Erschienen in: Current Biology ; 28 (2018), 22. - S. 3625-3633.e3
https://dx.doi.org/10.1016/j.cub.2018.09.024

Worldwide Occurrence and Activity of the Reef-Building Coral Symbiont *Symbiodinium* in the Open Ocean

Johan Decelle,[1,2,3,11,*] Quentin Carradec,[3,4] Xavier Pochon,[5,6] Nicolas Henry,[1,3] Sarah Romac,[1,3] Frédéric Mahé,[7] Micah Dunthorn,[8] Artem Kourlaiev,[9] Christian R. Voolstra,[10] Patrick Wincker,[3,4] and Colomban de Vargas[1,3,*]
[1]Sorbonne Université, CNRS, Station Biologique de Roscoff, AD2M, UMR 7144, 29680 Roscoff, France
[2]Cell & Plant Physiology laboratory UMR 5168, University of Grenoble Alpes, CNRS, CEA, INRA, 38054 Grenoble Cedex 9, France
[3]Research Federation for the study of Global Ocean Systems Ecology and Evolution, FR2022/GOSEE, 3 rue Michel-Ange, 75016 Paris, France
[4]Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Université Evry, Université Paris-Saclay, 91057 Evry, France
[5]Coastal and Freshwater Group, Cawthron Institute, Private Bag 2, Nelson 7042, New Zealand
[6]Institute of Marine Science, University of Auckland, Private Bag 349, Warkworth 0941, New Zealand
[7]CIRAD, UMR LSTM, 34398 Montpellier, France
[8]Department of Ecology, Technische Universität Kaiserslautern, Kaiserslautern, Germany
[9]Genoscope, Institut François Jacob, CEA, 91057 Evry, France
[10]Red Sea Research Center, Division of Biological and Environmental Science and Engineering (BESE), King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia
[11]Lead Contact
*Correspondence: johan.decelle@univ-grenoble-alpes.fr (J.D.), vargas@sb-roscoff.fr (C.d.V.)

SUMMARY

The dinoflagellate microalga *Symbiodinium* sustains coral reefs, one of the most diverse ecosystems of the biosphere, through mutualistic endosymbioses with a wide diversity of benthic hosts [1]. Despite its ecological and economic importance, the presence of *Symbiodinium* in open oceanic waters remains unknown, which represents a significant knowledge gap to fully understand the eco-evolutionary trajectory and resilience of endangered *Symbiodinium*-based symbioses. Here, we document the existence of *Symbiodinium* (i.e., now the family Symbiodiniaceae [2]) in tropical- and temperate-surface oceans using DNA and RNA metabarcoding of size-fractionated plankton samples collected at 109 stations across the globe. *Symbiodinium* from clades A and C were, by far, the most prevalent and widely distributed lineages (representing 0.1% of phytoplankton reads), while other lineages (clades B, D, E, F, and G) were present but rare. Concurrent metatranscriptomics analyses using the *Tara* Oceans gene catalog [3] revealed that *Symbiodinium* clades A and C were transcriptionally active in the open ocean and expressed core metabolic pathways (e.g., photosynthesis, carbon fixation, glycolysis, and ammonium uptake). Metabarcodes and expressed genes of clades A and C were detected in small and large plankton size fractions, suggesting the existence of a free-living population and a symbiotic lifestyle within planktonic hosts, respectively. However, high-resolution genetic markers and microscopy are required to confirm the life history of oceanic *Symbiodinium*. Overall, the previously unknown, metabolically active presence of *Symbiodinium* in oceanic waters opens up new avenues for investigating the potential of this oceanic reservoir to repopulate coral reefs following stress-induced bleaching.

RESULTS AND DISCUSSION

Presence and Diversity of *Symbiodinium* in Open Oceanic Waters

Endosymbioses involving the dinoflagellate genus *Symbiodinium* (now representing the family Symbiodiniaceae [2]) are the foundation of coral reef ecosystems that provide close to $375 billion in goods and services each year [1, 4, 5]. Because of widespread coral reef mortality and degradation [6, 7], there is an urgent need to comprehensively understand the life history, ecology, and evolution of *Symbiodinium*. To date, most research on the diversity and biogeography of *Symbiodinium* has focused on symbiotic populations from coastal benthic ecosystems [8, 9]. Studies of free-living *Symbiodinium* in the environment (*ex hospite*) are comparatively rare [10, 11], despite the importance of free-living stages for sexual recombination [12, 13] and as a reservoir for larval hosts that must acquire new symbionts for their growth [14]. It is currently not known whether this microalga is endemic to coastal benthic ecosystems or can also exist as a free-living planktonic alga in the open ocean between coastal reef ecosystems. While population genetic studies on *in hospite Symbiodinium* tend to report low connectivity between reefs [8, 15], the extended genetic repertoire of

Symbiodinium (i.e., up to 49,000 genes) [16, 17], its mixotrophic ability [18], and the motile stages observed in culture together argue for adaptation to pelagic waters. Tracing the occurrence of *Symbiodinium* in oceanic waters is crucial in evaluating whether there is a planetary reservoir of cells that could potentially recolonize corals and restart carbon fixation in near-shore reef ecosystems following mass bleaching.

Here, we generated and analyzed DNA/RNA metabarcoding datasets from plankton samples collected during the *Tara* Oceans expedition [19] in order to assess the presence of *Symbiodinium* (i.e., the family Symbiodiniaceae) in the world's open ocean and thereby reinterpret its ecology through a macroevolutionary lens. We analyzed >3.6 million eukaryotic rDNA metabarcodes (i.e., the V9 region of the nuclear 18S rRNA) from 324 size-fractionated plankton communities collected in surface waters (0–10 m deep) at 121 stations (Figure 1A). *Symbiodinium* reads were extracted using a reference database containing V9 rDNA sequences from the nine *Symbiodinium* clades described so far (clades A–I, which can represent different genera) ([2]; see also [9, 20]) (Figure S1; Table S1). *Symbiodinium* reads were found in 90% (109 out of 121) of the pelagic stations, revealing an unexpected pandemic distribution of this microalga far from benthic coastal ecosystems (i.e., up to 3,000 km away) (Figure 1; Tables S2 and S3). The relative abundance of *Symbiodinium* reads among the total reads of the entire eukaryotic planktonic community (protists and metazoans) was globally more pronounced in tropical and sub-tropical waters (up to 0.1%), and decreased or even became undetectable at higher latitudes (e.g., Southern Ocean) (Figure 1), indicating an ecological preference for warm and oligotrophic oceanic waters.

In total, we found 44,607 *Symbiodinium* reads distributed in the piconanoplankton (*piconano*) (0.8–5 μm), the microplankton (*micro*) (20–180 μm), and, to a lesser extent, the mesoplankton (*meso*) (180–2,000 μm) size fractions (Figures 1B and 1C). In the smallest size fraction, which putatively corresponds to free-living cells, *Symbiodinium* reads represented, on average, 0.1% of the phytoplankton reads and, in some pelagic stations, reached abundances comparable to those of other known phytoplankton genera, such as the haptophytes *Phaeocystis* and *Emiliania* or the diatom *Skeletonema* (Data S2). In the *micro* and *meso* plankton size fractions that likely include symbiotic stages within host organisms, *Symbiodinium* reads represented 0.3% and 1.4% of phytoplankton reads, respectively. Overall, a large fraction of *Symbiodinium* reads (68%) were identical to 18 reference sequences of *Symbiodinium* clades A, B, C, D, F, and G (*sensu* [21]). Clade A, which is the earliest diverging *Symbiodinium* lineage in phylogenetic reconstructions [21, 22], was, by far, the most dominant clade in the open ocean (77% of total *Symbiodinium* reads in all size fractions and stations), followed by clades C (18%), F (1%), and B (0.4%) at much lower abundances. The dominance of clade A among *Symbiodinium* reads decreased from the smallest (93%) to the largest size fractions (43.4%), whereas the contribution of clade C (the genus *Cladocopium*) increased (45%). In the large size fractions (20–180 μm and 180–2,000 μm), *Symbiodinium* reads likely correspond to a symbiotic stage within planktonic host organisms. For clade A, most reads (80%) were identical to V9 rDNA sequences of *Symbiodinium* recently found in endosymbiosis within the pelagic ciliate *Tiarina* sp. from the same samples [23]. Co-occurrence analyses showed a significant correlation between clade A and the ciliate (Spearman's rho = 0.64; Data S2). These clade A sequences were also found in the smallest size fraction (Figure 2B), suggesting that they may correspond to free-living stages of *Tiarina* symbionts. However, the sequencing of a barcode with a higher taxonomic resolution [24] will be needed to verify this hypothesis. For clade C, co-occurrence analyses showed a correlation with anthozoans (Spearman's rho = 0.73; Data S2), suggesting a symbiotic stage within the larvae of anemones and/or corals drifting in the open ocean. Clade B (the genus *Breviolum*) was specifically found in the Mediterranean Sea in all size fractions, while clade D was mainly detected in the Indian Ocean in the *piconano* size fraction (Figure 1C). The relatively rare clades F and G were found in the Indian and South Pacific Oceans up to 500 km away from coral reef habitats, mainly in the larger plankton size fractions.

