Plankton Planet: 'seatizen' oceanography to assess open ocean life at the planetary scale

Colomban de Vargas[1,2,3] [#], Thibaut Pollina[2,4], Sarah Romac[1,2,3], Noan Le Bescot[1,2], Nicolas Henry[1,2,3], Calixte Berger[2], Sébastien Colin[1,2], Nils Haëntjens[5,2], Margaux Carmichael[2], David Le Guen[2], Johan Decelle[6], Frédéric Mahé[7], Emmanuel Malpot[8], Carole Beaumont[9], Michel Hardy[10], *the planktonauts*, the *Plankton Planet team*, Damien Guiffant[2], Ian Probert[1], David F. Gruber[11], Andy Allen[12], Gabriel Gorsky[13,2], Mick Follows[14], Barry B. Cael[15], Xavier Pochon[16,17], Romain Troublé[18,2] [#], Fabien Lombard[2,13,19], Emmanuel Boss[5,2], Manu Prakash[4,2] [#]

[1] Sorbonne Université, CNRS, Station Biologique de Roscoff, UMR7144, ECOMAP, 29680 Roscoff, France.
[2] Plankton Planet NGO, Station Biologique de Roscoff & Atelier PontonZ Morlaix, 29680 Roscoff, France
[3] Research Federation for the study of Global Ocean Systems Ecology and Evolution, FR2022/Tara GOSEE, Paris, France
[4] Stanford University, Department of Bioengineering, Stanford, CA 94305, USA.
[5] University of Maine, School of Marine Sciences, 5706 Aubert Hall, Orono, ME 04473, USA
[6] Laboratoire de Physiologie Cellulaire et Végétale, Université Grenoble Alpes, CNRS, CEA, INRA; 38054, Grenoble, France
[7] CIRAD, UMR GBPI, 34398, Montpellier, France
[8] Moana Fisheries Ltd, Cawthron Aquaculture Park, Nelson, New Zealand
[9] On board 'Folligou'
[10] On board 'Taravana'
[11] Baruch College and the Graduate Center, Department of Natural Sciences, City University of New York, USA
[12] J. Craig Venter Institute, Microbial and Environmental Genomics Group, La Jolla, CA, USA & Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA, USA.
[13] Sorbonne Université, CNRS, Laboratoire d'Océanographie de Villefranche, 06230 Villefranche-sur-mer, France.
[14] Department of Earth, Atmosphere and Planetary Sciences, MIT, Cambridge, MA, USA.
[15] Ocean Biogeochemistry and Ecosystems, National Oceanography Centre, Southampton, UK.
[16] Coastal and Freshwater Group, Cawthron Institute, Nelson, New Zealand
[17] Department of Civil and Environmental Engineering, University of Auckland, Auckland, New Zealand
[18] Tara Ocean Foundation, 75004 Paris, France
[19] Institut Universitaire de France (IUF), Paris, France

[#] Correspondence should be addressed to C.d.V. (c2vargas@gmail.com), M.P. (manup@stanford.edu), and R.T. (romain@taraexpeditions.or)

Abstract

In every liter of seawater there are between 10 and 100 billion life forms, mostly invisible, called plankton, which form the largest and most dynamic ecosystem on our planet, at the heart of global ecological and economic processes. While physical and chemical parameters of planktonic ecosystems are fairly well measured and modelled at the planetary scale, but biological data are still scarce due to the extreme cost and relative inflexibility of the classical vessels and instruments used to explore marine biodiversity. Here we introduce '*Plankton Planet*', an initiative whose goal is to merge the creativity of researchers, makers, and mariners to (*i*) develop frugal scientific instrumentation and protocols to assess the genetic and morphological diversity of plankton life, and (*ii*) organize their systematic deployment through fleets of volunteer sailors, fishermen, or cargo-ships to generate comparable and open-access plankton data across global and long-term spatio-temporal scales. As proof-of-concept, we show how 20 crews of sailors ("planktonauts") were abl to sample plankton biomass from the world surface ocean in a single year, generating the first citizen-based, planetary dataset of plankton biodiversity based on DNA barcodes. The quality of this dataset is comparable to that generated by *Tara* Oceans and is not biased by the multiplication of samplers. This dataset has unveiled significant genetic novelty and can be used to explore the taxonomic and ecological diversity of plankton at both regional and global scales. This pilot project paves the way for construction of a miniaturized, modular, evolvable, affordable and open-source citizen field-platform that will allow systematic assessment of the eco/morpho/genetic variation of aquatic ecosystems across the dimensions of the Earth system.

Global surveys of marine plankton life: urgency and challenges

The ocean contains 97% of all water on our planet. In every liter of seawater there are between 10 and 100 billion, mostly invisible planktonic life forms. These form a continuous global ecosystem that generates approximately half of planetary oxygen, sustains the large majority of marine life, and regulates atmospheric CO_2 and climate. Understanding and modelling the structure, dynamics, and evolution of global plankton populations is critical for predicting the future of our biosphere and learning how to live in symbiosis with our spaceship, the Earth.

Plankton populations comprise, like in terrestrial biomes, organisms from across the tree of life (viruses, bacteria, archaea, protists, and animals) which interact in complex networks (*1*, *2*) that are also shaped by ocean currents and associated physico-chemical environmental parameters (*3*, *4*) – i.e. the *seascape* (*5*). But in contrast to terrestrial ecosystems, there are no plants and trees in the plankton - primary production is driven by a large and ancient diversity of photosynthetic bacteria and protists (called phytoplankton) - and the pelagic ecosystem is much more dynamic in terms of both organism life cycle and strategies (e.g. mixotrophy) (*6*) and transport (advection and mixing). The self-organization of local plankton biota into complex ecosystems (*7*) determines their impact on the carbon cycle. For instance, in some regions and seasons, blooms of relatively large cells with mineral components lead to a vigorous sinking flux of organic carbon into the deep sea (*8*, *9*). Overall, these fundamental properties of the plankton ecosystem make it arguably the most reactive and proactive compartment of the biosphere to climate change and pollution. Changes in plankton communities can have dramatic effects on global biogeochemical cycles (e.g. *10*), climate (e.g. *11*, *12*), and major human societal and economic activities (i.e. fisheries, aquaculture, tourism, etc. (*13*, *14*).

Current models aimed at predicting global ocean ecological changes (e.g. (*15, 16*)) are fairly well constrained in terms of physics and chemistry, but are by comparison heavily oversimplified and unrealistic in terms of biology. In fact, they simply lack good quality, high resolution data on the nature and dynamics of oceanic plankton biodiversity *at a planetary scale*. While quantitative global data on ocean physics and biogeochemistry are abundantly available by satellites (e.g. *17*), *in situ* floats (e.g.*18*, *19*)) and sail drones (e.g. *20*), as well as research (e.g. *21*) and citizen (e.g. *22*) vessels, standardized biological data are still scarce due to the significant challenge of sampling and assessing complex communities of fragile plankton in a harmonized and comparable manner (*23*). Despite the availability of a century's worth of recorded data on oceanic plankton ((*24*, *25*), and see for instance https://www.st.nmfs.noaa.gov/copepod/)), some areas of the ocean (including much of the south and tropical Pacific Ocean) are almost devoid of any biological observations and we do not yet have an informed vision of the global distribution and variation of plankton communities.

The largest and longest homogenous survey of plankton life has been conducted using the Continuous Plankton Recorder (CPR), a visionary plankton-scroll instrument created by Sir Alister Hardy in 1931 (*26*), and towed since then behind ferries and cargo ships, particularly in the North Atlantic, North Pacific, and Southern Ocean south of Australia (*27*). The CPR database is currently the only basin-wide standardized historical record of ocean plankton life, and it has given rise to keystone studies describing and modelling basin-scale dynamics of plankton community over time and global climate change (e.g. *13*, *28*, *29*). However, CPR data also has major drawbacks: (*i*) the instrument mainly recovers zooplankton >300μm, thus missing most plankton diversity (see below), (*ii*) it is rather destructive for soft or gelatinous taxa, (*iii*) data analyses rely on taxonomic experts identifying and counting a restricted number of morpho-taxa, (*iv*) the formalin preservation of the silk severely limits subsequent light microscopy observations for some groups and analysis of nucleic acids in general, and (*v*) data come from ships navigating over a restricted number of (mostly northern) commercial routes. CPR data are therefore semi-quantitative and taxonomically and geographically limited.

Over the past 20 years, the revolution in environmental DNA/RNA sequencing has stimulated a new era of global-scale open ocean cruises led by molecular and cellular biologists, notably the 'Global Ocean Sampling' (GOS - *30*), *Tara* Oceans (*31*), and Malaspina (*32*) expeditions. Interestingly, the first two expeditions were private or semi-private enterprises undertaken by sailing boats. The *Tara* Oceans sailing expedition was the longest (2009 – 2013) and most comprehensive: its team developed an eco-systems biology strategy to explore plankton diversity from genes to communities, from viruses to animals, and across coarse but planetary spatial and seasonal scales (*33*). The combination of *standardized* DNA metabarcoding (*34, 35*), metagenomic (*36, 37*) and metatranscriptomic (*38, 39*) datasets has unveiled the basic structure of plankton taxonomic and metabolic diversity (*33, 40*), generating hypotheses about its interactions (*1*), dynamics in the seascape (*3*), and role in emerging ecosystem functions such as the carbon pump (*2, 41*). Notably, the *Tara* Oceans team discovered that the great majority of plankton biodiversity is found in organismal size fractions <100μm, and above all in eukaryotes rather than viruses or prokaryotes (*34, 38*).

Given the enormous local (e.g. plastics, pollutants (*42, 43*)) and global (e.g. deoxygenation, warming, acidification, freshening, ocean circulation changes (*44, 45*)) anthropogenic pressures on our ocean, we urgently need comprehensive global plankton surveys that merge the spatio-temporal sampling power of the CPR with the systems-biology approach of *Tara* Oceans. Only application of a

comprehensive and homogenous measure of plankton life from micro- to meso- to global ocean scales (*7*, *23*, *46*) will provide the data necessary not only to unveil fundamental principles of ecology and evolution of marine life at the ecosystem level of organization, but also to feed mathematical models of the ocean system that integrate physical, chemical, and biological processes.

Plankton Planet: toward a participative Oceanography 3.0

However, this endeavor is hindered by the extreme cost and limited logistical flexibility of classical oceanographic research vessels and instruments, together with the current impossibility to use autonomous samplers (e.g. floats) to sample the biocomplexity of plankton and generate high-quality data from it. In this context, the thousands of citizen sailing boats (*47*), professional sailing yachts, the >50,000 cargo ships (https://www.ics-shipping.org/shipping-facts/shipping-and-world-trade) and about the same number of global fishing vessels (https://globalfishingwatch.org/datasets-and-code/vessel-identity/) which are navigating the world ocean every day represent an outstanding opportunity.

With the miniaturization of sequencing (e.g. *48*) and imaging (e.g. *49*) devices, the power of cloud computing and artificial intelligence, and the possibilities offered by a participative science engaging actors and volunteers across disciplines and societal/economic activities (*22*, *50*), five years ago we - a team of researchers, engineers, makers and sailors from France, the US, and New Zealand - started to develop a frugal approach and effective protocols to sample the world ocean plankton for the production of high quality eco/morpho/genetic data in collaboration with recreational and professional mariners. We created '*Plankton Planet*' (P2, https://planktonplanet.org), an international initiative that promotes long-term citizen-based sampling of marine plankton to assess the biodiversity and health of the world's open and coastal oceans (see P2 vision & mission, box 1).

BOX 1. Plankton Planet Vision & Mission

Vision: to harness the creativity of mariners, scientists, and makers, for global evaluation of ocean health, biodiversity, and evolution.

Mission: to develop user-friendly affordable tools to underpin a sustainable citizen Oceanography 3.0 providing critical new knowledge on global plankton morphology, genetics, and ecology, that will be universally accessible and feed mathematical modelling toward a predictable ocean.

The Oceanography 3.0 is independent and universal, based on interconnected individuals and instruments.
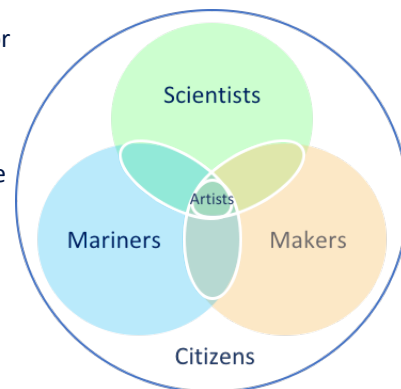
We strongly believe that mariners, makers, and researchers share a passionate curiosity and will to explore and discover their environment, such that erasing the boundaries between these worlds will generate the necessary synergies to achieve our objectives. Our practical goal is to co-construct a new generation of frugal yet robust scientific instruments and protocols, which will allow all volunteer mariners – from recreative or professional sailors to fishermen or cargo-ship crews - to

collect comparable eco-morpho-genetic data on plankton diversity and abundance at a planetary scale, a concept coined *Oceanography 3.0* (Box 1).

In this paper, we present the first steps of this cost-effective, eco-friendly, and society-engaging Oceanography 3.0. We show how the recruitment of the first 20 crews of citizen sailors – the planktonauts-, equipped with a simple kit to sample plankton for DNA-metabarcoding, were able to help generate a scientifically sound, planetary dataset of plankton biodiversity in less than a year. We finally discuss recent developments (see also (*51*)), as well as the 'Plankton Planet' perspectives toward the large-scale deployment of integrated, affordable and portable 'field-eco-scopes' for long-term assessment of ocean life and ecosystems at an unprecedented level of sensitivity.

Plankton Planet: Proof-of-Concept

Between November 2014 and January 2016, with a total budget of $70,000, we provided proof-of-concept for the P2 vision and mission (Box 1). Our primary goal was to demonstrate that, with the goodwill of mariners, we can sample ocean waters from the entire planet and obtain high quality plankton DNA data for global ecological analyses. We designed a general functional strategy (Sup. Fig.1) based on robust methods linking affordable and user-friendly instruments for on-board citizen plankton sampling to cutting-edge DNA sequencing and bioinformatic pipelines developed previously in *Tara* Oceans (*34*).

