Population dynamics of the cosmopolitan eukaryotic picophytoplankton Bathycoccus during seasonal blooms in the bay of Banyuls sur Mer (North Western Mediterranean sea)

Martine Devic*[1], Cédric Mariac[2], Valérie Vergé[1], Philipe Schatt[1], Louis Dennu[1], Jean-Claude Lozano[1], François-Yves Bouget*[1] and François Sabot*[2]

Author affiliations :

[1] Laboratoire d'Océanographie Microbienne (LOMIC), CNRS/Sorbonne Université, UMR7621, Observatoire Océanologique, 66650 Banyuls/mer, France.

[2] Diversité, adaptation et développement des plantes (DIADE), UMR 232 IRD/UM/CIRAD, Centre IRD de Montpellier, 911 avenue Agropolis, BP 604501, 34394, Montpellier Cedex 5, France.

* corresponding authors : martine.devic@obs-banyuls.fr; francois-yves.bouget@obs-banyuls.fr; francois.sabot@ird.fr

Running head : Markers of genetic diversity in Bathycoccus

Key words : Phytoplankton, Bathycoccus, intra-specific diversity, marker, bloom

word count for Abstract, Introduction, Materials and Methods, Results and Discussion : 7569 words

Abstract (251 words)

Although Bathycoccus is one of the most abundant picophytoplankton, little is known about the genetic diversity underlying its adaptation to ecological niches. In this study, the diversity of Bathycoccus populations during their annual bloom in the Mediterranean bay of Banyuls France was assessed by an INDEL based approach. Oxford Nanopore Technology (ONT) was used to characterise structural variants (SV) among the genomes of Bathycoccus sampled from geographically distinct regions in the world ocean. Markers derived from INDEL were validated by PCR and sequencing in the world-wide strains. These markers were then used to genotype 55 Bathycoccus strains isolated during the winter bloom 2018-2019 in Banyuls. With five markers, eight Multi Loci Genotypes (MLG) were determined, two of which represented 53% and 29% of the isolates. Physiological studies confirmed that isolates are phenotypically different, cells isolated in February growing better at low temperature than those isolated in December. When tested on environmental samples, two diversity markers showed a similar allele frequency in sea water as in individual Bathycoccus strains isolated at the same period. We conclude that these markers constitute a resource to identify the most abundant variant alleles in a given bloom. A follow-up on three consecutive blooms revealed differences in allele abundance during the course of a bloom, particularly at initiation, and between years. In addition to *Bathycoccus prasinos*, two other species of Bathycoccus were identified including the recently described species *B. calidus* and a novel species *B. catiminus*, suggesting that species diversity of the genus Bathycoccus may be underestimated.

Introduction (929 words)

Marine phytoplankton, including picoalgae, is responsible for a large fraction of primary production (Li et al., 1983). In temperate regions, the abundance and diversity of the phytoplankton is often seasonal and occurs in bursts, as algal blooms. Per se, blooms have a large impact on global primary production and therefore the understanding of the effects of ocean warming on phytoplanktonic blooms is of the utmost importance.

In the bay of Banyuls, Mamiellophyceae (Bathycoccus and Micromonas) bloom yearly from November to April. Previous work centred on the occurrence and abundance of different species during the course of a bloom showed that Bathycoccus was one of the most abundant species (Lambert et al., 2019). Bathycoccus together with Ostreococcus and Micromonas are picoalgae in the order of Mamiellales belonging to the green lineage. Widely distributed from the equator to arctic and antarctic poles with a marked seasonality in temperate and polar regions (Joli et al., 2017, Tragin et al., 2018, Lambert et al. 2019, Leconte et al., 2020), this cosmopolite presence illustrates a high capacity for adaptation to a wide range of contrasting environments. The highly reproducible yearly reoccurrence in the Banyuls bay during the last decade (Lambert et al., 2019) also raises the question of the persistence of a Bathycoccus population adapted to the bay or of a variation of the population structure each year. In addition, since outside of the bloom period, Mamiellales are virtually absent from the bay, is the Bathycoccus bloom initiated by an uptake of resident “resting cells” in the sediment or by a fresh input carried by North western Mediterranean currents along the gulf of Lion? At present no resting stages that can act as inoculum of subsequent blooms have been described for Bathcoccus.

Assessing interspecies diversity uniquely underestimates the diversity of the populations. Equally, a single isolate cannot represent the diversity of a population. Since natural selection

acts on variation among individuals within populations, it is essential to incorporate both intra- and inter-specific trait variability into community ecology (Violle et al. 2012). Raffard et al (2019) demonstrated that intraspecific variation has significant ecological effects across a large set of 52 species, confirming a previous estimate based on a more restricted species set (Des Roches et al., 2018). Furthermore, it has been shown that diversity within species is rapidly decreasing, making them more homogenous and highlighting the need to preserve intraspecies variations (Des Roches et al. 2018) since intra-species diversity reinforces the overall population stability in the face of environmental change.

In most studies, phytoplankton diversity and abundance have been primarily determined by microscopy at species level and by metabarcoding on the nuclear or plastidial 18/16S rRNA gene. It is now imperative to incorporate intraspecific variability into population studies.

In diatoms, intraspecific variation has been shown to play a key role in the responses of the species to several important environmental factors such as light, salinity, temperature and nutrients (Godhe et al. 2017). Modelling efforts indicate that this variation within species extends bloom periods and likely provides sufficient variability in competitive interactions between species under variable conditions. The intraspecific variation most likely corresponds to optimal fitness in temporary microhabitats. This rich intraspecific genetic diversity allows for the possibility of local adaptation and for differentiation in important physiological characteristics that produces local populations that are exceptionally fit and competitive in their respective local habitat.

To date, very little information is available on intraspecies diversity of Bathycoccaceae with the exception of Ostreococcus (Blanc-Mathieu et al, 2017). Bathycoccus may be divided in two clades or species, the polar and temperate *Bathycoccus prasinos* type B1 genome (Moreau et al. 2012; Joli et al. 2017) and the tropical *Bathycoccus calidus* type B2 genome (Vannier et al. 2016; Limardo et al. 2017; Bachy et al. 2021). Thus the cosmopolitan nature of

Bathycoccus from poles to equator might be due to the combination of the B1 and B2 clades or species. Several studies suggest that previously recognized cosmopolitan species are actually composed of multiple populations or even multiple species (Kashtan et al., 2014). Genotypes or species can either replace each other temporally (but with overlap) as in the case of the marine diatoms *Pseudo-nitzschia multistriata* (Tesson et al., 2014) and *Skeletonema costatum* (Gallagher, 1980) or co-exist sympatrically as in the freshwater *Asterionella formosa* (Van Den Wyngaert et al., 2015). Furthermore, the frontier between variant genotype and species is thin and knowledge will be gained by whole genome sequencing of a large number of accessions. Read et al. (2013) reveal a pan genome of the coccolithophore *Emiliania sp* suggesting that what was previously considered a single species, is actually composed of multiple species.

One of the major challenges in the study of marine phytoplankton intraspecies variation is the difficulty to isolate individuals in sufficient number for classical diversity analysis (microsatellites : Srivastava et al. 2019, mitochondrial DNA : Galtier et al. 2009, chloroplastic DNA : Wheeler et al. 2014, nuclear Simple Sequence Repeats SSR, Single Nucleotide Variant (SNV). Microsatellites have been described in some diatoms (Tesson et al. 2011) but not yet in Mamiellales. Meta-ribosomal barcoding has opened the access to massive data in time and space and has accelerated the study of natural communities, inter-species co-occurrence and potential biotrophic interactions. Now it is imperative to develop new tools to study intraspecies diversity in environmental samples. To address this problem, we combined an efficient method to isolate Mamiellales with whole genome sequencing by Oxford Nanopore Technology (ONT) in order to identify Structural Variants (SV) in the Bathycoccus genome. Diversity markers designed from INDEL were used to genotype Bathycoccus accessions and populations from environmental samples.

Materials and Methods (885 words)

Algal strains and culture conditions

World-wide Bathycoccus strains were obtained at the Roscoff Culture Collection (RCC) centre: RCC4222, RCC5417, RCC1613, RCC685, RCC1615, RCC1868, RCC4752 and RCC716. The strains were cultivated in 100 mL flasks in filtered artificial seawater (24.55 g/ L NaCl, 0.75 g/L KCl, 4.07 g/L $MgCl_2$ $6H_2O$, 1.47 g/L $CaCl_2$ $2H_2O$, 6.04 g/L $MgSO_4$ $7H_2O$, 0.21 g/L $NaHCO_3$, 0.138 g/L NaH_2PO_4 and 0.75 g/L $NaNO_3$) supplemented with trace metals and vitamins. Cultures were maintained under constant gentle agitation in an orbital platform shaker (Heidoph shaker and mixer unimax 1010). Realistic sunlight irradiation curves were applied during the light period in temperature-controlled incubators (Panasomic MIR-154-PE).

Cell isolation

Surface water was collected at 3 meters depth at SOLA buoy in Banyuls bay, North Western Mediterranean Sea, France (42°31′N, 03°11′E) approximately every week from December 2018 to March 2019, November 2019 to March 2020 and October 2020 to April 2021. Two ml aliquots were used to determine the quantity and size of phytoplankton by flow cytometry. For the bloom 2018/2019, 50 ml was filtered through a 1.2-µm pore-size acrodisc (FP 30/1.2 CA-S cat N° 10462260 Whatman GE Healthcare Sciences) and used to inoculate 4 culture flasks with 10 ml of filtrate each. The sea water was supplemented by vitamins, NaH2PO4, NaNO3 and metal traces at the same concentration then artificial sea water (ASW), antibiotics (Streptomycine sulfate and Penicillin at 50 µg/ml) were added in half of the cultures. The cultures were incubated under light and temperature conditions similar to those at sampling date for 3-4 weeks. The presence of picophytoplankton was analysed by a BD accury C6 flow cytometer. In general, superior results were obtained without antibiotics. Cultures containing

at least 90% of picophytoplankton with only residual nanophytoplankton were used for plating on agarose. Colonies appearing after 10 days were hand picked and further cultured in 2 ml ASW in deepwell plates (Nunc, Perkin Elmer, Hessen, Germany) for 10 days. Cells were cryopreserved at this stage. Circa 500 clones were cryopreserved. At the same time, DNA extraction and PCR were performed in order to identify Bathycoccus clones.

DNA extraction, genome sequencing, assembly and PCR amplification

For PCR analysis, total DNA was extracted from 4 ml Bathycoccus cell cultures according to the Plant DNA easy Qiagen protocol. For whole genome sequencing by Oxford Nanopore technology (ONT), DNA was extracted by a CTAB method from 100 ml culture principally based on Debladis et al. (2017). ONT libraries were barcoded using the Rapid Barcoding Sequencing (SQK-RBK004) and deposited on R9.4 flow cell. For environmental samples, 5 litres of seawater at SOLA 3 meters depth were passed through 3 micron and 0.8 micron pore filters. DNA from cells collected on the 0.8 micron filters were extracted using the Plant DNA easy Qiagen protocol with the addition of a proteinase K treatment in the AP1 buffer. PCR was performed using the Red Taq polymerase Master mix (VWR) with the required primers (Supplemental data 2) and corresponding DNA. For sequencing, the PCR products were purified using the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel reference 740609.50) and the filtrate was sent to GENEWIZ for Sanger sequencing.

Raw ONT Fast5 data were basecalled using Guppy 4.0.5 (https://nanoporetech.com) and the HAC model, and QC performed using NanoPlot 1.38.1 (De Coster et al. 2018). All reads with a QPHRED higher than 8 were retained and subject to genome assembly using Flye 2.8 (Kolmogorov et al. 2019) under standard options. Raw assemblies were then polished with 3 turns of standard Racon (Vaser et al. 217) after mapping of raw reads on the previous sequence using minimap2 (-ax map-ont mode; Li et al. 2021). Final scaffolding was performed using Ragoo 1.1 (Alonge et al. 2019) upon the original B. prasineos reference

genome (GCA_002220235.1, Moreau et al. 2012). Final QC of assemblies was performed using QUAST 5.0 (Mikheenko et al. 2018).

