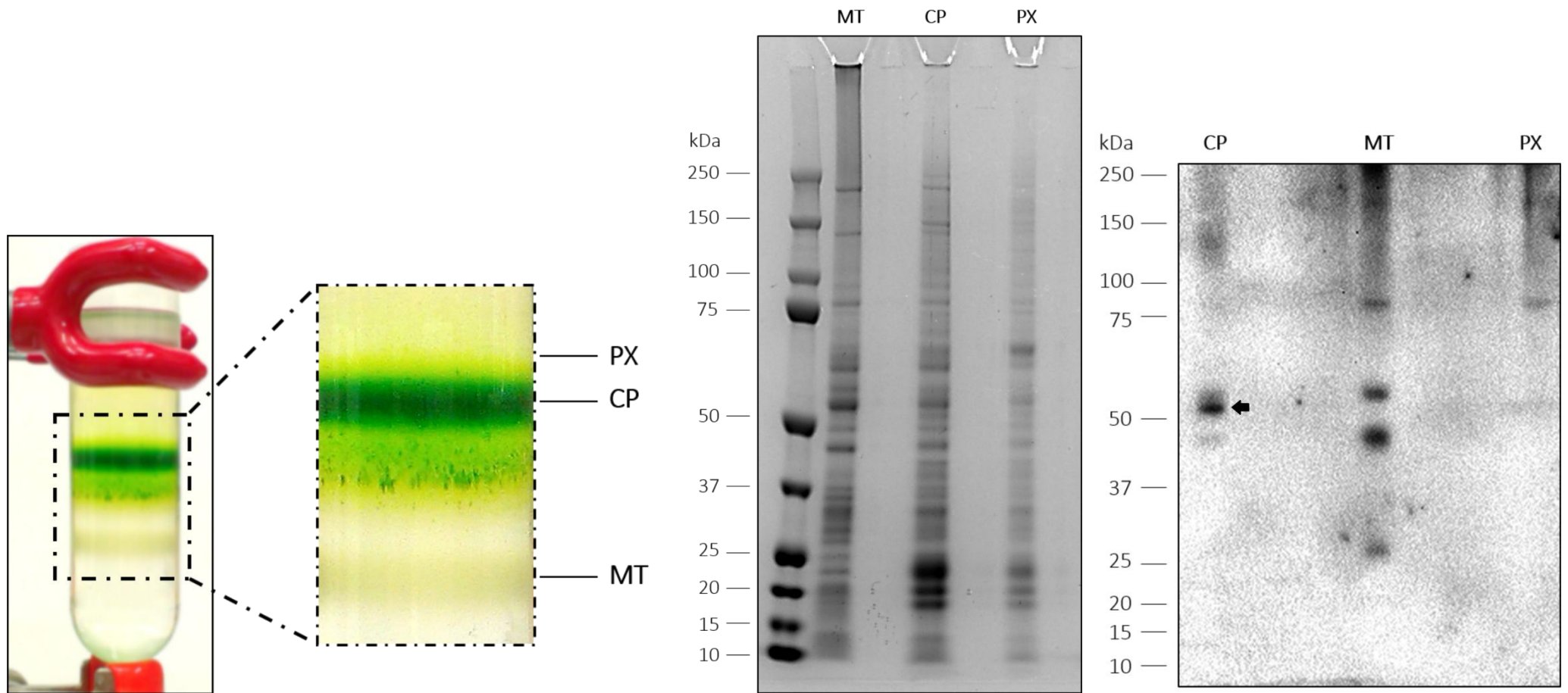


# Supplementary data



S1.1: The result of cell fractionation on a sucrose gradient: the plastidial fraction is marked as CP, while the mitochondrial and peroxisomal fractions are marked MT and PX, respectively.

