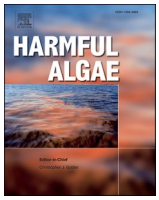
Harmful Algae 142 (2025) 102791

Contents lists available at ScienceDirect

Harmful Algae

journal homepage: www.elsevier.com/locate/hal



Genomic diversity in time and space in the toxic diatom *Pseudo-nitzschia multistriata*

Svenja Mager [a,1], Francesco Manfellotto [a,1], Antonella Ruggiero [a,1], Viviana Di Tuccio [a], Federica Cerino [b], Stefano Accoroni [c,d], Tomohiro Nishimura [e], Marta Mikhno [a], Neri Fattorini [a], Timotej Turk Dermastia [f], Jasna Arapov [g], Sanda Skejic [g], Lesley Rhodes [e], Kirsty Smith [e,h], Antonio Longo [a], Caterina Manzari [i], Lisa Campbell [j], Graziano Pesole [i,k], Remo Sanges [l], Francesca Raffini [a], Maria Valeria Ruggiero [a], Monia Teresa Russo [a], Marina Montresor [a], Maria Immacolata Ferrante [a,b,*]

[a] *Stazione Zoologica Anton Dohrn, Naples, Italy*
[b] *Oceanography Section, National Institute of Oceanography and Applied Geophysics (OGS), Trieste, Italy*
[c] *Dipartimento di Scienze della Vita e dell'Ambiente, Marche Polytechnic University, Ancona, Italy*
[d] *Fano Marine Center, The Inter-Institute Center for Research on Marine Biodiversity, Resources and Biotechnologies (FMC), Fano, Italy*
[e] *Cawthron Institute, Nelson, New Zealand*
[f] *Marine Biology Station Piran, National Institute of Biology, Piran, Slovenia*
[g] *Institute of Oceanography and Fisheries, Split, Croatia*
[h] *School of Biological Sciences, University of Auckland, Auckland, New Zealand*
[i] *Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies, National Research Council, Bari, Italy*
[j] *Department of Oceanography, Texas A&M University, TX, United States*
[k] *Department of Biosciences, Biotechnology and Biopharmaceutics, University of Bari "A. Moro", Bari, Italy*
[l] *Computational Genomics Laboratory, Neuroscience Area, International School for Advanced Studies (SISSA), Trieste, Italy*

A R T I C L E I N F O

Keywords:
Diatom
Pseudo-nitzschia multistriata
Genome resequencing
Population genomics
Genetic variation
Domoic acid
Dab genes

A B S T R A C T

Understanding the origin and maintenance of genetic diversity is crucial to elucidate population dynamics of unicellular microalgae, their microevolutionary history and their adaptive ability. The planktonic, domoic acid-producing diatom *Pseudo-nitzschia multistriata* has a ubiquitous distribution in the world oceans and past population genetics studies, based on few genomic loci, have shown a clear temporal structure over different years in the Gulf of Naples (Italy). Despite the ecological and toxicological importance of this organism, detailed information on its diversity across the whole genome and at the population level is still lacking. We collected *P. multistriata* strains in the Gulf of Naples in five different years, obtained strains from the Adriatic Sea, the Gulf of Mexico and New Zealand coasts, and resequenced the whole genomes of a total of 28 strains at high coverage. While strains from the first three geographical areas were capable of producing the toxin domoic acid, the New Zealand strains had been reported to be non-toxic. A comparison of the domoic acid biosynthetic (dab) genes sequences between toxic and non-toxic strains showed very little variation among the strains, and no disrupting mutation was found in the *dab* genes in the non-toxic strains. On the other hand, the *dab* genes showed higher levels of expression in toxic strains than in non-toxic strains, suggesting that, in this species, absence of toxicity is explained by gene regulation rather than *dab* sequence divergence. Variant analysis showed stronger spatial than temporal genetic structuring and a clear separation was observed between the New Zealand strains and the others, the former having a greater content of genes under selection. Overall, the genomes of the different groups, including strains from a clonal bloom, did not appear to contain major rearrangements. Our findings contribute to enlarging our understanding of diatom diversity, a key factor underlying diatom success, and provide novel data on the longstanding problem of *Pseudo-nitzschia* toxicity.

* Corresponding author.
E-mail address: mariella.ferrante@szn.it (M.I. Ferrante).
[1] S. Mager, F. Manfellotto and A. Ruggiero contributed equally.

https://doi.org/10.1016/j.hal.2024.102791
Received 8 August 2024; Received in revised form 21 November 2024; Accepted 20 December 2024
Available online 22 December 2024

1. Introduction

Diatoms are a highly diverse group of microscopic single-celled algae that play an important role as primary producers in the oceans (Benoiston et al., 2017; Tréguer et al. 2018). In the last decades, several studies mainly based on genotyping approaches of microsatellite markers showed that species of unicellular microalgae are not always panmictic. Instead, they include genetically distinct populations at a spatial, *i.e.*, different geographical locations/sites (Benoiston et al. 2017; Casteleyn et al. 2010; Sefbom et al. 2018; Ruggiero et al. 2024), and temporal scale, *i.e.*, within a single bloom season or between blooms occurring in different years (Godhe et al. 2016; Sassenhagen et al. 2018). Population genetics studies have been carried out since 2008 also with the marine planktonic diatom *Pseudo-nitzschia multistriata*, recorded regularly in the Gulf of Naples during summer and autumn (D'Alelio et al. 2010; Tesson et al. 2014; Ruggiero et al. 2018, 2024). As most of the pennate diatoms, *P. multistriata* has a heterothallic mating system, with separate sexes called mating types (MT) (D'Alelio et al. 2010; Scalco et al. 2016). Analyses based on microsatellites indicated that different populations appeared in the Gulf of Naples over the years and the genetic loci tested in those studies were generally in linkage equilibrium, suggesting that sexual reproduction contributes to maintaining genetic homogenization. These studies also revealed occasional clonal expansions, and the occurrence of two cases in which an unbalance in the sex ratio was observed. Specifically, a *P. multistriata* bloom in 2013 was shown to be dominated by a single genotype (Ruggiero et al. 2018; 2024), while in general diatom blooms often maintain high genotypic diversity (Rynearson and Armbrust 2005; Tesson et al. 2014; Godhe et al. 2016; Ruggiero et al. 2024).

P. multistriata has been selected as a model for studies on sexual reproduction (Bilcke et al. 2022), and a reference genome and several transcriptomics datasets are available for the species (Basu et al. 2017; Annunziata et al. 2022; Ferrante et al. 2023), yet detailed data on genetic diversity at the population level is still lacking. Here we aimed at assessing the relationships at the genomic level by obtaining and analyzing whole genomes of multiple strains from different geographic sites: New Zealand, the Gulf of Mexico and the Adriatic Sea (Mediterranean Sea), that derived from opportunistic sampling with different purposes. Additionally, we included the Gulf of Naples and sampled it repeatedly over different years, providing also a temporal resolution to our study. While the generally low levels of *P. multistriata* abundance worldwide, the narrow temporal windows of presence at the LTER MareChiara in Naples, and the difficulty of maintaining many strains in axenic conditions for long enough to obtain high quality DNA did not allow to obtain large sample sizes generally used for population genomics, our resequencing dataset currently represents one of the largest for diatoms.

Pseudo-nitzschia species are regarded as harmful microalgae, with half of the 60 putative species capable of producing the neurotoxin domoic acid (DA) (Bates et al. 2018). DA can have detrimental effects on various marine organisms as well as on human health (Petroff et al. 2021), causing a syndrome known as Amnesic Shellfish Poisoning (ASP). *Pseudo-nitzschia* cell content of DA can be variable and in the same strain DA production can be stimulated by different triggers, such as phosphate or silica limitation, or the presence of grazers (Bates et al. 2018; Harðardóttir et al. 2019). *P. multistriata* is one of the toxic species, with its strains capable of producing variable levels of toxin (Orsini et al. 2002; Turk Dermastia et al. 2022) even though there are reports of non-toxic strains (Rhodes et al. 2000; Amato et al. 2010). DA is a small molecule synthesised from the precursors L-glutamic acid and geranyl diphosphate, a C10 isoprenoid (Chekan et al. 2020). Biochemical and transcriptomic studies had indicated that enzymes belonging to the isoprenoid pathway, specifically those involved in the methyl-erythrito phosphate (MEP) route, are implicated in the toxin biosynthesis (Douglas et al. 1992; Harðardóttir et al. 2019). The identification of a four-genes genomic cluster for DA biosynthesis (dab), highly conserved between *P. multistriata* and its sister species *P. multiseries,* was crucial to refine the DA biosynthetic (dab) pathway (Brunson et al. 2018). However, the underlying differences between toxin- and non-toxin-producing *Pseudo-nitzschia* strains, and the relation to the dab pathway, are not yet understood.

In this study, we aimed at exploring intraspecific diversity at the genomic level, and in toxin production related genes. Specifically, we explored *dab* genes sequence conservation at a large spatial scale across the Gulf of Naples, New Zealand, Adriatic Sea and the Gulf of Mexico populations, including toxic and non-toxic strains, revealing the presence of intact *dab* genes in all samples. We tested gene expression levels on a subset of these samples and show that non-toxic strains have very low expression of their potentially functional *dab* genes. Using the whole genomes of 28 natural strains, we also aimed at following up past population genetics studies, looking at whether a genetic structure could be found on a temporal and/or spatial scale, and at identifying major rearrangements and genes and genomic regions under selection, with a special focus on genomes from strains collected during the 2013 clonal bloom. Knowledge of diatom gene function, still limited for a large fraction of genes, is gradually improving, and our dataset will serve for many purposes, including a better refinement of the relationship between traits and their genetic basis. Our findings provide answers to fundamental questions on *Pseudo-nitzschia* biology and ecology.

2. Materials and methods

2.1. Strain collection, culture conditions and crosses

For this study, 15 *Pseudo-nitzschia multistriata* wild strains were collected between 2013 and 2020 at the LTER site MareChiara in the Gulf of Naples, Italy (DEIMS iD: https://deims.org/0b87459a-da3c-45af-a3e1-cb1508519411). Five strains were obtained from the Adriatic Sea, along the Italian, Slovenian (Turk Dermastia et al. 2022) and Croatian coasts, two from the Gulf of Mexico, US (Russo et al. 2021), and six from the coasts of New Zealand (Nishimura et al. 2021) (Table 1). The strain B856, used to generate the reference genome, is an F2 strain derived from a cross of two strains isolated at LTER-MC in the Gulf of Naples in 2009 (Basu et al. 2017). One additional strain (1364-5) collected at LTER-MC in 2020 was used only for domoic acid (DA) assessment and for gene expression analysis. Individual cells were isolated with a micropipette and grown in f/2 medium supplemented with silicates (Guillard 1975) at a temperature of 18 °C, an irradiance of 50 µmol photons m^{-2} s^{-1} and a photoperiod of 12 L:12 D h. We adjusted the f/2 medium salinity according to the strains' isolation place (36 ppt for the Gulf of Naples and Adriatic Sea populations; 33 ppt for the Gulf of Mexico and New Zealand populations). To test sexual compatibility and assess the mating type in the collected strains, they were crossed both with MT- and MT+ strains isolated at LTER-MC (Table S1) using the same procedure described in (Scalco et al. 2016). For genome resequencing, strains were grown in f/2 medium supplemented with antibiotics, as described in the protocol "*Axenic Diatoms cultures protocol" (dx.doi.org/10.17504/protocols.io.bgudjws6)*. The absence of bacteria was tested as described in the protocol "*Detection of bacteria in antibiotic-treated diatom cultures and cell harvesting*" *(dx.doi.org/10.17504/protocols.io.btt5nnq6)*.