Relative allelic abundance in environmental samples

Amplifications were performed twice with a difference of 2 cycles in order to obtain clear bands on ethidium bromide stained agarose gels for each sample. Similarly the gels were photographed after different exposure times in order to obtain a non-saturated image for each sample. Relative abundance of each variant within the same DNA sample was performed using ImageJ software Analyse Gel.

Determination of Growth Rates

For each culture condition the cell number was determined by flow cytometry daily, for 10 days. The growth rate in batch culture was determined as Ln(N)/dT, where N is the cell concentration per ml and T the time (days). The maximal growth rate (μmax) was determined according to Guyon et al. (2018) on a graph expressing the neperian logarithm of cell concentration as a function of time of culture. Mmax corresponded to the slope of the linear part of the growth curve (i.e., excluding the lag phase and the stationary phase).

$$\mu_{max} = Log(N_{fmax}) - Log(N_0)/Log(2) \times T$$

Measurement of photosynthetic capacities

Cultures were acclimated to the temperature and light rhythm and intensity for 7-10 days before subculturing at 10^6 cells /ml in triplicates. After 3-4 days of growth, the photosynthetic activities were recorded with PHYTO-PAM-II (Walz). After 20 min in obscurity, the samples were transfer into the PHYTO-PAM and Fv/Fm, ETRmax and NPQ were measured.

Ultrastructure of Bathycoccus species determined by transmission electronic microscopy

Cells were prepared according to Chrétionot-Dinet et al. (1995). Thin sections were stained

with uranyl acetate and lead citrate and observed with a 7500 Hitachi transmission electronic microscope.

Results (3164 words)

Search for intraspecies diversity markers

With the aim to differentiate Bathycoccus isolates, we undertook a search for genetic determinants of diversity. Since only two Mediterranean strains were available in the Roscoff culture collection at the beginning of the project, we examined the world-wide diversity of Bathycoccus and selected the most geographically dispersed accessions (Supplemental Table 1). Bathycoccus cells differ at the genome level and can be differentiated in two types, B1 and B2 genomes (Limardo et al. 2017). Roughly the accessions possessing the B1 genome run along a latitude gradient from the Baffin bay (67°) to the Mediterranean sea (40°). The most tropical accession, RCC716 from the Indian Ocean (-14°) was not included in this analysis since it has a B2 genome, recently described a novel species named *Bathycoccus calidus* (Bachy et al. 2021). Oxford Nanopore Technology (ONT) was used to sequence the genome of 7 B1 accessions (RCC4222 clone replaces RCC1105 which was lost). After *de novo* assembly, each genome was compared to the reference genome of Banyuls isolate RCC1105. There were some large chromosomal rearrangements but for the design of diversity markers, we only considered INDEL inferior or equal to 2 kb within regions mapping on the RCC1105 genome. The number and size of INDEL are detailed in Table 1. The goal was to identify INDEL instead of SNP (Single Nucleotide Polymorphism) that could be used to genotype the strains directly by PCR.

Validation of sequence variations in the genomes of world-wide Bathycoccus

Putative markers were selected on several criteria. The insertion should be found at the same or close location in the genome of at least three accessions and of different sizes in at least two genomes, preferably three. In addition, the size of the amplified fragment should be between 200 bp and 2 kb (this size restriction reduced the mean number of insertions from 88 to 35, Table 1) and sufficiently different among the genomes of the accessions to be unambiguously visualised on agarose gel after amplification with a single set of primers. The aim of this drastic selection was to identify the most divergent markers among the largest available genetic diversity of Bathycoccus with the expectation that some of this variation would be found in local communities of Bathycoccus in the Banyuls bay. Only five candidate markers met these criteria and were experimentally tested. For two markers targeting variations of the number of amino acid repeats in open reading frames, primers positioned at proximity of the repeats did not produce a single amplimer and these two predictions from ONT could not be validated nor invalidated. Marker on chromosome 15 (the number of repeats in a zinc finger protein), marker chromosome 14 (variation in repeat number in a flavodoxin-like protein) and marker chromosome 1 (insertion and deletion into the promoter of yrdC gene) were validated (Table 2). To increase the number of markers, the striking insertion of 1.5 kb into the promoter of the clock gene TOC1 on chromosome 17 was included even though its diversity was below three (Table 2). The fifth marker was selected as marker of the Big outlier Chromosome (BOC) on chromosome 14 (Moreau et al. 2012).

Typical PCR results for the 5 validated diversity markers are presented in Figure 1. The detailed description of each marker is provided as supplementary data 1.

Bathy01g04300 encodes a yrdC domain-containing protein of unknown function. In *Escherichia coli*, yrdC binds preferentially to double-stranded RNA, consistent with a role of the protein in translation (Teplova et al. 2000). A diverse organisation was identified in the

promoter region of Bathy01g04300 in comparison to the reference genome RCC1105 and was visualised by amplification with primers MDB33 an MDB34 (Figure 1, Table 2, Supplemental data 1A). This set of primers showed marked differences in its amplification success among the accessions indicating important nucleotide variations.

The intergenic region upstream Bathy17g01510 encoding an homolog of TOC1 involved in the control of circadian rhythm was not similar in all accessions. A 2.2 kb insertion was identified in RCC1613, RCC1615 and RCC4752 after amplification, whilst the other accessions were similar to RCC4222 (Figure 1, Table 2, Supplemental data 1B).

Bathy03g02080 encodes a protein containing a flavodoxin-like domain, a flavin mononucleotide (FMN)-binding site and 6 imperfect repeats of 25 amino acids. In comparison to RCC1105, ONT sequencing revealed insertions of 147 bp in RCC4222, 74 bp in RCC1613 and 375 bp RCC685, while RCC1868 and RCC4752 were unchanged. These predictions verified by PCR and sequencing confirm that substantial INDEL can also occur within coding regions (Figure 1, Table 2, Supplemental data 1C).

Bathy15g02320 encodes a protein with Zinc Finger repeats (ZF) of a greater length in Arctic (9ZF) and RCC685 (7ZF) than in RCC4222 (6ZF). Primers were designed to amplify the array of ZF motifs (Figure 1, Table 2, Supplemental data 1D). Since MDB68 and MDB69 did not amplify a single fragment in RCC5417 and RCC1868, the corresponding region was obtained by the ONT data (Table2). Zinc finger C2H2 protein are numerous (53 genes) and highly conserved in the Bathycoccus reference genome (Moreau et al. 2012).

Bathy14g30100 encodes the protein containing a TIM domain found in the protein Timeless involved in circadian rhythm control in Drosophila (Sehgal et al. 1995). This gene is located in the BOC region of chromosome 14, a part of the chromosome potentially involved in mating type (Moreau et al. 2012). The ORF of TIM was found conserved among the tested

strains (Figure 1, Table 2). This fifth marker was non discriminating among the subset of world-wide Bathycoccus strains.

Accessory genes in Bathycoccus prasinos genome

The large insertions found in the promoters of yrdC and TOC1 were sequenced and analysed in detail (supplementary data 1A-B).

In yrdC promoter, a gene encoding a protein of unknown function possessing ANK repeats similar to the gene products of Bathy01g04610 (68% amino acid identity), Bathy01g04570 (67%) and Bathy11g02720 (57%) was inserted between an Evening Element-like (EEL) *cis* element and the start of Bathy01g04300 in RCC5417, 1613, 685 and 4752. In order to gain information on this additional gene, the Ocean Gene Atlas (OGA) website (http://tara-oceans.mio.osupytheas.fr/ocean-gene-atlas/) (Villar et al., 2018, Vernette et al. 2022) was interrogated with the additional protein and its three homologs. The only sequence retrieved from OGA shared significant similarity with the 4 proteins but was not identical. So the geographical distribution that we obtained is not the one corresponding to the accessory ANK gene.

In TOC1 promoter, the 2.2 kb insertion identified in RCC1613, RCC1615 and RCC4752 encodes a Methyltransferase-like protein (AdoMTase *METTL24* IPR026913 IPR029063) of 320 aa with 41.7% amino acids identity to the predicted *Ostreococcus lucimarinus* CCE9901 protein (XP_001422352). Searches were performed at high stringency in OGA (Expect threshold 1e-300) so that only the presence of the near-identical sequences was retrieved. TOC1 and AdoMTase sequences were not strictly co-occuring. At one station located near Chile (arrow Figure 2), TOC1 sequences were abundant but no AdoMTase was recorded which suggests that in this particular area the Bathycoccus genomes are almost devoid of AdoMTase. Surprisingly, it was possible to also find AdoMTase sequences not associated to

Bathycoccus TOC1 sequences indicating that this accessory gene can be found in other unknown microorganisms suggesting possible horizontal gene transfer. Most AdoMTase hits were found near the equator at temperature higher than 25°C. As a control, the co-occurrence of TOC1 and CCA1 (Bathy06g4380) sequences was complete (Supplemental Figure 1). Transcription of the gene encoding this AdoMTase was confirmed in the metaT database at OGA (data not shown).

In conclusion, two large INDEL comprised additional genes and these should be considered as structural rearrangements rather then simple INDEL. These accessory genes could potentially be beneficial for adaptation.

Isolation of Bathycoccus during the 2018/2019 winter bloom

Surface water was collected weekly from December 3rd to March 19th. During this period, the sea temperature rose to 15.87°C in December and did not descent below 10.68°C in February (Figure 3A). According to the 10-year study at the same location (SOLA Buoy in Banyuls bay) for 2007-2015 (Lambert et al. 2019) and 2015-2017, this period can be considered as an average climatic year in term of temperature (Lambert et al. 2021). The presence and abundance of phytoplankton were determined by flow cytometry after filtration on 3 µm (Figure 3). Whilst cyanobacteria were the most abundant at all time with a peak at the end of February, picophytoplankton was the second most abundant category with a first peak in December and a second in February (Supplemental Figure 2). At each sampling date, collected seawater was also filtered through 1.2 µm pore-size and transferred to culture flasks. After a period of acclimation of two weeks in supplemented sea water, cells were isolated by plating on agarose. Light green-yellow coloured colonies were picked and sub-cultured. In order to accelerate the identification of Bathycoccus among the isolated cells, amplifications of a fragment of the *LOV-HK* (Bathy10g02360) gene were performed. These primers were specific to the *Bathycoccus prasinos* genome and did not amplify the homologous gene in

Ostreococcus or Micromonas nor in any other species. The identity of these clones was further confirmed by ribotyping (amplification of a 2kb ribosomal DNA fragment followed by sequencing). In total, 55 Bathycoccus B1 genome isolates were recovered at nine sampling dates (Supplemental Table2). Additional isolates producing ambiguous amplification of *LOV-HK* were also ribotyped. Eleven were actually Bathyoccus, though not B1 type, and two Nanochloropsis and four Cylindrotheca (Supplemental Table2).

