *Protoperidinium* spp. can be larger than 180 μm; however, the highest number of metabarcodes was assigned to reference sequences of the genera *Blastodinium* and *Brandtodinium* which comprise only nano-/micro-dinoflagellate species (Table 1). *Blastodinium* includes parasitic species of copepods (Skovgaard *et al*., 2007). *Brandtodinium* is a recently erected genus that includes symbionts of polycystine radiolarians (Probert *et al*., 2014), whose type species, *Brandtodinium nutricula* (Brandt) Probert & Siano, is an emendation of *Scrippsiella nutricola* (Brandt) A.T. Banaszak, R. Iglesias-Prieto & R.K. Trench (Banaszak *et al*., 1993; Probert *et al*., 2014). Relatively high abundances of metabarcodes assigned to *Pelagodinium* spp., the dinoflagellate genus symbiotic of planktonic foraminifera (Siano *et al*., 2010), were also detected in meso-planktonic communities of stations 52, 70 and 72. Thus, the large majority of meso-planktonic dinoflagellate taxa unveiled by our metabarcoding survey correspond in fact to nano-/micro-dinoflagellate symbionts of mesoplanktonic heterotrophic protists or metazoans. The ecological role of this important community of symbiotic (parasitic, mutualistic or commensal) dinoflagellates is still poorly understood. The symbiotic dinoflagellate metabarcodes from the meso-plankton are also found in smaller size fractions, corresponding likely to the endosymbiotic and free-living life cycle stages of the same

taxa respectively. Again, the putatively complex life cycles of symbiotic dinoflagellates are barely characterized. Single-host organisms may harbour, as adult, several symbiotic and morphologically transformed dinoflagellate cells (Trench and Blank, 1987; Caron *et al*., 2000; Siano *et al*., 2010), multi-nucleated trophont and physiologically differentiated cells (Skovgaard *et al*., 2012) or even morphologically and genetically distant species (Decelle *et al*., 2012). However, the ontogeny of host–symbiont interactions (holobionts) is usually ignored and much morpho-molecular work across size fractions remains to be done to understand the ecological and biological features and significance of pelagic symbioses involving dinoflagellates. The biogeographic pattern of a higher prevalence of *Blastodinium* and *Brandtodinium* metabarcodes north and south of 10°North latitude, respectively (Fig. 4), may be linked to the seasonal distribution of host species.

Concluding remarks

Despite the lack of taxonomic resolution of V9 rDNA barcodes in dinoflagellates, our first deep metabarcoding study of this key phytoplanktonic lineage unveiled solid and globally coherent patterns of biodiversity at higher taxonomic levels across protist size fractions, space and time, supporting the value of next-generation metabarcoding as an alternative tool for quantitative molecular ecology and biogeography of marine protists. Our data suggest that the importance of pelagic dinoflagellates relative to all other protists has been largely underestimated. Exploring the frontier of dinoflagellate biodiversity, our study unveiled rich and unique community structures in both the largest and smallest size fractions, with abundant Peridiniales symbionts *sensu lato* in the meso-plankton and a phenomenal diversity and abundance of Gymnodiniales assigned metabarcodes in the pico-nano-plankton. Clearly, metabarcoding highlights the degree of our ignorance concerning the diversity of tiny and essentially symbiotic species escaping traditional microscopy detection.

Despite their extreme diversity, abundance and functional complexity, dinoflagellate display strong and differential modes of community structure across organismal size fractions, suggesting that some orders have evolved their diversity within specific size ranges independently of extrinsic eco-evolutionary pressures. At finer systemic levels, the abundance of specific metabarcodes within a particular order at a given station is less uniform across the sampled ecosystems, likely responding to local ecological triggers. Thus, in biogeochemical models, instead of being lumped in one single group, dinoflagellates could be considered as different taxo-functional (e.g. autotrophic, heterotrophic) groups along the different size classes and help studying abiotic resources use and carbon production across size classes as well as the trophic relation with higher levels (e.g. copepods) of the marine food web. However, V9 rDNA metabarcoding will clearly need to be complemented by other metabarcodes targeting finer diversity levels and functional features (such as the diversity of group-specific chloroplasts) in order to fully assess the eco-functional dynamics of plankton communities. Finally, the development of morphogenetic protocols allowing linking of unknown metabarcode diversity to phenotypic and cellular complexity will be a key to anchor the unveiled biological diversity to its biogeochemical context.

