Supplementary Information

Clonal expansion behind a marine diatom bloom

Maria Valeria Ruggiero, Domenico D'Alelio, Maria Immacolata Ferrante, Mariano Santoro, Laura Vitale, Gabriele Procaccini & Marina Montresor

Department of Integrative Marine Ecology, Stazione Zoologica Anton Dohrn
Villa Comunale 80121 Napoli (Italy)

This document includes:

- Supplementary Methods
- Supplementary Figure 1
- Supplementary Text
- Supplementary Table S1
- Supplementary Table S2
- Supplementary Table S3

Supplementary Methods

Network analysis of allelic distance

The allelic distance matrix included 7,140 links, which represent the number of alleles by which the nodes reciprocally differed (allelic distance, w). Allelic distance ranged from 1 to 39 and its frequency distribution was bi-modal. The least copious sub-distribution included only 14% of all links but this fraction was composed by links connecting closely related MLGs that differed for less than 10 alleles. We thus chose to transform the weights of the above-mentioned distance matrix as follows:

$$w' = 1/w$$

where w is the allelic distance and w' is the allelic similarity derived by applying a logarithmic function dependent from w. This transformation allowed smoothing the noise produced by links connecting distantly related MLGs (about 86%). The new links-matrix was an allelic similarity instead of an allelic distance matrix. The allelic similarity network was produced and analyzed using the software Gephi (Bastian et al., 2009). Network layout was produced according to the 'ForceAtlas 2' algorithm (Jacomy et al., 2014). The latter allowed to represent the network based on a force-directed layout, i.e., the repulsion between weakly linked nodes was enhanced and the reverse was true for the more strongly linked nodes. The modularity of the network – a measure of how well a network decomposes into 'modular communities' – was analyzed according to a standard algorithm (Blondel et al., 2008). The degree of interconnection of a node was calculated on Gephi environment (parameterized as weighted-degree). This latter metric is based on the number of links for a node, but it is pondered by the weight of each link, i.e., it accounts for the sum of the weight of the links.

Testing for the presence of chytrids.

In order to test if the observed population genetics pattern could be due to selective forces exerted by parasitic infections, we tested for the presence of chytrid parasitoids that have been reported to infect diatom species and specifically *Pseudo-nitzschia* (Hanic *et al.*, 2009). Three ml of fixed samples collected at station LTER-MC before and during the bloom dominated by the clonal strain were placed in an Utermöhl chamber (Hydro-Bios, Kiel, Germany) and stained with 75 µL of Bactidrop®White Calcofluor (Remel Microbiology Products, Thermo Fisher Scientific, Kansas U.S.A.) that binds to the chitin layer of chytrids and fungi following the protocol illustrated in (Gerphagnon *et al.*, 2013). At least 100 cells were screened for each sample.

Comparison of growth rates.

In order to test if strains with the genetic profile of the dominant MLG #86 had higher growth rates as compared to those with a different genetic profile, we assessed the maximum growth rate of individual strains at the same experimental conditions. Four strains, belonging to Mt+, were selected for each group and exponentially-growing cells were inoculated at a final concentration of about 5,000 cell·ml^{-1} in eight 250 ml culture flasks containing 100 ml of f/2 culture medium. Culture flasks were incubated in a growth chamber at a temperature of 20°C, an irradiance of 100

µmol photons $m^{-2} \cdot s^{-1}$, and a photocycle of 12D:12L h. Every two days and for a total of 8 days, a subsample of 3 ml was collected and fixed with two drops of Lugol solution. Cell concentration was estimated with a Zeiss Axiophot light microscope (Carl Zeiss, Oberkochen, Germany) after settling 1 ml of the fixed sample in a Sedgewick–Rafter chamber. Growth rate, calculated by the linear regression over the exponential portion of the growth curve was expressed as divisions·day^{-1}.

Competition experiment in co-culture.

One strain with the genetic profile of the dominant MLG #86 (1075_22) and three strains with a different genetic profile (strains 1068_21, 1068_37, and 1075_13 belonging to MLGs #101, 79 and 66, respectively), all belonging to Mt+, were grown in monoculture and in triplicate co-cultures with the same final concentration as the mono-cultures. The same experimental set up illustrated above. On day 8, 10 single cells were isolated from each co-culture flasks and clonal cultures were obtained from 26 of them. DNA was extracted and strains were genotyped.