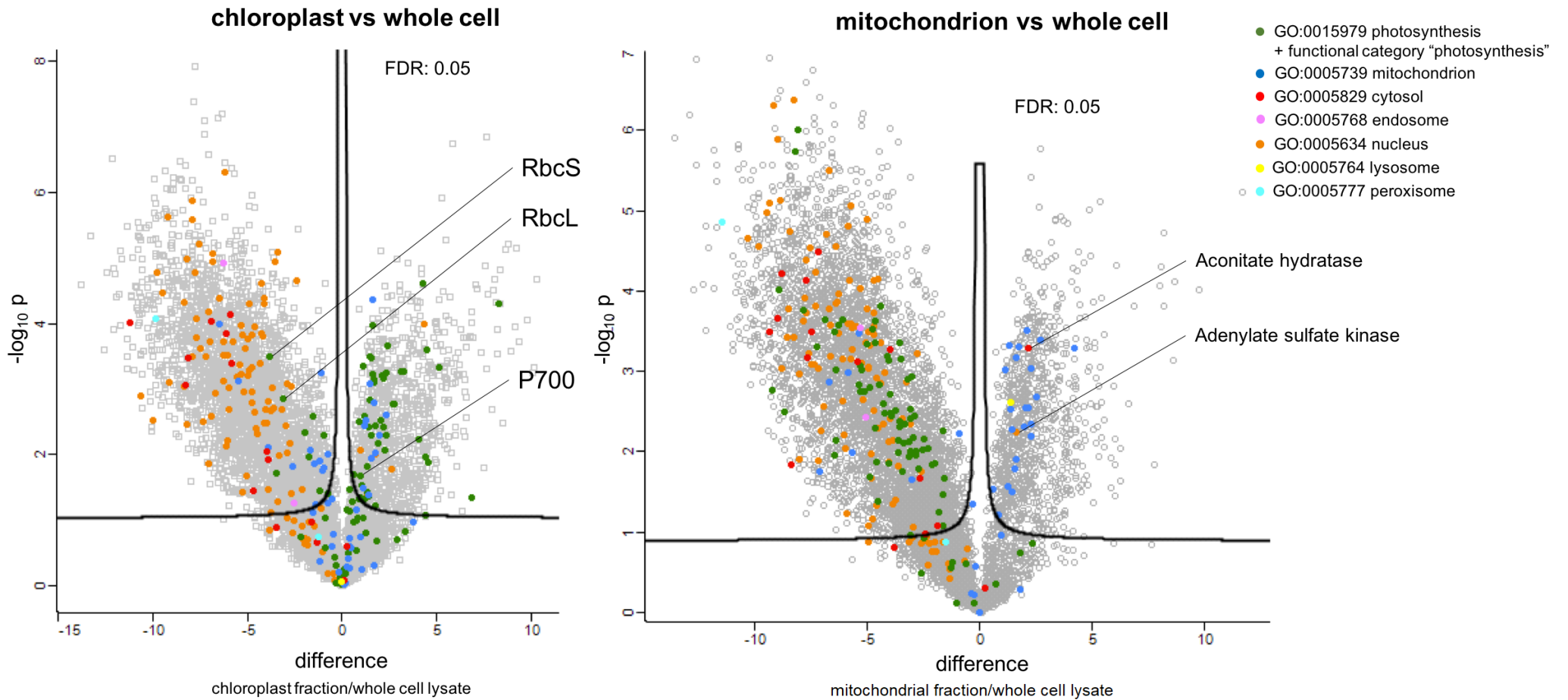
S1.2: Coomassie-stained gel (Nupage Bis-Tris Mini Gels 4-12%, IM-8042, Life Technologies) of the cellular fractions with Precision Plus Protein™ Dual Color Standards (#1610374, Bio-Rad).

S1.3: Immunoblot of the cellular fractions with rabbit anti-RbcL (1:5000, AS03 037, Agrisera, protein size: 52 kDa) as primary antibody and goat anti-Rabbit (1:2000, A6154-5x, Sigma) as secondary antibody, and Protein Ladder (161-0374, Biorad), the positive band of expected size is marked by arrow.

