Reduced representation sequencing accurately quantifies relative abundance and reveals population-level variation in *Pseudo-nitzschia* spp.

Carly D. Kenkel*[a], Jayme Smith[b], Katherine A. Hubbard[c,d], Christina Chadwick[c], Nico Lorenzen[a], Avery O. Tatters[e] & David A. Caron[a]

[a]Department of Biological Sciences, University of Southern California, 3616 Trousdale Parkway, Los Angeles, CA 90089, USA

[b]Southern California Coastal Water Research Project, 3535 Harbor Boulevard, Suite 110, Costa Mesa, CA, 92626, USA

[c]Florida Fish and Wildlife Conservation Commission-Fish and Wildlife Research Institute (FWC-FWRI), 100 8th Ave. SE, St. Petersburg, FL 33701, USA

[d]Biology Department, Woods Hole Oceanographic Institution, 266 Woods Hole Road, Woods Hole, MA 02543, USA

[e]U.S. Environmental Protection Agency, Gulf Ecosystem Measurement and Modeling Division, 1 Sabine Island Drive, Gulf Breeze, FL, 32561, USA

*corresponding author: ckenkel@usc.edu, phone: 213-821-1705, fax: 213-740-8123

Keywords: harmful algal bloom, community composition, population genomics, next-generation sequencing, 2b-RAD

Highlights: (3-5 bullet points, max 85 chars)

- 2bRAD method facilitates species- and population-level analysis of the same sample
- Method accurately quantifies species relative abundance with low false positives
- Consistent shifts in allele frequencies were detected between high and low DA years
- Certain *Pseudo-nitzschia* spp. populations may be more associated with DA presence

Abstract

Certain species within the genus *Pseudo-nitzschia* are able to produce the neurotoxin domoic acid (DA), which can cause illness in humans, mass-mortality of marine animals, and closure of commercial and recreational shellfisheries during toxic events. Understanding and forecasting blooms of these harmful species is a primary management goal. However, accurately predicting the onset and severity of bloom events remains difficult, in part because the underlying drivers of bloom formation have not been fully resolved. Furthermore, *Pseudo-nitzschia* species often co-occur, and recent work suggests that the genetic composition of a *Pseudo-nitzschia* bloom may be a better predictor of toxicity than prevailing environmental conditions. We developed a novel next-generation sequencing assay using restriction site-associated DNA (2b-RAD) genotyping and applied it to mock *Pseudo-nitzschia* communities generated by mixing cultures of different species in known abundances. On average, 94% of the variance in observed species abundance was explained by the expected abundance. In addition, the false positive rate was low (0.45% on average) and unrelated to read depth, and false negatives were never observed. Application of this method to environmental DNA samples collected during natural *Pseudo-nitzschia* spp. bloom events in Southern California revealed that increases in DA were associated with increases in the relative abundance of *P. australis*. Although the absolute correlation across time-points was weak, an independent species fingerprinting assay (Automated Ribosomal Intergenic Spacer Analysis) supported this and identified other potentially toxic species. Finally, we assessed population-level genomic variation by mining SNPs from the environmental 2bRAD dataset. Consistent shifts in allele frequencies in *P. pungens* and *P. subpacifica* were detected between high and low DA years, suggesting that different intraspecific variants may be associated with prevailing environmental conditions or the presence of DA. Taken together, this method presents a potentially cost-effective and high-throughput approach for studies aiming to evaluate both population and species dynamics in mixed samples.

1. Introduction

Multiple species of the diatom genus *Pseudo-nitzschia* produce the water-soluble neurotoxin, domoic acid (DA). The toxin bioaccumulates via food web transfer, leading to illness and mass mortality of marine animals (Bejarano et al., 2008; Kvitek et al., 2008; Moriarty et al., 2021), significant economic costs to commercial and recreational shellfisheries (Moore et al., 2020; Wessells et al., 1995), and amnesic shellfish poisoning in humans who consume contaminated seafood (Trainer et al., 2012). Blooms of toxigenic *Pseudo-nitzschia* cause significant negative impacts in many regions of the world, including along the west coast of North America, where fishery closures and animal mortality occur frequently (Scholin et al., 2000; Smith et al., 2018a; Trainer et al., 2012). In California, blooms have been increasing in frequency and severity in recent years (Schnetzer et al., 2013; Trainer et al., 2010). In 2007, DA concentrations in mussels collected in California reached 610 µg g^{-1}, more than 30 times higher than the U.S. Food and Drug Administration (FDA) tissue safety level of 20 µg g^{-1} (Trainer et al., 2012). In 2014, multiple mussel samples from Marin County contained DA concentrations in excess of 1000 µg g^{-1} (Langlois et al., 2014). *Pseudo-nitzschia* spp. bloom 'hot spots' in California are spread along the coast, from Monterey Bay in the north to San Luis Obispo and Point Conception, and the San Pedro/Long Beach Harbor area (Schnetzer et al., 2013; Smith et al., 2018a; Trainer et al., 2012), but it is not clear what attributes unite these geographically distant sites to promote HAB development.

Environmental factors implicated in bloom formation include nutrient enrichment from terrestrial sources, coastal upwelling, and mesoscale eddies (Anderson et al., 2008; Schnetzer et al., 2013). The grand majority of DA impacts on coastal ecosystems have been described in eastern boundary upwelling regions, including along the U.S. west coast where the California Current prevails (Trainer et al., 2010). Prior work has suggested that in California, blooms are associated with weak upwelling, lower salinity, temperature gradients and low macronutrient levels (Kudela et al., 2002). Culture and field studies have shown that phosphorus and silicate limitation can increase DA production (Fehling et al., 2004). The form of nitrogen can also play a role in DA production (Cochlan et al., 2008; Howard et al., 2007; Kudela et al., 2008), as can trace metal availability, including iron limitation and copper toxicity (Wells et al., 2005). Increased pCO_2 has also been shown to increase DA production, and increased pCO_2 in combination with silicate limitation has a synergistic effect on DA production (Tatters et al., 2012). In addition to the abiotic environment, biotic interactions, such as bacterial associations have also been implicated in toxin production (Bates et al., 1995). However, in field studies, consistent relationships between DA levels and specific environmental conditions have been less clear, with one study finding an association with colder, more saline water (Schnetzer et al., 2013); another, with dissolved silicic acid concentration (Smith et al., 2018b). While bloom risk mapping models have been developed for the California coast (C-HARM, (Anderson et al., 2016)), it remains challenging to predict the onset and severity of *Pseudo-nitzschia* bloom and toxic events, in part because the underlying drivers have not been fully resolved (Lelong et al., 2012; Smith et al., 2018b) and also because blooms are often a mixture of co-occurring species not all of which are toxigenic, or actively producing toxins for those that are.

More than 52 species of *Pseudo-nitzschia* have been described, but not all are toxigenic (Bates et al., 2018). Among the 26 species currently known to produce DA, toxin production was reported to be non-constitutive (Bates et al., 2018; Lelong et al., 2012; Trainer et al., 2012). Thus, variation in toxin concentrations during *Pseudo-nitzschia* blooms may stem from the presence or absence of exacerbating environmental factors and the ability to accurately measure such factors. This difference may also be due to an inability to accurately identify species- and population-level variation in *Pseudo-nitzschia* bloom composition. Recent work in several systems has indicated that the presence of certain species may be an important predictor of bloom toxicity, in addition to prevailing physiochemical factors such as temperature, salinity and nutrients (Clark et al., 2019; Smith et al., 2018b). In the San Pedro/Long Beach Harbor bloom hot spot, elevated DA concentration has been associated with community-level dominance of *Pseudo-nitzschia australis* and/or *P. seriata (Schnetzer et al., 2013; Smith et al., 2018b). P. multiseries* has also been reported as a major toxin producer in California (Trainer et al., 2000), but further north in Washington, *P.* cf. *pseudodelicatissima* and *P. cuspidata* join *P. australis* and *P. multiseries* as dominant toxin producers (Trainer et al., 2009).

Taxonomic identification of protists, such as *Pseudo-nitzschia,* is largely based on morphology, but salient features may not be sufficient to distinguish species with similar morphologies (Adl et al., 2007). Ultrastructural features of diatom cell walls visualized using scanning electron microscopy have long been the gold standard for species delineation (Round et al., 1990). However, the application of DNA sequencing revealed significant cryptic diversity – morphologically similar algae can have different genetic profiles (Alverson, 2008). Various genetic assays have been developed to genotype *Pseudo-nitzschia* spp., including direct sequencing of ribosomal gene regions (Casteleyn et al., 2010; Lim et al., 2014; Orsini et al., 2004), and DNA fingerprinting (Bornet et al., 2005) including microsatellites (Evans et al., 2004; Evans and Hayes, 2004). Automated Ribosomal Intergenic Spacer Analysis (ARISA) was more recently developed to distinguish *Pseudo-nitzschia* species within a heterogeneous bloom without the need for culturing (Hubbard et al., 2008), but this method still relies on a single locus, which limits the resolution of population genetic analyses because estimates of relatedness based on single loci are noisy (Lynch and Milligan, 1994). Furthermore, in spite of its broader field applicability, ARISA is still unable to fully resolve differences among certain *Pseudo-nitzschia* species in California, including *P. australis*, putatively one of the most toxic species, and another potentially toxic species, *P. seriata* (Hubbard et al., 2008).

Metagenomic analyses have the potential to provide increased taxonomic resolution, but as *Pseudo-nitzschia* spp. genomes can be large (e.g. *P. multiseries*, 218 Mbp, (Osuna-Cruz et al., 2020)) it will be challenging to cost-effectively obtain sufficient coverage for high-resolution spatial and temporal datasets. Reduced representation sequencing using Restriction site-Associated DNA (RAD) was originally developed for population genetic analyses of multicellular organisms (Hohenlohe et al., 2010). However, this method has the potential to cost-effectively facilitate understanding of both community and population-level variation in single-cell eukaryotes. RAD can identify and score thousands of genetic markers from individual or pooled samples (Davey and Blaxter, 2010; Schlötterer et al., 2014). It is analogous to other molecular genotyping methodologies, such as restriction fragment length polymorphisms (RFLPs) and

amplified fragment length polymorphisms (AFLPs), which have been used to study phytoplankton population genetics since the late 1970s (Bruin et al., 2003), in that it reduces the complexity of the genome by subsampling only restriction enzyme recognition sites (Davey and Blaxter, 2010). However, RAD greatly surpasses these and other methods, including microsatellites, in its ability to identify, verify and score markers simultaneously and without an extensive developmental process – genotyping can be accomplished '*de novo*', even on species with no prior sequence data (Davey and Blaxter, 2010; Matz, 2018). 2b-RAD is a type of RAD sequencing which uses type IIB restriction enzymes (Wang et al., 2012). These enzymes cut upstream and downstream of their recognition sites, generating uniform 36-bp DNA fragments, or tags, ideally suited for next-generation short read sequencing technologies. Analysis of within-species (~population) genomic variation is based on identifying single nucleotide polymorphisms (SNPs) that occur within these 36-bp fragments and calculating allele frequencies based on read counts of tags containing alternate alleles. Here we develop a novel next-generation sequencing based assay using 2b-RAD and apply it to mock *Pseudo-nitzschia* communities generated by mixing cultures of different species in known abundances and field samples collected during natural *Pseudo-nitzschia* spp. bloom events.