2.2. DNA extraction, genome sequencing and mapping

DNA was extracted following the protocol *"Genomic DNA extraction from the diatom Pseudo-nitzschia multistriata for Illumina sequencing" (dx.doi.org/10.17504/protocols.io.byk7puzn)*. Nucleic acids quantification was done by Qubit® 2.0 Fluorometer (Life Technologies) while integrity was assessed by 1 % agarose gel electrophoresis. The ITS sequence for strain 1020F was obtained in this study (Supplementary Fig. S1), while barcoding information for some of the other strains is provided in Nishimura et al. (2021), Turk Dermastia et al. (2022) and Russo et al.

(2021).

All strains were sequenced through Illumina Nextseq, producing 100 bp paired-end reads for strains 1119-15,1078-25, 1078-30, 1120-32 and 1120-47 and 150 bp paired-end reads for all other strains. The reference genome was assembled from three different Illumina libraries of strain B856 as described in (Basu et al. 2017). To add the reference B856 as an additional sample for variant calling, original Illumina HiSeq 100 bp paired-end reads were used and treated in the same way as the other 28 strains. The raw sequencing data of the newly sequenced strains are available under the accession number PRJNA882609. Filtering, mapping and quality controls were done as outlined in Note S1 in Supplementary Materials.

2.3. Variant calling

The identification of genetic polymorphisms in comparison to the reference genome B856 was done using gatk v4.2.3.0 (Poplin et al. 2018). For several analyses a file containing the variant sites among all 29 strains (including the reference B856) was used. Several multisample vcf files were produced including varying sets of strains, depending on the downstream analysis performed as described in Supplementary Note S2. In short, a GATK pipeline was used by first running Haplotypecaller to identify variants in individual samples. With combinegvcf the single gvcf files of the desired strains were combined and the combined data was then jointly genotyped via genotypegvcfs, producing the multisample vcf files that were used in the respective downstream analyses. All sets of variants were filtered to get high quality variant sets as described in Note S2.

2.4. Domoic acid quantification

DA production was assessed for the *P. multistriata* strains grown under the same condition stated above, with different methods since experiments were performed at different times during the project (Table 2). Data for strains 1119-15, 1120-32 and 1120-47 were obtained from cells collected in the stationary phase with an ELISA kit (*Europroxima©*) following manufacturer instructions. DA cellular content (fg

Table 2
DA content (fg cell^{-1}) in *P. multistriata* strains.

Strain	DA concentration (fg/Cell^{-1})	Test method
1119-15	3.47	TOF-MS
1119-15	8.04	ELISA
1120-32	26.55	ELISA
1120-47	40.11	ELISA
1364-5	0.926	HPLC
1365-11	0.97	HPLC
P4C2	4.35	HPLC
P4C5	0.255	HPLC
NZ5	N.D.	HPLC
NZ1	N.D.	LC-MS/MS*
NZ3	Trace	LC-MS/MS*
NZ11	Trace	LC-MS/MS*
MS3-SLO	up to 20.7	ELISA/HPLC/LC-MS #
A4-SLO	up to 114	ELISA/HPLC #

* Nishimura et al. 2021.
\# Turk Dermastia et al. 2022

In bold are strains used for the gene expression analysis. TOF-MS = Time-of-flight mass spectrometry, ELISA = ELISA kit (Europroxima©), HPLC = High-pressure liquid chromatography, LC-MS/MS = liquid chromatography-tandem mass spectrometry. N.D., not detected.

Table 1
Information on the 29 genomes analysed in the study.

Strains	Isolation Date	Geographic location of origin	MT	Coverage[a]	Mapping Rate [%][a]	Reference bases covered by ≥ 10 Reads [%][a]
B856	02/08/2011	B855 x B854, Italy, Gulf of Naples, LTER-MC	+	246	74	90
1119-15	02/09/2014	Italy, Gulf of Naples, LTER-MC	+	23	73	81
1078-25	08/10/2013	Italy, Gulf of Naples, LTER-MC	+	56	76	85
1078-30	28/10/2013	Italy, Gulf of Naples, LTER-MC	+	70	73	86
1120-32	09/09/2014	Italy, Gulf of Naples, LTER-MC	–	73	70	86
1120-47	02/09/2014	Italy, Gulf of Naples, LTER-MC	–	87	71	86
1264-D1	29/08/2017	Italy, Gulf of Naples, LTER-MC	–	37	30	85
1264-D5	29/08/2017	Italy, Gulf of Naples, LTER-MC	+	29	18	85
1334-1	21/06/2019	Italy, Gulf of Naples, LTER-MC	–	123	75	87
1334-4	21/06/2019	Italy, Gulf of Naples, LTER-MC	ND	93	65	87
1334-5	21/06/2019	Italy, Gulf of Naples, LTER-MC	+	116	77	87
1340-6	05/08/2019	Italy, Gulf of Naples, LTER-MC	ND	23	17	81
1342-30	03/09/2019	Italy, Gulf of Naples, LTER-MC	+	109	65	87
1343-46	10/09/2019	Italy, Gulf of Naples, LTER-MC	+	107	70	87
1364-6	07/07/2020	Italy, Gulf of Naples, LTER-MC	+	172	63	87
1365-11	14/07/2020	Italy, Gulf of Naples, LTER-MC	–	173	70	88
NZ1 (PA014Ps09)	14/01/2020	New Zealand, Kerikeri, Far North, Northland	+	191	75	86
NZ3 (G027Ps06)	06/07/2020	New Zealand, Brightlands Bay, Marlborough	+	167	70	84
NZ5 (G013Ps22)	30/06/2020	New Zealand, Oyster Bay, Marlborough	+	171	73	85
NZ7 (PB024Ps02)	20/07/2020	New Zealand, Bream Bay, Whangārei, Northland	–	197	72	84
NZ10 (PB024Ps08)	20/07/2020	New Zealand, Bream Bay, Whangārei, Northland	–	180	73	83
NZ11 (G001Ps01)	21/06/2020	New Zealand, Collingwood Farms, Tasman Bay	–	195	72	87
OGS-PS62	09/12/2019	Italy, Adriatic Sea, Gulf of Trieste, LTER-C1	–	56	66	85
MS3-SLO (MS3)	16/12/2019	Slovenia, Adriatic Sea, OB2 aquaculture site	+	82	72	85
A4-SLO (119 A4)	13/01/2019	Slovenia, Adriatic Sea, Slovenian LTER station 00F	–	92	70	85
1020F	19/10/2020	Italy, Adriatic Sea, Ancona coast	ND	154	64	87
IOR20-05	01/12/2020	Croatia, Adriatic Sea, Krka River estuary	ND	161	60	85
P4C2	25/07/2018	Texas US, Gulf of Mexico, Surfside Beach	+	14	13	38
P4C5	25/07/2018	Texas US, Gulf of Mexico, Surfside Beach	+	39	31	81

[a] Numbers depict mapping information after quality filtering.

Abbreviations: MT = Mating type, LTER = Long-Term Ecological Research, MC = MareChiara, ND = MT could not be determined by experimental crosses as the strains failed to produce sexual stages.

cell^{-1}) was assessed using the following formula:

$$DA = \left[(Ev)(CV)^{-1}\right]10^{6}$$

E is the DA concentration determined by ELISA kit, v is the volume of solvent used for extraction, C is the culture concentration expressed in cells/ mL^{-1}, V is the culture volume used for DA assessment. DA content for strain 1119-15 was also assessed by time-of-flight mass spectrometry (TOF-MS) assay for a comparison with the ELISA data. Cellular content of DA for strains 1365-11, 1364-5, NZ5, P4C2 and P4C5 was measured from cells collected in the late exponential phase via high-performance liquid-chromatography (HPLC) of the fluorenylmethoxycarbonyl derivative as described by (Pocklington et al. 1990). RNA samples used for DA genes expression analysis were collected simultaneously with samples for DA quantification.

2.5. Protein alignment and modelling

To investigate sequence variation for the DA biosynthetic pathway genes, as well as for the gene *MRP3,* consensus sequences were isolated with bcftools v1.14 (Danecek et al. 2011) from genome sequencing data of the 29 strains. Corresponding amino acid sequences were obtained using the EditSeq®, Version 5.00 (DNASTAR. Madison, WI) software and were aligned using the multiple sequence alignment online software Clustal Omega (Sievers and Higgins 2018).

Uncertain residues (X and B on the original sequences) were changed to the smallest amino acid possible, alanine, based on the consensus in the assembled sequences. Sequences were annotated using InterPro scan (https://www.ebi.ac.uk/interpro/). Signal peptide was cleaved from Daba and Dabd, and sequences submitted to Colabfold v1.5.5 (Mirdita et al. 2022) to predict protein fold. The shallowness of the mmseq2-based MSA (<5 sequences) was balanced by the increase of the number of recycles for Dabb and Dabc. Structures were visualised with ChimeraX v1.7 (Meng et al. 2023).

2.6. RNA extraction and genes expression analysis

To compare the levels of *dab* and selected isoprenoid biosynthetic genes expression among selected strains, RNA extractions were performed as described in the protocol "*RNA extraction from the diatom P. multistriata*" (dx.doi.org/10.17504/protocols.io.261gen627g47/v1). 0.5 µg of total RNA was reverse-transcribed using the QuantiTect® Reverse Transcription Kit (Qiagen) following manufacturer's instructions. *dab* and isoprenoid genes expression levels were evaluated by qPCR. Gene primers, designed on the *P. multistriata* genome, are reported in Table S2.

The qPCR experiments were performed in triplicate in a ViiA 7 Real-Time PCR System (Applied Biosystems, Foster City, CA, USA) using Fast SYBR Green Master Mix (Applied Biosystems, Foster City, CA, USA), following manufacturer instructions. The reference gene used was *TUB-A* (Adelfi et al. 2014). The relative gene expression was calculated with the following formula: relative expression ratio of the gene of interest is $2^{-\Delta CT}$ with $\Delta CT = Ct_{(\text{target gene})}$ minus $CT_{(\text{reference gene})}$. The results SD are the mean of at least three separate experiments, measuring each parameter by triplicate ($n = 3$).

2.7. Genetic diversity in space and time

A pairwise distance matrix (using absolute distance between the two vectors as distance measure) based on the called genomic variants of all 29 strains was produced in R v4.2.0 (R Core Team, 2021) with vcfR v. 1.12.0 (Knaus and Grünwald 2017) and adegenet v. 2.1.3 (Jombart and Ahmed 2011) and a clustering tree deduced from the distance matrix with MEGA v11.0.11 with the neighbor joining tree function (Tamura et al. 2021). The distance scale of the tree indicates the dissimilarity for the specified length of a branch. The sum of all branch lengths between each two strains is the dissimilarity between these two strains.

A Pearson's test was done in R 4.2.0 using the cor.test function of the ggpubr package v. 0.4.0 to test for correlation between genetic distance and geographic distance at a large scale (among the four major sampling sites: Gulf of Naples, Adriatic Sea, New Zealand and Gulf of Mexico) as well as smaller spatial scale (New Zealand, four sampling sites). Geographic distances were measured in Google Maps by measuring the length of the shortest water route between two sampling sites.

A principle coordinate analysis was performed for the 15 strains from the Gulf of Naples in Genalex v. 6.5 (Peakall and Smouse 2006).

For the ADMIXTURE analysis, the vcf file was further filtered to remove singletons, retain only biallelic and unlinked (LD pruning $r^2 < 0.6$) SNPs, and a minor allele frequency higher than 0.05 using vcftools v. 0.1.16 (Danecek et al. 2011) and bcftools v. 1.15.1 (Danecek et al. 2021). Strain ancestries were estimated through a maximum likelihood approach and cross validation procedures using ADMIXTURE v. 1.3.0, PLINK v. 2.00a3.7 (Purcell et al. 2007) and custom scripts. K values ranging from 1 to 15, corresponding to the number of geographic areas and sampling years in the dataset, were tested. To investigate the influence of large spatial scale divergence, this analysis was repeated after excluding the New Zealand strains.

The genome-wide heterozygosity rate was calculated by counting the total number of heterozygous sites in the set of high-quality variants among the 29 strains for each sample and dividing the number by the callable genome size as described in Note S3. To see whether characteristics of subgroups of strains, especially the dominance of the two strains from the 2013 bloom and strains from different populations, could be attributed to differences in the heterozygosity level in certain scaffolds, the heterozygosity level of each scaffold was compared between pairs of subgroups (Note S3).

