

# Supporting Information

Peschke et al. 10.1073/pnas.1301945110

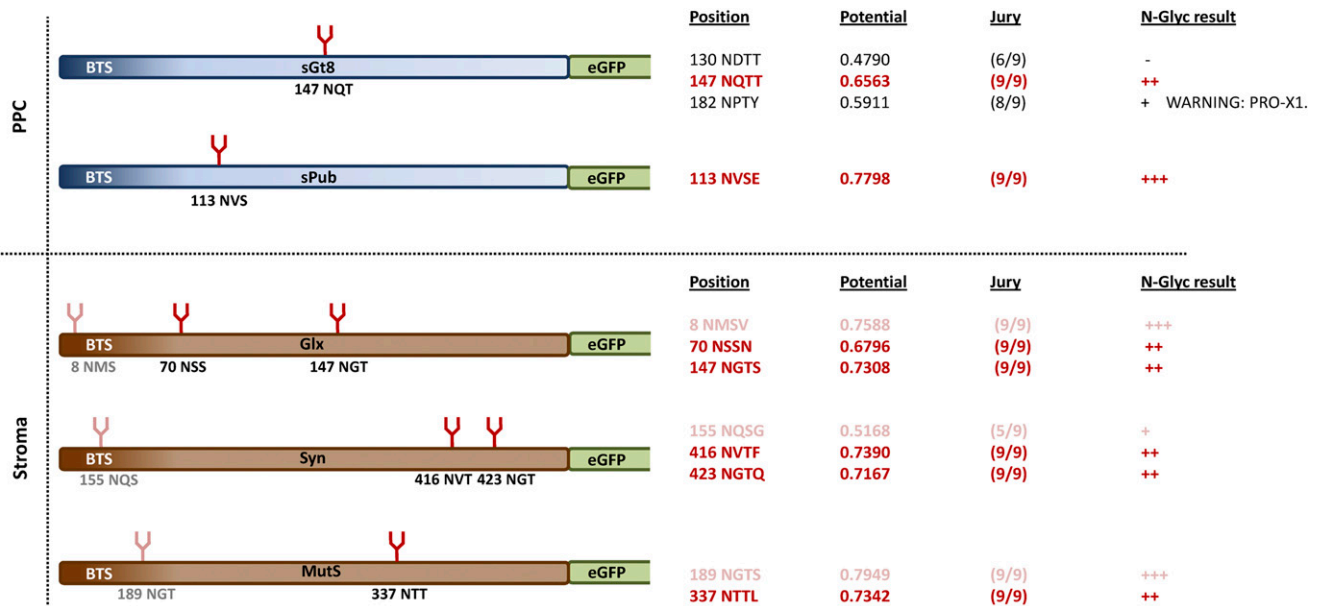