2. Methods

2.1 *Reference Cultures*

Pseudo-nitzschia cultures were initiated from individual chains collected from multiple locations in the North Pacific. Seven strains representing four species, *P. australis* (isolated at Gaviota Beach, Santa Barbara County, California, 34.47˚ N, 120.23˚ W in April 2016), *P.* sp. (2 cultures, isolated from near Station Aloha 22.75°N, 158°W in May 2016), *P. pungens* (3 cultures, isolated from Newport Beach, Orange County, California, 33.61˚ N, 117.93˚ W in March 2017), and *P. subpacifica* (also isolated from Newport Beach, Orange County, California in March 2017), were grown in a modified F medium with 50% nutrient concentration (35.3 µM NO^{3-}, 4.24 µM SiO_4^{2-}, 1 µM PO_4^{3-}) in autoclaved 0.2 µm filtered seawater. Incubators maintained cultures at 15°C, except for the *P.* sp. cultures which were maintained at 22°C. Cells were maintained under a 12h:12h light:dark cycle using cool white fluorescent illumination (~110 µmol photons m^{-2} s^{-1}). Cultures were visually inspected every 7-10 days with a dissecting scope for morphological signatures of health including proper cell shape and chain formation and transferred to new media on this same schedule.

Pseudo-nitzschia species were initially distinguished by overall size, shape of valve ends, and degree of valve overlap within chains by examination up to 400x magnification using light microscopy. *P. australis* (Fig. S1A), *P. pungens* (Fig. S1B) and *P. subpacifica* isolates were identified using frustule features including poroids, interstriae and fibulae observed via scanning electron microscopy (SEM). Respective samples were processed according to Hasle and Syvertsen (Hasle and Syvertsen, 1997). The cleaned frustules were observed with a Philips XL 30 S, FEG SEM. The small-celled *P.* sp. isolates were evaluated using light microscopy.

2.2 *Generating Mock Communities*

Mock communities were generated with cultured isolates identified to species level. Replicate community mixes with individual species representatives ranging in relative abundance from 10 (0.01%) to 99,990 cells (99.99%) in 100,000 total cells (Table 1) were generated to approximate the minimum bloom threshold of 80,000 cells/L previously described at Newport Pier, CA (Seubert et al., 2013). Triplicate cell counts were completed using a Sedgwick Rafter for each reference species after staining a 4-mL aliquot with 50 µL/mL Lugol's Solution. A minimum of 400 cells were counted in each count and the relative standard deviation of the triplicate counts ranged from 2-10% (Table S1). Culture volumes were pooled within 4 hours of cell enumeration and immediately deposited onto 25mm GF/F filters (nominal pore size 0.7 microns) which were stored at -20°C until DNA extractions were completed. The level of detection for cellular enumeration was 3,000 cells/L for each individual count and so mix fractions represented by less than 3,000 cells (0.1 and 0.01%) were made through serial dilution prior to mixing. There were 6 mix types in addition to pure culture samples as detailed in Table 1.

Table 1. Replication of mock culture sample mixing scheme.

Targeted cell abundance ratio	Percent abundance ratio	N replicate mixes per ratio
100,000	100%	3-5 per species
99,990 : 10	99.99% : 0.01%	8
99,900 : 100	99.9% : 0.1%	4
75,000 : 20,000 : 5,000	75% : 20% : 5%	3
66,000 : 33,000	66% : 33%	1
60,000 : 30,000 : 10,000	60% : 30% :10%	4
40,000 : 30,000 : 30,000	40% : 30% : 30%	4
33,000 : 33,000 : 33,000	33% : 33% : 33%	3
50,000 : 50,000	50% : 50%	3

2.3 *DNA extraction from mock community mixes*

Pure culture and mock community mix samples were extracted using a phenol-chloroform protocol modeled after (Countway et al., 2005). Briefly, 2 mL of lysis buffer comprised of 40 mM EDTA (ph 8), 100 mM Tris (ph 8), 100mM NaCl, and 1% SDS weight/volume was added to frozen filters in 15 mL Falcon tubes. Tubes were thawed at 70°C in a water bath after which 200 µL of 0.5 mm zirconia/silica beads (BioSpec Products) were added. The mixture was vortexed for 1 minute then heated at 70°C for 5 minutes for four repetitions to lyse the cells and break down the filter. The lysate and filter mixture was transferred to a 10 mL syringe and lysate was dispensed into a new 15 mL Falcon tube. 2.5 M NaCL and 10% CTAB were added to yield a 0.7 M NaCl 1% CTAB solution and the full mixture was incubated for 10 minutes at 70°C.

An equal volume of 25:24:1 phenol:chloroform:isoamyl alcohol (Sigma) was added and the mixture was vortexed and centrifuged at 4750 RPM for 10 minutes. After the layers settled the bottom layer was removed via pipetting and discarded. An equal volume of phenol:chloroform:isoamyl alcohol was again added and the mix was vortexed and centrifuged at 4750 RPM for 10 minutes. The supernatant was removed and placed in a new Falcon tube.

The supernatant was extracted by performing two repetitions of adding an equal volume addition of 24:1 chloroform: isoamyl alcohol (Sigma), vortexing, centrifuging at 4750 RPM for 10

minutes and discarding the bottom layer up to the interface. The supernatant was distributed into 1.5 mL Eppendorf tubes in 400 µL aliquots. Each aliquot was precipitated in 800 µL of cold 95% ethanol and 0.1X volume of 10.5 M ammonium acetate (40 µL), vortexed and incubated overnight at -20°C.

Following incubation samples were centrifuged at 14000 RPM for 30 minutes at 4°C. Liquid was decanted and the pellet was rinsed in 70% cold ethanol and centrifuged at 14000 RPM for 15 minutes at 4°C after which the liquid was decanted and the pellet was allowed to air dry. The pellet of each aliquot was suspended in 30 µL MilliQ H_2O and aliquots from the same sample were combined for DNA quantitation.

Extractions were further purified using a Genomic DNA Clean & Concentrator-10 kit (Zymo Research) according to the manufacturer's instructions. Briefly, the eluted DNA was combined with buffer and placed into a spin column. Samples were washed and centrifuged before being eluted in 10 µL elution buffer. Samples were vacuum-concentrated using a Speed Vac (Thermo Fisher, USA) on low speed for 1 min to yield a sufficient volume for the 2bRAD protocol (~8 µL).

2.4 *Natural Time-series Samples*

Weekly water samples were obtained as part of long-term SCCOOS & CeNCOOS Harmful Algal Bloom Monitoring Alert Program (https://sccoos.org/harmful-algal-bloom/) at Newport Beach Pier, CA (Kudela et al., 2015; Seubert et al., 2013; Smith et al., 2018a). Samples have been routinely collected since 2008 for the quantification of domoic acid and enumeration of harmful phytoplankton taxa, as well for basic physicochemical conditions. Particulate domoic acid (pDA) samples were collected via the filtration of 200 mL of sea water onto a GF/F filter and were frozen at -20°C until extraction. Samples were quantified via Enzyme-Linked ImmunoSorbant Assay (ELISA: Mercury Science, Durham, NC) with a detection limit of 0.02 ng/mL. Samples below the detection limit were assumed to be zero for all calculations and statistical analyses. Seawater samples were collected and preserved with 3% formaldehyde (final concentration) for the enumeration of harmful phytoplankton taxa including *Pseudo-nitzschia*. Cells were enumerated using a Leica DM IRBE inverted light microscope (Leica Microsystems, Buffalo Grove, IL) at 400x after settling 25 mL of the sample in Utermöhl chambers for approximately 24 hours (Utermöhl, 1958). *Pseudo-nitzschia* cells were categorized into seriata size class (> 3 µm) and delicatissima (< 3 µm) size classes. Light microscopy based size sorting was conducted to differentiate between larger species more often associated with detectable pDA in the region and smaller cells that are more rarely associated with high toxin concentrations, as described in Seubert et al. (2013).

DNA was extracted from archived HABMAP sampling filters for 2bRAD sequencing. Samples were selected from the spring bloom in four different years in which *Pseudo-nitzschia* spp. cells were quantified, two in which in pDA was observed (2017 and 2019) and two in which no DA was detected (2015 and 2018). Briefly, 500 mL of water were filtered onto a 25mm GF/F filter and subsequently frozen at -20°C. DNA was extracted from filters using a Qiagen Plant kit (Qiagen, USA).

2.5 *Library Preparation and Sequencing*

DNA samples from the mock community mixes were prepared for 2bRAD sequencing using the BcgI restriction endonuclease targeting 100% of restriction sites (Wang et al., 2012). The same protocol was used for the natural time-series samples, save that a 1/64 site reduction scheme was used. The full library preparation protocol is available at https://github.com/z0on/2bRAD_denovo. Culture reference and mock community mix libraries were run on two lanes of Illumina HiSeq 2500, SR 50 format. Natural bloom libraries were run on four lanes of Illumina Nextseq 550, SR 75 format. All sequencing runs were carried out by the University of Southern California Genome core.

2.6 *Generation of RAD reference libraries*

All reads originating from pure culture samples were used to create independent RAD tag reference libraries for subsequent read mapping. 2bRAD tag references were also generated for the *Pseudo-nitzschia multiseries* reference genome (CLN-47, JGI) and a whole genome shotgun sequence set for *Pseudo-nitzschia multistriata* (CAACVS01, GCA_900660405, ENA). Full scaffolds, mitochondria, and plastid assemblies were concatenated and a custom perl script was used to extract tags exhibiting the Bcg I restriction site motif. Full bioinformatic protocols and scripts can be found at https://github.com/ckenkel/Pseudo-nitzschia2bRAD. Briefly, for sequenced cultures, a custom perl script was used to trim sequencing adapters from raw reads, retaining only the 36-bp insert. These trimmed files were quality filtered using the fastx_toolkit (Assaf and Hannon, 2010) requiring a minimum phred quality of 20 over 100% of the read. The bbduk command of the BBMap package (Bushnell, 2014) was then used to filter reads matching to Illumina sequencing adapters. Reads passing these filtering criteria were clustered at 100% identity using a custom perl script to remove perfect duplicates, and only reads exhibiting a 100% match to the restriction site were retained (Table S1).

As cultures were not axenic, reads were then filtered for contaminants based on a BLAST search against the NCBI nt database ((Sayers et al., 2019), downloaded 17 Jun 2020). Up to five alignments were reported for each read for matches below the e-value threshold of 10^{-5}. A custom perl script was then used to summarize best hits using the NCBI taxonomy (Schoch et al., 2020). Contaminant reads were flagged as those exhibiting significant similarity to any non-Bacillariophyte sequence. Reads for each species were then clustered at 91% identity, which allows for up to 3 mismatches, or SNPs, per tag, using cd-hit (Fu et al., 2012; Li and Godzik, 2006). Species clusters containing any reads previously identified as matching to a contaminant were removed using a custom bash script. Species specific references and known contaminants were then concatenated into a ‘global’ reference, retaining a species-level identifier within each tag.