2.8. Determination of genes under selection

To better understand the adaptation of populations and influences on structuring, the selective pressure acting on genes was measured applying the $\pi N/\pi S$ ratio (Note S4), which determines the rate of substitutions at nonsynonymous (missense) sites to the rate of substitutions at synonymous (silent) sites. It is a measure to investigate the mode of selection acting on a gene, with values < 1 indicating more constraint (purifying selection) while values > 1 indicate positive selection (Jeffares et al. 2015). It has to be noted that positive selection can also lead to reduction of polymorphism at non-silent sites (and linked silent sites) when an adaptive allele is spreading, and higher polymorphism at non-silent sites may be created by other specific types of selection, such as balancing selection at a set on non-synonymous sites. SNPs were annotated with snpEff v5.0 (Cingolani et al. 2012) and the $\pi N/\pi S$ ratio calculated for callable genes as described in Note S4.

The overlap between these genes and the genes under selective pressure determined by (Basu et al. 2017) for orthologous genes between *P. multistriata* and *P. multiseries* (Ka/Ks > 1) was determined and a Fisher's exact test applied to test for significance in the overlap. For the statistical test to be meaningful, genes had to be filtered to take into account only genes considered as callable in both studies.

Callable genes as well as nucleotide diversity and genes under selection were determined for several subsets of strains among our dataset besides for the whole set of 29 strains, namely for the strains from the Gulf of Naples (excluding 1078-30 to include only one of the two highly identical strains from the bloom in 2013), the strains from New Zealand as well as for the strains from the Adriatic Sea (Note S4).

2.9. Determination of dispensable genes

Since homozygous high impact mutations can cause disruption of a gene, this gene can be considered as not crucial for the survival of the strain. To identify the fraction of dispensable genes, high impact variants showing a homozygous genotype in at least one strain among the

variant dataset of the 29 strains were determined as explained in Note S5. With the list of dispensable genes, a GO enrichment analysis was done with TOPGO v2.44.0 (Alexa 2019) to determine overrepresented gene categories.

2.10. Determination of candidate loci under selection

Since regions under selection are key determinants of genetic variation that influences fitness in different environments, we identified such regions in the *P. multistriata* genome using differences in allele frequencies between the Gulf of Naples, the Adriatic Sea, the Gulf of Mexico and New Zealand (i.e., excluding the reference strain) in BayeScan v2.1 (Fischer et al. 2011) as outlined in Note S6. Although ideally more strains would have been needed, the BayeScan analysis was repeated on the strains from the Gulf of Naples only, specifying the two dominant strains as one population and the other strains as another population, to check if outlier sites varying in the allele frequency between the two defined populations could be found, potentially contributing to understanding the dominance of the 2013 genotype.

To check if there are genes with both $\pi N/\pi S > 1$ and outlier loci determined with Bayescan, solidifying the presumed selective pressure, a Fisher's exact test applied to test for significance in the overlap. For the statistical test to be meaningful, the genes with outlier sites had to be filtered for retaining only callable genes as was done for determination of genes with $\pi N/\pi S > 1$.

3. Results

3.1. Strain collection and genome resequencing

We resequenced the genomes of 28 *P. multistriata* strains isolated between 2013 and 2020 from different sites (Fig. 1, Table 1).

P. multistriata cells can be easily recognised in natural samples examined in light microscopy due to their distinctive sigmoid shape (Orsini et al. 2002). In order to confirm that all the strains collected belonged to the same 'biological' species *sensu* (Mayr 2000), we set up crosses and demonstrated sexual compatibility among them. Crosses with reference strains of known MT from the Gulf of Naples and strains from the other geographic locations also allowed MT assignment (Table 1, Table S1).

3.2. Sequences and expression of domoic acid biosynthetic genes

P. multistriata strains from the Gulf of Naples are known to produce domoic acid (Orsini et al. 2002; Amato et al. 2010). The toxin was also detected in Adriatic strains (Turk Dermastia et al. 2022), but not in strains from New Zealand (Nishimura et al. 2021) (Table 2). We tested DA production in five strains from the Gulf of Naples, in the strains from the Gulf of Mexico and in the New Zealand strain NZ5, confirming that in this last strain the toxin was undetectable, while all the other tested strains were moderately toxic (Table 2).

We found minor differences in the amino acid sequences of the four *dab* genes (Brunson et al. 2018) among the 29 strains (including the reference strain used for genome sequencing (Basu et al. 2017)) but none that could suggest loss of function (Fig. 2B, Figs. S2-S5). It is interesting to highlight that variations in *dabd* are all in the peptide signal and transmembrane regions. We measured the *dab* genes expression in two toxin-producing (1365-11, 1364-5) and three non- or trace-producing (NZ1, NZ3 and NZ5) strains, the only three NZ strains that survived long enough to carry out this study. A positive trend between *daba, dabb, dabc and dabd* genes expression levels and toxin production was detected, with an upregulation ranging between 5 and 14 folds in toxic strains with respect to non-toxic strains (Fig. 2C). In addition to the *dab* genes, we measured expression levels of four genes belonging to the isoprenoid pathway, shown to be upregulated when toxicity is induced by grazers cues (Harðardóttir et al. 2019) and to be coregulated with the *dab* genes in *in situ* data (Brunson et al. 2024). We found that *DXS* and *HDR* have higher expression levels in the toxic strains than in the non-toxic ones, whereas the expression levels of *HDS* and *IDI* were similar in all strains (Fig. 2D).

The expression of these eight genes was also tested in an independent set of four Neapolitan strains collected and analysed together in a previous experiment, confirming a general trend of higher *dab* genes expression levels in the two strains with high DA production compared

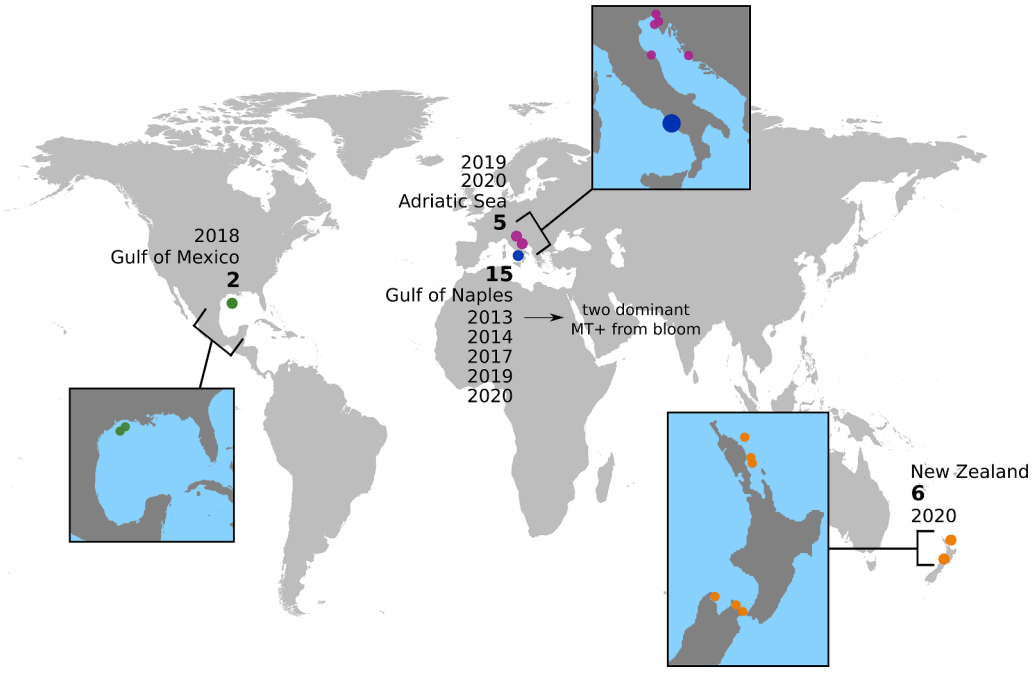


Fig. 1. Sampling locations of 28 *P. multistriata* strains isolated from the Gulf of Naples (15 strains collected between 2013 and 2020, blue dot), the Adriatic Sea (five strains collected in 2019 and 2020, purple dots), New Zealand (six strains collected in 2020, orange dots) and the Gulf of Mexico (two strains collected in 2018, green dot).

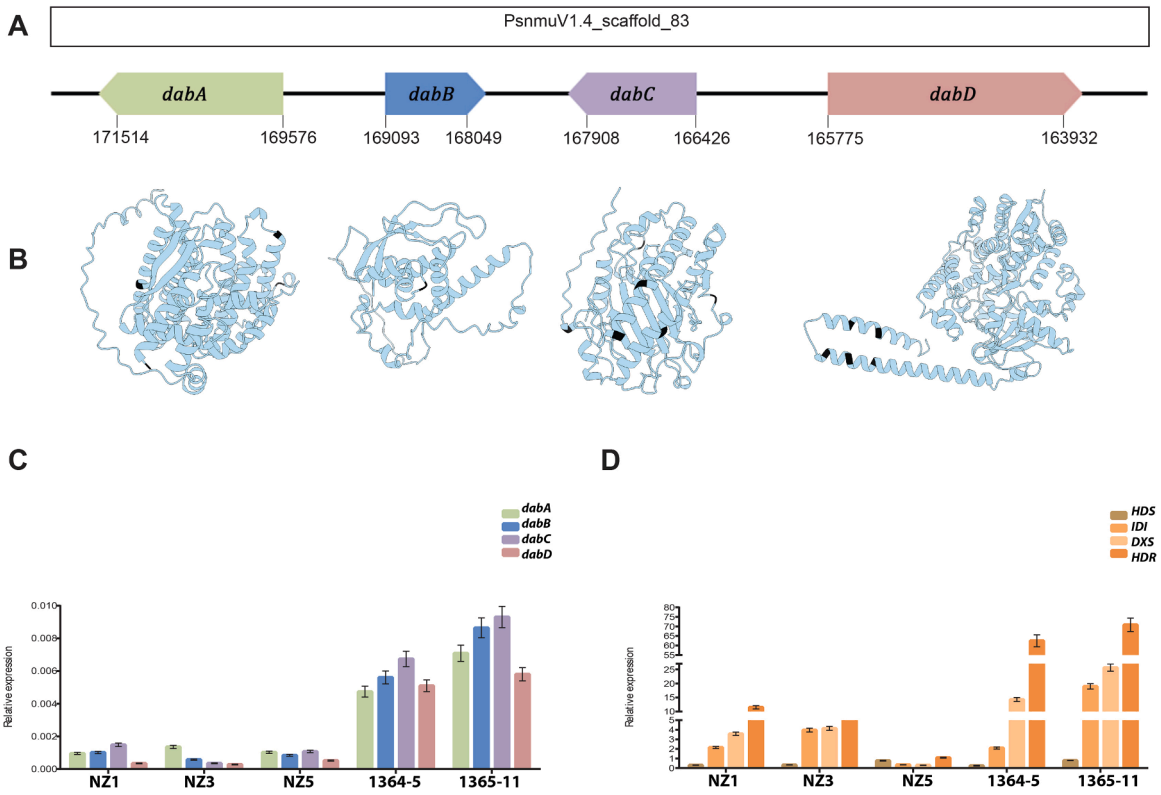


Fig. 2. **A**. DA biosynthesis (dab) gene cluster graphic representation. **B**. 3D predicted structures of the DAB proteins, amino acid variations found in the 29 strains are represented at their position in each model (in black), for details see Figs S2-S5. **C**. Relative gene expression of *dab* genes in DA-producing (1364-5 and 1365-11) and non-producing strains (NZ1, NZ3 and NZ5). **D**. Relative gene expression of isoprenoid precursor genes in DA-producing (1364-5 and 1365-11) and non-producing strains (NZ1, NZ3 and NZ5). Gene expression levels were normalised with respect to the internal control *TUB-A* gene. Error bars represent the standard deviation of three biological replicates.