Frugal and global plankton sampling by citizen-sailors ('planktonauts')

The P2 PlanktoKit. We first assembled a simple plankton sampling kit including a small net to collect plankton (>20µm) and a manual pumping system to rapidly transfer the freshly collected plankton onto a filter membrane (Sup. Fig. 2). Of note, this sampling protocol does not rely on toxic chemicals or electricity, which are typically required in a regular laboratory. In order to avoid the need for high-energy storage of the filter membranes (plankton samples) in freezers on board, as well as complex frozen-shipping to the lab, the planktonauts were asked to place the filter on a clean aluminum cove and carefully heat-dry the filter membranes in a pan on the boat gas-cooker, store the dried plankton samples in labeled zip-lock bags with granular desiccants, and send them to the lab via standard postal services (Sup. Fig. 3, Sup. Text 1).

Plankton DNA preservation. The protocol for plankton DNA preservation by heating and desiccation was first tested during two field campaigns along the French Atlantic coast and compared with gold-standard cryo-fixation and preservation. Six citizen crews sampled plankton at different locations (Sup Fig. 4A) using the P2 protocol (Sup Fig. 2, 3). The concentrated plankton samples were equally divided into two subsamples: one was flash-frozen in liquid nitrogen as classically performed on oceanographic vessels for genetics analyses, while the other was heat-dried as implemented in P2 (Sup Fig. 3). On land, samples were preserved in a -80°C freezer (flash-frozen samples) and at room temperature (heat-dried samples) for a couple of months before total DNA extraction and sequencing of 1.3 ±0.07 million V9 SSU rDNA amplicons per sample (Sup. Text 1). Bioinformatic clustering of the plankton communities (defined by types and abundance of rDNA Operational

Taxonomic Units (OTUs, see Sup. Text 1) using different dissimilarity indices indicated that the sub-samples preserved by desiccation and flash-freezing systematically clustered together (Sup. Fig. 4B, C). Plankton communities segregated then first by sampling location, then by their distance to fresh-water input within each bay, irrespective of preservation method. These results proved that the heat drying and subsequent desiccation preservation method do not alter the community composition as measured by DNA metabarcoding and can be applied globally.

Empowering pioneer planktonauts to sample the world oceans. With modest seed-funding, we were able to assemble a first set of 20 PlanktoKits, and our primary strategy was to maximize geographic coverage and sampling conditions. We promoted the idea amongst the French sailing community and encountered great enthusiasm to participate. The antique proverb –'*There are three sorts of people: the living, the dead, and those who sail the sea*' is profound: mariners are natural engineers, explorers, and planet-lovers. Word of mouth is powerful in their close-knit community and, limited by the low number of sampling kits available, we soon had to start declining requests. The 27 selected pioneer yacht crews, who we call 'planktonauts' (Fig. 1A, and https://planktonplanet.org/the-planktonauts/) represented a wide variety of boats and sailing modes, from multi-year expeditions around the world (e.g. Race4Water, Taravana, Folligou), to family cruises (e.g. Manevai, Nika, or Zigomar - see kidsforsea.over-blog.com/), recreation sailing in the same zone over the year (e.g. Suhail), explorers of the poles (e.g. Vagabond or Podorange), or participants in a New Zealand yachting rally across the South Pacific Islands (Fig. 1B, bottom-right insert).

The planktonauts were trained to perform the P2 protocol (Sup. Fig.3), individually or in small groups, and regular internet dialogues were established with them during their voyage to answer their questions and follow their progress. They regularly sent movies and pictures (Fig. 1A, https://planktonplanet.org) of their actions at sea, allowing us to improve our training protocols and outreach. Their feedback on all steps of the protocol at sea (Sup. Fig. 3) have been key to identify the bottlenecks and challenges to overcome for the implementation stage of the scientific program (see below).

Geographical coverage and cost of P2-pilot samples: Despite the relatively low number of samples recovered per boat (2 to 18), the widespread routes of many planktonauts (Fig. 1B) yielded plankton samples from 258 sites across the world surface oceans in less than 1 year (Fig. 2B and Sup. Fig.5). This is remarkable compared to the few tens of spatio-temporally relatively restricted stations usually sampled during classical oceanographic cruises, or even to the 147 sites sampled by the schooner *Tara* in 3 years during her first circumglobal expedition (*40*) (Fig. 2B). Furthermore, the total cost of ~200 US$ per sample (including the price of the kit, training, sampling, shipping, and DNA extractions), which could be significantly reduced with increasing sampling frequency, is one to several orders of magnitude lower than sample cost on oceanographic vessels. The running budget of an open oceanographic research vessel is on the order of 30,000 US$/day. Therefore, a single research vessel traveling at 10 knots and stopping for one hour at each sampling station would have taken 8,6 months - and thus a minimum cost of 7,9 million US$ - to cover 56% of the total P2 sites sampled in slightly more than a year by the planktonauts (Sup. Fig.5). Note that the average cost to equip and run a sailing boat for a transoceanic route is ~25,000 US$; all together, volunteer planktonauts thus offered ~90% of the cost of field work in sailing charges. This being said, we

acknowledge that planktonauts are logistically constrained and will never be able to gather the sort of comprehensive bathymetric, oceanographic and physico-chemical data that oceanographic vessels routinely collect. P2 is not a substitute, but a valuable complement to the intensive research cruise manned by experts.

High-quality data for global plankton biodiversity and ecology

18 months after the launch of P2, we had extracted total DNA from 214 plankton samples collected by 27 boats, PCR amplified rDNA metabarcodes from each sample, and generated a total of ±453 million rDNA reads to assess the diversity of eukaryotic plankton (>20 µm) in the explored surface water masses. The methods used for DNA extraction, sequencing, and analyses (Sup. Text 1) were essentially developed in *Tara* Oceans (*34*, *52*); the metabarcode used (V9 SSU rDNA) has proven successful to measure the ecological diversity of total eukaryotic plankton (*34*), focus eco-evolutionary analyses on specific groups (*53–56*), reconstruct plankton ecological networks (*1*), or revisit plankton macro-ecological patterns (*35*, *57*), or biogeochemical processes (*2*).

Integration of P2 data into Tara Oceans data. Tara Oceans (*T*O) data are today a gold standard for ocean plankton ecology, and we first merged the P2 data into the global *T*O metabarcoding dataset (Sup. Text 1) for quality checks and comparison of content. As P2 samples were collected by different sailors, sometimes in harsh conditions, we first checked for the presence of obvious biases, such as the sequencing of bacterial contaminants. After taxonomic assignment of the rDNA reads (Sup. Text 1), we found that the percentage of prokaryotic reads per boat matched the expected value observed in *T*O (given the very large taxonomic spectrum of the eukaryotic PCR primers used that also amplify prokaryotic genes), with just a few outlier samples typically explained by on-board major processing errors (Sup. Fig 6).

We then compared plankton community composition between P2 and *T*O samples. Principal component analysis confirmed the primary influence of organism size on community structuring (*34*), with the pico-nanoplankton (0.8-5 µm) displaying stronger cohesiveness than micro- (20-180 µm) and meso- (180-2000 µm) planktonic communities (Fig. 2A). P2 samples clearly fell within the range of variability of *T*O micro- and meso-plankton samples, reflecting the P2 protocol that uses a 20µm mesh-size net and does not apply any pre-filtration. P2 and *T*O micro/meso- plankton size fractions samples spread together along the second axis of the PCA whose variance (7%) is the result of multiple physico-chemical (dispersal and mixing, variations in temperature, light, nutrients, etc) and biological (species interactions, life cycles, behavior, acclimation/adaptation) processes.

Despite the overall similarity between P2 and *T*O micro/meso-plankton samples (Fig. 2A), the P2 sampling effort did unveil significant novelty in global plankton diversity. All P2 samples produced rDNA OTUs that had not been reported in the *T*O world ocean survey (Fig. 2B), and the number of discovered OTUs clearly increased with sampling effort, both per area (Fig. 2B) and per boat (Fig. 2C). Each planktonaut crew unveiled between 62 and 3,907 unique OTUs, unseen by either *Tara* or any other P2 boats, with a discovery rate partly explained by the eccentricity of the sampled areas relative to *T*O sampling sites. Note for instance the high percentage of boat-specific OTUs (>75%) in the New Zealand area (Fig. 2C).

We further dug into the composition of plankton diversity (rDNA OTUs) unveiled by P2 versus *TO* samples. P2 samples yielded a total of 57,994 OTUs, as compared to 158,716 for *TO* surface ocean samples. Phylogenetic breakdown of the rDNA data (Sup. Fig. 7) shows the overall taxonomic similarity, in both abundance and richness, between data from P2 and the *TO* micro- and meso-planktonic size fractions. While 8,220 OTUs (mostly metazoans and alveolates) were shared exclusively between P2 and the *TO* larger (>20 μm) organismal size fractions, only 1,981 were common with the smaller (<20 μm) *TO* plankton size fractions (Fig. 2D). This emphasizes the consistency of the organisms harvested by P2 and *TO* in comparable plankton size-fractions, and provides organismal data support for our principal component analysis (Fig. 2A). Remarkably, however, more than half of all P2 OTUs (37,163) were actually not seen in any plankton size fractions of the *TO* circumglobal dataset (Fig. 3D). These correspond mostly (~78%) to Alveolata and Opisthokonta, with an overwhelming majority (98%) of relatively large metazoans (copepods, other arthropods, mollusks, see insert in Fig. 2D). This major difference is most likely explained by the fact that the P2 sampling protocol does not involve an upper-size filtration while the *TO* larger size-fraction was constrained by a sieving at 2mm (micro-plankton: 180-2000 μm), resulting in a more complete survey of larger plankton in P2.

In order to check the value of P2 data at finer-grained taxonomic resolution, we explored the phylogenetic and organismal size-fraction distribution of P2 and *TO* OTUs assigned to a well-known phytoplankton group, the dinoflagellate order Peridiniales (Sup. Fig. 8). Most Peridiniales OTUs observed in *TO* were also found in P2, except a few taxa particularly abundant in the piconano- and nano- size fractions (<20 μm). The relatively large, microplanktonic taxa from the genus *Protoperidinium* were particularly well represented in the P2 dataset, with four OTUs unseen in the *TO* dataset. Note also the presence of many *Blastodinium* and *Brandtodinium* OTUs, which are well-known parasites (*58*) and photosymbionts (*59*) of respectively copepods and radiolarians, and are thus part of the meso- and macro-plankton in their symbiotic stage. Overall, these analyses confirm the quality of the P2 data to assess plankton diversity from the OTU to the community level.

Insights into multiscale plankton ecology using P2 data. We finally used P2 metabarcoding data to explore the consistency of macroecological patterns at both global and local scales, across the data collected by the different boats and planktonauts who sampled plankton independently. At the world ocean scale, the P2 data displayed an increase in alpha diversity (Shannon index) from both poles to the tropics (Fig. 3A), reflecting the Latitudinal Diversity Gradient that is well known from terrestrial and marine ecosystems (*60*), and has been previously recorded in several more restricted planktonic groups (e.g. *61–63*)) and across plankton kingdoms (*35*). On a more regional scale, the five planktonaut crews from New Zealand collected enough samples to highlight a biogeographical pattern in this area. Clustering of the plankton communities based on their OTU composition into three groups fitted with the three legs of the navigation loop: New Zealand - Fiji ; Fiji - New Caledonia, New Caledonia - New Zealand (Fig. 3B). The first area was characterized by high abundances of copepods (Fig. 3D) and important richness of indicator OTUs (*64*) amongst Marine Alveolates (MALV), Acantharia, Dinophyceae and also copepods (Sup. Fig. 9A). By contrast, the last portion of the navigation loop (genetic cluster 3) was characterized by low abundances of copepods and high abundances of collodarians (Fig. 3C), with indicator OTUs belonging to collodarians and

8

spumellarians, diatoms (Bacillariophyta), copepods and choanoflagellates. Group 2 (Fiji-New Caledonia) had a lower diversity of indicator OTUs and was genetically less homogeneous (Sup. Fig. 9B), which may be due to the variability of habitats characterizing this island region. All boats (except *Nika* who is represented by one sample in the analysis) have sampled plankton in at least two different eco-genetic regions, indicating there was no apparent bias linked to the multiplication of samplers.

Contextual environmental data.

Given the importance of the abiotic environment in structuring planktonic ecosystems (*3*, *4*), we also included the measure of basic environmental parameters at each P2 sampling site. Besides UTC date/time for each sampling event, the planktonauts were asked to record surface water temperature using the water temperature sensor of their boat and/or mercury thermometer (Fig. 3A), and broad-band spectral water reflectance using the Hydrocolor App (*65*). *In situ* temperature data were compared with those from the NASA's MODIS-Aqua satellite, and a good match – on average better than 1°C – was observed. We also developed automated procedures to extract remote sensing ocean color data at each sampling site (https://github.com/OceanOptics/getOC), providing bulk information on chlorophyll-a concentration, as well as suspended dissolved and particulate materials in the target water. In the future, this data can not only assist planning of sampling stations at sea in near-real time, but also allow analyses linking ocean color to phytoplankton functional type measured by DNA metabarcoding. On the other hand, the Hydrocolor App has been compared to commercial instrumentation and found to correlate well (*66*); it is hoped it will help relate water reflectance to *in-situ* plankton communities, increasing further the utility of remote sensed ocean color.

Plankton Planet: perspectives

Operational PlanktoKit integrating eco-morpho-genetic data

In the pilot stage of 'Plankton Planet', we have witnessed the enormous desire of sailors to sample ocean life during their voyages around the globe. We have also generated what is, to our knowledge, the first citizen-based homogenous planetary dataset to explore plankton biodiversity and ecology. Our objective is now to optimize the P2 sampling protocol and tools for large-scale deployment on navigation loops and transects crossing critical ecological and geographical dimensions of the Earth system. Along the way, we have identified critical challenges to overcome.