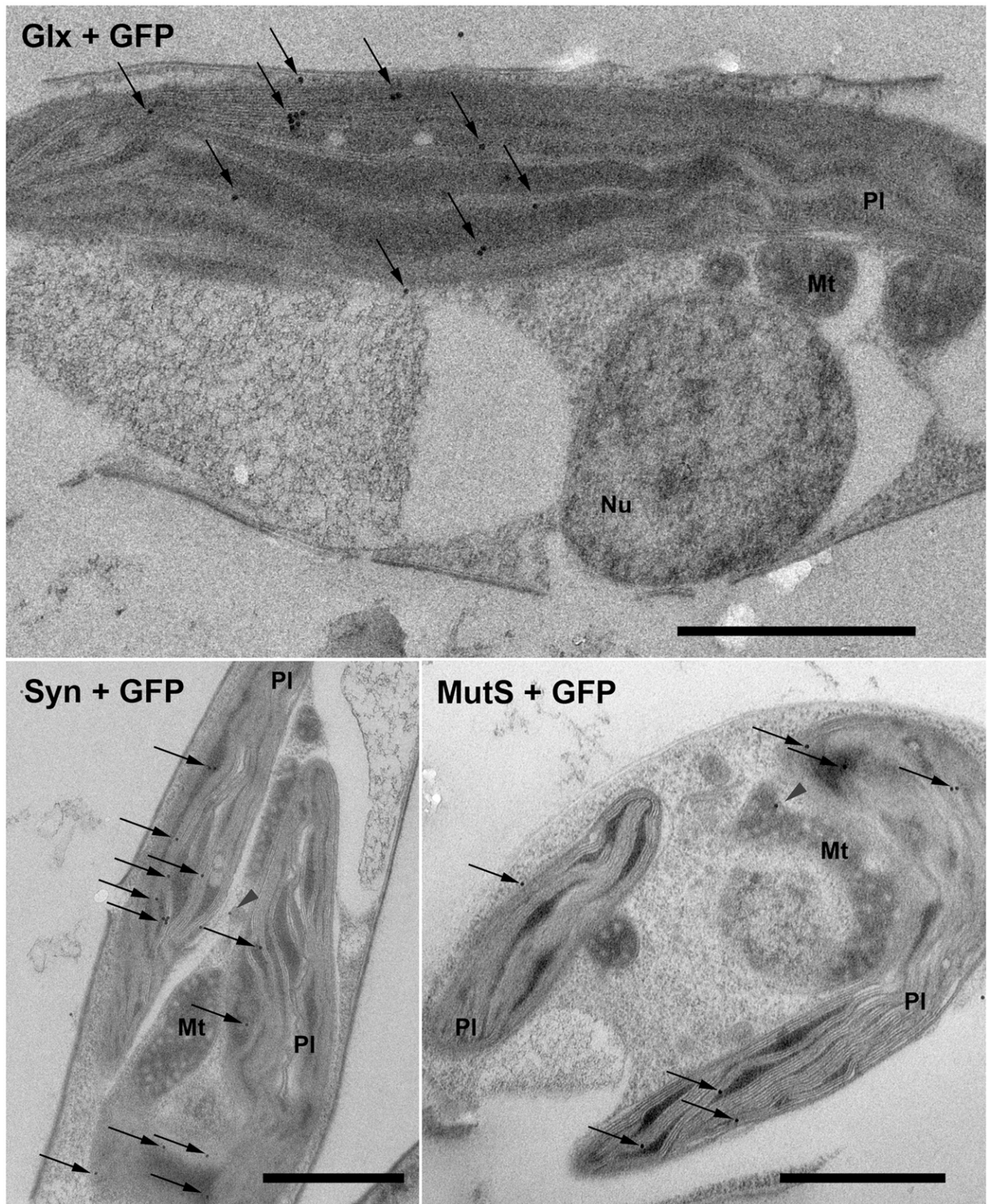
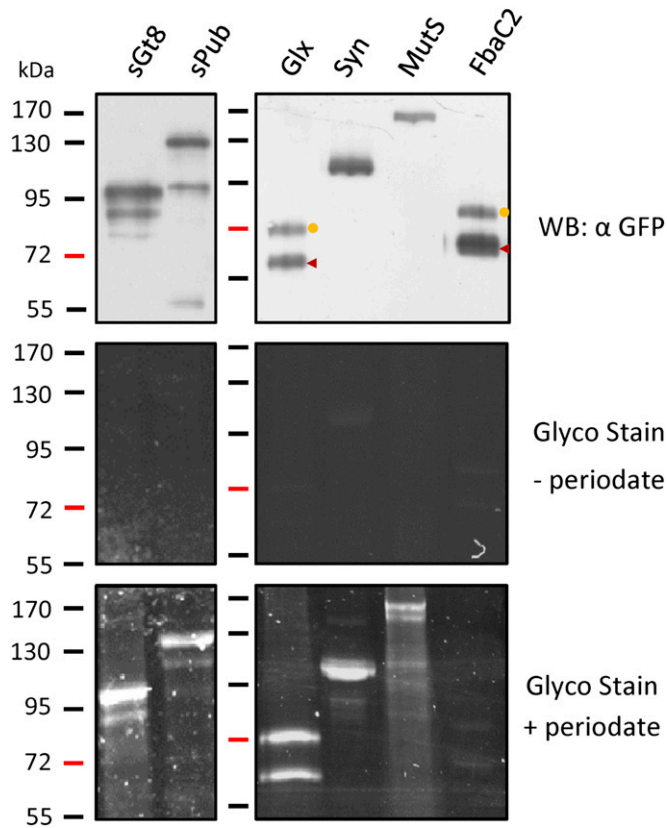