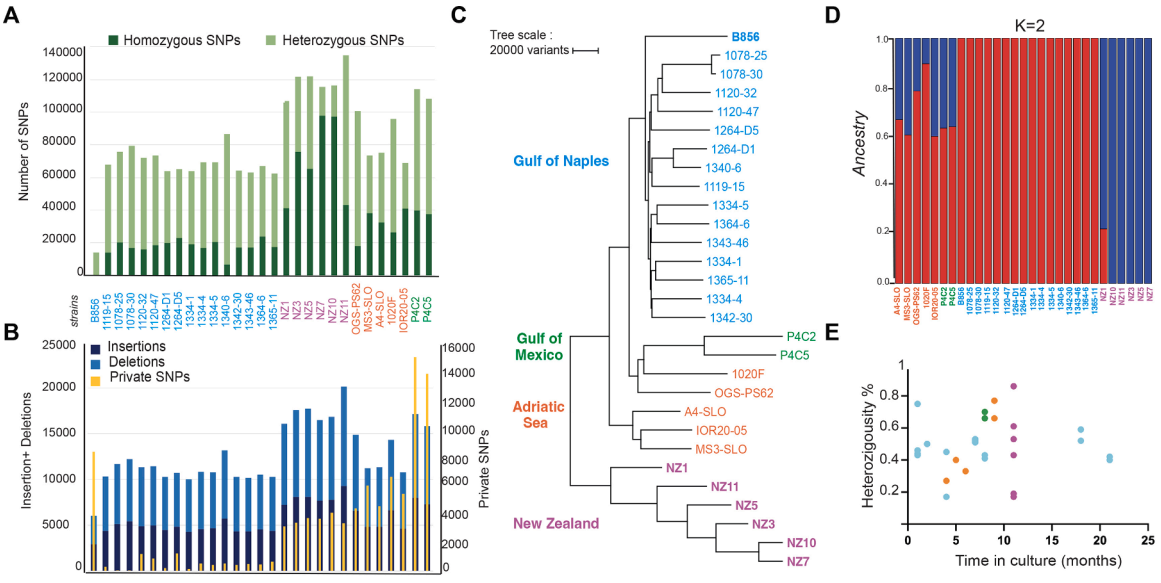


Fig. 3. Variant profile of the different strains and population structure. **A** The number of homozygous (dark green bars) and heterozygous (light green bars) SNPs among the set of high-quality variants. **B** The number of insertions (dark blue bars), deletions (light blue bars) and private SNPs (yellow bars), i.e. variants that occur only in that specific strain. **C**. Distance between the 29 *P multistriata* strains based on the high confidence set of variant sites, shown as a clustering tree. Scale bar = 20,000 variants, the scale represents dissimilarity for the specified length of a branch, the sum of all branch lengths between each two strains is the dissimilarity between these two strains. Different colors are given to strains from the different geographic locations. **D**. Admixture analysis, shown for K = 2 and with genetic clusters indicated by the two colours. **E**. Genome-wide heterozygosity level on the y-axis related to the time, in months, between strain isolation and DNA extraction, on the x-axis. Color code as in C.

to the two with low DA production (Figure S6).

3.3. Genetic divergence across space and time

The *P. multistriata* reference genome was assembled from strain B856 using five different Illumina sequencing datasets (Basu et al. 2017). In order to include the reference strain B856 in the variant analysis, we mapped part of the original Illumina data against the reference assembly, leading to a total number of 29 strains for the variant calling (Table 1). We obtained a mapping rate of at least 60 % for 24 of the 29 strains, while values were lower than 31 % for three strains from the Gulf of Naples and two from the Gulf of Mexico (details in Table 1 and Note S2). Unmapped reads were mainly represented by bacterial sequences, and to a minor extent by sequences of the organelles, not included in the reference assembly (Table S3).

Filtering resulted in a high quality, reliable set of 385 029 variants across the 29 genomes with a read depth between 23 and 137. More SNPs and Indels were found in New Zealand and Gulf of Mexico strains than in the other two locations (Fig. 3A) and, as expected, fewest variants in the reference strain. New Zealand strains showed a high variability in the number of heterozygous SNPs. The number of strain private SNPs tended to be higher in locations with a small sample size (Fig. 3B).

A clustering tree based on the variant sites detected across the 29 strains showed that strains clustered mostly according to their geographic distance, with the reference B856 expectedly closest to the Gulf of Naples strains (Fig. 3C). Note that the sum of all branch lengths between each two strains is the dissimilarity between these two strains, therefore New Zealand and the Mexican strains are among those more distant from the rest.

After further stringent filtering to fit the model's assumptions, a total of 27 323 SNPs were used for ADMIXTURE analysis. The cross-validation results indicated the occurrence of two distinct genetic clusters, corresponding to New Zealand and the other strains (Fig. 3D), in line with the clustering tree (Fig. 3C). The Gulf of Naples did not show temporal structuring, as indicated by a PCoA analysis of samples isolated between the 2013 and 2020, with the exception of the two clonal strains (1078-25, 1078-30) from the 2013 bloom (Fig. 4).

The Gulf of Naples did not show evidence of admixture with the Adriatic and Mexican population, although it was not clearly supported as a distinct genetic group. This result remained consistent after removing the New Zealand strains, which led to the identification of a single cluster in the remaining dataset (Fig. S7).

The pairwise comparison between the four groups of strains from the same geographic area showed that genetic distance (based on the variant sites) increased with increasing geographic distance (Fig. S8), verified by a Pearson's test showing highly significant correlation (correlation coefficient 0.99, p-value 0.0001) (Fig. S8). On a regional scale, among the strains sampled around New Zealand, no significant correlation (p-value 0.69) was found (Fig. S8C). A pairwise distance matrix applied between all 29 strains showed that the two genetically closest strains were expectedly those isolated from the clonal bloom in 2013 (Ruggiero et al. 2018) (Fig. S9). Calculating from the distance matrix (Fig. S9), the lowest average genetic distance for all pairs of strains in one location was found in the Gulf of Naples (average distance 73 452), the highest (112 041) within the Adriatic Sea, while the average distance was 95 515 within New Zealand.

The range of genome-wide heterozygosity level was broad in the Adriatic Sea and especially in New Zealand strains (Table S4), which showed both the highest and the lowest values. A much narrower range was detected in the Gulf of Naples, while the two strains from the Gulf of Mexico showed both a relatively high heterozygosity level (Fig. 2E, Table S4). The time spent in culture since the date of strain isolation has been suggested to cause ample changes in the genome structure of model diatom *P. tricornutum* (Bulankova et al. 2021). In our study, months of cultivation before DNA collection did not seem to correlate with heterozygosity nor to influence clustering (Fig. 2E).

3.4. Heterozygosity, genetic diversity and dispensable genes

We were interested in exploring major genomic changes and variations that could be distinctive of the different geographic groups, with a specific interest in the genomes of the non-toxic strains and in the two clonal strains isolated from the clonal bloom from 2013.

To determine if the two dominant clonal strains showed localised enrichment in heterozygosity in specific genomic regions compared to other strains, we determined scaffolds with diverging (higher/lower) heterozygosity level from other strains for different subgroups, including comparing the two dominant strains to the other strains from the Gulf of Naples (Fig. S10). No specific enrichment of regions with diverging heterozygosity level was found for the two dominant clonal strains when comparing to other combinations of strains (Table S5).

The comparison of multiple strains for a species allows to define core genes that are necessary to allow survival, and dispensable genes that

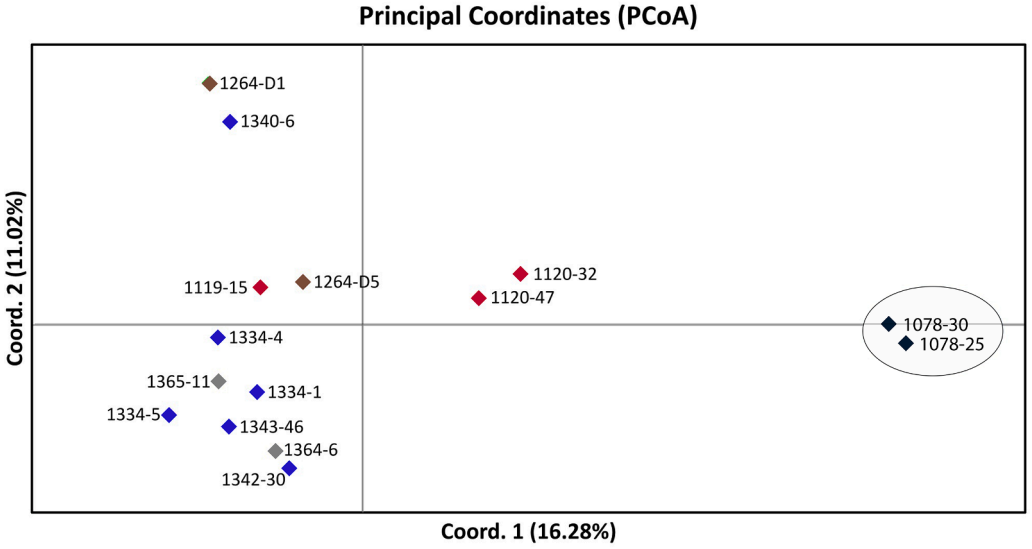


Fig. 4. Principal coordinates analysis for the 15 strains from the Gulf of Naples based on the high confidence set of variant sites. The two strains isolated from the clonal bloom of 2013 are circled and in black, other colours indicate samples from 2014, in red, from 2017, in brown, from 2019, in blue, and from 2020, in grey.

are not necessary and can therefore be lost in some of the strains. We used two approaches to define dispensable genes in *P. multistriata*. In one case, dispensable genes were determined based on a horizontal exon coverage of <5 % in at least one of the strains (Osuna-Cruz et al. 2020). A significant negative correlation (Pearson's correlation coefficient = -0.38, p-value = 0.04) between the number of lost genes and the number of mapped reads was found (Fig. S11), indicating a technical bias in this approach. As an alternative, genes with a homozygous high impact mutation, rendering the gene non-functional in at least one of the 29 strains, were considered as dispensable. A total of 1 257 genes were found with this criterion (Table S6). The dispensable genes were slightly enriched in non-annotated genes (about 44 % in comparison to 38 % genes without functional annotation among the reference genes). GO enrichment analysis indicated 14 significantly overrepresented categories in the dispensable genes set (Table S7). Like in *S. robusta* (Osuna-Cruz et al. 2020), we found the occurrence of several transporter proteins among the dispensable genes. Repeating the analyses for individual groups, we found that the highest percentage of unique dispensable genes was found in the New Zealand strains (~45 %) compared to the Gulf of Naples (23 %) and the Adriatic Sea (20 %). We then assessed the πN/πS as a measure of the level of selective pressure acting on the species. The πN/πS ratio calculates the rate of substitutions at nonsynonymous (missense) sites to the rate of substitutions at synonymous (silent) sites. For the 6 685 callable genes (~56 % of the ~12 000 reference genes, see Materials and Methods) across all 29 strains, *P. multistriata* was found to have a non-synonymous pairwise diversity (πN) value of 0.012, a synonymous diversity (πS) value of 0.035 and a πN/πS ratio of 0.34 (Table 2).

A total of 194 genes (2.9 % of the callable genes) with πN/πS > 1 were found. Like dispensable genes, they were found to be enriched in genes without annotation (56 % in comparison to 36 % among the callable reference genes) but no overrepresented GO category with more than one gene was found among the annotated genes. When comparing different species instead of strains of the same species, Ka/Ks is used, which measures the ratio of the number of nonsynonymous substitutions per non-synonymous site (Ka), in a given period of time, to the number of synonymous substitutions per synonymous site (Ks), in the same period. In a previous study, genes under potential selective pressure were determined measuring the Ka/Ks ratio for orthologous genes between *P. multistriata* and *P. multiseries*, finding 131 genes to have a Ka/Ks ratio > 1 (Basu et al. 2017). Among the 3590 reference genes determined as callable in both studies, nine overlapped between the sets of genes, with a highly significant overlap indicated by a Fisher's exact test (p-value = 1.6^{-6}) (Table S8). Four out of these nine genes have no annotation and the others cannot be linked to specific functions.

