

1 **Variation in prostaglandin metabolism during growth of the diatom *Thalassiosira rotula***
2 **Valeria Di Dato, Roberta Barbarinaldi, Alberto Amato, Federica Di Costanzo, Carolina**
3 **Fontanarosa, Anna Perna, Angela Amoresano, Francesco Esposito, Adele Cutignano,**
4 **Adrianna Ianora, Giovanna Romano**

5

6

7

8

9

Supplementary Information

10

11

12

13

14

15 **Supplementary information content:**

16	✓ Supplementary figure 1:	4
17	<i>Phyre</i> ² best hits structures from TrotCOX protein blast.	
18	✓ Supplementary figure 2:	10
19	<i>Phyre</i> ² alignment between <i>Thalassiosira rotula</i> (TrotCOX) and <i>Bos taurus</i> lactoperoxidase	
20	(BtLPO).	
21	✓ Supplementary figure 3:	11
22	<i>Phyre</i> ² alignment between <i>Thalassiosira rotula</i> (TrotCOX) and <i>Ovis aries</i> cyclooxygenase	
23	(OaCOX).	
24	✓ Supplementary figure 4:	12
25	Alignment of the diatoms sequences used in the phylogenetic analysis.	
26	✓ Supplementary table 1:	13
27	List of species and sequences ID used in the phylogenetic analysis.	
28	✓ External standard calibration method for PG quantization.	15
29	✓ Supplementary figure 5:	17
30	Calibration curve of deuterated prostaglandin E2 (PGE2-d4) used as internal standard.	
31	✓ Supplementary figure 6:	18
32	Calibration curve of prostaglandin E2 (PGE2) used as standard.	
33	✓ Supplementary figure 7:	19
34	Calibration curve of prostaglandin B2 (PGB2) used as standard.	
35	✓ Supplementary figure 8:	20

36	Calibration curve of 15-Deoxy-Delta-12,14-prostaglandin J2 (15dPGJ2) used as standard.	
37	✓ Supplementary figure 9:	21
38	TIC chromatogram (A) of a representative sample and XIC chromatogram (B) of PGEM (327	
39	m/z precursor ion; 291 m/z and 309 m/z daughter ion).	
40	✓ Supplementary table 3:	22
41	List of PGs identified in the culture medium with their LC-MS/MS parameters.	
42	✓ Aligned sequences, with gap, used in the phylogenetic analysis:	23
43	✓ Sequences used in the phylogenetic analysis, without gap	38
44	✓ References	48
45		

46 **Supplementary figure 1.** Phyre2¹ best hits structures from TrotCOX protein blast.



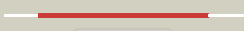
















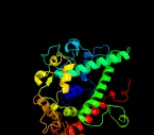


47

48

Phyre2

Email	alberto.amato@cea.fr
Description	Thalassiosira_rotula_TR754_c0_g1_i1_m_3431
Date	Thu Nov 7 16:33:40 GMT 2019
Unique Job ID	ca99ec1885b17cd9

Detailed template information

#	Template	Alignment Coverage	3D Model	Confidence	% i.d.	Template Information
1	c2gjmA_	 Alignment		100.0	19	PDB header: oxidoreductase Chain: A: PDB Molecule: lactoperoxidase; PDBTitle: crystal structure of buffalo lactoperoxidase at 2.75a resolution
2	c2oyuP_	 Alignment		100.0	28	PDB header: oxidoreductase Chain: P: PDB Molecule: prostaglandin g/h synthase 1; PDBTitle: indomethacin-(s)-alpha-ethyl-ethanolamide bound to cyclooxygenase-1
3	c1pggB_	 Alignment		100.0	28	PDB header: oxidoreductase Chain: B: PDB Molecule: prostaglandin h2 synthase-1; PDBTitle: prostaglandin h2 synthase-1 complexed with 1-(4-iodobenzoyl)-5-2 methoxy-2-methylindole-3-acetic acid (iodoindomethacin), trans model
4	c1ht8B_	 Alignment		100.0	28	PDB header: oxidoreductase Chain: B: PDB Molecule: prostaglandin h2 synthase-1; PDBTitle: the 2.7 angstrom resolution model of ovine cox-1 complexed with2 alclufenac
5	d1q4ga1	 Alignment		100.0	28	Fold: Heme-dependent peroxidases Superfamily: Heme-dependent peroxidases Family: Myeloperoxidase-like
6	c6ercA_	 Alignment		100.0	26	PDB header: oxidoreductase Chain: A: PDB Molecule: peroxinectin a; PDBTitle: peroxidase a from dictyostelium discoideum (ddpoxa)
7	c5mfaA_	 Alignment		100.0	20	PDB header: oxidoreductase Chain: A: PDB Molecule: myeloperoxidase; PDBTitle: crystal structure of human promyeloperoxidase (prompo)
8	c1ddxA_	 Alignment		100.0	27	PDB header: oxidoreductase Chain: A: PDB Molecule: protein (prostaglandin h2 synthase-2); PDBTitle: crystal structure of a mixture of arachidonic acid and prostaglandin2 bound to the cyclooxygenase active site of cox-2: prostaglandin3 structure
9	c3pghD_	 Alignment		100.0	27	PDB header: oxidoreductase Chain: D: PDB Molecule: cyclooxygenase-2; PDBTitle: cyclooxygenase-2 (prostaglandin synthase-2) complexed with a non-2 selective inhibitor, flurbiprofen
10	d1cvua1	 Alignment		100.0	27	Fold: Heme-dependent peroxidases Superfamily: Heme-dependent peroxidases Family: Myeloperoxidase-like
11	c1d2vD_	 Alignment		100.0	20	PDB header: oxidoreductase Chain: D: PDB Molecule: myeloperoxidase; PDBTitle: crystal structure of bromide-bound human myeloperoxidase isoform c at2 ph 5.5

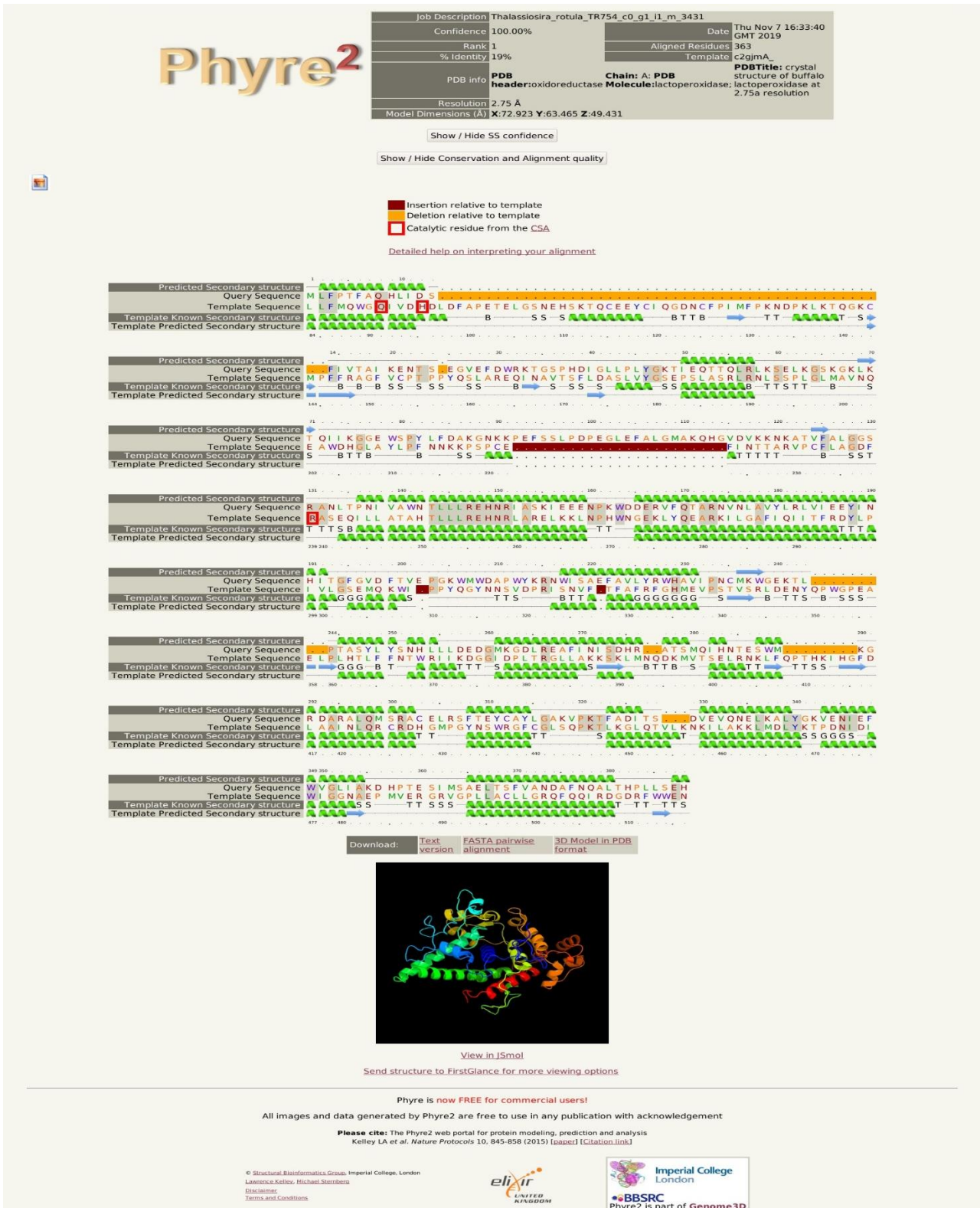
12	c4hhsA_	Alignment		100.0	19	PDB header: oxidoreductase Chain: A: PDB Molecule: alpha-dioxygenase; PDBTitle: crystal structure of fatty acid alpha-dioxygenase (arabidopsis2 thaliana)
13	c4kvjA_	Alignment		100.0	19	PDB header: oxidoreductase Chain: A: PDB Molecule: fatty acid alpha-oxidase; PDBTitle: crystal structure of oryza sativa fatty acid alpha-dioxygenase with2 hydrogen peroxide
14	c5gvrA_	Alignment		36.3	17	PDB header: hydrolase Chain: A: PDB Molecule: probable atp-dependent rna helicase ddx41; PDBTitle: crystal structure of the ddx41 dead domain in an apo closed form
15	c3brcA_	Alignment		35.0	18	PDB header: structural genomics, unknown function Chain: A: PDB Molecule: conserved protein of unknown function; PDBTitle: crystal structure of a conserved protein of unknown function from2 methanobacterium thermoautotrophicum
16	c1vk5A_	Alignment		33.5	20	PDB header: structural genomics, unknown function Chain: A: PDB Molecule: expressed protein; PDBTitle: x-ray structure of gene product from arabidopsis thaliana at3g22680
17	d1vk5a_	Alignment		33.5	20	Fold: Hypothetical protein At3g22680 Superfamily: Hypothetical protein At3g22680 Family: Hypothetical protein At3g22680
18	d1u7ka_	Alignment		28.8	33	Fold: Retrovirus capsid protein, N-terminal core domain Superfamily: Retrovirus capsid protein, N-terminal core domain Family: Retrovirus capsid protein, N-terminal core domain
19	c3fe2B_	Alignment		22.4	13	PDB header: hydrolase Chain: B: PDB Molecule: probable atp-dependent rna helicase ddx5; PDBTitle: human dead-box rna helicase ddx5 (p68), conserved domain i in complex2 with adp
20	c4px9C_	Alignment		21.3	12	PDB header: translation, rna binding protein Chain: C: PDB Molecule: atp-dependent rna helicase ddx3x; PDBTitle: dead-box rna helicase ddx3x domain 1 with n-terminal atp-binding loop
21	c3evyB_	Alignment	not modelled	17.7	9	PDB header: hydrolase Chain: B: PDB Molecule: putative type i restriction enzyme r protein; PDBTitle: crystal structure of a fragment of a putative type i restriction2 enzyme r protein from bacteroides fragilis
22	d1ou9a_	Alignment	not modelled	17.3	21	Fold: SspB-like Superfamily: SspB-like Family: Stringent starvation protein B, SspB
23	c2noxP_	Alignment	not modelled	17.3	14	PDB header: oxidoreductase Chain: P: PDB Molecule: tryptophan 2,3-dioxygenase; PDBTitle: crystal structure of tryptophan 2,3-dioxygenase from ralstonia2 metallidurans
24	c6hwwA_	Alignment	not modelled	16.6	33	PDB header: viral protein Chain: A: PDB Molecule: putative gag polyprotein; PDBTitle: immature mlv capsid hexamer structure in intact virus particles
25	d1zszc1	Alignment	not modelled	16.5	16	Fold: SspB-like Superfamily: SspB-like Family: Stringent starvation protein B, SspB
26	c6hwyB_	Alignment	not modelled	15.9	33	PDB header: viral protein Chain: B: PDB Molecule: putative gag polyprotein; PDBTitle: mature mlv capsid pentamer structure in intact virus particles
27	c3fhcB_	Alignment	not modelled	15.6	9	PDB header: transport protein/hydrolase Chain: B: PDB Molecule: atp-dependent rna helicase ddx19b; PDBTitle: crystal structure of human dbp5 in complex with nup214
28	d1or7b2	Alignment	not modelled	15.0	32	Fold: Sigma2 domain of RNA polymerase sigma factors Superfamily: Sigma2 domain of RNA polymerase sigma factors Family: Sigma2 domain of RNA polymerase sigma factors

29	c4cmxB	Alignment	not modelled	14.8	14	PDB header: nuclear protein Chain: B: PDB Molecule: rv3378c; PDBTitle: crystal structure of rv3378c
30	d1g1xc	Alignment	not modelled	14.3	25	Fold: DNA/RNA-binding 3-helical bundle Superfamily: Ribosomal protein S18 Family: Ribosomal protein S18
31	c4qicB	Alignment	not modelled	14.2	18	PDB header: signaling protein/dna binding protein Chain: B: PDB Molecule: anti-sigma factor nepr; PDBTitle: co-crystal structure of anti-anti-sigma factor phyr complexed with2 anti-sigma factor nepr from bartonella quintana
32	d1mn3a	Alignment	not modelled	13.8	14	Fold: RuvA C-terminal domain-like Superfamily: UBA-like Family: CUE domain
33	d1ou8a	Alignment	not modelled	12.8	21	Fold: SspB-like Superfamily: SspB-like Family: Stringent starvation protein B, SspB
34	c2c2xB	Alignment	not modelled	12.6	14	PDB header: oxidoreductase Chain: B: PDB Molecule: methylenetetrahydrofolate dehydrogenase- PDBTitle: three dimensional structure of bifunctional methylenetetrahydrofolate2 dehydrogenase-cyclohydrolase from mycobacterium tuberculosis
35	c1fcuA	Alignment	not modelled	12.6	36	PDB header: hydrolase Chain: A: PDB Molecule: hyaluronoglucosaminidase; PDBTitle: crystal structure (trigonal) of bee venom hyaluronidase
36	d1fcqa	Alignment	not modelled	12.5	36	Fold: TIM beta/alpha-barrel Superfamily: (Trans)glycosidases Family: Bee venom hyaluronidase
37	c3fmpD	Alignment	not modelled	12.2	9	PDB header: oncoprotein/hydrolase Chain: D: PDB Molecule: atp-dependent rna helicase ddx19b; PDBTitle: crystal structure of the nucleoporin nup214 in complex with the dead-2 box helicase ddx19
38	d1yfna1	Alignment	not modelled	11.9	21	Fold: SspB-like Superfamily: SspB-like Family: Stringent starvation protein B, SspB
39	c2b19A	Alignment	not modelled	11.5	42	PDB header: neuropeptide Chain: A: PDB Molecule: neuropeptide k; PDBTitle: solution structure of mammalian tachykinin peptide,2 neuropeptide k
40	c3rfaB	Alignment	not modelled	11.0	13	PDB header: oxidoreductase Chain: B: PDB Molecule: ribosomal rna large subunit methyltransferase n; PDBTitle: x-ray structure of rlmn from escherichia coli in complex with s-2 adenosylmethionine
41	c2kneB	Alignment	not modelled	10.8	32	PDB header: metal transport Chain: B: PDB Molecule: atpase, ca++ transporting, plasma membrane 4; PDBTitle: calmodulin wraps around its binding domain in the plasma2 membrane ca2+ pump anchored by a novel 18-1 motif
42	d1upka	Alignment	not modelled	10.6	12	Fold: alpha-alpha superhelix Superfamily: ARM repeat Family: Mo25 protein
43	c3rfaA	Alignment	not modelled	10.2	13	PDB header: oxidoreductase Chain: A: PDB Molecule: ribosomal rna large subunit methyltransferase n; PDBTitle: x-ray structure of rlmn from escherichia coli in complex with s-2 adenosylmethionine
44	d2bmfa1	Alignment	not modelled	10.1	21	Fold: P-loop containing nucleoside triphosphate hydrolases Superfamily: P-loop containing nucleoside triphosphate hydrolases Family: RNA helicase
45	d2qalr1	Alignment	not modelled	10.1	29	Fold: DNA/RNA-binding 3-helical bundle Superfamily: Ribosomal protein S18 Family: Ribosomal protein S18
46	c3df1R	Alignment	not modelled	10.1	29	PDB header: ribosome Chain: R: PDB Molecule: 30s ribosomal protein s18; PDBTitle: crystal structure of the bacterial ribosome from escherichia coli in2 complex with hygromycin b. this file contains the 30s subunit of the3 first 70s ribosome, with hygromycin b bound. the entire crystal4 structure contains two 70s ribosomes.
47	d1ho8a	Alignment	not modelled	9.9	21	Fold: alpha-alpha superhelix Superfamily: ARM repeat Family: Regulatory subunit H of the V-type ATPase
48	c3cuza	Alignment	not modelled	9.9	9	PDB header: transferase Chain: A: PDB Molecule: malate synthase a; PDBTitle: atomic resolution structures of escherichia coli and2 bacillus anthracis malate synthase a: comparison with3 isoform g and implications for structure based drug design
49	c2g2bA	Alignment	not modelled	9.9	7	PDB header: immune system Chain: A: PDB Molecule: allograft inflammatory factor 1; PDBTitle: nmr structure of the human allograft inflammatory factor 1
50	c5gjuA	Alignment	not modelled	9.5	26	PDB header: hydrolase Chain: A: PDB Molecule: atp-dependent rna helicase dead; PDBTitle: dead-box rna helicase
51	c5cofA	Alignment	not modelled	9.4	11	PDB header: unknown function Chain: A: PDB Molecule: uncharacterized protein; PDBTitle: crystal structure of uncharacterised protein q1r1x2 from escherichia2 coli uti89
52	c5o5jR	Alignment	not modelled	9.3	31	PDB header: ribosome Chain: R: PDB Molecule: 30s ribosomal protein s18 2; PDBTitle: structure of the 30s small ribosomal subunit from mycobacterium2 smegmatis
53	c3a0bT	Alignment	not modelled	8.9	20	PDB header: electron transport Chain: T: PDB Molecule: photosystem ii reaction center protein t; PDBTitle: crystal structure of br-substituted photosystem ii complex