References

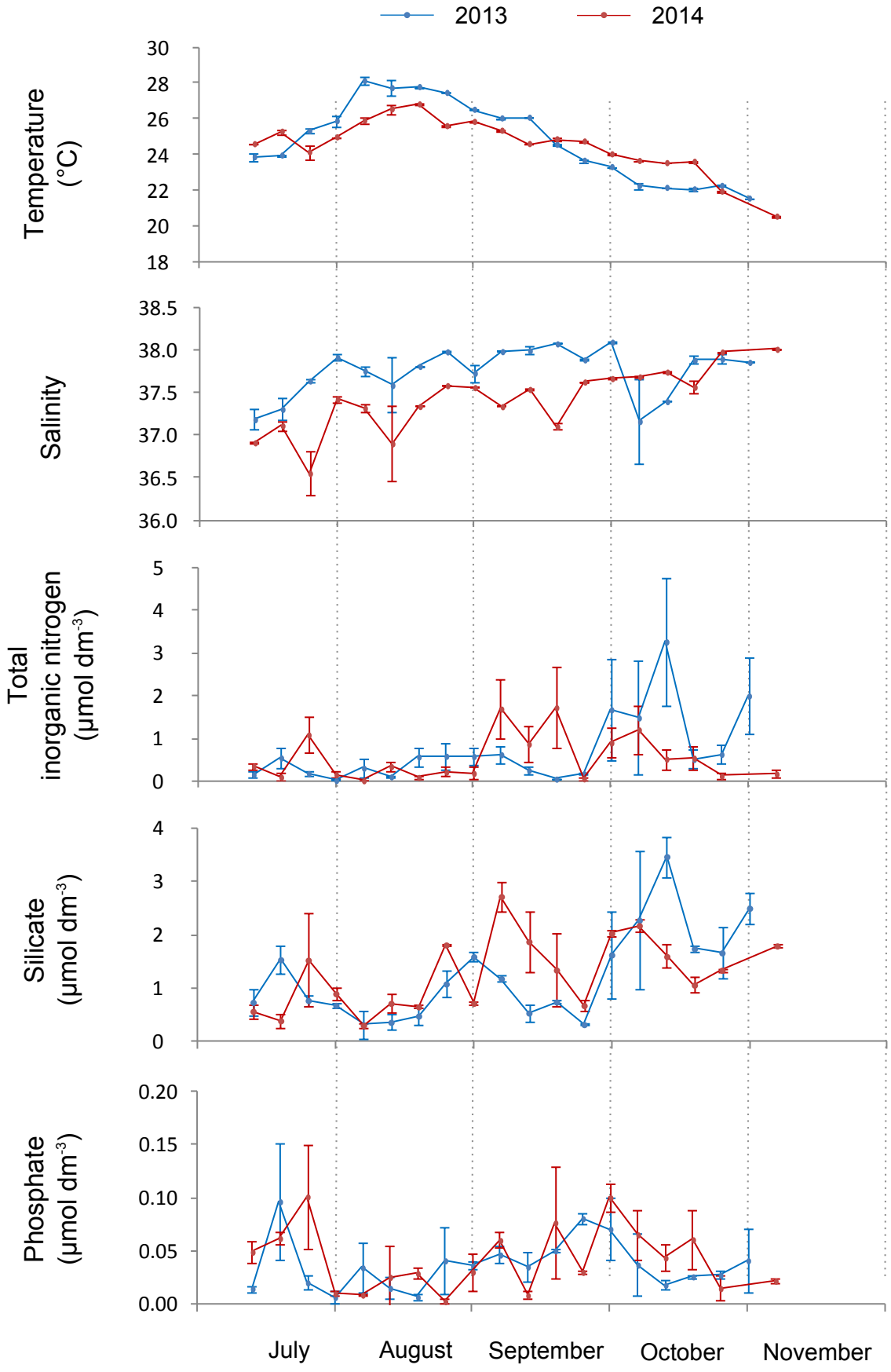
Bastian M, Heymann S, Jacomy M. (2009). Gephi: an open source software for exploring and manipulating networks. *ICWSM* **8**: 361–362.

Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. (2008). Fast unfolding of communities in large networks. *J Stat Mech theory Exp* **2008**: P10008

Gerphagnon M, Latour D, Colombet J, Sime-Ngando T. (2013). A double staining method using SYTOX green and calcofluor white for studying fungal parasites of phytoplankton. *Appl Environ Microbiol* **79**: 3943–3951.