	<b>Cytosol</b>	<b>Acidocalcisome</b>	<b>Endosome</b>	<b>Peroxisome</b>	<b>Lysosome</b>	<b>Golgi</b>	<b>Nucleus</b>	<b>Surface</b>
Protein ID <u>EG_transcript_</u>	<b>21524</b>	<b>2633</b>	<b>181</b>	<b>15991</b>	<b>10514</b>	<b>5712</b>	<b>53416</b>	<b>32527</b>
Annotation	Aldolase	VP1 Vacuolar proton translocating pyrophosphatase	CHC Clathrin heavy chain	PEX2 Peroxisomal biogenesis factor 2	Lysosomal aspartic protease	Coatomer subunit $\gamma$ 2	Histone H4	Articulon 80 kDa
<b>Ratio CP/W</b>	<b>0.01</b>	<b>1.95</b>	<b>0.21</b>	<b>W only</b>	<b>3.30</b>	<b>0.01</b>	<b>0.04</b>	<b>0.01</b>
-log10p	5.9	1.2	3.1	1.4	2.0	3.4	3.3	4.7
Ratio Mt/W	0.02	1.46	0.53	W only	7.4	0.01	0.02	0.04
-log10p	3.3	0.5	1.9	2.2	2.6	3.3	3.6	4.4
Ratio CP/Mt	0.26	1.33	0.40	NaN	0.44	0.74	2.0	0.15
-log10p	1.5	0.6	2.8	NaN	1.5	0.3	0.8	2.9