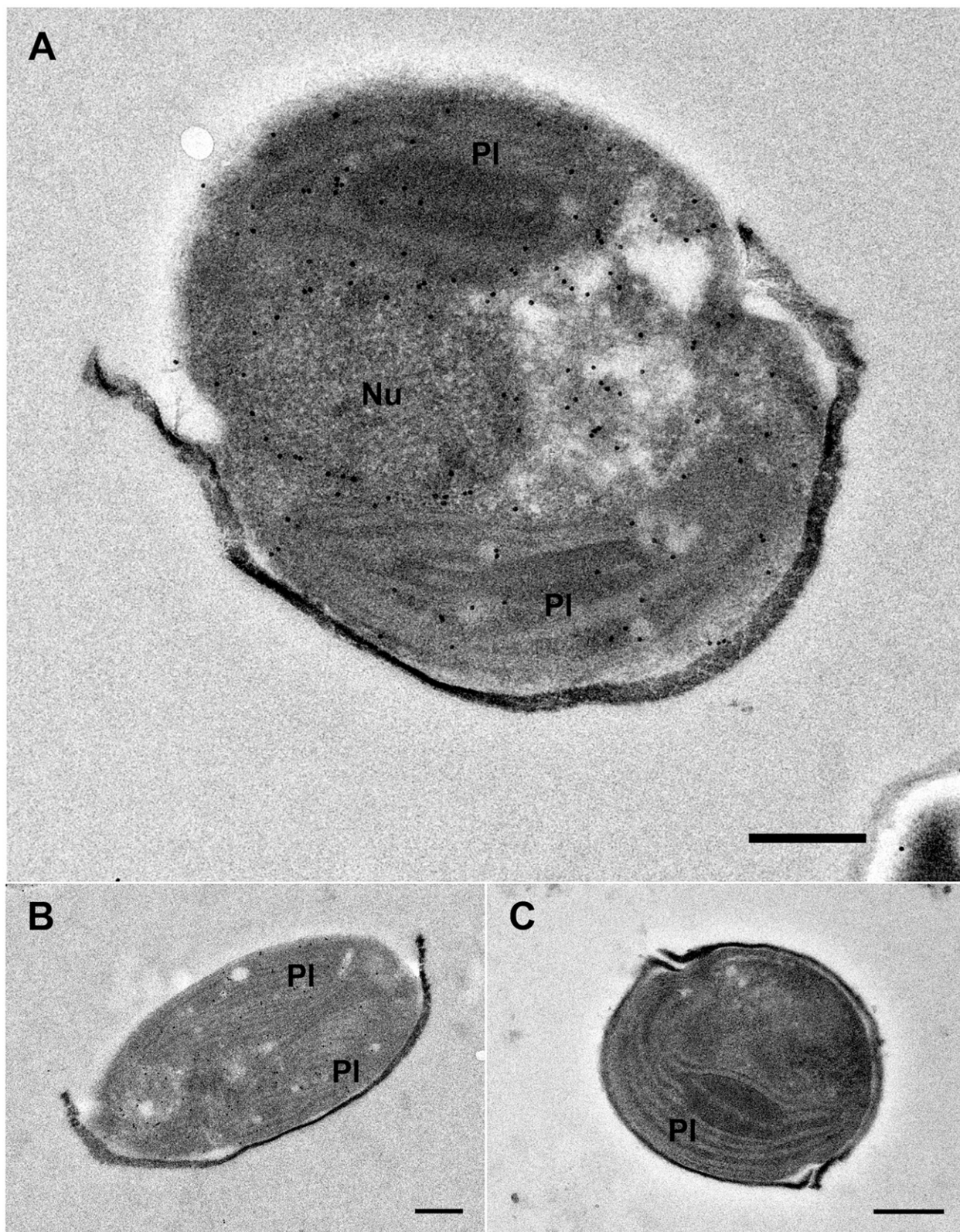
Fig. S1. NetNGlyc 1.0 prediction for endogenous PPC and stroma specific glycoproteins. Sequences of the periplastidal proteins sGt8 (symbiont-specific glycosyltransferase family 8), sPub (symbiont-specific PNGase/UBA or UBX), and the stromal proteins Glx (Glyoxalase), Syn (t-RNA synthetase), and MutS (Mutator S) were analysed with the program NetNGlyc 1.0 from the Center for Biological Sequence Analysis prediction server (<http://cbs.dtu.dk/services/NetNGlyc>).