Identification of dominant Bathycoccus Multi Loci Genotypes in Banyuls bay in 2018/2019

The five diversity markers were used in a combination of PCR and sequencing in order to distinguish the different isolates of Bathycoccus sampled during the 2018/2019 winter bloom. For the *yrdC* promoter, the inserted gene encoding a protein with ANK repeats was not detected in the Banyuls samples. In one isolate only (B1 February 25th), the yrdC promoter was near identical to the one in RCC4222 (99% identity in 332 bp, Supplemental data 1A). In 54 other isolates, a fragment of 200 bp similar in size to that in RCC1868 was amplified and sequenced (100% identity, Supplemental data 1A). The insertion present in the *TOC1* promoter was more prevalent in the Banyuls isolates (91%) than in the world-wide strains (43%) (Table 2 and 3). A high degree of similarity (98.6%) was observed over the entire insertion whilst the core promoter of *TOC1* containing the essential Evening Element-Like (EEL) *cis* element had a minimum of 71.6% nucleotide identity in 400 bp. A phylogenetic tree was constructed with these sequences (Supplemental data 1B). The 2 kb inserted sequence was not included in the study but its occurrence is indicated between brackets for each isolate. Most of the Mediterranean isolates were found in the main clade which divided further into two subclades Ia and Ib. Subclade Ia contained all the Mediterranean reference type and RCCC685, while subclade Ib most of the Mediterranean 2kb type and RCC1615. RCC1613 and RCC4752 from Naples lie in a separate clade II. RCC1868 and the Arctic RCC5417 contained the most divergent sequences in clade III. As a consequence of the 2kb

343 insertion, the *TOC1* promoter activity is most likely not abolished since the crucial EEL *cis*
344 element is still present but the analysis of the core promoter sequences indicates that its
345 activity could differ between the clades. Thus the *TOC1* promoter could constitute a
346 functional marker as well as a diversity marker.
347 The gene encoding a flavodoxin-like protein was more diverse in size than were the intergenic
348 regions (Supplemental data 1C,Table 2, 3). The maximum size difference increases to 245
349 amino acids between B1, a February Banyuls isolate and the Arctic RCC5417 strain. The
350 function related to this sequence variation is unknown. It forms a coiled-coil structure of
351 several alpha-helices. Coiled-coil domains have been identified in enzymes where they
352 function as molecular spacers positioning catalytic activities. So the variable length of the
353 repeats could influence the activity of the flavodoxin-like protein.
354 The genome of most Banyuls isolates encoded a Zinc finger C2H2 protein with 6 ZF motifs
355 (Bathy15g02320, Supplemental data 1D, Table 3). Only three isolates from February
356 sampling were similar to RCC685 with an additional ZF motif.
357 Amplification of a fragment of 530 bp of TIM (TIMa) was observed in 65% (approximatly
358 two thirds) of the isolates (Table 3). Based on ONT data for the A8 isolate, we designed an
359 additional primer in order to amplify a variant of TIM present in the remaining third of the
360 isolates, TIMb. The alignment of the two predicted variant proteins TIM4222 and TIMA8
361 showed that one third of the protein is well conserved while two thirds were more variable
362 (Supplemental data 1E). A phylogenetic tree further supported this dichotomy (Figure 4).
363 OGA MetaG database was interrogated with TIM4222 and TIMA8 and each retrieved a
364 single hit, respectively OGATIM4222 and OGATIMA8 confirming the existence of the 2
365 isoforms of TIM world-wide (Figure 4B). OGATIMA8 has a marked abundance for high
366 latitudes and cold temperatures in the Northern hemisphere while OGATIM4222 is more
367 widely distributed.

Based on the results described above, the Banyuls isolates were classified in eight MLG (Table 3). MLG 1 and MLG 2 represent respectively 53% and 29% of the population. Since the number of isolates from each sampling date was not identical (Supplemental Table 2), this percentage may not be entirely representative. However the presence of MLG 1 and 2 in five out of nine independent samplings rules out a bias due to experimental cloning during aisolation, Figure 3). We can thus confidently state that these two MLG were dominant in Banyuls bay during the 2018/2019 bloom. No isolate with a MLG identical to RCC4222 was found.

Determination of major allelic variants in environmental samples : a three year follow-up

The identification of dominant MLG during the bloom 2018/2019 raises the question of their yearly or occasional prevalence in Banyuls bay. Since isolating strains is highly time consuming, an alternative approach was developed to estimate local diversity. Five litres of seawater were sampled once a week and filtered between 3 and 0.8 µm. DNA extracted from 0.8 µm filters was used as template for PCR analysis using our set of diversity markers. Samplings were performed during 3 successive blooms from 2018 to 2021 (Figure 3, Supplemental Figure 2).

Variations in the yrdC and TOC1 promoters and TIM ORF sequences were analysed on these environmental DNA samples. For markers of chromosome 3 and 15, despite the high specificity of the primers on DNA of individual Bathycoccus (Table3), they could not be used on complex environmental DNA samples due to the presence of high background for flavodoxin sequences and to the marginal size difference for Zinc finger C2H2 variants.

The time of sampling is indicated as week of the month (e.g. Oct-01 = 1st week of October) to facilitate the comparison between years. The complete dataset is presented in Supplemental Figure 3 and Supplemental Table 3.

The initiation of the bloom was analysed during two consecutive years (Figure 5). In 2019, the presence of Bathycoccus was detected in the third week of November with a clear predominance of the 200 bp allele of yrdC promoter and the presence of both TIMa and TIMb. In 2020, Bathycoccus was detected in October, was barely detected or absent in November and reappeared in December, consistent with the recording the decrease of abundance of picophytoplankton by flow cytometry (Supplemental Figure 2). In 2020, the allelic ratios of 200/400 bp of the yrdC promoter were clearly different from those in 2019 and TIMb was not detected. We conclude that the populations at the onset of the bloom were different in both timing and diversity.

The diversity of populations was also assessed during the bloom (Figure 5Figure 3). The abundance of alleles of yrdC promoter and TIM were clearly different in November-December compared to February.

In summary, the study of three successive blooms showed changes in the diversity of Bathycoccus populations during a bloom, particularly at its onset and between years.

Physiological characteristic of isolated Bathycoccus strains

Since alleles abundance showed that November-December and February populations are different and since temperature and light intensity are significantly different in December (15.9°C, 9h15 light, maximum intensity 540 µE/m^2/s) and February (11.8°C, 10h30 light, maximum intensity 830 µE/m^2/s), the physiological parameters of the “December “ and “February” isolates were determined. C2, an isolate from January 28th was included in the experiment. Cells were grown at 13°c or 16°C under December or February illumination. At low temperature, all the “February” isolates grew better than the “December” cells independently of the light regime (Figure 6). At 16°C, the fitness of “December” isolates B9 and G11 was significantly improved and reached the growth rate of the “February” cells.

February illumination conditions further reduced the fitness difference between "December" and "February" isolates. Overall, the "February" isolates were more performing than the December" isolates particularly at low temperature. Photosynthesis parameters were determined by PhytoPAM. For most isolates, no significant differences in indicators of photosynthesis parameters were observed with the exception of the B1 February isolate with the unique MLG 7 (Supplemental Figure 4).

Together, the results showed a clear difference in the growth rate of the Bathycoccus isolated in December and February mainly due their capacity of adaptation to low temperature.

Presence of additional Bathycoccus species in Banyuls

Eleven isolates from December were characterised as Bathycoccus by their ribosomal 18S sequences. However since specific primers for the LOV-HK gene and the TOC1 promoter did not amplify a fragment using their DNA as template they were not B1 type. They could be of B2 type. The presence of B2 Bathycoccus in Banyuls bay has been suggested by the TARA metagenomic data analysis (Vannier et al. 2016). A closer examination of ITS2 sequences showed that these 11 Bathycoccus were neither B1 nor B2 but a new type that we named B3 (Figure 7A). We included in the analysis a potential new Bathycoccus type identified among uncultured seawater samples form Russian seas (Belevich et al. 2021) that we named B4 and we showed that B3 is different from the 3 others.

In addition, primers were designed for the TOC1 promoter and the intein inserted into the PRP8 gene in RCC716 B2 type genome and used them on Banyuls seawater. These 2 gene portions (validated by sequencing, Supplemental data 1F and 1G) confirmed the presence of Bathycoccus B2 type although none were isolated. Specific primers for B3 (TOC1 ORF) were also used on environmental samples and taking together these results show that Bathycoccus B1, B2 and B3 type were present from January to March 2019 in Banyuls bay (Figure 7B). They relative abundance compared to the control DNA suggests that B2 and B3 were less

abundant than B1 in Banyuls. To provide additional evidence to classify Bathycoccus B3 type as a new species, the cell ultrastructure were determined by electronic microscopy compared to the one of B1 type. B3 cells showed the characteristic features of Bathycoccus with a single chloroplast containing a single starch granule and scales at the surface. We named this new species *Bathycoccus catiminus* (Figure 7C).

Discussion (2340 words)

Contribution of ONT sequencing to the identification of Bathycoccus molecular markers

Bathycoccus has a small genome of approximately 15 Mb distributed among 19 chromosomes and has only been found in an haploid phase (Moreau et al. 2012). This organisation makes it suitable for Oxford Nanopore Technology Rapid barcoding libraries and sequencing. On a single flow cell, it was possible to obtain sufficient coverage for the genome of up to 3 strains. The main difficulty was to obtain good quality genomic DNA for each strain, a criteria particularly important when pooling barcoded libraries. With the exception of RCC1615, all ONT data were superior to 10 times the size of the Bathycoccus genome (Table 1). When comparing ONT data from RCC4222 to the clonal RCC1105 reference, 19 INDEL were found. Two main reasons can be proposed. Firstly, RCC14222 is not strictly identical to RCC1105. Secondly, ONT is particularly suitable to identify structural variations previously undetected by conventional sequencing method (Michael et al. 2018, Mantere et al. 2019). Therefore, most of the INDEL identified between RCC1105 and RCC4222 genomes could result from the use of different sequencing techniques. To design diversity markers, INDEL in intergenic regions (promoters) as well as ORF encoding repeat amino acids sequences were selected. Most predictions were accurate and validated by PCR and sequencing, except in two cases, where we could not design specific primers and the marker status was unresolved (data not shown).

Identification of Bathycoccus local diversity using INDEL markers

The 55 Mediterranean B1 Bathycoccus isolates were genotyped using five INDEL markers that were based on ONT sequencing of strains coming from contrasted geographic locations between arctic and temperate regions. The percentage of the presence of a marker type was quite different between the world-wide strains and the Banyuls strains (Table 2 and 3). For example, the insertion of a gene encoding ANK repeats was found in four out of seven world strains but was absent from Banyuls strains, while the insertion of the AdoMTase into the TOC1 promoter was much more prevalent in Banyuls strains. In general, the genome of Banyuls Bathycoccus isolates possesses allelic variants common to most of them (65-98%, Table 3) representing the dominant MLG 1 and 2. Remarkably, these three dominant INDEL/rearrangements (200 bp yrdC, 2.2 Kb TOC1, 1.2 Kb Flavodoxin) were not found in the reference genome RCC4222 that was isolated in the Banyuls bay in 2006. In addition, no 2018-2019 isolates share the same five marker types with either the Banyuls reference nor the RCC4752 Napoli genome that was isolated in 1986 (Table 3). Thus it is clear that RCC4222 isolated in Banyuls bay in 2006 was certainly not abundant and probably not present during the bloom 2018/2019. Overall we did not observe any obvious consistent pattern of occurrence between specific markers and the geographic origin of the world-wide strains (Table 2 and 3).

The determination of eight MLG in Bathycoccus is probably an underestimation. For example in diatoms, where more than 600 individuals have been genotyped using microsatellites, it was estimated that the blooming population was comprised of at least 2400 different genotypes (Rynearson and Armbrust 2005). However the dominance of two MLG among Bathycoccus isolates is probably accurate, despite being based on a small number.

The existence of major MLG highlights an apparent paradox : how can blooms be diverse, given that the best genotype should prevail? Blooms are predicted to quickly become dominated by a few particularly well-adapted genotypes (De Meester, 1996). Nevertheless,

most studies describing genetic diversity of blooming phytoplankton populations report high intraspecific variation (Rynearson and Armbrust, 2005; Alpermann et al.,2009 ; Lebret et al., 2012 ; Dia et al., 2014). Indeed, our results revealed such diversity by the detection of six minor MLG beside the two major ones. Remarkably, the different fitness observed between December and February 2018/2019 isolates correlates with differences in allelic frequencies at onset of a bloom and during the course of a bloom as determined on seawater samples, suggesting that best seasonal MLG may become dominant at specific times of the year. Similarly, temporal succession of two genetically distinct sub-populations was observed during the bloom of the haploid *Alexandrium dinoflagellate* in Gulf of Maine (Erdner et al., 2011).

Finally, in addition to eight *Bathycoccus prasinos* MLG (B1), we identified two other Bathycoccus species, including the previously described species *Bathycoccus calidus* (B2) and a yet not described Bathycoccus B3 species that we name *B. catiminus*. Our results clearly demonstrate that under a single Bathycoccus 18S ribotype are hidden at least three distinct species. This emphasises the necessity of isolating new Bathycoccus to design interspecific markers that could be used on environmental samples and metagenomic datasets in order to understand the dynamics of Bathycoccus blooms and the adaptation of this cosmopolitan genus in the world ocean.