Experimental procedures

Sampling

Samples analysed in this study were collected in the frame of the 30 month-long international expedition Tara Oceans (Karsenti *et al*., 2011; Pesant *et al*., 2015), which explored seven oceanic regions and 12 Longhurst's provinces (Longhurst, 2007), allowing it to collect worldwide plankton samples from a complete range of planktonic ecosystems (coastal, tropical, oceanic, upwelling, etc.) and through a wide inter-annual and seasonal variability. Plankton from 10 size classes, from viruses to fish larvae, was collected together with a series of physico-chemical contextual data (Karsenti *et al*., 2011; Pesant *et al*., 2015). For this study focusing on dinoflagellates, we used samples collected at 33 Tara Oceans stations over a 2-year period from 15 September 2009 to 13 May 2011. These samples concern four organismal size fractions (four samples per station, one per size fraction) from the smallest to the largest protists (0.8–5 μm; 5–20 μm; 20–180 μm; 180–2000 μm) and cover seven oceanic basins [the North Atlantic Ocean (one station), the Mediterranean Sea (six), the Red Sea (four), the Indian Ocean (eight), the South Atlantic Ocean (eight), the Southern Ocean (two) and the South Pacific Ocean (four) (Table 2)]. For convenience, we will call these four size fractions respectively pico-nano-, nano-, micro- and meso-plankton, although rigorous size boundaries are classically set at 2, 20, and 200 μm. Samples were collected in subsurface (< 5m) waters using an industrial peristaltic pump for the smallest size fraction (0.8–5 μm) and a series of plankton nets with different mesh size for the three larger size fractions (5–2000 μm). The pumped water was pre-filtered over 200 μm pre-filter than passed through a Gravity Plankton Sieving System (GPSS, a superposition of three nets with successive mesh sizes of 20 μm, 5 μm and 5 μm again, in order to carefully separate the larger and often fragile plankton from the cells < 5 μm. This system was not expressly designed to assess dinoflagellate but rather total protist diversity. The risk of any filtering strategy is that it may create a bias via the disruption of cells and their dispersion across smaller size fractions. This risk is particularly high for dinoflagellates, which contain large and fragile unarmoured species. However, GPSS was designed to concentrate large volumes of water as gently as possible in order to obtain sufficient biomass and nucleic acids to explore plankton diversity even in ultra-oligotrophic water masses. Two samples of 100L of recovered < 5 μm

Table 2. Geographical coordinates and Longhurst provinces (Longhurst, 2007) of sampling stations (4–109).

Station number	Latitude	Longitude	Sampling day	Season	Longhurst Province ID/name	Ecosystem
4	N 036° 34′ 22″	W 006° 32′ 19″	15 Sep. 2009	Summer	NASE	Westerlies
7	N 037° 02′ 19″	E 001° 56′ 59″	23 Sep. 2009	Autumn	MEDI	Westerlies
9	N 039° 04′ 24″	E 005° 51′ 35″	28 Sep. 2009	Autumn	MEDI	Westerlies
18	N 035° 45′ 37″	E 014° 15′ 07″	2 Nov. 2009	Autumn	MEDI	Westerlies
23	N 042° 09′ 40″	E 017° 43′ 52″	18 Nov. 2009	Autumn	MEDI	Westerlies
25	N 039° 24′ 30″	E 019° 22′ 50″	23 Nov. 2009	Autumn	MEDI	Westerlies
30	N 033° 55′ 31″	E 032° 46′ 26″	15 Dec. 2009	Autumn	MEDI	Westerlies
31	N 027° 08′ 59″	E 034° 49′ 05″	09 Jan. 2010	Winter	REDS	Coastal
32	N 023° 22′ 59″	E 037° 15′ 08″	11 Jan. 2010	Winter	REDS	Coastal
33	N 022° 03′ 06″	E 038° 13′ 00″	13 Jan. 2010	Winter	REDS	Coastal
34	N 018° 23′ 53″	E 039° 51′ 42″	20 Jan. 2010	Winter	REDS	Coastal
36	N 020° 49′ 03″	E 063° 30′ 43″	12 Mar. 2010	Winter	ARAB	Coastal
38	N 019° 01′ 32″	E 064° 37′ 21″	15 Mar. 2010	Winter	ARAB	Coastal
41	N 014° 33′ 48″	E 070° 02′ 23″	30 Mar. 2010	Spring	ARAB	Coastal
42	N 005° 59′ 33″	E 073° 54′ 16″	04 Apr. 2010	Spring	MONS	Trades
45	N 001° 47′ 13″	E 071° 29′ 40″	12 Apr. 2010	Spring	MONS	Trades
52	S 016° 57′ 24″	E 054° 00′ 37″	17 May. 2010	Autumn	EAFR	Coastal
64	S 029° 32′ 36″	E 037° 56′ 08″	08 Jul. 2010	Winter	EAFR	Coastal
65	S 035° 14′ 48″	E 026° 18′ 34″	12 Jul. 2010	Winter	EAFR	Coastal
66	S 034° 53′ 35″	E 018° 04′ 22″	15 Jul. 2010	Winter	EAFR	Coastal
67	S 032° 13′ 53″	E 017° 42′ 29″	07 Sep. 2010	Winter	SATL	Trades
68	S 031° 01′ 48″	E 004° 41′ 16″	14 Sep. 2010	Winter	SATL	Trades
70	S 020° 26′ 12″	W 003° 11′ 09″	21 Sep. 2010	Winter	SATL	Trades
72	S 008° 42′ 09″	W 017° 56′ 23″	5 Oct. 2010	Spring	SATL	Trades
76	S 021° 02′ 44″	W 035° 22′ 07″	16 Oct. 2010	Spring	SATL	Trades
78	S 030° 08′ 47″	W 043° 15′ 13″	4 Nov. 2010	Spring	SATL	Trades
82	S 047° 09′ 55″	W 057° 53′ 44″	06 Dec. 2010	Spring	FKLD	Coastal
84	S 060° 13′ 56″	W 060° 38′ 42″	03 Jan. 2011	Summer	ANTA	Polar
85	S 062° 14′ 37″	W 049° 10′ 57″	06 Jan. 2011	Summer	APLR	Polar
98	S 025° 51′ 03″	W 111° 46′ 21″	04 Apr. 2011	Autumn	SPSG	Westerlies
100	S 012° 59′ 40″	W 095° 59′ 07″	15 Apr. 2011	Autumn	SPSG	Westerlies
102	S 005° 16′ 10″	W 085° 13′ 43″	22 Apr. 2011	Autumn	PEQD	Trades
109	N 002° 04′ 36″	W 084° 31′ 13″	13 May. 2011	Spring	PEQD	Trades