To evaluate whether the putative free-living *Symbiodinium* cells are active in the ocean, we extracted total RNA from the smallest fractions (0.8–20 μm) of 16 plankton communities from surface waters in the Indian, North Atlantic, and South Pacific Oceans, from 50 to 3,000 km away from coral reef ecosystems. RNA was targeted, as it better represents active organisms in the water column, as compared to DNA [25, 26]. The plastidial 23S rRNA gene was obtained and taxonomically assigned with a newly constructed reference database (Data S1) [27]. Corroborating the V9 rDNA barcode, plastidial RNA metabarcodes were largely dominated by *Symbiodinium* clades A and C (Figure 1D). Both clades were represented by few highly abundant metabarcodes strictly identical or highly similar (>96%) to *Symbiodinium* reference sequences from benthic reef hosts (Figure S2; Data S1). In much lower abundance, plastidial metabarcodes of clades D and B were found evenly across stations (Figure 1E). Although detected by only 11 reads, a metabarcode strictly identical to the reference sequence of *Symbiodinium voratum* [28] in clade E (the genus *Effrenium*) was found in the Indian and North Atlantic Oceans. This species is known to feed on bacteria and other microalgae with its well-developed peduncle and is presumed to be free living and not involved in stable symbioses [18].

In the water columns of coral reefs in the Indian and South Pacific Oceans, the relative abundance and diversity patterns of *Symbiodinium* V9 rDNA were similar to those observed in oceanic waters (outside reefs), with dominance of clades A and C (Figure 2). Clade C was the dominant *Symbiodinium* lineage (73%), with a lower contribution of clade A (18%), compared to pelagic waters, followed by clade D (2.6%). In the smallest size fraction, the presence of clades C and A (60% and 31%, respectively) corroborates previous tropical reef studies showing the presence of free-living clade C (the most speciose and common clade of symbionts of reef-building corals [22, 29]) and clade A in reefs' water columns and sediment [10, 11, 30–32]. The other clades (B, D, F, and G) were relatively more prevalent in reef waters (4.5% of *Symbiodinium* reads) than open oceans (1.7%), particularly in larger size fractions (Figure 2A). Of note, clades E, H, and I, which are assumed to have a narrow geographic distribution and host spectra [20], were not detected in the reef and oceanic V9 rDNA datasets.

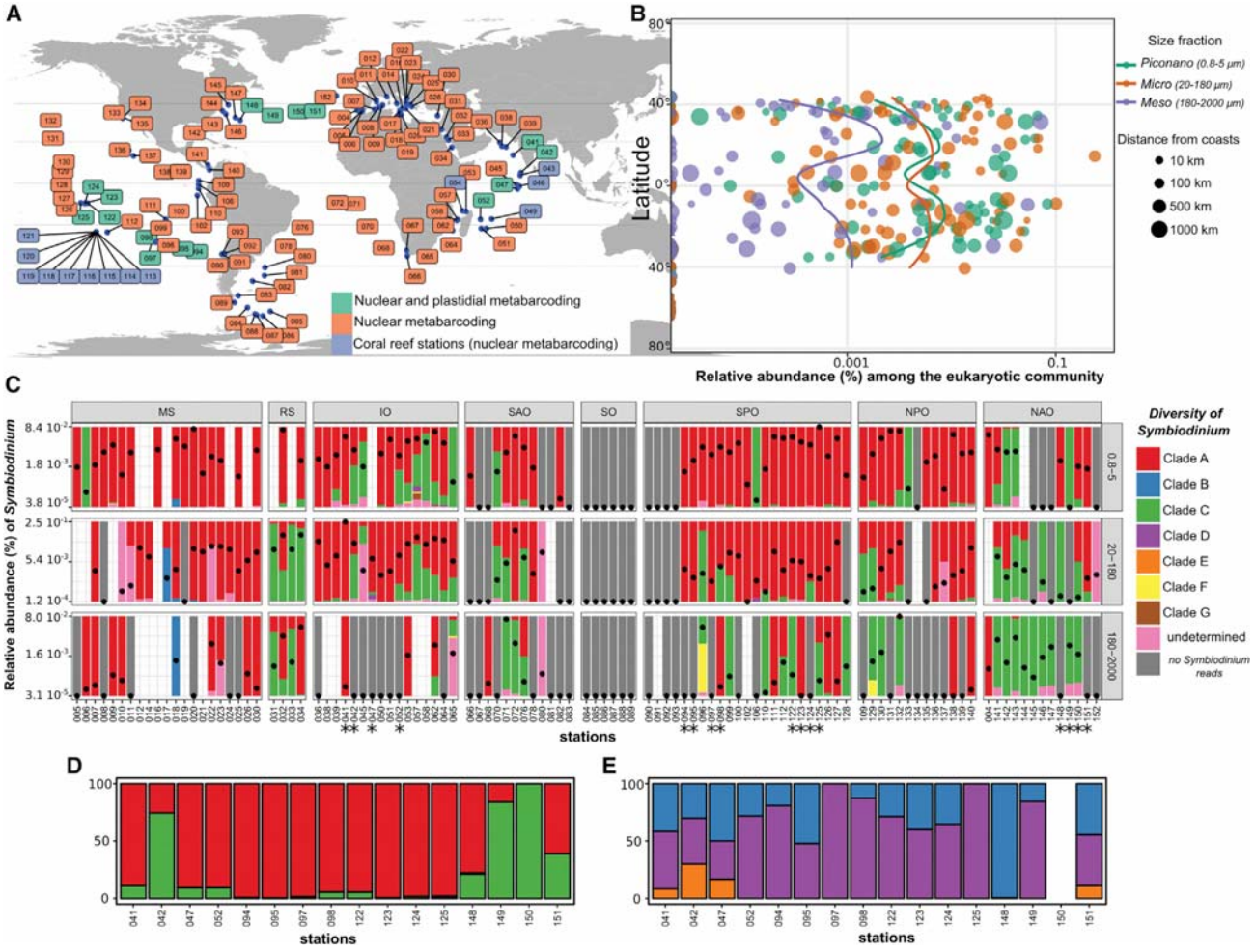

Figure 1. Biogeography and Diversity of *Symbiodinium* in the Open Ocean

(A) World map showing the different oceanic and reef sampling stations of the *Tara* Oceans expedition, from which the metabarcoding and metatranscriptomics datasets were analyzed in this study. Based on a newly built reference database (see also Figure S1 and Table S1), the environmental nuclear 18S rDNA metabarcodes (V9 region) of the microalga *Symbiodinium* were searched in the oceanic and reef stations, which are highlighted in orange and blue, respectively. In addition to the V9 rDNA, the plastidial 23S rRNA marker (hypervariable region cp23S-HVR) of *Symbiodinium* was also obtained in the oceanic stations (far from reef waters) and is highlighted in green.

(B) Scatterplot showing the relative abundance of *Symbiodinium* V9 rDNA reads (in percentages) in the open ocean among the total reads of the eukaryotic community (metazoans and protists) across different latitudes and plankton size fractions (*piconano*: 0.8–5 µm; *micro*: 20–180 µm; *meso*: 180–2,000 µm). The size of each circle represents the distance from the coast (calculated from the ETOPO1 Global Relief Model from NOAA [4-arcminute resolution]: https://doi.org/10.7289/V5C8276M using the marmap R library), including reef ecosystems. The lines are defined by a regression analysis for each size fraction (LOESS, LOcally Estimated Scatterplot Smoother). (See also Table S2 and Data S2.)