Structural variants as intraspecies markers to follow intra-specific diversity in environmental samples

Historically, the intraspecies markers corresponded to small nucleotide repeats (microsatellites) or chloroplastic or mitochondrial genes and a few nuclear genes that were applied to several hundreds of isolated individuals. The development of Restriction-site-Associated DNA sequencing (RADseq) techniques has allowed the discovery and genotyping

of thousands of genetic markers for any given species at relatively low-cost (Andrews et al. 2016). Some of these approaches have been used to analyse the diversity of populations during algal blooms (Regenfors et al. 2017). The recent dramatic increase of the number of sequenced genomes led to large-scale diversity studies with large sets of nuclear genes or whole genome comparisons. Most of these approaches required the isolation of a large number of individuals. As a consequence, intraspecies diversity has been poorly documented in the past in marine phytoplankton. Novel, rapid and cheap sequencing technologies have given access to Mamiellales diversity by the sequencing of the genomes of Ostreococcus isolates (Blanc-Mathieu et al. 2017) or by metagenomic approaches (Leconte et al. 2020, Da Silva et al. 2022; Richter et al. 2022) or metatranscriptomic approaches (Simmons et al. 2016).

We aimed to develop a rapid and cost effective alternative which did not rely on isolated individuals since it is very challenging and time consuming to isolate marine microalgae from complex microbial communities. Compared to short read sequencings, the recent ONT and PACBIO sequencing technology provided information on structural variants, in particular on relatively large INDEL and on repeated/low complexity sequences, with some of which previously overlooked (Wellenreuther et al. 2019). This knowledge was particularly useful to develop a novel type of marker for assessing intraspecies diversity that will rely only on PCR amplification of variable size fragment without the need of sequencing. Furthermore, with INDEL based on ONT we are reaching a higher level of population structure compared to microsatellite or SNP markers. Structural variants are expected to be less neutral and more stable than microsatellites (Mérot et al. 2020). Microsatellites can change in clonal strains of *P. multistriata* in the laboratory over several months (Ruggiero et al. 2018), while our structural markers are still identical in the reference genome published in 2012 and its clonal strain RCC4222 sequenced in 2018. Thus INDEL markers can identify large subpopulations

543 rather than small groups of individuals. Our studies on three successive blooms in Banyuls
544 showed that the ONT designed markers are capable of determining the dominant allelic
545 variants and their perennial occurrence. Similarly these INDEL markers could be used as
546 query on metagenomic databases for a wider and large analysis of Bathycoccus populations.
547
548 Variations in allele frequencies were observed for three consecutive years, raising the
549 question of the nature of the highly reproducible yearly re-occurrence of Bathycoccus in the
550 bay of Banyuls. Seasonal blooms may result either from re-activation of "dormant/survivor"
551 cells from the water column (whose genetic fingerprint will determine the genetic profile of
552 the next bloom) or by yearly *de novo* fertilisation by cells carried by the north Mediterranean
553 current along the gulf of Lion. At the first glance, our preliminary results are in favour of the
554 introduction of a new population rather than "resuscitation" of cells of the previous bloom
555 since allelic frequencies are distinct between the end of a bloom o the onset of the next
556 (Figure 5). By monitored the temporal population structures of the dinoflagellate *Alexandrium*
557 *minutum* in two estuaries in France, Dia et al. (2014) showed that interannual genetic
558 differentiation was greater than intra-bloom differentiation. Alternation of
559 genotypes/populations has also been observed with diatoms in the dominance of one of the
560 two sympatric populations of *Pseudonitzschia multistriata* which could be due either to
561 environmental factors favouring one population over the other or intrinsic factors coupled to
562 the obligate sexual life cycle of *P. multistriata* (D'Alelio et al. 2010). Thus the observed
563 differences in alleles frequencies could equally be the result of new inoculatum from currents
564 or of sexual reproduction. Even though sexual reproduction has not been demonstrated in
565 Bathycoccus, there is genomic evidence that it may occur (Benites et al. 2021). Sexual
566 recombination generates new combinations of alleles, whereas clonality favours the spread of
567 the fittest genotype through the entire population (Dia et al. 2014). Erdner et al. (2011)

propose for *A. fundyense* that mitosis is the primary mode of multiplication during blooms whereas mating is triggered presumably in response to unfavourable conditions at the end of blooms, with vegetative cells not overwintering in the water column. In Banyuls bay, the abundance during the bloom is followed by severe bottlenecks in which Bathycoccus are hardly detected in the water column (Lambert et al. 2019).

Knowing that (1) Bathycoccus blooms are followed by severe bottlenecks between one bloom and the next, (2) allelic frequencies were not similar at the end of one bloom and at the onset of the next and (3) structural markers are very stable in mitotic dividing cells, the hypothesis of rare vegetative cells remaining in the water column between the blooms is unlikely except if those remaining cells were produced by sexual reproduction. The other hypothesis of new strains brought by current is equally probable.

Structural variants versus functional variants

Our principal interest was to identity markers of intraspecies diversity in order to follow the dynamics of Bathycoccus population during annual blooms in the bay of Banyuls. However structural variants are probably not neutral markers unlike microsatellites. Most of the selected markers could also represent functionally significant variants such as an additional gene or a modified promoter activity or protein function, that could correspond to an adaptation to contrasted intra- and interranual variations in environmental parameters in the Banyuls Bay (Lambert et al., 2021).

In only five worldwide INDEL, we discovered two additional genes in the genome of *Bathycoccus prasinos* and a particular protein structure. The additional ANK repeat encoding gene in chromosome 1 has probably arisen by gene duplication or gene loss since it belongs to a multigenic family. The origin of the AdoMTase could be the result of Horizontal Gene

Transfer. The Flavodoxin-like protein has an organisation specific to Bathycoccus with a coiled-coil domain of variable size with similarity to Eukaryotes parasites and toxic bacteria proteins and a flavodoxin domain found in the 7 other Bathycoccus flavodoxin-like proteins. Bathycoccus culture strains do not possess a flavodoxin per se while it was found in uncultured Bathycoccus (Pierella Karlusich et al. 2015). This peculiar flavodoxin-like protein could represent a case of neofunctionalisation in Bathycoccus. The core promoter of the central circadian clock *TOC1* gene has a conserved evening element like box (EEL box) that has been experimentally demonstrated as essential in the central oscillator of *Ostreococcus tauri* (Corellou et al. 2009). Although the EEL box is found in all the accessions sequenced, the distance between the *cis* element and the initiation codon is variable. In addition, the phylogenetic tree of the core promoter sequences clearly discriminated the Arctic RCC5417 and RCC1868 and to a lesser extent, RCC1613 and the second Mediterranean strain, RCC4752 (Suppl. data 1B). An insertion of the AdoMTase was found about 100 bp upstream the EEL box. This insertion could potentially modify the promoter activity and ultimately the expression pattern of TOC1. Such a natural variation of promoter length modulates the photoperiodic response of FLOWERING LOCUS T by differentially spacing two interdependent regulatory regions (Liu et al. 2014). Although the presence of the AdoMTase was not correlated with the latitude or the temperature in the Ocean Gene Atlas (Figure 2), it could still be associated with a seasonal niche. Less information is available for the promoter and function of the yrdC gene. The rearrangments are more drastic, specially with the displacement of the EEL box by insertion or deletion, and could lead to the inactivation of the promoter. The most striking feature concerns the TIM protein where only one third of the protein is conserved between TIMa and TIMb. Due to its position in the BOC of chromosome 14 putatively involved in mating, this raises the question of the mating types of cells with genome containing TIMa or a genome with TIMb.

Altogether, this analysis suggests that our allelic markers may be more than diversity markers, being potentially involved in adaptation to changing environmental conditions.

Conclusions and perspectives

In this paper we analysed the populations of Bathycoccus blooming in the Banyuls bay. Based on structural variants, the isolates were categorised into eight MLG with two dominant classes. The survey of Bathycoccus diversity in environmental samples during three successive blooms confirms the yearly presence of dominant allelic variants that were different within and between years. This pioneer study on Bathycoccus diversity in the bay of Banyuls now paves the way to an in depth analysis of multiple markers present in more than a decade of bimonthly sampled metagenomic data at a discrete location (Lambert et al. 2019; Lambert et al. 2021). The sequencing of the whole genomes of the different MLG, together with the assessment of their physiological performances will bring additional information contributing to the local diversity of Bathycoccus and provide insight their seasonal pattern of abundance. In addition, these diversity markers represent an essential tool for grasping the maximum diversity of newly isolated Bathycoccus world-wide and identify putative molecular mechanisms involved in adaptation to environmental niches of this cosmopolitan genus.

Acknowledgments

We are grateful to the captain and the crew of the RV ‘Nereis II’ for their help in acquiring the samples. Additional ONT were performed with the help of Christel Llauro and Marie Mirouze LGDP. The authors acknowledge the ISO 9001 certified IRD itrop HPC (member of the South Green Platform) at IRD Montpellier for providing HPC resources that have contributed to the research results reported within this paper (URL: https://bioinfo.ird.fr/ - http://www.southgreen.fr). The work was financed by a internal LOMIC Microprojet and the ANR Climaclock 2020-2024. We are thankful to Marie-Line Escande and the platform

643 PioPIC at the Banyuls Oceanographic Observatory for electron microscopy service. We thank
644 Thomas Roscoe for critical reading the manuscript.

645

646 **References (1875 words)**

647 Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, Lippman ZB,
648 Schatz MC. (2019) RaGOO: fast and accurate reference-guided scaffolding of draft genomes.
649 *Genome Biol*. 20(1):224. doi: 10.1186/s13059-019-1829-6.

650 Alpermann TJ, Beszteri B, John U, Tillmann U, Cembella AD. (2009) Implications of life-
651 history transitions on the population genetic structure of the toxigenic marine dinoflagellate
652 Alexandrium tamarense. *Molecular Ecology*, 18(10):2122-33. doi: 10.1111/j.1365-
653 294X.2009.04165.x.

654 Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. (2016) Harnessing the power
655 of RADseq for ecological and evolutionary genomics. *Nature Reviews Genet*ics, 17(2):81-92.
656 doi: 10.1038/nrg.2015.28.

657 Bachy C, Yung CCM, Needham DM, Gazitúa MC, Roux S, Limardo AJ, Choi CJ, Jorgens
658 DM, Sullivan MB, Worden AZ. (2021) Viruses infecting a warm water picoeukaryote shed
659 light on spatial co-occurrence dynamics of marine viruses and their hosts. *The ISME Journal*,
660 doi: 10.1038/s41396-021-00989-9.

661 Belevich TA, Milyutina IA, Abyzova GA, Troitsky AV. (2021) The pico-sized
662 Mamiellophyceae and a novel Bathycoccus clade from the summer plankton of Russian
663 Arctic Seas and adjacent waters. *FEMS Microbiology Ecology*, 97(2):fiaa251. doi:
664 10.1093/femsec/fiaa251.

665 Benites LF, Bucchini F, Sanchez-Brosseau S, Grimsley N, Vandepoele K, Piganeau G. (2021)
666 Evolutionary Genomics of Sex-Related Chromosomes at the Base of the Green Lineage.
667 *Genome Biology Evolution*, 13(10):evab216. doi: 10.1093/gbe/evab216.

668 Blanc-Mathieu R, Krasovec M, Hebrard M, Yau S, Desgranges E, Martin J, Schackwitz W,
669 Kuo A, Salin G, Donnadieu C, Desdevises Y, Sanchez-Ferandin S, Moreau H, Rivals E,
670 Grigoriev IV, Grimsley N, Eyre-Walker A, Piganeau G. (2017) Population genomics of
671 picophytoplankton unveils novel chromosome hypervariability. *Science Advances*,
672 3(7):e1700239. doi: 10.1126/sciadv.1700239.

673 Chrétionot-Dinet MJ, Courties C, Vaquer A, Neveux J, laustre H, autrier J and Machado MC
674 1995. A new marine picoeucaryote:Ostreococcus tauri gen. et sp. nov. (Chlorophyta,
675 Prasinophyceae). *Phycologia*, 34(4): 285–292.

676 Corellou F, Schwartz C, Motta JP, Djouani-Tahri el B, Sanchez F, Bouget FY. (2009) Clocks
677 in the green lineage: comparative functional analysis of the circadian architecture of the
678 picoeukaryote ostreococcus. *The Plant Cell*, 21(11):3436-49. doi: 10.1105/tpc.109.068825.