Unique peptides	19	31	90	1	7	40	20	13

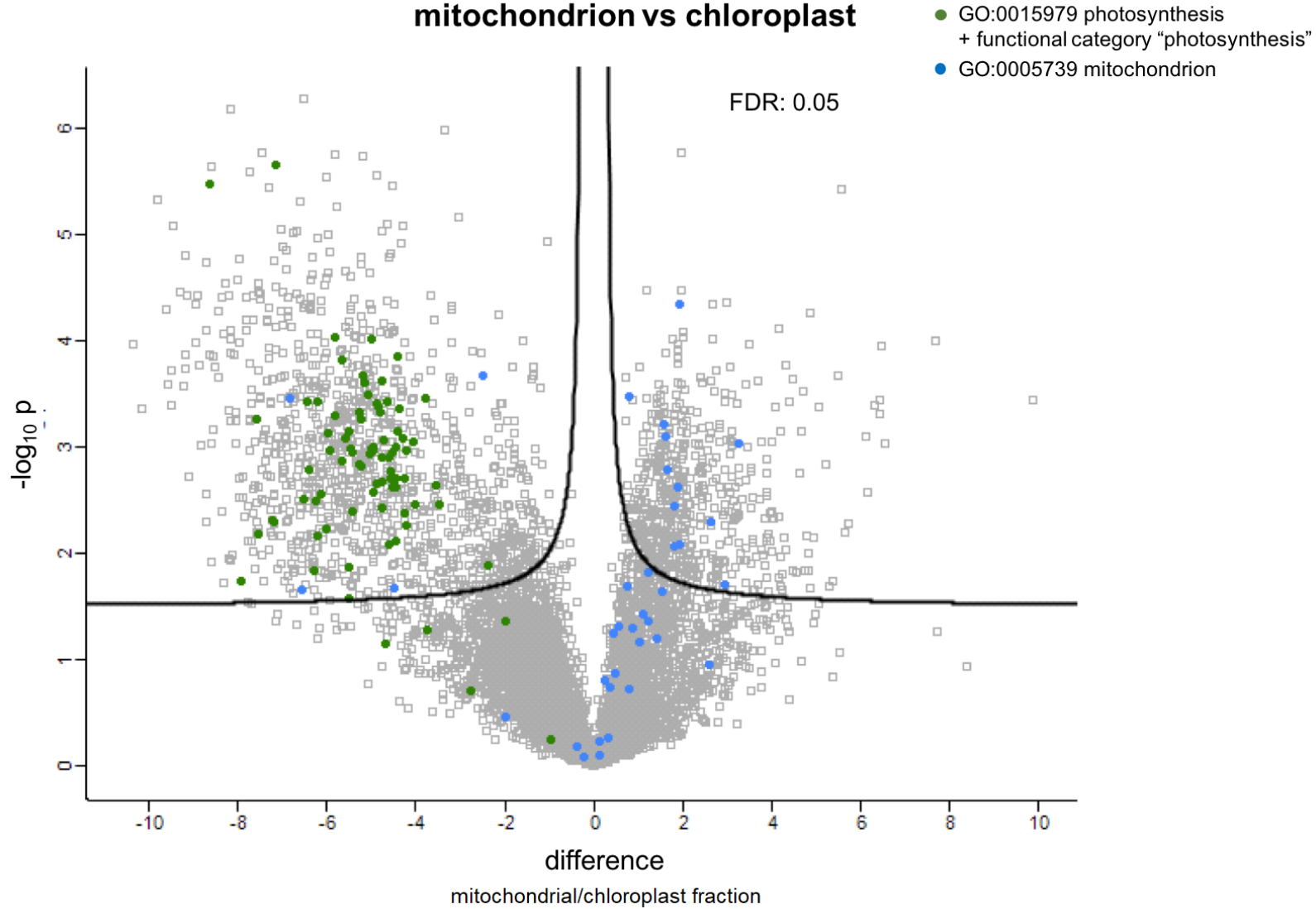
	<b>Chloroplast</b>			<b>Mitochondrion</b>		
Protein ID <u>EG_transcript_</u>	<b>40006</b>	<b>158</b>	<b>25897</b>	<b>2112</b>	<b>23844</b>	<b>8912</b>
Annotation	light-harvesting complex I protein precursor LhcB5	Photosystem I P700 chlorophyll a apoprotein A1	light-harvesting complex I protein precursor Lhca2	Pyruvate dehydrogenase [NADP(+)], mitochondrial	ubiquinol-cytochrome c reductase iron-sulfur subunit	F-type H <sup>+</sup> -transporting ATPase subunit beta
<b>Ratio CP/W</b>	<b>4.5</b>	<b>1.95</b>	<b>CP only</b>	<b>1.31</b>	<b>3.10</b>	<b>4.6</b>
-log10p	3.2	1.7	1.9	0.6	4.4	3.8
Ratio Mt/W	0.2	0.10	NaN	4.7	4.27	4.8
-log10p	1.8	2.3	NaN	2.6	3.5	3.4
Ratio CP/Mt	24.2	20.5	CP only	0.28	0.73	0.95
-log10p	2.7	2.8	2.2	2.1	1.3	0.1
Unique peptides	4	30	2	76	35	32