**Fig. S2.** Electron microscopic analyses on the localization of plastidal glycoproteins. Immunoelectron microscopic analyses using an  $\alpha$ GFP antibody confirm the fluorescence microscopic data, demonstrating that the GFP fusion proteins of Glx, Syn, and MutS accumulate within the plastid stroma. Gold particles (10 nm) fused to the secondary antibody are highlighted with arrows. Arrowheads indicate two nonstromal presumably unspecific labels. (Scale bar, 1  $\mu$ m.) Mt, mitochondrion; Nu, nucleus; PI, plastid.



**Fig. S3.** Control experiments for Glyco-Staining of endogenous periplastidal and stromal GFP fusion proteins. Purified samples of the periplastidal proteins sGt8 and sPub as well as the stromal proteins Glx, Syn, MutS, and FbaC2 (fructose-1,6-bisphosphate aldolase C2) were separated by gel electrophoresis and stained with the Staining Pro-Q Emerald 300 Glycoprotein Gel kit (Invitrogen). As an additional control, gels were treated either with (+) or without (-) periodic acid before fluorescence labeling, demonstrating that staining is specific for periodate labile glycans. FbaC2 is not a glycoprotein and serves as a further control. As expected, no signal is detected in the Glyco-Stain. Once again different processing stadia are observed in the case of Glx and FbaC2. The completely processed variant is marked with an arrowhead; the version still carrying the transit peptide is marked with a circle.



**Fig. 54.** Electron microscopic analyses using concanavalinA (ConA) labeling on *Phaeodactylum tricornutum* wild-type cells. Electron microscopic analyses reveal that the lectin ConA, which is commonly used for detection of glycoproteins and binds to  $\alpha$ -D-glucose and  $\alpha$ -D-mannose residues, labels not only the secretory system throughout the cell but also the plastid stroma with high affinity (A and B), suggesting once again, that glycoproteins exist in the complex plastid of *P. tricornutum*. Importantly, and in contrast to many other phototrophs, the plastid of *P. tricornutum* does not contain any starch, which would have been recognized by ConA as well. ConA-biotin was used together with streptavidin gold (20 nm). No labeling is observed within the nucleus (Nu) of the cells. In the negative control without ConA-biotin, no binding of streptavidin-gold to endogenous biotin occurs (C). Cells were embedded in Lowicryl resin as glycan moieties were highly sensitive to Epon embedding procedures. (Scale bars, 500 nm.) Nu, nucleus; PI, plastid.

Table S1. N-glycosylation predictions for a set of periplastidal and stromal proteins of *P. tricornutum*