679 D'Alelio D, Ribera d'Alcala M, Dubroca L, Sarno D, Zingone A, Montresor M (2010) The
680 time for sex: A biennial life cycle in a marine planktonic diatom *Limnol.ogy Oceanography*,
681 55(1) 106–114.

682 Da Silva O, Ayata SD, Ser-Giacomi E, Leconte J, Pelletier E, Fauvelot C, Madoui MA, Guidi
683 L, Lombard F, Bittner L (2022) Genomic differentiation of three pico-phytoplankton species
684 in the Mediterranean Sea. *Environmental Microbiology*, Aug 16. doi: 10.1111/1462-
685 2920.16171.

686 Debladis E, Llauro C, Carpentier MC, Mirouze M, Panaud O. (2017) Detection of active
687 transposable elements in Arabidopsis thaliana using Oxford Nanopore Sequencing
688 technology. *BMC Genomics*, 18(1):537. doi: 10.1186/s12864-017-3753-z.

689 De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. (2018) NanoPack:
690 visualizing and processing long-read sequencing data. *Bioinformatics* 34(15):2666-2669. doi:
691 10.1093/bioinformatics/bty149.

De Meester L. (1996) Evolutionary potential and local genetic differentiation in a phenotypically plastic trait of a cyclical ârthenogen Daphnia Magna. *Evolution*, 50(3):1293-1298. doi: 10.1111/j.1558-5646.1996.tb02369.x.

Des Roches S, Post DM, Turley NE, Bailey JK, Hendry AP, Kinnison MT, Schweitzer JA, Palkovacs EP (2018) The ecological importance of intraspecific variation. *Nature Ecology Evolution*, 2(1):57-64. doi: 10.1038/s41559-017-0402-5.

Dia A, Guillou L, Mauger S, Bigeard E, Marie D, Valero M, Destombe C. (2014) Spatiotemporal changes in the genetic diversity of harmful algal blooms caused by the toxic dinoflagellate Alexandrium minutum. *Molecular Ecology*, 23(3):549-60. doi: 10.1111/mec.12617.

Erdner DL, Richlen M, McCauley LA, Anderson DM. (2011) Diversity and dynamics of a widespread bloom of the toxic dinoflagellate Alexandrium fundyense *PLoS One*, 6(7):e22965. doi: 10.1371/journal.pone.0022965.

Gallagher JC (1980) Population genetics of Skeletonema costatum (Bacillariophyceae) in Narragansett bay. *Journal of Phycology*, 16, 464-474 doi.org/10.1111/j.1529-8817.1980.tb03061.x

Galtier N, Nabholz B, Glémin S, Hurst GD. (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology*, 18(22):4541-50. doi: 10.1111/j.1365-294X.2009.04380.x.

Godhe A, Rynearson T. (2017) The role of intraspecific variation in the ecological and evolutionary success of diatoms in changing environments. *Philosophical Transaction of the Royal Society London B Biological Sciences*, 5;372(1728):20160399. doi: 10.1098/rstb.2016.0399.

Guyon JB, Vergé V, Schatt P, Lozano JC, Liennard M, Bouget FY. (2018) Comparative Analysis of Culture Conditions for the Optimization of Carotenoid Production in Several

717 Strains of the Picoeukaryote Ostreococcus. *Marine Drugs*, 16(3):76. doi:
718 10.3390/md16030076.

719 Joli N, Monier A, Logares R, Lovejoy C. (2017) Seasonal patterns in Arctic prasinophytes
720 and inferred ecology of Bathycoccus unveiled in an Arctic winter metagenome. *The ISME*
721 *Journal*, 11(6):1372-1385. doi: 10.1038/ismej.2017.7.

722 Kashtan N, Roggensack SE, Rodrigue S, Thompson JW, Biller SJ, Coe A, Ding H, Marttinen
723 P, Malmstrom RR, Stocker R, Follows MJ, Stepanauskas R, Chisholm SW. (2014) Single-cell
724 genomics reveals hundreds of coexisting subpopulations in wild Prochlorococcus. *Science*,
725 344(6182):416-20. doi: 10.1126/science.1248575.

726 Kolmogorov M, Yuan J, Lin Y, Pevzner PA. (2019) Assembly of long, error-prone reads
727 using repeat graphs. *Nat Biotechnol*. 37(5):540-546. doi: 10.1038/s41587-019-0072-8.

728 Lambert S, Tragin M, Lozano JC, Ghiglione JF, Vaulot D, Bouget FY, Galand PE. (2019)
729 Rhythmicity of coastal marine picoeukaryotes, bacteria and archaea despite irregular
730 environmental perturbations. *The ISME Journal*, 13(2):388-401. doi: 10.1038/s41396-018-
731 0281-z.

732 Lambert S, Lozano JC, Bouget FY, Galand PE. (2021) Seasonal marine microorganisms
733 change neighbours under contrasting environmental conditions. *Environmental Microbiology*,
734 23(5):2592-2604. doi: 10.1111/1462-2920.15482. Epub 2021

735 Lebret K, Kritzberg ES, Figueroa R, Rengefors K. Environ Microbiol. (2012) Genetic
736 diversity within and genetic differentiation between blooms of a microalgal species.
737 *Environmental Microbiology*, 14(9):2395-404. doi: 10.1111/j.1462-2920.2012.02769.x.

738 Leconte J, Benites LF, Vannier T, Wincker P, Piganeau G, Jaillon O. (2020) Genome
739 Resolved Biogeography of Mamiellales. *Genes (Basel)*, 11(1):66. doi:
740 10.3390/genes11010066.

741 Li H. (2021) New strategies to improve minimap2 alignment accuracy. *Bioinformatics*
742 37(23):4572-4. doi: 10.1093/bioinformatics/btab705.

743 Li WK, Rao DV, Harrison WG, Smith JC, Cullen JJ, Irwin B, Platt T. (1983) Autotrophic
744 picoplankton in the tropical ocean. *Science*, **219**: 292-5.

745 Limardo AJ, Sudek S, Choi CJ, Poirier C, Rii YM, Blum M, Roth R, Goodenough U, Church
746 MJ, Worden AZ. (2017) Quantitative biogeography of picoprasinophytes establishes ecotype
747 distributions and significant contributions to marine phytoplankton. *Environmental*
748 *Microbiology*, 19(8):3219-3234. doi: 10.1111/1462-2920.13812.
749 Liu L, Adrian J, Pankin A, Hu J, Dong X, von Korff M, Turck F. (2014) Induced and natural
750 variation of promoter length modulates the photoperiodic response of FLOWERING LOCUS
751 T. *Nature Communications*, 5:4558. doi: 10.1038/ncomms5558.
752 Mantere T, Kersten S, Hoischen A. Long-Read Sequencing Emerging in Medical Genetics.
753 (2019) *Frontiers in Genetics*, 10:426. doi: 10.3389/fgene.2019.00426.
754 Mérot C, Oomen RA, Tigano A, Wellenreuther M. (2020) A Roadmap for Understanding the
755 Evolutionary Significance of Structural Genomic Variation. Trends Ecol Evol. 35(7):561-
756 572. doi: 10.1016/j.tree.2020.03.002.
757 Michael TP, Jupe F, Bemm F, Motley ST, Sandoval JP, Lanz C, Loudet O, Weigel D, Ecker
758 JR. (2018) High contiguity Arabidopsis thaliana genome assembly with a single nanopore
759 flow cell. *Nature Communications*, 9(1):541. doi: 10.1038/s41467-018-03016-2.

760 Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. (2018) Versatile genome
761 assembly evaluation with QUAST-LG. *Bioinformatics* 34(13):i142-i150. doi:
762 10.1093/bioinformatics/bty266.

763 Monier A, Sudek S, Fast NM, Worden AZ. (2013) Gene invasion in distant eukaryotic
764 lineages: discovery of mutually exclusive genetic elements reveals marine biodiversity. *The*
765 *ISME Journal*, 7(9):1764-74. doi: 10.1038/ismej.2013.70.
766 Moreau H, Verhelst B, Couloux A, Derelle E, Rombauts S, Grimsley N, Van Bel M, Poulain
767 J, Katinka M, Hohmann-Marriott MF, Piganeau G, Rouzé P, Da Silva C, Wincker P, Van de
768 Peer Y, Vandepoele K. (2012) Gene functionalities and genome structure in Bathycoccus
769 prasinos reflect cellular specializations at the base of the green lineage. *Genome Biology*,
770 13(8):R74. doi: 10.1186/gb-2012-13-8-r74.
771 Pierella Karlusich JJ, Ceccoli RD, Graña M, Romero H, Carrillo N. (2015) Environmental
772 selection pressures related to iron utilization are involved in the loss of the flavodoxin gene
773 from the plant genome. *Genome Biology and Evolution*, 7(3):750-67. doi:
774 10.1093/gbe/evv031.
775 Raffard A, Santoul F, Cucherousset J, Blanchet S. (2019) The community and ecosystem
776 consequences of intraspecific diversity: a meta-analysis. *Biological Reviews of the Cambridge*
777 *Philosophical Society*, 94(2):648-661. doi: 10.1111/brv.12472.
778 Read BA, Kegel J, Klute MJ, Kuo A, Lefebvre SC, Maumus F, Mayer C, Miller J, Monier A,
779 Salamov A, Young J, Aguilar M, Claverie JM, Frickenhaus S, Gonzalez K, Herman EK, Lin
780 YC, Napier J, Ogata H, Sarno AF, Shmutz J, Schroeder D, de Vargas C, Verret F, von
781 Dassow P, Valentin K, Van de Peer Y, Wheeler G; Emiliania huxleyi Annotation Consortium,
782 Dacks JB, Delwiche CF, Dyhrman ST, Glöckner G, John U, Richards T, Worden AZ, Zhang
783 X, Grigoriev IV. Pan genome of the phytoplankton Emiliania underpins its global
784 distribution. (2013) *Nature*, 499(7457):209-13. doi: 10.1038/nature12221.
785 Rengefors, K., Kremp, A., Reusch, T.B.H. and Wood A. M. (2017) Genetic diversity and
786 evolution ineukaryotic phytoplankton: revelations from population genetic studies. *Journal of*
787 *Plankton Research*, (2017) 39(2): 165– 179.

788 Richter DJ, Watteaux R, Vannier T, Leconte J, Frémont P, Reygondeau G, Maillet N, Henry
789 N, Benoit G, Da Silva O, Delmont TO, Fernàndez-Guerra A, Suweis S, Narci R, Berney C,
790 Eveillard D, Gavory F, Guidi L, Labadie K, Mahieu E, Poulain J, Romac S, Roux S, Dimier
791 C, Kandels S, Picheral M, Searson S; Tara Oceans Coordinators, Pesant S, Aury JM, Brum
792 JR, Lemaitre C, Pelletier E, Bork P, Sunagawa S, Lombard F, Karp-Boss L, Bowler C,
793 Sullivan MB, Karsenti E, Mariadassou M, Probert I, Peterlongo P, Wincker P, de Vargas C,
794 Ribera d'Alcalà M, Iudicone D, Jaillon O. (2022) Genomic evidence for global ocean
795 plankton biogeography shaped by large-scale current systems. *Elife*, 11:e78129. doi:
796 10.7554/eLife.78129.
797 Ruggiero MV, D'Alelio D, Ferrante MI, Santoro M, Vitale L, Procaccini G, Montresor M.
798 (2018) Clonal expansion behind a marine diatom bloom. *The ISME Journal*, 2018
799 Feb;12(2):463-472. doi: 10.1038/ismej.2017.181.
800 Rynearson TA, Armbrust EV. (2005) Maintenance of clonal diversity during a spring bloom
801 of the centric diatom Ditylum brightwellii. *Molecular Ecology*, 14(6):1631-40. doi:
802 10.1111/j.1365-294X.2005.02526.x.
803 Sehgal A, Rothenfluh-Hilfiker A, Hunter-Ensor M, Chen Y, Myers MP, Young MW (1995).
804 "Rhythmic expression of timeless: a basis for promoting circadian cycles in period gene
805 autoregulation". *Science*, 270 (5237): 808–10. doi:10.1126/science.270.5237.808.
806 Simmons MP, Sudek S, Monier A, Limardo AJ, Jimenez V, Perle CR, Elrod VA, Pennington
807 JT, Worden AZ. (2016) Abundance and Biogeography of Picoprasinophyte Ecotypes and
808 Other Phytoplankton in the Eastern North Pacific Ocean. *Applied Environmental*
809 *Microbiology*, 82(6):1693-1705. doi: 10.1128/AEM.02730-15.
810 Srivastava S, Avvaru AK, Sowpati DT, Mishra RK. (2019) Patterns of microsatellite
811 distribution across eukaryotic genomes. *BMC Genomics*, 20(1):153. doi: 10.1186/s12864-
812 019-5516-5.