2.7 *Analysis of mock community mixes and natural time-series samples*

All bioinformatic and statistical scripts can be found at https://github.com/ckenkel/Pseudo-nitzschia2bRAD. A custom perl script was used to filter and trim raw reads. Reads were discarded if they did not contain the correct restriction site motif or adapter sequence, and those exhibiting the expected sequence construct were subsequently trimmed to remove adapter sequences. In addition, PCR duplicates were discarded by retaining only one representative

read for sequences sharing the same 64-fold degenerate adapter sequence, the first 34 bases of the insert, and the secondary barcode. Adapter trimmed reads were further filtered for quality, retaining only reads exhibiting Q20 over 100% of bases and exhibiting no match to Illumina adaptors using the fastx_toolkit (Assaf and Hannon, 2010) and BBMap (Bushnell, 2014)(Table S2). Reads were mapped against the global species reference library using bowtie2 (Langmead and Salzberg, 2012), default parameters, including the --very-sensitive flag) and a custom bash script was used to count the number of reads exhibiting high quality matches (reads with MAPQ<23, discarding any ambiguous matches) to each species reference.

Samples with fewer than 1000 high quality mapped reads remaining were removed. The false positive rate (FPR) was calculated as the sum of reads matching species known to be in each sample relative to the total number of high quality mapped reads for that sample. Accuracy was calculated as the absolute value of the difference between observed and expected percent abundance values for each focal species within each sample.

Coverage of natural time-series sample reads mapping to *Pseudo-nitzschia* spp. in the global reference library was lower than the 80-100x minimum coverage recommended for pooled sequencing (Schlötterer et al., 2014). Therefore, to achieve sufficient read depth to assess allelic diversity within species over time, mapped reads in SAM files were split by species and concatenated by year. Tags with three or more SNPs were excluded from further analysis and remaining SNPs were thinned to one per tag. A Cochran-Mantel-Haenszel test as implemented in Popoolation2 (Kofler et al., 2011) was subsequently used to test for consistent changes in allelic diversity between high DA and low DA years. We required the minimum allele count to be 12, and the minimum and maximum coverage to be 50x and 200x respectively. P-values were calculated from pairwise comparisons of 2015 vs 2017 and 2018 vs 2019. Plotting and statistical analyses were carried out using the R language environment (v4.1.0, (R Core Team and Others, 2017).

2.8 Amplification and identification using ARISA

Pseudo-nitzschia species community composition in natural time series samples were also analyzed via ARISA. Genomic DNA concentrations were quantified on a plate reader (Biotek Instruments) using Picogreen (Invitrogen) and then standardized to 1ng μL^{-1} prior to amplification. Genus-specific oligonucleotides PnAll F (5′-TCTTCATTGTGAATCTGA-3′) and FAM-labeled PnAll R (5′- CTTTAGGTCATTTGGTT-3′) were used to amplify the ITS1 region (Hubbard et al., 2014, 2008). For PCR, 10 ng of genomic DNA was added to replicate (duplicate or triplicate) 20 µL reactions consisting of 2.5 mM deoxynucleoside triphosphates, 0.4 mM of each primer, 0.75 U of Apex Taq polymerase, 2 mM of $MgCl_2$, and 1 x standard reaction buffer (Apex Bioresearch Products). Amplification was conducted using a 2–minute denaturation step at 94˚C, followed by 32 cycles of 30 seconds at 95˚C, 30 seconds at 50.6˚C, and 60 seconds at 72˚C, and ending with a 10 minute extension at 72˚C. Resulting products were purified using MultiScreen PCRµ96 filter plates (Millipore). Replicate reactions were pooled, quantified using Picogreen, and diluted to 1 ng μL^{-1}. For fragment analysis, 1 ng of PCR product was processed using an Applied Biosystems 3730 XL DNA Analyzer (University of Illinois DNA Core Sequencing Facility) with a LIZ600 size standard. Electropherograms were analyzed using DAx

software (Van Mierlo Software Consultancy) to determine peak height and size (in base pairs or bp). Relative abundance of each amplicon was calculated by dividing the height of each individual peak by total peak height, and only those peaks that exceeded 3.0% of the total peak height were used in the final dataset following (Hubbard et al., 2014, 2008). Furthermore, only samples with a total peak height of >1000 relative fluorescent units (RFUs) were used. Relative peak height was previously determined to be correlated with the proportion of ITS1 copies added to the PCR, recognizing that larger-celled species generally have larger genomes and more ITS1 copies per cell (Hubbard et al., 2014) and was thus used to provide semi-quantitative data about changes in species contributions over time.

For each peak, 4 bp were added manually such that amplicon sizes for ARISA matched those based on *in sillico* sequence comparisons. A subset of samples was selected for confirmatory ITS1 sequencing where 1 µL of genomic DNA was amplified as described above (with an annealing temperature of 67˚C) using 18SF-euk (5′-CTTATCATTTAGAGGAAGGTGAAGTCG-3′) and 5.8SR-euk (5′-CTGCGTTCTTCATCGTTGTGG-3′) oligonucleotides. The resulting PCR product was run on a 3% agarose gel. Using sterile micropipette tips, distinct bands were individually picked from the gel and then gel picks were dissolved in 10 µL of sterile molecular grade water for 5 minutes at 40˚C prior to PCR amplification. The PnAll F/R primer set was utilized to amplify 1 µL of the melted gel/sterile water mixture and the resulting PCR product was run on a 3% agarose gel to confirm the presence of a single band and assess approximate product size; gel picks from these products were used in another round of PCR with the PnAll F/R primer set. A 1.5% agarose gel was used to confirm single band product sizes and remaining products were purified using ExoSap-It Express PCR Product Cleanup Reaction (Applied Biosystems). Purified products were then sequenced on a 3730 XL DNA Analyzer (Applied Biosystems) by Eurofins Genomics LLC. Sequences were analyzed using Sequencher software (Gene Codes Corporation) and identified using BLASTn to query the National Center for Biotechnology Information's (NCBI) GenBank nucleotide database (Sayers et al., 2019) to verify sizes determined by ARISA. Linear models were used to compare the relative abundance of focal species derived using the different methods (ARISA vs 2bRAD) and a Fisher's exact test was applied to compare presence/absence based detection using a contingency table comparing the overlap between positive and negative calls for each method.

3. RESULTS

3.1 Reference Library Construction

Sequencing yielded a total of 133,370-452,059 raw reads for each species (Table 2). Of these, between 6.5-24.4% were identified as non-Bacilliariophyte contaminants. The majority of contaminants exhibited best matches to Proteobacteria (Fig. S2). Some bacteria were previously identified as putative associates of diatoms and *Pseudo-nitzschia* spp. including *Sulfitobacter pseudonitzschiae*, which comprised 0.1-0.5% of the total contaminant sequences identified across species and *Marinobacter salarius*, which occurred at low abundance in most species (<5%) but dominated contaminant sequences in *P. australis* cultures, comprising more than 93% (76,814) of the tags identified as contaminants. Removal of any clusters with individual sequences exhibiting matches to contaminants yielded a final set of 71,889-113,898

read clusters by species, on par with the number of BcgI restriction sites identified in the *P. multiseries* reference genome (81,121, Table 2).

Table 2. Sequencing effort, clustering, and contaminant removal to generate 2bRAD reference libraries per species. Tags: reads trimmed to retain only the BcgI restriction site motif; HQ: high quality, defined as reads exhibiting a minimum phred quality of 20 over 100% of the read and not containing any Illumina adapters.

Species	N sequencing replicates	N HQ unique tags	N contaminant tags	N initial tag clusters	N clusters removed as contaminants	N clusters in final reference
P. australis	5	336,268	82,192	83,852	10,211	73,641
P. sp.	3	133,370	19,974	87,639	15,750	71,889
P. pungens	4	452,059	29,356	135,628	21,730	113,898
P. subpacifica	3	367,356	24,631	81,732	5,660	76,072
P. multiseries (CLN 47)	-	-	-	-	-	81,121
P. multistriata (WGS set)	-	-	-	-	-	27,881

3.2 *Accuracy and precision of reduced representation sequencing based estimates of species abundance*

The false positive rate (FPR), reflecting samples in which reads mapped to a reference species not included in that mock community mix, was low, 0.45% on average, and was unrelated to read depth (Fig. 1a). There was a significant difference in the FPR between mix types, with samples derived from multi-species pools exhibiting an 0.68% FPR on average (range: 0.045 - 4.1%) whereas the FPR of 'pure' culture samples was 0.04% on average (range: 0 - 0.1%, $t(26.15)=3.48$, $p=0.0018$, Fig. 1b). The highest false positive rates were observed in the most extreme mixes (ratios greater than 99:1, Fig. 1). False negatives were never observed. That is to say, species known to be included in the sample mix were always detected, even at very low relative abundances, e.g. 0.01% (Fig. 1).

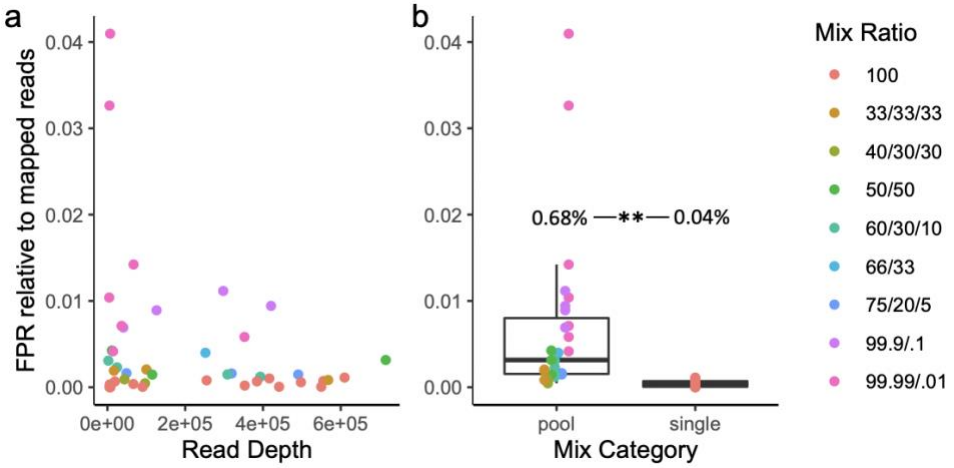

Figure 1. False positive rate (FPR) relative to high quality mapped reads as a function of read depth (a) and whether the sample consisted of a single species or multi-species pool (b). Individual samples are colored according to their mix type ratio (Table 1).

This pattern was also apparent when examining FPR by species, with significant increases in the FPR for reads mapping to *P.* sp. (t(26.1)=10.5, p=4e-11), *P. multiseries* (t(39.1)=2.5, p=0.008), *P. multistriata* (t(26)=2.89, p=0.004), and *P. pungens* (t(4.05)=4.45, p=0.005) references in multi-species mixes relative to individual species samples and similar trends for the remaining species (Fig. S3).

On average, 94% of the variance in observed species abundance was explained by the expected abundance based on the target mix ratios (R2=0.94, F(1,79)=1350, p<0.001, Fig. 2A). Expected abundances were also a strong predictor of observed abundances, and the intercept was not significantly different from zero (α=0.24, β=0.95, t(79)=36.7, p<0.001). However, fit did differ between species, with *P. australis* having a higher intercept on average than *P. pungens* and *P. subpacifica* (Fig. 2a). Accuracy was not a function of species (F(3,77)=2.45, p=0.07), although mixes with *P. subpacifica* tended to be more accurate (Fig. 2b). Accuracy did differ among mix types (F(8,72)=12.54, p=4.3e-11, Fig. 2c). More equal mixes, for example those with a ratio of 33:33:33 of three species, were less accurate than more extreme mixes, or those with a ratio >99:1 of two species (Tukey's HSD < 0.001, Fig. 2c). Accuracy also differed among culture replicates (F(8,72)=4.38, p=0.0002, Fig. 3d). For example, more error was observed for the C5 culture replicate of *P. pungens* than for C3 (Tukey's HSD=0.05, Fig. 2d).