(C) Bar chart showing the V9 rDNA diversity and relative abundance of *Symbiodinium* at the clade level in three plankton size fractions, across pelagic stations of different oceanic basins (MS, Mediterranean Sea; RS, Red Sea; IO, Indian Ocean; SAO, South Atlantic Ocean; SO, Southern Ocean; SPO, South Pacific Ocean; NPO, North Pacific Ocean; NAO, North Atlantic Ocean). Each *Symbiodinium* clade is highlighted by a distinct color, while gray bars indicate that no *Symbiodinium* reads have been detected in a given station. Black dots represent the relative abundance (in percentages) of *Symbiodinium* V9 rDNA reads among the total eukaryotic plankton community. (See also Table S2 and Data S2.)

(D and E) Bar charts showing plastidial 23S rRNA diversity of *Symbiodinium* (hypervariable region cp23S-HVR) in the 0.8- to 20-µm plankton size fractions across oceanic stations (highlighted by asterisks in C) considering all clades (D) and excluding the dominant clades A and C (E).

See also Figures S1 and S2, Tables S1 and S2, Data S1, and Data S2.

Symbiodinium Gene Diversity and Function in the Open Ocean

We next explored whether *Symbiodinium* populations in oceanic waters are metabolically active by analyzing the global ocean atlas of eukaryotic genes [3], containing 116 million expressed unigenes. We used 23 reference transcriptomes of *Symbiodinium* from clades A to F to extract all environmental *Symbiodinium* unigenes and analyzed their function and expression across stations and size fractions (Table S3). We found 5,119 unigenes that shared >95% identity to reference transcriptomes of clades A and C, but we did not detect unigenes from any other clades (Figure 3A). This confirms that clades A and C are not

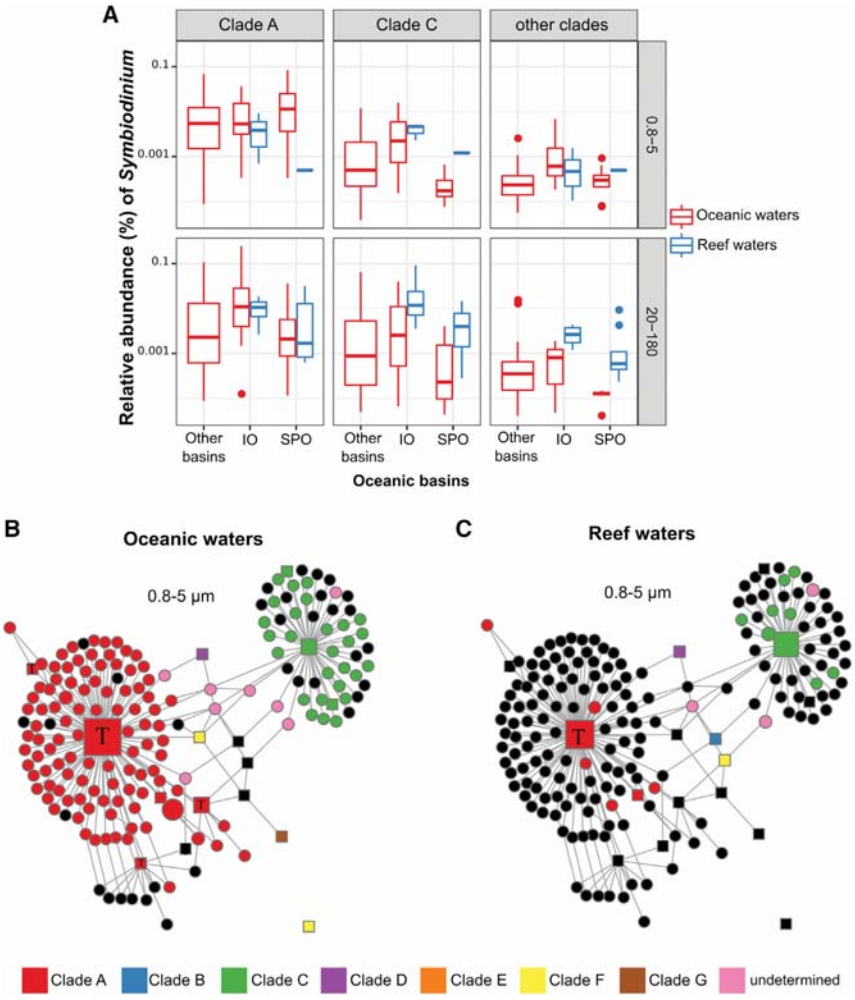

Figure 2. Genetic Diversity and Relative Abundance of *Symbiodinium* in Oceanic Waters versus Reef Waters

(A) Boxplot showing the relative abundance (in percentages) of V9 rDNA reads assigned to *Symbiodinium* clades A and C and other clades (B, D, F, and G) from oceanic and reef waters in the Indian (IO) and South Pacific (SPO) Oceans and in the *piconano* (0.8–5 μm) and *micro* (20–180 μm) size fractions.

(B and C) Haplotype networks showing the different V9 rDNA metabarcodes assigned at ⩾99% nucleic similarity in oceanic (B) and reef (C) waters and present in at least five samples in the piconano plankton size fraction. Each square symbol and round symbol corresponds to one metabarcode assigned at 100% and at ⩾99% to a reference sequence, respectively. The size of the symbol indicates the abundance of the metabarcode in the dataset, and its color indicates a specific *Symbiodinium* clade. Black symbols indicate that the metabarcode is not present in oceanic but in reef waters or vice versa. The letter T inside symbols indicates that the V9 rDNA metabarcode is identical to a sequence of *Symbiodinium* found in symbiosis with the pelagic ciliate *Tiarina* [23]. Haplotype networks were built with the R igraph library.

See also Figure S1.

only the most abundant *Symbiodinium* in the open ocean but also the most transcriptionally active. More specifically, we unveiled the partial transcriptome of clade A containing 2,983 unigenes assigned to type A1 at 97.1% identity on average (Figure 3B). Like metabarcodes, expressed genes of clade A were found in the small (*piconano*) size fraction (Figures 3C and S2), supporting the existence of clade A as an active autonomous microalga in the open ocean. However, unigenes were relatively less prevalent in the small size fraction (up to 0.0017% of all mapped reads) than in the *micro* size fraction (up to 0.021%), suggesting a lower abundance and/or lower gene expression of clade A in the free-living stage. We also detected the partial transcriptome of *Symbiodinium* clade C composed of 2,136 unigenes, of which 1,337 were assigned to type C1 [33] with 98.9% identity, and 1,601 were assigned to type C15 [34] with 96.9% identity (i.e., 802 unigenes were aligned on both transcriptomes) (Figures 3A and 3B). These unigenes were detected in two stations in the *piconano* size fraction (up to 0.002% of all mapped reads) and were relatively more prevalent in the *micro* and *meso* size fractions (up to 0.015% of all mapped reads) (Figure 3C).