NASE = North Atlantic Subtropical Gyral; MEDI = Mediterranean Sea; REDS = Red Sea; ARAB = NW Arabian Upwelling Province; MONS = Indian Monsoon Gyres Province; EAFR = East Africa Coastal Province; SATL = South Atlantic Gyral Province; FKLD = SW Atlantic Shelves Province; ANTA = Antarctic Province; APLR = Austral Polar Province; SPSG = South Pacific Subtropical Gyre; PEQD = Pacific Equatorial Divergence.

sieved seawater were then filtered through two parallel 142 mm diameter, 0.8 μm porosity polycarbonate membranes to recover organisms of the 0.8–5 μm size fraction. The membranes (four per sample) were immediately folded into 5 ml barcoded cryotubes stored in liquid nitrogen on board. For the three > 5 μm size fractions, 5, 20 and 180 μm meshed plankton nets were towed for ~15 min in subsurface waters, and rinsed from the outside with filtered (0.1 μm) seawater when back on board. Plankton samples were then poured from the cod ends through the appropriate sieves into 8 L bottles. The volume was adjusted to 3 L with 0.1 μm filtered seawater. 0.5 L of concentrated plankton was then filtered through two 47 mm polycarbonate membranes and recovered together into a single 5 ml barcoded cryotube which was immediately stored in liquid nitrogen. All cryotubes were stored at −80°C back in the laboratory until further molecular processing. Detailed protocols are available in de Vargas and colleagues (2015).

Genomic DNA extraction, purification and sequencing

Total nucleic acids (DNA + RNA) were extracted from all 132 samples collected (33 stations × 4 size fractions), using the Nucleospin DNA II kit (Macherey-Nagel, Hoerdt, France). The V9 hyper-variable regions of the 18S ribosomal RNA gene were PCR amplified with the eukaryotic primers 1389f 5′-TTGTACACACCGCCC-3′ and 1510r 5′-CCTTCYGC AGGTTCACCTAC-3′ (Amaral-Zettler *et al*., 2009). Amplifications were conducted with the Phusion High-Fidelity DNA Polymerase (Finnzymes). Briefly, the PCR mixture (final volume of 25 μL) contained 5 ng of template with forward and reverse primers at a final concentration of 0.35 μM, 3% of dimethylsulphoxide and 2× of GC buffer Phusion Master Mix (Finnzymes). Amplifications consisted in an initial denaturation step at 98°C for 30 s followed by 25 cycles of 10 s at 98°C, 30 s at 57°C, 30 s at 72°C and a final elongation step at 72°C for 10 min. Each sample was amplified in triplicates to obtain sufficient PCR products, which were systematically checked on 1.5% agarose gels for positive bands of the expected length. The amplification was successful for 106 out of 132 samples collected. Amplicons were then pooled, purified using the NucleoSpin Extract II kit (Macherey-Nagel, Hoerdt, France) and sequenced using a Genome Analyser IIx system (Illumina, San Diego, CA, USA) from both sides (de Vargas *et al*., 2015). Merged paired reads were examined for the exact presence of the forward and reverse primer

sequences and checked if they were chimeras. Chimeric sequences were detected using the chimera search module of the USEARCH program [version 4.2; (Edgar *et al*., 2011)], looking for chimeras that could be derived from sequences in the reference data set as well as from more abundant sequences in the same sample. Singleton sequences showing a poor quality score (more than 1% expected error) were discarded. The hypervariable V9 regions sequenced and retained after data cleaning are here called 'metabarcodes' as also designed in de Vargas and colleagues (2015). A total of 177 821 359 high-quality and complete V9 rDNA protistan metabarcodes were recovered, including 4 402 312 distinct metabarcodes.