3.3 *Abundance estimate of species in natural* Pseudo-nitzschia *spp. blooms using 2b-RAD*

The dynamics of *Pseudo-nitzschia* spp. blooms differed across years. *Pseudo-nitzschia* spp. cells were observed every year, but pDA was only detected in 2017 and 2019 (Fig. 3a-c). Although samples were sequenced for 55 out of 56 total time-points, the overall percent of high quality reads mapping to the six species reference was generally very low (Fig. 3d). Consequently, we were unable to quantify species composition in some samples due to insufficient read depth (Fig. 1e, empty columns). Superficially, increases in pDA appeared to be associated with increases in the relative abundance of *P. australis*. However, the correlation across time-points was weak (R2=0.05, t(44)=1.8, p=0.08, Fig. 3,S4). The relative abundance of *P. australis* did explain a significant portion of the variation in the percent of reads mapping to the six species reference (R2=0.69, β=0.56, t(44)=10.06, p<0.001, Fig. S5).

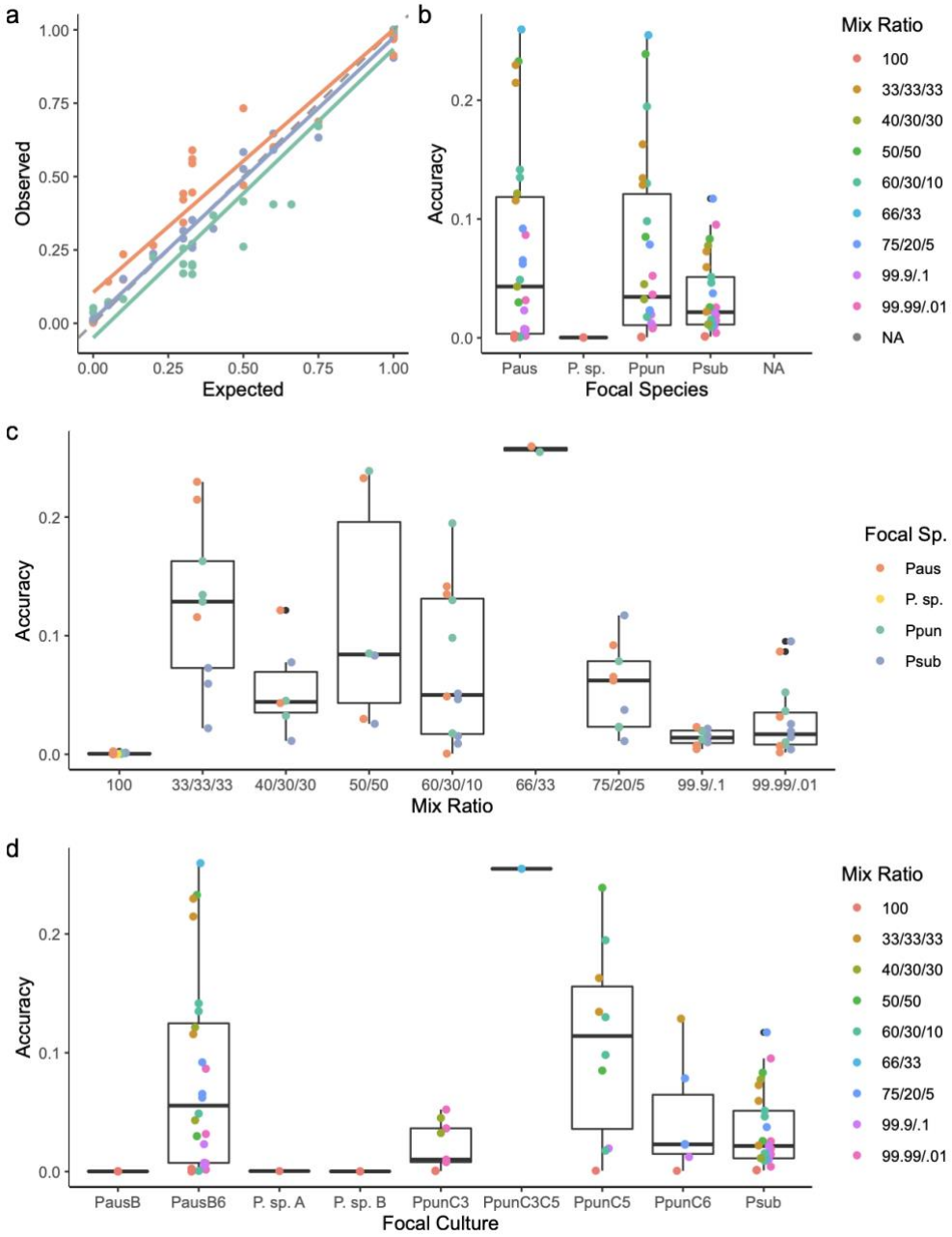

Figure 2. Accuracy of a reduced representation sequencing (2bRAD) method for quantifying species relative abundance. (a) Observed abundance, calculated as the proportion of reads mapping to the reference library, as a function of expected abundance based on the target proportion in the mock community mix varies by species. The dashed gray line is the 1:1 line. (b) Accuracy, or the deviation between observed and expected values, as a function of species and mix ratio. (c) Accuracy as a function of mix ratio and species. (d) Accuracy as a function of culture and mix ratio.

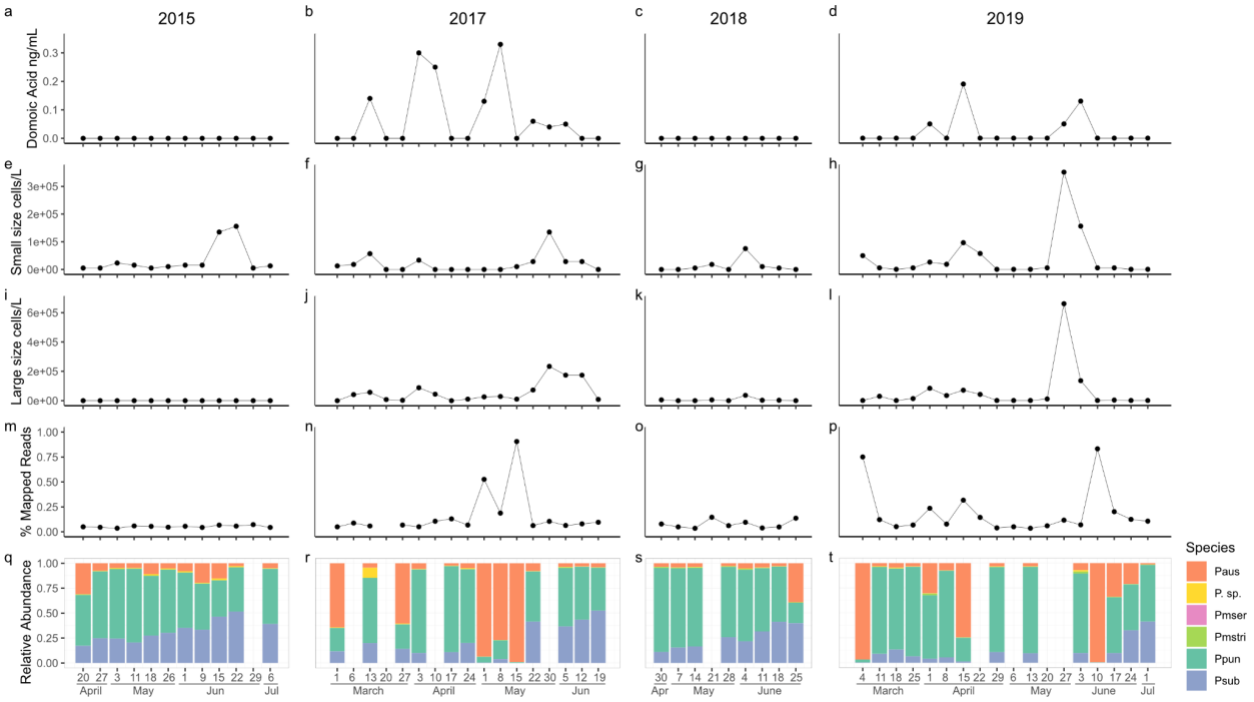

Figure 3. Dynamics of *Pseudo-nitzschia* spp. blooms sampled from Newport Beach Pier, CA. (a-d) The concentration of particulate domoic acid (ng/mL) in seawater. (e-h) The concentration of small size class (< 3 µm) cells per liter. (i-l) The concentration of large size class (> 3 µm) cells per liter. (m-p) The percent of high quality reads exhibiting high quality mapping to the *Pseudo-nitzschia* spp. reference library. (q-t) The relative abundance of *Pseudo-nitzschia* spp. calculated as the proportion of high quality reads exhibiting high quality mapping to each individual species relative to the total number of reads mapping to the *Pseudo-nitzschia* spp. reference library.

3.4 *Abundance of species in natural* Pseudo-nitzschia *spp. blooms using ARISA*

Seventeen distinctly-sized amplicons were observed with ARISA (Fig. 4). Ten were associated with species previously reported along the US Pacific Coast based on reference sequences available in GenBank and prior studies using ARISA (Carlson et al., 2016; Hubbard et al., 2014, 2008; Smith et al., 2018b). Half were putatively identified based on prior observations as *P. australis/P. seriata* (which share the same fragment size, 150 base pairs [bp]), *P. inflatula* (156 bp), *P. delicatissima* (168 bp), *P. fraudulenta* (203 bp), and *P. fryxelliana* (207 bp). The other half were associated with species based on prior observations and direct sequencing of PCR products as part of the present study: *P. galaxiae* (140 bp), *P. pungens* (142 bp), *P. multiseries* (144 bp), *P. subpacifica* (196 bp) and *P. cuspidata* (233 bp); all were 100% similar to previously observed sequences. The sequence obtained for *P. pungens* corresponded to *P. pungens* var. *cingulata.* Seven additional amplicons of unknown identity were characterized by ARISA only (170 bp, 177 bp, 200 bp, 210 bp, 213 bp, 219 bp, and 223/224 bp). A few samples (n=3) with low total peak height were excluded from analysis (indicated by * in Fig. 4).