We tested if nucleotide diversity varied among different populations of *P. multistriata.* This was also useful to discern whether the higher πN/πS diversity in *P. multistriata* compared to other single-celled algal species (Table S9) might be caused by the higher number of strains considered in our study. πN/πS was found to be comparable when testing subsets of strains (Table 3), moreover among the New Zealand strains, the highest πN/πS ratio as well as the highest percentage of genes with πN/πS > 1 (Table 3, Table S10) were found.

Table 3
Nucleotide diversity, πN/πS ratio and the number of genes under selection for different *P. multistriata* groups of strains.

P. multistriata subset	πN	πS	πN/πS	Genes with πN/πS > 1
All 29 strains	0.012	0.035	0.34	194 (2.9 % of 6685 callable genes)
GoN[a], 14 strains	0.008	0.023	0.35	269 (2.6 % of 10,295 callable genes)
NZ, 6 strains	0.010	0.026	0.38	356 (3.2 % of 11,069 callable genes)
AS, 5 strains	0.010	0.027	0.37	316 (2.9 % of 10,888 callable genes)

[a] excluding 1078-30 to include only one of the two highly identical strains from the bloom in 2013

Abbreviations: GoN = Gulf of Naples, NZ = New Zealand, AS = Adriatic Sea.

Loci under selection are a key for the genetic variation that influences fitness in different environments. Therefore, we did a genome scan for outlier candidates based on differences in allele frequencies between the four groups of strains from the different geographic areas to identify regions of the genome under purifying or diversifying selection. By analyzing the Bayscan Fixation index values (Fst), a total of 49 outliers were detected (Fig. 5), located in or very close to 30 different genes (Table S11).

About 60 % (29) of the outliers were distributed on only three scaffolds (Fig. 5), all but one being under positive diversifying selection as indicated by the positive alpha value, comprising three larger variable regions harboring 10 genes (Table S11). The remaining 20 outliers were scattered along different scaffolds. When comparing the set of 30 genes on and around the 49 outliers a significant overlap (nine genes) was found with the 1257 dispensable genes among all strains (Fisher's exact test with a p-value of 0.003) (Table S12).

No significant candidate outlier sites between the dominant strains from the 2013 clonal expansion event and the other 13 strains collected in the Gulf of Naples could be found.

It is interesting to note that the list of genes with πN/πS > 1 in the Adriatic strains (Table S10) contains the mating type determining gene *MRP3*. This gene is expressed only in MT+ strains, from only one of the two alleles, and is required for the specification of the MT+ identity (Russo et al., 2018a). The alignment of the *MRP3* protein product obtained from the coding sequence extracted from each genome indeed showed more amino acid changes in the Adriatic strains (Fig. S12). Changes in the gene sequence include a nucleotide change generating a stop codon, leading to a truncated protein product in the MT- strains OGS-PS62 and A4-SLO, and small deletions in the A4-SLO and in the NZ5 strains causing a frameshift and leading to a truncated protein product (Fig. S13).

4. Discussion

The exploration of differences in the genomic structure of microalgal populations isolated from distinct geographic regions is still in its infancy. Here we explored the extent of similarity in the genomic features of *P. multistriata* strains from multiple locations around the world, revealing i) overall high levels of variation in this species, compared to other phytoplankton taxa, ii) a clear differentiation of strains sampled in New Zealand, the most distant geographic site, and iii) a weaker population structuring when excluding these from the analysis. The New Zealand strains were reported to be non-toxic, offering the opportunity to address the pending question of whether the genetic potential for toxin biosynthesis is present or not in non-toxic cells. Moreover, genomes of strains sampled from a rare clonal bloom detected in the Gulf of Naples in 2013 did not show signs of loss of heterozygosity nor of major rearrangements which could explain their sudden dominance.

4.1. *Dab* genes sequence variation and expression levels in different strains

Not all *Pseudo-nitzschia* species can produce DA, and in the species reported to be toxic, such as *P. multistriata,* non-toxic strains can also be found (Amato et al. 2010). In the set used for this study, either no DA production or only traces were reported for the New Zealand strains (Nishimura et al. 2021). The mechanism underlying this difference is still unknown. The *dab* gene cluster has been identified as responsible for toxin production (Brunson et al. 2018). The comparison of the *dab* genes sequences among the 29 strains did not show any gene-interrupting mutations that could explain lack of DA production in the New Zealand strains. Instead, we found that the *dab* genes expression levels were

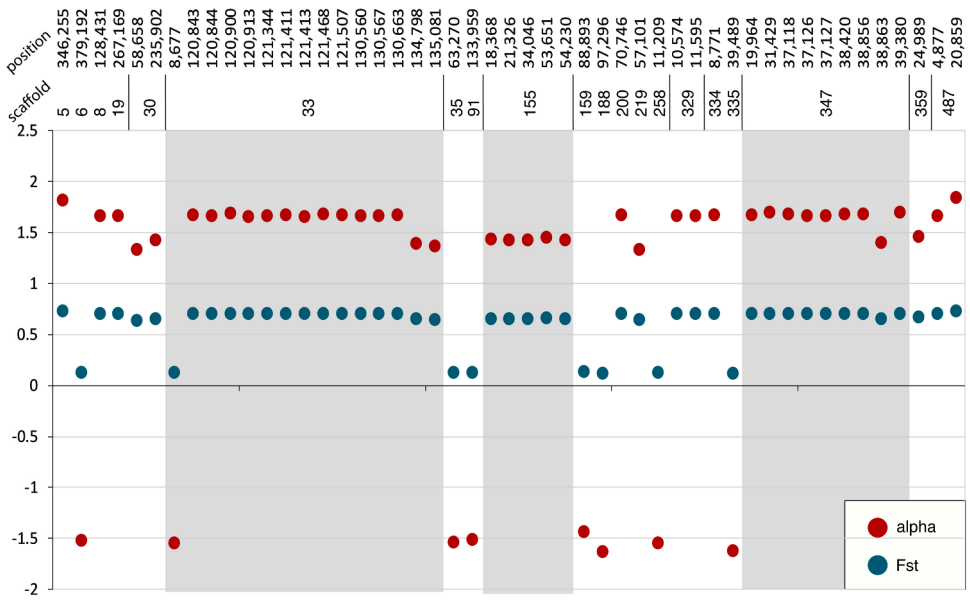


Fig. 5. Outlier sites determined among all strains (excluding the reference) with their alpha and Fst values. In the top part of the figure the location (scaffold and basepair position in the respective scaffold) is reported. Three scaffolds have more than two outlier sites, indicating larger regions under positive selective pressure as suggested by the positive alpha values (except for a negative alpha in one outlier on scaffold 33 far away from the other sites) in these outlier sites.

lower in the non-toxic strains compared to the DA producing strains 1364-5 and 1365-11, suggesting that the *dab* genes expression is modulated by still undescribed regulatory mechanisms. A tendency to have higher *dab* genes expression in high DA producing strains versus low DA producing strains was also observed in another set of Neapolitan strains, although we note that in general the *dab* genes expression levels are never very high when compared to other *P. multistriata* genes (see Figure S6 as example). The methyl-erythrito phosphate (MEP) route for isoprenoids has been shown to be implicated in DA levels increase under condition of phosphate limitation and elevated pCO_2 in *P. multiseries* (Brunson et al. 2018), and upon grazer exposure in *P. seriata* (Harðardóttir et al. 2019). Although physiological metabolic differences among strains could determine differences in the expression of MEP genes, required for the synthesis of many other compounds, we note that in the *P. multistriata* strains examined, the MEP gene *HDR* showed a remarkable difference in the expression levels between the two groups, coherently with the fact that it is one of the isoprenoid genes induced in all three conditions mentioned above (Harðardóttir et al. 2019). *DXS*, encoding the first enzyme in the MEP pathway, also displayed higher levels of expression in the DA producing strains, while the *HDS* and *IDI* genes had comparable levels among strains. Our data demonstrate that *P. multistriata* strains for which DA production was undetectable have the genetic potential to produce DA. Whether this can be extrapolated in general to non-toxic *Pseudo-nitzschia* species remains to be assessed, but the hypothesis of *dab* genes transcriptional repression would be coherent with the finding that the putatively non-toxic species *P. obtusa* unexpectedly began to produce DA when exposed to copepod cues (Harðardóttir et al. 2019; Tammilehto et al. 2015). The production of *Pseudo-nitzschia* genomes from toxic and non-toxic species will confirm presence/absence and sequence variation of the *dab* cluster. We note a very recent study on the topic (He et al., 2024). Additional studies on gene expression regulation are needed to refine the regulatory mechanisms, including experiments where DA production in multiple toxic and non-toxic *P. multistriata* strains is stimulated by external cues, such as copepodamides from copepod grazers (Grebner et al. 2019), or by specific growth conditions. Combined with studies aimed at following toxin levels and gene expression *in situ*, results of such studies will contribute to a better understanding of toxic *Pseudo-nitzschia* blooms and will be instrumental for the implementation of bioassays needed for improved monitoring and forecasting of harmful events (Brunson et al. 2024). As far as questions related to the *dab* gene cluster evolution, additional sequence data from this and other *Pseudo-nitzschia* species will allow to explore evolutionary rates and selection forces acting on the four genes.

4.2. Genetic structuring in *P. multistriata*

Variant analyses showed that the sampling location has a bigger influence on the genomic differences between strains than the sampling date or culture age.

Past studies of population genetic structure in diatoms relied on microsatellites, neutral markers with high mutation rates. With this tool, genetic structures were recorded at various spatial scales for planktonic diatom species, *e.g.*, *Pseudo-nitzschia pungens* (Casteleyn et al. 2010) and *Skeletonema marinoi* (Sjöqvist et al. 2015). More recent analyses explored genetic variation and the basis of adaptation to different environmental conditions with genome-wide approaches coupled with ecophysiological tests. Populations of the pelagic diatom *Fragilariopsis kerguelensis* sampled across the Southern Polar Front showed the presence of two ecotypes apparently adapted to the different environmental conditions above and below the Polar Front and a third ecotype with a broader distribution. Results suggested adaptive expansion and an incipient speciation of the northern and southern ecotypes supported by reproductive isolation (Postel et al. 2020). The combination of high-resolution population genomics and transcriptomics data showed evidence that the salinity gradient of the Baltic Sea is a major driver of adaptive change in the diatom *Skeletonema marinoi*, but also highlighted the complex role of additional metabolic processes in shaping population structure in this highly variable environment (Pinseel et al. 2023).

Chaetoceros genetic diversity has been assessed using metagenome assembled genomes (MAGs). Single-nucleotide variants identified within the different MAG populations indicated the presence of genetic structure in some Arctic Ocean populations, and distinct abiotic factors shaped populations more than geographic distance (Nef et al. 2022), supporting previous results in *Thalassiosira rotula*, indicating that distance and time can have equally strong influences on genetic structuring, but that environmental selection can play an important role in maintaining genetic divergence (Whittaker and Rynearson 2017).

In *P. multistriata*, different analyses supported the differentiation between the samples collected in New Zealand compared to the other samples. Without taking into account currents, mixing between

populations from the different locations would require overcoming distances of about 25 000 km for the most distant sites, suggesting that the gene flow between the different populations could be limited. The Neapolitan strains showed lower intra-population diversity despite strains having been isolated along a time span of several years, than the ones from the Adriatic Sea and New Zealand, possibly due to the more dispersed sampling sites accompanied by wider range of environmental conditions in the latter two. Reference strain B856 was a lab strain obtained by crossing *P. multistriata* strains from 2009 (Tesson et al. 2014) and thus expectedly has the greatest distance from the other Gulf of Naples strains.