Sampling total plankton at sailing-speed.   In the P2 pilot project, the planktonauts were asked to maneuver their boats at a speed of less than 2 knots when towing the 20μm-mesh size plankton net. This requests uncomfortable sailing operations impacting the cruising speed, and it was identified as the primary limiting factor for denser sampling. We have therefore been working on the design of new miniaturized *high-speed nets*, inspired by our successful experience during the *Tara* Pacific expedition (*67*), as well as simple manual pumping systems allowing aspiration of seawater at cruising speed, followed by filtration through a small net system installed on board. The latter device has the advantage of being able to collect pristine water that can also be used to extract DNA/RNA

from the smaller plankton size fractions enriched in bacteria, archaea and viruses. Indeed, long-term monitoring of marine plankton systems will ultimately require a sampling protocol covering the 8 orders of plankton organismal size-magnitude, from viruses to animals (*23*, *33*), in order to assess both top-down and bottom-up ecological mechanisms shaping the ecosystem.

Collecting morphological and behavioral plankton data at sea.   The revolution in environmental DNA/RNA sequencing provides the power to comprehensively assess plankton taxonomic and metabolic diversity (*33*). However, meta-omics data convey relatively poor information on the shapes, structures and behaviors of cells and organisms, which may well be the primary drivers for the self-organization of contemporary ecosystems and their emergent functions (*68*). Fundamental phenotypic mechanisms such as symbioses *sensu lato*, selective feeding, or vertical motions and migrations (*69*), and several other unknown complex cellular and organismal behaviors, need to be identified and quantified in the context of the seascape if we ever want to reach a mechanistic pheno-genomic understanding of ecosystem patterning and functioning in the ocean.

Over the last few years, we have thus developed the *PlanktonScope* (see companion paper by Pollina et al., this issue (*51*)), an affordable, miniaturized, modular and evolvable, open source imaging platform for citizen oceanography. For a total cost of less than 400 US$ in parts, the *PlanktonScope* allows both quantitative imaging of plankton communities through a fluidic module before their storage for total DNA/RNA extraction and genetic analyses in the lab, and high-quality imaging/filming of individual cells or organisms under different types of illumination. The scarcity of live images and movies of plankton at sea constitutes arguably the major knowledge-gap in oceanography as today's oceanographers spend most of their time on land behind their computer screens or analyzing dead plankton in the lab. The *PlanktonScope* will not only bring planetary imaging of plankton life through its global seatizen deployment, but also, we hope, trigger an emotional shock in each planktonaut discovering the vibrant beauty of the marine microbiome, changing their view of the ocean forever and further driving their will to explore and preserve the invisible world thriving under their boat.

Toward a perennial, self-sustainable standardized citizen monitoring of global ocean life

The cost of the upcoming P2 PlanktoKit 1.0 (plankton collecting system, PlanktonScope, DNA-kit, software App to visualize, integrate, and store eco/morpho/genetic information) is estimated at 2,000-3,000 US$ per boat, allowing its deployment on hundreds of seatizen boats. The deployment of the first 100 kits will start in 2021 in the Atlantic and Pacific oceans, and is designed to address topical questions in ocean ecology and to test current models and theories that predict patterns of ocean biodiversity and functionality in relation to the physics of marine water and climate.

Based on feedback between these unique data and modelling efforts, we will then fine tune the optimal density of sailboats (or fish boats or cargo ships) and sampling frequency for longer-term deployment of the P2 protocol on key navigation loops and routes that cover the world oceans. Our ultimate goal is to implement, over the next decade, a self-sustainable and standardized global survey of the diversity and evolution of surface ocean plankton ecosystems. We believe that this endeavor can only happen through collaboration at the interface between the worlds of sailors, makers, and researchers, and it can fundamentally change the way oceanography is performed,

while enhancing the collective knowledge and consciousness of the open ocean invisible life at the core of global ecology (Box 2).

BOX 2. Perspective outputs of Plankton Planet

Oceanography 3.0 ecosystem
- An international fleet of planktonauts to act as sentinels and the collective consciousness of the biological health of our oceans.
- An evolvable toolkit of affordable, scientifically relevant instruments for seatizen-based assessment of aquatic (marine and freshwater) biodiversity and ecosystems.
- An ever-growing cryo-bank of global ocean DNA samples, archiving the memory of our changing oceans for future generations and technologies.
- A continuous flow of standardized ocean imaging and genetic data at unprecedented spatial, temporal, and taxonomic scales, freely available for fundamental and applied science, policy-makers, and education.

Blue-sky science
- Exceptional long-term monitoring of the distribution and evolution of total plankton biodiversity in our fast-changing ocean; including analysis of the global impacts of warming, acidification, and de-oxygenation.
- Novel understanding of abundance, structures, functions, and behaviors of plankton life via the billions of images and movies of planktonic organisms generated using PlanktonScopes.
- Incorporation of high-resolution global-scale comprehensive biological data into efforts to model the dynamics of ocean ecosystems and ocean-climate interactions, from genes to biogeochemical cycles.

Applied science & policies
- Detection of invasive, toxic, or economically relevant species at a planetary scale.
- Linking of data on plankton communities (including fish gametes and larvae) to fish-catch data for robust prediction of fish stocks.
- Assessment of ocean biogeographic zones ("seascapes") for optimal design of marine protected areas in the high seas (e.g. ecosystems with high capacity for carbon pumping).
- Evaluation of the health of oceanic regions based on species content, species richness, and trends in time.

Education through 'Plankton & Arts'
- Collective awareness of oceanic microbiomes and their planetary impact, both directly (through the PlanktonScope) and indirectly via the shared images, movies, and 3D prints of plankton. Educational tools will notably be distributed by planktonauts in remote countries and islands whose communities interact with and depend on the marine ecosystem.

Citizen ecoscopes for collective field assessment of the biosphere

At the core of our long-term vision, the principles applied to development of the PlanktonScope (*51*) will be extended to co-construct a field-EcoScope for citizen exploration of aquatic eco-systems, and the concept could be easily extended for the exploration of other planetary biomes. Like the PlanktonScope, the field-EcoScope is a miniaturized platform that can host *modules* and sensors to

measure the critical biological, but also (bio)chemical, and physical parameters of any eco-system. Each module (including the PlanktonScope) will be co-developed by small international and interdisciplinary teams of (*i*) researchers who will define the scientific specifications to address fundamental question in global ecology, (*ii*) mariners and citizen samplers who will assess practical user-constraints linked to specific field conditions (e.g. *on-board* sailing boats or cargo-ships), and (*iii*) makers and freelance engineers who will design and construct the frugal modules to measure given parameters. The Plankton Planet ecosystem will manage communication and provide the overall coordination to ensure that all modules can be fully integrated into the miniaturized field-laboratory and generate good-quality standard data that are immediately shared in public databases.

We believe that it is a matter of urgency to quantify and understand the structure, dynamics, and evolution of *eco-systems*. Over the last century, societies have spent phenomenal amounts of money to discover the properties of the infinitely small (atoms and cells) and the infinitely large (weather, planets and stars, and the universe). But it is in between these extremes that we find arguably the most complex object in the known universe: eco-systems, i.e. self-organized and evolving life networks that interact with physico-chemical processes at micro- to planetary scales. Ecosystems have shaped the Earth atmosphere and biogeochemical cycles, they currently buffer climate change while providing sources of food and medicine, they will determine the future habitability of our planet and the fate of humanity. Due to their extreme complexity, integrating biology, chemistry, and physics, ecosystems have long escaped holistic quantitative assessment. Today the tools and methods exist, in particular automated sequencing and imaging, as well as artificial intelligence and massive computing to collect and integrate the layers of big data needed to understand eco-systems. The major challenge is to develop the new generation of affordable tools that can be deployed in a systematic manner across the spatio-temporal dimension of the Earth system. Plankton Planet proposes a coherent and frugal approach to overcome this challenge, at the interface of science and society, using one of the simplest biomes – marine waters- as a case study. Clearly being at the onset of planetary biology, we hope that P2 will contribute to reaching a profound understanding of our habitat in the decades to come, and to learning how to live in synergy with the biosphere.

Acknowledgments

P^2 on the web:

- Web site: http://planktonplanet.org
- YouTube channel: http://youtube.planktonplanet.org
- Vimeo channel: http://vimeo.planktonplanet.org
- Twitter account: https://twitter.com/PlanktonPlanet (@PlanktonPlanet)

References

1. G. Lima-mendez, K. Faust, N. Henry, S. Colin, F. Carcillo, S. Chaffron, J. C. Ignacio-Espinoza, S. Roux, F. Vincent, L. Bittner, Y. Darzi, J. Wang, S. Audic, L. Berline, G. Bontempi, A.-M. Cabello, L. Coppola, F. M. Cornejo-castillo, F. D'Ovidio, L. De Meester, I. Ferrera, M.-J. Garret-Delmas, L. Guidi, E. Lara, S. Pesant, M. Royo-llonch, G. Salazar, P. Sánchez, M. Sebastian, C. Souffreau, C. Dimier, M. Picheral, S. Searson, S. Kandels-lewis, C. Tara Oceans, G. Gorsky, F. Not, H. Ogata, S. Speich, L. Stemmann, J. Weissenbach, P. Wincker, S. G. Acinas, S. Sunagawa, P. Bork, M. B. Sullivan, E. Karsenti, C. Bowler, C. de Vargas, J. Raes, Determinants of community structure in the global plankton interactome. *Science*. **348**, 1262073 (2015).

2. L. Guidi, S. Chaffron, L. Bittner, D. Eveillard, A. Larhlimi, S. Roux, Y. Darzi, S. Audic, L. Berline, J. R. Brum, L. P. Coelho, J. C. Ignacio-Espinoza, S. Malviya, S. Sunagawa, C. Dimier, S. Kandels-Lewis, M. Picheral, J. Poulain, S. Searson, T. O. Coordinators, L. Stemmann, F. Not, P. Hingamp, S. Speich, M. Follows, L. Karp-Boss, E. Boss, H. Ogata, S. Pesant, J. Weissenbach, P. Wincker, S. G. Acinas, P. Bork, C. de Vargas, D. Iudicone, M. B. Sullivan, J. Raes, E. Karsenti, C. Bowler, G. Gorsky, Plankton networks driving carbon export in the oligotrophic ocean. *Nature*. **532**, 465–470 (2016).

3. D. J. Richter, R. Watteaux, T. Vannier, J. Leconte, P. Frémont, G. Reygondeau, N. Maillet, N. Henry, G. Benoit, A. Fernàndez-Guerra, S. Suweis, R. Narci, C. Berney, D. Eveillard, F. Gavory, L. Guidi, K. Labadie, E. Mahieu, J. Poulain, S. Romac, S. Roux, C. Dimier, S. Kandels, M. Picheral, S. Searson, T. O. Coordinators, S. Pesant, J.-M. Aury, J. R. Brum, C. Lemaitre, E. Pelletier, P. Bork, S. Sunagawa, L. Karp-Boss, C. Bowler, M. B. Sullivan, E. Karsenti, M. Mariadassou, I. Probert, P. Peterlongo, P. Wincker, C. de Vargas, M. R. d'Alcalà, D. Iudicone, O. Jaillon, T. O. Coordinators, Genomic evidence for global ocean plankton biogeography shaped by large-scale current systems. *bioRxiv*, 867739 (2019).

4. R. Logares, I. M. Deutschmann, P. C. Junger, C. R. Giner, A. K. Krabberød, T. S. B. Schmidt, L. Rubinat-Ripoll, M. Mestre, G. Salazar, C. Ruiz-González, M. Sebastián, C. de Vargas, S. G. Acinas, C. M. Duarte, J. M. Gasol, R. Massana, Disentangling the mechanisms shaping the surface ocean microbiota. *Microbiome*. **8**, 55 (2020).

5. S. J. Pittman, *Seascape Ecology* (Oxford, Wiley Blac., 2018).

6. P. Falkowski, The power of plankton. *Science (80-.)*. **483**, 17–20 (2012).

7. M. J. Follows, S. Dutkiewicz, S. Grant, S. W. Chisholm, Emergent biogeography of microbial communities in a model ocean. *Science (80-.)*. **315**, 1843–1846 (2007).

8.   C. A. Durkin, B. A. S. Van Mooy, S. T. Dyhrman, K. O. Buesseler, Sinking phytoplankton associated with carbon flux in the Atlantic Ocean. *Limnol. Oceanogr.* **61**, 1172–1187 (2016).

9.   J. Decelle, P. Martin, K. Paborstava, D. W. Pond, G. Tarling, F. Mahé, C. de Vargas, R. Lampitt, F. Not, Diversity, Ecology and Biogeochemistry of Cyst-Forming Acantharia (Radiolaria) in the Oceans. *PLoS One*. **8** (2013), doi:10.1371/journal.pone.0053598.

10.   P. G. Falkowski, T. Fenchel, E. F. Delong, The microbial engines that drive earth's biogeochemical cycles. *Science (80-.).* **320**, 1034–1039 (2008).

11.   E. Y. Kwon, F. Primeau, J. L. Sarmiento, The impact of remineralization depth on the air-sea carbon balance. *Nat. Geosci.* **2**, 630–635 (2009).

12.   K. O. Buesseler, P. W. Boyd, E. E. Black, D. A. Siegel, Metrics that matter for assessing the ocean biological carbon pump. *Proc. Natl. Acad. Sci.* **117**, 201918114 (2020).

13.   G. Beaugrand, R. R. Kirby, Climate, plankton and cod. *Glob. Chang. Biol.* **16**, 1268–1280 (2010).

14.   K. L. Morgan, S. L. Larkin, C. M. Adams, Red tides and participation in marine-based activities: Estimating the response of Southwest Florida residents. *Harmful Algae*. **9**, 333–341 (2010).

15.   M. J. Follows, S. Dutkiewicz, Modeling Diverse Communities of Marine Microbes. *Ann. Rev. Mar. Sci.* **3**, 427–451 (2011).