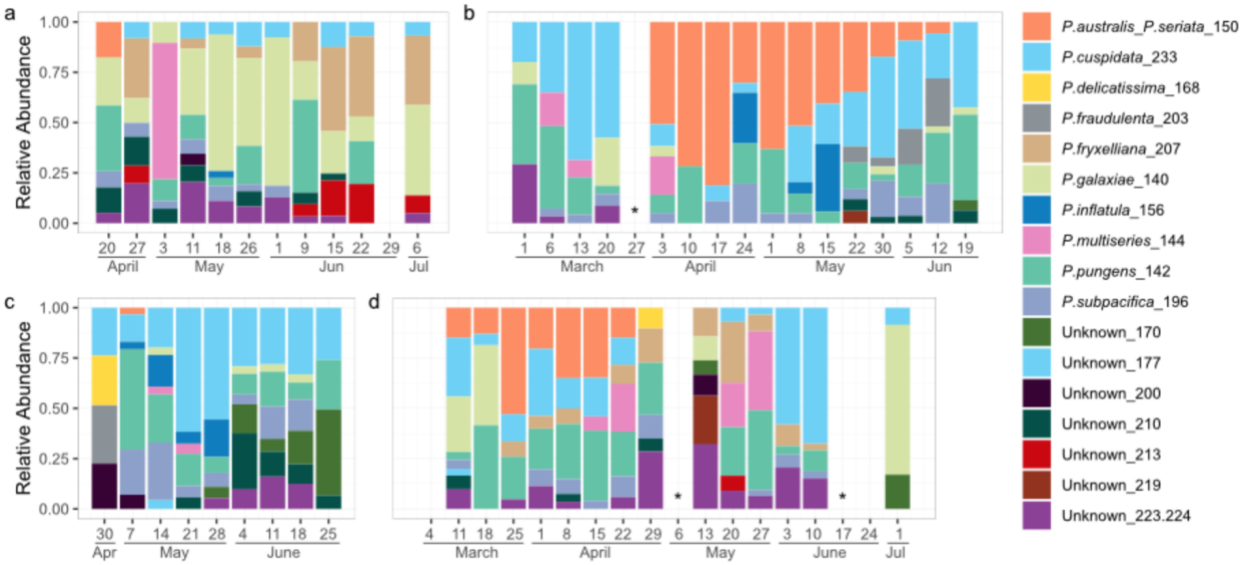

Figure 4. ARISA-based diversity profile of *Pseudo-nitzschia* spp. blooms sampled from Newport Beach Pier, CA in (a) 2015, (b) 2017, (c) 2018, and (d) 2019. * samples with low total peak height.

The *P. australis/P. seriata* fragment was observed in spring in all four years but the duration varied from year to year: 2015 (April, one week), 2017 (April-June, 11 weeks), 2018 (May, one week), and 2019 (March through April, seven weeks, Fig. 4). *Pseudo-nitzschia* cellular abundances and pDA concentrations generally tracked well with *P. australis/P. seriata*. For example, in 2015 and 2018, cellular abundances and pDA concentrations were minimal, and increases in the former only were noted after the single observations of *P. australis*/*P. seriata.* In contrast, in 2017 and 2019, cellular abundances and pDA concentrations were elevated and dynamic for the time frame when the *P. australis/P. seriata* amplicon was present, and even in a few samples where this amplicon was absent but other toxic species like *P. cuspidata* were present (Fig. 3,4). *P. cuspidata* was observed in nearly every sample, but appeared to be quite dynamic based on the relativized ARISA signal, and was detected before, during, and after *P. australis/P. seriata* in all years except 2015, the year that *P. cuspidata* generally appeared to comprise a smaller proportion of the *Pseudo-nitzschia* assemblage and when *P. galaxie* was more prevalent. *P. multiseries* is another species recognized as producing high pDA levels and was detected in all years, but was never the dominant taxon in the ARISA. *P. multiseries* was observed after *P. australis* in 2015, before and with *P. australis* in 2017, after *P. australis* in 2018, and with and after *P. australis* in 2019.

Although both methods identified an increase in *P. australis* in the high domoic acid years (2017 and 2019), correlations between relative abundance estimates for individual species obtained using the different methods were weak or nonexistent. A significant correlation was detected between the relative abundance of *P. australis* estimated via 2bRAD and the ARISA-based estimate of *P. australis/P. seriata*, but the overall variance explained was less than 10% (R2=0.096, t(40)=2.3, p=0.03, Fig. S6). No relationships were detected for any of the other species that could be compared between methods (*P.* sp., *P. pungens*, *P. multiseries*, *P.*

subpacifica) nor were any relationships evident when converting relative abundance to simple presence/absence calls.

3.6 *Population-level variation in* Pseudo-nitzschia *spp. over time*

The low percentage of reads mapping to the *Pseudo-nitzschia* spp. reference in general (Fig. 3) resulted in insufficient coverage for SNP calling in individual samples. For example, even for *P. pungens*, the most abundant species across sampling time-points and years, 75% of samples had fewer than 7 tags with at least 80x coverage, the minimum recommended threshold for analysis of pooled samples (Schlötterer et al., 2014) (Fig. S7). Therefore, to assess the potential for the 2bRAD method to yield information on population-level variation, we pooled samples by year and compared changes in allele frequency between low domoic acid (DA, 2015 and 2018) and high domoic acid years (2017 and 2019, Fig. 3) following (Bendall et al., 2016).

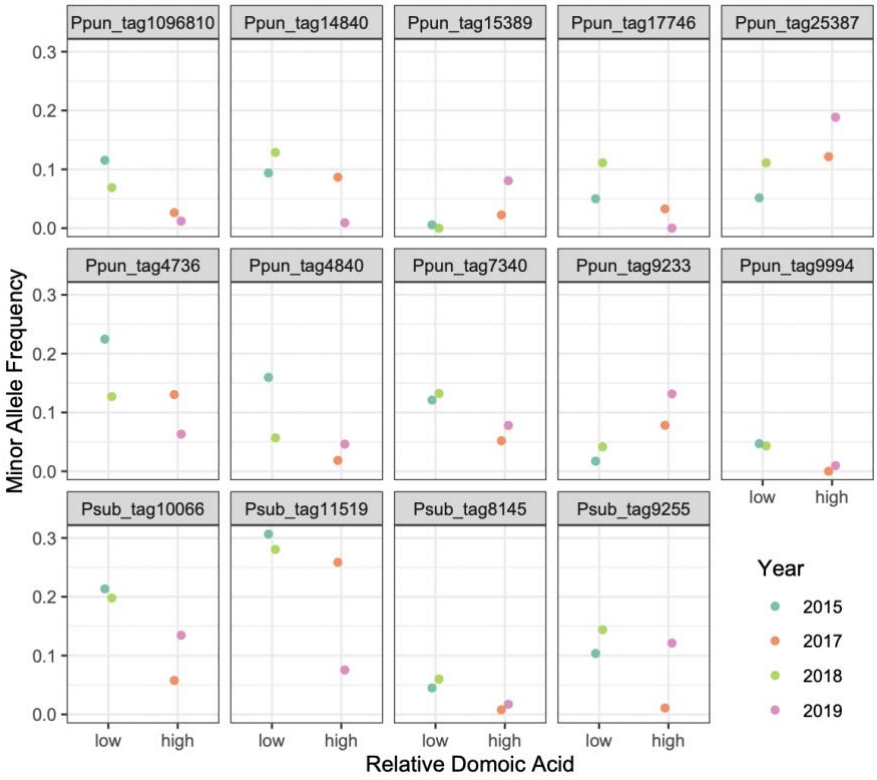

Figure 5. Minor allele frequency at select loci (CMH<0.05) by sampling year, stratified by the relative amount of domoic acid detected over the course of the sampling window. Ppun = *P. pungens*, Psub = *P. subpacifica*.

The relative abundances of *P. delicatissima*, *P. multiseries*, and *P. multistriata* were too low to conduct population-level analyses with confidence (Fig. 3). No consistent changes in allele frequencies were detected in *P. australis* populations over time. We identified 17 SNPs in *P. pungens* and 4 SNPs in *P. subpacifica* that showed differences in allele frequencies across years (CMH<0.05). Of these, 10 in *P. pungens* and all of the *P. subpacifica* SNPs exhibited consistent changes between years with low and high pDA concentrations. In *P. subpacifica*, the

minor allele frequency (MAF) at all loci consistently decreased in years where pDA was observed whereas in *P. pungens*, MAF decreased at 7 loci and increased at 3 loci (Fig. 5, Table S3).

4. DISCUSSION

In culture, the toxicity of *Pseudo-nitzschia* is influenced by both abiotic (Trainer et al., 2012) and biotic (Sison-Mangus et al., 2014) factors. In the Southern California Bight, while large-scale drivers of domoic acid outbreaks have been identified, the factors contributing to interannual variability in toxin production remain unresolved (Smith et al., 2018a). Inter- and intraspecies level variation in *Pseudo-nitzschia* is thought to play a key role in DA production (Clark et al., 2019; Fernandes et al., 2014; Guannel et al., 2015; Hubbard et al., 2014; Smith et al., 2018b). Here we show that a reduced representation based sequencing approach can be used to accurately quantify the relative abundance of *Pseudo-nitzschia* spp. in a mixed community. In addition, we were able to assess population-level genomic variation within focal *Pseudo-nitzschia* spp. using the same dataset. This method presents a potentially cost-effective approach for large-scale studies aiming to evaluate population and community dynamics in mixed samples, although additional development work is needed.

4.1. *The dynamics of Pseudo-nitzschia blooms at Newport Beach Pier, CA*.

Analysis of a field sample set over high and low domoic acid years suggest a role for *P. australis* in DA production. However, the direct linear relationship between the relative abundance of *P. australis* as estimated through 2bRAD and pDA concentrations is weak, although corroborated by ARISA, suggesting that the toxicity of *P. australis* is not a linear function of its abundance, or that other members of the *Pseudo-nitzschia* spp. community or variability in their abundances may be involved. The parallel ARISA dataset generated for the same samples supports this latter explanation: in addition to the high prevalence of *P. australis*/*P. seriata*, additional toxigenic or suspected toxigenic species such as *P. cuspidata, P. pungens, P. multiseries, P. fraudulenta, P. delicatissima, P. subpacifica and P. galaxiae* were also identified. An earlier ARISA based analysis of relative species abundance also identified *P. australis/P. seriata* in association with a strong DA event in 2013, but these species were absent in a similar bloom the following year (Smith et al., 2018b). ARISA cannot distinguish *P. australis* from *P. seriata* on the US west coast (Hubbard et al., 2008). Although *P. seriata* was not included in our reference library, the 2bRAD method is highly specific in terms of read recruitment and false positive rate, and given the strength of read recruitment to the *P. australis* reference when this species is dominant (Fig. S4), it is likely that *P. australis* rather than *P. seriata* is the dominant member of the *Pseudo-nitzschia* spp. assemblage for the years analyzed here, as false positives between sister species using the 2bRAD method, such as *P. pungens* and *P. multiseries*, are minimal (~0.005%). If we can interpret the overall percent mapping as a reflection of the total composition (Fig. 3m-p), both for species within our *Pseudo-nitzschia* reference set and others, then *P. australis* may indeed dominate blooms in this region at certain sampling times and is often dominant when DA is detected. Indeed, both ARISA and 2bRAD detected dynamic assemblages over time and a coincident increase in *P. australis* as