Database of reference dinoflagellate 18S rDNA and metabarcode annotation

All 18S rDNA dinoflagellate sequences available in National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/) (Geer *et al*., 2009) on 30 June 2012, were used to construct an *ad hoc* reference database needed for the annotation of Tara Oceans metabarcodes. Reference dinoflagellate sequences were aligned using MAFFT (Katoh and Toh, 2010; Katoh and Standley, 2013) and only those containing both the V4 and V9 rDNA fragments were kept for detailed phylogenetic analyses. Key DNA Tool (http://keydnatools.com) (Guillou *et al*., 2013) was used to detect the presence of chimeras in all remaining sequences, resulting in a total final number of 1191 dinoflagellate reference sequences. The taxonomic status of each sequence was manually checked using phylogenetics (neighbour joining) reconstructions as implemented in SEAVIEW, (Gouy *et al*., 2010) and comparison to public taxonomic reference databases [AlgaeBase (Guiry, 2014) and WoRMS (WoRMS Editorial Board, 2012)]. We imposed a hierarchical nomenclature (order, family, genus and species) on all clades generated by our phylogeny. The polyphyletic clades typically found in dinoflagellate phylogenies were given the same name but were kept distinct through the addition of successive, clade-specific numbers. All reference sequences used herein were integrated into the *PR2* – Protist Ribosomal Reference – database (Guillou *et al*., 2013), and our dinoflagellate-centered database (DinR2, Dinoflagellate ribosomal reference database) is available upon request. All V9 rDNA metabarcodes obtained from the world surface oceans were individually compared with each of the 28 547 PR2 reference sequences (including the 1191 DinR2 sequences). An exact global pairwise alignment algorithm (Needleman and Wunsch, 1970), as implemented in GGSEARCH (Pearson and Lipman, 1988) was used. Metabarcodes were assigned to dinoflagellates and received the taxonomic name of its nearest DinR2 neighbour or of the last common ancestor in case of a tie (de Vargas *et al*., 2015).

Assessment of molecular biodiversity

The entire taxonomically assigned raw data set was reduced by two subsequent cut-offs based first on the number of copies obtained per metabarcode, then on their percentage of identity (% of identity: ID) to reference sequences. In order to minimize the sequencing error and taxonomic assignation biases, only metabarcodes present in 10 or more copies and with a %ID to any dinoflagellate reference sequence equal to or greater than 90% were retained in our final data set. These restrictive criteria did not affect the overall distribution of raw dinoflagellate metabarcodes obtained without any cut-offs (Fig. S4) and limited all biases linked to sequencing errors. We ended up with 83 860 different dinoflagellate metabarcodes, which constituted our raw material for further biodiversity and community analyses. The choice to work with metabarcodes and not with pairwise distance (OTUs: Operational Taxonomic Units) was motivated by the fact that rDNA evolves particularly slowly in dinoflagellates and single substitutions in V9 rDNA can often distinguish different genera or families. Clustering at lower %ID would result in significantly lower taxonomic resolution, blurring the biodiversity patterns. To characterize the relative importance of dinoflagellates over the whole protist community, all known protist classes (72), except the Dinophyceae, were pooled together. We applied to this protist database the same strict and conservative cut-offs used for dinoflagellates. We retained only protist metabarcodes present in 10 or more copies and with a %ID $\geq$ 90% to any protist reference sequence. Only 14 Tara Oceans stations with a complete set of data for all the four size fractions were used for this analysis. Further biodiversity analyses for each individual size fraction were carried out essentially at the taxonomic level of order. Metabarcodes were assigned to the 11 dinoflagellate orders (Dinophysiales, Gonyaulacales, Gymnodiniales, Lophodiniales, Noctilucales, Peridiniales, Phytodiniales, Prorocentrales, Pyrocystales, Suessiales, Thoracosphaerales) plus two undefined groups (uncultured and undetermined). The group 'uncultured' includes environmental metabarcodes that do not match any specific order but are phylogenetically clearly classified as dinoflagellate. The group 'undetermined' contains metabarcodes that are genetically strictly equally distant to two or more taxonomically unrelated reference sequence at the order level. Dinoflagellate biodiversity was calculated using richness and abundance metabarcode values, where abundances are the numbers of copies for a given metabarcode. The relative importance of single group of metabarcodes (size fractions, orders or genera) were calculated as the ratio of the metabarcodes belonging to the group over the total number of metabarcodes either of the station, sample, size class or order depending on the analysis.