Previous studies based on microsatellite data identified four distinct genetic clusters in the Gulf of Naples using a total of over 1100 strains isolated during the years 2008 to 2020. These populations appeared over different time intervals, one was dominant in 2008–2010, three appeared in 2013–2014, one of which became dominant in 2017–2020 (Ruggiero et al. 2024). Our genome-wide data including strains from 2013, 2014, 2017, 2019 and 2020 are coherent with this separation, with most of the Neapolitan strains from those last four years clustering together in a PCoA analysis, and a clear separation of only the two clonal strains from 2013. The Admixture analysis, conducted with stringent filtering to fit the model predictions, did not pick up this separation, but strongly supported a different genetic makeup in the New Zealand strains. If the New Zealand strains are excluded from the analysis, ADMIXTURE still supports the inclusion of all other strains in one population. As for the reason why the Gulf of Mexico strains appear close to some of the Adriatic strains, we excluded major technical issues, we note that many invasive species arrive in the North Adriatic Sea thanks to ballast waters (Gollasch et al. 2019) but suggest that the sample size should be increased before we can draw any definite conclusions on these observations. The Adriatic samples showed high diversity, with the greatest average genetic distance. A possible explanation lies in the fact that the North Adriatic Sea, despite being a relatively small basin, is exposed to a strongly fluctuating nutrient concentrations due to the Po River freshwater input, resulting in very dynamic and variable environmental conditions (Zavatarelli et al. 1998). In addition to individual strain variability (Lema et al. 2019; Kremp et al. 2012), culture age is another factor to be considered when assessing ecophysiological responses or microevolution dynamics, for instance time spent in culture explained part of the transcriptomic differences in *Leptocylindrus aporus* strains exposed to different temperature (Pargana et al. 2020). For our study, strains were collected after a variable number of months from initial isolation, spanning from one to over 20 months for some of the Neapolitan strains, and we note that there is no specific effect of the culture age on clustering based on variants.

Finally, mapping rate for the strains in this study was generally around 70 %, lower than 94–99 % reported for 17 *P. tricornutum* strains (Chaumier et al. 2024), and more similar to values found for the haptophyte *Emiliania huxleyi* (now *Gephyrocapsa*), which were reported around 95–65 % (Read et al. 2013) and 25–69 % (Bendif et al. 2023). Bacterial sequences appeared to be the major source for unmapped reads, suggesting the presence of bacterial strains resistant to the antibiotic cocktail used to treat cultures. We note that it is not possible to maintain *P. multistriata* in axenic conditions for longer than a few weeks, due either to toxicity of the antibiotics or perhaps to the need of *P. multistriata* cells for the presence of bacterial compounds.

4.3. Genomic features in the different groups

In the diatom *Thalassiosira pseudonana* regional loss of heterozygosity was found in isolates from several locations (Koester et al. 2018). Our exploration of the genome in search of regions with distinctive features, such as areas with loss of heterozygosity, could not point to any specific event, although these analyses will be more sensitive when a more refined genome with improved contiguity is available.

In line with what mentioned above, we could not find any specific genomic feature associated with cultivation time. The few studies in which the genome structure was studied in strains kept in clonal culture for variable periods of time provide contrasting results. Genome-wide data for 10 *P. tricornutum* strains, where the sampling time spanned almost one century and the strains were kept in culture for extended times, showed temporal stability (Rastogi et al. 2020). A more recent analysis with seven additional *P. tricornutum* strains, for a total of 17 strains, failed to establish a clear link between age of the strain and the extent of genetic diversity (Chaumier et al. 2024). As in our case, the loss of heterozygosity in cultured strains over time was not supported. However, in a comparison between a wild type and a mutant *P. tricornutum* strain separated for over one year of cultivation, major loss of heterozygosity events were observed (Russo et al. 2018b). Another report indicated that *P. tricornutum* clonal cultures rapidly accumulate genomic diversity, with increased copy number variation and loss of heterozygosity (Bulankova et al. 2021). Dedicated experiments, with repetitive sampling of selected strains along longer cultivation times, would add more on genome stability in *P. multistriata*.

Non-synonymous to synonymous substitution ratio was determined for orthologous genes between *P. multistriata* and *P. multiseries* in our previous study, in which 131 genes had a Ka/Ks ratio > 1 (Basu et al. 2017). Here we focused on finding genes with a high rate of non-synonymous polymorphisms within the dataset of *P. multistriata* strains. The comparison of two species considers mainly substitutions (mutations) that got fixed in the genome. Within-species comparison regards polymorphisms that will not necessarily get fixed (Rocha et al., 2006; Kryazhimskiy and Plotkin, 2008) and allows insights into population dynamics on a short time-scale. The significant overlap of genes under selection found in the interspecific analysis of (Basu et al. 2017) and the genes under selection found in our intraspecific analysis indicates that looking at a “snapshot” view of nucleotide diversity can give hints also on long-term evolutionary aspects of a species. The availability of both kinds of analyses indicated that the set of constraint genes was rather stable in *P. multistriata*, considering the significant overlap of genes under selection between both studies.

The relatively high πN and πS values for *P. multistriata* among other diatoms and single-celled marine organisms (Table S10) was interpreted as indication for a small effective population size and/or a low sexual reproduction rate since a high recombination rate is linked to stronger purifying selection (Rastogi et al. 2020; Osuna-Cruz et al. 2020; Cavassim et al. 2021).

Genes tend to be less constrained and consequently more variable if they are non-essential for the survival of a species (Jordan et al. 2002; Osuna-Cruz et al. 2020). Dispensable genes are therefore more frequently found to be evolving than essential core genes, presumably to increase the possibility for beneficial new characteristics to be developed. In the *P. multistriata* strains analysed in this study, a few dispensable genes were significantly overlapping with genes in which candidate sites under diversifying selection were located.

Overall, the New Zealand strains showed highest variability in heterozygosity levels, the highest $\pi N/\pi S$ ratio, the highest number of genes with $\pi N/\pi S > 1$ as well as the highest number of unique dispensable genes, and together with the results of the ADMIXTURE analysis, appear to be distinct from the other samples obtained for this study. This probably is due to the geographic distance limiting the genetic flow, and while there are differences in the specific environmental conditions, the four sites selected were all from comparable coastal sites and from temperate areas. We expect that part of the adaptive response of these strains relies on gene expression modulation. Future metagenomics and metetranscriptomics studies will help in confirming this idea and in providing more information on the role of the putative genes under selection identified in this study.

Interesting results have been obtained with the coccolithophore *E. huxleyi* (now *Gephyrocapsa huxleyi*), where a study based on the analysis of the genome of 17 isolates first showed that all strains shared a core-genome but also harbour considerable genome variability (Read

et al. 2013). In this case the variability was supported by the presence of different metabolic repertoires and a further study led to the description of three distinct species for which authors could also depict the *tempo* of speciation. In the case of *P. multistriata*, we could identify regions that appear under selection, but not relevant differences at the level of genes or gene families. Nevertheless, having a statistically significant sample size of sequenced genomes for the different geographic regions, analytical tools are available to address relevant questions for understanding the mechanisms that drive diversification of unicellular microalgae.

4.4. Sequence variation in the MT determining gene

P. multistriata is the only diatom for which a sex determination mechanism has been described. The *MRP3* gene responsible for MT+ specification is expressed from only one of the two alleles in MT+, and is not expressed in the MT- (Russo et al., 2018a). Resequencing strains of both MTs provided new opportunities to add information on the sex determination system, another key trait which is still vastly unexplored in diatoms. While long reads sequencing will be needed to resolve the putative 34 kb MT locus structure in opposite MTs, availability of the new sequencing data provided new insight into the *MRP3* sequence variation. We observed three cases (Figure S11) in which non expressed alleles accumulated mutations that would translate into mutated protein products. Moreover, *MRP3* appeared to be enriched in non-synonymous changes in the Adriatic population. Without data on sex determination mechanisms from any other diatom species, comparisons are currently not possible but the lack of conservation of *MRP3* outside of the *Pseudo-nitzschia* and *Fragilariopsis* genera (Russo et al., 2018a) suggests that more than one solution for determining the MT has evolved in diatoms. Therefore, the *P. multistriata* MT locus is predicted to be young and evolving. Our results, showing accumulation of mutations on alleles where the gene is inactive, are coherent with this hypothesis. Indeed, in systems with genetic sex determination, one of the two sex chromosomes often degenerates over time, losing gene function and accumulating repetitive elements, due to the lack of genetic exchange between the regions involved in sex determination (Bachtrog et al. 2014). We hypothesise that this is happening in the *P. multistriata* MT determining region.

5. Conclusions

Whole-genome resequencing of several *P. multistriata* strains sampled over many years in various sampling sites allowed a simultaneous investigation of the spatial and temporal effect on genetic diversity in diatoms, indicating spatial structure in the four groups and no temporal structure in the Gulf of Naples over a five years interval. Two dominant strains from the 2013 clonal bloom were expectedly separated but did not show specific genomic features, such as LOH events or outlier sites differing in the allele frequency, suggesting that the explanation for their sudden increase does not lie in major changes in their genome structure, at least at the resolution level currently possible with existing resources. Plasticity at the regulatory level is a likely alternative explanation. The investigated SNP dataset creates an important baseline for projects following phytoplankton at sea over time, like the Naples Ecological Research Augmented Observatory (NEREA, (Campese et al. 2024)), as it will allow to follow the fate of *P. multistriata* in the Gulf of Naples and add observations of the evolution and long-term dynamics of this species. The rare availability of non-toxic strains allowed to show that *dab* genes are present and potentially functional, but are expressed at very low levels when compared to toxic strains. Dedicated experiments will be needed to explore the mechanisms for *dab* gene expression modulation. Mutations accumulating in the non-transcribed copies of the sex determination gene *MRP3* are in line with what is expected for evolving sex determination regions and calls for additional exploration of the *Pseudo-nitzschia* genus to define emergence, evolution and conservation of the MT determination system found in *P. multistriata*. Overall, our findings will enable deeper investigations of adaptation and genetic regulatory mechanisms in a toxic species.

CRediT authorship contribution statement

Svenja Mager: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation. **Francesco Manfellotto:** Writing – review & editing, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Antonella Ruggiero:** Writing – review & editing, Visualization, Validation, Methodology, Investigation, Formal analysis. **Viviana Di Tuccio:** Methodology, Investigation. **Federica Cerino:** Resources. **Stefano Accoroni:** Resources. **Tomohiro Nishimura:** Resources. **Marta Mikhno:** Methodology, Investigation. **Neri Fattorini:** Investigation, Methodology. **Timotej Turk Dermastia:** Resources. **Jasna Arapov:** Resources. **Sanda Skejic:** Resources. **Lesley Rhodes:** Resources. **Kirsty Smith:** Resources. **Antonio Longo:** Visualization, Formal analysis. **Caterina Manzari:** Methodology, Investigation. **Lisa Campbell:** Resources. **Graziano Pesole:** Methodology, Investigation. **Remo Sanges:** Supervision, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Francesca Raffini:** Writing – review & editing, Visualization, Software, Investigation, Formal analysis. **Maria Valeria Ruggiero:** Visualization, Methodology, Investigation, Formal analysis. **Monia Teresa Russo:** Writing – review & editing, Methodology, Investigation, Formal analysis, Data curation. **Marina Montresor:** Writing – review & editing, Writing – original draft, Supervision, Resources, Investigation, Data curation, Conceptualization. **Maria Immacolata Ferrante:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors wish to thank Dr Valeria Di Dato, Dr. Monica Cangini, Dr. Vittorio Soprano for the DA measurements, and Prof Dmitry Filatov for helpful discussion on the manuscript. We acknowledge support from the SZN Sequencing and molecular analyses center and the EMBL Gene Core facility. We thank Ms Carmen Minucci, Mr Ferdinando Tramontano, Ms Sarah Challenger and Ms Juliette Butler for their assistance with culture management. This work was supported by the Gordon and Betty Moore Foundation through grant no 7978 to MIF, and by the MoBiLab - Lifewatch - ITA initiative. This study was also partially supported by the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.4—Call for tender No. 3138 of 16 December 2021, rectified by Decree n.3175 of 18 December 2021 of the Italian Ministry of University and Research funded by the European Union – NextGenerationEU; CN_00000033, Concession Decree No. 1034 of 17 June 2022 adopted by the Italian Ministry of University and Research, CUP C63C22000520001, “National Biodiversity Future Center - NBFC”. The Long-Term Ecological Research site MareChiara program is funded by Stazione Zoologica Anton Dohrn, Naples (Italy).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.hal.2024.102791.