16.   B. A. Ward, S. Dutkiewicz, M. J. Follows, Modelling spatial and temporal patterns in size-structured marine plankton communities: Top-down and bottom-up controls. *J. Plankton Res.* **36**, 31–47 (2014).

17.   W. Stanley Wilson, E. J. Lindstrom, J. K. Willis, *Satellite oceanography-history and introductory concepts* (Elsevier Ltd., ed. 3, 2019; http://dx.doi.org/10.1016/B978-0-12-409548-9.10816-4).

18.   D. Roemmich, M. H. Alford, H. Claustre, K. S. Johnson, B. King, J. Moum, P. R. Oke, W. B. Owens, S. Pouliquen, S. Purkey, M. Scanderbeg, T. Suga, S. E. Wijffels, N. Zilberman, D. Bakker, M. O. Baringer, M. Belbeoch, H. C. Bittig, E. Boss, P. Calil, F. Carse, T. Carval, F. Chai, D. O. Conchubhair, F. D'Ortenzio, G. Dall'Olmo, D. Desbruyères, K. Fennel, I. Fer, R. Ferrari, G. Forget, H. Freeland, T. Fujiki, M. Gehlen, B. Greenan, R. Hallberg, T. Hibiya, S. Hosoda, S. Jayne, M. Jochum, G. C. Johnson, K. R. Kang, N. Kolodziejczyk, A. Koertzinger, P. Y. Le Traon, Y. D. Lenn, G. Maze, K. A. Mork, T. Morris, T. Nagai, J. Nash, A. N. Garabato, A. Olsen, R. R. Pattabhi, S. Prakash, S. Riser, C. Schmechtig, E. Shroyer, A. Sterl, P. Sutton, L. Talley, T. Tanhua, V. Thierry, S. Thomalla, J. Toole, A. Troisi, T. Trull, J. D. Turton, P. J. Velez-Belchi, W. Walczowski, H. Wang, R. Wanninkhof, A. Waterhouse, A. Watson, C. Wilson, A. P. Wong, J. Xu, I. Yasuda, On the future of Argo: A global, full-depth, multi-disciplinary array. *Front. Mar. Sci.* **6**, 1–28 (2019).

19.   H. Claustre, K. S. Johnson, Y. Takeshita, Observing the Global Ocean with Biogeochemical-Argo. *Ann. Rev. Mar. Sci.* **12**, 23–48 (2020).

20.   J. Vazquez-Cuervo, J. Gomez-Valdes, M. Bouali, L. E. Miranda, T. Van der Stocken, W. Tang, C. Gentemann, Using saildrones to validate satellite-derived sea surface salinity and sea surface temperature along the California/Baja coast. *Remote Sens.* **11** (2019), doi:10.3390/rs11171964.

21.   B. M. Sloyan, R. Wanninkhof, M. Kramp, G. C. Johnson, L. Talley, T. Tanhua, E. McDonagh, C. Cusack, E. A. O'Rourke, E. McGovern, K. Katsumata, S. Diggs, J. Hummon, M. Ishii, K. Azetsu-Scott, E. Boss, I. Ansorge, F. F. Pérez, H. Mercier, M. Williams, L. Anderson, J. H. Lee, A. Murata, S. Kouketsu, The global ocean ship-base hydrographic investigations program (GO-SHIP): A platform for integrated multidisciplinary ocean science. *Front. Mar. Sci.* **6**, 1–21 (2019).

22.   C. Simoniello, J. Jencks, F. M. Lauro, J. D. Loftis, J. M. Weslawski, K. Deja, D. R. Forrest, S. Gossett, T. C. Jeffries, R. M. Jensen, S. Kobara, L. Nolan, M. Ostrowski, D. Pounds, G. Roseman, O. Basco, S. Gosselin, A. Reed, P. Wills, D. Wyatt, Citizen-science for the future: Advisory case studies from around the globe. *Front. Mar. Sci.* **6** (2019), pp. 1–15.

23.   F. Lombard, E. Boss, A. M. Waite, J. Uitz, L. Stemmann, H. M. Sosik, J. Schulz, J. B. Romagnan, M. Picheral, J. Pearlman, M. D. Ohman, B. Niehoff, K. O. Möller, P. Miloslavich, A. Lara-Lopez, R. M. Kudela, R. M. Lopes, L. Karp-Boss, R. Kiko, J. S. Jaffe, M. H. Iversen, J. O. Irisson, H. Hauss, L. Guidi, G. Gorsky, S. L. C. Giering, P. Gaube, S. Gallager, G. Dubelaar, R. K. Cowen, F. Carlotti, C. Briseño-Avena, L. Berline, K. J. Benoit-Bird, N. J. Bax, S. D. Batten, S. D. Ayata, W. Appeltans, Globally consistent quantitative observations of planktonic ecosystems. *Front. Mar. Sci.* **6** (2019), doi:10.3389/fmars.2019.00196.

24.   E. T. Buitenhuis, M. Vogt, R. Moriarty, N. Bednaršek, S. C. Doney, K. Leblanc, C. Le Quéré, Y. W. Luo, C. O'Brien, T. O'Brien, J. Peloquin, R. Schiebel, C. Swan, MAREDAT: Towards a world atlas of MARine Ecosystem DATa. *Earth Syst. Sci. Data*. **5**, 227–239 (2013).

25.   A. J. Richardson, A. W. Walne, A. W. G. John, T. D. Jonas, J. A. Lindley, D. W. Sims, D. Stevens, M. Witt, Using continuous plankton recorder data. *Prog. Oceanogr.* **68**, 27–74 (2006).

26.   P. C. Reid, J. M. Colebrook, J. B. L. Matthews, J. Aiken, The Continuous Plankton Recorder: concepts and

history, from Plankton Indicator to undulating recorders. *Prog. Oceanogr.* **58**, 117–173 (2003).

27. S. D. Batten, R. Abu-Alhaija, S. Chiba, M. Edwards, G. Graham, R. Jyothibabu, J. A. Kitchener, P. Koubbi, A. McQuatters-Gollop, E. Muxagata, C. Ostle, A. J. Richardson, K. V. Robinson, K. T. Takahashi, H. M. Verheye, W. Wilson, A global Plankton diversity monitoring program. *Front. Mar. Sci.* **6**, 1–14 (2019).

28. G. Beaugrand, P. C. Reid, F. Ibañez, J. A. Lindley, M. Edwards, Reorganization of North Atlantic marine copepod biodiversity and climate. *Science (80-.).* **296**, 1692–1694 (2002).

29. W. J. Chivers, A. W. Walne, G. C. Hays, Mismatch between marine plankton range movements and the velocity of climate change. *Nat. Commun.* **8**, 1–8 (2017).

30. J. C. Venter, K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Wu, I. Paulsen, K. E. Nelson, W. Nelson, D. E. Fouts, S. Levy, A. H. Knap, M. W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y. H. Rogers, H. O. Smith, Environmental Genome Shotgun Sequencing of the Sargasso Sea. *Science (80-.).* **304**, 66–74 (2004).

31. E. Karsenti, S. G. Acinas, P. Bork, C. Bowler, C. de Vargas, J. Raes, M. Sullivan, D. Arendt, F. Benzoni, J. M. Claverie, M. Follows, G. Gorsky, P. Hingamp, D. Iudicone, O. Jaillon, S. Kandels-Lewis, U. Krzic, F. Not, H. Ogata, S. Pesant, E. G. Reynaud, C. Sardet, M. E. Sieracki, S. Speich, D. Velayoudon, J. Weissenbach, P. Wincker, C. Abergel, D. Arslan, S. Audic, J. M. Aury, N. Babic, L. Beaufort, L. Bittner, E. Boss, C. Boutte, J. Brum, M. Carmichael, R. Casotti, A. Chambouvet, P. Chang, C. Chica, C. Clerissi, S. Colin, F. M. Cornejo-Castillo, C. Da Silva, S. De Monte, J. Decelle, Y. Desdevises, C. Dimier, J. Dolan, M. Duhaime, X. Durrieu de Madron, F. d'Ortenzio, F. d'Ovidio, I. Ferrera, L. Garczarek, M. J. Garet-Delmas, S. Gasmi, J. M. Gasol, N. Grimsley, R. Heilig, J. Ignacio-Espinoza, J. L. Jamet, L. Karp-Boss, M. Katinka, H. Khalili, Z. Kolber, N. Le Bescot, H. Le Goff, G. Lima-Mendez, F. Mahe´, M. G. Mazzocchi, M. Montresor, P. Morin, B. Noel, C. Pedro´s-Alio´, E. Pelletier, Y. Perez, M. Picheral, G. Piganeau, O. Poirot, J. Poulain, N. Poulton, F. Prejger, J. Prihoda, I. Probert, J. Rampal, G. Reverdin, S. Romac, J. B. Romagnan, F. Roullier, C. Rouviere, G. Samson, S. Santini, H. Sarmento, A. Sciandra, S. Solonenko, L. Stemmann, L. Subirana, S. Sunagawa, A. Tanaka, P. Testor, A. Thompson, V. Tichanne´-Seltzer, L. Tirichine, E. Toulza, S. Tozzi, A. Veluchamy, A. Zingone, A holistic approach to marine Eco-systems biology. *PLoS Biol.* **9**, 7–11 (2011).

32. C. M. Duarte, Seafaring in the 21st century: The Malaspina 2010 circumnavigation expedition. *Limnol. Oceanogr. Bull.* **24**, 11–14 (2015).

33. S. Sunagawa, S. G. Acinas, P. Bork, C. Bowler, D. Tara Oceans Coordinators, Eveillard, G. Gorsky, L. Guidi, D. Iudicone, E. Karsenti, F. Lombard, H. Ogata, S. Pesant, M. B. Sullivan, P. Wincker, C. de Vargas, Tara Oceans: towards global ocean eco-systems biology. *Nat. Rev. Microbiol.* (2020), doi:10.1038/s41579-020-0364-5.

34. C. De Vargas, S. Audic, N. Henry, J. Decelle, F. Mahé, R. Logares, E. Lara, C. Berney, N. Le Bescot, I. Probert, M. Carmichael, J. Poulain, S. Romac, S. Colin, J. M. Aury, L. Bittner, S. Chaffron, M. Dunthorn, S. Engelen, O. Flegontova, L. Guidi, A. Horak, O. Jaillon, G. Lima-Mendez, J. Lukes, S. Malviya, R. Morard, M. Mulot, E. Scalco, R. Siano, F. Vincent, A. Zingone, C. Dimier, M. Picheral, S. Searson, S. Kandels-Lewis, T. O. Coordinators, S. G. Acinas, P. Bork, C. Bowler, G. Gorsky, N. Grimsley, P. Hingamp, D. Iudicone, F. Not, H. Ogata, S. Pesant, J. Raes, M. Sieracki, S. Speich, L. Stemmann, S. Sunagawa, J. Weissenbach, P. Wincker, E. Karsenti, Eukaryotic plankton diversity in the sunlit ocean. *Science (80-.).* **348**, 1261605–1/11 (2015).

35. F. M. Ibarbalz, N. Henry, M. C. Brandão, S. Martini, G. Busseni, H. Byrne, L. P. Coelho, H. Endo, J. M. Gasol, A. C. Gregory, F. Mahé, J. Rigonato, M. Royo-Llonch, G. Salazar, I. Sanz-Sáez, E. Scalco, D. Soviadan, A. A. Zayed, A. Zingone, K. Labadie, J. Ferland, C. Marec, S. Kandels, M. Picheral, C. Dimier, J. Poulain, S. Pisarev, M. Carmichael, S. Pesant, S. G. Acinas, M. Babin, P. Bork, E. Boss, C. Bowler, G. Cochrane, C. de Vargas, M. Follows, G. Gorsky, N. Grimsley, L. Guidi, P. Hingamp, D. Iudicone, O. Jaillon, L. Karp-Boss, E. Karsenti, F. Not, H. Ogata, N. Poulton, J. Raes, C. Sardet, S. Speich, L. Stemmann, M. B. Sullivan, S. Sunagawa, P. Wincker, E. Pelletier, L. Bopp, F. Lombard, L. Zinger, Global Trends in Marine Plankton Diversity across Kingdoms of Life. *Cell*. **179**, 1084-1097.e21 (2019).

36. A. C. Gregory, A. A. Zayed, N. Conceição-Neto, B. Temperton, B. Bolduc, A. Alberti, M. Ardyna, K. Arkhipova, M. Carmichael, C. Cruaud, C. Dimier, G. Domínguez-Huerta, J. Ferland, S. Kandels, Y. Liu, C. Marec, S. Pesant, M. Picheral, S. Pisarev, J. Poulain, J. É. Tremblay, D. Vik, S. G. Acinas, M. Babin, P. Bork, E. Boss, C. Bowler, G. Cochrane, C. de Vargas, M. Follows, G. Gorsky, N. Grimsley, L. Guidi, P. Hingamp, D. Iudicone, O. Jaillon, S. Kandels-Lewis, L. Karp-Boss, E. Karsenti, F. Not, H. Ogata, N. Poulton, J. Raes, C. Sardet, S. Speich, L. Stemmann, M. B. Sullivan, S. Sunagawa, P. Wincker, A. I. Culley, B. E. Dutilh, S. Roux, Marine DNA Viral Macro- and Microdiversity from Pole to Pole. *Cell*. **177**, 1109-1123.e14 (2019).