Acknowledgements

This research was supported by a Région Bretagne PhD grant to NLB, the CNRS GDR3280 Tara Oceans and the ANR (Agence National de la Recherche) grants *POSEIDON* (ANR-09-BLAN-0348) and France Génomique (ANR-10-INPS-09). It is part of the 'Investissement d'Avenir' program *OCEANOMICS* funded by the French government under the grant agreement ANR-11-BTBR-0008. We thank the coordinators and members of the *Tara Oceans* consortium for organizing samples and data collection, and for their continuous support and collaborative energy. We thank the commitment of the following people and sponsors who made this singular expedition possible: Agnès B. and Etienne Bourgois, the Veolia Environment Foundation, Région

Bretagne, World Courier, Illumina, Cap L'Orient, the EDF Foundation Diversiterre, the Fondation pour la Recherche sur la Biodiversité (FRB), the Prince Albert II de Monaco Foundation, the captain, crew and logistic team of *Tara* and *Tara* Expeditions. *Tara Oceans* would not exist without the continuous support of the 23 participating institutes (see http://oceans.taraexpeditions.org). This article is contribution number 29 of the *Tara Oceans* Expedition 2009–2012.

References

Adl, S.M., Simpson, A.G.B., Farmer, M.A., Andersen, R.A., Anderson, O.R., Barta, J.R., *et al*. (2005) The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J Eukaryot Microbiol* **52:** 399–451.

Amaral-Zettler, L., McCliment, E.A., Ducklow, H.W., and Huse, S.M. (2009) A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *PLoS ONE* **4:** e6372.

Balzano, S., Marie, D., Gourvil, P., and Vaulot, D. (2012) Composition of the summer photosynthetic pico and nanoplankton communities in the Beaufort Sea assessed by T-RFLP and sequences of the 18S rRNA gene from flow cytometry sorted samples. *ISME J* **6:** 1480–1498.

Banaszak, A.T., Prieto, R.I., and Trench, R.K. (1993) *Scrippsiella velellae* sp. nov. (Peridiniales) and *Gloeokinium viscum* sp. nov. (Phytodiniales), Dinoflagellate symbionts of two hydrozoans (Cnidaria). *J Phycol* **29:** 517–528.

Berney, C., Romac, S., Mahé, F., Santini, S., Siano, R., and Bass, D. (2013) Vampires in the oceans: predatory cercozoan amoebae in marine habitats. *ISME J* **7:** 2387–2399.

Caron, D.A., Dennett, M.R., Lonsdale, D.J., Moran, D.M., and Shalapyonok, L. (2000) Microzooplankton herbivory in the Ross Sea, Antarctica. *Deep-Sea Res II* **47:** 3249–3272.

Cheung, M.K., Chu, K.H., Li, C.P., Kwan, H.S., and Wong, C.K. (2008) Genetic diversity of picoeukaryotes in a semi-enclosed harbour in the subtropical western Pacific Ocean. *Aquat Microb Ecol* **53:** 295–305.

Cheung, M.K., Au, C.H., Chu, K.H., Kwan, H.S., and Wong, C.K. (2010) Composition and genetic diversity of picoeukaryotes in subtropical coastal waters as revealed by 454 pyrosequencing. *ISME J* **4:** 1053–1059.

Coffroth, M.-A., and Santos, S.R. (2005) Genetic diversity of symbiotic dinoflagellates in the genus *Symbiodinium*. *Protist* **156:** 19–34.

Decelle, J., Siano, R., Probert, I., Poirier, C., and Not, F. (2012) Multiple microalgal partners in symbiosis with the acantharian *Acanthochiasma* sp. (Radiolaria). *Symbiosis* **58:** 233–244.

Decelle, J., Romac, S., Sasaki, E., Not, F., and Mahé, F. (2014) Intracellular diversity of the V4 and V9 regions of the 18S rRNA in marine protists (Radiolarians) assessed by high-throughput sequencing. *PLoS ONE* **9:** e104297.

Diez, B., Pedros-Alio, C., and Massana, R. (2001) 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.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., and Knight, R. (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27:** 2194–2200.

Edvardsen, B., Jakobsen, K.S., Shalchian Tabrizi, K., Medlin, L.K., Dahl, E., Brubak, S., and Paasche, E. (2003) Genetic variability and molecular phylogeny of *Dinophysis* species (Dinophyceae) from Norwegian waters inferred from single cell analyses of rDNA. *J Phycol* **39:** 395–408.