Protein ID no.	Protein function	Position	Potential	Jury agreement	N-Glyc result	Pro C	TMDs
PPC-specific proteins							
19162 <sup>†</sup>	AAA-ATPase, cell division cycle protein 48						–
37661 <sup>†</sup>	Thioredoxin and PUB domain-containing protein	113 NVSE	0.7798	(9/9)	+++		–
bd_206 <sup>†</sup>	N-terminal nucleophile (Ntn)-hydrolase superfamily	246 NQSS	0.5382	(5/9)	+		–
45475 <sup>†</sup>	N-terminal nucleophile (Ntn)-hydrolase superfamily	119 NSTT	0.5733	(7/9)	+		–
42884 <sup>†</sup>	N-terminal nucleophile (Ntn)-hydrolase superfamily						–
17388 <sup>†</sup>	N-terminal nucleophile (Ntn)-hydrolase superfamily						–
43079 <sup>†</sup>	N-terminal nucleophile (Ntn)-hydrolase superfamily	62 NISP	0.2209	(9/9)	–		+ (1, SP)
		254 NITN	0.6603	(9/9)	++		
		257 NQTP	0.1156	(9/9)	–		
144662970*	N-terminal nucleophile (Ntn)-hydrolase superfamily						–
38228 <sup>†</sup>	N-terminal nucleophile (Ntn)-hydrolase superfamily						–
35028 <sup>†</sup>	N-terminal nucleophile (Ntn)-hydrolase superfamily						–
12731 <sup>†</sup>	N-terminal nucleophile (Ntn)-hydrolase superfamily	179 NSTT	0.5006	(4/9)	+		–
		224 NLSG	0.4978	(5/9)	–		–
49432 <sup>†</sup>	N-terminal nucleophile (Ntn)-hydrolase superfamily						–
22110 <sup>†</sup>	N-terminal nucleophile (Ntn)-hydrolase superfamily	147 NISL	0.5866	(7/9)	+		+ (1, SP)
23414 <sup>†</sup>	Trypsin-like serine protease	104 NETL	0.7070	(9/9)	++		+ (1, SP)
		262 NSTA	0.6362	(9/9)	++		
		582 NSTS	0.6496	(9/9)	++		
		621 NSTK	0.4546	(7/9)	–		
35803 <sup>†</sup>	Trypsin-like serine protease	198 NLTY	0.6893	(9/9)	++		+ (1)
		206 NISM	0.4787	(4/9)	–		
40345 <sup>†</sup>	Trypsin-like serine protease	176 NLTV	0.7224	(9/9)	++		+ (1)
		354 NQSY	0.4397	(6/9)	–		
		404 NTTK	0.7138	(9/9)	++		
		444 NPSD	0.6081	(8/9)	+	PRO-X1	
40553 <sup>†</sup>	Trypsin-like serine protease	96 NPPI	0.7684	(9/9)	+++	PRO-X1	–
		502 NPTR	0.5653	(7/9)	+	PRO-X1	
49698 <sup>†</sup>	Trypsin-like serine protease	59 NSSI	0.7340	(9/9)	++		–
		75 NLTA	0.6507	(8/9)	+		
		520 NWTT	0.4317	(7/9)	–		
		560 NDTN	0.5058	(4/9)	+		
34512 <sup>†</sup>	DnaJ and TPR domain-containing protein						–
48633 <sup>†</sup>	DnaJ and PDI domain-containing protein	369 NASS	0.5297	(5/9)	+		–
		415 NLSK	0.7417	(9/9)	++		
50480 <sup>†</sup>	Heat shock protein 70, prokaryotic type	287 NKTV	0.7339	(9/9)	++		–
		437 NETP	0.1553	(9/9)	–		
40314 <sup>†</sup>	Glycosyl transferase family 8	130 NDTT	0.4790	(6/9)	–		+ (1, SP)
		147 NQTT	0.6563	(9/9)	++		
		182 NPTY	0.5911	(8/9)	+	PRO-X1	
37379 <sup>†</sup>	Dynamain-related GTPase						+ (1, SP)
35370 <sup>†</sup>	Alpha carbonic anhydrase	33 NDTS	0.6102	(7/9)	+		
		177 NPTR	0.6033	(8/9)	+	PRO-X1	
		358 NMTS	0.4176	(6/9)	–		
42574 <sup>†</sup>	Alpha carbonic anhydrase						–
31704 <sup>†</sup>	Similar to <i>Guillardia theta</i> nucleomorph encoded ORF139						–

Table S1. Cont.