573 DA increased. However, an expansion of the 2bRAD reference library set and additional work
574 will enable this supposition to be tested.
575
576 Interestingly, we also identified apparent cyclical shifts in allele frequency in both *P. pungens*
577 and *P. subpacifica* populations wherein the minor allele frequency increased or decreased
578 corresponding with high vs low DA years (Fig. 5). Prior temporal analyses of population level
579 variation in SNPs from natural bacterial populations have identified unidirectional shifts in allele
580 frequency over time (Bendall et al., 2016). Population genetic analysis of *Pseudo-nitzschia* spp.
581 has been limited to microsatellite-based characterization (Casteleyn et al., 2010). Globally
582 sampled populations of *P. pungens* were found to exhibit strong isolation by distance, however,
583 time-series samples were pooled by region (Casteleyn et al., 2010), potentially masking relevant
584 temporal shifts which have been repeatedly identified in other marine diatoms (Godhe and
585 Härnström, 2010; Rynearson et al., 2006; Whittaker and Rynearson, 2017). Interestingly,
586 morphologically and genetically distinct *P. pungens* varieties have been identified along the US
587 West Coast (Carlson et al., 2016; Hubbard et al., 2014, 2008; Villac and Fryxell, 1998). Here we
588 show that not only do allele frequencies show temporal shifts, in some instances these shifts
589 correlate to some extent with the presence of DA (Fig. 5). It is unlikely that the SNPs identified
590 here are causal, in the sense that they are responsible for DA production, but it is possible that
591 certain *Pseudo-nitzschia* spp. populations are more associated with DA presence and/or
592 production than others. The presence of competitors can alter both ecological and evolutionary
593 outcomes (Agrawal et al., 2012). A competitive role for DA has been proposed, but the
594 mechanism of action remains debated. DA has been hypothesized to play a role in iron uptake
595 (Maldonado et al., 2002; Rue and Bruland, 2001; Wells et al., 2005). DA production has also
596 been shown to facilitate the competitive ability of *P. delicatissima* in co-culture with *S. marinoi*
597 (Prince et al., 2013) supporting an allelopathic role (Xu et al., 2015). Alternatively, the change in
598 allele frequency may be the result of other co-occurring biotic or abiotic drivers (including ocean
599 circulation) that influence DA production but also exert a selective force on other species.
600
601 Whatever the reason, greater temporal and spatial resolution will help address these
602 hypotheses. This can be accomplished by an increase in sequencing depth to obtain sufficient
603 coverage for calling SNPs in an individual sample, which will increase per-sample sequencing
604 costs. If the goal is to assess population level variation, it may be possible to enrich for the
605 eukaryotic fraction by targeting not DNA, but RNA. Use of poly-A priming in the cDNA synthesis
606 reaction can then be used to enrich for eukaryotes. This should not distort the relative
607 abundance ratios of target taxa as the RNA to cDNA conversion does not involve any
608 amplification, however, additional genotyping errors could be introduced during the conversion
609 that could contribute to null alleles. Paired comparison of RNA and DNA based abundance
610 estimates will be needed to evaluate the potential effectiveness of this modification.
611
612 4.2 *Detection limit is driven by the false positive rate*
613 The false positive rate greatly exceeded the false negative rate, but both were low on average,
614 with less than 1% of reads identified as false positives in the mock community mix experiment.
615 Although the method was able to detect the presence of species at very low abundances (tested
616 down to 0.01%), an average false positive rate of 0.45% suggests that the detection limit for

calling true positives should be set at 0.5-1%. Consequently, in the natural bloom sample set, although there were reads mapping to all species (Fig. 3q-t) only *P. australis, P. pungens*, and *P. subpacifica* were present at detectable levels. Interestingly, far fewer false positives were observed for reads mapping to the *P. multiseries* and *P. multistriata* reference species, which were *in silico* derived by extracting BcgI restriction sites from previously sequenced genomes. This despite *P. pungens* and *P. multiseries* being sibling species, which should increase overlap due to greater genetic similarity (Lim et al., 2018; Manhart et al., 1995). This suggests that some aspect of the culturing process may have contributed to artificially inflating the false positive rate. One explanation may be unintentional cross-contamination during the process of creating artificial community mixes, although the use of sterile technique makes this unlikely. Alternatively, false positives could still be the result of additional unknown contaminants (prokaryotic or eukaryotic), representatives of which have not yet been incorporated into the NCBI database. As 2b-RAD tags are only 36-bp in length, single base pair differences can be the difference between a match and a 'no hits' outcome. While we were able to identify and exclude the most contaminants using a global clustering approach, the vast majority of microbial diversity remains underexplored (Salazar and Sunagawa, 2017), and consequently cannot serve as a homology-based reference for filtering out contaminant reads. An alternative approach to excluding contaminants could be to focus not on homology based searches, which require a pre-existing reference database, but to use clustering algorithms, as have been developed for binning metagenomic sequence datasets using, for example tetranucleotide frequency and percent GC content (Graham et al., 2017). Future work should aim to explore if such alternative approaches can be applied to short read RAD datasets.

Accuracy should also be taken into consideration when setting detection limits. Cell size was not measured, but would be expected to vary considerably, along with genome size, across the different species utilized herein (Hubbard et al., 2014). It is possible that these differences contribute to variability in the expected vs. observed signal. In the mock community mix experiment, we observed significant variation in accuracy among replicates. Perhaps counter-intuitively, more equal mixes (e.g. ratios of 50:50) exhibited lower accuracy than more extreme mixes (e.g. ratios of 99.9:0.1) and pure cultures. In addition, accuracy varied significantly among replicate cultures of the same species. Taken together, this suggests that the variation in accuracy is likely driven by inaccuracies during mix creation rather than errors in the assignment of reads during mapping or systematic biases among species. As mixes were generated using replicate Sedgwick Rafter-based counts of cell densities in culture replicates, any inaccuracy during mix creation would propagate in proportion to the final volume of culture added to each sample mix. More equal mixes necessitated mixing of larger volumes for each individual culture, increasing the chance for variation. As it will be impossible in field samples to quantify the absolute abundance of any given focal species without a universal reference library, and given the strong association between observed and expected abundance (Fig. 2a), so long as the organism passes the minimum detection limit, then our results suggest its relative abundance within the focal community can be reliably assessed.

4.3. *Low agreement of relative abundance estimates across methods*.

Both methods implicate *P. australis* as a driver of DA production in the natural sample set (Fig. 3q-t, Fig. 4) reinforcing earlier reports from this region *(Schnetzer et al., 2013; Smith et al., 2018b)*. However, direct comparison of the relative abundance estimates obtained using the different methods and even simple presence/absence determinations largely disagree. This discrepancy may be attributable to different methodological biases. After excluding samples below the 1% detection limit, 2bRAD still identified low abundances of *P. australis* (<5%) at almost all time-points (Fig. 3q-t) which were not evident in the ARISA analysis (Fig. 4). One explanation is false negatives in the ARISA analysis, as this method applies a 3% peak threshold. Lowering the ARISA limit of detection to 1% results in detection of *P. australis* in four additional samples, but a lowered threshold is not recommended without further quantitative validation. Alternatively, we cannot rule out the possibility that these low abundance detections are false positives in the 2bRAD analysis, as this method has a tendency to overestimate the abundance of species when rare, which is particularly evident in *P. australis* (Fig. 2a). Low agreement between methods is likely also driven by the limited diversity in the current 2bRAD reference library. The 2bRAD relative abundance estimates of all species will be artificially inflated by the inability to capture all the relevant diversity in *Pseudo-nitzschia* spp. Expansion of the reference library set to include additional relevant taxa for the region will help remedy this problem and shed more light on the ecology of this genus.

4.4 *Proteobacteria*: *contaminants or symbionts?*

Interactions between phytoplankton and bacteria are critical for nutrient cycling in aquatic environments, but recent work suggests that microscale interactions in the phycosphere, such as the exchange of metabolites, extend beyond general food web interactions into the realm of mutualism (Amin et al., 2012; Seymour et al., 2017). The essential relationship between diatoms and their associated bacteria may be why it is so difficult to obtain axenic phytoplankton cultures (Töpel et al., 2019). Bacteria have also been shown to enhance toxin production in *P. multiseries* cultures (Bates et al., 1995) suggesting a key role for other microbes in the ecology of *Pseudo-nitzschia* spp. We identified a substantial community of *Proteobacteria* in our unialgal *Pseudo-nitzschia* spp. cultures, the profiles of which differed among species. All cultures contained representatives of taxa previously identified as putative associates of diatoms and *Pseudo-nitzschia* spp. including *Sulfitobacter pseudonitzschiae* and *Marinobacter salarius* (Hong et al., 2015; Johansson et al., 2019) which occurred at low abundance (<5%) in most species. However, *P. australis* exhibited the highest level of 'contamination' with over 24% of the total RAD tags identified as *Proteobacteria* (Fig. S2) and of these contaminant tags, more than 93% exhibited blast matches to *M. salarius* in NCBI's nt database. *M. salarius* has been shown to stimulate the growth of the diatom *Skeletonema marinoi* in culture (Johansson et al., 2019). Genomic analysis suggests that *M. salarius* may produce a growth factor in addition to siderophores, which may increase iron availability for its host diatom (Töpel et al., 2019). Future culture-based work should aim to test whether *M. salarius* fulfills a similar role for *P. australis*. From a methods development perspective, since 2bRAD captures both prokaryotic and eukaryotic tags, an interesting next step would be to expand the reference to include key prokaryotic taxa to assess the change in relative abundance of putative proteobacterial symbionts in addition to target *Pseudo-nitzschia* spp. over the course of natural bloom events.

4.5 Conclusions

Taken together, our results show that a reduced representation based sequencing approach can both quantify the relative abundance of *Pseudo-nitzschia* spp. in a mixed community and assess population-level genomic variation within species using the same dataset. Advantages of this method include simplicity of the library preparation protocol and cost-effectiveness (Puritz et al., 2014). Depending on the level of coverage desired, costs for both library preparation and sequencing range from ~$10-$40 a sample, whereas full metagenomic libraries, which also require additional computational power to analyze, are on the order of $100 a sample. While other RAD methods could in principle be used, we advocate for 2bRAD for community level analyses as the fixed size of tags precludes the need for genome size corrections that would be necessary with other methods (Puritz et al., 2014), given that abundance is estimated from the total number of mapped reads, which is proportional to the per-taxa sequencing effort, rather than the mean coverage per tag. There are, however, some limitations that remain to be overcome. New species cannot be identified *a priori* in a mixed community. A RAD tag reference must first be generated. Culturing is a rate-limiting step for many methods, but with advances in single cell sequencing (Gawad et al., 2016), it may be possible to skip this step altogether in the near future in favor of cell sorting and direct sequencing (Cuvelier et al., 2010; Marie et al., 2017). By mining existing metagenomic datasets and other genome databases it may be possible to create a large-scale reference library that includes prokaryotes, archaea, and eukaryotes, to competitively recruit RAD tags from natural samples, thereby greatly expanding the taxonomic resolution of the method. Combined with the ability to secondarily assess population-scale variation for the most abundant members of the community, this method has the potential to greatly expand the scope of large-scale spatial and temporal monitoring studies, warranting additional development.

Acknowledgements

We thank Carmelo Tomas for assistance in identification of *Pseudo-nitzschia* strains.

Funding: This publication was supported by NOAA Grant #NOA18OAR4170073, California Sea Grant College Program Project #115697431, through NOAA'S National Sea Grant College Program, U.S. Dept. of Commerce. The statements, findings, conclusions and recommendations are those of the author(s) and do not necessarily reflect the views of California Sea Grant, NOAA, the U.S. Environmental Protection Agency, or the U.S. Dept. of Commerce. ARISA analysis was supported via National Science Foundation Grant Number OCE-1840381, the National Institute of Environmental Health Sciences Grant Number 1P01ES028938, and the Woods Hole Center for Oceans and Human Health.

Data Accessibility: FASTQ reads for both the Mock Community Mix and Natural Sample libraries can be obtained from NCBI's SRA under BioProject PRJNA749297. The final *Pseudo-nitzschia* spp. 2bRAD reference library is hosted at https://dornsife.usc.edu/labs/carlslab/data/. Bioinformatic and statistical scripts necessary to re-create analyses, as well as raw input data used to generate figures used in this study are available at https://github.com/ckenkel/Pseudo-nitzschia2bRAD.