Despite the incompleteness of the two open ocean *Symbiodinium* clade A and C transcriptomes, we analyzed the most highly prevalent functions by counting the number of unigenes carrying gene ontology (GO) terms (Figures 4A and 4C). For both clades, genes associated to translational activity were the most abundant, followed by oxidation-reduction processes, photosynthesis, proteolysis, and carbohydrate metabolic processes (e.g., glycolysis), indicating that the main central metabolic pathways of a microalga are expressed in oceanic *Symbiodinium*. In order to assess the functional expression of *Symbiodinium* unigenes, we calculated the expression level of unigenes carrying the same Pfam domains and compared their average relative expressions (Figures 4B and 4D; Table S4). In clades A and C, light-harvesting-complex coding genes (Pfam: Chloroa_b-bind) were the most expressed, with up to 9% of total gene expression. Clade A also exhibited high transcription levels for genes involved in (1) photosynthesis (e.g., photosystems I and II), such as PsbH (1.46%, on average, across size fractions), Rieske domain-containing proteins (1.46%), PsbU (1.08%), Ferredoxin (Fer2; 0.80%), PsbY (0.76%), and PsaE (0.47%); (2) carbon fixation, such as bicarbonate transporter (HCO3_cotransp; 0.80%) and carbonic anhydrase (cdCA1; 0.59%); (3) glycolysis, such as glyceraldehyde 3-phosphate dehydrogenase (Gp_dh:C; 0.74%) and fructose-bisphosphate aldolase (F_bP_aldolase; 1.20%); and (4) nitrogen metabolism, such as NAD-specific glutamate dehydrogenase (NAD-GH; 0.69%). These results demonstrate that clade A is photosynthetically active and capable of fixing carbon and assimilating inorganic nitrogen in the small (free-living) and large size fractions. In clade C, in addition to genes involved in the photosynthetic machinery, carbonic

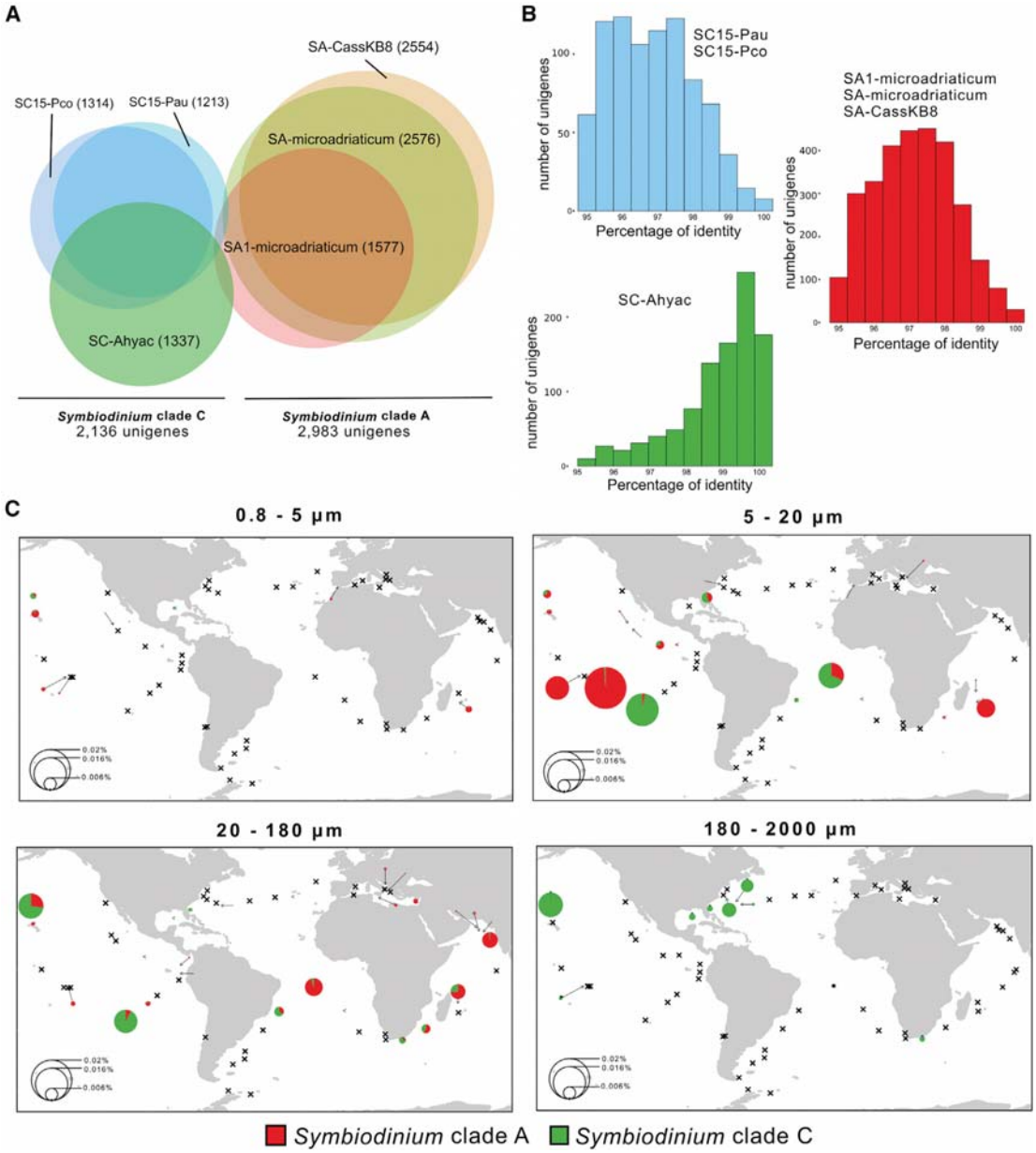

Figure 3. Taxonomic Origin and Biogeography of the *Symbiodinium* Transcriptomes Expressed in the Open Ocean

(A) Euler diagram showing the number of environmental unigenes with nucleic similarities (>95% of identity and >50% of unigene length) to one or several reference *Symbiodinium* transcriptomes (see also Table S3). Each circle represents a set of unigenes matching one *Symbiodinium* reference transcriptome, and the circle size is proportional to the number of unigenes expressed in the open ocean (indicated in parentheses). Overlapping circles represent unigenes shared between two or more reference transcriptomes. *Symbiodinium* transcriptomes matching fewer than 500 unigenes were removed from this figure.

(B) Distribution of the percentages of identity between unigenes and *Symbiodinium* reference transcriptomes. Unigenes matching clade C15 and A1 transcriptomes were pooled (blue and red graphs, respectively).

(C) World map representation of *Symbiodinium* unigenes (clades A and C) detected in the global ocean atlas of the eukaryotic genes of *Tara* Oceans [3], across different plankton size fractions (see also Figure S2). Circle sizes represent the percentages of reads aligned on all unigenes of *Symbiodinium* clade A (red) or clade C (green) transcriptomes.

See also Figure S2 and Table S3.

anhydrase (Pro_CA: 1.4%) and ammonium transporters (1.25%) were highly expressed (Figure 4D). Of note, ammonium transporter genes in clades A and C tended to be more prevalent in the small size fraction (clade C, 2.48%; and clade A, 0.47%) than in larger size fractions (clade C, 0.69%; clade A, 0.16%). This suggests that nitrogen uptake and recycling are different between the free-living and symbiotic lifestyles. These results also reflect recent analysis of *Symbiodinium* genomes that found