Data availability

Sequencing data have been deposited under accession number PRJNA882609. Other data, material and information are available upon request.

References

Adelfi, M.G., Borra, M., Sanges, R., Montresor, M., Fontana, A., Ferrante, M.I., 2014. Selection and validation of reference genes for qPCR analysis in the pennate diatoms Pseudo-nitzschia multistriata and P. Arenysensis. J. Exp. Mar. Biol. Ecol. 451, 74–81. https://doi.org/10.1016/j.jembe.2013.11.003. February.

Alexa. 2019. 'Alexa A, Rahnenfuhrer J (2024). topGO: enrichment analysis for gene ontology. R Package Version 2.58.0'.

Amato, A., Lüdeking, A., Kooistra, W.H.C.F., 2010. Intracellular domoic acid production in Pseudo-nitzschia multistriata isolated from the Gulf of Naples (Tyrrhenian Sea, Italy). Toxicon. 55 (1), 157–161. https://doi.org/10.1016/j.toxicon.2009.07.005.

Annunziata, R., Hay Mele, B., Marotta, P., Volpe, M., Entrambasaguas, L., Mager, S., Stec, K., et al., 2022. Trade-off between sex and growth in diatoms: molecular mechanisms and demographic implications. Sci. Adv. 8 (3), eabj9466. https://doi.org/10.1126/sciadv.abj9466.

Bachtrog, D., Mank, J.E., Peichel, C.L., Kirkpatrick, M., Otto, S.P., Ashman, T.-L., Hahn, M.W., et al., 2014. Sex determination: why so many ways of doing it? PLoS Biol. 12 (7), e1001899. https://doi.org/10.1371/journal.pbio.1001899.

Basu, S., Patil, S., Mapleson, D., Russo, M.T., Vitale, L., Fevola, C., Maumus, F., et al., 2017. Finding a partner in the ocean: molecular and evolutionary bases of the response to sexual cues in a planktonic diatom. New Phytologist 215 (1), 140–156. https://doi.org/10.1111/nph.14557.

Bates, Stephen S., Katherine A.H., N. Lundholm, M. Montresor, and C. Pin Leaw. 2018. 'Pseudo-nitzschia, nitzschia, and domoic acid: new research since 2011'. *Harmful. Algae* 79:3–43. https://doi.org/10.1016/j.hal.2018.06.001.

Bendif, E.M., Probert, I., Archontikis, O.A., Young, J.R., Beaufort, L., Rickaby, R.E., Filatov, D., 2023. Rapid diversification underlying the global dominance of a cosmopolitan phytoplankton. ISME J. 17 (4), 630–640. https://doi.org/10.1038/s41396-023-01365-5.

Benoiston, A.-S., Ibarbalz, F.M., Bittner, L., Guidi, L., Jahn, O., Dutkiewicz, S., Bowler, C., 2017. The evolution of diatoms and their biogeochemical functions. Philos. Trans. R. Soc. B: Biol. Sci. 372 (1728), 20160397. https://doi.org/10.1098/rstb.2016.0397.

Bilcke, G., Ferrante, M.I., Montresor, M., Decker, S.De, Veylder, L.De, Vyverman, W., 2022. Life cycle regulation. In: Falciatore, A., Mock, T (Eds.), The Molecular Life of Diatoms. Springer, pp. 205–228.

Brunson, J.K., McKinnie, S.M.K., Chekan, J.R., McCrow, J.P., Miles, Z.D., Bertrand, E.M., Bielinski, V.A., et al., 2018. Biosynthesis of the neurotoxin domoic acid in a bloom-forming diatom. Science 361 (6409), 1356–1358. https://doi.org/10.1126/science.aau0382.

Brunson, J.K., Thukral, M., Ryan, J.P., Anderson, C.R., Kolody, B.C., James, C.C., Chavez, F.P., et al., 2024. Molecular forecasting of domoic acid during a pervasive toxic diatom bloom. Proc. Natl. Acad. Sci. 121 (40), e2319177121. https://doi.org/10.1073/pnas.2319177121.

Bulankova, P., Sekulić, M., Jallet, D., Nef, C., van Oosterhout, C., Delmont, T.O., Vercauteren, I., et al., 2021. Mitotic recombination between homologous chromosomes drives genomic diversity in diatoms. Curr. Biol. 31 (15), 3221–3232. https://doi.org/10.1016/j.cub.2021.05.013 e9.

Campese, L., Russo, L., Abagnale, M., Alberti, A., Bachi, G., Balestra, C., Bellardini, D., et al., 2024. The NEREA augmented observatory: an integrative approach to marine coastal ecology. Sci. Data 11 (1), 989. https://doi.org/10.1038/s41597-024-03787-y.

Casteleyn, G., Leliaert, F., Backeljau, T., Debeer, A.-E., Kotaki, Y., Rhodes, L., Lundholm, N., Sabbe, K., Vyverman, W., 2010. Limits to gene flow in a cosmopolitan marine planktonic diatom. Proc. Natl. Acad. Sci. 107 (29), 12952–12957. https://doi.org/10.1073/pnas.1001380107.

Cavassim, M.I.A., Andersen, S.U., Bataillon, T., Schierup, M.H., 2021. Recombination facilitates adaptive evolution in rhizobial soil bacteria. Mol. Biol. Evol. 38 (12), 5480–5490. https://doi.org/10.1093/molbev/msab247.

Chaumier, T., Yang, F., Manirakiza, E., Ait-Mohamed, O., Wu, Y., Chandola, U., Jesus, B., Piganeau, G., Groisillier, A., Tirichine, L., 2024. Genome-wide assessment of genetic diversity and transcript variations in 17 accessions of the model diatom phaeodactylum tricornutum. ISMe Commun. 4 (1), ycad008. https://doi.org/10.1093/ismeco/ycad008.

Chekan, J.R., McKinnie, S.M.K., Noel, J.P., Moore, B.S., 2020. Algal neurotoxin biosynthesis repurposes the terpene cyclase structural fold into an N-prenyltransferase. Proc. Natl. Acad. Sci. 117 (23), 12799–12805. https://doi.org/10.1073/pnas.2001325117.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: sNPs in the genome of drosophila melanogaster strain W1118; Iso-2; Iso-3′. Fly (Austin) 6 (2), 80–92. https://doi.org/10.4161/fly.19695.

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

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., et al., 2011. The variant call format and VCFtools. Bioinformatics. 27 (15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., et al., 2021. Twelve years of SAMtools and BCFtools. Gigascience 10 (2), giab008. https://doi.org/10.1093/gigascience/giab008.

Douglas, D.J., Ramsey, U.P., Walter, J.A., Wright, JL.C., 1992. Biosynthesis of the neurotoxin domoic acid by the marine diatom nitzschia pungens forma multiseries, determined with [13C]-labelled precursors and nuclear magnetic resonance. J. Chem. Soc., Chem. Commun. (9), 714–716. https://doi.org/10.1039/C39920000714. January.

Ferrante, M.I., Broccoli, A., Montresor, M., 2023. The pennate diatom Pseudo-nitzschia multistriata as a model for diatom life cycles, from the laboratory to the sea. J. Phycol. 59 (4), 637–643. https://doi.org/10.1111/jpy.13342.

Fischer, M.C., Foll, M., Excoffier, L., Heckel, G., 2011. Enhanced AFLP genome scans detect local adaptation in high-altitude populations of a small rodent (Microtus Arvalis). Mol. Ecol. 20 (7), 1450–1462. https://doi.org/10.1111/j.1365-294X.2011.05015.x.

Godhe, A., Sjöqvist, C., Sildever, S., Sefbom, J., Harðardóttir, S., Bertos-Fortis, M., Bunse, C., et al., 2016. Physical barriers and environmental gradients cause spatial and temporal genetic differentiation of an extensive algal bloom. J. Biogeogr. 43 (6), 1130–1142. https://doi.org/10.1111/jbi.12722.

Gollasch, S., Hewitt, C.L., Bailey, S., David, M., 2019. Introductions and transfers of species by ballast water in the adriatic sea. Mar. Pollut. Bull. 147, 8–15. https://doi.org/10.1016/j.marpolbul.2018.08.054. Vulnerability to bioinvasions: current status, risk assessment and management of ballast water through a regional approach – the Adriatic SeaOctober.

Grebner, W., Carina Berglund, E., Berggren, F., Eklund, J., Harðadóttir, S., Andersson, M. X., Selander, E., 2019. Induction of defensive traits in marine plankton—new copepodamide structures. Limnol. Oceanogr. 64 (2), 820–831. https://doi.org/10.1002/lno.11077.

Guillard, R.R.L., 1975. Culture of phytoplankton for feeding marine invertebrates. In: Smith, W.L., Chanley, M.H. (Eds.), Culture of Marine Invertebrate Animals. Springer US, pp. 29–60 edited by. http://link.springer.com/chapter/10.1007/978-1-4615-8714-9_3.

Harðardóttir, S., Wohlrab, S., Hjort, D.M., Krock, B., Nielsen, T.G., John, U., Lundholm, N., 2019. Transcriptomic responses to grazing reveal the metabolic pathway leading to the biosynthesis of domoic acid and highlight different defense strategies in diatoms. BMC Mol. Biol. 20 (1), 7. https://doi.org/10.1186/s12867-019-0124-0.

He, Z., Xu, Q., Chen, Y., Liu, S., Song, H., Wang, H., Leaw, C.P., Chen, N., 2024. Acquisition and evolution of the neurotoxin domoic acid biosynthesis gene cluster in Pseudo-nitzschia species. Commun. Biol. 7, 1–12. https://doi.org/10.1038/s42003-024-07068-7.

Jeffares, D.C., Tomiczek, B., Sojo, V., Reis, M., 2015. A beginners guide to estimating the non-synonymous to synonymous rate ratio of all protein-coding genes in a genome. In: Peacock, C (Ed.), Parasite Genomics Protocols. Springer, New York, NY, pp. 65–90. https://doi.org/10.1007/978-1-4939-1438-8_4 edited by.

Jombart, T., Ahmed, I., 2011. Adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics. 27 (21), 3070–3071. https://doi.org/10.1093/bioinformatics/btr521.

Jordan, I.K., Rogozin, I.B., Wolf, Y.I., Koonin, E.V., 2002. Essential genes are more evolutionarily conserved than are nonessential genes in bacteria'. Genome Res. 12 (6), 962–968. https://doi.org/10.1101/gr.87702.

Knaus, B.J., Grünwald, N.J., 2017. Vcfr: a package to manipulate and visualize variant call format data in R'. Mol. Ecol. Resour. 17 (1), 44–53. https://doi.org/10.1111/1755-0998.12549.

Koester, J.A., Berthiaume, C.T., Hiranuma, N., Parker, M.S., Iverson, V., Morales, R., Ruzzo, W.L., Virginia Armbrust, E., 2018. Sexual ancestors generated an obligate asexual and globally dispersed clone within the model diatom species thalassiosira pseudonana. Sci. Rep. 8 (1), 10492. https://doi.org/10.1038/s41598-018-28630-4.

Kremp, A., Godhe, A., Egardt, J., Dupont, S., Suikkanen, S., Casabianca, S., Penna, A., 2012. Intraspecific variability in the response of bloom-forming marine microalgae to changed climate conditions. Ecol. Evol. 2 (6), 1195–1207. https://doi.org/10.1002/ece3.245.