References

Adl, S.M., Leander, B.S., Simpson, A.G.B., Archibald, J.M., Anderson, O.R., Bass, D., Bowser, S.S., Brugerolle, G., Farmer, M.A., Karpov, S., Kolisko, M., Lane, C.E., Lodge, D.J., Mann, D.G., Meisterfeld, R., Mendoza, L., Moestrup, Ø., Mozley-Standridge, S.E., Smirnov, A.V., Spiegel, F., Collins, T., Sullivan, J., 2007. Diversity, Nomenclature, and Taxonomy of Protists. Syst. Biol. 56, 684–689.

Agrawal, A.A., Hastings, A.P., Johnson, M.T.J., Maron, J.L., Salminen, J.-P., 2012. Insect herbivores drive real-time ecological and evolutionary change in plant populations. Science 338, 113–116.

Alverson, A.J., 2008. Molecular Systematics and the Diatom Species. Protist 159, 339–353.

Amin, S.A., Parker, M.S., Armbrust, E.V., 2012. Interactions between diatoms and bacteria. Microbiol. Mol. Biol. Rev. 76, 667–684.

Anderson, C.R., Kudela, R.M., Kahru, M., Chao, Y., Rosenfeld, L.K., Bahr, F.L., Anderson, D.M., Norris, T.A., 2016. Initial skill assessment of the California Harmful Algae Risk Mapping (C-HARM) system. Harmful Algae 59, 1–18.

Anderson, D.M., Burkholder, J.M., Cochlan, W.P., Glibert, P.M., Gobler, C.J., Heil, C.A., Kudela, R.M., Parsons, M.L., Rensel, J.E.J., Townsend, D.W., Trainer, V.L., Vargo, G.A., 2008. Harmful algal blooms and eutrophication: Examining linkages from selected coastal regions of the United States. Harmful Algae 8, 39–53.

Assaf, G., Hannon, G.J., 2010. FASTX-toolkit. FASTX-Toolkit.

Bates, S.S., Douglas, D.J., Doucette, G.J., Léger, C., 1995. Enhancement of domoic acid production by reintroducing bacteria to axenic cultures of the diatom Pseudo-nitzschia multiseries. Nat. Toxins 3, 428–435.

Bates, S.S., Hubbard, K.A., Lundholm, N., Montresor, M., Leaw, C.P., 2018. Pseudo-nitzschia, Nitzschia, and domoic acid: New research since 2011. Harmful Algae 79, 3–43.

Bejarano, A.C., VanDola, F.M., Gulland, F.M., Rowles, T.K., Schwacke, L.H., 2008. Production and Toxicity of the Marine Biotoxin Domoic Acid and Its Effects on Wildlife: A Review. Human and Ecological Risk Assessment: An International Journal 14, 544–567.

Bendall, M.L., Stevens, S.L., Chan, L.-K., Malfatti, S., Schwientek, P., Tremblay, J., Schackwitz, W., Martin, J., Pati, A., Bushnell, B., Froula, J., Kang, D., Tringe, S.G., Bertilsson, S., Moran, M.A., Shade, A., Newton, R.J., McMahon, K.D., Malmstrom, R.R., 2016. Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. ISME J. 10, 1589–1601.

Bornet, B., Antoine, E., Françoise, S., Baut, C.M., 2005. Development of sequence characterized amplified region markers from intersimple sequence repeat fingerprints for the molecular detection of toxic phytoplankton alexandrium catenella (dinophyceae) and pseudo-nitzschia pseudodelicatissima (bacillariophyceae) from french coastal waters1. J. Phycol. 41, 704–711.

Bruin, A. de, Ibelings, B.W., Van Donk, E., 2003. Molecular techniques in phytoplankton research: from allozyme electrophoresis to genomics. Hydrobiologia 491, 47–63.

Bushnell, B., 2014. BBMap: A Fast, Accurate, Splice-Aware Aligner (No. LBNL-7065E). Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States).

Carlson, M.C.G., McCary, N.D., Leach, T.S., Rocap, G., 2016. Pseudo-nitzschia Challenged with Co-occurring Viral Communities Display Diverse Infection Phenotypes. Front. Microbiol. 7, 527.

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. U. S. A. 107, 12952–12957.

Clark, S., Hubbard, K.A., Anderson, D.M., McGillicuddy, D.J., Ralston, D.K., Townsend, D.W., 2019. Pseudo-nitzschia bloom dynamics in the Gulf of Maine: 2012–2016. Harmful Algae

88, 101656.
Cochlan, W.P., Herndon, J., Kudela, R.M., 2008. Inorganic and organic nitrogen uptake by the toxigenic diatom Pseudo-nitzschia australis (Bacillariophyceae). Harmful Algae 8, 111–118.
Countway, P.D., Gast, R.J., Savai, P., Caron, D.A., 2005. Protistan diversity estimates based on 18S rDNA from seawater incubations in the Western North Atlantic. J. Eukaryot. Microbiol. 52, 95–106.
Cuvelier, M.L., Allen, A.E., Monier, A., McCrow, J.P., Messié, M., Tringe, S.G., Woyke, T., Welsh, R.M., Ishoey, T., Lee, J.-H., Binder, B.J., DuPont, C.L., Latasa, M., Guigand, C., Buck, K.R., Hilton, J., Thiagarajan, M., Caler, E., Read, B., Lasken, R.S., Chavez, F.P., Worden, A.Z., 2010. Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton. Proc. Natl. Acad. Sci. U. S. A. 107, 14679–14684.
Davey, J.W., Blaxter, M.L., 2010. RADSeq: next-generation population genetics. Brief. Funct. Genomics 9, 416–423.
Evans, K.M., Bates, S.S., Medlin, L.K., Hayes, P.K., 2004. Microsatellite marker development and genetic variation in the toxic marine diatom pseudo-Nitzschia multiseries (Bacillariophyceae)1. J. Phycol. 40, 911–920.
Evans, K.M., Hayes, P.K., 2004. Microsatellite markers for the cosmopolitan marine diatom Pseudo-nitzschia pungens. Mol. Ecol. Notes 4, 125–126.
Fehling, J., Davidson, K., Bolch, C.J., Bates, S.S., 2004. Growth and domoic acid production by pseudo-nitzschia seriata (bacillariophyceae) under phosphate and silicate limitation. J. Phycol. 40, 674–683.
Fernandes, L.F., Hubbard, K.A., Richlen, M.L., Smith, J., Bates, S.S., Ehrman, J., Léger, C., Mafra, L.L., Kulis, D., Quilliam, M., Libera, K., McCauley, L., Anderson, D.M., 2014. Diversity and toxicity of the diatom Pseudo-nitzschia Peragallo in the Gulf of Maine, Northwestern Atlantic Ocean. Deep Sea Res. Part 2 Top. Stud. Oceanogr. 103, 139–162.
Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W., 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28, 3150–3152.
Gawad, C., Koh, W., Quake, S.R., 2016. Single-cell genome sequencing: current state of the science. Nat. Rev. Genet. 17, 175–188.
Godhe, A., Härnström, K., 2010. Linking the planktonic and benthic habitat: genetic structure of the marine diatom Skeletonema marinoi. Mol. Ecol. 19, 4478–4490.
Graham, E.D., Heidelberg, J.F., Tully, B.J., 2017. BinSanity: unsupervised clustering of environmental microbial assemblies using coverage and affinity propagation. PeerJ 5, e3035.
Guannel, M.L., Haring, D., Twiner, M.J., Wang, Z., Noble, A.E., Lee, P.A., Saito, M.A., Rocap, G., 2015. Toxigenicity and biogeography of the diatom Pseudo-nitzschia across distinct environmental regimes in the South Atlantic Ocean. Mar. Ecol. Prog. Ser. 526, 67–87.
Hasle, G.R., Syvertsen, E.E., 1997. Chapter 2 - Marine Diatoms, in: Tomas, C.R. (Ed.), Identifying Marine Phytoplankton. Academic Press, San Diego, pp. 5–385.
Hohenlohe, P.A., Bassham, S., Etter, P.D., Stiffler, N., Johnson, E.A., Cresko, W.A., 2010. Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. PLoS Genet. 6, e1000862.
Hong, Z., Lai, Q., Luo, Q., Jiang, S., Zhu, R., Liang, J., Gao, Y., 2015. Sulfitobacter pseudonitzschiae sp. nov., isolated from the toxic marine diatom Pseudo-nitzschia multiseries. Int. J. Syst. Evol. Microbiol. 65, 95–100.
Howard, M.D.A., Cochlan, W.P., Ladizinsky, N., Kudela, R.M., 2007. Nitrogenous preference of toxigenic Pseudo-nitzschia australis (Bacillariophyceae) from field and laboratory experiments. Harmful Algae 6, 206–217.
Hubbard, K.A., Olson, C.H., Armbrust, E.V., 2014. Molecular characterization of Pseudo-nitzschia community structure and species ecology in a hydrographically complex estuarine system (Puget Sound, Washington, USA). Mar. Ecol. Prog. Ser. 507, 39–55.

Hubbard, K.A., Rocap, G., Armbrust, V.E., 2008. Inter- and intraspecific community structure within the diatom genus pseudo-nitzschia (bacillariophyceae). J. Phycol. 44, 637–649.

Johansson, O.N., Pinder, M.I.M., Ohlsson, F., Egardt, J., Töpel, M., Clarke, A.K., 2019. Friends With Benefits: Exploring the Phycosphere of the Marine Diatom Skeletonema marinoi. Front. Microbiol. 10, 1828.

Kofler, R., Pandey, R.V., Schlötterer, C., 2011. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics 27, 3435–3436.

Kudela, R.M., Bickel, A., Carter, M.L., Howard, M.D.A., Rosenfeld, L., 2015. Chapter 5 - The Monitoring of Harmful Algal Blooms through Ocean Observing: The Development of the California Harmful Algal Bloom Monitoring and Alert Program, in: Liu, Y., Kerkering, H., Weisberg, R.H. (Eds.), Coastal Ocean Observing Systems. Academic Press, Boston, pp. 58–75.

Kudela, R.M., Cochlan, W.P., Roberts, A., 2002. Spatial and temporal patterns of Pseudo-nitzschia species in central California related to regional oceanography, in: Harmful Algae. Florida Fish and Wildlife Conservation Commission, Florida Institute of Oceanography, and Intergovernmental Oceanographic Commission of UNESCO, St. Petersburg, FL.

Kudela, R.M., Lane, J.Q., Cochlan, W.P., 2008. The potential role of anthropogenically derived nitrogen in the growth of harmful algae in California, USA. Harmful Algae 8, 103–110.

Kvitek, R.G., Goldberg, J.D., Smith, G.J., Doucette, G.J., Silver, M.W., 2008. Domoic acid contamination within eight representative species from the benthic food web of Monterey Bay, California, USA. Mar. Ecol. Prog. Ser. 367, 35–47.

Langlois, G., Zubkousky-White, V., Christen, J., Rankin, S., 2014. Marine Biotoxin Monitoring Program Annual Report. California Department of Public Health.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359.

Lelong, A., Hégaret, H., Soudant, P., Bates, S.S., 2012. Pseudo-nitzschia (Bacillariophyceae) species, domoic acid and amnesic shellfish poisoning: revisiting previous paradigms. Phycologia 51, 168–216.