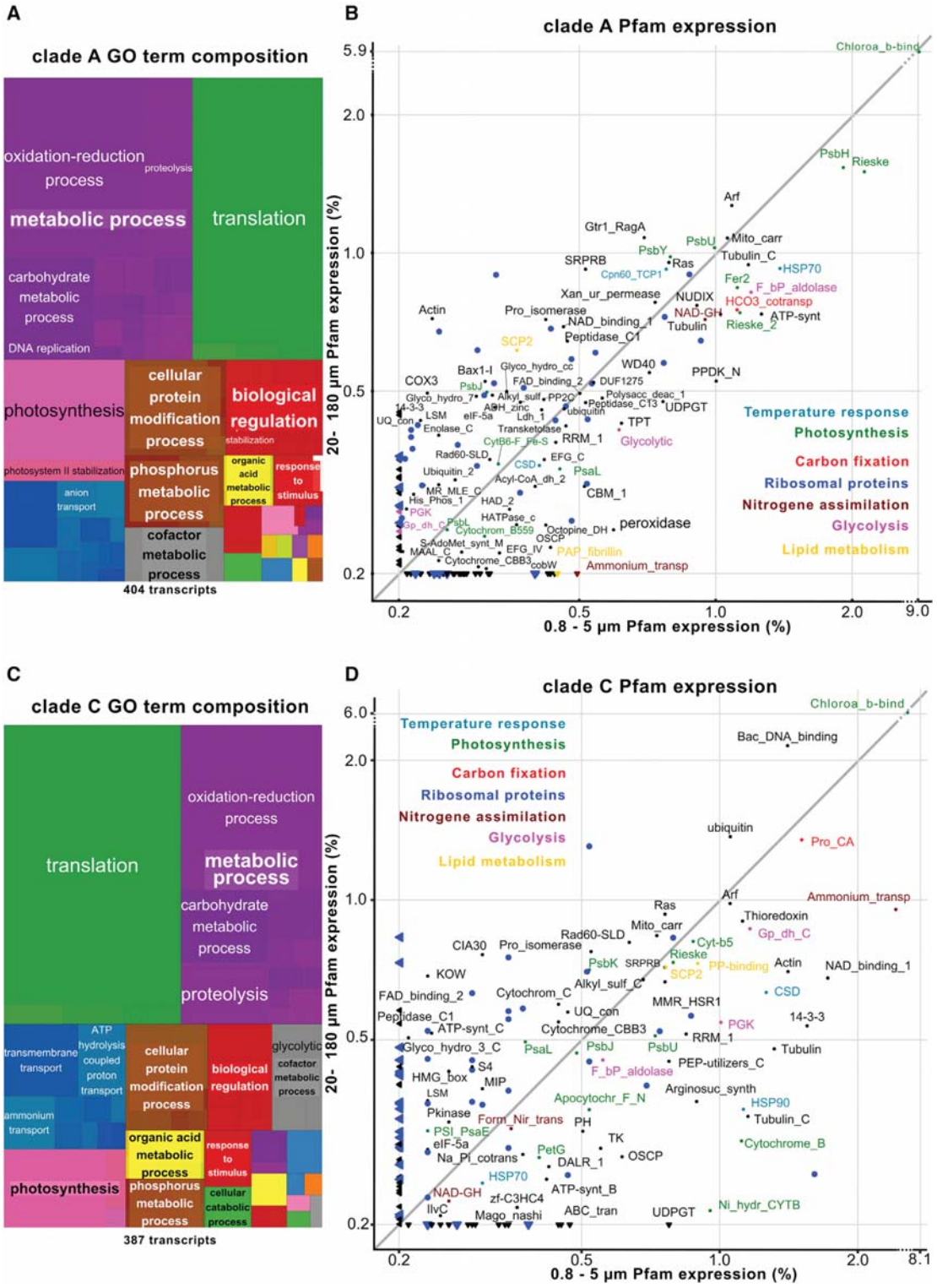

Figure 4. Functions Expressed in the Transcriptomes of *Symbiodinium* Clades A and C in the Open Ocean

(A and C) Treemaps present the most prevalent biological processes (GO terms) of unigenes from *Symbiodinium* clade A (A) and clade C (C). Unigenes without a GO term were removed from this analysis, and the number of unigenes with a GO term is indicated under the treemaps.

(legend continued on next page)

a high prevalence of bicarbonate- and ammonium-related transporter genes in comparison to other dinoflagellates [35].

Conclusions

This study provides a first global picture of the diversity and activity of *Symbiodinium* in the open ocean, outside the realms of shallow benthic reef ecosystems, at large geographical and macro-evolutionary scales (i.e., lineage level). Clades A and C constitute most of the *Symbiodinium* diversity and perform typical metabolic pathways of a microalga. The pervasive presence of *Symbiodinium* in the small size fraction, mainly clade A, suggests a free-living lifestyle in oceanic waters. Free-living cells may rely on their extensive genetic repertoire, mixotrophy, and physiological mechanisms [18, 35], as well as the metabolic pathways unveiled in our study. The most prevalent clade A is also the most ancestral *Symbiodinium* lineage [21], reinforcing the hypothesis that *Symbiodinium* originally inhabited the open ocean as an autonomous, free-living microalga [36], prior to colonizing reef ecosystems through the establishment of symbioses with benthic hosts during the Jurassic Period [2]. The dominance of clade A in the ocean may also explain why this clade is the most slowly evolving lineage among *Symbiodinium* lineages [21]. Immense exogenous populations in the oceans would allow increased frequency of genetic exchange through sexual recombination conserved in *Symbiodinium* [13, 37], as opposed to the clonal propagation of smaller populations in reefs experiencing recurrent bottlenecks *in hospite*. Clades A and C may also be present in the open ocean as endosymbionts within planktonic hosts, such as the ciliate *Tiarina*, jellyfishes, or coral and anemone larvae [23]. The free-living and symbiotic stages of *Symbiodinium* will have to be confirmed by microscopy observations. Other *Symbiodinium* lineages (B, D, F, and G) were detected in oceanic waters at much lower abundance and may, thus, be more restricted to coastal benthic ecosystems, compared to clades A and C. We speculate that these rare clades may be drifting passively from freshly released symbiotic forms or through meroplanktonic larvae.

This study raises the question of whether *Symbiodinium* in the open ocean constitutes a beneficial reservoir of functional diversity that could potentially repopulate highly impacted reef ecosystems and, therefore, engage in maintaining biodiversity and resilience of this key habitat. While *Symbiodinium* clade C (*Cladocopium*) generally has a high rate of carbon fixation and subsequent translocation to hosts, some types of clade A are recognized as opportunistic or even parasitic, since they grow rapidly in corals recovering from bleaching events [38, 39]. Thus, future studies should investigate the dispersal of *Symbiodinium* clades in the ocean at finer taxonomic (i.e., ITS2 marker [24]), spatial, and temporal resolutions and also evaluate their capacity at establishing symbiotic partnerships upon arrival in reef ecosystems. Understanding of the global bentho-pelagic connectivity is critical for effective management of vital reef ecosystems in the face of anthropogenic threats.

STAR★METHODS

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

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Analysis of the V9 rDNA metabarcoding dataset of the Tara Oceans expedition
  - RNA extraction, PCR amplification and assignment of environmental plastidial 23S rRNA sequences of Symbiodinium
  - Transcriptomic analyses
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

ACKNOWLEDGMENTS

This work was supported by the project OCEANOMICS and France Génomique, which received funding from the French government, managed by the Agence Nationale de la Recherche, under the grant agreement "Investissement d'Avenir" (grants ANR-11-BTBR-0008 and ANR-10-INBS-09). J.D. has been supported by the European Union's Horizon 2020 research and innovation programme, under the Marie Sklodowska-Curie grant agreement 658442 (MSCA-IF-2014), and by the LabEx GRAL (ANR-10-LABX-49-01) and Pôle CBS from the University of Grenoble Alpes. We thank the coordinators and members of the *Tara* Oceans consortium and Giovanni Finazzi and Ian Probert for critical reading of the manuscript. This article is contribution number 78 of *Tara* Oceans.

AUTHOR CONTRIBUTIONS

Conceptualization, J.D. and C.d.V.; Methodology, J.D. and S.R.; Funding Acquisition, C.d.V. and P.W.; Formal Analysis, J.D., F.M., N.H., Q.C., and A.K.; Investigation, J.D., C.d.V., F.M., N.H., Q.C., and X.P.; Visualization, J.D., N.H., and Q.C.; Writing – Original Draft, J.D. and Q.C.; Writing – Review & Editing, J.D., Q.C., M.D., X.P., C.R.V., and C.d.V.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Stanley, G.D.J., Jr. (2006). Ecology. Photosymbiosis and the evolution of modern coral reefs. Science *312*, 857–858.

(B and D) Correlation of Pfam domain expression between the 0.8–5-μm and the 20–180-μm size fractions for *Symbiodinium* clade A (B) and clade C (D). The level of expression for each Pfam domain was computed as percentage of all reads mapped on the transcriptome for each sample (see also Table S4). All samples from the same size fraction were grouped, and the mean value for each Pfam was kept. Pfam domains with an expression lower than 0.02% in one size fraction are indicated with an arrow on the border of each graph. Dots in the graph were colored according to their biological function.
See also Table S4.