Jacomy M, Venturini T, Heymann S, Bastian M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS One* **9**: e98679.

Supplementary figure 1



Supplementary figure 1. Trends of environmental parameters at LTER-MC during the blooming season of *P. multistriata* (years 2013-2014). Average values ± standard deviation. Sampling and water analyses were carried out based on methods described in Ribera d'Alcalà et al. 2004, Scientia Marina, Vol. 68, pages 65-83.

Supplementary Text: Mating type attribution to *P. multistriata* strains

Pseudo-nitzschia multistriata, as many pennate diatoms, has a heterothallic mating system, which implies that sex can occur only when strains of opposite mating type are in contact. However, there is no morphological character that can help in identifying the different mating types and the only way to proceed is to carry out a matrix of crosses between multiple strains. The crosses in which we see sexual stages belong to opposite mating type. The sexual phase in *P. multistriata* includes pairing of gametangia, formation of gametes (two for each gametangium), conjugation of gametes, and development of a particular stage – the auxospore – within which the large initial cell is produced (see Scalco et al., 2016*). The gamete migration pattern is uni-directional, and the auxospore will thus develop on one of the two paired gametangia. If we carry out crosses with two strains differing in size, we will be able to identify the one that bears the auxospore. We conventionally attribute Mt- to this latter strain and Mt+ to the opposite strain. This couple of strains can be used as a reference to attribute the mating type to strains isolated to the natural environment.

*Scalco, E., Amato, A., Ferrante, M. & Montresor, M. 2016. The sexual phase of the diatom *Pseudo-nitzschia multistriata*: cytological and time-lapse cinematography characterization. Protoplasma 253:1421–31.

Mating type attribution of *P. multistriata* strains isolated in different years (- = not tested).

	2008	**2009**	**2010**	**2011**	**2013**	**2014**
Mt+	94	16	23	-	199	14
Mt-	2	28	35	-	16	15
% Mt+	97.9	36.4	40.0		92.6	48.3
% Mt-	2.1	63.6	60.0		7.4	51.7

Supplementary Table 1: The 22 microsatellite loci for *Pseudo-nitzschia multistriata*. For each locus are given: repeat motif, location on the genome sequence (Basu et al. 2017), primer sequences, the fluorophore, the mix #, the allele size range, the number of alleles detected in the complete dataset (N), presence/absence (YES/NO) of null alleles (Nul), frequency of null alleles according to Brookfield's estimator 1 (r), presence/absence of stuttering (STU) and total expected heterozygosity (Ht).