Protein ID no.	Protein function	Position	Potential	Jury agreement	N-Glyc result	Pro C	TMDs
47811 <sup>†</sup>	Similar to <i>G. theta</i> nucleomorph encoded ORF261						–
45935 <sup>†</sup>	Similar to <i>G. theta</i> nucleomorph encoded ORF532a	157 NNSS 158 NSSN	0.4326 0.6942	(5/9) (9/9)	– ++		–
48879 <sup>†</sup>	Similar to <i>G. theta</i> nucleomorph encoded ORF534	5 NVSK 52 NDTI 316 NLSV	0.8146 0.5178 0.5017	(9/9) (7/9) (5/9)	+++ (SP) + +		+ (1, SP)
44766 <sup>†</sup>	Sec14p-like protein, PtdIns/PtdCho transfer protein	37 NASV 105 NHTL 138 NTTT	0.5751 0.6761 0.4679	(5/9) (9/9) (3/9)	+ ++ –		–
41316 <sup>†</sup>	Pentapeptide repeats containing protein						–
47766 <sup>†</sup>	Prolyl 4-hydroxylase and TPR domain-containing protein	425 NFSL	0.5532	(7/9)	+		–
37424 <sup>†</sup>	Pectin esterase-like domain-containing protein, PT-repeats	258 NPSD 390 NPST 697 NPTS 781 NPTS 865 NPTS 943 NVSP 1022 NPTQ	0.5651 0.4077 0.5849 0.5716 0.5699 0.1580 0.5019	(6/9) (8/9) (8/9) (8/9) (7/9) (9/9) (5/9)	+ – + + + – +	PRO-X1 PRO-X1 PRO-X1 PRO-X1	+ (1, SP)
42675 <sup>†</sup>	SMC (structural maintenance of chromosomes)-domain-containing protein	455 NTTH 546 NETS	0.6380 0.6053	(9/9) (9/9)	++ ++		+ (1)
54460 <sup>†</sup>	Ubiquitin-activating enzyme E1	140 NPSV 211 NDTD 530 NLSR 1069 NRST 1092 NGSE	0.8012 0.6112 0.5444 0.5200 0.5247	(9/9) (8/9) (7/9) (6/9) (6/9)	+++ + + + +	PRO-X1	–
34112 <sup>†</sup>	Ubiquitin-conjugating enzyme E2	151 NLSV	0.5213	(6/9)	+		–
56431 <sup>†</sup>	Ubiquitin-conjugating enzyme E2						+ (1)
54323 <sup>†</sup>	Ubiquitin						–
31697 <sup>†</sup>	Derlin, degradation at the ER, rhomboid family member						+ (4)
35965 <sup>†</sup>	Derlin, degradation at the ER, rhomboid family member	237 NLST	0.6908	(9/9)	++		+ (3)
49319 <sup>†</sup>	Ubiquitin fusion degradation						–
20914 <sup>†</sup>	AAA-ATPase, cell division cycle protein 48	921 NGSL	0.5968	(7/9)	+		–
55890 <sup>†</sup>	Heat shock protein 70	148 NPTG	0.7009	(9/9)	++	PRO-X1	–
48035 <sup>†</sup>	<i>P. tricornutum</i> E3 enzyme of the PPC	80 NLTV	0.7664	(9/9)	+++		+ (1,SP) <sup>†</sup>
56710 <sup>†</sup>	<i>P. tricornutum</i> deubiquitinating enzyme of the PPC	87 NDSG 283 NLSL	0.4853 0.7048	(5/9) (9/9)	– ++		–
45333 <sup>†</sup>	6-Phosphogluconate dehydrogenase						–
48539 <sup>†</sup>	Thioredoxin h						–
37815 <sup>†</sup>	NADPH dependent thioredoxin reductase containing N-terminal thioredoxin domain	386 NATG 489 NATV	0.6549 0.7099	(9/9) (9/9)	++ ++		–
34750 <sup>†</sup>	Nuclear protein localization 4	27 NKSI	0.6287	(8/9)	+		–
50113 <sup>†</sup>	Peptide:N-glycanase	88 NRSS 202 NDTT 211 NSTI 238 NITK 298 NATA 444 NASS	0.7604 0.5764 0.5376 0.7312 0.6448 0.4303	(9/9) (7/9) (7/9) (9/9) (9/9) (8/9)	+++ + + ++ ++ –		+ (1, SP)

Table S1. Cont.