37.	S. Sunagawa, L. P. Coelho, S. Chaffron, J. R. Kultima, K. Labadie, G. Salazar, B. Djahanschiri, G. Zeller, D. R. Mende, A. Alberti, F. M. Cornejo-Castillo, P. I. Costea, C. Cruaud, F. D'Ovidio, S. Engelen, I. Ferrera, J. M. Gasol, L. Guidi, F. Hildebrand, F. Kokoszka, C. Lepoivre, G. Lima-Mendez, J. Poulain, B. T. Poulos, M. Royo-Llonch, H. Sarmento, S. Vieira-Silva, C. Dimier, M. Picheral, S. Searson, S. Kandels-Lewis, E. Boss, M. Follows, L. Karp-Boss, U. Krzic, E. G. Reynaud, C. Sardet, M. Sieracki, D. Velayoudon, C. Bowler, C. De Vargas, G. Gorsky, N. Grimsley, P. Hingamp, D. Iudicone, O. Jaillon, F. Not, H. Ogata, S. Pesant, S. Speich, L. Stemmann, M. B. Sullivan, J. Weissenbach, P. Wincker, E. Karsenti, J. Raes, S. G. Acinas, P. Bork, Structure and function of the global ocean microbiome. *Science (80-.).* **348**, 1–10 (2015).

38.	Q. Carradec, E. Pelletier, C. Da Silva, A. Alberti, Y. Seeleuthner, R. Blanc-Mathieu, G. Lima-Mendez, F. Rocha, L. Tirichine, K. Labadie, A. Kirilovsky, A. Bertrand, S. Engelen, M. A. Madoui, R. Méheust, J. Poulain, S. Romac, D. J. Richter, G. Yoshikawa, C. Dimier, S. Kandels-Lewis, M. Picheral, S. Searson, S. G. Acinas, E. Boss, M. Follows, G. Gorsky, N. Grimsley, L. Karp-Boss, U. Krzic, S. Pesant, E. G. Reynaud, C. Sardet, M. Sieracki, S. Speich, L. Stemmann, D. Velayoudon, J. Weissenbach, O. Jaillon, J. M. Aury, E. Karsenti, M. B. Sullivan, S. Sunagawa, P. Bork, F. Not, P. Hingamp, J. Raes, L. Guidi, H. Ogata, C. De Vargas, D. Iudicone, C. Bowler, P. Wincker, A global ocean atlas of eukaryotic genes (SupMat). *Nat. Commun.* **9** (2018).

39.	G. Salazar, L. Paoli, A. Alberti, J. Huerta-Cepas, H. J. Ruscheweyh, M. Cuenca, C. M. Field, L. P. Coelho, C. Cruaud, S. Engelen, A. C. Gregory, K. Labadie, C. Marec, E. Pelletier, M. Royo-Llonch, S. Roux, P. Sánchez, H. Uehara, A. A. Zayed, G. Zeller, M. Carmichael, C. Dimier, J. Ferland, S. Kandels, M. Picheral, S. Pisarev, J. Poulain, S. G. Acinas, M. Babin, P. Bork, E. Boss, C. Bowler, G. Cochrane, C. de Vargas, M. Follows, G. Gorsky, N. Grimsley, L. Guidi, P. Hingamp, D. Iudicone, O. Jaillon, S. Kandels-Lewis, L. Karp-Boss, E. Karsenti, F. Not, H. Ogata, S. Pesant, N. Poulton, J. Raes, C. Sardet, S. Speich, L. Stemmann, M. B. Sullivan, S. Sunagawa, P. Wincker, Gene Expression Changes and Community Turnover Differentially Shape the Global Ocean Metatranscriptome. *Cell.* **179**, 1068-1083.e21 (2019).

40.	P. Bork, C. Bowler, C. De Vargas, G. Gorsky, E. Karsenti, P. Wincker, Tara Oceans studies plankton at Planetary scale. *Science (80-.).* **348**, 873 (2015).

41.	L. Caputi, Q. Carradec, D. Eveillard, A. Kirilovsky, E. Pelletier, J. J. Pierella Karlusich, F. Rocha Jimenez Vieira, E. Villar, S. Chaffron, S. Malviya, E. Scalco, S. G. Acinas, A. Alberti, J. M. Aury, A. S. Benoiston, A. Bertrand, T. Biard, L. Bittner, M. Boccara, J. R. Brum, C. Brunet, G. Busseni, A. Carratalà, H. Claustre, L. P. Coelho, S. Colin, S. D'Aniello, C. Da Silva, M. Del Core, H. Doré, S. Gasparini, F. Kokoszka, J. L. Jamet, C. Lejeusne, C. Lepoivre, M. Lescot, G. Lima-Mendez, F. Lombard, J. Lukeš, N. Maillet, M. A. Madoui, E. Martinez, M. G. Mazzocchi, M. B. Néou, J. Paz-Yepes, J. Poulain, S. Ramondenc, J. B. Romagnan, S. Roux, D. Salvagio Manta, R. Sanges, S. Speich, M. Sprovieri, S. Sunagawa, V. Taillandier, A. Tanaka, L. Tirichine, C. Trottier, J. Uitz, A. Veluchamy, J. Veselá, F. Vincent, S. Yau, S. Kandels-Lewis, S. Searson, C. Dimier, M. Picheral, P. Bork, E. Boss, C. de Vargas, M. J. Follows, N. Grimsley, L. Guidi, P. Hingamp, E. Karsenti, P. Sordino, L. Stemmann, M. B. Sullivan, A. Tagliabue, A. Zingone, L. Garczarek, F. d'Ortenzio, P. Testor, F. Not, M. R. d'Alcalà, P. Wincker, C. Bowler, D. Iudicone, Community-Level Responses to Iron Availability in Open Ocean Plankton Ecosystems. *Global Biogeochem. Cycles.* **33**, 391–419 (2019).

42.	J. Jacquin, J. Cheng, C. Odobel, C. Pandin, P. Conan, M. Pujo-Pay, V. Barbe, A. L. Meistertzheim, J. F. Ghiglione, Microbial ecotoxicology of marine plastic debris: A review on colonization and biodegradation by the "plastisphere." *Front. Microbiol.* **10**, 1–16 (2019).

43.	V. Tornero, G. Hanke, Chemical contaminants entering the marine environment from sea-based sources: A review with a focus on European seas. *Mar. Pollut. Bull.* **112**, 17–38 (2016).

44.	G. C. Hays, A. J. Richardson, C. Robinson, Climate change and marine plankton. *Trends Ecol. Evol.* **20**, 337–344 (2005).

45.	O. A. & P. J. Pörtner H.-O., Roberts D. C., Masson-Delmotte V., Zhai P., M T., Poloczanska E., Mintenbeck K., Nicolai M., "IPCC Special Report on the Ocean and Cryosphere in a Changing Climate" (Geneva, 2019).

46.	M. Lévy, P. J. S. Franks, K. S. Smith, The role of submesoscale currents in structuring marine ecosystems. *Nat. Commun.* **9** (2018), doi:10.1038/s41467-018-07059-3.

47.	F. M. Lauro, S. J. Senstius, J. Cullen, R. Neches, R. M. Jensen, M. V. Brown, A. E. Darling, M. Givskov, D. McDougald, R. Hoeke, M. Ostrowski, G. K. Philip, I. T. Paulsen, J. J. Grzymski, The Common Oceanographer: Crowdsourcing the Collection of Oceanographic Data. *PLoS Biol.* **12** (2014), doi:10.1371/journal.pbio.1001947.

48.	L. Urban, A. Holzer, J. J. Baronas, M. Hall, P. Braeuninger-Weimer, M. J. Scherm, D. J. Kunz, S. N. Perera, D. E. Martin-Herranz, E. T. Tipper, S. J. Salter, M. R. Stammnitz, *Freshwater monitoring by nanopore sequencing* (2020).

49.   J. S. Cybulski, J. Clements, M. Prakash, Foldscope: Origami-based paper microscope. *PLoS One*. **9**, e98781 (2014).

50.   C. Garcia-Soto, G. I. Van der Meeren, J. A. Busch, J. Delany, C. Domegan, K. Dubsky, G. Fauville, G. Gorsky, K. von Juterzenska, F. Malfatti, G. Mannaerts, P. McHugh, P. Monestiez, J. Seys, J. M. Weslawsky, O. Zielinky, *Advancing Citizen Science for Coastal and Ocean Research* (Position P., 2017; https://brage.bibsys.no/xmlui/bitstream/handle/11250/2476796/EMB_PP23__Citizen_Science_web.pdf?sequence=2).

51.   T. Pollina, A. G. Larson, F. Lombard, L. Hongquan, D. Le Guen, S. Colin, C. de Vargas, M. Prakash, PlanktonScope: Affordable modular imaging platform for citizen oceanography. *One Earth* (2020).

52.   A. Alberti, J. Poulain, S. Engelen, K. Labadie, S. Romac, I. Ferrera, G. Albini, J. M. Aury, C. Belser, A. Bertrand, C. Cruaud, C. Da Silva, C. Dossat, F. Gavory, S. Gas, J. Guy, M. Haquelle, E. Jacoby, O. Jaillon, A. Lemainque, E. Pelletier, G. Samson, M. Wessner, S. G. Acinas, M. Royo-Llonch, F. M. Cornejo-Castillo, R. Logares, B. Fernández-Gómez, C. Bowler, G. Cochrane, C. Amid, P. Ten Hoopen, C. De Vargas, N. Grimsley, E. Desgranges, S. Kandels-Lewis, H. Ogata, N. Poulton, M. E. Sieracki, R. Stepanauskas, M. B. Sullivan, J. R. Brum, M. B. Duhaime, B. T. Poulos, B. L. Hurwitz, S. Pesant, E. Karsenti, P. Wincker, P. Bazire, O. Beluche, L. Bertrand, M. Besnard-Gonnet, I. Bordelais, M. Boutard, M. Dubois, C. Dumont, E. Ettedgui, P. Fernandez, E. Garcia, N. G. Aiach, T. Guerin, C. Hamon, E. Brun, S. Lebled, P. Lenoble, C. Louesse, E. Mahieu, B. Mairey, N. Martins, C. Megret, C. Milani, J. Muanga, C. Orvain, E. Payen, P. Perroud, E. Petit, D. Robert, M. Ronsin, B. Vacherie, P. Bork, E. Boss, M. Follows, G. Gorsky, P. Hingamp, D. Iudicone, L. Karp-Boss, F. Not, J. Raes, C. Sardet, S. Speich, L. Stemmann, S. Sunagawa, Viral to metazoan marine plankton nucleotide sequences from the Tara Oceans expedition. *Sci. Data*. **4**, 1–20 (2017).

53.   O. Flegontova, P. Flegontov, S. Malviya, S. Audic, P. Wincker, C. de Vargas, C. Bowler, J. Lukeš, A. Horák, Extreme Diversity of Diplonemid Eukaryotes in the Ocean. *Curr. Biol.* **26**, 3060–3065 (2016).

54.   Mordret S, Romac S, Henry N, Colin S, Carmichael M, Berney C, Audic S, Richter DJ, Pochon X, de Vargas C, Decelle J, The symbiotic life of Symbiodinium in the open ocean within a new species of calcifying ciliate (Tiarina sp.). *ISME J.* **10**, 1424–1436 (2016).

55.   S. Malviya, E. Scalco, S. Audic, F. Vincent, A. Veluchamy, L. Bittner, J. Poulain, P. Wincker, D. Iudicone, C. de Vargas, A. Zingone, C. Bowler, Insights into global diatom distribution and diversity in the world's ocean. *Proc. Natl. Acad. Sci.* **348** (2015).

56.   A. M. Cabello, F. M. Cornejo-Castillo, N. Raho, D. Blasco, M. Vidal, S. Audic, C. de Vargas, M. Latasa, S. G. Acinas, R. Massana, Global distribution and vertical patterns of a prymnesiophyte–cyanobacteria obligate symbiosis. *ISME J.* **10**, 693–706 (2016).

57.   E. Ser-Giacomi, L. Zinger, S. Malviya, C. De Vargas, E. Karsenti, C. Bowler, S. De Monte, Ubiquitous abundance distribution of non-dominant plankton across the world's ocean. *NEE (in Press.* (2018).

58.   A. Skovgaard, S. A. Karpov, L. Guillou, The parasitic dinoflagellates Blastodinium spp. Inhabiting the gut of marine, Planktonic copepods: Morphology, ecology, and unrecognized species diversity. *Front. Microbiol.* **3**, 1–22 (2012).

59.   I. Probert, R. Siano, C. Poirier, J. Decelle, T. Biard, T. Akihiro, S. Noritoshi, F. Not, Brandtodinium Gen. Nov. (Dinophyceae), a dinoflagellates commonly found in symbiosis with polycystines radiolarians. *J. Phycol.* **50**, 388–399 (2014).

60.   M. R. Willig, D. M. Kaufman, R. D. Stevens, Latitudinal Gradients of Biodiversity: Pattern, Process, Scale, and Synthesis. *Annu. Rev. Ecol. Evol. Syst.* **34**, 273–309 (2003).

61.   J. A. Fuhrman, J. A. Steele, I. Hewson, M. S. Schwalbach, M. V. Brown, J. L. Green, J. H. Brown, A latitudinal diversity gradient in planktonic marine bacteria. *Proc. Natl. Acad. Sci.* **105**, 7774–8 (2008).

62.   D. Boltovskoy, N. Correa, Planktonic equatorial diversity troughs: Fact or artifact? Latitudinal diversity gradients in Radiolaria. *Ecology*. **98**, 112–124 (2017).

63.   J. R. Dolan, E. J. Yang, S. H. Kang, T. S. Rhee, Declines in both redundant and trace species characterize the latitudinal diversity gradient in tintinnid ciliates. *ISME J.* **10**, 2174–2183 (2016).

64.   M. Dufrene, P. Legendre, Species assemblages and indicator species: The need for a flexible asymmetrical approach. *Ecol. Monogr.* **67**, 345–366 (1997).

65.   T. Leeuw, E. Boss, The HydroColor app: Above water measurements of remote sensing reflectance and turbidity using a smartphone camera. *Sensors (Switzerland)*. **18** (2018), doi:10.3390/s18010256.

66.   Y. Yang, L. L. E. Cowen, M. Costa, Is ocean reflectance acquired by citizen scientists robust for science applications? *Remote Sens.* **10**, 835 (2018).