Lim, H.C., Lim, P.T., Teng, S.T., Bates, S.S., Leaw, C.P., 2014. Genetic structure of Pseudo-nitzschia pungens (Bacillariophyceae) populations: Implications of a global diversification of the diatom. Harmful Algae 37, 142–152.

Lim, H.C., Tan, S.N., Teng, S.T., Lundholm, N., Orive, E., David, H., Quijano-Scheggia, S., Leong, S.C.Y., Wolf, M., Bates, S.S., Lim, P.T., Leaw, C.P., 2018. Phylogeny and species delineation in the marine diatom Pseudo-nitzschia (Bacillariophyta) using cox1, LSU, and ITS2 rRNA genes: A perspective in character evolution. J. Phycol. 54, 234–248.

Li, W., Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22, 1658–1659.

Lynch, M., Milligan, B.G., 1994. Analysis of population genetic structure with RAPD markers. Mol. Ecol. 3, 91–99.

Maldonado, M.T., Hughes, M.P., Rue, E.L., Wells, M.L., 2002. The effect of Fe and Cu on growth and domoic acid production byPseudo-nitzschia multiseriesandPseudo-nitzschia australis. Limnol. Oceanogr. 47, 515–526.

Manhart, J.R., Fryxell, G.A., Villac, M.C., Segura, L.Y., 1995. Pseudo-Nitzschia pungens and p. Multiseries (Bacillariophyceae): Nuclear ribosomal dnas and species differences1. J. Phycol. 31, 421–427.

Marie, D., Le Gall, F., Edern, R., Gourvil, P., Vaulot, D., 2017. Improvement of phytoplankton culture isolation using single cell sorting by flow cytometry. J. Phycol. 53, 271–282.

Matz, M.V., 2018. Fantastic Beasts and How To Sequence Them: Ecological Genomics for Obscure Model Organisms. Trends Genet. 34, 121–132.

Moore, S.K., Dreyer, S.J., Ekstrom, J.A., Moore, K., Norman, K., Klinger, T., Allison, E.H.,

Jardine, S.L., 2020. Harmful algal blooms and coastal communities: Socioeconomic impacts and actions taken to cope with the 2015 U.S. West Coast domoic acid event. Harmful Algae 96, 101799.

Moriarty, M.E., Tinker, M.T., Miller, M.A., Tomoleoni, J.A., Staedler, M.M., Fujii, J.A., Batac, F.I., Dodd, E.M., Kudela, R.M., Zubkousky-White, V., Johnson, C.K., 2021. Exposure to domoic acid is an ecological driver of cardiac disease in southern sea otters☆. Harmful Algae 101, 101973.

Orsini, L., Procaccini, G., Sarno, D., Montresor, M., 2004. Multiple rDNA ITS-types within the diatom Pseudo-nitzschia delicatissima (Bacillariophyceae) and their relative abundances across a spring bloom in the Gulf of Naples. Mar. Ecol. Prog. Ser. 271, 87–98.

Osuna-Cruz, C.M., Bilcke, G., Vancaester, E., De Decker, S., Bones, A.M., Winge, P., Poulsen, N., Bulankova, P., Verhelst, B., Audoor, S., Belisova, D., Pargana, A., Russo, M., Stock, F., Cirri, E., Brembu, T., Pohnert, G., Piganeau, G., Ferrante, M.I., Mock, T., Sterck, L., Sabbe, K., De Veylder, L., Vyverman, W., Vandepoele, K., 2020. The Seminavis robusta genome provides insights into the evolutionary adaptations of benthic diatoms. Nat. Commun. 11, 3320.

Prince, E.K., Irmer, F., Pohnert, G., 2013. Domoic acid improves the competitive ability of Pseudo-nitzschia delicatissima against the diatom Skeletonema marinoi. Mar. Drugs 11, 2398–2412.

Puritz, J.B., Matz, M.V., Toonen, R.J., Weber, J.N., Bolnick, D.I., Bird, C.E., 2014. Demystifying the RAD fad. Mol. Ecol. 23, 5937–5942.

R Core Team, Others, 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Vienna, Austria.

Round, F.E., Crawford, R.M., Mann, D.G., 1990. Diatoms: Biology and Morphology of the Genera. Cambridge University Press.

Rue, E., Bruland, K., 2001. Domoic acid binds iron and copper: a possible role for the toxin produced by the marine diatom Pseudo-nitzschia. Mar. Chem. 76, 127–134.

Rynearson, T.A., Newton, J.A., Armbrust, E.V., 2006. Spring bloom development, genetic variation, and population succession in the planktonic diatom Ditylum brightwellii. Limnol. Oceanogr. 51, 1249–1261.

Salazar, G., Sunagawa, S., 2017. Marine microbial diversity. Curr. Biol. 27, R489–R494.

Sayers, E.W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K.D., Karsch-Mizrachi, I., 2019. GenBank. Nucleic Acids Res. 47, D94–D99.

Schlötterer, C., Tobler, R., Kofler, R., Nolte, V., 2014. Sequencing pools of individuals — mining genome-wide polymorphism data without big funding. Nat. Rev. Genet. 15, 749–763.

Schnetzer, A., Jones, B.H., Schaffner, R.A., Cetinic, I., Fitzpatrick, E., Miller, P.E., Seubert, E.L., Caron, D.A., 2013. Coastal upwelling linked to toxic Pseudo-nitzschia australis blooms in Los Angeles coastal waters, 2005–2007. J. Plankton Res. 35, 1080–1092.

Schoch, C.L., Ciufo, S., Domrachev, M., Hotton, C.L., Kannan, S., Khovanskaya, R., Leipe, D., Mcveigh, R., O'Neill, K., Robbertse, B., Sharma, S., Soussov, V., Sullivan, J.P., Sun, L., Turner, S., Karsch-Mizrachi, I., 2020. NCBI Taxonomy: a comprehensive update on curation, resources and tools. Database 2020. https://doi.org/10.1093/database/baaa062

Scholin, C.A., Gulland, F., Doucette, G.J., Benson, S., Busman, M., Chavez, F.P., Cordaro, J., DeLong, R., De Vogelaere, A., Harvey, J., Haulena, M., Lefebvre, K., Lipscomb, T., Loscutoff, S., Lowenstine, L.J., Marin, R., 3rd, Miller, P.E., McLellan, W.A., Moeller, P.D., Powell, C.L., Rowles, T., Silvagni, P., Silver, M., Spraker, T., Trainer, V., Van Dolah, F.M., 2000. Mortality of sea lions along the central California coast linked to a toxic diatom bloom. Nature 403, 80–84.

Seubert, E.L., Gellene, A.G., Howard, M.D.A., Connell, P., Ragan, M., Jones, B.H., Runyan, J., Caron, D.A., 2013. Seasonal and annual dynamics of harmful algae and algal toxins revealed through weekly monitoring at two coastal ocean sites off southern California,

952 USA. Environ. Sci. Pollut. Res. Int. 20, 6878–6895.
953 Seymour, J.R., Amin, S.A., Raina, J.-B., Stocker, R., 2017. Zooming in on the phycosphere: the
954 ecological interface for phytoplankton–bacteria relationships. Nature Microbiology 2, 1–12.
955 Sison-Mangus, M.P., Jiang, S., Tran, K.N., Kudela, R.M., 2014. Host-specific adaptation
956 governs the interaction of the marine diatom, Pseudo-nitzschia and their microbiota. ISME
957 J. 8, 63–76.
958 Smith, J., Connell, P., Evans, R.H., Gellene, A.G., Howard, M.D.A., Jones, B.H., Kaveggia, S.,
959 Palmer, L., Schnetzer, A., Seegers, B.N., Seubert, E.L., Tatters, A.O., Caron, D.A., 2018a.
960 A decade and a half of Pseudo-nitzschia spp. and domoic acid along the coast of southern
961 California. Harmful Algae 79, 87–104.
962 Smith, J., Gellene, A.G., Hubbard, K.A., Bowers, H.A., Kudela, R.M., Hayashi, K., Caron, D.A.,
963 2018b. Pseudo-nitzschia species composition varies concurrently with domoic acid
964 concentrations during two different bloom events in the Southern California Bight. J.
965 Plankton Res. 40, 29–45.
966 Tatters, A.O., Fu, F.-X., Hutchins, D.A., 2012. High CO2 and silicate limitation synergistically
967 increase the toxicity of Pseudo-nitzschia fraudulenta. PLoS One 7, e32116.
968 Töpel, M., Pinder, M.I.M., Johansson, O.N., Kourtchenko, O., Godhe, A., Clarke, A.K., 2019.
969 Whole Genome Sequence of Marinobacter salarius Strain SMR5, Shown to Promote
970 Growth in its Diatom Host. J Genomics 7, 60–63.
971 Trainer, V.L., Adams, N.G., Bill, B.D., Stehr, Wekell John C., Moeller, P., Busman, M., Woodruff,
972 D., 2000. Domoic acid production near California coastal upwelling zones, June 1998.
973 Limnol. Oceanogr. 45, 1818–1833.
974 Trainer, V.L., Bates, S.S., Lundholm, N., Thessen, A.E., Cochlan, W.P., Adams, N.G., Trick,
975 C.G., 2012. Pseudo-nitzschia physiological ecology, phylogeny, toxicity, monitoring and
976 impacts on ecosystem health. Harmful Algae 14, 271–300.
977 Trainer, V.L., Hickey, B.M., Lessard, E.J., Cochlan, W.P., Trick, C.G., Wells Mark L.,
978 MacFadyen Amoreena, Moore Stephanie K., 2009. Variability of Pseudo-nitzschia and
979 domoic acid in the Juan de Fuca eddy region and its adjacent shelves. Limnol. Oceanogr.
980 54, 289–308.
981 Trainer, V.L., Pitcher, G.C., Reguera, B., Smayda, T.J., 2010. The distribution and impacts of
982 harmful algal bloom species in eastern boundary upwelling systems. Prog. Oceanogr. 85,
983 33–52.
984 Villac, M.C., Fryxell, G.A., 1998. Pseudo-nitzschia pungens var. cingulata var. nov.
985 (Bacillariophyceae) based on field and culture observations. Phycologia 37, 269–274.
986 Wang, S., Meyer, E., McKay, J.K., Matz, M.V., 2012. 2b-RAD: a simple and flexible method for
987 genome-wide genotyping. Nat. Methods 9, 808–810.
988 Wells, M.L., Trick, C.G., Cochlan, W.P., Hughes, M.P., Trainer, V.L., 2005. Domoic acid: The
989 synergy of iron, copper, and the toxicity of diatoms. Limnol. Oceanogr. 50, 1908–1917.
990 Wessells, C.R., Miller, C.J., Brooks, P.M., 1995. Toxic Algae Contamination and Demand for
991 Shellfish: A Case Study of Demand for Mussels in Montreal. Mar. Resour. Econ. 10, 143–
992 159.
993 Whittaker, K.A., Rynearson, T.A., 2017. Evidence for environmental and ecological selection in
994 a microbe with no geographic limits to gene flow. Proceedings of the National Academy of
995 Sciences 114, 2651–2656.
996 Xu, N., Tang, Y.Z., Qin, J., Duan, S., Gobler, C.J., 2015. Ability of the marine diatoms Pseudo-
997 nitzschia multiseries and P. pungens to inhibit the growth of co-occurring phytoplankton via
998 allelopathy. Aquat. Microb. Ecol. 74, 29–41.

999