S2.1: Assessment of the purity of isolated fractions using selected marker proteins and their relative abundance compared against the whole cell lysate and between the organellar fractions.



S2.2: Volcano plots from p-values versus the corresponding t-test difference of 8216 protein groups quantified in the two organellar fractions and whole cell lysate. Green and blue dots represent proteins assigned to "photosynthetic" and "mitochondrial" GO categories, respectively. The remaining colours represent other selected GO categories (indicated at the top right) associated with other cellular compartments. Stringent cutoff curves for statistically significant enrichment (black curves) were calculated from the estimated false discovery rate (FDR).

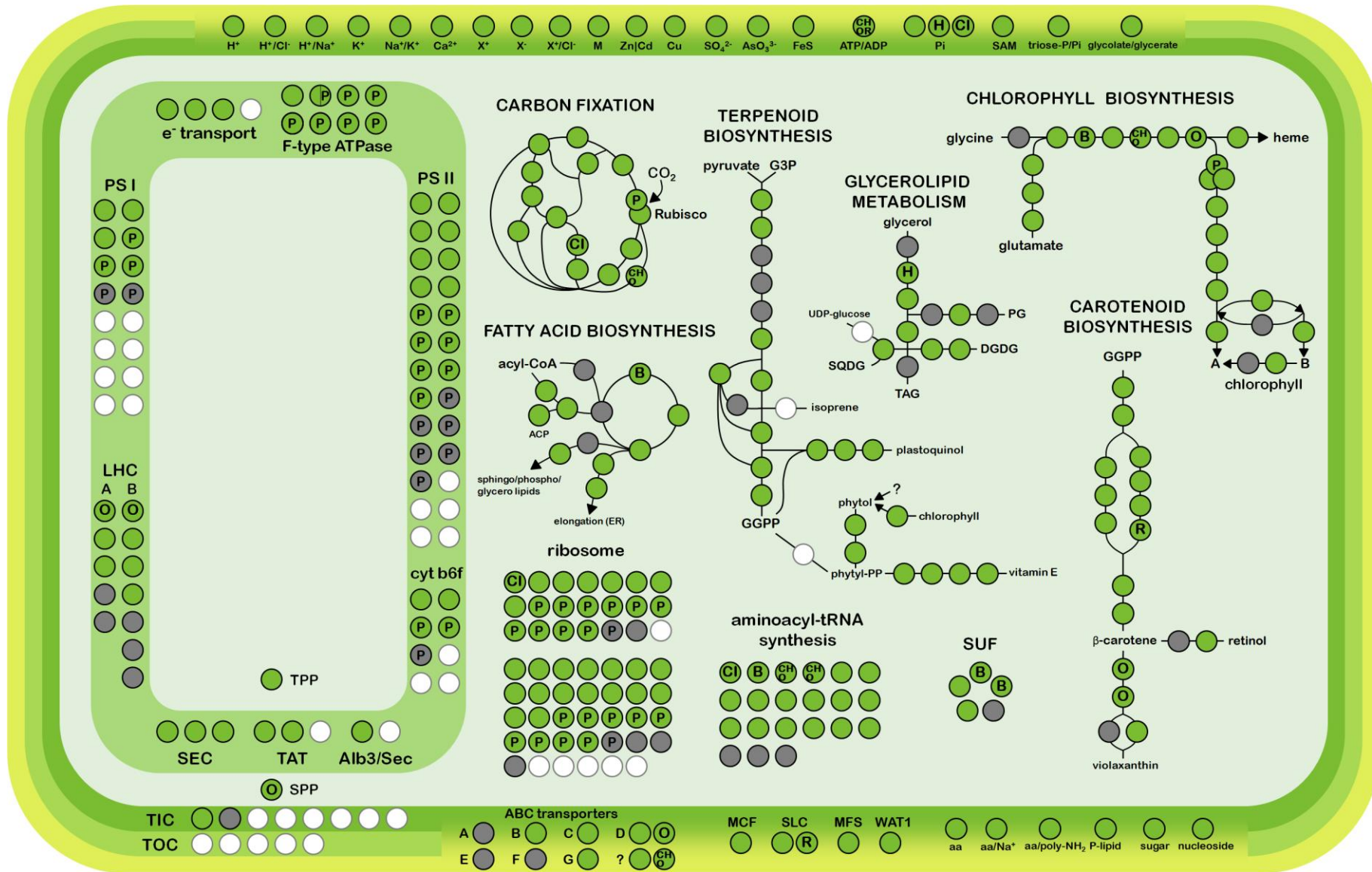
### mitochondrion vs chloroplast



S2.3: Volcano plot from p-values versus the corresponding t-test difference of 3,736 protein groups quantified in the mitochondrial and chloroplast fraction. Green and blue dots represent proteins assigned to "photosynthetic" and "mitochondrial" GO categories, respectively. Stringent cutoff curves for statistically significant enrichment (black curves) were calculated from the estimated false discovery rate (FDR).