Protein ID no.	Protein function	Position	Potential	Jury agreement	N-Glyc result	Pro C	TMDs
47962 <sup>†</sup>	Ubiquilin						–
44122 <sup>†</sup>	UBX domain-containing protein	208 NSSR	0.5558	(7/9)	+		+ (1, SP)
52 periplastidal sequences, 23% with at least one high confidence prediction N-Glyc site (+/+/+) within the mature, soluble protein							
Stroma-specific proteins							
20349 <sup>†</sup>	Glutathione reductase	165 NITF	0.7229	(9/9)	++		–
219112363*	Chorismate synthase	308 NRSNG 430 NGTL	0.3626 0.5608	(8/9) (5/9)	– +		+ (1, SP)
219127660*	Ribosome recycling factor						–
50082 <sup>†</sup>	Member of the clp superfamily, regulatory gamma subunit	4 NVSR 846 NGSP	0.6381 0.1417	(9/9) (9/9)	++ –		+ (1)
219110927*	FtsH, ATP-dependent Zn protease	34 NNSA 603 NNSY	0.5137 0.3920	(6/9) (9/9)	+ –		–
219121191*	Cytochrome b6-f complex iron-sulfur subunit, chloroplast precursor						+ (1, SP)
219119472*	Nitrite reductase-ferredoxin dependent	213 NQTS	0.7256	(9/9)	++		–
219117802*	Protochlorophyllide reductase A						–
219113211*	Member of the clp superfamily, regulatory beta subunit	842 NGSS	0.5888	(7/9)	+		–
219123872*	Plastid division protein FtsZ						–
219112433*	Mutase phosphoglucomutase	17 NAST 314 NLTY 334 NSSM 476 NDDT	0.6146 0.6566 0.4434 0.4537	(7/9) (9/9) (7/9) (5/9)	+ ++ – –		–
219120381*	3-Isopropylmalate dehydrogenase						–
219117493*	Ferrochelatase						+ (1)
219112651*	Alb3 homolog, thylakoidal inner membrane insertase	411 NSSK	0.5958	(7/9)	+		+ (3)
18019 <sup>†</sup>	Nuclear-encoded-like protein of chloroplast gro						–
219115519*	Farnesyltransferase						–
19363 <sup>†</sup>	PEFG, plastid translation elongation factor EF-G	733 NYSM	0.5270	(4/9)	+		+ (1, SP)
219109864*	Zeta-carotene desaturase	58 NQSF 320 NLSS	0.3857 0.6971	(7/9) (9/9)	– ++		–
219120510*	Light-dependent NADPH: protochlorophyllide oxidoreductase						–
219123807*	Glutamine synthetase	232 NISG 299 NVST	0.5836 0.4689	(6/9) (5/9)	+ –		–
219125083*	Pyruvate kinase 1	111 NFSH	0.5447	(5/9)	+		–
219109858*	cbbX protein	52 NSSS 397 NYSA	0.5598 0.4198	(7/9) (7/9)	+ –		–
219129854*	Ketol-acid reductoisomerase	227 NSSF 437 NESV 462 NCSF	0.3248 0.5472 0.5032	(9/9) (4/9) (4/9)	– + +		–
219118809*	Phytoene dehydrogenase	525 NFTL	0.6420	(8/9)	+		+ (1, SP)
219128124*	Nucleotide transporter 1	148 NYTI 254 NWSF 434 NFSS	0.7328 0.4138 0.5911	(9/9) (8/9) (6/9)	++ – +		+ (11)
219114431*	Precursor of synthase DXS 1-deoxy-d-xylulose-5-phosphate synthase	584 NQSS	0.6208	(6/9)	+		–
219121065*	Serine type protease, similar to Protease Do-like 1, chloroplast precursor	159 NLSL	0.6713	(9/9)	++		–
219122837*	Ascorbate peroxidase	157 NCSA	0.5750	(7/9)	+		–
219112735*	Mg chelatase, subunitD						–