Kryazhimskiy, S., Plotkin, J.B., 2008. The Population Genetics of dN/dS. PLoS Genet 4. https://doi.org/10.1371/journal.pgen.1000304.

Lema, K.A., Metegnier, G., Quéré, J., Latimier, M., Youenou, A., Lambert, C., Fauchot, J., Gac, M.Le, 2019. Inter- and Intra-specific transcriptional and phenotypic responses of Pseudo-nitzschia under different nutrient conditions. Genome Biol. Evol. 11 (3), 731–747. https://doi.org/10.1093/gbe/evz030.

Mayr, E., Wheeler, Q.D., Meier, R., 2000. Species Concepts and Phylogenetic Theory, A Debate. Columbia University Press.

Meng, E.C., Goddard, T.D., Pettersen, E.F., Couch, G.S., Pearson, Z.J., Morris, J.H., Ferrin, T.E., 2023. UCSF ChimeraX: tools for structure building and analysis'. Protein Sci. 32 (11), e4792. https://doi.org/10.1002/pro.4792.

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., Steinegger, M., 2022. ColabFold: making protein folding accessible to all'. Nat. Methods 19 (6), 679–682. https://doi.org/10.1038/s41592-022-01488-1.

Nef, C., Madoui, M.-A., Pelletier, É., Bowler, C., 2022. Whole-genome scanning reveals environmental selection mechanisms that shape diversity in populations of the epipelagic diatom chaetoceros. PLoS Biol. 20 (11), e3001893. https://doi.org/10.1371/journal.pbio.3001893.

Nishimura, T., Murray, J.S., Boundy, M.J., Balci, M., Bowers, H.A., Smith, K.F., Harwood, D.T., Rhodes, L.L., 2021. Update of the planktonic diatom genus Pseudo-nitzschia in aotearoa New Zealand coastal waters: genetic diversity and toxin production. Toxins (Basel) 13 (9), 637. https://doi.org/10.3390/toxins13090637.

Orsini, L., Sarno, D., Procaccini, G., Poletti, R., Dahlmann, J., Montresor, M., 2002. Toxic Pseudo-nitzschia multistriata (Bacillariophyceae) from the Gulf of Naples:

morphology, toxin analysis and phylogenetic relationships with other Pseudo-nitzschia species. Eur. J. Phycol. 37 (2), 247–257. https://doi.org/10.1017/S0967026202003608.
Osuna-Cruz, C.M., Bilcke, G., Vancaester, E., Decker, S.De, Bones, A.M., Winge, P., Poulsen, N., et al., 2020. The seminavis robusta genome provides insights into the evolutionary adaptations of benthic diatoms'. Nat. Commun. 11 (1), 3320. https://doi.org/10.1038/s41467-020-17191-8.
Pargana, A., Musacchia, F., Sanges, R., Russo, M.T., Ferrante, M.I., Bowler, C., Zingone, A., 2020. Intraspecific diversity in the cold stress response of transposable elements in the diatom leptocylindrus aporus. Genes (Basel) 11 (1), 9. https://doi.org/10.3390/genes11010009.
Peakall, R., Smouse, P.E., 2006. Genalex 6: genetic analysis in excel. population genetic software for teaching and research. Mol. Ecol. Notes 6 (1), 288–295. https://doi.org/10.1111/j.1471-8286.2005.01155.x.
Petroff, R., Hendrix, A., Shum, S., Grant, K.S., Lefebvre, KA., Burbacher, T.M., 2021. Public health risks associated with chronic, low-level domoic acid exposure: a review of the evidence'. Pharmacol. Ther. 227, 107865. https://doi.org/10.1016/j.pharmthera.2021.107865. November.
Pinseel, E., E.C. Ruck, T. Nakov, P.R. Jonsson, O. Kourtchenko, A. Kremp, M.I.M. Pinder, et al. 2023. 'Local adaptation of a marine diatom is governed by genome-wide changes in diverse metabolic processes'. bioRxiv. https://doi.org/10.1101/2023.09.22.559080.
Pocklington, R., Milley, J.E., Bates, S.S., Bird, C.J., De Freitas, A.S.W., Quilliam, M.A., 1990. Trace determination of domoic acid in sea water and phytoplankton by high-performance liquid chromatography of the fluorenylmethoxycarbonyl (FMOC) derivative'. Int. J. Environ. Anal. Chem. 38 (3), 351–368. https://doi.org/10.1080/03067319008026940.
Poplin, R., V. Ruano-Rubio, M.A. DePristo, T.J. Fennell, M.O. Carneiro, G.A. Van der Auwera, D.E. Kling, et al. 2018. 'Scaling accurate genetic variant discovery to tens of thousands of samples'. bioRxiv. https://doi.org/10.1101/201178.
Postel, U., Glemser, B., Salazar Alekseyeva, K., Eggers, S.L., Groth, M., Glöckner, G., John, U., et al., 2020. Adaptive divergence across southern ocean gradients in the pelagic diatom fragilariopsis kerguelensis. Mol. Ecol. 29 (24), 4913–4924. https://doi.org/10.1111/mec.15554.
Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., et al., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81 (3), 559–575. https://doi.org/10.1086/519795.
Rastogi, A., Vieira, F.R.J., Deton-Cabanillas, A.-F., Veluchamy, A., Cantrel, C., Wang, G., Vanormelingen, P., et al., 2020. A genomics approach reveals the global genetic polymorphism, structure, and functional diversity of ten accessions of the marine model diatom phaeodactylum tricornutum. ISME J. 14 (2), 347–363. https://doi.org/10.1038/s41396-019-0528-3.
Read, B.A., Kegel, J., Klute, M.J., Kuo, A., Lefebvre, S.C., Maumus, F., Mayer, C., et al., 2013. Pan genome of the phytoplankton emiliania underpins its global distribution. Nature 499 (7457), 209–213. https://doi.org/10.1038/nature12221.
Rhodes, L.L., Adamson, J., Scholin, C., 2000. Pseudo-nitzschia Multistriata (Bacillariophyceae) in New Zealand. N. Z. J. Mar. Freshwater Res. 34 (3), 463–467. https://doi.org/10.1080/00288330.2000.9516948.
Rocha, E.P.C., Smith, J.M., Hurst, L.D., Holden, M.T.G., Cooper, J.E., Smith, N.H., Feil, E. J., 2006. Comparisons of dN/dS are time dependent for closely related bacterial genomes, 239. Memory of John Maynard Smith, pp. 226–235. https://doi.org/10.1016/j.jtbi.2005.08.037.
Ruggiero, M.V., Buffoli, M., Wolf, K.K.E., D'Alelio, D., Tuccio, V.Di, Lombardi, E., Manfellotto, F., et al., 2024. Multiannual patterns of genetic structure and mating type ratios highlight the complex bloom dynamics of a marine planktonic diatom. Sci. Rep. 14 (1), 6028. https://doi.org/10.1038/s41598-024-56292-y.
Ruggiero, M.V., D'Alelio, D., Ferrante, M.I., Santoro, M., Vitale, L., Procaccini, G., Montresor, M., 2018. Clonal expansion behind a marine diatom bloom. ISME J. 12, 463–472. https://doi.org/10.1038/ismej.2017.181.
Russo, MT., Vitale, L., Entrambasaguas, L., Anestis, K., Fattorini, N., Romano, F., Minucci, C., et al., 2018a. MRP3 is a sex determining gene in the diatom pseudo-nitzschia multistriata'. Nat. Commun. 9 (1), 5050. https://doi.org/10.1038/s41467-018-07496-0.
Russo, M.T., W. Sanseverino, R.A. Cigliano, and M.I. Ferrante. 2018b. 'Assessment of genomic changes in a CRISPR/Cas9 phaeodactylum tricornutum mutant through whole genome resequencing'. https://doi.org/10.7717/peerj.5507.
R Core Team, 2021. R: A language and environment for statistical computing.
Russo, M.T., Ruggiero, M.V., Manfellotto, F., Scriven, V., Campbell, L., Montresor, M., Ferrante, M.I., 2021. New alleles in the mating type determination region of west atlantic strains of pseudo-nitzschia multistriata. Harmful. Algae 103, 101995. https://doi.org/10.1016/j.hal.2021.101995. March.
Rynearson, T.A., Virginia Armbrust, E., 2005. Maintenance of clonal diversity during a spring bloom of the centric diatom ditylum brightwellii. Mol. Ecol. 14 (6), 1631–1640. https://doi.org/10.1111/j.1365-294X.2005.02526.x.
Sassenhagen, I., Gao, Y., Lozano-Duque, Y., Parsons, M.L., Smith, T.B., Erdner, D.L., 2018. Comparison of spatial and temporal genetic differentiation in a harmful dinoflagellate species emphasizes impact of local processes. Front. Mar. Sci. 5. https://doi.org/10.3389/fmars.2018.00393. October.
Scalco, E., Amato, A., Immacolata Ferrante, M., Montresor, M., 2016. The sexual phase of the diatom Pseudo-nitzschia multistriata: cytological and time-lapse cinematography characterization. Protoplasma 253 (6), 1421–1431. https://doi.org/10.1007/s00709-015-0891-5.
Sefbom, J., Kremp, A., Rengefors, K., Jonsson, P.R., Sjöqvist, C., Godhe, A., 2018. A planktonic diatom displays genetic structure over small spatial scales. Environ. Microbiol. 20 (8), 2783–2795. https://doi.org/10.1111/1462-2920.14117.
Sievers, F., Higgins, D.G., 2018. Clustal omega for making accurate alignments of many protein sequences'. Protein Sci. 27 (1), 135–145. https://doi.org/10.1002/pro.3290.
Sjöqvist, C., Godhe, A., Jonsson, P.R., Sundqvist, L., Kremp, A., 2015. Local adaptation and oceanographic connectivity patterns explain genetic differentiation of a marine diatom across the north sea–Baltic sea salinity gradient'. Mol. Ecol. 24 (11), 2871–2885. https://doi.org/10.1111/mec.13208.
Tammilehto, A., Gissel Nielsen, T., Krock, B., Friis Møller, E., Lundholm, N., 2015. Induction of Domoic Acid Production in the Toxic Diatom Pseudo-Nitzschia Seriata by Calanoid Copepods. Aquat. Toxicol. 159, 52–61. https://doi.org/10.1016/j.aquatox.2014.11.026. February.
Tamura, K., G. Stecher, and S. Kumar. 2021. 'MEGA11: molecular evolutionary genetics analysis version 11'. *Mol. Biol. Evol.* 38 (7): 3022–27. https://doi.org/10.1093/molbev/msab120.
Tesson, S.V.M., Montresor, M., Procaccini, G., Kooistra, W.H.C.F., 2014. Temporal changes in population structure of a marine planktonic diatom. PLoS. One 9 (12), e114984. https://doi.org/10.1371/journal.pone.0114984.
Tréguer, P., Bowler, C., Moriceau, B., Dutkiewicz, S., Gehlen, M., Aumont, O., Bittner, L., et al., 2018. Influence of diatom diversity on the ocean biological carbon pump'. Nat. Geosci. 11 (1), 27–37. https://doi.org/10.1038/s41561-017-0028-x.
Turk Dermastia, T., Dall'Ara, S., Dolenc, Jož, Mozetič, P., 2022. Toxicity of the diatom genus Pseudo-nitzschia (Bacillariophyceae): insights from toxicity tests and genetic screening in the northern Adriatic sea. Toxins (Basel) 14 (1), 60. https://doi.org/10.3390/toxins14010060.
Whittaker, K.A., Rynearson, T.A., 2017. Evidence for environmental and ecological selection in a microbe with no geographic limits to gene flow. Proc. Natl. Acad. Sci. 114 (10), 2651–2656. https://doi.org/10.1073/pnas.1612346114.
Zavatarelli, M., Raicich, F., Bregant, D., Russo, A., Artegiani, A., 1998. Climatological biogeochemical characteristics of the Adriatic sea. J. Mar. Syst. 18 (1), 227–263. https://doi.org/10.1016/S0924-7963(98)00014-1.