67.   G. Gorsky, G. Bourdin, F. Lombard, M. L. Pedrotti, S. Audrain, N. Bin, E. Boss, C. Bowler, N. Cassar, L. Caudan, G. Chabot, N. R. Cohen, D. Cron, C. De Vargas, J. R. Dolan, E. Douville, A. Elineau, J. M. Flores, J.

F. Ghiglione, N. Haëntjens, M. Hertau, S. G. John, R. L. Kelly, I. Koren, Y. Lin, D. Marie, C. Moulin, Y. Moucherie, S. Pesant, M. Picheral, J. Poulain, M. Pujo-Pay, G. Reverdin, S. Romac, M. B. Sullivan, M. Trainic, M. Tressol, R. Troublé, A. Vardi, C. R. Voolstra, P. Wincker, S. Agostini, B. Banaigs, E. Boissin, D. Forcioli, P. Furla, P. E. Galand, E. Gilson, S. Reynaud, S. Sunagawa, O. P. Thomas, R. L. V. Thurber, D. Zoccola, S. Planes, D. Allemand, E. Karsenti, Expanding Tara Oceans Protocols for Underway, Ecosystemic Sampling of the Ocean-Atmosphere Interface During Tara Pacific Expedition (2016–2018). *Front. Mar. Sci.* **6** (2019), doi:10.3389/fmars.2019.00750.

68. E. Karsenti, A journey from reductionist to systemic cell biology aboard the schooner Tara. *Mol. Biol. Cell.* **23**, 2403–2406 (2012).

69. D. Krishnamurthy, H. Li, F. B. du Rey, P. Cambournac, A. Larson, M. Prakash, Scale-free Vertical Tracking Microscopy: Towards Bridging Scales in Biological Oceanography. *bioRxiv*, 610246 (2019).

70. C. Amante, B. W. Eakins, in *OAA Technical Memorandum NESDIS NGDC-24. National Geophysical Data Center, NOAA* (2009).

71. E. Pante, B. Simon-Bouhet, marmap: A Package for Importing, Plotting and Analyzing Bathymetric and Topographic Data in R. *PLoS One.* **8**, 6–9 (2013).

72. M. Hahsler, K. Hornik, TSP - Infrastructure for the Traveling Salesperson Problem. *J. Stat. Softw.* **23** (2007).

73. T. Rognes, T. Flouri, B. Nichols, C. Quince, F. Mahé, VSEARCH: a versatile open source tool for metagenomics. *PeerJ.* **4**, e2584 (2016).

74. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal.* **17**, 10 (2011).

75. F. Mahé, T. Rognes, C. Quince, C. de Vargas, M. Dunthorn, Swarm v2: highly-scalable and high-resolution amplicon clustering. *PeerJ.* **3**, e1420 (2015).

76. L. Guillou, D. Bachar, S. Audic, D. Bass, C. Berney, L. Bittner, C. Boutte, G. Burgaud, C. De Vargas, J. Decelle, J. Del Campo, J. R. Dolan, M. Dunthorn, B. Edvardsen, M. Holzmann, W. H. C. F. Kooistra, E. Lara, N. Le Bescot, R. Logares, F. Mahé, R. Massana, M. Montresor, R. Morard, F. Not, J. Pawlowski, I. Probert, A. L. Sauvadet, R. Siano, T. Stoeck, D. Vaulot, P. Zimmermann, R. Christen, The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.* **41**, 597–604 (2013).

Figures Caption

Figure 1. *Planktonauts and their routes across the world ocean.* **A.** Examples of pictures (sketch-like filtered) sent by the planktonauts illustrating their actions at sea*:* towing and recovering the P2 plankton net, filtering, drying, and storing plankton samples, recording contextual data and observing plankton through a microscope. **B.** Routes of the main 20 (of 27) pioneer planktonaut crews recruited during the pilot phase of the project, selected to maximize the geographic coverage and oceanographic and sampling conditions. Note the 5 boats from New Zealand who participated to a rally (May to November 2015) organized in collaboration with the 'Island Cruising Association' (http://www.islandcruising.co.nz).

Figure 2. *The extent, novelty, and community composition of P2 metabarcoding data*. **A**. Grouping of P2 (black dots) and *TO* (colored triangles) plankton communities from surface water according to taxonomic compositional similarity (PCA of Hellinger standardized abundances). Colors correspond to the different plankton size-fractions sampled in *TO*. **B**. Geographic distribution and novelty of P2 sequenced samples as compared to *TO* samples. Sampled sites are aggregated in pre-defined geographic area (hexagons) for readability. Hexagons containing P2 samples are filled with a color

gradient corresponding to the number of rDNA OTUs that were not detected in TO (see colored scale). Number inside the colored hexagons indicate the number of samples collected in the area. Empty hexagons are area with only *TO* samples. Hexagons with red-line borders are area with both P2 and *TO* samples. Horizontal dotted red lines indicate the Northern and Southern 60 Degree latitudes; only samples comprised between these lines were kept for ecological comparative analyses between P2 and TO data. **C**. Novel plankton diversity (rDNA OTU) uncovered by each boat. Each dot represents a boat, with its position along the X and Y axes corresponding, respectively, to the number of samples collected, and the number of OTUs recovered by the boat that were not detected in TO. The color-gradient indicates the % of novel P2 OTUs that are unique to the particular boat. Note the higher values for the New Zealander crews (blue characters). **D**. UpSetR plot displaying the taxonomic richness, divided by eukaryotic super-groups, shared between P2 (>20µm) and *TO* size-fractionated samples. The horizontal bars show each individual complete dataset, while vertical bars correspond to the number of OTUs shared between particular datasets (intersections given by the dots under the vertical bars). The tree map shows the taxonomic composition of the 18,430 unique Opisthokonta OTU unveiled in P2, with a dominance (46%) of copepods.

Figure 3: P2 insights into plankton ecology. **A**: *Global ocean latitudinal diversity gradient*. Scatter plot showing the evolution of the Shannon index (alpha diversity) of all P2 samples (colored dots) along latitudes. The color gradient is related to the sea surface temperature measured by the planktonauts during plankton sampling. LOESS regression curve[42] fitting the data. **B-D**: *Plankton biogeography at the scale of a navigation loop* (the 'kiwi loop', around New Zealand, Fiji and New Caledonia, Fig. 2B). **B:**The samples from this loop (colored dots) were clustered into three groups based on their OTU composition (Jaccard distance) using the Partitioning Around Medoids (PAM) algorithm [43]. See Sup. Fig. 9 for details. Two samples from the *Nika* boat, considered as outliers (and collected very close to the coast), were excluded from the analysis. Squared numbers close to each sample site correspond to the sampling boat (1: Astral Express; 2: Celine; 3: Nika; 4: On the Double; 5: Windflower); the colors represent the genetic clusters. **C-D**: Relative abundances of the two most abundant taxonomic groups, the collodarians (**C**) and the copepods (**D**). The color gradient and the size of the circles both correspond to the relative abundance.

Supplementary Figure 1. *Original Plankton Planet strategy*, showing the successive steps of both the seatizen (right loop, a to f) and researchers (left loop, g to j) operations, respectively at sea and in the lab.

Supplementary Figure 2. *The original Plankton Planet sampling kit*. A. The P2 plankton net (25cm diameter, 20µm mesh size) with a 2kg weight and a small float to maintain it at a maximum depth of 3m. B. The manual vacuum pumping system used to transfer plankton from the cod-end of the net onto a 10µm filter membrane. The total cost for one kit approximates 700 US$ but could be largely reduced for mass production via tinkering of home-made parts in FabLabs.

Supplementary Figure 3. *Pictures illustrating various steps of the sampling protocol for planktonauts at sea.* 1. Training of a group of planktonauts, here in Auckland with Dr. Pochon. 2. After 15-20min of towing at a maximum speed of 2 knots, the plankton net is recovered on board, here onboard *Tethys*. 3. Recording of contextual parameters, here ocean color using the *Hydrocolor App* on board *Zigomar*. Note that 3 families participated to the pilot project, showing that even kids can realize parts of the protocol (see for instance: https://vimeo.com/219660346); 4. Pouring of the concentrated plankton from the net cod-end into the manual vacuum-pumping, onboard *Taravana*. 5. Manual recovery of the 10µm filter membrane full of plankton biomass. 6. Gentle drying of the filter membrane in a pan on the boat (*Taravana*) gas cooker. 7. Storage and labeling of dried plankton samples into zip lock plastic bags before shipping to the laboratory by regular mail.

Supplementary Figure 4: *Testing the effect of dried (P2) versus flash-frozen plankton samples on DNA-based community composition*. A. Geographic location of plankton samples collected by volunteer citizen sailors along the French coast during the 'Objectif Plankton' initiative (https://riem-asso.com/objectif-plancton/). Bay of Brest, June 21st 2016, on the left, and Bay of Lorient, June 14th 2016, on the right; OPB and OPL indicate samples from Brest and Lorient bays, respectively. Each sample was preserved by both desiccation (Dry) and Flash Freezing into Liquid Nitrogen (FF), before deep Illumina sequencing of V9 SSU rDNA amplicons. B. Heatmap of the Hellinger distance (Hellinger standardization followed by Euclidian distance) between samples and sub-samples. C. Same as B, using Jaccard distances (based on OTUs presence/absence). The relatively high values of the Jaccard distances between preservation methods from the same samples comes from the fact that low abundance OTUs create a noisy background that inflates dissimilarities. The Hellinger distances are less sensitive to this issue, and the dissimilarities values range from 0.15 to 0.37 between preservation methods, while they are close to 1 between samples. The residual dissimilarities observed between subsamples are due to Illumina sequencing errors and are expected between technical replicates.

Supplementary Figure 5: *Geographical coverage and putative cost of P2-pilot samples*. Location of 233 (out of 258) ocean sites sampled by the pioneer planktonauts in 14 months during the Plankton Planet pilot project. The blue dots are the sampling sites taken into account to compute the shortest possible route by a putative single oceanographic vessel in the Indo-Pacific and Atlantic basins, respectively ('traveling salesman problem', see Sup. Text 1). The green dots are sampling sites considered as land sample (depth >0 m) because of their proximity to islands and the lack of resolution of the bathymetric database used to compute the shortest path without crossing the land between samples (see Sup. Text 1) . The red dots in the Mediterranean sea and Arctic ocean, correspond to sampling sites that were not integrated into the 'traveling salesman problem'. The Atlantic loop, starting and ending in Brest, pass through 62 sites and is 44,972 km long; the Pacific loop, starting and ending in San Diego, pass through 82 sites and is 68,985 km long. Using a conventional oceanographic vessel (30,000 $US a day) travelling at 10 knots and stopping at each station for 1 hour, sampling these two loops would cost circa 7,9 million $US in total.

Supplementary Figure 6: Box plot representation of the ***percentage of prokaryotic (Archaea and Bacteria) rDNA reads P2 samples*** for each boat, and for *Tara* Oceans samples (micro- and meso-planktonic size fractions). 85% and 89% of P^2 and *TO* samples, respectively, contain less than 1% of prokaryotic reads, the difference being explained by a few outliers. Note the higher value for a few samples from *Amasia* and *Race for Water*, which correspond to samples that were kept on board at room temperature for several days before processing.

Supplementary Figure 7: ***Phylogenetic breakdown of the Tara Oceans and Plankton Planet global ocean metabarcoding datasets at the eukaryotic supergroup and 'taxogroup' (sensu[1]) levels***. All V9 rDNA reads and OTUs with genetic similarity to a eukaryotic reference sequence ≥80% were retained and taxonomically assigned. The tree-maps display the relative abundance (upper part) and richness (lower part) of the different taxonomic groups in *T*O plankton size fractions and P2. The category '*Orphans'* contains the known but phylogenetically uncertain deep-branching lineages (i.e. Haptophyta, Telonemida, Picomonadida, Katablepharidida, Cryptophyta, Centrohelida and Apusozoan); '*Unknown'* corresponds to OTUs assigned to two different supergroups.

Supplementary Figure 8: ***Phylogenetic and plankton size fraction distribution of the most abundant OTUs assigned to Peridiniales dinoflagellates in both the P2 and TO metabarcoding datasets***. The 100 most abundant Peridiniales OTUs in TO and the 50 most abundant in P2 were selected, together representing 119 unique OTUs. Colored bars indicate the proportion of mean relative abundance of each OTU amongst the four *T*O plankton size fractions. Black bars indicate the mean abundance (log transformed and scaled to 1) of each OTU amongst all P2 samples. Flag and red crosses symbols indicate OTUs observed respectively only in P2 and only in TO.

Supplementary Figure 9: ***Details about the three plankton community clusters generated for the South-West Pacific loop samples (Fig. 4B)***. **A**: Geographic distribution of indicator OTUs among taxonomic groups for each cluster. For each OTU, the indicator values, i.e. the product of the relative frequency and relative average abundance in clusters, are computed from Hellinger transformed abundances for each of the three clusters using the indval() function (labdsv R package, [49]). OTUs are considered as indicator of a specific cluster when p-value are below 0.01. **B**: Consistency of the clusters using the silhouette width (measure of how similar a sample is to its own cluster compared to other clusters). The silhouette width has been computed using the silhouette() function (cluster R package[43]).

Supplementary Text 1. Material & Methods

P2 protocols: plankton DNA sampling, storing, shipping, extracting, and sequencing

After 15-20 min of net towing at a speed of ~2 knots (Sup. Fig. 2), 500mL of concentrated seawater were split and manually filtered onto two replicate 10µm polycarbonate filter-membranes (diameter 47mm), dried at 70°C for 5 min in a pan (on the boat gas-cooker), and then stored at room temperature in a hermetic Ziploc bag with granular desiccants (Sup. Fig. 2, 3). Upon arrival in a port, samples were shipped to the Roscoff Marine Laboratory (Brittany, France) in a simple envelope via regular mail. In the laboratory, the samples were stored at -80°C and information related to samples was archived in a database together with contextual data.