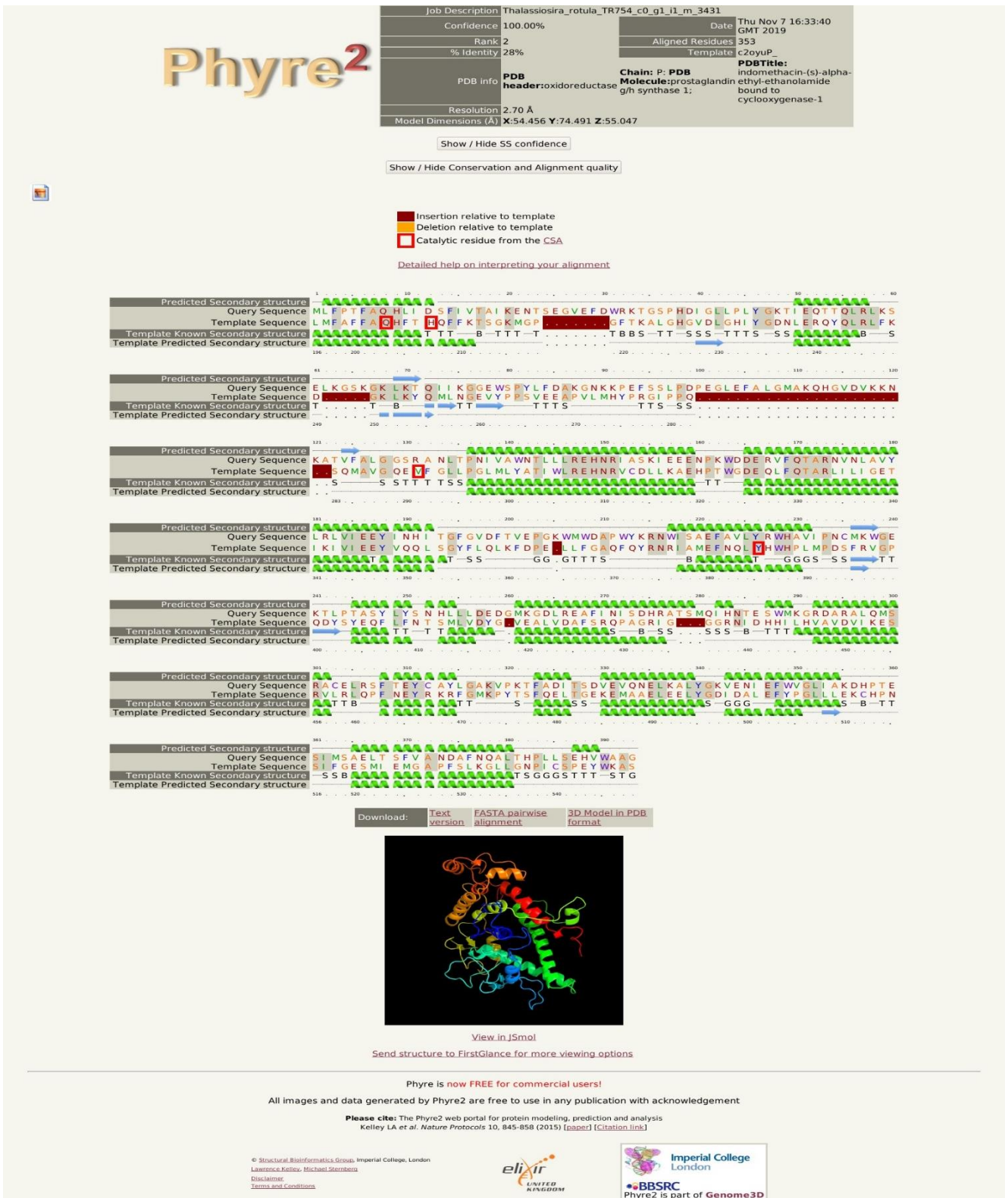
54	c5mx2T	Alignment	not modelled	8.2	18	PDB header: oxidoreductase Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: photosystem ii depleted of the mn4cao5 cluster at 2.55 a resolution
55	c2lv1J	Alignment	not modelled	8.2	17	PDB header: lyase Chain: J; PDB Molecule: cyanate hydratase; PDBTitle: site directed mutagenesis of key residues involved in the catalytic2 mechanism of cyanase
56	d1rfya	Alignment	not modelled	8.2	24	Fold: Long alpha-hairpin Superfamily: Transcriptional repressor TraM Family: Transcriptional repressor TraM
57	c5aj3R	Alignment	not modelled	8.0	6	PDB header: ribosome Chain: R; PDB Molecule: mitoribosomal protein bs18m, mrps18c; PDBTitle: structure of the small subunit of the mammalian mitoribosome
58	c2c0zA	Alignment	not modelled	7.9	22	PDB header: isomerase Chain: A; PDB Molecule: novw; PDBTitle: the 1.6 a resolution crystal structure of novw: a 4-keto-6-2 deoxy sugar epimerase from the novobiocin biosynthetic3 gene cluster of streptomyces spheroides
59	c3j0xU	Alignment	not modelled	7.8	31	PDB header: ribosome Chain: U; PDB Molecule: 30s ribosomal protein s18; PDBTitle: structural characterization of mrna-trna translocation intermediates2 (30s ribosome of class 4b of the six classes)
60	d1wrka1	Alignment	not modelled	7.8	7	Fold: EF Hand-like Superfamily: EF-hand Family: Calmodulin-like
61	c1upiA	Alignment	not modelled	7.8	16	PDB header: epimerase Chain: A; PDB Molecule: ddtp-4-dehydrorhamnose 3,5-epimerase; PDBTitle: mycobacterium tuberculosis rmlc epimerase (rv3465)
62	c5b66T	Alignment	not modelled	7.7	18	PDB header: electron transport, photosynthesis Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: crystal structure analysis of photosystem ii complex
63	c3bbnR	Alignment	not modelled	7.7	27	PDB header: ribosome Chain: R; PDB Molecule: ribosomal protein s18; PDBTitle: homology model for the spinach chloroplast 30s subunit fitted to 9.4a2 cryo-em map of the 70s chlororibosome.
64	c5gtht	Alignment	not modelled	7.6	18	PDB header: photosynthesis Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: native xfel structure of photosystem ii (dark dataset)
65	c5ws5t	Alignment	not modelled	7.6	18	PDB header: photosynthesis Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: native xfel structure of photosystem ii (preflash dark dataset)
66	c5ws6t	Alignment	not modelled	7.6	18	PDB header: photosynthesis Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: native xfel structure of photosystem ii (preflash two-flash dataset)
67	c5ws5T	Alignment	not modelled	7.6	18	PDB header: photosynthesis Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: native xfel structure of photosystem ii (preflash dark dataset)
68	c5gthT	Alignment	not modelled	7.6	18	PDB header: photosynthesis Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: native xfel structure of photosystem ii (dark dataset)
69	c3wu2T	Alignment	not modelled	7.6	18	PDB header: electron transport, photosynthesis Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: crystal structure analysis of photosystem ii complex
70	c5h2fT	Alignment	not modelled	7.6	18	PDB header: photosynthesis Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: crystal structure of the psbm-deletion mutant of photosystem ii
71	c5kait	Alignment	not modelled	7.6	18	PDB header: electron transport Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: nh3-bound rt xfel structure of photosystem ii 500 ms after the 2nd2 illumination (2f) at 2.8 a resolution
72	c5h2ft	Alignment	not modelled	7.6	18	PDB header: photosynthesis Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: crystal structure of the psbm-deletion mutant of photosystem ii
73	c4ub8t	Alignment	not modelled	7.6	18	PDB header: electron transport, photosynthesis Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: native structure of photosystem ii (dataset-2) by a femtosecond x-ray2 laser
74	c3wu2t	Alignment	not modelled	7.6	18	PDB header: electron transport, photosynthesis Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: crystal structure analysis of photosystem ii complex
75	c5gtiT	Alignment	not modelled	7.6	18	PDB header: photosynthesis Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: native xfel structure of photosystem ii (two flash dataset)
76	c5v2cT	Alignment	not modelled	7.6	18	PDB header: electron transport, photosynthesis Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: re-refinement of crystal structure of photosystem ii complex
77	c5gtit	Alignment	not modelled	7.6	18	PDB header: photosynthesis Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: native xfel structure of photosystem ii (two flash dataset)
78	c5v2ct	Alignment	not modelled	7.6	18	PDB header: electron transport, photosynthesis Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: re-refinement of crystal structure of photosystem ii complex
79	c6j1pt	Alignment	not modelled	7.6	18	PDB header: photosynthesis Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: xfel structure of cyanobacterial photosystem ii (3f state, dataset2)
						PDB header: photosynthesis

80	c6jlpT	Alignment	not modelled	7.6	18	Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: xfel structure of cyanobacterial photosystem ii (3f state, dataset2)
81	c6jlmT	Alignment	not modelled	7.6	18	PDB header: photosynthesis Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: xfel structure of cyanobacterial photosystem ii (dark state, dataset2)
82	c6jlkT	Alignment	not modelled	7.6	18	PDB header: photosynthesis Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: xfel structure of cyanobacterial photosystem ii (1f state, dataset1)
83	c6jloT	Alignment	not modelled	7.6	18	PDB header: photosynthesis Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: xfel structure of cyanobacterial photosystem ii (2f state, dataset2)
84	c6jltT	Alignment	not modelled	7.6	18	PDB header: photosynthesis Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: xfel structure of cyanobacterial photosystem ii (dark state, dataset1)
85	c6jllT	Alignment	not modelled	7.6	18	PDB header: photosynthesis Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: xfel structure of cyanobacterial photosystem ii (2f state, dataset1)
86	c6jlot	Alignment	not modelled	7.6	18	PDB header: photosynthesis Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: xfel structure of cyanobacterial photosystem ii (2f state, dataset2)
87	c6jlnT	Alignment	not modelled	7.6	18	PDB header: photosynthesis Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: xfel structure of cyanobacterial photosystem ii (1f state, dataset2)
88	c6jlnT	Alignment	not modelled	7.6	18	PDB header: photosynthesis Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: xfel structure of cyanobacterial photosystem ii (1f state, dataset2)
89	c6jllT	Alignment	not modelled	7.6	18	PDB header: photosynthesis Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: xfel structure of cyanobacterial photosystem ii (2f state, dataset1)
90	c6jikt	Alignment	not modelled	7.6	18	PDB header: photosynthesis Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: xfel structure of cyanobacterial photosystem ii (1f state, dataset1)
91	c6jimt	Alignment	not modelled	7.6	18	PDB header: photosynthesis Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: xfel structure of cyanobacterial photosystem ii (dark state, dataset2)
92	c6jjtT	Alignment	not modelled	7.6	18	PDB header: photosynthesis Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: xfel structure of cyanobacterial photosystem ii (dark state, dataset1)
93	c5tisT	Alignment	not modelled	7.6	18	PDB header: photosynthesis Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: room temperature xfel structure of the native, doubly-illuminated2 photosystem ii complex
94	c5zznT	Alignment	not modelled	7.6	18	PDB header: electron transport Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: crystal structure of photosystem ii from an sqdg-deficient mutant of2 thermosynechococcus elongatus
95	c5kaiT	Alignment	not modelled	7.6	18	PDB header: electron transport Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: nh3-bound rt xfel structure of photosystem ii 500 ms after the 2nd2 illumination (2f) at 2.8 a resolution
96	c5kafT	Alignment	not modelled	7.6	18	PDB header: electron transport Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: rt xfel structure of photosystem ii in the dark state at 3.0 a2 resolution
97	c4ub6T	Alignment	not modelled	7.6	18	PDB header: electron transport, photosynthesis Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: native structure of photosystem ii (dataset-1) by a femtosecond x-ray2 laser
98	c5zznT	Alignment	not modelled	7.6	18	PDB header: electron transport Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: crystal structure of photosystem ii from an sqdg-deficient mutant of2 thermosynechococcus elongatus
99	c4il6T	Alignment	not modelled	7.6	18	PDB header: electron transport Chain: T; PDB Molecule: photosystem ii reaction center protein t; PDBTitle: structure of sr-substituted photosystem ii

54 **Supplementary figure 2.** Phyre2¹ alignment between *Thalassiosira rotula* (TrotCOX) and *Bos*
 55 *taurus* lactoperoxidase (BtLPO).



57 **Supplementary figure 3.** Phyre2¹ alignment between *Thalassiosira rotula* (TrotCOX) and *Ovis aries*
 58 cyclooxygenase (OaCOX).

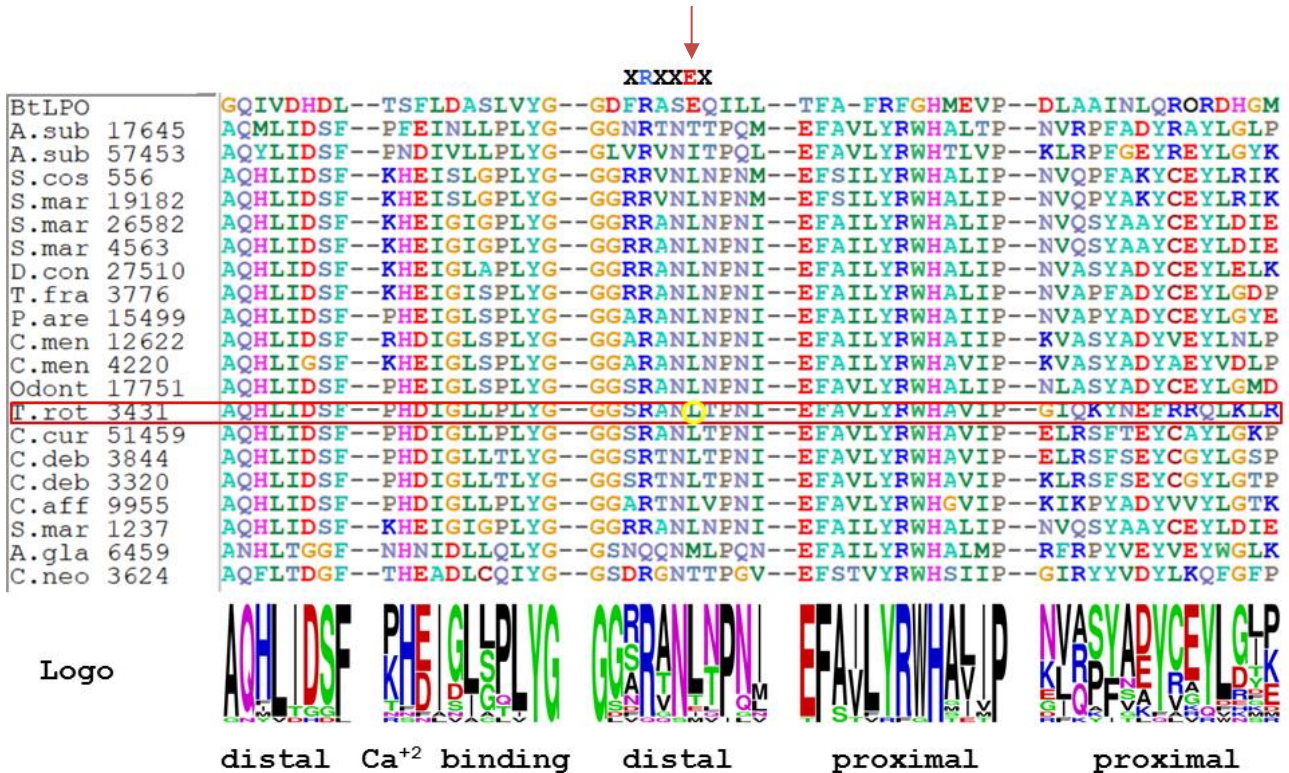


60 **Supplementary figure 4.** Alignment of the diatoms sequences used in the phylogenetic analysis
 61 (Figure 3, diatoms clade).

62 In red box the *T. rotula* sequence. The red arrow indicate the glutamate residue transformed in
 63 leucine in diatoms (yellow circle).

64 BtLPO = *Bos taurus* LPO; A.sub = *Aulacoseira subarctica* strain; S.cos = *Skeletonema costatum*;
 65 S.mar = *Skeletonema marinoi*; D.con = *Detonula confervacea*; T.fra = *Thalassionema frauenfeldii*;
 66 P.are = *Pseudo-nitzschia arenisensis*; C.men = *Cyclotella meneghiniana*; Odont = *Odontella*; T.rot
 67 = *Thalassiosira rotula*; C.cur = *Chaetoceros curvisetus*; C.deb = *Chaetoceros debilis*; C.aff =
 68 *Chaetoceros affinis*; A.gla = *Asterionellopsis glacialis*; C.neo = *Chaetoceros neogracile*.

69
 70



71

72 **Supplementary table 1: List of species and sequences IDs used in the phylogenetic analysis.**

Species	Annotation	Sequence ID
<i>Chaetoceros neogracile</i>		MMETSP0751-20121128_3624
<i>Asterionellopsis glacialis</i>		MMETSP0707-20130614_6459
<i>Aulacoseira subarctica</i>		MMETSP1064-20121228_17645
<i>Aulacoseira subarctica</i>		MMETSP1064-20121228_57453
<i>Chaetoceros affine</i>		MMETSP0090-20130426_9955
<i>Chaetoceros curvisetus</i>		MMETSP0717-20131115_51459
<i>Chaetoceros debilis</i>		MMETSP0149-20130528_3320
<i>Chaetoceros debilis</i>		MMETSP0150-20130528_3844
<i>Odontella</i>		MMETSP0015_2-20120614_17751
<i>Cyclotella meneghiniana</i>		MMETSP1057-20121228_4220
<i>Cyclotella meneghiniana</i>		MMETSP1057-20121228_12622
<i>Pseudo-nitzschia arenysensis</i>		MMETSP0329-20121206_15499
<i>Thalassionema frauenfeldii</i>		MMETSP0786-20121207_3776
<i>Skeletonema marinoi</i>		MMETSP0320-20121206_19182
<i>Skeletonema costatum</i>		MMETSP0013_2-20120614_556
<i>Detonula confervacea</i>		MMETSP1058-20130122_27510
<i>Skeletonema marinoi</i>		MMETSP1039-20121108_4563
<i>Skeletonema marinoi</i>		MMETSP0319-20121206_26582
<i>Skeletonema marinoi</i>		MMETSP0562-20121206_1237
<i>Skeletonema marinoi</i>		MMETSP1428-20130617_783
<i>Synedropsis recta</i>		MMETSP1176-20130426_10672
<i>Nostoc</i> sp.	COX-1	WP015113127
<i>Chaetoceros</i> cf. <i>neogracile</i>		MMETSP1336-20130426_1470
<i>Grammatophora oceanica</i>		MMETSP0009_2-20130614_5553
<i>Gersemia fruticosa</i>	COX-A	AAF93168
<i>Gersemia fruticosa</i>	COX-B	AAS48061
<i>Plexaura homomalla</i>	5S-specific-cyclooxygenase	AAU87497
<i>Plexaura homomalla</i>	COX-1	AAF93169
<i>Gammarus</i> sp.	COX-1	ADB65786
<i>Caprella</i> sp.	COX-1	ADB65785
<i>Nitzschia</i> sp		MMETSP0014_2-20120614_6887
<i>Skeletonema marinoi</i>		MMETSP0319-20121206_1511
<i>Pseudo-nitzschia pungens</i>		MMETSP1061-20121228_74475
<i>Thalassiosira rotula</i>		TR754 c0_g1_i1 m.3431
<i>Skeletonema marinoi</i>		MMETSP0920-20130426_33289
<i>Fragilariopsis kerguelensis</i>		MMETSP0907-20130614_13332
<i>Skeletonema grethae</i>		MMETSP0578-20130828_1168
<i>Staurosira</i> sp.		MMETSP1361-20130828_23824

<i>Skeletonema marinoi</i>		MMETSP1428-20130617_31509
<i>Homo sapiens</i>	COX-1	NP000953
<i>Ovis aries</i>	COX-1	NP001009476
<i>Homo sapiens</i>	COX-2	NP000954
<i>Coccotylus truncatus</i>	COX-1	AFN20596
<i>Ciona intestinalis</i>	COX-a	XP002127674
<i>Ciona intestinalis</i>	COX-b	XP002123273
<i>Oncorhynchus mykiss</i>	COX-1	CAC10360.1
<i>Fistulifera solaris</i>	Alpha-dioxygenase	GAX23950.1
<i>Cyanothece</i> sp.	Heme peroxidase	ACB53655.1
<i>Fistulifera solaris</i>		GAX18717.1
<i>Alcanivorax nanhaiticus</i>		WP_035233415.1
<i>Hordeum vulgare</i> subsp. <i>vulgare</i>		BAJ90503.1
<i>Triticum aestivum</i>		CDM84254.1
<i>Rhizophagus clarus</i>		GBC07128.1
<i>Emiliana huxleyi</i>		XP_005780718.1
<i>Rhodococcus gordoniae</i>		WP_064063209.1
<i>Herbidospora mongoliensis</i>		WP_066371138.1
<i>Eutrema halophilum</i>		BAJ34623.1
<i>Pisum sativum</i>	Alpha-dioxygenase	CAH05011.1

74 **External standard calibration method for PG quantization.**

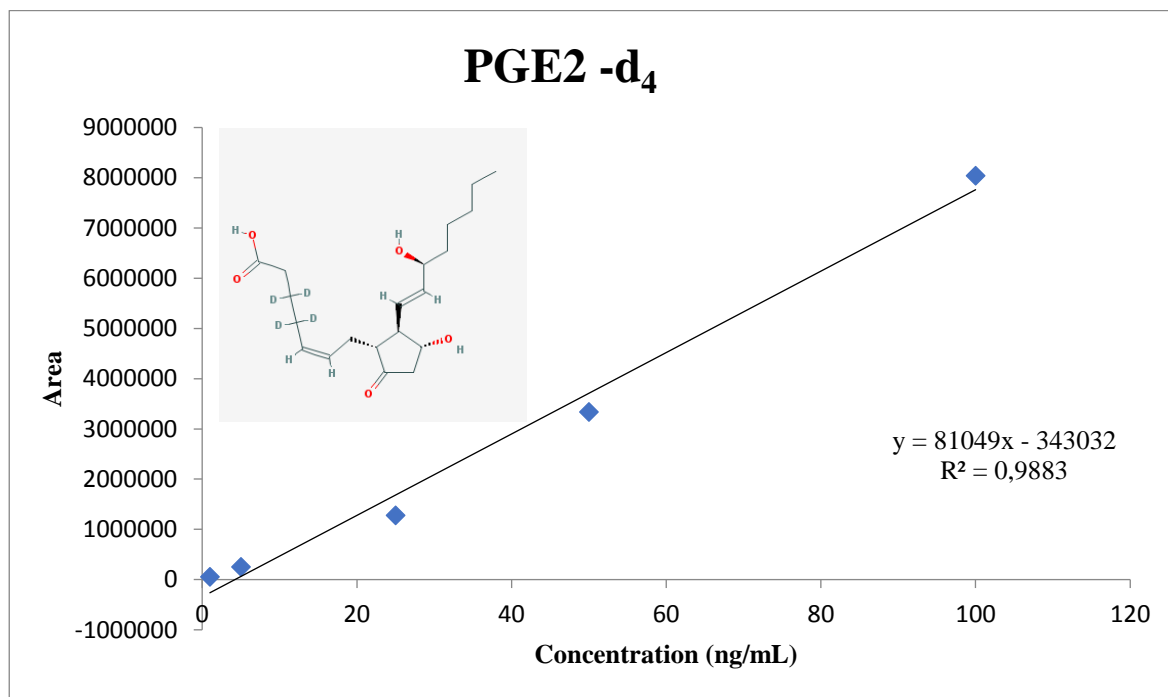
75 For an external standard quantitation, known data from a calibration standard and unknown data from
76 the sample are combined to generate a quantitative report. Working standard solutions containing all
77 the analytes at 1 $\mu\text{g}/\text{mL}$ were prepared, from stock solutions of 100 $\mu\text{g}/\text{mL}$ for each standard: PGE₂,
78 15-d-PGJ₂, PGB₂, PGD₂, PGJ₂ (SIGMA ALDRICH, Saint Louis, MO), 20-oh-PGE₂, 6-keto-PGF_{1 α} ,
79 DHK (CAYMAN CHEMICAL, Michigan, USA). From these standard solutions, five solutions at
80 different concentration levels were prepared: 1, 5, 25, 50, 100 ng/mL for the construction of the
81 calibration curves. The coefficients of determination (R²) were used to evaluate the linearity of the
82 calibration ranges. All R² were greater than 0.99, demonstrating that the method was linear in the
83 range of calibration for all analytes.

84 The amount of prostaglandins was normalized by the sampled volume extracted for each daily
85 replicate and then by the number of cells per mL for each replicate. The final value reported in the
86 graph were pg cell⁻¹.