2. LaJeunesse, T.C., Parkinson, J.E., Gabrielson, P.W., Jeong, H.J., Reimer, J.D., Voolstra, C.R., and Santos, S.R. (2018). Systematic revision of Symbiodiniaceae highlights the antiquity and diversity of coral endosymbionts. Curr. Biol. *28*, 2570–2580.e6.
3. Carradec, Q., Pelletier, E., Da Silva, C., Alberti, A., Seeleuthner, Y., Blanc-Mathieu, R., Lima-Mendez, G., Rocha, F., Tirichine, L., Labadie, K., et al.; Tara Oceans Coordinators (2018). A global ocean atlas of eukaryotic genes. Nat. Commun. *9*, 373.
4. Costanza, R., d'Arge, R., de Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neill, R.V., Paruelo, J., et al. (1997). The value of the world's ecosystem services and natural capital. Nature *387*, 253–260.
5. Moberg, F., and Folke, C. (1999). Ecological goods and services of coral reef ecosystems. Ecol. Econ. *29*, 215–233.
6. Hoegh-Guldberg, O. (1999). Climate change, coral bleaching and the future of the world's coral reefs. Mar. Freshw. Res. *50*, 839–866.
7. Hughes, T.P., Anderson, K.D., Connolly, S.R., Heron, S.F., Kerry, J.T., Lough, J.M., Baird, A.H., Baum, J.K., Berumen, M.L., Bridge, T.C., et al. (2018). Spatial and temporal patterns of mass bleaching of corals in the Anthropocene. Science *359*, 80–83.
8. Thornhill, D.J., Howells, E.J., Wham, D.C., Steury, T.D., and Santos, S.R. (2017). Population genetics of reef coral endosymbionts (*Symbiodinium*, Dinophyceae). Mol. Ecol. *26*, 2640–2659.
9. Coffroth, M.A., and Santos, S.R. (2005). Genetic diversity of symbiotic dinoflagellates in the genus *Symbiodinium*. Protist *156*, 19–34.
10. Coffroth, M.A., Lewis, C.F., Santos, S.R., and Weaver, J.L. (2006). Environmental populations of symbiotic dinoflagellates in the genus *Symbiodinium* can initiate symbioses with reef cnidarians. Curr. Biol. *16*, R985–R987.
11. Pochon, X., Stat, M., Takabayashi, M., Chasqui, L., Chauka, L.J., Logan, D.D.K., and Gates, R.D. (2010). Comparison of endosymbiotic and free-living *Symbiodinium* (Dinophyceae) diversity in a Hawaiian reef environment. J. Phycol. *46*, 53–65.
12. LaJeunesse, T.C. (2002). Investigating the biodiversity, ecology, and phylogeny of endosymbiotic dinoflagellates in the genus *Symbiodinium* using the its region: in search of a "species" level marker. J. Phycol. *37*, 866–880.
13. Chi, J., Parrow, M.W., and Dunthorn, M. (2014). Cryptic sex in *Symbiodinium* (Alveolata, Dinoflagellata) is supported by an inventory of meiotic genes. J. Eukaryot. Microbiol. *61*, 322–327.
14. Baird, A.H., Bhagooli, R., Ralph, P.J., and Takahashi, S. (2009). Coral bleaching: the role of the host. Trends Ecol. Evol. *24*, 16–20.
15. Wirshing, H.H., Feldheim, K.A., and Baker, A.C. (2013). Vectored dispersal of *Symbiodinium* by larvae of a Caribbean gorgonian octocoral. Mol. Ecol. *22*, 4413–4432.
16. Shoguchi, E., Shinzato, C., Kawashima, T., Gyoja, F., Mungpakdee, S., Koyanagi, R., Takeuchi, T., Hisata, K., Tanaka, M., Fujiwara, M., et al. (2013). Draft assembly of the *Symbiodinium minutum* nuclear genome reveals dinoflagellate gene structure. Curr. Biol. *23*, 1399–1408.
17. Aranda, M., Li, Y., Liew, Y.J., Baumgarten, S., Simakov, O., Wilson, M.C., Piel, J., Ashoor, H., Bougouffa, S., Bajic, V.B., et al. (2016). Genomes of coral dinoflagellate symbionts highlight evolutionary adaptations conducive to a symbiotic lifestyle. Sci. Rep. *6*, 39734.
18. Jeong, H.J., Yoo, Y.D., Kang, N.S., Lim, A.S., Seong, K.A., Lee, S.Y., Lee, M.J., Lee, K.H., Kim, H.S., Shin, W., et al. (2012). Heterotrophic feeding as a newly identified survival strategy of the dinoflagellate *Symbiodinium*. Proc. Natl. Acad. Sci. USA *109*, 12604–12609.
19. de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., Probert, I., et al.; Tara Oceans Coordinators (2015). Ocean plankton. Eukaryotic plankton diversity in the sunlit ocean. Science *348*, 1261605.
20. Pochon, X., and Gates, R.D. (2010). A new *Symbiodinium* clade (Dinophyceae) from soritid foraminifera in Hawai'i. Mol. Phylogenet. Evol. *56*, 492–497.
21. Pochon, X., Putnam, H.M., and Gates, R.D. (2014). Multi-gene analysis of *Symbiodinium* dinoflagellates: a perspective on rarity, symbiosis, and evolution. PeerJ *2*, e394.
22. Pochon, X., Montoya-Burgos, J.I., Stadelmann, B., and Pawlowski, J. (2006). Molecular phylogeny, evolutionary rates, and divergence timing of the symbiotic dinoflagellate genus *Symbiodinium*. Mol. Phylogenet. Evol. *38*, 20–30.
23. Mordret, S., Romac, S., Henry, N., Colin, S., Carmichael, M., Berney, C., Audic, S., Richter, D.J., Pochon, X., de Vargas, C., and Decelle, J. (2016). The symbiotic life of *Symbiodinium* in the open ocean within a new species of calcifying ciliate (*Tiarina* sp.). ISME J. *10*, 1424–1436.
24. Hume, B.C.C., Ziegler, M., Poulain, J., Pochon, X., Romac, S., Boissin, E., de Vargas, C., Planes, S., Wincker, P., and Voolstra, C.R. (2018). An improved primer set and amplification protocol with increased specificity and sensitivity targeting the *Symbiodinium* ITS2 region. PeerJ *6*, e4816.
25. Stoeck, T., Kasper, J., Bunge, J., Leslin, C., Ilyin, V., and Epstein, S. (2007). Protistan diversity in the Arctic: a case of paleoclimate shaping modern biodiversity? PLoS ONE *2*, e728.
26. Pochon, X., Zaiko, A., Fletcher, L.M., Laroche, O., and Wood, S.A. (2017). Wanted dead or alive? Using metabarcoding of environmental DNA and RNA to distinguish living assemblages for biosecurity applications. PLoS ONE *12*, e0187636.
27. Takabayashi, M., Adams, L.M., Pochon, X., and Gates, R.D. (2012). Genetic diversity of free-living *Symbiodinium* in surface water and sediment of Hawai'i and Florida. Coral Reefs *31*, 157–167.
28. Jeong, H.J., Lee, S.Y., Kang, N.S., Yoo, Y.D., Lim, A.S., Lee, M.J., Kim, H.S., Yih, W., Yamashita, H., and LaJeunesse, T.C. (2014). Genetics and morphology characterize the dinoflagellate *Symbiodinium voratum*, n. sp., (Dinophyceae) as the sole representative of *Symbiodinium* clade E. J. Eukaryot. Microbiol. *61*, 75–94.
29. Stat, M., Carter, D., and Hoegh-Guldberg, O. (2006). The evolutionary history of *Symbiodinium* and scleractinian hosts—symbiosis, diversity, and the effect of climate change. Perspect. Plant Ecol. Evol. Syst. *8*, 23–43.
30. Manning, M.M., and Gates, R.D. (2008). Diversity in populations of free-living *Symbiodinium* from a Caribbean and Pacific reef. Limnol. Oceanogr. *53*, 1853–1861.
31. Hirose, M., Reimer, J.D., Hidaka, M., and Suda, S. (2008). Phylogenetic analyses of potentially free-living *Symbiodinium* spp. isolated from coral reef sand in Okinawa, Japan. Mar. Biol. *155*, 105–112.
32. Granados-Cifuentes, C., Neigel, J., Leberg, P., and Rodriguez-Lanetty, M. (2015). Genetic diversity of free-living *Symbiodinium* in the Caribbean: the importance of habitats and seasons. Coral Reefs *34*, 927–939.
33. Ladner, J.T., Barshis, D.J., and Palumbi, S.R. (2012). Protein evolution in two co-occurring types of *Symbiodinium*: an exploration into the genetic basis of thermal tolerance in *Symbiodinium* clade D. BMC Evol. Biol. *12*, 217.
34. Shinzato, C., Mungpakdee, S., Satoh, N., and Shoguchi, E. (2014). A genomic approach to coral-dinoflagellate symbiosis: studies of *Acropora digitifera* and *Symbiodinium minutum*. Front. Microbiol. *5*, 336.
35. Lin, S., Cheng, S., Song, B., Zhong, X., Lin, X., Li, W., Li, L., Zhang, Y., Zhang, H., Ji, Z., et al. (2015). The *Symbiodinium kawagutii* genome illuminates dinoflagellate gene expression and coral symbiosis. Science *350*, 691–694.
36. Shaked, Y., and de Vargas, C. (2006). Pelagic photosymbiosis: rDNA assessment of diversity and evolution of dinoflagellate symbionts and planktonic foraminiferal hosts. Mar. Ecol. Prog. Ser. *325*, 59–71.
37. Wilkinson, S.P., Fisher, P.L., van Oppen, M.J., and Davy, S.K. (2015). Intra-genomic variation in symbiotic dinoflagellates: recent divergence or recombination between lineages? BMC Evol. Biol. *15*, 46.
38. LaJeunesse, T.C., Lee, S.Y., Gil-Agudelo, D.L., Knowlton, N., and Jeong, H.J. (2015). *Symbiodinium necroappetens* sp. nov. (Dinophyceae): an opportunist "zooxanthella" found in bleached and diseased tissues of Caribbean reef corals. Eur. J. Phycol. *50*, 223–238.

