README File

Required information:

1. File names, directory structure (for zipped files), and brief description of each file or file type, including where in the research process each data file lies (e.g. raw/unanalyzed data, processed/analyzed data, rendered/visualized data)

Dataset 4.1.xls

Dataset 4.2.xls

2. Definitions of acronyms, site abbreviations, or other project-specific designations used in the data file names or documentation files, if applicable

Asterionellopsis glacialis (A. glacialis)

False Discovery Rate (FDR)

Karenia brevis (K. brevis)

Kyoto Encyclopedia of Genes and Genomes (KEGG)

National Center for Biotechnology Information (NCBI)

Thalassiosira pseudonana (T. pseudonana)

3. Description of the parameters/variables (column headings in the data files) and units of measure for each parameter/variable, including special codes, variable classes, GIS coverage attributes, etc. used in the data files themselves, including codes for missing data values, if applicable:

Parameters are described below in Methods. See (6).

4. Uncertainty, precision, and accuracy of measurements, if known

Not known.

5. Environmental or experimental conditions, if appropriate (e.g., cloud cover, atmospheric influences, etc.)

A coculture design was used to investigate antagonistic interactions between K. brevis and each of two competitor species. The diatoms *A. glacialis* and *T. pseudonana* were each cocultured with the dinoflagellate *K. brevis* (n = 15) with the latter placed within permeable dialysis membrane tubing. This design allowed for ongoing exudation of allelopathic compounds from *K. brevis* through the dialysis membrane over the course of the experiment, but prevented cell contact between species. For controls (n = 15 for each experiment), competitors were exposed to dialysis tubes filled with L1 media diluted to 65% (vol/vol) of full L1 nitrate, vitamins, and

trace metal concentrations, and 90% (vol/vol) of full L1 phosphate concentrations to mimic the nutritional environment of exponentially growing *K. brevis*. After a period of coculture with live *K. brevis* (treatments) or dilute media (controls), competitor cells were harvested, extracted, and their polar metabolite profiles compared using ^{1}H NMR spectroscopy and ultra performance liquid chromatography– mass spectrometry (UPLC-MS). Cell pellets of diatoms were also harvested for proteomics and analyzed via LC-MS.

6. Method(s) for collecting the data, as well as the methods for processing data, if data other than raw data are being contributed

Two biological replicate samples from each *T. pseudonana* treatment and three biological replicates from each *A. glacialis* treatment were collected for proteomic analyses (450 ml culture each). Cells were pelleted (10,000 *g*; 10 min) on ice and lysed using a titanium microtip sonicating probe. Each sample received 10 sonication events (10-15 s each) in 0.2% PPS silent surfactant (Agilent Technologies) in 50 mM ammonium bicarbonate. The details of the digestion were per the manufacturer's guidelines. Disulfide bonds were reduced with dithiothreitol and alkylated with iodoacetamide. Each sample received trypsin at an enzyme to protein ratio of 1:50, vortexed, and incubated on a Thermomixer (800 rpm) 4 h at 37 ^{o}C. Peptide concentrations were measured for each sample using a Thermo Scientific NanoDrop 2000/2000c spectrophotometer. The peptide bond absorbance was monitored at 205 nm and samples were diluted to yield a final concentration of 100 µg protein mL^{-1}.

Samples were separated and introduced into the mass spectrometer by reversed-phase chromatography using a 30 cm-long, 75 µm i.d. fused silica capillary column packed with C_{18} silica particles (Magic C18AQ, 100 Å, 5 m; Michrom, Bioresources, Inc., CA) fitted with a 2 cm long, 100 µm i.d. precolumn (Magic C18AQ, 200 Å, 5 m; Michrom). Peptides were eluted using an acidified (formic acid, 0.1% v/v) water/acetonitrile gradient (2-35% acetonitrile over 90 min). Mass spectrometry was performed on a Thermo Fisher (San Jose, CA) QExactive (QE). Based on peptide concentrations, a total of 1 µg of peptide digest in 10 µl of 2% ACN, 0.1% formic acid was sampled per LC-MS analysis. Biological duplicates were analyzed from the treatment and the control group of *T. pseudonana* using data-dependent acquisition (DDA). Biological triplicates of *A. glacialis* controls and treatment were analyzed using DDA on the QExactive. To determine relative expression of proteins with respect to the alternate cell state, *QSpec* software was utilized to provide statistical significance.

Tandem mass spectrometry results were searched and interpreted with SEQUEST (PVM v.27 20070905). SEQUEST parameters included: reverse concatenated sequence database search, trypsin enzyme specificity, cysteine modification of 57 Da (resulting from iodoacetamide), and modifications on methionine of 15.999 Da (oxidation). Minimum protein and peptide thresholds were set at $p > 0.95$ on ProteinProphet and PeptideProphet. Protein identifications from whole cell lysates were accepted by ProteinProphet if the above mentioned thresholds were passed, two or more peptides were identified (PeptideProphet), and at least one terminus was tryptic. Using concatenated target-decoy database searches, false discovery rates (FDR) were calculated according to Elias and Gygi 2007 and were all < 1%. The protein database used for correlating spectra from the *T. pseudonana* experiment with protein identification was

generated by combining the latest release of the *T. pseudonana* proteome (12914 proteins) and 50 common contaminants. Since the genome of *A. glacialis* has not been described, we first searched the *A. glacialis* data as for *T. pseudonana*. Second, we assembled a database to search for confident peptide matches to spectra from protein sequences associated with diatoms in the National Center for Biotechnology Information database (www.ncbi.nlm.nih.gov). Using the "nrdb90" script provided by Holm and Sander at a 95% identity threshold on the input diatom database of 93,744 protein sequence entries reduced redundancy and size down to 64,752 sequence entries. ProteinProphet and PeptideProphet were set at a higher threshold of p > 0.99, to increase confidence in peptide and protein assignments. Proteins from the *A.glacialis* treatments and controls were considered present if one of the biological replicates identified two or more unique peptides and the termini were tryptic.

To determine relative expression of proteins expressed in treatments with respect to the control cell state, *QSpec* software was utilized to provide statistical significance. The Database for Annotation, Visualization and Integrated Discovery (DAVID) v. 6.7 was used to examine enriched biological terms in the sets of proteins identified to be significantly enhanced or suppressed relative to control cell state. P-values <1E-3 are reported for DAVID analyses on *T. pseudonana* data.

7. Standards or calibrations that were used

N/A

8. Specialized software (including version number) used to produce, prepare, render, compress, analyze and/or needed to read the dataset, if applicable

See Methods (6) above.

9. Quality assurance and quality control that have been applied, if applicable

N/A

10. Known problems that limit the data's use or other caveats (e.g., uncertainty, sampling problems, blanks, QC samples)

No known problems.

11. Date dataset was last modified

20140206

12. Relationships with any ancillary datasets outside of this dataset, if applicable

N/A

Optional information:

14. Resources, such as books, articles, serials, and/or data files, if any, that served as source of this data collection

15. Methodology for sample treatment and/or analysis, if applicable

16. Example records for each data file (or file type)

17. Files names of other documentation that are being submitted along with the data and that would be helpful to a secondary data user, such as pertinent field notes or other companion files, publications, etc.

Jones, C. M. [‡]; Ellestad, K. L. [‡]; Roy, J.; Viant, M. R.; Fernández, F. M.; Kubanek, J.; Nunn, B. L., Metabolomics and Proteomics Reveal Impacts of Chemically Mediated Competition on Marine Plankton. *Proc. Natl. Acad. Sci. USA* 2014, *111* (24), 9009-9014.
[‡]equal contributing author