87

88 **Supplementary figure 5: calibration curve of deuterated prostaglandin E2 (PGE2-d4) used as**
89 **internal standard.**

90

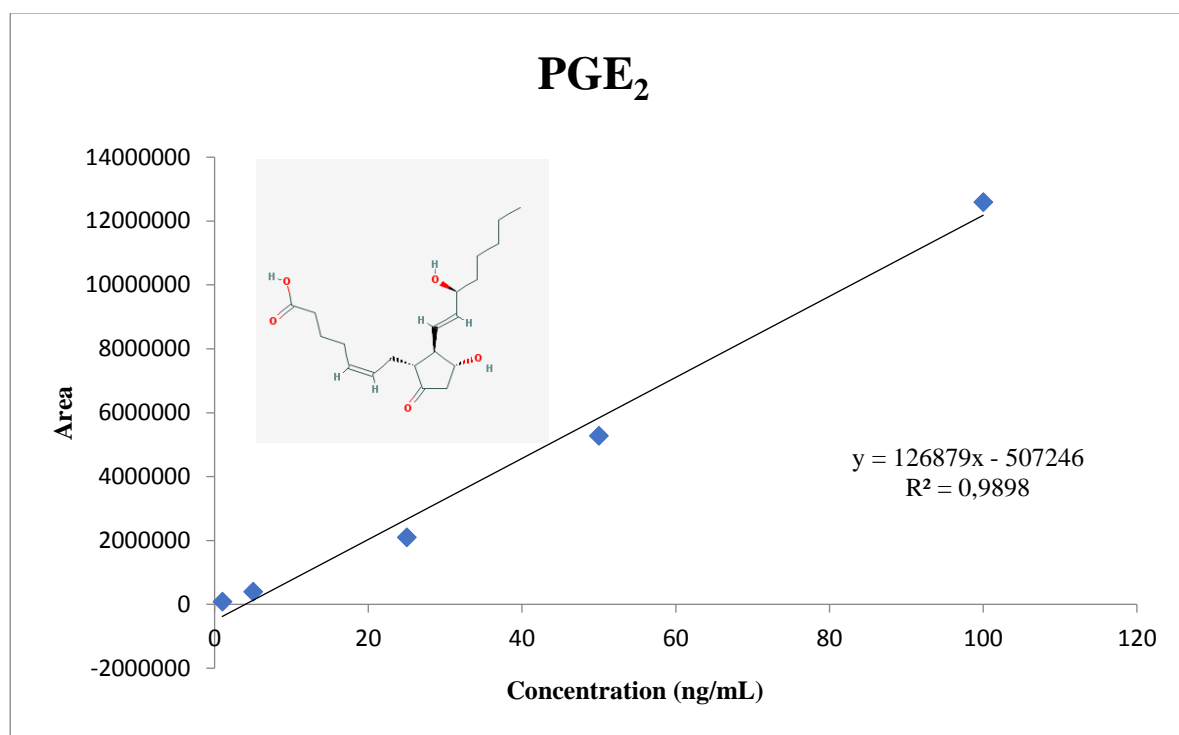


91

92

93 **Supplementary figure 6: calibration curve of prostaglandin E2 (PGE2) used as standard.**

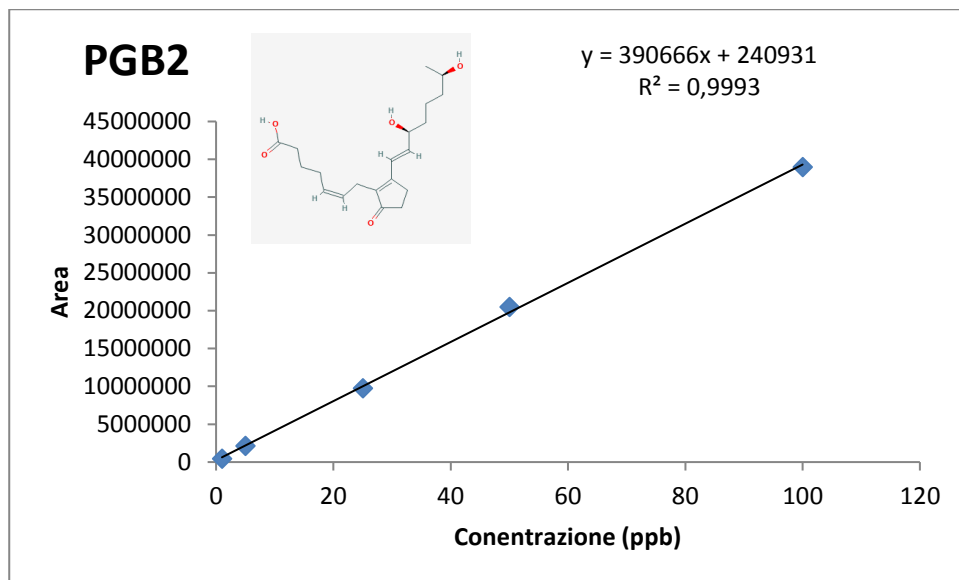
94



95

96

97 **Supplementary figure 7: calibration curve of prostaglandin B2 (PGB2) used as standard.**
98
99
100



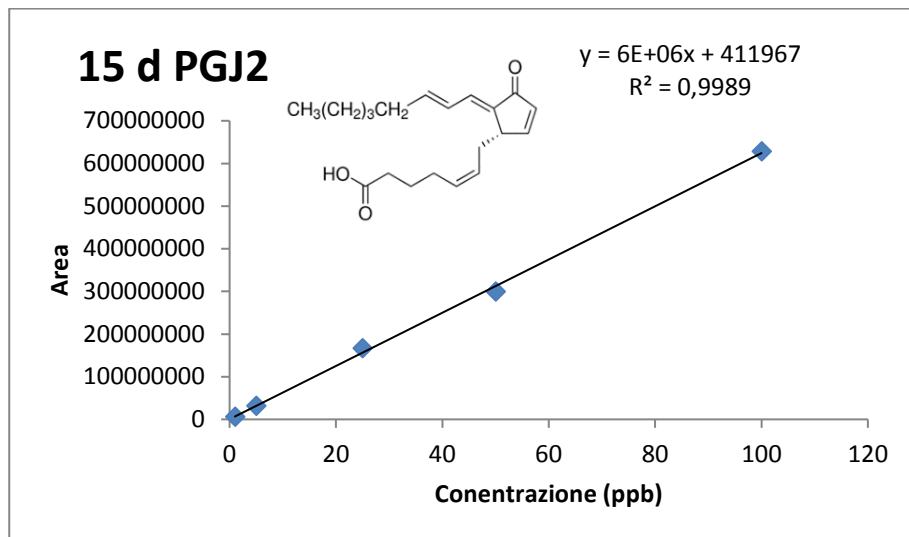
101

102

103

104 **Supplementary figure 8: calibration curve of 15-Deoxy-Delta-12,14-prostaglandin J2**
105 **(15dPGJ2) used as standard.**
106

107



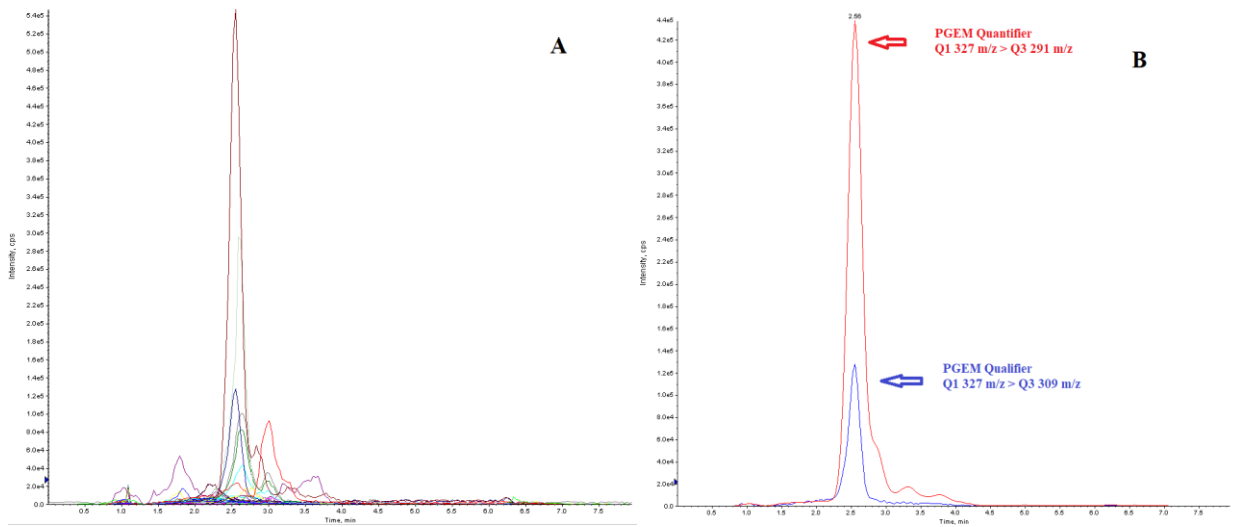
108

109 **Supplementary figure 9: examples of TIC chromatograms.**

110 **a.** a representative sample from the LC-MS/MS analysis of the culture medium and XIC
111 chromatogram;

112 **b.** PGEM (327 m/z precursor ion; 291 m/z and 309 m/z daughter ion) peaks.
113

114



115

116

117 **Supplementary table 3.** List of PGs identified in the culture medium with their LC-MS/MS
 118 parameters.

Molecule ID	rt	Q1	Q3	DP	CE
PGE₁	1,13	353	235	-60	-29
15-d-PGD₂	2,14	333	271	-60	-22
15-d-PGJ₂*	2,21	315	203	-60	-20
PGE₃	2,28	349	269	-60	-24
PGE-M	2,56	327	291	-60	-23
PGA₂	2,64	333	271	-60	-20
PGB₂*	2,69	333	271	-60	-25
PGF-M	2,7	329	293	-60	-25
2,3-dinor-11b-PGF₂	3,05	325	227	-60	-22
PGD₁	3,14	353	271	-60	-29
PGD₃	3,14	349	269	-60	-20
PGE₂*	3,23	351	272	-60	-23

119
 120 ID: official abbreviated prostaglandin names; Q1: precursor ion (*m/z*) selected in quadrupole Q1; Q3:
 121 fragment ion (*m/z*) selected in quadrupole Q3; Q1>Q3: transition pairs utilized for quantification; DP:
 122 declustering potential; CE: collision energy for each molecule; rt: Liquid Chromatography retention
 123 time in minutes. Asterisks indicate prostaglandins identified by comparison with their own standard
 124 reference. The non-asterisk one indicates the prostaglandins identified by comparison with literature
 125 data

126

127 **Aligned sequences, with gap, utilized in the phylogenetic analysis**

128

129 **Chaetoceros neogracile MMETSP0751-20121128_3624**

130 MVFPNFAQFLTDGFAK-----SIPG-----
131 TNFLRNNAHEADLCQIYGRNKDQTDCLRLMS-SSIGEEKGRLKSCMING-----
132 EEWPMPYFLESGEVDP-----QFERLDPPKFSFDHVIEQLSRIDPNGRGLVQ-
133 -----KIKKNIFA VGS DRGNTTPGVAAAFASLFLREHNRLAAEVENR-----NPA--
134 ----WDDGQIFQTARNINIVIYIRIIMEDYLNQLSHAP-----
135 ----FDFKLDPGPWTWDAKWNKK-----FRVA VEFSTVYR-WHSIIPNAIQIGE-----
136 GDPLPIMQAMFNNELLIDR-----TLLKSFEDSSSRATSFTPFNTAD-----
137 IMLPREYNTILQARQAGIRYYVDYLKQGFVPDLPKNYS DITS-DEKVQ-----
138 KMLEEMYGKG--NVDKVEFYIGAICATHERN-A----
139 PFSLFMNLQVAHDAIKAFYMNPLLQRSSWK----

140

141 **Asterionellopsis glacialis MMETSP0707-20130614_6459**

142 MILPTFANHLTGGFIK-----STRG--PDGS-----
143 YQYGRTHSNHNIDLLQLYGR TQEQTALRLRDNTTFGKKGRLKSQLNG----
144 GEEWPPFLYKDGRIDPQ-----FEVLDPIAHGMSHLVDIMEGD-----
145 -----RSKFFAMGSNQQNMLPQNIAMTTLFLREHN RVAGELERR-----YSD-----
146 WDDDRIFETARNIVIVIYLYKIVIEEYINHISPLP-----
147 DIVKFLVDPGPWMWNAEWNKP-----NWISAEFAILYR-WHALMPNKVMHA-----
148 GIVFDLSSTLFQHDLLLGED---RSLKKT FVDM SAQRSASCECLNTAG-----
149 PLVEREVESLRYSRALRFRPYVEYVEYWGLK--KPRRFEDITK-NAEVA-----
150 SMLRDLYGT---VDKVEFFVGLIASDHSKN-G----
151 IFGSSMNIGVGLDAFSQALTNPLLSEHVWK----

152

153 **Aulacoseira subarctica MMETSP1064-20121228_17645**

154 MLFPVFAQMLIDSFIS-----TVTYTNTTTK---
155 KLTFDWKRTNSPFEINLLPLYGRFEYQTDALRLKS-EKS--RGRLKSQIING-----
156 EEYAEFLYNDLGEIKD-----EFQVLGPPQA-FNQILSKVDKTEEEKR-----
157 -----KIKSKIFAFGGNRTNTTPQMAALNTLFLREHNRLAGQLEKH-----NED-----
158 WDDDRVFQTARNINLVIYLYKIIIEEYINHITSS-----
159 GARFKVQPKEWVWNADWNKP-----NWIAVEFAVL YR-WHALTPNSYLWE-----
160 GKRVKISDDL FNNALLLET---SGGLRQAI AEISKNPATIMAPFNTAL-----
161 ELLSQEQAAFAQTRQANVRPFADYRAYLGLP--LVKTFKDITQ-DKEVQ-----
162 QKLEELYKT---PDRVEFWAGLVAEDKDPK-A----
163 IFGPTLSTLVALDAFSQAL THPLLSEQVFN----

164

165 **Aulacoseira subarctica MMETSP1064-20121228_57453**

166 MLFPVFAQYLIDSFIL-----TKRIKTEGGT---VSTIDWKRSDSPNDIVLLPLYG-
167 FKNATDALRLKS-EEAQKRGKLSQLIDG-----EEYAPFLYDESGMVKE-----
168 EFKALG-EPESLGEILKAMMADSNTVK-----
169 KYKSKIFAFGLVRVNITPQLAAMNTLFLREHNRLAGVLEEN-----NPE-----
170 WDDERVFQTARNINLVIYLYKIVIEEYINHIS-----
171 GVQFKAKPEKWIWHASWNKP-----NWISVEFAVL YR-WHTLVPNSDEWN-----
172 GIRYDFQNEIFNNEILLD-----AGGLRHSFASISANRAPSICPFNTGG-----
173 FLLPREASALQQSRV NKL RPFGEYREYLYGK--KTKEFS DISS-DPQVQ-----

174 ETLRKLYKT---PDRVEFYAGLIAEDHVAK-G---
 175 ILAPIMTDLVAKDAFTQALHPLLSNVFN----
 176
 177 **Chaetoceros affine MMETSP0090-20130426_9955**
 178 MLFPTFAQHLIDSFIV-----TAVKSDGGSG---
 179 TEFEWKKTDSPHDIGLLPLYGRTFDQTKQLRVQN-PPRGKYGQLKSQIIHG-----
 180 EEYAPYLYDADGKVKK-----EFDLLE-TPQGLERSLSMLSPE--DAKA-----
 181 -----K-KSNIFAFGGARTNLVFNITAWNTLLLREHNRIAQTIEKE-----EPT-----
 182 WDDERVFQTARNVLLVIYLKLVVEEYINHITGY-----
 183 ---GIDFTVDPGKWMWNAPWYKR-----NWISAEFAVLYR-WHGVIPSCMKWG-----
 184 DKTLSTHESLFNNAVLTED--MKGSLRDTFINISNHRATQMNLFNTE-----
 185 MMVLRDMAALKQCRACKIKPYADYVVYLGTKR-RPTKFSISK-DKEVQ-----
 186 EALEKVYKK---VENVEFWTGLLASDNPPE-G---
 187 IMSPEMTTFVANDAFNQALCHPLLSNVWS----
 188
 189 **Chaetoceros curvisetus MMETSP0717-20131115_51459**
 190 MLFPTFAQHLIDSFIV-----TAIKENSSEG---
 191 VVFNWRKTGSPHDIGLLPLYGKTIEQTTQLRLKS-EVKGSKGRMKTQIIKG-----
 192 GEWSPYLFDSKGNKKA-----EFSSLP-DPEGLEFALMEAKKFGVDAEK-----
 193 -----N-KASVFALGGSRANLTPNIVAWNTLLLREHNSIAAKIEAE-----NPE-----
 194 WDDERVFQTARNVNLAVYLRLVIEEYINHITAF-----
 195 --GVDFTVEPGKWMWDAPWYKR-----NWISAEFAVLYR-WHAVIPNCMKWG-----
 196 EKTLPTAAYLYSNHLLDDT-GLKGDREAFINISNHRATSMEIHNSEK-----
 197 WMTGRDSRALQMSRECELSFTEYCAYLKGP--VPKTFADITS-DVEVQ-----
 198 NELKALYK---VENVEFWVGLIAKDHPTE-A---
 199 IMSAELTTFVANDAFNQALHPLLSSEHVWP----
 200
 201 **Chaetoceros debilis MMETSP0149-20130528_3320**
 202 MLFPTFAQHLIDSFID-----TVYHYDDDG---
 203 VVFDWKRKTETPHDIGLLTLYGKTIPETKQLRKQS-ETSGEKGLKLSQLVNG-----
 204 EEWAPFLYDSNGKVKE-----EFNELP-VPQGIDEKMYAARPQ-VQAKL-----
 205 -----K-ESIFAFFGGSRTNLTPNISAWNTLLLREHNRIAGLIEE-----NPT-----
 206 WDDERVFQTARNCTLVIYLRLVIEEYINHITAY-----
 207 -GVDFEIEPEKWMWDAPWYKR-----NWISAEFAVLYR-WHAVIPSLMKWG-----
 208 KNTHTTMDYLFNNLLSDD-GMKGNLRDCFHNICDHRAATNMQLHNSEGG-----
 209 FMVGRDKSALEMSRSCKLRSFSEYCGYLGTP--APESFADITQ-DKDLQ-----
 210 KELKDVEYGE---VKNVEFWTGLIAKDHSCA-A---
 211 IMSAELTKFVANDAFNQALHPLLSSEHVFN----
 212
 213 **Chaetoceros debilis MMETSP0150-20130528_3844**
 214 MLFPTFAQHLIDSFID-----TVYHYENDGN---
 215 IVFDRKLTGTPHDIGLLTLYGKTIPETKQLRKQS-DTSGEKEKICPNLLKE-----
 216 KSXAPFLYDSNGKVKE-----EFNKLP-VPQGIDGKMYAARPQ-VRAKL-----
 217 -----K-ESIFAFFGGSRTNLTPNISAWNTLLLREHNRIAGLIEE-----NPT-----
 218 WDDERVFQTARNCTLVIYLRLVIEEYINHITTY-----
 219 -GVDFKIEPEKWMWDAPWYKR-----NWISAEFAVLYR-WHAVIPSLMKWG-----
 220 KNTHTTMEYLFNNLLSDD-GMKGNLRDCFHNICDHRAATNMQLHNSEGG-----
 221 FMVGRDKSALEMSRSCELRSFSEYCGYLGSP--APESFADITQ-DKDLQ-----

222 KELKDVYGE----VKNVEFWTGLIAKDHSC-E-A----IMSAELTKFVANDAFNXGFD----
223 SSSPFR----
224
225 **Odontella MMETSP0015_2-20120614_17751**
226 LLFPTVAQHLIDSFN-----TKYTKTING-----
227 TFFDWARDTSPHEIGLSPLYGCPNKQTSQLRELN-EKIGCKGRLKTQILGD---
228 GGEEWAPFLYGDSGKIKAE-----EFNELD-EPVALDHILGFNSENAEMMR-----
229 -----KSIFAFGGSRANLNPNIWAWNTLLLREHNRVASKIEAS-----EPS-----
230 WDDERVFQTARNVVLVIYLVKIVIEDYVAHIS-----
231 GVNFKVQPGCEWWMWNAKWYKR-----NWMSVEFAVLYR-WHALIPNTTFWG-----
232 SKTASTKDALYNNTLLLNEQGCAGNLRNIMVQISEQRITSFQLNNTED-----
233 WLVDRLNALRQSRECNLASYADYCEYLGMD--RPKTFDISL-YPEVQ-----
234 EKLRSLYKT---ADKVEFYVGLIAADHSPGGK----
235 LFSEAMTRFVANDAFNQALANPLLESENWVK----
236
237 **Cyclotella meneghiniana MMETSP1057-20121228_4220**
238 LLFPTFAQHLIDSFN-----
239 TRIDHEATEKNGGRPVFDWARDTSKHEIGLSPLYGDTTEQLNQLRLLS-
240 NVNGFKGLMKTQIINE-----EEWAPFLYEVDGSKKK-----EFSAIW-
241 DPSGASYVLGMRRDVGERQK-----
242 RTLFAFGGARANLNPNIWAWNTLLLREHNRIAGEIEKS-----EPS-----
243 WDDERVFQTARNVLLVVIYLVKLVIEEYIAHIA-----
244 GVNFKVDPGEWWMWNAWAKT-----NWMSTEFAILYR-WHAVIPNTSSWGK-----
245 AKNLKVLDTLNFNDLLLDTKEGLSGNLRDAFVSISEERVTAHQLFNTEE-----
246 WMVDRELAAIKQGRANKVASADYAEYVDLP--RPKTFADISL-YPEVQ-----
247 KALEEVYGT---VDRVEFYVGLIAVDMGAGGK---IFSP-
248 MTKFVANDAFNQALTNPLLSMNWVN----
249
250 **Cyclotella meneghiniana MMETSP1057-20121228_12622**
251 LLFPTFAQHLIGSFN-----
252 TKIDHEATEKNGGQVIFDWAKTYSRHDIGLSPLYGDTTEQTDQLRLMS-
253 NVTGFKGRMKTQLIND-----EEWAPFLYRVDGTTKKP-----EFNAIW-
254 DPAGASHVLKMRGEEGFKQK-----
255 RTLFAFGGARANLNPNIWAWNTLLLREHNRLAGEIEKS-----EPS-----
256 WDDERVFQTARNVNIVIYLVKLVIEEYIAHIS-----
257 GADFKVDPGEWWMWNAWNA-----NWMSVEFAILYR-WHAIIPNTINWG-----
258 TSNMKVSDILFNDLLVKETEGLDANLRDVFVQMSEQRATAHQLFNTEE-----
259 WLLDRELAAVTQGRANKVASADYVEYLNLP--RPKTFADISL-YPEVQ-----
260 NALKEVYGT---VARVELYVGLIAADMAGGK---IFSL-
261 MTKFVANDAFNQALTNPLLSQNVVK----
262
263 **Pseudo-nitzschia arenysensis MMETSP0329-20121206_15499**
264 LLFPTVAQHLIDSFN-----THVDRKATEAKG-
265 SPVFNWAKTDSPEIGLSPLYGDSKEQTDQLRERS-EEVGRKGRLKSQMIEG-----
266 EEWSPFLYDSKGNKKT-----EFSKIP-DPSGMSHILKHVNAATAKKA-----
267 -----SIFAFGGGARANLNPNIWAWNVLLLREHNRLASMIEES-----EPT-----
268 WDDERVFQTARNVSIVIYLVKIVIEEYIKHIS-----
269 GANFRVAPGKWLWNAQWNKT-----NWMSVEFAILYR-WHAIIPNSSSWGP-----