39. Stat, M., Morris, E., and Gates, R.D. (2008). Functional diversity in coral-dinoflagellate symbiosis. Proc. Natl. Acad. Sci. USA *105*, 9256–9261.
40. Pesant, S., Not, F., Picheral, M., Kandels-Lewis, S., Le Bescot, N., Gorsky, G., Iudicone, D., Karsenti, E., Speich, S., Troublé, R., et al.; Tara Oceans Consortium Coordinators (2015). Open science resources for the discovery and analysis of Tara Oceans data. Sci. Data *2*, 150023.
41. Alberti, A., Poulain, J., Engelen, S., Labadie, K., Romac, S., Ferrera, I., Albini, G., Aury, J.-M., Belser, C., Bertrand, A., et al.; Genoscope Technical Team; Tara Oceans Consortium Coordinators (2017). Viral to metazoan marine plankton nucleotide sequences from the Tara Oceans expedition. Sci. Data *4*, 170093.
42. R Core Team (2018). R: A language and environment for statistical computing (R Foundation for Statistical Computing).
43. Mahé, F., Rognes, T., Quince, C., de Vargas, C., and Dunthorn, M. (2014). Swarm: robust and fast clustering method for amplicon-based studies. PeerJ *2*, e593.
44. Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. PeerJ *4*, e2584.
45. LaJeunesse, T.C., Smith, R.T., Finney, J., and Oxenford, H. (2009). Outbreak and persistence of opportunistic symbiotic dinoflagellates during the 2005 Caribbean mass coral 'bleaching' event. Proc. Biol. Sci. *276*, 4139–4148.
46. 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.
47. Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scand. J. Statist. *6*, 65–70.
48. Santos, S.R., Taylor, D.J., Kinzie, R.A., 3rd, Hidaka, M., Sakai, K., and Coffroth, M.A. (2002). Molecular phylogeny of symbiotic dinoflagellates inferred from partial chloroplast large subunit (23S)-rDNA sequences. Mol. Phylogenet. Evol. *23*, 97–111.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Symbiodinium cultures, Table S1	This study	N/A
Chemicals, Peptides, and Recombinant Proteins		
Total RNA extraction NucleoSpin RNA kit	Macherey-Nagel	Cat#740955
Turbo DNA-free kit	Ambion	Cat#AM1907
Reverse transcription using the RT Superscript III random primers kit	Invitrogen	Cat#18080051
Master Mix Phusion GC HighFidelity DNA Polymerase	Finnzymes	Cat#F553S
antIt PicoGreen double stranded DNA Assay kit	Invitrogen	Cat# P7589
NucleoSpin Gel and PCR CleanUp kit	Macherey-Nagel	Cat#740609
Deposited Data		
Ribosomal nuclear sequences (V9rDNA + ITS1) of cultured and uncultured *Symbiodinium* (provided in Table S1)	This paper	GenBank MH702341–MH702366
Environmental plastidial 23S rRNA sequences (Data S1)	This paper	EBI: PRJEB27503
V9 rDNA metabarcoding dataset of *Tara* Oceans	[19, 40, 41]	EBI: PRJEB16766
Abundance, expression, and Pfam of each unigene from the eukaryotic gene catalog	[3]	http://www.genoscope.cns.fr/tara/.
Oligonucleotides		
PCR Forward primers to amplify the V9 region of the 18S rDNA of *Symbiodinium* 1389F: 5′-TTGTACACACCGCCC-3′)	This paper.	N/A
PCR Reverse primers to amplify the V9 region of the 18S rDNA in the 5.8S (ITS_din_rev: 5′-GTGAATTGCCAGAACTCCGTG-3′)	This paper.	N/A
PCR forward primer of the hypervariable region of the chloroplast 23S domain V (cp23S-HVR, ~180 bp) 23SHYPERUP 5′- TCAGTACAAATAATATGCTG-3′	[30]	N/A
PCR reverse primer of the hypervariable region of the chloroplast 23S domain V (cp23S-HVR, ~180 bp) 23SHYPERDN 5′-TTATCGCCCCAATTAAACAGT-3′	[30]	N/A
Software and Algorithms		
R studio v:3.1.1	[42]	https://www.rstudio.com/
SWARM	[43]	https://github.com/torognes/swarm
vsearch v2.3.1	[44]	https://github.com/torognes/vsearch
R library ggplot2	Open access.	https://cran.r-project.org/web/packages/ggplot2/index.html
R library igraph	Open access	https://cran.r-project.org/web/packages/igraph/index.html
R library venneuler	Open access	https://cran.r-project.org/web/packages/venn/index.html
R library rgeos	Open access	https://cran.r-project.org/web/packages/rgeos/index.html
R library maptools	Open access	https://cran.r-project.org/web/packages/maptools/index.html
R library scatterpie	Open access	https://cran.r-project.org/web/packages/scatterpie/vignettes/scatterpie.html
R library Outliers	Open access	https://cran.r-project.org/web/packages/outliers/index.html

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Johan Decelle (johan.decelle@cea.fr).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

DNA from different *Symbiodinium* clades was used to construct the reference database of the rDNA V9 marker. *Symbiodinium* isolates were obtained from cultures (grown at 22°C in F/2 medium and 100 μmol quanta.m^{-2}.s^{-1}) at the Roscoff Culture Collection (http://roscoff-culture-collection.org/), and from a variety of hosts, such as Foraminifera (Table S1).

METHOD DETAILS

Analysis of the V9 rDNA metabarcoding dataset of the Tara Oceans expedition

Metabarcoding is particularly relevant in the case of *Symbiodinium*, since its putatively low natural abundance *ex hospite* and the lack of discernible morphological characters make the detection of this microalga within environmental microbial assemblages impossible using traditional microscopy [45]. We analyzed 3,575,329 rDNA metabarcodes (i.e., V9 region of the nuclear 18S rDNA) that are publicly available (EBI accession number PRJEB16766) from 385 size-fractionated plankton communities collected in surface waters at 121 stations across the world's oceans during the *Tara* Oceans expedition (see [19, 40, 41] for more details on the sampling and data processing). Sampling included reef waters (in the water column of reef ecosystems) from the Indian and Pacific Oceans, and oceanic waters (outside reefs, in the open ocean). Three plankton size fractions were considered: *piconano* (0.8-5 μm), *micro* (20-180 μm), and *meso* (180-2000 μm). For each community of eukaryotes (representing on average 1,128,579 V9 rDNA reads), the V9 metabarcoding dataset was interrogated in order to extract and taxonomically assign *Symbiodinium* reads. To do so, a dedicated V9 reference database was built based on a curated alignment of 107 public 18S rDNA sequences of *Symbiodinium* (obtained from Pr2 database [46]) and newly produced sequences from different *Symbiodinium* clades. To obtain these sequences, the V9 region was PCR sequenced from DNA extracts of cultured and uncultured *Symbiodinium* using a forward primer in the 18S rDNA gene (1389F: 5′-TTGTACACACCGCCC-3′) and a reverse primer in the 5.8S (ITS_din_rev: 5′-GTGAATTGCCAGAACTCCGTG-3′) using the same PCR conditions described below and as in [23]. The reference database contained 18 distinct V9 rDNA sequences representing the nine *Symbiodinium* clades described to date (A-I) [20], as well as intra-clade variants, such as D1 and D2 [21] (Table S1 and Figure S1). Although displaying few polymorphic sites, the V9 rDNA region can be effectively used as a clade-level barcode for *Symbiodinium* environmental surveys. Read extraction and assignment were performed using a conservative nucleic identity threshold of $\geq$ 99% to reference *Symbiodinium* V9 rDNA sequences on a global alignment (vsearch v2.3.1 using the–usearch_global command with–iddef 1). The different graphs presenting V9 rDNA metabarcoding data have been generated with the R libraries ggplot2 and igraph [42]. In order to assess co-occurrence patterns between *Symbiodinium* clades A and C and putative hosts (e.g., the ciliate *Tiarina*, Anthozoa, Scyphozoa, and Foraminifera) in large size fractions (20-180 μm and 180-2000 μm), multiple pairwise comparisons were performed as in [23], using Spearman's rank correlation tests based on the relative abundances of their V9 rDNA reads (Data S2). P values were adjusted for multiple comparisons using the Holm method [47]. Correlations between groups (i.e., *Symbiodinium* clades A and C, and their putative hosts) were considered significant when p values were < 0.05.