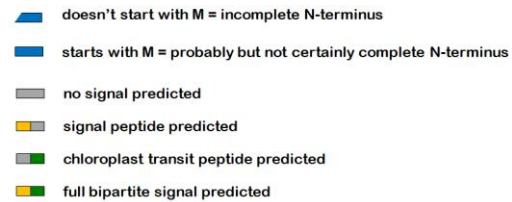
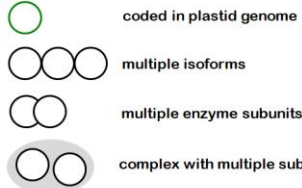
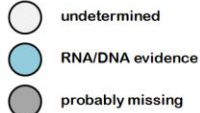
custom category name	description
<b>protein transport, folding, processing, and degradation</b>	protein translocases of plastid envelope and thylakoid membranes, signal peptidases, heat shock proteins, enzymes of post-translational protein modifications such as methylation, acetylation, glycosylation, proline cis/trans isomerisation, disulfide bond formation and breakage, and components of protein degradation systems
<b>metabolite and ion transport</b>	proteins involved in transport of all non-protein compounds, including predicted membrane transporters of undetermined substrates
<b>photosynthesis</b>	components of photosystems, light-harvesting antennae, cytochrome b6/f complex and proteins involved in their biogenesis
<b>ribosome, aminoacyl-tRNA biosynthesis and translation</b>	ribosomal proteins and proteins involved in ribosome biogenesis, aminoacyl-tRNA synthetases, translation regulators
<b>regulation and signal transduction</b>	proteins involved in other than transcriptional and translational regulation, signaling molecules and their receptors, not-further-specified protein kinases, phosphatases, adenylate cyclases and similar enzymes typically involved in signal transduction
<b>metabolism of cofactors and vitamins</b>	mostly proteins involved in chlorophyll biosynthesis, several enzymes of metabolism of ubiquinone and retinol
<b>lipid metabolism</b>	enzymes of fatty acid biosynthesis, elongation, modification and degradation, synthesis of glycerolipids and glycerophospholipids
<b>core metabolic pathways</b>	proteins of glycolysis, pentose phosphate pathway, pyruvate metabolism, carbon fixation, one-carbon and acetyl-CoA metabolism
<b>oxidative phosphorylation and electron transport</b>	components of ATP synthase and electron transport chain
<b>transcription and transcription regulation</b>	transcription and translation factors and other proteins involved in gene expression regulation
<b>RNA processing and degradation</b>	RNAses and other enzymes responsible for RNA splicing, maturation and degradation
<b>metabolism of terpenoids and polyketides</b>	proteins involved in biosynthesis of terpenoids and carotenoids
<b>DNA replication, recombination and repair</b>	DNA polymerases, ligases, helicases, proteins involved in DNA maintenance and repair
<b>reaction to oxidative and toxic stress</b>	enzymes involved in detoxification of xenobiotics and protection from reactive oxygen species and photo oxidative damage, proteins responsible for redox balance
<b>amino acid metabolism</b>	proteins involved in amino acid synthesis and interconversions, enzymes of shikimate pathway
<b>FeS cluster assembly and sulfur metabolism</b>	components of SUF pathway, proteins involved in metabolism of sulfur compounds
<b>carbohydrate metabolism</b>	enzymes of starch and other saccharides metabolism
<b>other</b>	members of other pathways which were minor in comparison to other functions

S3: List of the custom protein categories with their descriptions and examples.

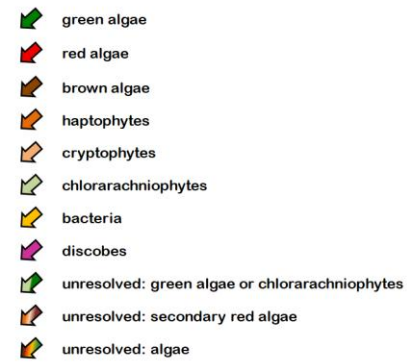


S4: Overview of the *E. gracilis* chloroplast metabolism as reconstructed from mass spectrometry-based proteome: Enzymes present in the plastid proteome in at least one isoform are marked as green circles, grey circles represent enzymes which were identified on the RNA or DNA level (in this study or previously) but are absent from the proteome; white circles represent genes completely absent in *Euglena*; circles marked by the letter “P” represent genes coded in the plastid genome while rest of the circles represent genes coded in the nucleus; circles marked by other letters represent genes with at least one of their isoforms gained via lateral transfer from one of the following donor groups: “Ch” for chlorarachniophytes, “Cr” for cryptophytes, “H” for haptophytes, “O” for ochrophytes, “R” for rhodophytes, “CHO” for unresolved secondary algae (cryptophytes, haptophytes or ochrophytes), CHOR for unresolved primary or secondary red algae (cryptophytes, haptophytes, ochrophytes or rhodophytes), and “B” for bacteria. Genes related to green algae and discobes are not marked, as well as genes with either “green or chlorarachniophyte” or completely unresolved algal origin. Multiple overlapping circles represent multiple subunits of certain enzymes.