270 SKDVEVRESLFNNTLLLDKTKGMGAKLADIFVQISNERTTSFELNNTK-----
 271 WLVDREMAAIKQGRNTNNVAPYADYCEYLGYE--RPKTFADINR-DPKVQ-----
 272 EQLKELYGT---PDKVEFYVGLIAGEHPSGGK---IFSK-
 273 MTSFVANDAFNQALGNPLLSQNVWE----
 274
 275 **Thalassionema frauenfeldii MMETSP0786-20121207_3776**
 276 LLFPTFAQHLIDSFN-----TRIKADTDKNK--
 277 PPEFEWTRTDSKHEIGISPLYGDEPAQTKQLRELS-ETSGRKGRLKTQMIEG-----
 278 EEWAPFLYNGN-NKKP-----EFS DIP-EPDGVRMIEKHG--FGDRS-----
 279 -----TIFAFGGRRANLNPNI VAWNTLLLREHNRLAAEIEKS-----EPE-----
 280 WDDERVFQTARNVLTVIYSKIVVEEYVGHIS-----
 281 GVPFEVQPGPMMWNAEWNKT-----NWM SVEFAILYR-WHALIPDTIRWGP-----
 282 KVDIGIMKQLFNNTLLLSKENGMGANLKDCFTEISRNRVTSFELFNTEGS-----
 283 YMATREMQAIRQCRAANVAPFADYCEYLGDP--RPKTFEDISR-KPEVQ-----
 284 KVLKELYGT---PDRVEFYVGLIAQDHSAGPK---
 285 IFGDVMTK FVANDAFNQALANPLLSQNVWE----
 286
 287 **Skeletonema marinoi MMETSP0320-20121206_19182**
 288 LLFPTFAQHLIDSFN-----TKLN-TETG-----
 289 QFEWDQTESKHEISLGPLYGDDVGQTNQLREKS-EVTGRRGRLKTQILDG---
 290 GEEWAHFLYDETGTKKE-----EFNLIR-DPDGMKHILKALYNSDPAAKS-----
 291 -----SIMQTIFAFGGRRVNLNPNMVAWNTLLLREHNRLAGEIERS-----EPS---
 292 ---WDDERVFQTARNVLIVIYLKLVIEYIGHIS-----
 293 --GVKFKLDPGEWMWNAPWYKT-----NWMSTEF SILYR-WHALIPNQNLGP-----
 294 SKDLGVLKSLFNPLLLDEEKGLGGNLRDVFVDICKARITSFQLFNTEK-----
 295 WMVGREAAATISQGRANNVQPYAKYCEYLRIK--PPKTFKDISL-VPEVQ-----
 296 QALEELYGT---PDRVEFYVGLIAADHAPG-K---IFSL-
 297 MTKFVANDAFNQALTNPLLSQNVWE----
 298
 299 **Skeletonema costatum MMETSP0013_2-20120614_556**
 300 LLFPTFAQHLIDSFN-----TKLN-TETG-----
 301 KFEWDQTESKHEISLGPLYGDDVGQTNQLREKS-EVTGRRGRLKTQILDG---
 302 GEEWAHFLYDETGTKKE-----EFNLIR-DPDGMKHILKALYSSDPAAKS-----
 303 -----SIMQTIFAFGGRRVNLNPNMVAWNTLLLREHNRLAGEIERS-----EPS---
 304 ---WDDERVFQTARNVLIVIYLKLVIEYIGHIS-----
 305 --GVKFKLDPGEWMWNAPWYKT-----NWMSTEF SILYR-WHALIPNQNLGP-----
 306 SKDLGVLKSLFNPLLLDKEKGLGGNLRDVFVDICKARITSFQLFNTEK-----
 307 WMVGREAAATINQGRANNVQPF AKYCEYLRIK--PPKTFKDISL-VPEVQ-----
 308 QALEELYGT---PDRVEFYVGLIAADHAPG-K---IFSL-
 309 MTKFVANDAFNQALTNPLLSQNVWE----
 310
 311 **Detonula confervacea MMETSP1058-20130122_27510**
 312 LLFPTFAQHLIDSFN-----TKTKPLESGRL----
 313 GFEWDQTF SKHEIGLAPLYGDYEEETLQLREKS-EASGRKGRLKTQVLDG---
 314 GEEWAPFLYDSEGTKKP-----EFSLLP-DPDGLQHIL---AMQPETME-----
 315 -----KKKKSIFAFGGRRANLNPNI VAWNTLLLREHNRLASEIEKS-----EPS-----
 316 WDDERIFQTSRNVVIVMYCKIIIEYIKHIS-----
 317 GVNLRVEPGEWMMWNASWYKP-----NWMSTEF FAILYR-WHALIPNTSSWGA-----

318 SKNIEVVDALFNNNLLISKDTGMGGNLRDAFVEISKTRITSFELFNTEK-----
 319 AMVQRETA AIEQGRFNNVASYADYCEYLELK--RPETFEDISL-KPEVQ-----
 320 QALKELYGS---PDRVEFYIGLIAADHPAGGK---
 321 VFSAAMTKFVANDAFNQALTNPLLSQNAWN---
 322
 323 **Skeletonema marinoi MMETSP1039-20121108_4563**
 324 LLFPTFAQHLIDSFN----TKIN-VETG-----EFEWDRDTSKHEIGIGPLYGDEVEQTSQLREKS-
 325 EKPGRRGRLKTQVLEG---GEEWAPFLYNEDGTKKK-----EFSAIH-
 326 DPDGMKITILGLVYSSDPTTKS-----
 327 SIEQSIFAFGGRRANLNPVAVWNTLLLREHNRLAGEIEKS-----EPG-----
 328 WDDERVFQTARNVLIVMYCKIVIEEYIKHIS-----
 329 GVNFKVEPGPMMWNAPWYKT-----NWMSTEFALYR-WHALIPNEAGLGP-----
 330 SKDAGVMEALFNNPMLLDDETGLGGNLRDIFVDISQTRVTSLQLFNTEK-----
 331 WMVERESAAINQGRANNVQSYAA YCEYLDIE--PPKTFEXISM-VPERQ-----
 332 QALKELYGT---PDRVEFYVGLIAADHPAGGK---IFSE-
 333 MTKFVANDAFNQALTNPLLSQNVWE---
 334
 335 **Skeletonema marinoi MMETSP0319-20121206_26582**
 336 LLFPTFAQHLIDSFN----TKIN-VETG-----EFEWDRDTSKHEIGIGPLYGDEVEQTSQLREKS-
 337 EKPGRRGRLKTQVLEG---GEEWAPFLYNEDGTKKKE-----FSAIEDGTKKEEFSAIH-
 338 DPDGMKITILGLVYSSDPTTKS-----
 339 SIEQSIFAFGGRRANLNPVAVWNTLLLREHNRLAGEIEKS-----EPG-----
 340 WDDERVFQTARNVLIVMYCKIVIEEYIKHIS-----
 341 GVNFKVEPGPMMWNAPWYKT-----NWMSTEFALYR-WHALIPNEAGLGP-----
 342 SKDAGVMEALFNNPMLLDDETGLGGNLRDIFVDISQTRVTSLQLFNTEK-----
 343 WMVERESAAINQGRANNVQSYAA YCEYLDIE--PPKTFEDISM-VPERQ-----
 344 QALKELYGT---PDRVEFYVGLIAADHPAGGK---IFSE-
 345 MTKFVANDAFNQALTNPLLSQNVWE---
 346
 347 **Skeletonema marinoi MMETSP0562-20121206_1237**
 348 LLFPTFAQHLIDSFN----TKIN-AETG-----EFEWDRDTSKHEIGIGPLYGDEVEQTSQLREKS-
 349 EKPGRRGRLKTQVREG---GEEWAPFLYNKDGTKKE-----EFSAIH-
 350 DPDGMKITILGLVYSSDPTTKS-----
 351 SIEQSIFAFGGRRANLNPVAVWNTLLLREHNRLAGEIEKS-----EPG-----
 352 WDDERVFQTARNVLIVMYCKIVIEEYIKHIS-----
 353 GVNFKVEPGPMMWNAPWYKT-----NWMSTEFALYR-WHALIPNEAGLGP-----
 354 SKDAGVMEALFNNPMLLDDETGLGGNLRDIFVDISQTRVTSFQLFNTEK-----
 355 WMVYRESAAINQGRANNVQSYAA YCEYLDIE--PPKTFEDITM-VPERQ-----
 356 QALKELYGT---PDRVEFYVGLIAADHPAGGK---
 357 IFSEAMTKFVANDAFNQALTNPLLSQNVWE---
 358
 359 **Skeletonema marinoi MMETSP1428-20130617_783**
 360 LLFPTFAQHLIDSFN----TKIN-AKTG-----
 361 EFEWDKNDKHEIGIGPLYGDEVEQTNQLREKS-EAPGRRGRLKTQVL-E---
 362 GGEEWAPFLYNEDGTKKK-----EFSAIH-DPDGMKITILGLVYSSDPTTKS---
 363 -----SIEQSIFAFGGRRANLNPVAVWNTLLLREHNRLAGEIEKS-----EPG---
 364 --WDDERVFQTARNVLIVMYCKIVIEEYIKHIS-----
 365 -GVNFKVDPGPMMWNAPWYKT-----NWMSTEFALYR-WHALIPNEAGLGP-----

366 SKDAGVMEALFNNPMLLDDKTGLGGNLRDIFVDISQTRVTSFQLFNTEK-----
367 WMVYRESAAINQGRANNVQSYAAAYCEYLDIE--PPKTFEDITM-VPERQ-----
368 QALKELYGT----PDRVEFYVGLIAADHPAGGK----
369 IFSEAMTKFVANDAFNQALTNPLLSQNVWE----
370
371 **Synedropsis recta MMETSP1176-20130426_10672**
372 PTKPWLQGQCRGMALR---
373 FPKRPLPDNLPDAKTLVEDFCIRPDDASFTPCANGVNSLTPYFALCVIHDFFRSDTGRSKV
374 GKLDLDR---PWVNLHSSYLQTLQTVYGYN-----KKSCASVRTQKHG-----
375 -----KLTEKKIVDHRLERMAVCHALILVMVKHHNFICDQLMER-----
376 YPD---KFKTDEMIFQTARLINCQVYINLIELYSCLFHVWN-----
377 -----EDGSNPVELRGLDYPRDIQGY-----HLSYEFNIMYR-FHAFIPKEWTPFYG-----
378 -----LKEAIAYLKKGDFTHSVPALSSDDAIKTLVDAMSHRAGANHVPSNVPR-----
379 -VMGPAEVKGIEDARLLGISTYNDFKEAVGEP--KYDTFLDMSGNRPELA-----
380 AKLEKHYPT---VDDVEFAVGMRVETRTPLRG----
381 AGFSTVGRAILADAFSSIRFDRFYSSQSEYH----
382
383 **Nostoc sp WP015113127**
384 MLFPYVWQWFTDSFLR----LDHTN-----
385 KLKNTSNHEIDL CNVYGLTRKQTHLLRSFQ-----
386 GGGKFKTQKLRQDGVVEEYPLFYADPAQG-----IVDPQFAGLYE-
387 PVNDE---KRQPAD-----
388 KKQYLFAMGVERANVQIGYVMLNTLCLREHNRLCDVLASN-----YPD-----
389 WDDERLFQTSRNILMAILKIIMEEYINHITPY-----
390 HFKLFADPEAFTK-ESWYRT-----NHMAIEFDFVYR-WHSAIPETFNCNG-----
391 KPTHVADTLWNNKILIDQG-----LGALMEETCSQAGTRIGLFNTPNI-----
392 LVDLAELPSIKLGRQLQLASYNDYREL CGFP--RVTSFDQITG-DEFTQ-----
393 QKLKEFYGH---VDNIEFFVGLYSEDVRQN-----STIPPVARLIGIDAFSQUALTNPLLSPKIFN---
394 -
395 **Chaetoceros cf neogracile MMETSP1336-20130426_1470**
396 LMLPIFAQWFTDSFLR----TKFRVDGPQD-----
397 FKENESNHEIDL CQIYGMTETQTAMLRSM-----NGFLKSQIIDK----EEFPAHLFEETASG---
398 LV-----QIKPEFAGLYT-EANVQRVFRRASKE-----
399 HKLNCFAVGIEHGNSTMGNTLMNIIWMREHNRIAREIAKA-----HPT-----
400 WDDERLFQTARNVNIVLLLKVVVEDYIFHIS-----
401 GPPFKMDN-KIAEGERWYRS-----NWMAAEFALLYR-WHDLIPDTVEFGG-----
402 ERKDSSALNKNNRWLMKVG-----VDRACLDASQEPSGKIMLGNTPNF-----
403 LVEVGKMS-LKQGRVSGLASFNDYCEHYGLK--RKKNFMDLTGNNKEVS-----
404 KKLEDVYGS---IDNLEWVGLWAEAYDDKTK---CMGP-
405 LTYMVGND AFTQALTNPLLA AEVFN----
406
407 **Grammatophora oceanica MMETSP0009_2-20130614_5553**
408 LLLPMFAQWFTDGFLR----TKWKPPTEQD-----
409 YVENESTHQIDL NQIYGATEVQTNILRAKK-----GGKLRSQMIKG----EEYPEYLFDTET-----
410 L-----QIKPEFEGLYS-DVNFTRVFQNVSDE-----
411 QKKVTFAVGLEHGNSTVANTIMNTLFLREHNRLAGIIEAA-----HKD-----
412 WDDERIFQTRNCLIVILMKIVINDYVAHIA-----
413 DAPVFLDAGGFAEDQPWYRE-----NWMSVEFNLLYR-WHDMIPDQVSFTSADQVSFS---

414 SEDEYWDMESSKTLVNNNTLMLELG-----VEQILLDASKQRAGQLGLFNTPDF-----
415 ---LLHVHKET-MKQCRMINLAPYVDYCKEYGLT--VPKTFRELTGGNLEYA-----
416 KALESVYDS---VEHVEWFVGLWAGG-KALGQ----
417 FSSDLLILMVGHDAVTQLFTNPLLSRRVYN----

418
419 **Gersemia fruticosa AAF93168**
420 WLFMFFAQHFHTEFFK-----TIYHS-PAFT-----WGN-
421 HGVDVSHIYGQDMERQNKLSFE-----DGKLSQTING-----EEWPPYLKDVDNVTMQY-----
422 -----PPNT-----
423 PEDQKFALGHFPFYSMLPGLFMYASIWLREHNRVCTILRKE-----HPH-----
424 WVDERLYQTGKLIITGELIKIVIEDYVNHLAN-----
425 YNLKLTYNP-ELVFDHGYDYD-----NRIHVEFNHMYH-WHPFSPDEYNISG-----
426 STYSIQDFMYHPEIVVKHG-----MSSFVDSMSKGLCGQMS-HHNHGA-----
427 YTLDVAVEVIKHQRELRMQSFNNYRKHFALE--PYKSFEELTG-DPKMS-----
428 AELQEVYGD---VNAVDLYVGGFFLEKGLTT-S---
429 PFGITMIAFGAPYSLRGLLSNPVSSPTYWK----

430
431 **Gersemia fruticosa AAS48061**
432 WLFMFFAQHFHTEFFK-----TIYHS-PAFT-----WGN-HGVDVSHIYGQDIERQNKLSRFQ--
433 ---DGKLSQMING-----EEWPPYLKDVDNVTMQY-----PPNT-----
434 -----SEDQKFALGHFPFYGMLPGLFMYASIWLREHNRVCTILRKE-----
435 -HPH-----WEDERLYHTGKLIITGELIKIVIEDYVNHLAN-----
436 -----YNLKITYDP-ELVFDHGYDYD-----NRIHVEFNHMYH-WHPFSPDEYNISG-----
437 -----STYSIHEFMYHPEIVVKHG-----MSSFVDSMSKGLCGQMS-HHNHGA-----
438 YTLDVAVEVIKHQRELRMQSFNNYREHFALE--PYKSFEEMTG-DPKMA-----
439 AELQETYGD---VNAVDLYVGGFFLEKGLTT-S---
440 PFGITMIAFGAPYSLRGLLSNPVSSPTYWK----

441
442 **Plexaura homomalla AAU87497**
443 WLFMFFAQHFTHQFFK-----TVHHS-PAFS-----WGN-
444 HGVDVSHIYGQGVRENKLRFAK-----DGKLSQMING-----EEYPPYLKDVDLKMQY-----
445 -----LENT-----
446 AEEQKFALGHPPFSMLPGLFMYATIWLREHNRVCMILRKE-----HPH-----
447 WEDERIYQTGKLIITGELIKIVIEDYVNHLAN-----
448 YNMKLRYDP-QLVFSRNYDYD-----NRIHLEFNHLYH-WHPFSPDQFNISG-----
449 TTYAIKDFMYHPEIVVKHG-----MSSFVNAMSSGLCGKMS-HHNHGQ-----
450 YTLDVAVEVIKYQRELRMQSFNXYRRHFGLH--AYKSFEEMTG-NPKMA-----
451 AELKEVYGD---VNAVDLYVGGFFLEKSLTT-S---
452 PFGITMIASGAPYSLRGLLSNPVSSPTYWK----

453
454 **Plexaura homomalla AAF93169**
455 WLFMFFAQHFTHQFFK-----TVHHS-PAFS-----WGN-
456 HGVDVSHIYGQGVRENKLRFAK-----DGKLSQMING-----EEYPPYLKDVDGLKMQY-----
457 -----LENT-----
458 AEEQKFALGHPPFSMLPGLFMFATLWLREHNRVCMILRKE-----HPH-----
459 WEDERIYQAKLIITGETIKIVIEDYINHLAN-----
460 YNMKLRYDP-QLVFSRNYDYD-----NRIHLEFNHLYH-WHPFSPDQFNISG-----
461 TTYTINDFMYHPEIVVKHG-----MSSFVNAMSSGLCGKMS-HHNHGQ-----

462 YTLDVAVEVIKYQRKLRMQSFNNYRRHFGLP--AYKSFEEMTG-DPKLA-----
 463 AELKEVYGD---VNAVDFYVGGFFLEKSLPT-S----
 464 PFGITMIASGAPYSLRGLLSNPVSSPTYWK----
 465
 466 **Gammarus sp ADB65786**
 467 LLFQYYAQHFTHQFFR----TNYTKGPQFT-----KGN-
 468 GGVDVSNIYGLTERQRRALRSNV-----DGKCLKFQING-----EHFPPYLKDVPGISMEY-----
 469 -----PPLP-----
 470 ITEDNKFALGHPPFALLPGLFVYSTIWMREHNRVCEVLKEQ-----HPH-----
 471 WDDERLYHTAKLIITGEVIKITIEDYVQHLSQ-----
 472 YKVDLKF KP-QVVHGTRFQFH-----NRINVEFDHLYH-WHPLIPEGIKVED-----
 473 SYYSLMDMAFSTKSVFTHG-----LDAFVKALVTNRAGKLT-SRNHSP-----
 474 VTVPVLKKMLENSRILRFQGVNQYRKKFNMR--PFRDFLDLTG-DEELA-----
 475 RDMEEMYGD---INAVEYYVGLIAEKDSP--S----
 476 LTPLTMVNVGGPWSVKGLIANPICSPHWWK----
 477
 478 **Caprella sp ADB65785**
 479 LLFQYYAQHFTHQFFR----TNYTMCPQFT-----KGN-
 480 GGVDVSNIYGLTEQHRRAIRMNS-----DGKCLKYQVIND-----EHYPPYLRDVQGIEMDY-----
 481 -----PCHIP-----
 482 ITEDNKFALGHPPFALLPGLFVVFSTIWMREHNRVCDVLKNQ-----HPD-----
 483 WDDERLYQTAKLIITGEVIKITIEDYVQHLSQ-----
 484 YKVDLKF KP-QVVHGTRFQFD-----NRINAEFNHLYH-WHPLIPDGIQVED-----
 485 KYYSLMDMAFSTKSVFTHG-----LDKFIESMATS RAGKLS-HSNHPL-----
 486 VTLPLVKKMMENGRKLR YQGINEYRKR FALK--PFKDFMDLTG-DEALA-----
 487 KDLQELYGH---VDAVEFYVGLL TEKDSP--S----
 488 LTPLTMVNVGGPWSVKGLIANPICSPHWWK----
 489
 490 **Nitzschia sp MMETSP0014_2-20120614_6887**
 491 PQLNFWMLSFVNWFHD----DNFRTLPTDTG-----
 492 AFTWSDRGLHMT HLYGHTEYRQAALRTMA----GDGKMKTS SRLG----
 493 WDYYPPLLMDVQADFPD-----FDMWTSQRGSKHKSTSAGQTSEQADEN----
 494 -----MPYYFAIGDPRFNLHLGHILWTSVGLYLHNTACVILQRE-----DPG-
 495 ----LTDEDIFQRRARVIVFHIIQKIRLQDFVMD SISSTR-----
 496 ---DHIRIPYDPKLLREEFAHHFAYSGGN-QPNFLEFNHLYQAWHALIPNGLVLNEDVD-----
 497 --DGEKDILPIRKT LWAPKLM TTNFT----IGEMATSFASTPLTLYSPHNFPVF-----
 498 LRGVTEAALKDERAQRMAPYNSYRELIGLD--PITSFEQFAVDDP-----
 499 QKMAELYNN---DVDSVDFIAGILADSNPHLPG---
 500 NFFGVQLV LVALFALQDLANNPLILD PVMS----
 501
 502 **Skeletonema marinoi MMETSP0319-20121206_1511**
 503 EADVALVANIMKRDTN--SFAPFNQLASAWIQFMT--HDWFQHDASSS--QGLK-----
 504 -----MKNVVTHWWDASQMYGS-----SQEEVDAVRAE-
 505 GGKHLHDVNDEIDYN-----
 506 ASLPITGFRENWWAGLHILHTIFAREHNHIVDILAQS-----YPS-----
 507 MTEDELFGTARNIIA VLLAKIHTVEWTP TLLDNAVSDMALNINWHGLQTVTSMYFKGKNIP
 508 DEVKDIIDEMK-----VPSVFGSNYTTTEQTLFNTP-----FYMTEEFVSVYR-
 509 MHTLLPDAMILEGGKT-----VSLQELAF TDARNLVSDPSKTTAT-----

510 LLQAFHAHTPAQALSCLKNYPKSLFNLQIGNGK-----
511 SINLAEIDISRDRARGIPRYNDARRQLLLT--PYKSMDDLTSDE-----ELKLLKSVY-
512 ---TDIEQVDFLVGCLVDKDRPDGF----AFGIVPYYIFVVMASRRLSDRFFQEGLTEENYS
513