RNA extraction, PCR amplification and assignment of environmental plastidial 23S rRNA sequences of Symbiodinium

In order to further investigate the presence of live *Symbiodinium* cells in the ocean, the hypervariable region of the chloroplast 23S domain V (cp23S-HVR) was amplified and sequenced from RNA-derived template. This genetic marker was selected because of the high specificity of the cp23S-HVR primers for distinguishing *Symbiodinium* from other closely related dinoflagellates in complex environmental samples [11, 30]. An additional advantage of this coding locus is that it may be directly linked to physiological performance [48]. Total RNA was extracted using the NucleoSpin RNA kit (Macherey-Nagel) from the *pico-nano* (0.8-5 μm) and *nano* (5-20 μm) plankton size fractions of 16 communities from oligotrophic surface waters of the Indian, North Atlantic and South Pacific Oceans (stations 41, 42, 47, 52, 94, 95, 97, 98, 122, 123, 124, 125, 148, 149, 150 and 151). Extracts were quantified and checked on a 1.5% agarose gel. Contaminating DNA was removed from RNA extracts using the Turbo DNA-free kit (Ambion). Complete DNA removal was verified by PCR using eukaryotic primers, and in case of positive PCR, an additional DNase treatment was performed. One hundred nanograms of extracted RNA were immediately reverse transcribed into cDNA using the RT Superscript III random primers kit (Invitrogen) and stored at −80°C. Environmental cDNA products were used as templates for PCR amplification of the hypervariable region of the chloroplast 23S domain V (cp23S-HVR, ~180 bp) using *Symbiodinium*-specific primers forward 23SHYPERUP 5′-TCAGTACAAATAATATGCTG-3′ and reverse 23SHYPERDN 5′-TTATCGCCCCAATTAAACAGT-3′ [30]. The forward primer was built with a 8 mers sequence tag on its 5′ extremity, and both primers were adapted for Illumina™ sequencing. PCR reactions (25 μl) contained 1x Master Mix Phusion GC HighFidelity DNA Polymerase (Finnzymes), 0.35 μM of each primer, 3% dimethylsulphoxide and 1 μL of cDNA. The PCR program had an initial denaturation step at 98°C during 30 s, 40 cycles of 10 s at 98°C, 30 s at 50°C and 30 s at 72°C, and a final step at 72°C for 10 min. Samples from the *piconano* and the *nano* size fractions from the same stations were amplified with the same sequence-tag in triplicates, and then pooled (0.8-20 μm). PCR products were purified and eluted (20 μl) with NucleoSpin Gel and PCR CleanUp kit (MachereyNagel), and quantified with the QuantIt PicoGreen double stranded DNA Assay kit (Invitrogen). About 1 μg of pooled amplicons were sent to Fasteris (Switzerland) for high throughput sequencing on a 2x100bp HiSeq Illumina. The accession number for the raw plastidial 23S rRNA reads reported in this paper is EBI: PRJEB27503.

Exact sequence matching was used to demultiplex the raw fastq files and to trim primers. Paired-end reads that did not cover the complete region were trimmed at 68 bp and 79 bp for R1 and R2 reads, respectively, and concatenated. The same process was applied to the reference 23S rRNA sequences, described below. Sequence dereplication was performed with vsearch [44],

and clustering of environmental 23S rRNA metabarcodes with Swarm [43], using the fastidious option and default parameters. Taxonomic assignment was performed with global pairwise alignment (as implemented in vsearch [44]) against a reference database of *Symbiodinium* modified from [27] (Data S1). This database contains 141 plastidial 23S rDNA sequences of diverse *Symbiodinium* clades and subclades obtained in the symbiotic and free-living stages, as well as sequences from diverse phytoplankton lineages (i.e., outgroups), such as other dinoflagellates, chlorophytes, rhodophytes, cryptophytes, land plants, diatoms, and cyanobacteria (Data S1).

Transcriptomic analyses

Transcript homologies between 23 reference transcriptomes of *Symbiodinium* (Table S3) and the eukaryotic unigene catalog of *Tara* Oceans [3] were detected with a nucleic BLAST v:2.2.28. Sequence alignments with more than 95% of identity and over 80% of the unigene length were kept. Abundance and expression of each unigene in each *Tara* Oceans sample were obtained here http://www.genoscope.cns.fr/tara/. Expression values of unigenes were computed as a percentage of all mapped reads for each *Tara* Oceans sample. Selected unigenes were filtered in order to remove outliers (contaminants, cross-mapped and/or high copy number genes) with two distinct methods: 1- unigenes with atypical metagenomic abundance in each sample relative to other unigenes of the same transcriptome (R v:3.1.1 package "outlier," z-score method, 99% of confidence). 2- unigenes expressed in too many samples in comparison to all unigenes of a *Symbiodinium* transcriptome (same tool and parameters). After these two methods, 84 unigenes were eliminated for clade A and 56 for clade C. The Euler diagram of Figure 3 was made with "venneuler" package v:1.1 after removing *Symbiodinium* references matching less than 300 unigenes. World maps were drawn with the following R packages: rgeos, maptools, scatterpie, and ggplot2.

Unigene Pfam annotations (available at http://www.genoscope.cns.fr/tara/) from the eukaryotic gene catalog were used to describe functional activity of *Symbiodinium*. We then used the Pfam2GO correspondence table (v 01/2018) to attribute a Gene Ontology (GO) term to unigenes when possible. Results were represented with the R package treemap using two levels of classification. For each transcriptome, unigene expression values normalized as the percentage of all reads mapped for each transcriptome were added up by Pfam domain for each sample. All samples from the same size-fraction were grouped and the mean value for each Pfam domain was kept. These normalized expression values are given in Table S4.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis and graph production were performed using R with specific libraries mentioned above and in the legends of the related figures and tables (e.g., ggplot2, igraph, rgeos, maptools). In total, 324 metabarcoding samples (the nuclear V9 rDNA barcode) have been analyzed. N numbers were as follows: 102 from the *piconano* size fraction (0.8-5 μm), 114 from the *micro* size fraction (20-180 μm), and 108 from the *meso* size fraction (180-2000 μm) (Figures 1 and 2). For the co-occurrence analyses, multiple pairwise comparisons were performed using Spearman's rank correlation tests (Data S2). P values were adjusted for multiple comparisons using the Holm method and correlations between groups were considered significant when p values were < 0.05. For the metatranscriptomics analysis, 308 samples have been analyzed. N numbers were as follows: 66 for the *piconano* −0.8 −5 μm - size fraction, 49 for the 0.8 – 2000 μm size fraction, 63 for the 5 −20 μm size fraction, 63 for the *micro* 20-180 μm size fraction, and 67 for the *meso* 180 – 2000 μm size fraction (Figures 3 and 4).

DATA AND SOFTWARE AVAILABILITY

The accession numbers for the ribosomal nuclear sequences (V9 rDNA and ITS1) and the environmental plastidial 23S rRNA sequences of *Symbiodinium* reported in this paper are GenBank: MH702341–MH702366 and EBI PRJEB27503, respectively.