Faust, M.A., and Gulledge, R.A. (2002) Identifying harmful marine dinoflagellates. Department of Botany, National Museum of Natural History.

Fensome, R.A., Taylor, F.J.R., Norris, G., Sarjeant, W.A.S., Wharton, D.I., and Williams, G.L. (1993) A classification of living and fossil dinoflagellates. *Micropaleontology Press* **7:** 1–351.

Gaines, G., and Elbrächter, M. (1987) Heterotrophic nutrition. In *The Biology of Dinoflagellates*. Taylor, F.J.R. (ed.). Oxford: Blackwell Scientific Publications, pp. 224–268.

Galluzzi, L., Penna, A., Bertozzini, E., Vila, M., Garcés, E., and Magnani, M. (2004) Development of a real-time PCR assay for rapid detection and quantification of *Alexandrium minutum* (a Dinoflagellate). *Appl Environ Microbiol* **70:** 1199–1206.

Geer, L.Y., Marchler-Bauer, A., Geer, R.C., Han, L., He, J., He, S., *et al*. (2009) The NCBI bioSystems database. *Nucleic Acids Res* **38:** D492–D496.

Gómez, F. (2012a) A checklist and classification of living dinoflagellates (Dinoflagellata, Alveolata). *CICIMAR Oceánides* **27:** 65–140.

Gómez, F. (2012b) A quantitative review of the lifestyle, habitat and trophic diversity of dinoflagellates (Dinoflagellata, Alveolata). *Syst Biodivers* **10:** 267–275.

Gómez, F. (2013) Problematic biases in the availability of molecular markers in protists: the example of the dinoflagellates. *Acta Protozool* **52:** 21–33.

Gómez, F., Moreira, D., and López-García, P. (2010) Molecular phylogeny of Noctilucoid dinoflagellates (Noctilucales, Dinophyceae). *Protist* **161:** 466–478.

Godhe, A., Asplund, M.E., Harnstrom, K., Saravanan, V., Tyagi, A., and Karunasagar, I. (2008) Quantification of diatom and dinoflagellate biomasses in coastal marine seawater samples by real-time PCR. *Appl Environ Microbiol* **74:** 7174–7182.

Gong, J., Dong, J., Liu, X., and Massana, R. (2013) Extremely high copy numbers and polymorphisms of the rDNA operon estimated from single cell analysis of oligotrich and peritrich ciliates. *Ann Anat* **164:** 369–379.

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.

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., *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.

Guiry, M.D. (2012) How many species of algae are there? *J Phycol* **48:** 1057–1063.

Guiry, M.D. (2014) *AlgaeBase*. Guiry, M.D., and Guiry, G.M. (eds). Galway: World-wide electronic publication, National University of Ireland. http://www.algaebase.org.

Harrison, P.J., Furuya, K., Glibert, P.M., Xu, J., Liu, H.B., Yin, K., *et al*. (2011) Geographical distribution of red and green *Noctiluca scintillans*. *Chin J Ocean Limnol* **29:** 807–831.

Jeong, H.J. (1999) The ecological roles of heterotrophic dinoflagellates in marine planktonic community. *J Eukaryot Microbiol* **46:** 390–396.

Karsenti, E., Sullivan, M.B., Acinas, S.G., Bork, P., Bowler, C., de Vargas, C., *et al*. (2011) A holistic approach to marine eco-systems biology. *PLoS Biol* **9:** e1001177.

Katoh, K., and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30:** 772–780.

Katoh, K., and Toh, H. (2010) Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics* **26:** 1899–1900.

Ki, J.-S. (2011) Hypervariable regions (V1–V9) of the dinoflagellate 18S rRNA using a large dataset for marker considerations. *J Appl Phycol* **24:** 1035–1043.

Kohli, G.S., Neilan, B.A., Brown, M.V., Hoppenrath, M., and Murray, S.A. (2013) *Cob* gene pyrosequencing enables characterization of benthic dinoflagellate diversity and biogeography. *Environ Microbiol* **16:** 467–485.

Lin, S., Zhang, H., Hou, Y., Zhuang, Y., and Miranda, L. (2009) High-level diversity of dinoflagellates in the natural environment, revealed by assessment of mitochondrial *cox1* and *cob* genes for dinoflagellate DNA barcoding. *App Environ Microbiol* **75:** 1279–1290.

Litaker, R.W., Vandersea, M.W., Kibler, S.R., Reece, K.S., Stokes, N.A., Lutzoni, F.M., *et al*. (2007) Recognizing dinoflagellate species using ITS rDNA sequences. *J Phycol* **43:** 344–355.