514 **Pseudo-nitzschia pungens MMETSP1061-20121228_74475**

515 DNVQVIAQRLAREGFKPAGNQLNIAAAWIQAQV--
516 HSWIQHLDGAPTSIEATAEAVGPVCPVKK--MNFETE-ERP-----GE----
517 YNSFRTQWWDASFVYGQ-----
518 NREQVHLGRLYKDGKLVNESNPDTLSFV-----
519 EEGKSKIDVVGQDQSNVWVGTVLQVLFKEHNYCAEMIKKE-----NPQ-----
520 LTDEEYGHCRNIALVAKIHTIDWTVLLKTEQLRVAMEINWKG-----
521 ATKA VFGDKAPF-----HP--LRLINKPKADNKGVP-----FCLTEEFAAVYR-
522 LHPLLPPGLVVEHG-----EGKEEFIEKLLTTKGRDKMREPG---
523 MAKKIMFSAFHYSCHLSSNYPFIMRKFPTDTHKGVLDLQPPEDRVVDMAAIDLHRDRER
524 GIQKYNEFRRQLKLR--PFKTWEALTGEENKSDN-----
525 ELTDAKKLELIYGPAPEGIEKIDLLVGDLYERKIADGF----
526 ALSETSFMIFLLMASRRIDSDPYLNEYYN EYYT
527

528 **Thalassiosira rotula TR754|c0_g1_i1|m.3431**

529 MLFPTFAQHLIDSFIV-----TAIKENTSEG----
530 VEFDWRTGSPHDIGLLPLYGKTIEQTTQLRLKS-ELKGSKGKLTQIIKG-----
531 GEWSPYLFDAKGNKKP-----EFSSLP-DPEGLEFALGMAKQHGVDVKK-----
532 -----N-KATVFALGGSRANLTPNIVAWNTLLLREHNRIASKIEE-----NPK----
533 --WDDERVFQTARNVNLA VYLRVIEEYINHITGF-----
534 ---GVDFTVEPGKWMWDAPWYKR-----NWISAEFAVLYR-WHAVIPNCMKWG-----
535 EKTLPASYLYSNHLLDED-GMKGDLREAFINISDHRATSMQIHNTES-----
536 WMKGRDARALQMSRACELRSFTEYCA YLGAK--VPKTFADITS-DVEVQ-----
537 NELKALY GK---VENIEFWVGLIAKDHPTE-S---
538 IMSAELTSFVANDAFNQAL THPLLSEHVWAAGPE
539

540 **Skeletonema marinoi MMETSP0920-20130426_33289**

541 LLFPTFAQHLIDSFIN-----TKIN-VETG-----EFEWDRDTSKHEIGIGPLYGDEVEQTSQLREKS-
542 EKPGRRGRLKTQVLEG---DEEWAPFLYNEDGTKKE-----EFSAIH-
543 DPDGMKITILGLVYSSDPTTKS-----
544 SIEQSIFAFGGRRANLNPVAVWNTLLLREHNRLAGEIEKS-----EPG-----
545 WDDERVFQTARNVLIVMYCKIVIEEYIKHIS-----
546 GVNFKVDPGPWMWNAPWYKT-----NWMSTEFALYR-WHALIPNEAGLGP-----
547 SKDAGVMESLFNNPVLLDDETGLGGNLRDIFVDISQTRITSFQLFNTEK-----
548 WMVERESAANQGRANNVQSYAA YCEYLDIE--PPKTFEDISM-VPERQ-----
549 QALKELYGT---PDRVEFYVGLIAADHPAGGK---
550 IFSEAMTKFVANDAFNQALTNPLLSQNVWE----
551

552 **Fragilariopsis kerguelensis MMETSP0907-20130614_13332**

553 PNVQVVAQRLARESFTPAGDQLNIVAAGWIQAMV--
554 HDWMKHEDGKKTSMVTPAVVGSQCPLHR--FNLFETK-ERP-----
555 YNSERTNWWDA SFVYGQ-----
556 NAEVQVQNSRAVGGKLVNEKNPDTLPSR-----DDG---
557 TDLTGQDQSNVWVGSVLQTLFLYEHNYCAEQIAKE-----NPN-----

558 LTDHQIYGHCRNIIAALVAKIHTIDWTVELLKTPQLKIGMRTNWMG-----
559 IIQAITGLKIPF-----LDRLRLRIKKKENNEGVP-----FCLTEEFAAVYR-
560 LHPLLPPGLIIEGEGA----GDKDEDEDTFIDLRTLTTKGRDLMRKSG---
561 MAKKVMKAVFTYPCGNMAPSNYPDVMRDFHPTDLLGNNL----
562 DDRIDLAAIDLFRDRERGIQYFNNFRKLSMK--
563 PFQTWEELTGDDKMTEEALAAFVAGTGNLTNAKKLELVYGAAPKIEHCDDLVDLYEK
564 KIP-GF----AISETSFMIFLLMASRRLDADPYLNEYFDEEHYT
565
566 **Skeletonema grethea MMETSP0578-20130828_1168**
567 ILVAYSIXSFQYVLDFKPAASQLNVLAASWIQAMV--HDWIGHFDGDET--
568 ETLDRGGESLCPFAKSPFSFKNTKTEKIDGVPF-----SPSERTNWWDASFVYGN-----
569 -----NSEQIDRARTMQGGKMVTS-IPHALAED-----KDG---
570 VYFAGDNKNSWVGVALLODLFIREHNYICDQIAAE-----AKEEGK--
571 EMTDEELFGKARVVVAALVAKIHTVDWTVELLKTKLLAIGMKTNWDG-----
572 LLKAVG-IPPG-----ILSQMGEKKGRVSDNEGTP-----FCLTEEFAAVYR-
573 LHSLSPGLILGDG-----DAKDKFIGLEDLLGDEGRKQMRETKT--
574 RPKEMMKSLHWPCGALMSSNYPNAFRDVAPTDDYGKDLK---
575 SQNIDLAALDLFRDRERGILKFNEFRQLNLK--PYRTWLELTENEE-----
576 DARKLELIYGPGQEGIERCDLLVGDMYERKVQPSF----
577 ALSETSFIFLLMASRRLSADPFLNELYNEETYS
578
579 **Staurosira sp MMETSP1361-20130828_23824**
580 LLFPFFAQWFVDSFLR-----TKWKPLAEQD-----
581 FKENESNHDIDLNQIYGTSEIQTDMRLSMK-----GGRLKSQIIDG-----EEYPVFLFDQKT-----
582 A-----TLKPEFRGLYT-EENFKRVFGNASKE-----
583 HKLHSFAVGLEHGNSTIGNTVMNTLFLREHNRVAGVISAA-----HPE-----
584 WDDERVFQTTTRNVMIVLLIKIVLADYIYHIS-----
585 GAAVFADPGGFAEDELWYRE-----NWMSVEFSLLYR-WHDLIPSSVTFDG-----
586 ETRDAVDLQNNNRWLLKAG-----MDSVIQDASNQKAGVMGLGNTPDF-----
587 LLSVTKMS-LHMARTCKLSSYVEYCKEYQGD--PPEDFMDLTG-DQDSA-----
588 SKLERVYGS---IDKVEWVGLFAQK-RDV-----
589 FSGLLMTLMVGNDAVTQAFTNPLLAKRVYN----
590 **Skeletonema marinoi MMETSP1428-20130617_31509**
591 EADVALVANMMKRDTN--SFAPFNQLASAWIQFMT--HDWFQHDASS--QGLK-----
592 -----MQNVVTHWWDASQMYGS-----SQEEVDAVRAE-
593 GGKLHLDVNDEIDYN-----
594 ASLPITGFRENWWAGLHILHTVFAREHNHIVDILAQS-----YPS-----
595 MTEDELFGTARNIIAALLAKIHTVEWTP-----KNIPDEVKDIIDEMK-----
596 -----VPSVFGSNYTTEQTLFNT-----FYMTEEFVSVYR-MHHLLPDEMILEGGK-----
597 -LQELAFTDARNLVSDPSKTTAT-----LLQAFAPHTPAQALSLSKNYPKSLFNLQIGNGK-----
598 --LINLAEIDISRDRARGIPRYNDARRQLLLT--PYKSMDDLTSKDE-----
599 ELKLLKSVY---ADIEQVDFLVGCLVDKDRPDGF---AFGI--YYIFVVMASRRLSDRFFQ----
600 ----
601 **Homo sapiens NP000953**
602 LMFAFFAQHFTHQFFK-----TSGKMGPGFT-----
603 KALGHGVDLGHYGDNLERQYQLRLFK-----DGKLYQVLDG-----EMYPPSVVEEAP-
604 VLMHY-----PRGI-----
605 PPQSQMAVGQEVFGLLPGLMLYATLWLREHNRVCDLLKAE-----HPT-----

606 WGDEQLFQTTRLILIGETIKIVIEEYVQQLSG-----
 607 YFLQLKFDL-ELLFGVQFQYR-----NRIAMEFNHLYH-WHPLMPDSFKVGS-----
 608 QEYSYEQLFNLSMLVDYG-----VEALVDAFSRQIAGRIGGGRNMDH-----
 609 HILHVAVDVIRESRMRLQPFNEYRKRFGMK--PYTSFQELVG-EKEMA-----
 610 AELEELYGD---IDALEFYPLGKLLLEKCHPN-S---IFGESMIEIGAPFSLKGLLGNPICSPEYWK--
 611 --
 612
 613 **Ovis aries NP001009476**
 614 LMFAFFAQHFTHQFFK----TSGKMGPGFT-----
 615 KALGHGVDLGHYGDNLERQYQLRLFK-----DGKLYQMLNG-----EVYPPSVEEAP-
 616 VLMHY-----PRGI-----
 617 PPQSQMAVGQEVFGLLPGLMLYATIWLREHNRVCDLLKAE-----HPT-----
 618 WGDEQLFQTARLILIGETIKIVIEEYVQQLSG-----
 619 YFLQLKFDL-ELLFGAQFQYR-----NRIAMEFNQLYH-WHPLMPDSFRVGP-----
 620 QDYSYEQLFNLSMLVDYG-----VEALVDAFSRQPAGRIGGGRNIDH-----
 621 HILHVAVDVIKESRVLRLQPFNEYRKRFGMK--PYTSFQELTG-EKEMA-----
 622 AELEELYGD---IDALEFYPLGKLLLEKCHPN-S---
 623 IFGESMIEMGAPFSLKGLLGNPICSPEYWK---
 624
 625 **Homo sapiens NP000954**
 626 MMFAFFAQHFTHQFFK----TDHKRGPAFT-----
 627 NGLGHGVDLNHIYGETLARQRKRLFK-----DGKMKYQIIDG-----EMYPTVKDTQ-
 628 AEMHY-----PPQV-----
 629 PEHLRFAVGQEVFGLVPLMMYATIWLREHNRVCDVVKQE-----HPE-----
 630 WGDEQLFQTSRLILIGETIKIVIEDYVQHLSG-----
 631 YHFKLKFDP-ELLFNKQFQYQ-----NRIAAEFNTLYH-WHPLLPDTFQIHD-----
 632 QKYNQYQFIYNNLSILLEHG-----ITQFVESFTRQIAGRVAGGRNVPP-----
 633 AVQKVSQASIDQSRQMKYQSFNEYRKRFMLK--PYESFEELTG-EKEMS-----
 634 AELEALYGD---IDAVELYPALLVEKPRPD-A---
 635 IFGETMVEVGAPFSLKGLMGNVICSPAYWK---
 636
 637 **Coccyzus truncatus AFN20596**
 638 VLIAYYAQWVTHQFFN----TDESPTGHS-----
 639 VKQPVGVNMSMLYGSKQEVKSVRAYK-----GGLLKSTIKNG-----QEFPEIMPCQEGSRIP--
 640 -----
 641 GKEMFNMPILIANMIPGFAAIHVLFFRRHQYICRELAKW-----AEAQGK--
 642 NIDDEELFQKAKLIVTVNMLRITMHDYVSRALQSSHAKMR-----
 643 -----FDQKVKQSRIWKMFDPYFPPS----NAIQFEFNIFYR-WHQFYPTTKIMKR-----
 644 IDDLKFPKSKQQLDEKWNVAVRWIAD---PDGMERVLFSASSQRAGKLSLLNTNQW-----
 645 -----IVEHVVKPGLARCREHQLASYNDYREKVGFP--RLTTFEQVTS-NPALL-----
 646 EKLKRVYRN---VDQIEYYPGVFAEDKHFG-----
 647 NVHGPLTFGSSMTFTGIFSSRLFETALDE---
 648
 649 **Ciona intestinalis XP002127674**
 650 VLFPFFAQHFTHMFFK----TDPKMGMPYQ-----WGD-
 651 QLVDSLQIYGHGKQRQHELRSV-----NGKLVSLVDG-----HEFPP-LSNQTTANMSN-----
 652 -----INLL-----
 653 PQEYQFVFGHQGFSMLPTFLIWSTIWLREHNRICDLIKEE-----NPA-----

654 WDDERIFQTARLVLTGETIKVVIEDYVQHLSG-----
655 FHYKLLYDP-ELVQGGSHSFH-----NQIHVEFQLLYH-WHALMPDQIEFNG-----
656 KSYTMKRLLFNPEPVVKGK-----LKRTIEDLSNQWAGQVAGGKTQGA-----
657 ATLVHAGLAIKNGRDLRMQSFNAYKEKFEMK--KYTTFQELTG-EEEMA-----
658 AELQKLYGD---IDAVEYYIGIMLEKRRSP-Q---
659 LFGETLTEMGSPYSLKGLYSNPINHKDWWK----
660
661 **Ciona intestinalis XP002123273**
662 VFFAFFAQHFTHQFFK-----TNTIKGMPFQ-----WGE-HSVDLSHVYGHYTIQRQHELRSHI--
663 ---DGKLVFETNG-----EVFPP-LTESANVTMSG-----EKLM-----
664 -----RGR-KFAIGHPGFGAFPSFFVIATLWLREHNRVCDILKDL-----
665 HPD-----WDDERLFQTARLILTGETLKIIVEDYVQHVSG-----
666 -----FHFQLSYDP-EILHKSTFSYN-----NQIHAEFHILYH-WHMLMPDFIELGE-----
667 --HVYPLKELLFNVDPVVEIG-----METVLKQLSNQFAGKVVGGRNQGP-----
668 ELVAVVELALKQTRQMRMCSFNKYRERFGMK--PYTSFEELTG-ETEVA-----
669 ALLRNLYYD---IDALELFVGYFVEHRRNR-Q---
670 VLGATMLEMGAPYSLKGVFGNPIGSPAWWK----
671
672 **Oncorhynchus mykiss CAC10360.1**
673 LMFAFFAQHFTHQFFK-----TRNSMGLGFT-----
674 RALGHGVDAGNVYGDNLVRQLNLRLLK-----DGKMKYQVVKG-----EVYPPTVAEAA-
675 VNMRY-----PQET-----
676 PVGQRMAIGQEVFGLLPGLTMYATLWLREHNRVCDILKAE-----HPT-----
677 WGDEQLFQTARLIVIGETIRIVIEEYVQHLSG-----
678 YLLDLKFDP-VLLFKSTFQYR-----NRIAVEFKQLYH-WHPLMPDSFHIDG-----
679 DEVPYSQFIFNTSIVTHYG-----VEKLVDAFSRQCAGQIGGGRNHHP-----
680 VVTNVAEGVIEESRTLRLQPFNEYRKRFLNK--PYTSFSDFTG-EEEMA-----
681 RELEELYGD---IDALEFYPAIMLEKTRPN-A---
682 IFGESMVEMGAPFSLKGLLGNPICSPEYWK----
683
684 **Fistulifera solaris GAX23950.1**
685 LPPVKDVAAILRRPMNPTSVAPFNQIAVAWIQMMT--HDWFQHDP AHPDQK-----
686 -----MNRVTHWWDASQLYGS-----
687 TLAQQTAVRVPNTGKVRDLQHQELNYTT-----
688 TGIPITGFADNWWAGLHMMHTLFRVREHNWLVDQFERQ-----YPG-----
689 VYTANDKFQLARLCLLSALLAKIHTVEWTPTLNDNPVAALGLHTNWRG---
690 VDAILEYGTRFELQLAYRIVGGD-----QSVPHAGNGTTRETLYNTT-----
691 FAMTEEFVAVYR-MHPLLPEMEIEGKTFS-----LNDLSFVDARTLTKS-VKTTQT-----
692 LLQAFGMTPANTLSLQNYPRQLYGLEKPGMS-----
693 QPVNLAEIDLQRDRERNLPRYNDMRRQLLLK--PYKRLEDLTDDDET-----
694 ELNLLKSVY---QDMDQVDLMVGCLVDKDRPYGF---
695 AFGIVPFHVFLVMASRRILNDRFFMEDFNAKVYT
696
697 **Cyanothece sp ACB53655.1**
698 PNPRVVSRTLMTREDFKPATILNLLAAAWIQFEN--
699 HDWFHSHGDNKPEDKLEIPLEANDPWPEEYRPLEVGKTLPTDSRPEGAKPP--
700 TFINTVTHWWDGSQLYGS-----NPETVDQLRSHEDGKLIIGENGLLPVDP---
701 -----ETGVDITGFNDNWWIGLGMHLTLFTREHNLICDHLKQE-----YSQ---

702 ---WSDDDLFDHARLINAALMAKIHTVDWTPAILPLPATDIALNVNWNWNGFLG-----
 703 EDIKQVLGTVG-----EGEIADLLTGIIGS-DKNHHTAP-----YYLTEEFVSVYR-
 704 MHPLIPDELEFRSLEGD-----KFLQEVNFFEMSGKRTRALLESISL----
 705 PDLFYSFGITHPGAITLHNYPRFLQQLVRDNGE-----
 706 VFDLAAVDILRDRERGVPRYNRFREIMGRG--RVKCFEEISSNKQ-----
 707 WVEEMRRVYN---DNIDQVDLMVGLFAEDT-PEGF----
 708 GFSDTAFRVFILMASRRLKSDRFFTTDYRAEIYT
 709
 710 **Fistulifera solaris GAX18717.1**
 711 LPPVKDVAAILRRPENPSTVAPFNQIAVAWIQMMT--HDWFQHDPTNPDSK-----
 712 -----MNRVTHWWDASQLYGS-----
 713 TLTQQRAVRVPNTGKLRLDKHQELNYTS-----
 714 TGIPITGFADNWWAGLHMMHTLFVREHNWLADQFELQ-----YPG-----
 715 VYTANDKFQLTRLCLSALLAKIHTIEWTPTLDDNPVAALGLHTNWRG---
 716 VDAILEYGTQFELQLAYRIVGGD-----QSVPHAGNGTTRETMYNTT-----
 717 FAMTEEFVAVYR-MHPLLDDMEIEGITLT-----LNDLSFVDARKLTKS-VKTTQT-----
 718 LLQAFGTTPANTLSLQNYPRQLYGLEKPGMS-----
 719 QPINLAEIDLQRDRERNLPRYNMRRQQLLK--PYKRLEDLTDDDET-----
 720 ELNLLKSVY---QDIDQVDLMVGLVDKDRPYGF----
 721 AFGIVPFHVFLVMASRRILNDRFFMEDFNAKVYT
 722
 723 **Alcanivorax nanhaiticus WP_035233415.1**
 724 PNPREVSNLIMSRGGDFKPATTLNFIATSWIQFMV--
 725 HDWFDHGPRTDANPIEFPLPAGDVLGGGTMSVQRTRPDPDVSDEGIITY--ENIN--
 726 THWWDGSQLYGS-----SKEKNDEVRSFVDGKLVGDGRLPTEF-----
 727 -----FSGKPVTGFNENWWVGLSMLHHIFTQEHNAIADMLVAN-----YPG-----
 728 QSDQWYFDKARLINSALMAKIHTVEWTPAILANPVLERAMYANWWGLGGDRDKRDYQ
 729 DDLDNLNNNLGQIGLLDLVGIDNGLGDSPTGSLEHALAGLVGSRTPNNYNVP-----
 730 YTLTEEFVSVYR-MHPLLDRDEIKVYDIGSN-----VVDEEIALEDTRNGDAEDLLGDIGQ---
 731 DRLWYSFGITHPGALTLNNYPDFLRNLSMPLIG-----
 732 DIDMAAIDILRDRERGVPRYNEFRRQIGLK--PLTSFEQLTSDPQ-----
 733 LLADLKALYN---NDIELVDTLVGQLGEETRPEGF----
 734 GFGETSFQIFILNASRRLMTDRFFTTDYTDEVYT
 735
 736 **Hordeum vulgare subsp. Vulgare BAJ90503.1**
 737 PDPFVVATKLLARREYKDTGKQFNILAAAWIQFMV--
 738 HDWMDHMEDTKQIEITAPKEVANECPLKS--FKFYATKEQPTNSDGIKT---
 739 GYHNIRTAWWDGSAVYGN-----NEKQEKKIRTYADGKLVIGDD-GLLLHE-
 740 -----ENG---VPLSGDVRNGWVGISILQALFVKEHNAVCDAIKEE-----
 741 HPN-----LSDEELYRYAKLVTSAVIAKIHTIDWTVELLKTKTLRAGMRANWYG-----
 742 LLGKKIKDTFGHIG-----GTALGGLVGLKKPINHGVP-----YSLTEEFVSVYR-
 743 MHPLIPSTLKLDRDPTGQPA--ADNSPPYLEDIDIGELVGLKGEDQLSKIG----
 744 FEKQTLSMGYQACGALELWNYPSSFRDLIPQNLDTGTRNS----
 745 DRIDLAALEVYRDRERSVPRYNEFRRLFLI--PIKCWEDLTSDND-----
 746 AIEAIRAIYG---DDVEKLDLLVGLLAEKKIK-GF----
 747 ASETAFNIFILMASRRLEADRFLTSNFNEKTYT
 748
 749