Locus	Core repeat motif	Genomic location	Primer sequences (5'-3')	Fluoro-phor	mix #	Allele size range	N	Nul	r	STU	Ht
PNm1*	$(AG)_{23}$	Scaffold_196 (54655 - 54526)	F: CACCAATTGCATCCTAAAAGGG R: TCCGTCTAAGCCTGTATTTGTGAC	PET™	2	109-193	13	YES	0.118	YES	0.587
PNm2*	$(AC)_{17}$	Scaffold_306 (20001 - 19814)	F: GGGATCGATTCGTGAAAGAGC R: GCATAGAAGCACGGCACAGTG	NED™	3	170-248	8	NO	-0.216	NO	0.518
PNm3*	$(GAC)_{8}$	Scaffold_115 (11211 - 11418)	F: GGATCGAATAGGGGATGAATACG R: GGAGCTTGCATCATCATCACAG	VIC®	1	201-213	5	NO	0.007	NO	0.156
PNm5*	$(GT)_{11}$	Scaffold_279 (7841 - 8077)	F: GAACAGAACTGCCCGAAGGAC R: AGGATCACCCACGAGACACTG	NED™	2	195-241	9	NO	-0.018	NO	0.577
PNm6*	$(CT)_{9}$	Scaffold_95 (110675 - 110401)	F: AGCGAAAGCGACAAATAGCATC R: TGAGCAAAAGGACGAAACGAG	PET™	1	247-279	13	NO	-0.025	NO	0.678
PNm7*	$(CA)_{8}$	Scaffold_18 (53140 - 53402)	F: GTTGGCACCGGTGGTCTAAC R: CTTCGACGCTCCATTGGTG	6FAM™	4	247-265	4	NO	-0.098	NO	0.635
PNm16*	$(GTC)_{7}$	Scaffold_696 (4296 - 3975)	F: GGATCATACTGGAGGGGAACAAG R: GCTTTCACATCCAGAAGACAACAG	NED™	2	284-353	11	YES	0.064	NO	0.556
PNm90	$(GT)_{56}$	Scaffold_362 (20079 - 20433)	F: TTGCTGTGGGTGTGACAAAT R: CACTGCCAAGCTAACCACAA	PET™	3	357-445	18	YES	0.073	NO	0.446
PNm95	$(CA)_{59}$	Scaffold_110 (56037 - 56429)	F: TGCACCACTTGTGACTTGGT R: CTGCCCTTCATTTTGCATGT	NED™	1	321-441	20	NO	-0.045	NO	0.622
PNm254	$(GT)_{63}$	Scaffold_182 (83356 - 83643)	F: GCCCATCCTGTAAGCATTGT R: AGCTTTGCATCGTGGTTCTT	VIC®	1	247-383	18	NO	-0.080	NO	0.715
PNm349	$(AAG)_{17}$	Scaffold_172 (12946 - 13186)	F: TTCCAAGTGACTGCTCATGC R: CACCAACAGCAGCAAAGATG	VIC®	2	207-252	10	NO	-0.124	NO	0.748
PNm583	$(ATT)_{7}(AGT)_{14}$	Scaffold_342 (17162 - 17491)	F: TGCGTAGTACGGTGGAATGA R: CCAATCACTCAGTGGCTGAA	6FAM™	3	339-426	14	NO	-0.113	NO	0.663
PNm907	$(CA)_{11}(TA)_{53}$	Scaffold_26 (95244 - 95391)	F: TTGCGACCTATGCACAAACT R: TGGCCATGGTGTCTGTTCTA	PET™	4	90-152	9	NO	0.009	NO	0.405
PNm934	$(ATT)_{12}(AGT)_{15}$	Scaffold_568	F: AATTGTTTCCTTGGCCTTTG	VIC®	3	128-224	9	NO	-0.001	NO	0.210

		(13308 - 13492)	R: AGGCAGCCTTCTTAGAGCAT								
PNm1286	$(AC)_6(GC)_3(AC)_{62}$	Scaffold_956 (922 - 1084)	F: AGCCACTCCGCGATGTATAA R: GTGCAGTCCATGTTTCGTTG	6FAM™	4	153-285	15	YES	0.182	YES	0.591
PNm1493	$(CT)_{18}$	Scaffold_42 (25100 - 25344)	F: CGATTGTGCAGTGACGAGTT R: AACCCACAACGAGCAAAAAC	PET™	4	222-248	10	YES	0.190	YES	0.478
PNm1821	$(CT)_{19}$	Scaffold_137 (21283 - 21658)	F: AATTCATGCAAGCATCCACA R: CCTTCTGGGGAGAAGAATCC	VIC®	4	240-378	18	NO	-0.012	NO	0.715
PNm2198	$(CT)_{17}$	Scaffold_11 (2030 - 2273)	F: TGGAAGAAGCAAAGAACAGGA R: GAGTAGGGGTGGATCACCAG	VIC®	3	222-356	15	YES	0.110	YES	0.517
PNm2694	$(AGG)_{12}$	Scaffold_64 (30893 - 31200)	F: TGGAGGAATCAAAGGAGTGG R: AAGTCTCCCCCTGCTCCTAC	NED™	4	288-346	7	NO	-0.129	NO	0.539
PNm3011	$(AT)_{24}$	Scaffold_222 (5413 - 5606)	F: ATGCTCCCTCAGAATGGATG R: TCTTTGTTCTTGGCAAGGTG	PET™	1	137-197	13	NO	-0.134	NO	0.690
PNm6420	$(AGC)_{45}$	Scaffold_312 (66113 - 66277)	F: GAAGCCTCCTATTGCTGCAT R: ACTGCATTCCAGGATTGGTC	6FAM™	1	113-245	14	NO	-0.065	NO	0.637
PNm7546	$(AT)_{18}$	Scaffold_591 (12887 - 13245)	F: CAAGTGCAGCTCACCGATTA R: AGTCACCTGAGGGACCATGA	6FAM™	2	230-442	18	NO	-0.024	NO	0.474