Teplova M, Tereshko V, Sanishvili R, Joachimiak A, Bushueva T, Anderson WF, Egli M. (2000) The structure of the yrdC gene product from Escherichia coli reveals a new fold and suggests a role in RNA binding. *Protein Science*, 9(12):2557-66. doi: 10.1110/ps.9.12.2557.

Tesson SV, Borra M, Kooistra WH, Procaccini G. (2011) Microsatellite primers in the planktonic diatom Pseudo-nitzschia multistriata (Bacillariophyceae). *American Journal of Botany*, 98(2):e33-5.

Tesson SV, Montresor M, Procaccini G, Kooistra WH. (2014) Temporal changes in population structure of a marine planktonic diatom. *PLoS One*, 9(12):e114984. doi: 10.1371/journal.pone.0114984.

Tragin M, Zingone A, Vaulot D (2018) Comparison of coastal phytoplankton composition estimated from the V4 and V9 regions of the 18S rRNA gene with a focus on photosynthetic groups and especially Chlorophyta. *Environmental Microbiology*, 20(2):506-520.

Van den Wyngaert S, Möst M, Freimann R, Ibelings BW, Spaak P. (2015) Hidden diversity in the freshwater planktonic diatom Asterionella formosa. *Molecular Ecology*, 24(12):2955-72. doi: 10.1111/mec.13218.

Vannier T, Leconte J, Seeleuthner Y, Mondy S, Pelletier E, Aury JM, de Vargas C, Sieracki M, Iudicone D, Vaulot D, Wincker P, Jaillon O. Survey of the green picoalga Bathycoccus genomes in the global ocean. (2016) *Scientific Reports*, 6:37900. doi: 10.1038/srep37900.

Vaser R, Sović I, Nagarajan N, Šikić M. (2017) Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* 27(5):737-746. doi: 10.1101/gr.214270.116.

Vernette C, Lecubin J, Sánchez P; Tara Oceans Coordinators, Sunagawa S, Delmont TO, Acinas SG, Pelletier E, Hingamp P, Lescot M. (2022) The Ocean Gene Atlas v2.0: online exploration of the biogeography and phylogeny of plankton genes. *Nucleic Acids Research*, 50(W1):W516-26. doi: 10.1093/nar/gkac420

837 .Villar E, T. Vannier, C. Vernette, M. Lescot, M. Cuenca, A. Alexandre, P. Bachelerie, T.
838 Rosnet, E. Pelletier, S. Sunagawa, P. Hingamp. (2018) The Ocean Gene Atlas: exploring the
839 biogeography of plankton genes online. *Nucleic Acids Research*, doi: 10.1093/nar/gky376.
840 Violle C, Enquist BJ, McGill BJ, Jiang L, Albert CH, Hulshof C, Jung V, Messier J. (2012)
841 The return of the variance: intraspecific variability in community ecology. *Trends in Ecology*
842 *and Evol.ution*, 27(4):244-52. doi: 10.1016/j.tree.2011.11.014.
843 Wellenreuther M, Mérot C, Berdan E, Bernatchez L (2019) Going beyond SNPs: The role of
844 structural genomic variants in adaptive evolution and species diversification. *Molecular*
845 *Ecology*, 28(6):1203-1209. doi: 10.1111/mec.15066.
846 Wheeler GL, Dorman HE, Buchanan A, Challagundla L, Wallace LE. (2014) A review of the
847 prevalence, utility, and caveats of using chloroplast simple sequence repeats for studies of
848 plant biology. *Applications in Plant Sciences*, 2(12):apps.1400059. doi:
849 10.3732/apps.1400059.

850 **Data accessibility**

851 Strains isolated in Banyuls have been sent to Roscoff Collection Centre and will be publicly
852 available after curation by the Centre. For whole genome sequences, basecalled reads and
853 RaGOO output for the 7 RCC strains used in this study are available at ????

854 **Author contribution**

855 MD, FYB and FS conceived the work and acquired funding. MD extracted high molecular
856 weighy genomic DNA. ONT sequencing was performed by CM and MD and sequence
857 analysis by LD and FS. Diversity markers were designed and validated by MD. PS analysed
858 seawater samplings by flow cytometry. MD and VV isolated Banyuls strains during the
859 winter bloom, genotyped by MD and JCL. MD determined diversity of seawater. MD and
860 FYB wrote the article and all authors participated in critical revisions and approved the final
861 version for submission.

862 **Figure 1**. Diversity markers among world wide accessions

863 **Figure 2**. Geographical distribution of TOC1 and AdoMTase proteins

864 **Figure 3**. Abundance of phytoplankton during three successive blooms

865 **Figure 4**. Distribution and abundance of TIM variant proteins in MetaG database

866 **Figure 5**. Diversity markers in seawater

867 **Figure 6**. Growth of Mediterranean Bathycoccus isolates

868 **Figure 7**. Presence of two additional Bathycoccus species in Banyuls bay

869 **Table 1.** Number and sizes of INDEL

870 **Table 2.** Distribution of diversity markers in world-wide strains

871 **Table 3.** Multi Loci Genotypes of Banyuls isolates

872 **Supplemental Table1.** Strains used in this study

873 **Supplemental Table 2.** Isolation of Bathycoccus strains during 2018/ 2019 winter bloom in

874 the Banyuls bay

875 **Supplemental Table 3.** Relative abundance of diversity markers in sea water

876 **Supplemental data 1.** Sequences and Alignments

877 **Supplemental data 2.** Primers used in this study

878 **Supplemental Figure 1.** Geographical distribution of TOC1 and CCA1 (Bathy05g02420)

879 proteins

880 **Supplemental Figure 2.** Abundance of picophytoplankton during winter blooms

881 **Supplemental Figure 3.** Images of sea water PCR amplification for yrdC, TOC1 and TIM

882 used for Supplemental Table 3

883 **Supplemental Figure 4.** Photosynthesis parameters of Mediterranean Bathycoccus strains

884

885

Table 1. Number and sizes of INDEL

Strain	Insertion		Insert 0.2-2kb	deletion		Coverage
	number	size range	number	number	size range	
RCC5417	101	36-15979 bp	36	80	50-19025bp	x9
RCC1613	149	51-18972 bp	37	67	50-13730 bp	x239
RCC685	116	51-16349 bp	60	69	51-19026 bp	x52
RCC1615	62	72-4620 bp	23	69	73-13730 bp	x4
RCC1868	49	50-5509 bp	22	43	51-13732 bp	x19
RCC4222	14	76-3011 bp	11	5	51-3999 bp	x30
RCC4752	53	50-15411 bp	34	60	50-13730 bp	x14
mean*	88		35			
*without RCC4222						
INDEL within aligment						

886

887

Table 2. Distribution of diversity markers in world-wide strains

Isolates	**yrcD prom chr1**	**TOC1 prom chr17**	**Flavodoxin chr3**	**Zinc finger chr15**	**TIMa chr14**
primers	MDB33+MDB34	MDB7+MDB9	MDB40+MDB41	MDB68+MDB69	MDB57+MDB58
RCC4222	400 bp	700 bp	800 bp	730 bp	530bp
RCC5417	1400 bp	700 bp	600 bp	900 bp	530bp
RCC1613	1400 bp	2.2 kb	1000 bp	730 bp	530bp
RCC685	1400 bp	700 bp	1200 bp	820 bp	530bp
RCC1615	400 bp	2.2 kb	1400 bp	730 bp	530bp
RCC1868	200 bp	700 bp	800 bp	730 bp	530bp
RCC4752	1400 bp	2.2 kb	1200 bp	730 bp	530bp
	3 size variants	2 size variants	5 size variants	3 size variants	1 size variant
	14.28% (200 bp)	42.86% (2.2 kb)	28.57% (1.2 kb)	71.5% (730 bp)	100% (530 bp)

Table 3. Multi Loci Genotypes (MLG) of Banyuls isol

Strains	Chr1 yrcD prom	Chr17 TOC1 prom	Chr3 flavodoxin	Chr15 C2H2	chr14 TIMa			
4222 Banyuls	400bp	700bp	800bp	730 bp	530bp			
4752 Naples	1.4kb	2.2kb	1200bp	730 bp	530bp			
Banyuls 18/19								
MLG 1	200bp	2.2 kb	1200bp	730 bp	530bp	29 strains	5/9 samplings	53%
MLG 2	200bp	2.2kb	1200bp	730 bp	0	16 strains	5/9 samplings	29%
MLG 3	200bp	2.2 kb	800bp	730 bp	530bp	4 strains	3/9 samplings	8%
MLG 4	200bp	700 bp	1200bp	730 bp	530bp	2 strains	2/9 samplings	4%
MLG 5	200bp	2.2kb	1200bp	820 bp	0	1 strain	1/9 samplings	2%
MLG 6	200bp	700 bp	1200bp	820 bp	0	1 strain	1/9 samplings	2%
MLG 7	400bp	700 bp	1400bp	820 bp	530bp	1 strain	1/9 samplings	2%
MLG 8	200bp	2.2kb	1600bp	730 bp	530bp	1 strain	1/9 samplings	2%
Mediterranean								
Size number	2*	2	4	2	2			
Frequency	98% (200bp)	91% (2.2 kb)	89% (1.2kb)	95% (730 bp)	66% (530 bp)			
Globally dispersed								
Size number	3	2	4	3	1			
Frequency	14.28% (200 bp)	42.86% (2.2 kb)	14.28% (1.2 kb)	71.5% (730 bp)	100% (530 bp)			

* no insertion of ANK gene tested with MDB33+MDB35

similar to RCC4222

similar to RCC 4752

888

889 **Figure 1.** Diversity markers among world-wide accessions

890 Samples are arranged on a latitudinal gradient from Arctic (RCC5417) to Mediterranean

891 (RCC4750) sea. The primers used for amplification are specified at the top of the gel. Arrows

892 indicate the size of the various fragments. In 1C TIMb, B1-B7 are 7 seven Bathycoccus

893 isolates from Banyuls during winter 2018/2019.

894

895

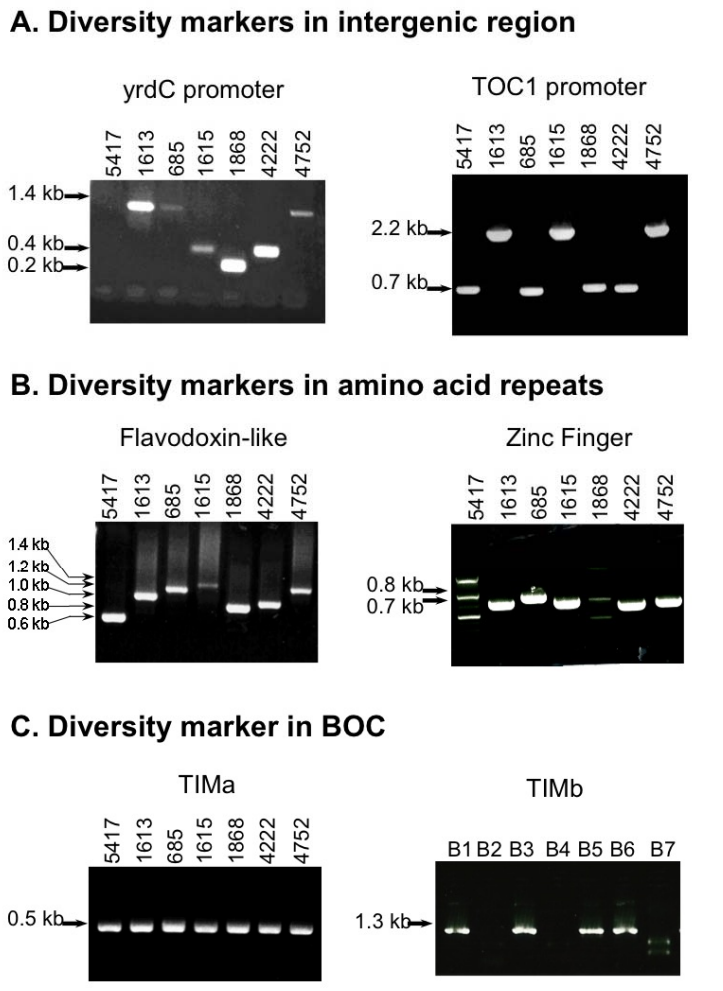


Figure 1. Diversity markers world-wide accessions

896

897

898 **Figure 2.** Geographical distribution of TOC1 and AdoMTase proteins

899 The protein sequences of TOC1 and AdoMTase were used as a query at high stringency in
900 OGA. Under these conditions, a single hit was found and its presence and abundance are
901 represented on the world map at surface upper water (SFR) and dense chlorophyll maximum
902 layer (DCM) and in relation to latitude and temperature. The arrows point to a station at the
903 Chilean coast where TOC1 is present but not AdoMTase.