750 **Triticum aestivum CDM84254.1**
751 PDPFVVATKLLARREYKDTGKQFNILAAAWIQFMV--
752 HDWMDHMEDTKQIEITAPKEVANECPLKS--FKFYATKEQPTNSDGIKT---
753 GYNNVRTAWWDGSAVYGN-----NEKQAEKTRTYVDGKLVIGDD-
754 GLLLHE-----ENG---VPLSGDVRNGWVGVSILQALFVKEHNAVCDAIKED-
755 -----HPN-----LSDEELYRYAKLVTSAVIAKIHTIDWTVELLKTKTMRAAMRANWYG-----
756 ----LLGKKFKDTFGHIG-----GTALGGLVGLKKPINHGVP-----YSLTEEFTSVYR-
757 MHSLIPTTLKLRDPTGQPA--ANNSPPYLEDIDIGELVGLKGEDQLSKIG----
758 FEKQTLSMGYQACGALELWNYPSPFRDLIPQNLDTGTRNS----
759 DRIDLAALEVYRDRERSVPRYNEFRRLFLI--PIKSWEDLTSDKD-----
760 AIESIRAIYG--DDVEKLDLLVGLMAEKKIK-GF----
761 ASETAFNIFILMASRRLEADRFITSNFNEKTYT
762
763 **Rhizophagus clarus GBC07128.1**
764 LELEHYITALTQLPLN----
765 TPEEKLLSILERLLVTQLWSDISKPPAMIAGDIYRSSDGSYGNRLIQSLGKANSRYSLSRSI
766 QYPIKLSVLPKSEDIFDKIMVQGGDFVEHPSGISSMLFYLAIIITHDLFHTSFADPNINLTSSYL
767 DLTPLYGSNDQEKSIRTLKG-----
768 GLLKPDTFADSRILLQPPGVSAVLVILFSRNHNFIQAQTLLKKNELGRFSVTNPNDPEQLKKQDE
769 DLFQATARLINCIFYINVILHNYLRILGLDR-----
770 TNSKWFVDPTVPYNKRGQLEPLPSGIGNIVSLEFNYYR-WHAATSKDDSKFVEDEFKTIFG-
771 ----
772 DDWENITIDEFKEKMGVWGRSIPKDPKWKFNHIERGSDNRFKDTDIAKEIINGTKKVSGA
773 FGANRIPKVFRPIELLGIESARVLGLSSLNDFRRSLNLK--PYESFMEMNP-DPLIA-----
774 -KKLEELYGS----
775 IENVELYPGLMTEKTKPDMLGSAIALPFTISRILSDAVNLVRNDRYYTNDFPSRNL
776
777 **Emiliana huxleyi XP_005780718.1**
778 PFQALHNLPSADEVVR-----LLYKRDAFK-----
779 KAPYGVNSLATWFANVAIHDFRTAT---GTDGGTHPERGSDK--
780 EWVNLHSSYLDLQPLYGYS-----KTTADATREWS-----
781 -----GGKLKAFAEDRMRRIPESRVIVELLRREHNYVAEQLAQR-----YPAQ----
782 FATDEELYQQARLIMGGVYINIILRAYGCQMFG-----
783 ---EIAPDGSGFCEL RQGYGGAGVG---NMCTFNFNLIYR-FHTSIPVEWSATD-----
784 PPPIDTDEQMR-----TLLNGILNWESGGFGPNNVPDSILG-----
785 ERARVSQRAIEAARLMGAPTLNTRRRRFTSG---YSSFEDMTGGDQATA-----
786 DTLRQLYPG---GIEDVELLVGCQVEKCMSSGW----
787 ALPSTIGQAIVADAFASIRQDRFYTQDWGASSYT
788
789 **Rhodococcus gordoniae WP_064063209.1**
790 VWFAFFAQWFTDSFLR-----TNSKD-----
791 PRKNDSNHEIDLQCQYGVNQAKTMLRAGY-----GGRLDSQVIDK-----
792 KEYPPFLFAARTPGEEL-----RFVVPKFEGLDREYLLDVTVLRRLCPDE-----
793 -----RKKSVFVAVGLEHGNSTIGNTVLNVFLREHNRIAGILEGA-----YPEWD---
794 ---DDRLFETTRNIMIVILLKIVIEEYIKHIG-----
795 PFDFPIEFVP-FMADNAPWNRT-----NWCAIEFNLLYR-WHSLIPDTVVFDS-----
796 QRVSTRILVDNNPLVLDRG-----IESIIDQCSRQKASRIGLGNTPAFLIDRHPMCPD-----
797 RESVEERT-VGLMRQARLRSFNDRYEAFLG--RLTSFTELTG-DVEVQ-----

798 QKLARLYGD----VDAVEWYVGIFAED-YPRHR----

799 MMGELLTTMVAHDAFTQAFTNPLLARHVYHEDTF

800

801 **Herbidospora mongoliensis WP_066371138.1**

802 VMFMFFAQWFTDSFLR-----TSRDD-----FRRNTSTQEIDLCQIYGLTEEKTHLLRAHR--

803 ----DGRLKSQLIDG----EYPEFLFRPRSSGEPP-----

804 VFKPEFEGLDHSEFIVSRLLLEDAPEK-----

805 QKDTFFAVGLEHGNSTIGNTIMNVVFLREHNRIAGLLKQA-----

806 HPEWAERPEGADARLFETTRNIMIVLLLKL VVEEYIRHIS-----

807 -----PYDFPLETVP-FIADGKRWNRS-----NWISIEFNLLYR-WHSLVPSTIGSGP-----

808 ----DRLDSTDFRNNNPLVLARG-----IESLVSQCSGERAGRIGLMNTPGFLVDRDPAHPE-----

809 ---RPSVQERT-IALSRKARLASYN DYRENFGLG--RLKDFGRLTQ-DAELR-----

810 ERLEKLYDD---IDKLEWYVGIFAED-YSRDE----

811 MMGRLMTTMVAYDAFTQALTNPLLARDVYNEATF

812

813 **Eutrema halophilum BAJ34623.1**

814 ---MVVATKLLTRRK MIDTGKQFN MIAASWIQFMI--

815 HDWVDHLEDTDQIELSAPKEAAKGCPLSS--FRFFKTKEVPTGFFEIKT---

816 GSLNTRTPWWDSSVIYGS-----NSKTLERVRTYKDGKLGKISEETGLLLHD---

817 -----DDG---LAISGDIRNSWVGVSALQALFIKEHNAVCDLLKKE-----YED--

818 ---LEDEDLYRHARLVTS AVIAKIHTIDWTVELLKTDTLLAGMRANWYG-----

819 LLGKKFKDTRFGHVG-----SSIFGGVVGMMKKPQNHGVP-----YSLTEEFTSVYR-

820 MHSLLPDQLHMRDIDVTPG--PNKSLPLTQEVSM EKLIGREGEETMSQIG----

821 FTKLMVSMGHQACGALELMNYPAWFRDLVPQDPNGHDRP----

822 DHIDLA ALEIYRDRERNVARYNDFRRAMFMI--PIKTWEDLTDDKE-----

823 AIELDDVYG---GDVDEL DLLVGLMAEKKIK-GF----

824 ASETAFNIFLLMATRRLEADRFFTSDFNEMTYT

825

826 **Pisum sativum CAH05011.1**

827 PDPMVVVTKLLERKTYKDTGTQFN VIAASWIQFMI--

828 HDWIDHMEDTKQVELSAPSEVASQCPLKS--FKFFKTKEIPTGFYDIKT---

829 GHANVRTPWWDGSSVYGS-----NEQVLNKVRTFKDGKLGKISKE-

830 GHLLHN-----EDG---TAISGDIRNSWAGVTTLQTLFVQEHN AVCDALKKE-

831 -----NSD-----LEDEDLYRHARLVTS AVIAKIHTIDWTVELLKTDTLLAGMRANWYG-----

832 ---LLGKQFKDRFGHVG-----NSILSGFVGMKRSENHGVP-----YSLTEEFATVYR-

833 MHPLLPDSLHLRDISASPG--PNKSPPLIKEIPMNDLIGLQGEKTLLEIG----

834 NAKKLVSMGHQACGALELWNYP SWLRNLVPHNIDGTERS----

835 DHVDLA ALEVYRDRERNVARYNQFRRGLLLI--PISKWEDLTDDDEE-----

836 AIKVLEEVYG---DDVEELDVLVGLMAEKKIK-GF----

837 ASETAFVIFLLMASRRLEADRFFTSNFNEETYT

838 **Original sequences, ungapped, utilized in the phylogenetic analysis**

839

840 **Chaetoceros neogracile MMETSP0751-20121128_3624**

841 MVFPNFAQFLTLDGFAKSIPGTNFLRNNATHEADLCQIYGRNKDQTDCLRLMSSSIGE
842 KGRLKSCMINGEWPMPYFLESGEVDPQFERLDPPKFSFDHVIEQLSRIDPNGRLGLVQKIK
843 KNIFAVGSDRGNTTPGVAAFASLFLREHNRLAAEVENRNPAWDDGQIFQTARNINIVYIRII
844 MEDYLNQLSHAPDFKLDPGPWTWDAKWNKKFRVAVEFSTVYRWHSIIPNAIQIGEGDPL
845 PIMQAMFNNELLIDRTLLKSFEDSSSQRATSFTPFNTADIMLPREYNTILQARQAGIRYYVD
846 YLKQFGFPVDLPKNYSDITSDEKVQKMLEEMYGKGNVDKVEFYIGAICATHERNAPFSLF
847 MNLQVAHDAIKAFYMNPLLQRSSWK

848

849 **Asterionellopsis glacialis MMETSP0707-20130614_6459**

850 MILPTFANHLTGGFIKSTRGPDGSYQYGRTHSNHNIDLLQLYGRTEQTLALRLRDN
851 TTFGKKGRLKSQILNGGEEWPPFLYKDGRIQPEVLDPIAHGMSHLVDIMEGDRSKFFAM
852 GSNQQNMLPQNIAMTTLFLREHNRVAGELERRYSDWDDDRIFETARNIVIVIYLYKIVIEEYI
853 NHISPLPDIVKFLVDPGPWMWNAEWNKPNWISAEFAILYRWHALMPNKVMHAGIVFDLSS
854 TLFQHDLLGEDRSLKKTFFVDMASQRSASCECLNTAGPLVEREVESLRYSRALRFRPYVEY
855 VEYWGLKKPRRFEDITKNAEVASMLRDLYGTVDKVEFFVGLIASDHSKNGIFGSSMNIGVG
856 LDAFSQALTNPLLSEHVWK

857

858 **Aulacoseira subarctica MMETSP1064-20121228_17645**

859 MLFPVFAQMLIDSFISTVYTNNTTKLTFDWKRTNSPFEINLLPLYGRFEYQTDALR
860 LKSEKSRGRLKSQIINGEEYAEFLYNDLGEIKDEFQVLGPPQAFNQILSKVDKTEEEKRRIKS
861 KIFAFGGNRTNTPQMAALNTLFLREHNRLAGQLEKHNEWDDDR VFQTARNINLVIYLYKI
862 IIEEYINHITSSGARFKVQPKWVWNADWNKPNWIAVEFAVLYRWHALTPNSYLWEGKRV
863 KISDDL FNNALLETSGGLRQAIAEISKNPATIMAPFN TALELLS QEQA AFAQTRQANVRPF
864 ADYRAYLGLPLVKTFKDITQDKEVQKLEELYKTPDRVEFWAGLVAEDKDPKAI FGPTLS
865 TLVALDAFSQAL THPLLSEQVFN

866

867 **Aulacoseira subarctica MMETSP1064-20121228_57453**

868 MLFPVFAQYLIDSFILTKRIKTEGGTVSTIDWKRS DSPNDIVLLPLYGFKNATDALRL
869 KSEEAQKRGKLSQLIDGEEYAPFLYDESGMVKEEFKALGEPESLGEILKAMMADSNTVK
870 KYKSKIFAFGLVRVNITPQLAAMNTLFLREHNRLAGVLEENNP EWDDERVFQTARNINLVI
871 YLKIVIEEYINHISG VQFKAKPEKWIWHASWNKPNWISVEFAVLYRWHTLVPNSDEWNGI
872 RYDFQNEIFNNEILLDAGGLRHSFASISANRAPSICPFNTGGFLLPREASALQ QSRV NKL RPF
873 GEYREYLG YKKTKEFS DISSDPQVQETLRKLYKTPDRVEFYAGLIAEDHVAKGILAPIMTDL
874 VAKDAFTQAL THPLLSENVFN

875

876 **Chaetoceros affine MMETSP0090-20130426_9955**

877 MLFPTFAQHLIDSFIVTAVKSDGGSGTEFEWKKTDS PHDIGLLPLYGRTFDQTKQLR
878 VQNPPRGKYGQLKSQIIHGEEYAPYLYDADGKVKKEFDLLETPQGLERSLSMLS PEDAKAK
879 KSNIFAFGGARTNLVPNITAWNTLLLREHNRIAQTIEKEEPTWDDERVFQTARNVLLVIYLYK
880 LVVEEYINHITGYGIDFTVDPGKWMWNAPWYKRNWISAEFAVLYRWHGVIPSCMKWGD
881 KTLSTHESL FNNAVLTEDMKGSLRDTFINISNH RATQMNLFNTESMMVLRDMAALKQCRA
882 CKIKPYADYVVYLGTKERPTKFS DISKDKEVQEAL EKVKVENVEFWTGLLASDN PPEGI
883 MSPEMTTFVANDAFNQALCHPLLSENVWS

884 **Chaetoceros curvisetus MMETSP0717-20131115_51459**
885 MLFPTFAQHLIDSFIVTAIKENSSEGVVFNWRKTGSPHDIGLLPLYGKTIEQTTQLRL
886 KSEVKGSKGRMKTQIIKGGEWSPYLFDSKGNKKAEFSSLPDPEGLEFALMEAKKFGVDAE
887 KNKASVFALGGSRLNTPNIVAWNTLLLREHNSIAAKIEAENPEWDDERVFQTARNVNLA
888 VYLRLVIEEYINHITAFGVDFTEVEPGKWMWDAPWYKRNWISAEFAVLRYRWHAVIPNCMK
889 WGEKTLPTAAYLYSNHLLDDTGLKGDLEAFINISNHRATSMEIHNSEKWMTGRDSRAL
890 QMSRECELSFTEYCAYLKGPVPKTFADITSDVEVQNELKALYGKVENVEFWVGLIAKDH
891 PTEAIMSAELTTFVANDAFNQALHPLLESHVWP

892
893 **Chaetoceros debilis MMETSP0149-20130528_3320**
894 MLFPTFAQHLIDSFIDTVYHYDDDGNVVFWDWKRTETPHDIGLLTLYGKTIPETKQLR
895 KQSETSGEKGKLSQLVNGEEWAPFLYDSNGKVKEEFNELPVPQGIDEKMYAARPQVQA
896 KLKESIFAFFGGSRTNLTNPISAWNTLLLREHNRIAGLIEENPTWDDERVFQTARNCTLVIY
897 LRLVIEEYINHITAYGVDFEIEPEKWMWDAPWYKRNWISAEFAVLRYRWHAVIPSLMKWGK
898 NTHTTMDYLFSNNLLSDDGMKGNLRDCFHNICDHRATNMQLHNSEGGFMVGRDKSALE
899 MSRSCKLRSFSEYCGYLGTPAPESFADITQDKDLQKELKD VYGEVKNVEFWTGLIAKDHSC
900 EAIMSAELTKFVANDAFNQALHPLLESHVFN

901
902 **Chaetoceros debilis MMETSP0150-20130528_3844**
903 MLFPTFAQHLIDSFIDTVYHYENDGNVVFDRKLTGTPHDIGLLTLYGKTIPETKQLRK
904 QSDTSGEKEKICPNLLKEKSXAPFLYDSNGKVKEEFNKLVPVQGDGKMYAARPQVRACL
905 KESIFAFFGGSRTNLTNPISAWNTLLLREHNRIAGLIEEDNPTWDDERVFQTARNCTLVIYLR
906 LVIEEYINHITTYGVDFKIEPEKWMWDAPWYKRNWISAEFAVLRYRWHAVIPSLMKWGKN
907 THTTMEYLFSNNLLSDDGMKGNLRDCFHNICDHRATNMQLHNSEGGFMVGRDKSALEM
908 SRSCELRSFSEYCGYLGSPAPESFADITQDKDLQKELKD VYGEVKNVEFWTGLIAKDHSC
909 AIMS AELTKFVANDAFNXGFDSSSPFR

910
911 **Odontella MMETSP0015_2-20120614_17751**
912 LLFPTVAQHLIDSFINTKYTKTINGTFFDWARDTSPHEIGLSPLYGCNPKQTSQLREL
913 NEKIGCKGRLKTQILGDGGEEWAPFLYGDGSKKAEFNELEPVALDHILGFNSENAEMMR
914 KSIFAFFGGSRLNTPNIVAWNTLLLREHNRVASKIEASEPSWDDERVFQTARNVVLVIYLKI
915 VIEDYVAHISGVNFKVQPGEW MWN AKWYKRNWMSVEFAVLRYRWHALIPNTTFWGSKTA
916 STKDALYNN TLLL NKEQGCAGNLRNIMVQISEQRITSFQLNNTEDWL VDRDLNALRQSRE
917 CNLASYADYCEYLGMDRPKTFTDISLYPEVQEKLRSLYKTADKVEFYVGLIAADHSPGGK
918 LFSEAMTRFVANDAFNQALANPLLESEN VWK

919
920 **Cyclotella meneghiniana MMETSP1057-20121228_4220**
921 LLFPTFAQHLIDSFINTRIDHEATEKNGGRPVFDWARDTSKHEIGLSPLYGDTEEQLN
922 QLRLLSNVNGFKGLMKTQIINEEEWAPFLYEVDGSKKKEFS AIWDPSGASYVLGMRRDVG
923 ERQKRTLFAFGGARANLNP NIVAWNTLLLREHNRIAGEIEKSEPSWDDERVFQTARNVLLV
924 IYLKLVIEEYIAHIAGVNFKVDPGEW MWNAPWAKTNWMSTEFALRYRWHAVIPNTSSWG
925 KAKNLKVLDTL FNNDLLLD TKEGLSGNLRDAFVSISEERVTAHQLFNTEEWMVDRELA AI
926 KQGRANKVAS YADYAEYVDLPRPKTFADISLYPEVQKALEEVYGTVD RVEFYVGLIAVD
927 MGAGGKIFSPMTKFVANDAFNQALTNPLLSMN VWN

928
929 **Cyclotella meneghiniana MMETSP1057-20121228_12622**
930 LLFPTFAQHLIGSFINTKIDHEATEKNGGQVIFDWAKTYSRHDIGLSPLYGDTTEQTD
931 QLRLMSNV TGFKGRMKTQLINDEEWAPFLYRVDGTTKPEFN AIWD PAGASHVLKMRGEE

932 GFKQKRTLFAFGGARANLNPNIWA WNTLLLREHNRLAGEIEKSEPSWDDERVFQTARNVNI
933 VIYLKLVIEEYIAHISGADFKVDPGEWMWNADWNKANWMSVEFAILYRWHAIIPNTINWG
934 TSNMKVSDILFNNDLLVKETEGLDANLRDVFVQMSEQRATAHQLFNTEEWLLDRELA AVT
935 QGRANKVASYADYVEYLNLPKPTFADISLYPEVQNALKEVYGTVARVELYVGLIAADM
936 GAGGKIFSLMTK FVANDAFNQALTNPLLSQNVWK
937

938 **Pseudo-nitzschia arenisensis MMETSP0329-20121206_15499**

939 LLFPTVAQHLIDSFINTHVDRKATEAKGSPVFNWAKTDS PHEIGLSPLYGDSKEQTD
940 QLRERSEEVGRKGRLKSQMIEGEEWSPFLYDSKGNKKTEFSKIPDPSGMSHILKHVNAATA
941 KKASIFAFGGARANLNPNIWA WNVLLLREHNRLASMIEESEPTWDDERVFQTARNVSIVIY
942 LKIVIEEYIKHISGANFRVAPGKWLWNAQWNKTNWMSVEFAILYRWHAIIPNSSSWGPSKD
943 VEVRESLFNNTLLLDKTKGMGAKLADIFVQISNERTTSFELNNTTEKWLVDREMAAIKQGR
944 NNVAPYADYCEYLG YERP KTFADINRDPKVQEQLKELYGTPDKVEFYVGLIAGEHPSGGKI
945 FSKMTSFVANDAFNQALGNPLLSQNVWE
946

947 **Thalassionema frauenfeldii MMETSP0786-20121207_3776**

948 LLFPTFAQHLIDSFINTRIKADTDKNKPPEFEWTRTDSKHEIGISPLYGDEPAQTKQLR
949 ELSETSGRKGRLKTQMIEGEEWAPFLYNGNKKPEFSDIPEPDGVRMIEKHGFGDRSTIFAF
950 GGRRANLNPNIWA WNTLLLREHNRLAAEIEKSEPEWDDERVFQTARNVLTVIYSKIVVEEY
951 VGHISGVPFEVQPGPMMWNAEWNKTNWMSVEFAILYRWHALIPDTRWGPVKVDIGIMKQ
952 LFNNTLLLSKENGMGANLKDCFTISRNRVTSFELFNTEGSYMATREMQAIRQCRAANVAP
953 FADYCEYLGDPKPTFEDISRKPEVQKVLKELYGTPDRVEFYVGLIAQDHSAGPKIFGDVM
954 TKFVANDAFNQALANPLLSQNVWE
955

956 **Skeletonema marinoi MMETSP0320-20121206_19182**

957 LLFPTFAQHLIDSFINTKLNTETGQFEWDQTESKHEISLGPLYGDDVGGQTNQLREKSE
958 VTGRRGRLKTQILDGGEEWAHFLYDETGTCKKEEFNLIRDPDGMKHKALYNSDPAKSSI
959 MQTIFAFGGRRVNLNPNMVA WNTLLLREHNRLAGEIERSEPSWDDERVFQTARNVLIVIYL
960 KLVIQEYIGHISGVKFKLDPGEWMWNAPWYKTNWMSTEF SILYRWHALIPNQNGLGPSKD
961 LGVLKSLFNPLLLDDEEKGLGGNLRDVFVDICKARITSFQLFNTEKWMVGREAAATISQGRA
962 NNVQPYAKYCEYLRIKPPKTFKDISLVPEVQQALEELYGTPDRVEFYVGLIAADHAPGKIFS
963 LMTK FVANDAFNQALTNPLLSQNVWE
964

965 **Skeletonema costatum MMETSP0013_2-20120614_556**

966 LLFPTFAQHLIDSFINTKLNTETGKFEWDQTESKHEISLGPLYGDDVGGQTNQLREKSE
967 VTGRRGRLKTQILDGGEEWAHFLYDETGTCKKEEFNLIRDPDGMKHKALYSSDPAKSSI
968 MQTIFAFGGRRVNLNPNMVA WNTLLLREHNRLAGEIERSEPSWDDERVFQTARNVLIVIYL
969 KLVIQEYIGHISGVKFKLDPGEWMWNAPWYKTNWMSTEF SILYRWHALIPNQNGLGPSKD
970 LGVLKSLFNPLLLDKEKGLGGNLRDVFVDICKARITSFQLFNTEKWMVGREAAATINQGR
971 ANNVQPF AKYCEYLRIKPPKTFKDISLVPEVQQALEELYGTPDRVEFYVGLIAADHAPGKIF
972 SLMTK FVANDAFNQALTNPLLSQNVWE
973