Liu, H., Aris-Brosou, S., Probert, I., and de Vargas, C. (2009) A timeline of the environmental genetics of the haptophytes. *Mol Biol Evol* **27:** 161–176.

Logares, R., Audic, S., Santini, S., Pernice, M.C., de Vargas, C., and Massana, R. (2012) Diversity patterns and activity of uncultured marine heterotrophic flagellates unveiled with pyrosequencing. *ISME J* **6:** 1823–1833.

Logares, R., Audic, S., Bass, D., Bittner, L., Boutte, C., Christen, R., *et al*. (2014) Patterns of rare and abundant marine microbial eukaryotes. *Curr Biol* **24:** 813–821.

Longhurst, A.R. (2007) *Ecological geography of the sea*. New York, USA: Academic Press.

Lopez-Garcia, 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.

Lundholm, N., and Moestrup, Ø. (2006) The biogeography of harmful algae. In *Ecology of harmful algae*. Granéli, E. and Turner, J.T. (eds). Heidelberg, Germany: Springer, pp. 23–35.

Massana, R., Balagué, V., Guillou, L., and Pedros-Alio, C. (2004) Picoeukaryotic diversity in an oligotrophic coastal site studied by molecular and culturing approaches. *FEMS Microbiol Ecol* **50:** 231–243.

Massana, R., Gobet, A., Audic, S., Bass, D., Bittner, L., Boutte, C., *et al*. (2015) Marine protist diversity in European coastal waters and sediments as revealed by high-throughput 5 sequencing. *Environ Microbiol*. doi:10.1111/1462-2920.12955.

Moon-van der Staay, S.Y., De Wachter, R., and Vaulot, D. (2001) Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature* **409:** 607–610.

Moreira, D., and Lopez-Garcia, P. (2002) The molecular ecology of microbial eukaryotes unveils a hidden world. *Trends Microbiol* **10:** 31–38.

Nanjappa, D., Audic, S., Romac, S., Kooistra, W.H.C.F., and Zingone, A. (2014) Assessment of species diversity and distribution of an ancient diatom lineage using a DNA metabarcoding approach. *PLoS ONE* **9:** e103810.

Needleman, S.B., and Wunsch, C.D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* **48:** 443–453.

Not, F., Siano, R., Kooistra, W.H.C.F., Simon, N., Vaulot, D., and Probert, I. (2012) Diversity and ecology of eukaryotic marine phytoplankton. In *Genomic insights into the biology of algae*. Jacquot, J.P., and Gadal, P. (eds). Amsterdam, the Netherlands: Elsevier, pp. 1–53.

Orr, R.J.S., Murray, S.A., Stüken, A., Rhodes, L., and Jakobsen, K.S. (2012) When naked became armored: an eight-gene phylogeny reveals monophyletic origin of theca in dinoflagellates. *PLoS ONE* **7:** e50004.

Pawlowski, J., Simpson, A.G.B., Audic, S., Adl, S.M., Bass, D., Belbahri, L., *et al*. (2012) CBOL Protist Working Group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLoS Biol* **10:** e1001419.

Pearson, W.R., and Lipman, D.J. (1988) Improved tools for biological sequence comparison. *PNAS* **85:** 2444–2448.

Percopo, I., Siano, R., Rossi, R., Soprano, V., Sarno, D., and Zingone, A. (2013) A new potentially toxic *Azadinium* species (Dinophyceae) from the Mediterranean Sea, *A. dexteroporum* sp. nov. *J Phycol* **49:** 950–966.

Pesant, S., Not, F., Picheral, M., Kandels-Lewis, S., Le Bescot, N., Sciandra, A., *et al*. (2015) Tara Oceans data: open science resources for the discovery and analysis of Tara Oceans data. *Sci Data* **2:** 150023. doi:10.1038/sdata.2015.23.

Pomroy, A.J. (1989) Scanning electron microscopy of *Heterocapsa minima* sp. nov. (Dinophyceae) and its seasonal distribution in the Celtic Sea. *Br Phycol J* **24:** 131–135.

Probert, I., Siano, R., Poirier, C., Decelle, J., Biard, T., Tuji, A., *et al*. (2014) *Brandtodinium* gen. nov. and *B. nutricula* comb. nov. (Dinophyceae), a dinoflagellate commonly found in symbiosis with polycystine (Radiolaria). *J Phycol* **50:** 388–399.

Puigserver, M., and Zingone, A. (2002) *Prorocentrum nux* sp. nov. (Dinophyceae), a small planktonic dinoflagellate from the Mediterranean Sea, and discussion of *P. nanum* and *P. pusillum*. *Phycologia* **41:** 29–38.