DNA extraction was performed using a protocol modified from the *NucleoSpin Plant Midi* kit (Macherey-Nagel). One replicate filter of each sample was cut in small pieces and incubated for 2h at 56°C with 3.6mL of the lysis buffer PL1 and 250µL of proteinase K. The lysate was transferred to a large capacity *NucleoSpin* Filter (*DNA Midi kit*) and centrifuged for 10 min at 1,500 g. The eluate was transferred to a new tube and 1 volume of PC buffer was added. The mixture was loaded into the appropriate spin column and washed 3 times with the DNA wash solution. Total DNA was finally eluted twice with 150 µl of DNA elution buffer, and stored in sterile microtubes at -20°C. The amount of recovered DNA was quantified by dsDNA-specific fluorimetry using a Qubit 2.0 Fluorometer with Qubit dsDNA Broad Range and High Sensitivity Assays (ThermoFisher Scientific, Waltham, MA). The DNA quality was double-checked in a subset of samples by running 1 µl on 1,2% agarose gel for 45 min at 120V.

To address general questions of eukaryotic biodiversity over extensive taxonomic and ecological scales, the hyper-variable loop V9 of the Small Sub-Unit (SSU) ribosomal (r) RNA gene was targeted for the generation of amplicons by Polymerase Chain Reaction (PCR). This barcode presents a combination of advantages: (*i*) it is universally conserved in length 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 amongst all known lineages. Amplifications of the V9 from SSU rDNA were conducted with Phusion® High-Fidelity DNA Polymerase (Finnzymes) using the PCR primers 1389f 5'- TTGTACACACCGCCC -3' and 1510r 5'- CCTTCYGCAGGTTCACCTAC -3'. The PCR mixture (25µL final volume) contained 5ng of template with 0.35µM final concentrations of each primer, 3% of DMSO and 2X of GC buffer Phusion Master Mix (Finnzymes). Amplifications were conducted following the PCR program: initial denaturation step at 98°C for 30 sec, followed by 25 cycles of 10sec at 98°C, 30sec at 57°C, 30sec at 72°C, and a final elongation step at 72°C for 10 min. Each sample was amplified in triplicate to get enough amounts of amplicons. Products of the reactions were run on a 1.5% agarose gel to check for successful amplification products of the expected length. Amplicons were then pooled and purified using the NucleoSpin® PCR Clean Up kit (Macherey-Nagel, Hoerdt, France), and finally quantified with the Quant-iT™ PicoGreen ® dsDNA kit (Invitrogen). Purified amplicons were sent to Genoscope for HiSeq sequencing.

Sampling coverage and putative cost

In order to compare ocean sampling cost between P2 sailing boats and classical oceanographic vessel, we measured the shortest possible route between P2 sampling sites (traveling salesman problem) by a putative oceanographic vessel. For convenience, 144 P2 sampling sites (56% of total sites) from the Pacific (83), Indian (1) and Atlantic (60) Oceans with a depth >0m according to the ETOPO1 NOAA database (*70*) were selected. We then considered two ports as starting points for two navigation loops: Brest for the Atlantic Ocean samples plus 2 Pacific Ocean samples located at the extreme South of South America (62 samples in total); and San Diego for the other Pacific and Indian oceans samples (82 samples in total). For each group of stations, the shortest possible route that visits each station and return to the starting port was defined. To avoid crossing lands, the least cost distances between locations were computed by constraining paths to depths below 0 m. The distance matrix was computed using the lc.dist() function from the marmap R library (*71*). From this matrix, the traveling salesman problem was solved for each basin with the arbitrary insertion algorithm followed by a two edge exchange improvement procedure using the solve_TSP() function with default options (TSP R library (*72*)).

Bioinformatic data processing

In order to compare P2 DNA metabarcoding data to the primary *Tara* Oceans (*TO*) eukaryotic metabarcoding dataset (*34*), we first merged raw reads from both datasets and applied the following bioinformatics steps. Paired Illumina™ MiSeq reads from the 214 P2 samples and the 883 samples from *Tara* Oceans (2009-2012) were assembled with vsearch v2.7.1 (*73*) using the command fastq_mergepairs and the option fastq_allowmergestagger. Demultiplexing and primer clipping were performed with cutadapt v1.9 (*74*) enforcing a full-length match for sample tags and allowing a 2/3-length partial match for forward and reverse primers. Only reads containing both primers were retained. For each trimmed read, the expected error was estimated with vsearch's command fastq_filter and the option *eeout*. Each sample was then dereplicated, i.e. strictly identical reads were merged, using vsearch's command derep_fulllength, and converted to FASTA format. To prepare for clustering, samples were pooled and submitted to another round of dereplication with vsearch. Files containing expected error estimations were also dereplicated to retain only the lowest expected error for each unique sequence. Clustering was performed with Swarm v2.2.2 (*75*), using a local threshold of one difference and the *fastidious* option. The representative sequences of each molecular operational taxonomic unit (OTU) were then searched for chimeras with the vsearch's command uchime_denovo(Edgar et al., 2011). In parallel, the OUT representative sequences were taxonomically assigned with the stampa pipeline (https://github.com/frederic-mahe/stampa/) and a custom version of the Protist Ribosomal Reference database PR2 (*76*). This custom reference database is an update of the V9_PR2 reference database used for the taxonomic assignation of the *Tara* Oceans metabarcodes (*34*).
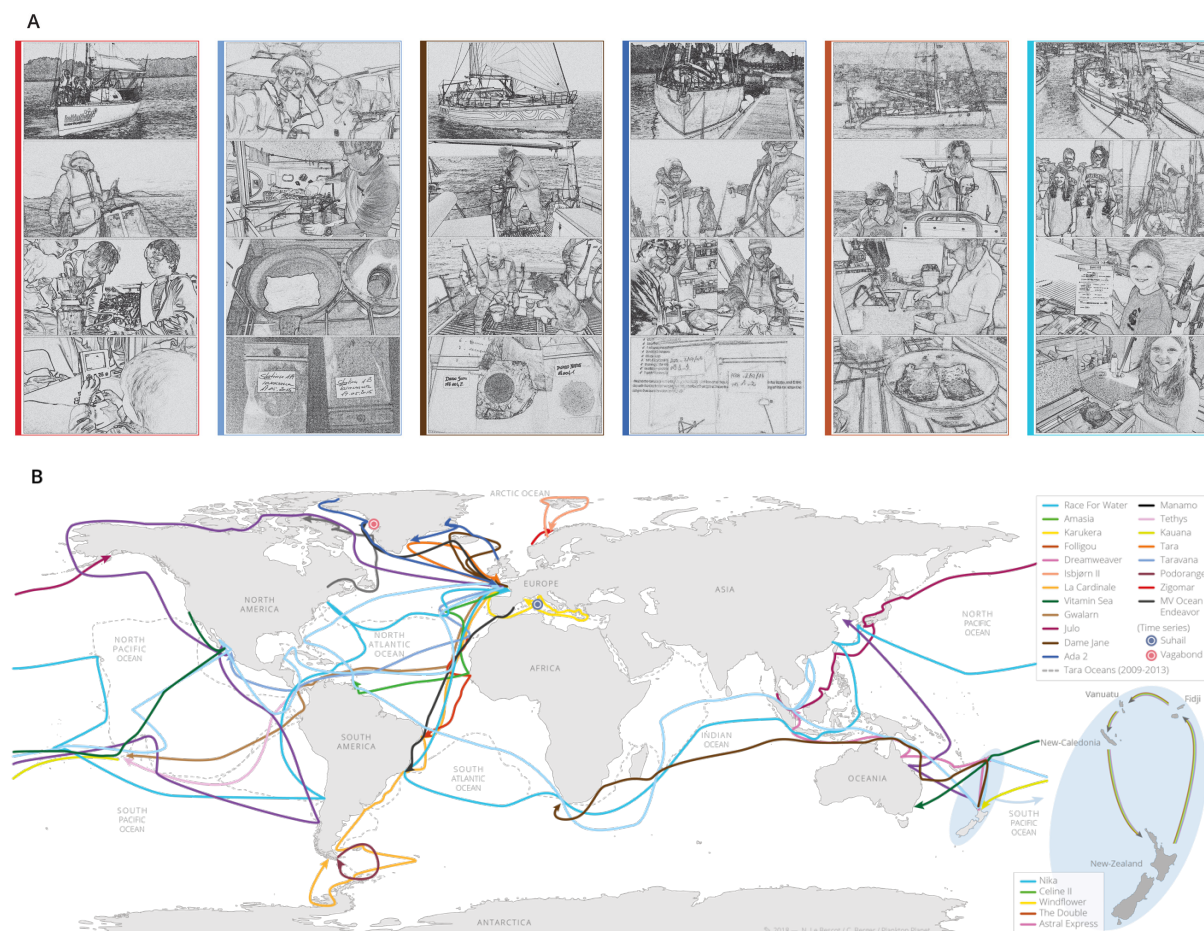
Clustering results, expected error values, taxonomic assignments and chimera detection results were used to build a raw OTU table. Up to that point, reads that could not be merged, reads without tags or primers, reads shorter than 32 nucleotides and reads with uncalled bases ("N") were eliminated. To create the "cleaned" OTU table, additional filters were applied to retain only: non-chimeric OTUs, OTUs with an expected error per nucleotide below 0.0002, and OTUs containing more than 3 reads or seen in 2 samples. The final OTU table contained integrated the 214 P2 samples together with the 386 *Tara* Oceans samples collected in surface waters for 4 organismal size fractions (0.8-5, 5-20, 20-

180, 180-2000 µm). This table was subsampled (rarefied) at the minimum number of reads observed for a sample (313,539 reads) for comparison of OTU richness between samples. Based on the rarefied OTU table, an OTU is considered as unique to P2 if it is present in at least one P2 sample and not present in any of the 386 *Tara* Oceans surface samples. The plankton samples from '*Objectif Plancton*' used to compare the preservation methods were processed independently using the same bioinformatic pipeline. Raw data are available under the bioproject XXXXX (https://www.ncbi.nlm.nih.gov/bioprojectXXXXX).

boilerplate: bioRxiv preprint doi: https://doi.org/10.1101/2020.08.31.263442; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.
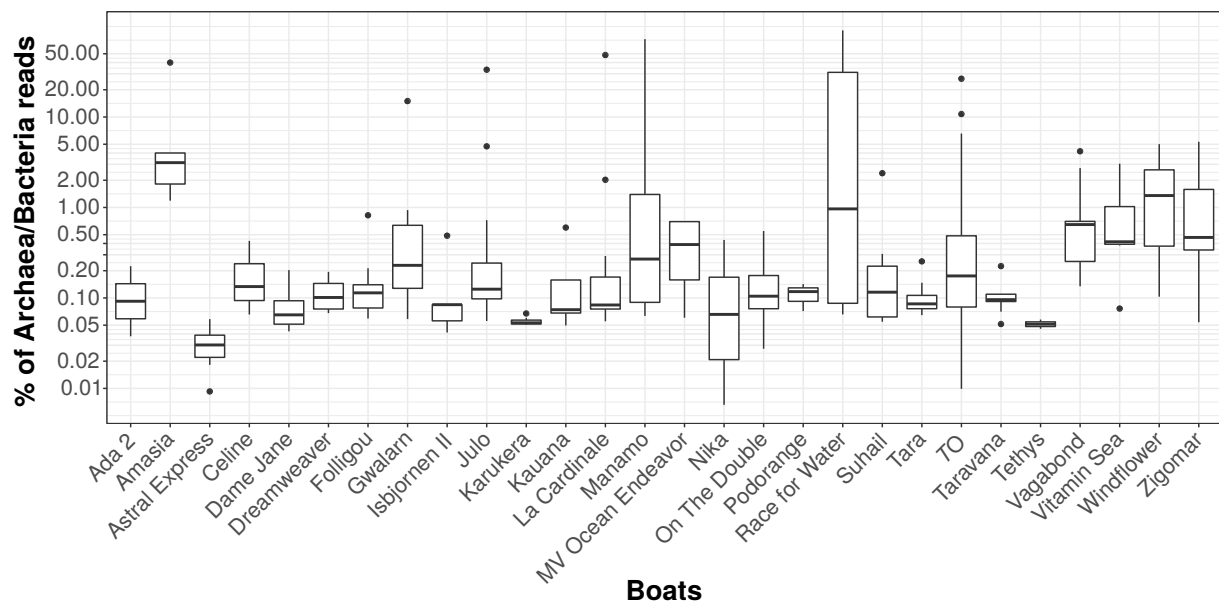
Plankton Planet – Proof-of-Concept & Perspectives
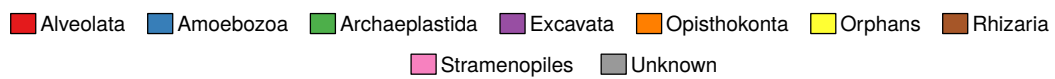
Figure 1