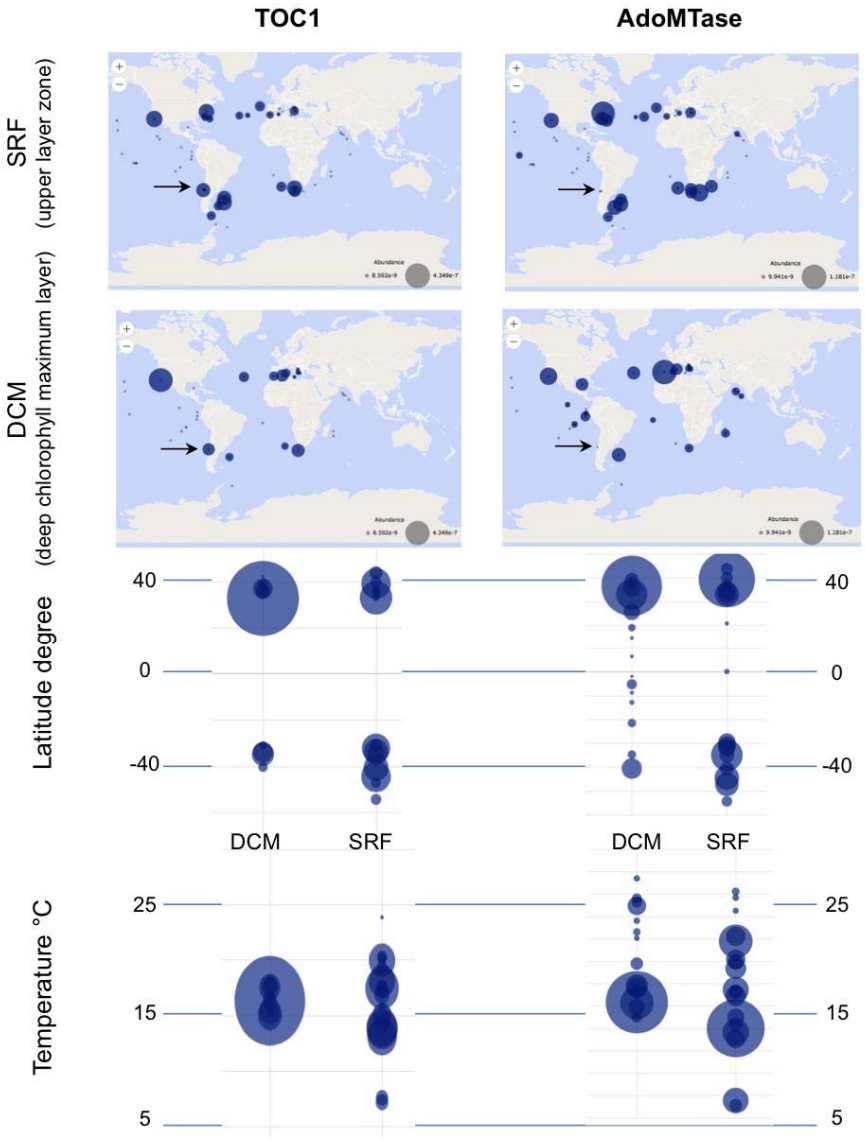


Figure 2 : Geographical distribution of TOC1 and AdoMTase proteins

904

905 **Figure 3.** Abundance of phytoplankton during three successive blooms
906 Seawater at SOLA buoy (Banyuls) was sampled at a depth of 3 meter. After passage on 3 µm
907 filter, the flow through was analysed by flow cytometry. At each sampling time, the
908 phytoplankton was categorised and quantified in function of cell size (pico- and nano-
909 phytoplankton, cyano-bacteria) with indication of the seawater temperature. The mean
910 seawater temperatures in the three years were in December 15.9°C in 2018, 14.8°C in 2019
911 and 14.7°C in 2020, in January/February 11.6°C in 2019, 12.8°C in 2020 and 11.8°C in 2021,
912 in March/April 12.8°C in 2019, 12.2°C in 2020 and 12.8°C in 2021. The main peak of
913 picophytoplankton abundance was in December 3th 2018 and February 19th 2019, January 7th
914 and February 11th 2020, January 27th and March 2nd 2021. The most striking difference
915 between the three years was the sudden abundance in nanophytoplankton in March 2021.

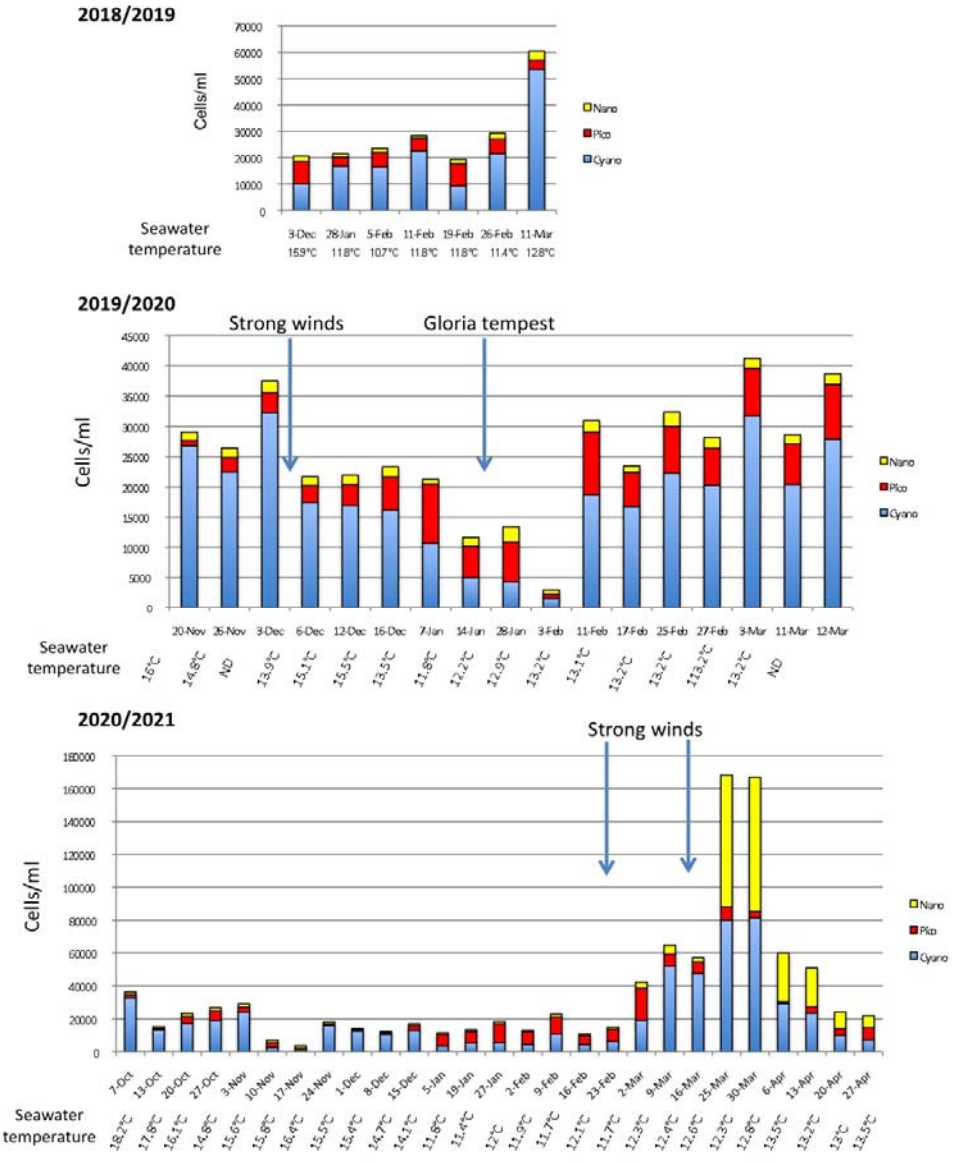


Figure 3. Abundance of phytoplankton during three successive blooms

916
917

Figure 4. Distribution and abundance of TIM variant proteins in MetaG database
A. Phylogenetic tree of TIM proteins from world-wide and Banyuls isolates presenting two main clades each containing one TARA OGA hit, OGATIM4222 or OGATIMA8. OGATIM4222 was retrieved after a query with the TIM protein from RCC4222 and OGATIMA8 with the variant TIM protein in the A8 isolate.
B. Geographical presence of TIM variants in OGA. The presence and abundance of each variant is represented in SRF and DCM samples with their correlation to the temperature and latitude parameters.

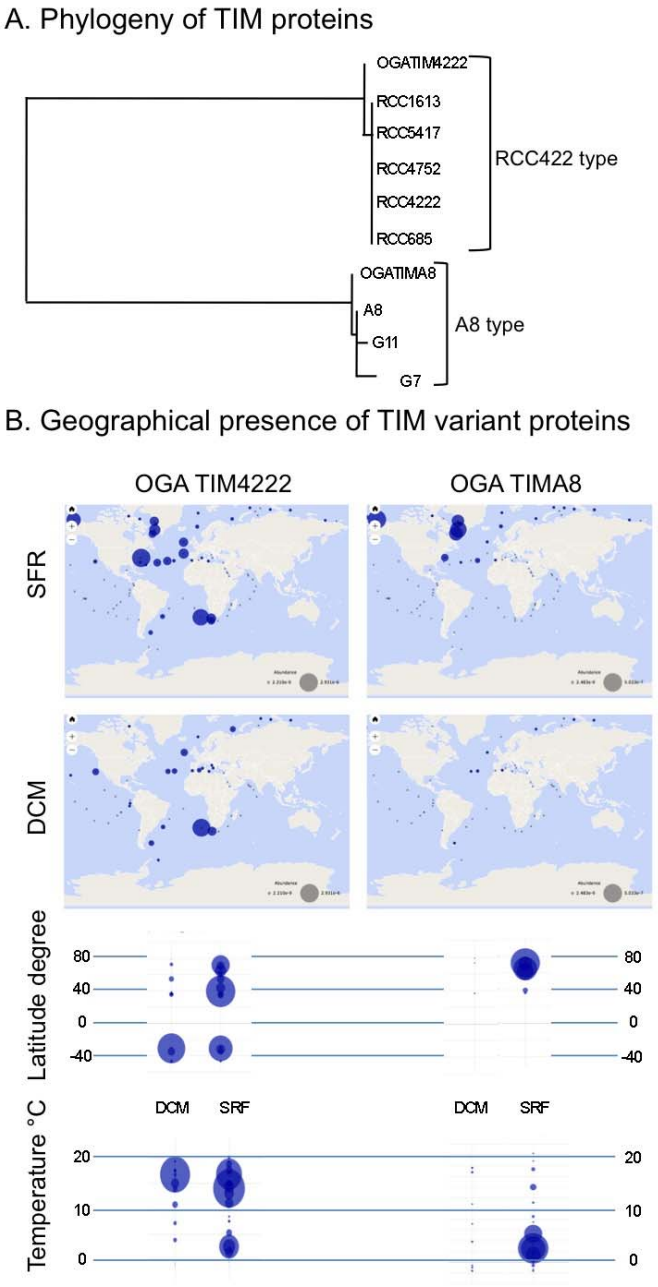


Figure 4. Distribution and abundance of TIM variant proteins in MetaG database

927 **Figure 5**. Diversity markers in seawater

928 For clarity, only extracts are presented in this Figure. The complete set of results is available
929 in Supplemental Figure3 and Supplemental Table3.
930 Seawater was filtered in autumn from end of November in 2019 and from October 2020. The
931 relative abundance of the 2 allelic variants of yrdC (200 bp, a deletion or 400 bp, the
932 reference type) and TIMa and b (absence or presence) were recorded. The onsets of the bloom
933 were different both their chronology and their population diversity. ND : not determined.