***From Tesson *et al.* (2011) [Tesson, S. V. M., Borra, M., Kooistra, W. & Procaccini, G. 2011. Microsatellite primers in the planktonic diatom *Pseudo-nitzschia multistriata* (Bacillariophyceae). Am. J. Bot. 98:E33-E35].**

Supplementary Table 2: For each MLG of *Pseudo-nitzschia multistriata* with more than one replicate are reported: the number of strains sharing the MLG; $P_{gen(fis)}$; and $P_{sex(fis)}$ for 1 re-encounter and for the maximum number of re-encounters (in brackets) in the data-set. $P_{gen(fis)}$; and $P_{sex(fis)}$ are, respectively, the probability of occurrence of a given MLG and the probability that repeated MLGs may arose by chance and not from distinct sexual events.

MLG	N strains	$P_{gen\ (fis)}$	$P_{sex\ (fis)}$ 1 re-encounter	$P_{sex\ (fis)}$ max re-encounter (N max re-encounter)
MLG 17	2	$-1.54 \bullet 10^{-35}$	$-5.80 \bullet 10^{-33}$	-
MLG 23	2	$-1.86 \bullet 10^{-29}$	$-7.03 \bullet 10^{-27}$	-
MLG 34	2	$5.81 \bullet 10^{-18}$	$2.19 \bullet 10^{-15}$	-
MLG 38	2	$-1.10 \bullet 10^{-21}$	$-4.15 \bullet 10^{-19}$	-
MLG 65	3	$5.00 \bullet 10^{-06}$	$1.88 \bullet 10^{-03}$	$1.77 \bullet 10^{-06}$(2)
MLG 66	6	$1.09 \bullet 10^{-05}$	$4.12 \bullet 10^{-03}$	$9.64 \bullet 10^{-15}$(5)
MLG 69	3	$6.11 \bullet 10^{-07}$	$2.30 \bullet 10^{-04}$	$2.64 \bullet 10^{-08}$(2)
MLG 73	2	$-4.31 \bullet 10^{-32}$	$-1.62 \bullet 10^{-29}$	-
MLG 85	7	$1.43 \bullet 10^{-05}$	$5.36 \bullet 10^{-03}$	$3.22 \bullet 10^{-17}$(6)
MLG 86	208	$3.12 \bullet 10^{-05}$	$1.17 \bullet 10^{-02}$	0(207)
MLG 89	2	$1.33 \bullet 10^{-06}$	$5.01 \bullet 10^{-04}$	-
MLG 92	7	$6.76 \bullet 10^{-06}$	$2.55 \bullet 10^{-03}$	$3.66 \bullet 10^{-19}$(6)
MLG 94	5	$8.55 \bullet 10^{-06}$	$3.22 \bullet 10^{-03}$	$4.42 \bullet 10^{-12}$(4)
MLG 100	4	$8.78 \bullet 10^{-06}$	$3.30 \bullet 10^{-03}$	$5.98 \bullet 10^{-09}$(3)
MLG 101	2	$2.40 \bullet 10^{-06}$	$9.06 \bullet 10^{-04}$	-
MLG 103	2	$2.82 \bullet 10^{-08}$	$1.06 \bullet 10^{-05}$	-
MLG 109	3	$7.35 \bullet 10^{-09}$	$2.77 \bullet 10^{-06}$	$3.83 \bullet 10^{-12}$(2)
MLG 110	2	$2.62 \bullet 10^{-06}$	$9.88 \bullet 10^{-04}$	-
MLG 111	7	$5.74 \bullet 10^{-06}$	$2.16 \bullet 10^{-03}$	$1.37 \bullet 10^{-19}$(6)
MLG 114	2	$6.27 \bullet 10^{-24}$	$2.36 \bullet 10^{-21}$	-

Supplementary Table 3. Maximum growth rates of four strains with distinct MLGs (black) and four strains belonging to the dominant MLG #86 (red) isolated in summer 2013. Strain code, average cell size (apical axis), growth rate of the single strains and average growth rate (± standard deviation) of the two groups of strains.

Strain code	Avg. cell size (µm)	Growth rate (div. day^{-1})	Avg. growth rate (± st. dev.)
1068_21	22	0.90	0.99 (±0.07)
1068_37	29	1.02	
1075_13	25	1.05	
VA1_12	20	1.00	
1075_22	29	0.96	0.87 (±0.09)
1075_25	27	0.76	
1078_17	31	0.85	
1078_36	21	0.91	