974 **Detonula confervacea MMETSP1058-20130122_27510**

975 LLFPTFAQHLIDSFINTKTKPLESGRLGFEWDQTF SKHEIGLAPLYGDYEEETLQLRE
976 KSEASGRKGRLKTQVLDGGEEWAPFLYDSEGTKKPEFSLLPDGLQHILAMQPETMEKK
977 KKSIFAFGGRRANLNPNIWA WNTLLLREHNRLASEIEKSEPSWDDERIFQTSRNVVIVMYCK
978 IIIIEYIKHISGVNLRVEPGEWMWNASWYKPNWISTEFAILYRWHALIPNTSSWGASKNIEV
979 VDALFNNTLLISKDTGMGGNLRDAFVEISKTRITSFELFNTEKAMVQRETA AIEQGRFNNV

980 ASYADYCEYLELKRPFEDISLKPEVQQALKELYGSPDRVEFYIGLIAADHPAGGKVFSAA
981 MTKFVANDAFNQALTNPLLSQNAWN

982

983 **Skeletonema marinoi MMETSP1039-20121108_4563**

984 LLFPTFAQHLIDSFINTKINVETGEFEWDRTDSKHEIGIGPLYGDEVEQTSQLREKSEK
985 PGRRGRLKTQVLEGGEWAPFLYNEDGTTKKEFSAIHDPDGMKTILGLVYSSDPTTKSSIE
986 QSIFAFGGRRANLNPNIWA WNTLLLREHNRLAGEIEKSEPGWDDERVFQTARNVLIVMYCK
987 IVIEEYIKHISGVNFKVEPGPMMWNAPWYKTNWMSTEFAILYRWHALIPNEAGLGPSKDA
988 GVMEALFNNPMLLDDETGLGGNLRDIFVDISQTRVTSLQLFNTEKWMVERESAAINQGRA
989 NNVQSYAAAYCEYLDIEPPKTFEXISMVPERQQALKELYGTPDRVEFYVGLIAADHPAGGKI
990 FSEMTKFVANDAFNQALTNPLLSQNVWE

991

992 **Skeletonema marinoi MMETSP0319-20121206_26582**

993 LLFPTFAQHLIDSFINTKINVETGEFEWDRTDSKHEIGIGPLYGDEVEQTSQLREKSEK
994 PGRRGRLKTQVLEGGEWAPFLYNEDGTTKKEFSAIEDGTTKKEFSAIHDPDGMKTILGLV
995 YSSDPTTKSSIEQSIFAFGGRRANLNPNIWA WNTLLLREHNRLAGEIEKSEPGWDDERVFQT
996 ARNVLIVMYCKIVIEEYIKHISGVNFKVEPGPMMWNAPWYKTNWMSTEFAILYRWHALIP
997 NEAGLGPSKDAGVMEALFNNPMLLDDETGLGGNLRDIFVDISQTRVTSLQLFNTEKWMVE
998 RESAAINQGRANNVQSYAAAYCEYLDIEPPKTFEDISMVPERQQALKELYGTPDRVEFYVGL
999 IAADHPAGGKIFSEMTKFVANDAFNQALTNPLLSQNVWE

1000

1001 **Skeletonema marinoi MMETSP0562-20121206_1237**

1002 LLFPTFAQHLIDSFINTKINAETGEFEWDRTDSKHEIGIGPLYGDEVEQTSQLREKSEK
1003 PGRRGRLKTQVREGGEWAPFLYNKDGTTKKEFSAIHDPDGMKTILGLVYSSDPTTKSSIE
1004 QSIFAFGGRRANLNPNIWA WNTLLLREHNRLAGEIEKSEPGWDDERVFQTARNVLIVMYCK
1005 IVIEEYIKHISGVNFKVEPGPMMWNAPWYKTNWMSTEFAILYRWHALIPNEAGLGPSKDA
1006 GVMEALFNNPMLLDDETGLGGNLRDIFVDISQTRVTSFQLFNTEKWMVYRESAAINQGRA
1007 NNVQSYAAAYCEYLDIEPPKTFEDITMVPERQQALKELYGTPDRVEFYVGLIAADHPAGGKI
1008 FSEAMTKFVANDAFNQALTNPLLSQNVWE

1009

1010 **Skeletonema marinoi MMETSP1428-20130617_783**

1011 LLFPTFAQHLIDSFINTKINAETGEFEWDKNDKHEIGIGPLYGDEVEQTNQLREKSE
1012 APGRRGRLKTQVLEGGEWAPFLYNEDGTTKKEFSAIHDPDGMKTILGLVYSSDPTTKSSI
1013 EQSIFAFGGRRANLNPNIWA WNTLLLREHNRLAGEIEKSEPGWDDERVFQTARNVLIVMYC
1014 KIVIEEYIKHISGVNFKVDPGPMMWNAPWYKTNWMSTEFAILYRWHALIPNEAGLGPSKD
1015 AGVMEALFNNPMLLDDETGLGGNLRDIFVDISQTRVTSFQLFNTEKWMVYRESAAINQGR
1016 ANNVQSYAAAYCEYLDIEPPKTFEDITMVPERQQALKELYGTPDRVEFYVGLIAADHPAGGK
1017 IFSEAMTKFVANDAFNQALTNPLLSQNVWE

1018

1019 **Synedropsis recta MMETSP1176-20130426_10672**

1020 PTKPWLQGQCRGMALRFPKRPLPDNLPAKTLVEDFCIRPDDASFTPCANGVNSL
1021 TPYFALCVIHDFFRSDTGRSKVGLDRPWVNLHSSYLDLQTVYGYNKKSCASVRTQKHGK
1022 LTEKKIVDHRLEMAVCHALILVMVKHHNFICDQLMERYPDKFKTDEMIFQTARLINC GV
1023 YINLIIELYSCLFHVWNEDGSNPVELRGLDYPRDIQGYHLSYEFNIMYRFHAFIPKEWTPFY
1024 GLKEAIAYLKKGDFTHSVPALSDDAIKTLVDAMSHRAGANHVPSNVPRVMGPAEVKGIE
1025 DARLLGISTYNDKFKA VGEPKYDTFLDMSGNRPELAAKLEKHYPTVDDVEFAVGMRVETR
1026 TPLRGAGFSTVGRAILADAFSSIRFDRFYQSSEYH

1027 **Nostoc WP015113127**
1028 MLFPYWVQWFTDSFLRLDHTNKLKNTSNHEIDL CNVYGLTRKQTHLLRSFQGGKF
1029 KTQKLKRQDGVVEEYPLFYADPAQGI VDPQFAGLYEPV NDEKRQPADKKQYLFAMGVE
1030 RANVQIGYVMLNTLCLREHNRLCDV LASNYPDWDDERLFQTSRNILMAILKIIMEEYINHI
1031 TPYHFKLFADPEAFTKESWYRTNHMAIEFDFVYR WHSAIPETFNCNGKPTHVADTLWNNK
1032 ILIDQGLGALMEETCSQAGTRIGLFNTPNILVDLAELPSIKLGRQLQLASYNDYRELCGFPRV
1033 TSFDQITGDEFTQQKLKEFYGHVDNIEFFVGLYSEDVRQNSTIPPVARLIGIDAFSQALTNP
1034 LSPKIFN
1035

1036 **Chaetoceros cf neogracile MMETSP1336-20130426_1470**
1037 LMLPIFAQWFTDSFLRTKFRVDGPQDFKENESNHEIDL CQIYGMTETQTAMLRSM
1038 NGFLKSQIIDKEEFPAHLFEETASGLVQIKPEFAGLYTEANVQRVFRRASKEHKLNCFAVGI
1039 EHGNTMGNTLMNIIWMREHNRIAREIAKAHPTWDDERLFQTARNVNIVLLLKVVVEDYI
1040 FHISGFPFKMDNKIAEGERWYRSNWMAAEFALLYRWHDLIPDTVEFGGERKDSSALNKNN
1041 RWLMKVGVDRACLDASQEPSGKIMLGNTPNFLVEVGKMSLKQGRVSGLASFN DYCEHYG
1042 LKRKKNFMDLTGNNKEVSKKLEDVYGSIDNLEWVGLWAEAYDDKTKCMGPLYMVG
1043 DAFTQALTNP LLAAEVFN
1044

1045 **Grammatophora oceanica MMETSP0009_2-20130614_5553**
1046 LLLPMFAQWFTDGLR TKWKPPTEQDYVENESTHQIDL NQIYGATEVQTNILRAKK
1047 GGKLR SQMIKGEEYPEYLFDTETLQIKPEFEGLYSDVNFTRVFQNV SDEQKKVTFAVGLEH
1048 GNSTVANTIMNTLFLREHNRLAGIIEAAHKDWDDERIFQTRNCLIVILMKIVINDYVAHIA
1049 DAPVFLDAGGFAEDQPWYRENWMSVEFNLLYRWHDMIPDQVSFTSADQVSFSSEDEYWD
1050 MESSKTLVNNNTLMLELGVEQILLDASKQRAGQLGLFNTPDFLLHVHKETMKQCRMINLA
1051 PYVDYCKEYGLTVPKTFRELTGGNLEYAKALESVYDSVEHVEWVFWGLWAGGKALGQFSS
1052 DLLILMVGHDAVTQLFTNPLLSRRVYN
1053

1054 **Gersemia fruticosa AAF93168**
1055 WLFMFFAQHF THEFFKTIYHSPAFTWGNHGV DVSHIYGQDMERQNKLR SFEDGKL
1056 KSQTINGEEWPPY LKDV DNVTM QYPPNTPEDQKFALGH PFYSMLPGLFMYASIWLREHNR
1057 VCTILRKEHPHWDERLYQTGKLIITGELIKIVIEDYVNHLANYNLKLTYNPELVFDHGYDY
1058 DNRIHVEFNHMYHWHPFSPDEYNISGSTYSIQDFMYHPEIVVKHGMSSFVDSMSKGLCGQ
1059 MSHHNHGAYTLDVAVEVIKHQRELRMQSFNNYRKHFAL EPYKSFEELTGDPKMSAELQE
1060 VYGDVNAV DLYVGGFFLEKGLTTS PFGITMIAFGAPYSLRGLLSNPVSSPTYWK
1061

1062 **Gersemia fruticosa AAS48061**
1063 WLFMFFAQHF THEFFKTIYHSPAFTWGNHGV DVSHIYGQDI ERQNKLR SFQD GKLK
1064 SQMINGEEWPPY LKDV DNVTM QYPPNTPEDQKFALGH PFYGMPLPGLFMYASIWLREHNR
1065 VCTILRKEHPHWEDERLYHTGKLIITGELIKIVIEDYVNHLANYNLKITYDPELVFDHGYDY
1066 DNRIHVEFNHMYHWHPFSPDEYNISGSTYSIHEFMYHPEIVVKHGMSSFVDSMSKGLCGQ
1067 MSHHNHGAYTLDVAVEVIKHQRELRMQSFNNYREHFALEPYKSFEEMTGDPKMAAELQE
1068 TYGDVNAV DLYVGGFFLEKGLTTS PFGITMIAFGAPYSLRGLLSNPVSSPTYWK
1069

1070 **Plexaura homomalla AAU87497**
1071 WLFMFFAQHF THQFFKTVHHS PAFSWG N HGV DVSHIYGQGV ERENKLRAF KD GKL
1072 KSQMINGEEYPPY LKDVDDLKM QYLENTAE EQKFALGH PFFSMLPGLFMYATIWLREHNR
1073 VCMILRKEHPHWEDERLYQTGKLIITGELIKIVIEDYVNHLANYNMKLR YDPQLVFSRNYD
1074 YDNRIHLEFNHLYHWHPFSPDQFNISGTTYAIKDFMYHPEIVVKHGMSSFVNAMSSGLCGK

1075 MSHHNHGQYTLDVAVEVIKYQRELRMQSFNXYRRHFGLHAYKSFEEMTGNPKMAAELK
1076 EVYGDVNAVDFYVGGFFLEKSLTTSFPGITMIASGAPYSLRGLLSNPVSSPTYWK

1077

1078 **Plexaura homomalla AAF93169**

1079 WLFMFFAQHFTHQFFKTVHHSWPAFSWGNHGVDVSHIYGQGVERENKLRAFKDGKL
1080 KSQMINGEEYPPYLKDVDGLKMQYLENTAEEQKFALGHPPFSMLPGLFMFATLWLRHNR
1081 VCMILRKEHPHWEDERIYQTAKLITGETIKIVIEDYINHLANYNMKLRYDPQLVFSRNYDY
1082 DNRIHLEFNHL YHWHPFSPDQFNISGTTYTINDFMYPHPEIVVKHGMSSFVNAMSSGLCGKM
1083 SHHNHGQYTLDVAVEVIKYQRKLRMQSFNNYRRHFGLPAYKSFEEMTGDPKLAAELKEV
1084 YGDVNAVDFYVGGFFLEKSLTTSFPGITMIASGAPYSLRGLLSNPVSSPTYWK

1085

1086 **Gammarus sp ADB65786**

1087 LLFQYYAQHFTHQFFRTNYTKGPQFTKGNNGVDVSNIYGLTERQRRALRSNVDGK
1088 LKFQIINGEHFPPYLKDVPGISMEYPPHLPITEDNKFALGHPPFALLPGLFVYSTIWMREHNR
1089 VCEVLKEQHPHWDDERLYHTAKLITGEVIKITYEDYVQHLSQYKVDLKFQVVGTRFQ
1090 FHNRIINVEFDHLYHWHPLIPEGIKVEDSYSLMDMAFSTKSVFTHGLDAFVKALVTNRAG
1091 KLTSRNHSPVTVPVLLKMLENSRILRFQGVNQRKKNMRPFRDFDLTGDEELARDMEE
1092 MYGDINAVEYYVGLIAEKDPSLTPMTMVNNGPWSVKGLIANPICSPHWWK

1093

1094 **Caprella sp. ADB65785**

1095 LLFQYYAQHFTHQFFRTNYTMCPQFTKGNNGVDVSNIYGLTEQHRRAIRMNSDGK
1096 LKYQVINDEHYPPYLKDVQGIEMDYPPHIPITEDNKFALGHPPFALLPGLFVSTIWMREHN
1097 RVCDVLKNQHPDWDDERLYHTAKLITGEVIKITYEDYVQHLSQYKVDLKFQVVGTRFQ
1098 QFDNRINAEFNHL YHWHPLIPDGIQVEDKYSLMDMAFSTKSVFTHGLDKFIESMATSRAG
1099 KLSHSNHPLVTLPLVLLKMMENGRKLR YQGINEYRKRFAKPFKDFMDLTGDEALAKDLQ
1100 ELYGHVDAVEFYVGLLTKDPSLTPMTMVNNGPWSVKGLIANPICSPHWWK

1101

1102 **Nitzschia sp. MMETSP0014_2-20120614_6887**

1103 PQLNFWMLSFVNWFHDDNFRTLPD TDGAF TWSDRGLHMT HLYGHTEYRQAALR
1104 TMAGDGKMKTSRSLGWDYYP LLM DVQADFPDFDMWTSQRGSKHKSTSAGQTSEQADE
1105 NMPYYFAIGDPRFNLHLGHILWTSVGLYLHNTACVILQREDPGLTDEDIFQRARVIVFHIIQ
1106 KIRLQDFVMSISSTRDHIRIPYDPKLLREEFAHHFAYS GGNQPNFLEFNHLYQAWHALIPN
1107 GLVLNEDVDDGEKDILPIRKT LWAPKLMTTNTFTIGEMATSFAS TPLTLYSPHNFPVFLRGVT
1108 EAALKDERAQRMAPYNSYRELIGLDPITSFEQFAVDDPQKMAELYNNDVDSVDFIAGILAD
1109 SNPHLPGNFFGVQLVLVALFALQDLANNPLILD PVMS

1110

1111 **Skeletonema marinoi MMETSP0319-20121206_1511**

1112 EADVALVANIMKRDTNSFAPFNQLASAWIQFMTHDWFQHDASSSQGLKMKNVVT
1113 HWWDASQMYGSSQEEVDAVRAEGGKLHLDVND EIDYNASLPITGFRENWWAGLHILHTI
1114 FAREHNHIVDILAQSYPSMTEDELFGTARNIIAVLLAKIHTVEWTP TLLDNAVSDMALNIN
1115 WHGLQTVTSMYFKGKNIPDEVKDIIDEMKVPSVFGSNYTT EQTLFNTPFYMTEEFVSVYR
1116 MHTLLPDAMILEGGKTVSLQELAF TDARNLVSDPSKTTATLLQAF AHTPAQALSLKNYPKS
1117 LFNLQIGNGK SINLAEIDISRDRARGIPRYNDARRQLL LTPYKSMDDLTS DKEELKLLKSVY
1118 TDIEQVDFLVGCLVDKDRPDGFAFGIVPYYIFVVMASRRLLSDRFFQEGLTEENYS

1119

1120 **Pseudo-nitzschia pungens MMETSP1061-20121228_74475**

1121 DNVQVIAQRLAREGFKPAGNQLNIIAAAWIQAQVHSWIQHLDGAPTSIEATAEAV
1122 GPVCPVKKMNFFETEERPDGEYNSFRTQWWDAS FVYGQNREQVHLGR LYKDGKLVNES

1123 NPDTLFVVEEGKSKIDVVGDQSNWSVGVTVLQVLFKEHNYCAEMIKKENPQLTDEEIYGH
1124 CRNIISALVAKIHTIDWTVELLKTEQLRVAMEINWKGATKAVFGDKAPFHPLRLINKPKAD
1125 NKGVPFCLTEEFAAVYRLHPLLPPGLVVEHGEKKEEFIEKLLTTKGRDKMREPGMAKKI
1126 MFSAFHYSCGHLKSSNYPFIMRKFTPTDHKGVDLQPPEDRVVDMAAIDLHRDRERGIQKY
1127 NEFRRQLKLRPFKTWEALTGEENKSDNELTDAKKLELIYGPAPEGIEKIDLLVGDLYERKIA
1128 DGFALSETSFMIPLLMASTRIDSDPYLNEYNEEYYT

1129

1130 **Thalassiosira rotula TR754|c0_g1_i1|m.3431**

1131 MLFPTFAQHLIDSFIVTAIKENTSEGVEFDWRKTGSPHDIGLLPLYGKTIEQTTQLRL
1132 KSELKGSKGLKTQIIKGGWSPYLFDAKGNKKPEFSSLPDPEGLEFALGMAKQHGVDVK
1133 KNKATVFALGGSANLTPNIVAWNTLLLREHNRIASKIEEENPKWDDERVFQTARNVNLA
1134 VYLRVIEEYINHITGFGVDFTVEPGKWMWDAPWYKRNWISAEFAVLRYRHAVIPNCMK
1135 WGEKTLPTASYLYSNHLLDDEDGMKGDLEAFINISDHRATSMQIHNTESWMKGRDARAL
1136 QMSRACELRSFTEYCAYLGAKEVPKTFADITSDVEVQNELKALYKVENIEFWVGLIAKDH
1137 PTESIMSAELTSFVANDAFNQALTHPLLSEHVWAAGPE

1138

1139 **Skeletonema marinoi MMETSP0920-20130426_33289**

1140 LLFPTFAQHLIDSFINTKINVETGEFEWDRDTSKHEIGIGPLYGDEVEQTSQLREKSEK
1141 PGRRGRLKTQVLEGDEEWAPFLYNEDGTKKEEFSAIHDPDGMKTILGLVYSSDPTTKSSIEQ
1142 SIFAFGGRRANLNPNIWA WNTLLLREHNRLAGEIEKSEPGWDDERVFQTARNVLIVMYCKI
1143 VIEEYIKHISGVNFKVDPGPWMWNAPWYKTNWMSTEFAILYRWHALIPNEAGLGPSKDAG
1144 VMESLFNPNVLLDDETGLGGNLRDIFVDISQTRITSFQLFNTEKWMVERESAAINQGRANN
1145 VQSYAAAYCEYLDIEPPKTFEDISMVPERQQALKELYGTPDRVEFYVGLIAADHPAGGKIFSE
1146 AMTKFVANDAFNQALTNPLLSQNVWE

1147

1148 **Fragilariopsis kerguelensis MMETSP0907-20130614_13332**

1149 PNVQVVAQRLLARESFTPAGDQLNIVAAGWIQAMVHDWMKHEDGKKTSMEVTPA
1150 VVGSQCPLHRFNLFETKERPDGHYNSERTNWWDASFVYQGNAEQVQNSRAFVGGKLVN
1151 EKNPDTLPSRDDGTDLTGDQSNWSVGVSVLQTLFLYEHNYCAEQIAKENPNLTDHQIYGH
1152 CRNIIAALVAKIHTIDWTVELLKTPQLKIGMRTNWMGIIQAITGLKIPFLDRLLRLIKKKENN
1153 NEGVPFCLTEEFAAVYRLHPLLPPGLIIEGEGAGDKDEDEDTFIDLRDTLTTKGRDLMRKSG
1154 MAKKVMKAVFTYPCGNMAPSNYPDVMRDFHPTDLLGNNLDDRIDLAAIDLFRDRERGIQ
1155 YFNNFRRKLSMKPFQTWEELTGDDKMTEEALAAFVAGTGNLNNAKKLELVYGAAPKGIE
1156 HCDLLVGDLYEKKIPGFAISETSFMIPLLMASTRILDADPYLNEYFDEEHYT

1157

1158 **Skeletonema grethea MMETSP0578-20130828_1168**

1159 ILVAYSIXSFQYVLDKPAASQLNVLAASWIQAMVHDWIGHFDGDETETLDRGGES
1160 LCPFAKSPFSFKNTKTEKIDGVFPSPERTNWWDASFVYGNNSAQIDRARTMQGGKMVTS
1161 DIPHALAEDKDG VYFAGDNKNSWVGVALQLDFIREHNYICDQIAAEAKEEGKEMTDEEL
1162 FGKARVVVAALVAKIHTVDWTVELLKTKLLAIGMKTNDWGLLKA VGIPIPGILSQMGEKK
1163 GRVSDNEGTPFCLTEEFAAVYRLHLSLPPGLILGDGDAKDKFIGLEDLLGDEGRKQMRETK
1164 TRPKEMMKSCLHWPCGALMSSNYPNAFRDVAPTDDYGKDLKSQNIDLAALDLFRDRERGI
1165 LKFNEFRRQLNLKPYRTWLELTENEEDARKLELIYGPQEGIERCDLLVGDMYERKVQPSF
1166 ALSETSFIIPLLMASTRILSADPFLNELYNEETYS