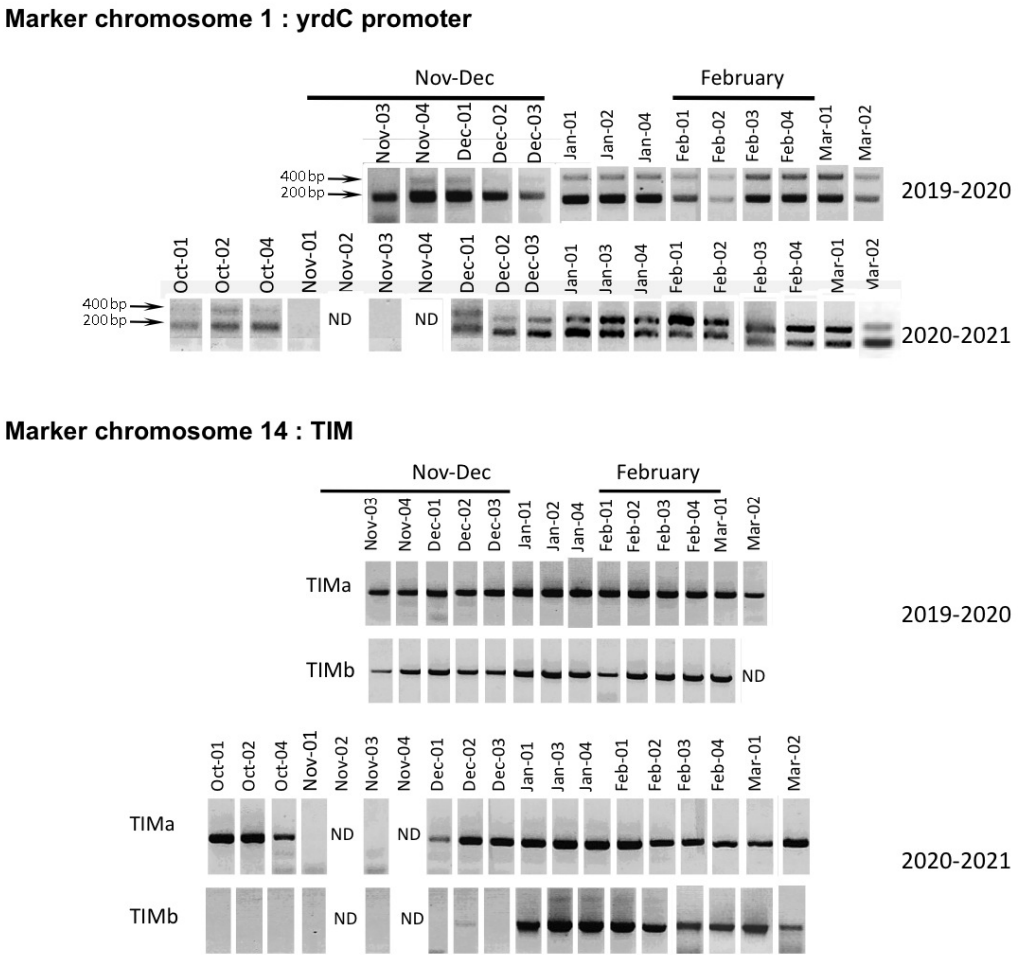


Figure 5. Diversity markers in seawater

934

935 **Figure 6.** Growth of Mediterranean Bathycoccus isolates

936 The growth curve of the "December" and "February" Bathycoccus isolates was determined
937 under 4 different conditions by sampling every day for 9 days. Cell concentration was
938 determined by flow cytometry and is expressed as 10^6 cells/ml.

939

940

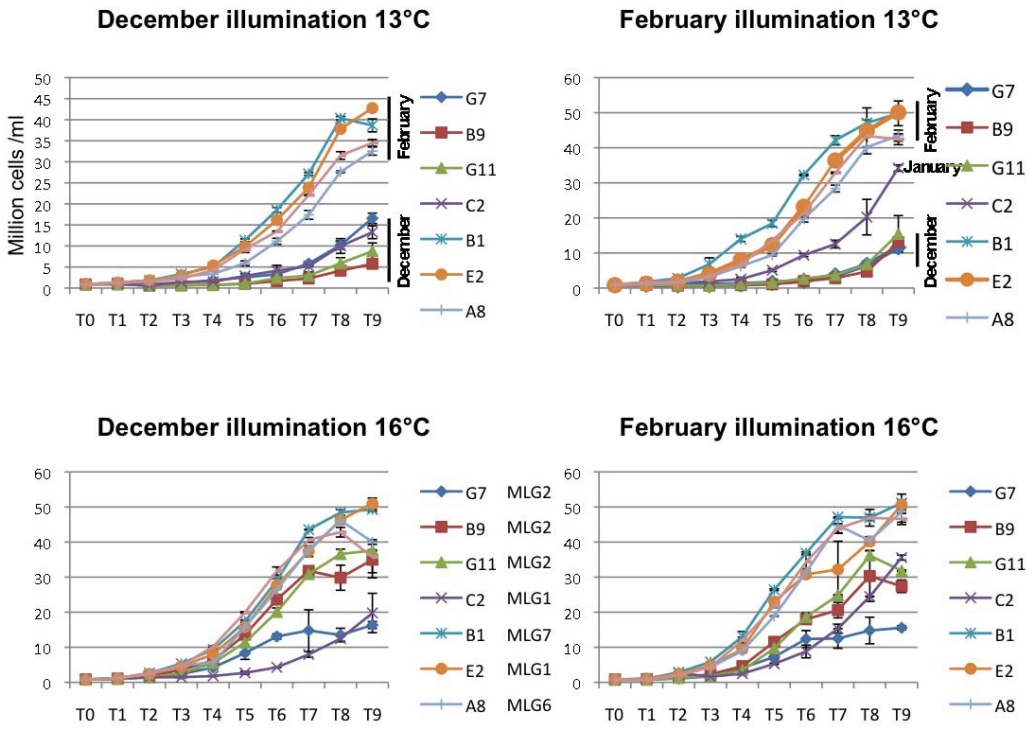


Figure 6 : Growth of Mediterranean Bathycoccus isolates

941

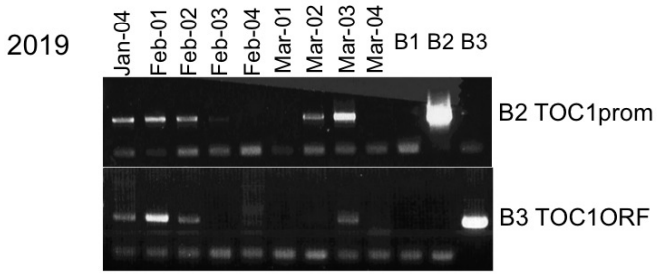
942

943 **Figure 7.** Presence of two additional Bathycoccus species in Banyuls bay
944 A. Alignment of ITS2 nucleotide sequences from B1 (RCC4752, B2 (RCC716) and B3 (C3) strains
945 and B4 from sea water samples in Kara sea (Belevich et al. 2021).
946 B. Presence of Bathycoccus B2 and B3 type in Banyuls seawater. Primers specific to B2 genome were
947 derived from TOC1 promoter and intein (Monier et al. 2013) in RCC716 and primers specific to B3
948 were designed from non-conserved sequences of TOC1 ORF in the C3 isolate. They were used on
949 seawater samples from 2019 and tested for their specificity on isolated DNA from B1, B2 and B3
950 strains.
951 C. Ultrastructure of *Bathycoccus catiminus* B3 type
952 Transmission electron microscopy image of (a) a B1 Banyuls isolate, (b) a B3 Banyuls isolate and (c)
953 a close up of detached spider web scales from B3 type Bathycocus. N: nucleus, Cp: chloroplast, St:
954 starch granule, Sc: scale. Bar = 100 nm.
955

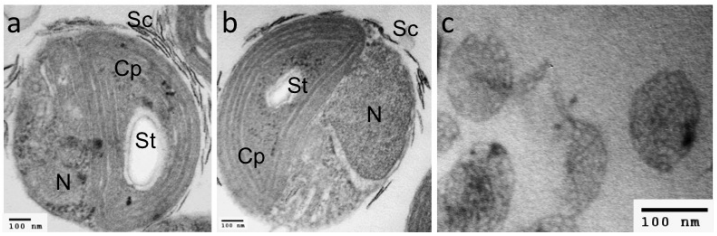
A. ITS2 Alignment

```
B1  TGTCTCCTCACCTCACTT-TTATTTT---------TGAGCGTGGATCTGGGCGTCCGGAA
B2  TGTCTCCTCACCTCTCCT-------T---------TGAGCGTGGATCTGGGCGTCCGGAA
B3  TGTCTCCTCACCTCACCT-TTATTGTATAAAGGTACGAGCGTGGATCTGGGCGTCCGGAG
B4  TGTCTCCTCACCTCTCCTATTATTGTATTTTAATAAAGAGGTGGATATGGGCGTCCGGAA
    ************** * *       *               ****** ************

B1
B2  TCTTTTGTTTTA----TTGAAAGA-CTCGGGTCGCCTGAAAAACAGTCGTACGTGCGACT
B3  TCTTTCGATTTT----TAGAAAGA-CTCGGGTCGCCTGAAAAACAGTCGTACGTGCGACT
B4  TCTTTCGTGTATATAATTGAAAGA-CTCGGGTCGCCTGAAAAACAGTCGTACGTGCGACT
    CATATTGTGTGTAAAAAACAAACAGTTCGKGTCGCCTGAAAAATAGTCGTACGTGCGACT
      * * *  *          *** *  *** ************* ****************

B1  GTCGCATAACCAACGTGGTAGACCACTCCGGTGGACGATCG-TTCGGTTTGACAGTTGTT
B2  GTCGCACAACCAACGTGGTAGACCACTCCGGTGGACGATCG-TTCGGTTTGACAGTTGTT
B3  GTCGCATAACCAACGTGGTAGACCACTCCGGTGGACGATCG-TTCGGTTTGACAGTTGTT
B4  GTCGCACAACCAGCGTGGTAGACCACTCCGGTGGACGATCGCTTCGGTTTGACAGTTGAT
    ****** ***** **************************** ***************** *

B1  TACCTAAGTATGATCTCGACCG-AATTTATTTCCGT-GTCGAAACCTGTGCTTTTTTACT
B2  TACCTAAGTATGATCTCGGCCG-AATTCATCTTCGC-GTCGAAACCTGTGCTTTCTTTAC
B3  TACCTAAGTATGATCTCGACCG-AATTGACTTTCGCTGTCGAAACCCGTGCTTTTTTACT
B4  TACTTA-GTACGATCTCGCTTTCGATTTAATTTATCGAGCGAAACCAGTGCTTTTATGAC
    *** ** *** *******       *** *  *        ******* *******  *

B1  CCGCACTTTTTCTT-TCACAAGAAAGAGAGCAAACCAAAAAAGCTTCAC
B2  CCGCACTCTTCATTCTTGTAATGCAGAGCGCACCCAAGGAAAGCTTCAC
B3  ---CACTCTTCTCT-TCATCGAAAAGAGTAC---CCAAAAAAGCTTCAC
B4  T----CTGTTTATT-TTATTATAAACAGGAC----AATAAATGCTTCAC
        ** **   * *        * **  *     *  ** *******
```

B. Presence of Bathycoccus B2 and B3 type in Banyuls seawater



C. Ultrastructure of *Bathycoccus catiminus* B3 type



956 **Figure 7.** Presence of two additional Bathycoccus species in Banyuls bay