**Table S1. Cont.**

Protein ID no.	Protein function	Position	Potential	Jury agreement	N-Glyc result	Pro C	TMDs
219110367*	Precursor of protein zeaxanthin epoxidase-like protein						–
219111891*	ALA dehydratase						–
219118280*	Kinase adenylate kinase						–
219125316*	Plastid division protein FtsZ	364 NLSL	0.6302	(8/9)	+		+ (1, SP)
219129317*	Glycerol-3-phosphate acyltransferase	126 NPSS 347 NKSK 428 NPTT	0.6658 0.6246 0.4480	(9/9) (7/9) (8/9)	++ + –	PRO-X1	+ (1, SP)
219111175*	Chloroplast-targeted nuclear-encoded recombinase	60 NDSN	0.6228	(6/9)	+		+ (1, SP)
219112561*	3-Phosphoshikimate 1-carboxyvinyltransferase	235 NISI	0.5610	(5/9)	+		–
219118943*	Zeaxanthin epoxidase						+ (2)
219110034*	Glutamate synthase	49 NPTS 922 NSSD	0.7803 0.4448	(9/9) (6/9)	+++ –	PRO-X1	–
219120458*	Glutamate 1-semialdehyde 2,1-aminomutase						–
27883911*	Delta 12 fatty acid desaturase	190 NRSL	0.5943	(6/9)	+		+ (4)
219111077*	Fructose-1,6-bisphosphatase						–
2498374*	Fucoxanthin-chlorophyll a-c binding protein E, chloroplastic						+ (1)
219125332*	Protein fucoxanthin chlorophyll a/c protein						–
219116130*	Protein fucoxanthin chlorophyll a/c protein						–
219117950*	Protein fucoxanthin chlorophyll a/c protein						–
219112111*	Phytanoyl-coa dioxygenase						–
2498375*	Fucoxanthin-chlorophyll a-c binding protein F, chloroplastic						–
219119491*	Precursor of ATPase ATPase gamma subunit	198 NPTA	0.5101	(4/9)	+	PRO-X1	–

48 stromal sequences, 10.4% with at least one high confidence prediction (++++) N-Glyc site within the mature, soluble protein

All known periplastidal proteins and a set of verified stromal proteins homolog to the training set of the heterokont targeting prediction program HECTAR (1) were analyzed for putative N-glycosylation sites with the program NetNGlyc 1.0. Proteins with at least one high confidence prediction (++++) for N-glycosylation within the mature, soluble protein were used for calculations (highlighted in red), suggesting that about 23% of periplastidal and 10.4% of stromal proteins might get N-glycosylated during passage of the chloroplast endoplasmic reticulum (cER). Membrane proteins and N-Glyc predictions within the signal peptide (SP) were excluded as the focus lies on the role of glycans on transport. The signal peptide is already cleaved off within the cER and in the case of membrane proteins a complete translocation does not necessarily occur. Column headings: "Protein function" indicates detected functional domains or annotation; "Position" indicates the number/position of Asn (N) in the protein sequence and letters indicate sequence context; "Potential" indicates N-glycosylation potential, averaged output of nine neural networks, default threshold of 0.5 (NetNGlyc 1.0); "Jury agreement" indicates how many of the nine networks support the prediction (NetNGlyc1.0); "N-Glyc result" indicates N-glycosylation prediction confidence (NetNGlyc 1.0); – indicates a negative result, + indicates low confidence prediction and ++++ indicates high confidence prediction "Pro C" indicates proline context; Pro-X1, proline is directly following Asn, N-glycosylation unlikely (NetNGlyc 1.0); "TMDs" indicates prediction of transmembrane domains by TMHMM v2.0 where + or – indicate a positive or negative result, the number in parentheses indicates the quantity of predicted TMDs, and SP (signal peptide) indicates that the SP likely was recognized as potential TMD.

\*National Center for Biotechnology Information.

<sup>†</sup>Phaeodactylum tricornutum data base version 2.0 (Phatr2DB).

<sup>‡</sup>TMD prediction with SOSUI +(1), membrane integration experimentally verified (2).

1. Gschloessl B, Guermeur Y, Cock JM (2008) HECTAR: A method to predict subcellular targeting in heterokonts. *BMC Bioinformatics* 9:393.

2. Hempel F, Felsner G, Maier UG (2010) New mechanistic insights into pre-protein transport across the second outermost plastid membrane of diatoms. *Mol Microbiol* 76(3):793–801.