1167

1168 **Staurosira sp MMETSP1361-20130828_23824**

1169 LLFPFFAQWFVDSFLRTKWKPLAEQDFKENESNHIDIDLNQIYGTSEIQTDMRLSMK
1170 GGRLKSQIIDGEEYPVFLFDQKTATLKPEFRGLYTEENFKRVFGNASKEHKLHSFAVGLEH

1171 GNSTIGNTVMNTLFLREHNRVAGVISAAHPEWDDERVFQTTTRNVMIVLLIKIVLADYIYHIS
1172 GAAVFADPGGFAEDELWYRENWMSVEFSLLYRWHDLIPSSVTFDGETRDAVDLQNNNRW
1173 LLKAGMDSVIQDASNQKAGVMGLGNTPDFLLSVTKMSLHMARTCKLSSYVEYCKEYGGQD
1174 PPEDFMDLTGDQDSASKLERVYGSIDKVEWVGLFAQKRDFVSGLLMTLMVGNDAVTQA
1175 FTNPLLAKRVYN

1176

1177 **Skeletonema marinoi MMETSP1428-20130617_31509**

1178 EADVALVANMMKRDTNSFAPFNQLASAWIQFMTHDWFQHDASSSQGLKMQNVV
1179 THWWDASQMYGSSQEEVDAVRAEGGKLHLVDNDEIDYNASLPITGFRENWWAGLHILHT
1180 VFAREHNHIVDILAQSYPSMTEDELFGTARNIIAALLAKIHTVEWTPKNIPDEVKDIIDEMKV
1181 PSVFGSNYTTEQTLFNTPFYMTEEFVSVYRMHLLPDEMILEGGKLQELAFDARNLVSDP
1182 SKTTATLLQAFAPHTPAQALSLKNYPKSLFNLQIGNGKLNLAEIDISRDRARGIPRYNDARR
1183 QLLLTPYKSMDDLTSKKEELKLLKSVYADIEQVDFLVGCLVDKDRPDGFAFGIYYIFVMA
1184 SRRLLSDRFFQ

1185

1186 **Homo sapiens NP000953**

1187 LMFAFFAQHFTHQFFKTSKGKMGPGFTKALGHGVLDLGHYGDNLERQYQLRFLKDG
1188 KLKYQVLDGEMYPPSVVEEAPVLMHYPRGIPPQSQMAVGQEVFGLLPGLMLYATLWLREH
1189 NRVCDLLKAEHPTWGDEQLFQTTRLILIGETIKIVIEEYVQQLSGYFLQLKFDPELLFGVQFQ
1190 YRNRIAMEFNHLYHWHPLMPDSFKVGSQEYSYEQFLFNTSMLVDYGVVEALVDAFSRQIAG
1191 RIGGGRNMDHHILHVAVDVIRESRMRLQPFNEYRKRFGMKPYTSFQELVGEKEMAAELE
1192 ELYGDIDALEFYPLGALLEKCHPNSIFGESMIEIGAPFSLKGLLGNPICSPEYWK

1193

1194 **Ovis aries NP001009476**

1195 LMFAFFAQHFTHQFFKTSKGKMGPGFTKALGHGVLDLGHYGDNLERQYQLRFLKDG
1196 KLKYQMLNGEVYPPSVVEEAPVLMHYPRGIPPQSQMAVGQEVFGLLPGLMLYATIWLREHN
1197 RVCDLLKAEHPTWGDEQLFQTARLILIGETIKIVIEEYVQQLSGYFLQLKFDPELLFGAQQFQ
1198 YRNRIAMEFNQLYHWHPLMPDSFRVGPQDYSYEQFLFNTSMLVDYGVVEALVDAFSRQPA
1199 GRIGGGRNIDHHILHVAVDVIKESRVLRLQPFNEYRKRFGMKPYTSFQELTGEKEMAAELE
1200 ELYGDIDALEFYPLGALLEKCHPNSIFGESMIEMGAPFSLKGLLGNPICSPEYWK

1201

1202 **Homo sapiens NP000954**

1203 MMFAFFAQHFTHQFFKTDHKGPAFTNGLGHGVLDLNHIYGETLARQRKLRLFLKDG
1204 KMKYQIIDGEMYPPPTVKDTQAEMIPPQVPEHLRFAVGQEVFGLVPGLMMYATIWLREHN
1205 RVCDVLKQEHPEWGDEQLFQTSRLILIGETIKIVIEDYVQHLSGYHFKLKFDPELLFNKQFQ
1206 YQNRIAAEFNTLYHWHPLLPDTFQIHDQKYNYYQFIYNNSSILLEHGITQFVESFTRQIAGRV
1207 AGGRNVPPAVQKVSQASIDQSRQMKYQSFNEYRKRFMLKPYESFEELTGEKEMSAELEAL
1208 YGDIDAVELYPALLVEKPRPDAIFGETMVEVGAPFSLKGLMGNVICSPAYWK

1209

1210 **Coccotylus truncatus AFN20596**

1211 VLIAYYAQWVTHQFFNTDESPTGHSVKQPVGVNMSMLYGSKQEVVEKSVRAYKG
1212 GLLKSTIKNGQEFPEIMPCQEGSRIPGKEMFNMPILIANMIPGFAAIHVLFFRRHQYICRELA
1213 KWAEAQGNIDDEELFQAKLIVTVNMLRITMHDYVSRALQSSHAKMRFDQKVKQSRIW
1214 KMFGPDYFPPSNAIQFEFNIFYRWHQFYPTTKIMKRIDDLKFPKSKQQLDEKWNVAVRWIA
1215 DEPDGMRVLFSASSQRAGKLSLLNTNQWIVEHVVKPGLARCREHQLASYNDYREKVGFP
1216 RLTTFEQVTSNPALLEKLRVYRNVDQIEYYPGVFAEDKHFGNVHGPLTFGSSMTFTGIFSS
1217 RLFETALDE

1218 **Ciona intestinalis XP002127674**
1219 VLFPFFAQHFTHMFFKTDPMKGMPIYQWGDQLVDLSQIYGHGKQRQHELRSVNG
1220 KLKVSLLVDGHEFPPLSNQTTANMSNINLLPQEYQFVFGHQGFSLMPTFLIWSTIWLREHNRI
1221 CDLIKEENPAWDDERIFQTARLVLTGETIKVVIEDYVQHLSGFHYKLLYDPELVQGGSHSF
1222 HNQHVEFQLLYHWHALMPDQIEFNGKSYTMKRLLFNPEPVVKGGGLKRTIEDLSNQWAGQ
1223 VAGGKTQGAATLHVAGLAIKNGRDLRMQSFNAYKEKFEMKKYTTFQELTGEEEMAELQ
1224 KLYGDIDAVEYYIGIMLEKRRSPQLFGETLTEMGSPYSLKGLYSNPINHKDWWK
1225

1226 **Ciona intestinalis XP002123273**
1227 VFFAFFAQHFTHQFFKTNITIKGMPFQWGEHSVDLSHVYGHITQRQHELRSIDGKL
1228 KVFETNGEVFPPLTESANVTMSGEKLMRGRKFAIGHPGFGAFPSFFVIATLWLREHNRVCDI
1229 LKDLHPDWDDERLFQTARLILTGETLKIIVEDYVQHVS GFHFQLSYDPEILHKSTFSYNNQI
1230 HAEFHILYHWHMLMPDFIELGEHVYPLKELLFNVDPVVEIGMETVLKQLSNQFAGKVVG
1231 RNQGPPELVAVVELALKQTRQMRMCSFNKYRERFRGMKPYTSFEELTGETEVAALLRNLYY
1232 DIDALELFGYFVEHRRNRQVLGATMLEMGAPYSLKGVFGNPIGSPAWWK
1233

1234 **Oncorhynchus mykiss CAC10360.1**
1235 LMFAFFAQHFTHQFFKTRNSMGLGFTRALGHGVDAGNVYGDNLVRQLNLRLLKD
1236 GKMKYQVVKGEVYPPTVAEAAVNMRYPQETPVGQRMAIGQEVFGLLPGLTMYATLWLR
1237 EHNRVCDILKAEHPTWGDEQLFQTARLIVIGETIRIVIEEYVQHLSGYLLDLKFDPVLLFKST
1238 FQYRNRIAVEFKQLYHWHPLMPDSFHIDGDEVPYSQFIFNTSIVTHYGVEKLVDAFSRQCA
1239 GQIGGGRNIHPVVTNVAEGVIEESRTLRLQPFNEYRKRFLNPKPYTSFSDFTGEEEMARELEE
1240 LYGDIDALEFYPAIMLEKTRPNAIFGESMVEMGAPFSLKGLLGNPICSPEYWK
1241

1242 **Fistulifera solaris GAX23950.1**
1243 LPPVKDVAAILRRPMNPTSVAPFNQIAVAWIQMMTHDWFQHDPAHPDQKMNRVT
1244 HWWDASQLYGSTLAQQTAVRVPNTGKVRDLQHQELNYTTTGIPITGFADNWWAGLHMM
1245 HTLFRVREHNWLVDQFERQYPGVYTANDKFQLARLCLSALLAKIHTVEWTPTLNDNPVAAL
1246 GLHTNWRGVDAILEYGTREFELQAYRIVGGDQSVPHAGNGTTRETLYNTTFAMTEEFVAV
1247 YRMHPLLPDEMEIEGKTFSLNDLSFVDARTLTKSVKTTQTLLQAFGMTANTLSLQNYPRQ
1248 LYGLEKPGMSQPVNLAEIDLQRDRERNLPRYNDMMRQQLLKPYPKRELDLTDDETELNLLK
1249 SVYQDMDQVDLMVGCLVDKDRPYGFAFGIVPFHVFLVMASRRILNDRFFMEDFNAKVYT
1250

1251 **Cyanothece sp Heam peroxidase ACB53655.1**
1252 PNPRVVSRTLMTREDFKPATILNLLAAAWIQFENHDWFSHGDNKPEDKLEIPLEAN
1253 DPWPPEEYRPLEVGKTLPTDTSRPEGAKPPTFINTVTHWWDGSGIYGSNPETVDQLRSHEDGK
1254 LIIGENGLLPVDPETGVDITGFNDNWWIGLGMHLTLFTREHNLICDHLKQEYSQWSDDDL
1255 DHARLINAALMAKIHTVDWTPAILPLPATDIALNVNWNGLGEDIKQVLGTVGEGEADLL
1256 TGIIGSDKNHHTAPYYLTFEFSVYRMHPLIPDELEFRSLEGDKFLQEVNFFEMSGKRTRAL
1257 LESISLPDLFYSGITHPGAITLHNYPRFLQQLVRDNGEVFDLAAVDILRDRERGVPRYNRFR
1258 EIMGRGRVKCFEEISSNKQWVEEMRRVYNDNIDQVDLMVGLFAEDTPEGFGFSDTAFRVFI
1259 LMASRRLKSDRFFTTDYRAEIYT
1260

1261 **Fistulifera solaris GAX18717.1**
1262 LPPVKDVAAILRRPENPSTVAPFNQIAVAWIQMMTHDWFQHDPTNPDSKMNRVTH
1263 WWDASQLYGSTLTQQRVAVRVPNTGKLRDLKHQELNYTSTGIPITGFADNWWAGLHMMH
1264 TLFVREHNWLADQFELQYPGVYTANDKFQLTRLCLSALLAKIHTIEWTPTLNDNPVAALGL
1265 HTNWRGVDAILEYGTQFELQAYRIVGGDQSVPHAGNGTTRETMYNTTFAMTEEFVAVY

1266 RMHPLLDDMEIEGITLTLNDLSFVDARKLTKSVKTTQTLLQAFGTTTPANTLSLQNYPRQL
1267 YGLEKPGMSQPINLAEIDLQRDRERNLPRYNMRRQLLLKPYKRLEDLTDDTELNLLKSV
1268 YQDIDQVDLMVGCLVDKDRPYGFAFGIVPFHVFLVMASRRILNDRFFMEDFNAKVYT
1269

1270 **Alcanivorax nanhaiticus WP_035233415.1**

1271 PNPREVSNLIMSRGGDFKPATTLNFIATSWIQFMVHDWFDHGPRTDANPIEFPLPAG
1272 DVLGGGTMSVQRTRPDPDVSDEGHIYENINTHWWDGSQLYGSSKEKNDEVRSFVDGKL
1273 KVDGDGRLPTEFFSGKPVTGFNENWWVGLSMLHHIFTQEHNADMLVANYPGQSDQWY
1274 FDKARLINSALMAKIHTVEWTPAILANPVLERAMYANWWGLGGDRDKRDKYQDDLDNL
1275 NNNLGQIGLLDLVGIDNGLGDSPTGSLEHALAGLVGSRTPNNNYNVPYTLTEEFVSVYRM
1276 HPLLREDEIKVYDIGSNVDEEIALEDTRNGDAEDLLGDIGQDRLWYSFGITHPGALTLNNYP
1277 DFLRNLSMPLIGDIDMAAIDILRDRERGVPYRNEFRRQIGLKPLTSFEQLTSDPQLLADLKAL
1278 YNNDIELVDTLVGQLGEETRPEGFVGFGETSFQIFILNASRRLMTDRFFTTDYTDEVYT
1279

1280 **Hordeum vulgare subsp. vulgare BAJ90503.1**

1281 PDPFVVATKLLARREYKDTGKQFNILAAAWIQFMVHDWMDHMEDTKQIEITAPKE
1282 VANECPLKSFKFYATKEQPTNSDGIKTGYHNIRTAWWDGSAVYGNNEKQEKKIRTYADG
1283 KLVIGDDGLLLHEENGVPVLSGDVRNGWVGISILQALFVKEHNAVCDAIKEEHPNLSDEELY
1284 RYAKLVTSAVIAKIHTIDWTVELLKTKTLRAGMRANWYGLLGKKIKDTFGHIGGTALGGL
1285 VGLKKPINHGVPYSLTEEFTSVYRMHPLIPSTLKLDPDTPGQPAADNSPPYLEIDIDIGELVGLK
1286 GEDQLSKIGFEKQTLSMGYQACGALELWNYPSSFRDLIPQNLDTNRSDRIDLAALEVYRD
1287 RERSVPRYNEFRRLFLIPIKCWEDLTSNDIAIEAIRAIYGDDVEKLDLLVGLLAEKKIKGFA
1288 ISETAFNIFILMASRRLEADRFLTSNFNEKTYT
1289

1290 **Triticum aestivum CDM84254.1**

1291 PDPFVVATKLLARREYKDTGKQFNILAAAWIQFMVHDWMDHMEDTKQIEITAPKE
1292 VANECPLKSFKFYATKEQPTNSDGIKTGYNNVRTAWWDGSAVYGNNEKQAEKTRTYVDG
1293 KLVIGDDGLLLHEENGVPVLSGDVRNGWVGVSILQALFVKEHNAVCDAIKEDHPNLSDEEL
1294 YRYAKLVTSAVIAKIHTIDWTVELLKTKTMRAAMRANWYGLLGKKFKDTFGHIGGTALG
1295 GLVGLKKPINHGVPYSLTEEFTSVYRMHSLIPTTLKLDPDTPGQPAANNSPPYLEIDIDIGELV
1296 LKGEDQLSKIGFEKQTLSMGYQACGALELWNYPSSFRDLIPQNLDTNRSDRIDLAALEVY
1297 RDRERSVPRYNEFRRLFLIPIKSWEDLTSKDAIESIRAIYGDDVEKLDLLVGLMAEKKIK
1298 GFAISETAFNIFILMASRRLEADRFLTSNFNEKTYT
1299

1300 **Rhizophagus clarus GBC07128.1**

1301 LELEHYITALTQLPLNTPEEKLLSILERLLVTQLWSDISKPPAMIAGDIYRSSDGSY
1302 NRLIQSLGKANSRYSLSRIQYPIKLSVLPKSEDIFDKIMVQGGDFVEHPSGISSMLFYLAIIIT
1303 HDLFHTSFADPNINLTSSYLDLTPLYGSNDQEQKSIRTLKGGLLKPDTFADSRILLQPPGVSA
1304 LVILFSRNHNFAIQTLLKKNELGRFSVTNPNDPEQLKKQDEDLFQTARLINCIFYINVILHNY
1305 LRTILGLDRNSKWFVDPTVPYNKRGQLEPLPSGIGNIVSLEFNYYRWAATSKDDSKFVE
1306 DEFKTIFGDDWENITIDEFKEKMGVWGRSIPKDPSKWKFNHIERGSDNRFKDTDIAKEIING
1307 TKKVSGAFGANRIPKVFRPIELLGIESARVLGLSSLNDFRRSLNLKPYESFMEMNPDPLIAKK
1308 LEELYGSIENVELYPGLMTEKTKPDMLGSAIALPFTISRILSDAVNLVRNDRYYTNDFSR
1309 NLT
1310

1311 **Emiliana huxleyi XP_005780718.1**

1312 PFQALHNLPSADEVVRLLYKRDAFKKAPYGVNSLATWFANVAIHDFFRATATGTDG
1313 GTHPERGSDKEWVNLHSSYLDLQPLYGYSKTTADATREWSGGKLLKAFEDRMRRIPESRV

1314 IVELLRREHNYVAEQLAQRYPAQFATDEELYQQARLIMGGVYINILRAYGCQMFGEIAPD
1315 GSGFCELRRQGYGGAGVGNMCTFNFNLIYRFHTSIPVEWSATDPPPIDTDEQMRTLNGILN
1316 WESGGFGPNNVPDSILGERARVSQRAIEAARLMGAPTLNTRRRFTSGYSSFEDMTGGDQA
1317 TADTLRQLYPPGGIEDVELVVGVCQVEKCMSGGWALPSTIGQAIVADAFASIRQDRFYTQDW
1318 GASSYT

1319

1320 **Rhodococcus gordoniae WP_064063209.1**

1321 VWFAFFAQWFTDSFLRTNSKDPKNDNSNHEIDLCQIYGVNQAKTMLRAGYGGRL
1322 DSQVIDKKEYPPFLFAARTPGEELRFVPKFEGLHDREYLLDTVLRLCPDERKKS VFAVGLE
1323 HGNSTIGNTVLNVFLREHNRIAGILEGAYPEWDDDRLFETTRNIMIVILLKIVIEEYIKHIGP
1324 FDFPIEFVPMADNAPWNRTNWCAIEFNLLYRWHSLIPDTVVFDSQRVSTRILVDNNPLVL
1325 DRGIESIIDQCSRQKASRIGLGNTPAFLIDRHPMCPDRESVEERTVGLMRQARLRSFNDYRE
1326 AFGLGRLTSFTELTGDVEVQQKLARLYGDVDAVEWYVGIFAEDYPRHRMMGELLTTMVA
1327 HDAFTQAFTNPLLARHVYHEDTF

1328

1329 **Herbidospora mongoliensis WP_066371138.1**

1330 VMFMFFAQWFTDSFLRTSRDDFRNTSTQEIDLCQIYGLTEEKTHLLRAHRDGRK
1331 SQLIDGEEYPEFLFRPRSSGEPVFKPEFEGLHDSEFIVSRILLEDAPKQKDTFFAVGLEHGN
1332 STIGNTIMNVVFLREHNRIAGLLKQAHPEWAERPEGADARLFETTRNIMIVLLLKLVVEEYI
1333 RHISPYDFPLETVPFIADGKRWNRSNWISIEFNLLYRWHSLVPSTIGSGPDRLDSTDFRNNNP
1334 LVLARGIESLVSQCSGERAGRIGLMNTPGFLVDRDPAHPERPSVQERTIALSRKARLASYN
1335 YRENFGLGRLKDFGRLTQDAELRERLEKLYDDIDKLEWYVGIFAEDYSRDEMMGRLMTT
1336 MVAYDAFTQALTNPLLARDVYNEATF

1337

1338 **Eutrema halophilum BAJ34623.1**

1339 MVVATKLLTRRKMIDTGKQFNMAASWIQFMIHDWVDHLEDTDQIELSAPKEAAK
1340 GCPLSSFRFFKTKEVPTGFFEIKTGSLNTRTPWWDSSVIYGSNSKTLERVRTYKDGKLIKISEE
1341 TGLLLHDDDDGLAISGDIRNSWVGVSALQALFIKEHNAVCDLLKKEYEDEDLRYRHARL
1342 VTSAVIAKIHTIDWTVELLKTDTLLAGMRANWYGLLGKFKDFTFGHVGSSIFGGVGMKK
1343 PQNHGVPYSLTEEFTSVYRMHSLLPDQLHMRDIDVTPGPNKSLPLTQEVSMKLGREGEE
1344 TMSQIGFTKLMVSMGHQACGALELMNYPAWFRDLVPQDPNGHDRPDHIDLAALEIYRDR
1345 ERNVARYNDFRRAMFMIPIKTWEDLTDDKEAIELDDVYGGDVDELDDLVLGLMAEKKIKG
1346 FAISETAFNIFLLMATRRLEADRFFTSDFNEMTYT

1347

1348 **Pisum sativum CAH05011.1**

1349 PDPMVVVTKLLERKTYKDTGTQFNVAASWIQFMIHDWIDHMEDTKQVELSAPSEV
1350 ASQCPLKSFKFFKTKEIPTGFYDIKTGHANVRTPWWDGSSVYGSNEQVLNKVVRTFKDGKL
1351 KISKEGHLLHNEDGTAISGDIRNSWAGVTTLQTLFVQEHNAVCDALKKENS DLEDEDLYR
1352 HARLVTSAVIAKIHTIDWTVELLKTDTLLAGMRANWYGLLGKQFKDRFGHVGNSILSGFV
1353 GMKRSENHGVPYSLTEEFATVYRMHPLLPDSLHLRDISASPGPNKSPPLIKEIPMNDLIGLQ
1354 GEKTLLEIGNAKKLVS MGHQACGALELWNYPSWLRNLVPHNIDGTERS DHVDLALEVY
1355 RDRERNVARYNQFRRGLLLIPISKWEDLTDDEEAIK VLEEVYGGDDVEELDVLVGLMAEKKI
1356 KGFAISETAFVIFLLMASRRLEADRFFTSNFNEETYT

1357

1358

1359

1360

1361

1362 References

1363 1. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal
1364 for protein modeling, prediction and analysis. *Nature Protocols* **10**, 845–858 (2015).

1365

1366