Rowan, R., and Powers, D.A. (1992) Ribosomal RNA sequences and the diversity of symbiotic dinoflagellates (zooxanthellae). *PNAS* **89:** 3639–3643.

Schnepf, E., and Elbrächter, M. (1992) Nutritional strategies in dinoflagellates: a review with emphasis on cell biological aspects. *Eur J Protistol* **28:** 3–24.

Siano, R., Kooistra, W.H.C.F., Montresor, M., and Zingone, A. (2009) Unarmoured and thin-walled dinoflagellates from

the Gulf of Naples, with the description of *Woloszynskia cincta* sp. nov. (Dinophyceae, Suessiales). *Phycologia* **48:** 44–65.

Siano, R., Montresor, M., Probert, I., Not, F., and de Vargas, C. (2010) *Pelagodinium* gen. nov. and *P. béii* comb. nov., a dinoflagellate symbiont of planktonic foraminifera. *Protist* **161:** 385–399.

Siokou-Frangou, I., Christaki, U., Mazzocchi, M.G., Montresor, M., Ribera d'Alcalá, M., Vaqué, D., and Zingone, A. (2010) Plankton in the open Mediterranean Sea: a review. *Biogeosciences* **7:** 1543–1586.

Skovgaard, A., Massana, R., and Saiz, E. (2007) Parasitic species of the genus *Blastodinium* (Blastodiniphyceae) are peridinioid dinoflagellates. *J Phycol* **43:** 553–560.

Skovgaard, A., Karpov, S.A., and Guillou, L. (2012) The parasitic dinoflagellates *Blastodinium* spp. inhabiting the Gut of Marine, planktonic copepods: morphology, ecology, and unrecognized species diversity. *Front Microbiol* **3:** 1–22.

Smayda, T.J., and Reynolds, C.S. (2003) Strategies of marine dinoflagellate survival and some rules of assembly. *J Sea Res* **49:** 95–106.

Stern, R.F., Horak, A., Andrew, R.L., Coffroth, M.-A., Andersen, R.A., Küpper, F.C., *et al*. (2010) Environmental barcoding reveals massive dinoflagellate diversity in marine environments. *PLoS ONE* **5:** e13991.

Stern, R.F., Andersen, R.A., Jameson, I., Küpper, F.C., Coffroth, M.-A., Vaulot, D., *et al*. (2012) Evaluating the ribosomal internal transcribed spacer (ITS) as a candidate dinoflagellate barcode marker. *PLoS ONE* **7:** e42780.

Taberlet, P., Coissac, E., and Pompanon, F. (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol Ecol* **21:** 2045–2050.

Thessen, A.E., Patterson, D.J., and Murray, S.A. (2012) The taxonomic significance of species that have only been observed once: the genus *Gymnodinium* (Dinoflagellata) as an example. *PLoS ONE* **7:** e44015.

Trench, R.K., and Blank, R.J. (1987) *Symbiodinium microadriaticum* Freudenthal, S. Goreauii sp. nov., *S. kawagutii* sp. nov. and *S. pilosum* sp. nov.: gymnodinioid dinoflagellate symbionts of marine invertebrates. *J Phycol* **23:** 469–481.

de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., *et al*. (2015) Eukaryotic plankton diversity in the sunlit global ocean. *Science* **348:** 1261605-1–1261605-11.

Worden, A.Z. (2006) Picoeukaryote diversity in coastal waters of the Pacific Ocean. *Aquat Microb Ecol* **43:** 165–175.

WoRMS Editorial Board. (2012) WoRMS – World Register of Marine Species. [WWW document]. URL http://www.marinespecies.org.

Zhu, F., Massana, R., Not, F., Marie, D., and Vaulot, D. (2005) Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiol Ecol* **52:** 79–92.

Zingone, A., Sarno, D., Siano, R., and Marino, D. (2010) The importance and distinctiveness of small-sized phytoplankton in the Magellan Straits. *Polar Biol* **34:** 1269–1284.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Fig. S1. Metabarcode abundances and richness assigned to the different dinoflagellate orders across size fractions per sampling station.

Fig. S2. Frequency of conflict assignation for undetermined dinoflagellates in size fraction 0.8–5 μm. The relative importance of orders is represented in each bar.

Fig. S3. Abundances of top 100 metabarcodes (orders/genus/%ID) found in fraction 0.8–5 μm across all size fractions.

Fig. S4. Number of dinoflagellate metabarcodes in relation to their percentage of identity (%ID) with taxonomic references. The gray bars represent the original data set. The green + yellow bars represent the database conserved after eliminating sequences present in less than 10 copies; the green bars only represent the final data set obtained keeping only sequences with %ID ≥ 90% to any dinoflagellate reference sequence.