Plankton Planet – Proof-of-Concept & Perspectives

Figure 2
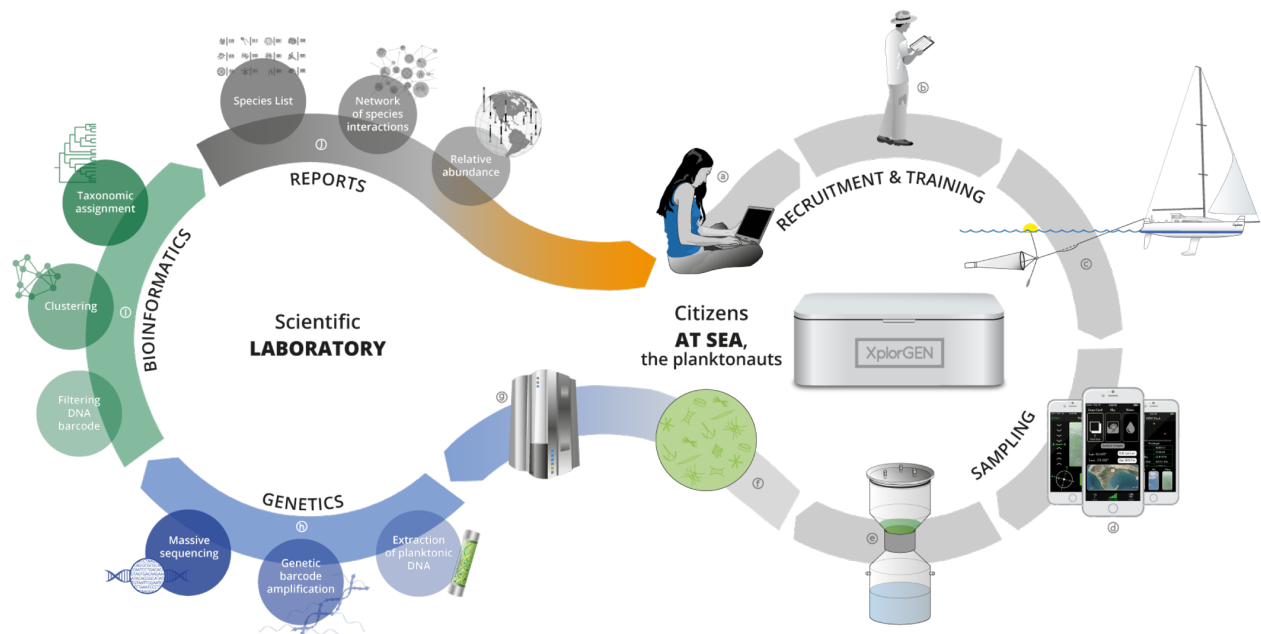
Figure 3

Supplementary Figure 1
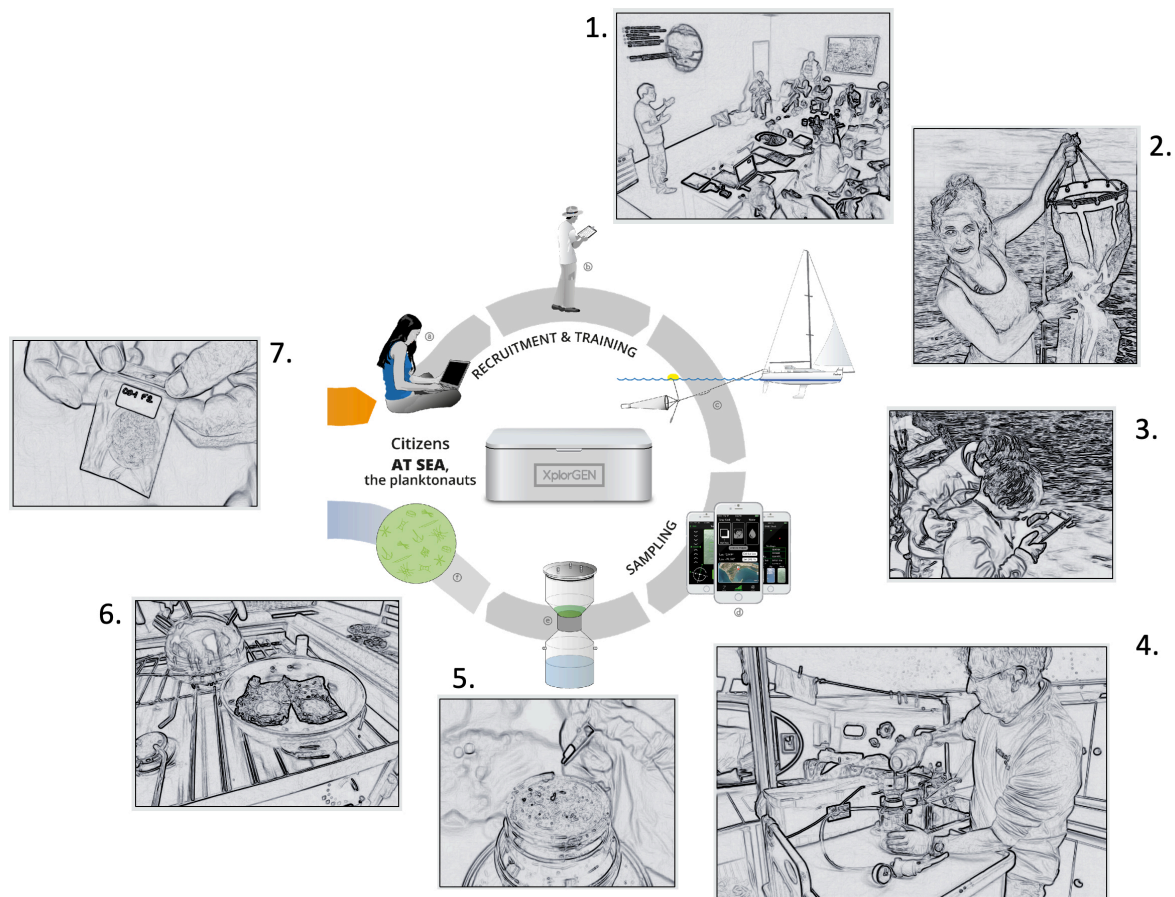
Supplementary Figure 2

Supplementary Figure 3

Supplementary Figure 4
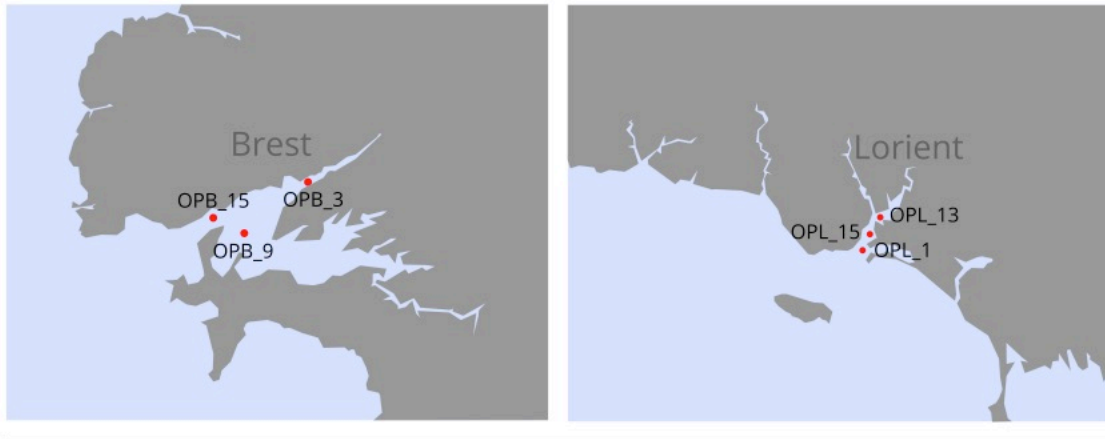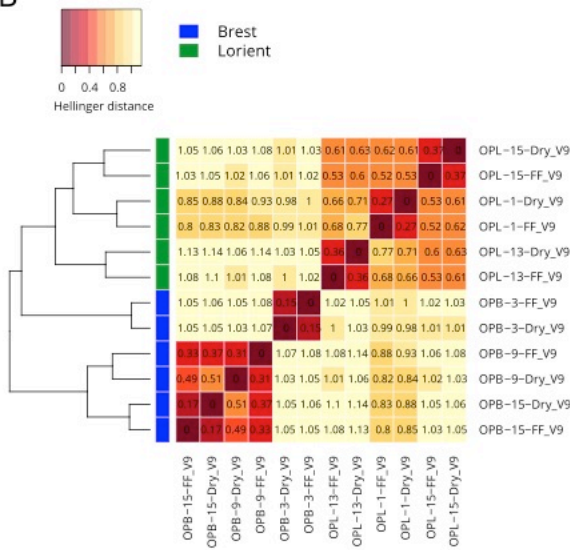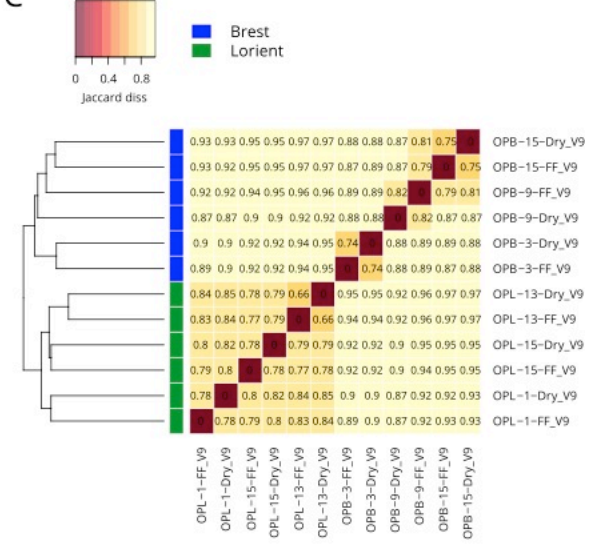
Plankton Planet – Proof-of-Concept & Perspectives

Supplementary Figure 5

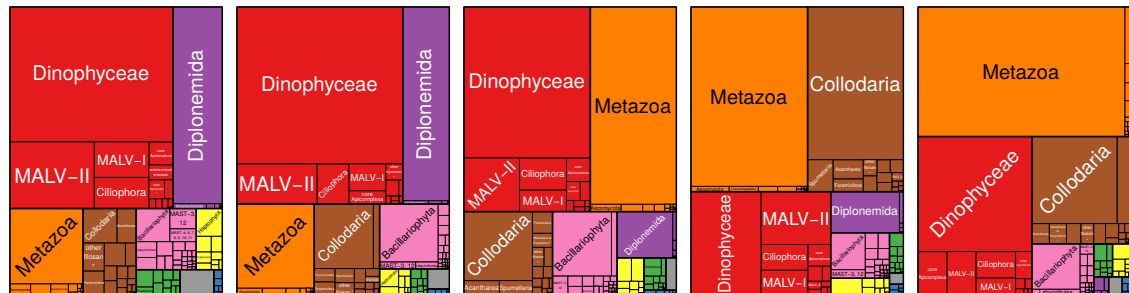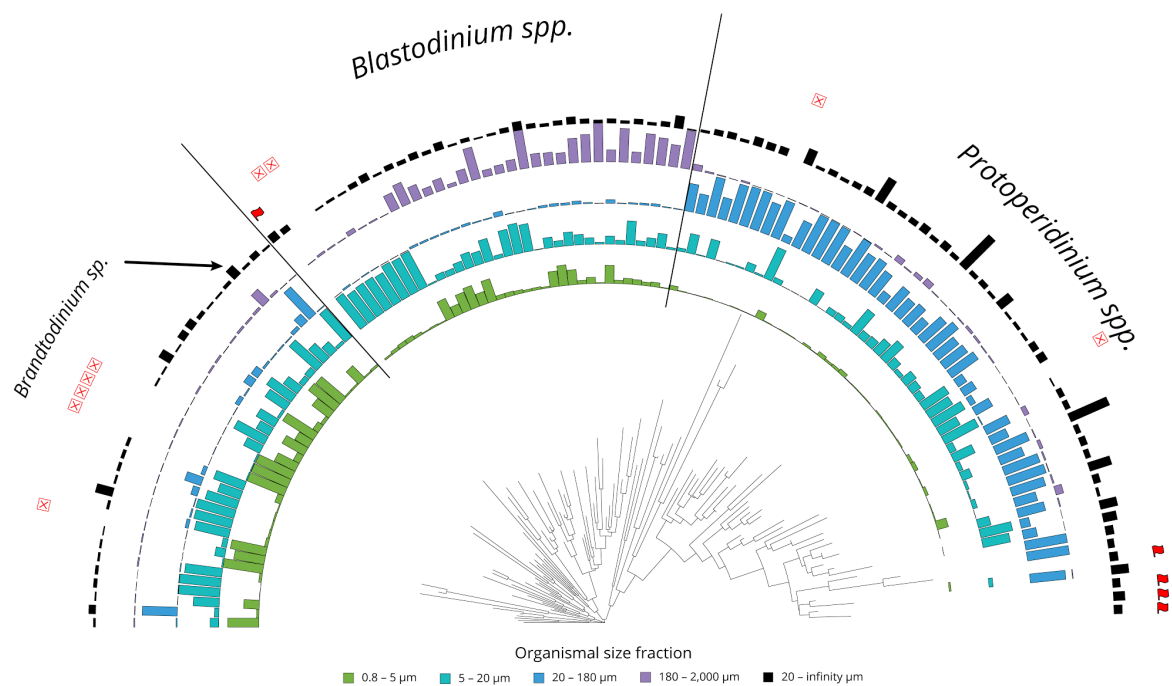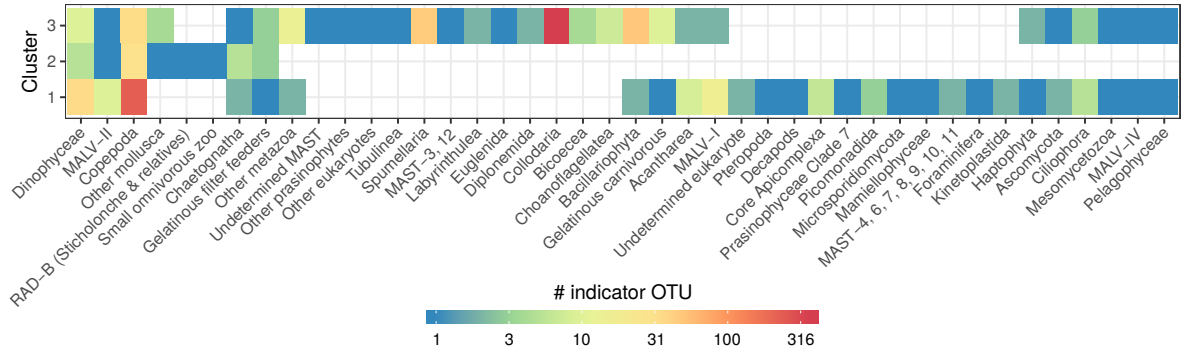Plankton Planet – Proof-of-Concept & Perspectives

Supplementary Figure 6

Supplementary Figure 7

Supplementary Figure 8

Supplementary Figure 9