present in plastid proteome:

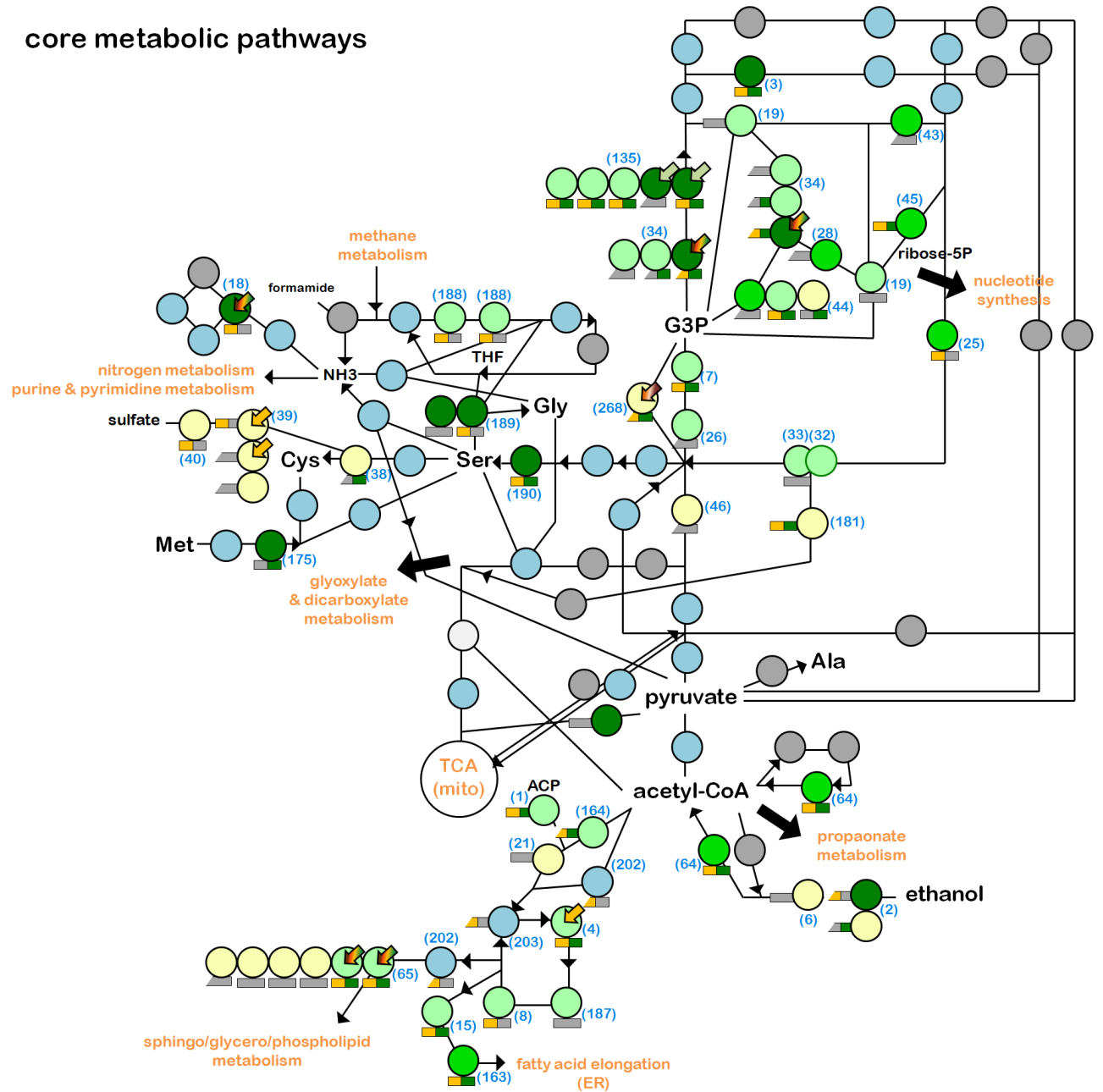


evolutionary origin:

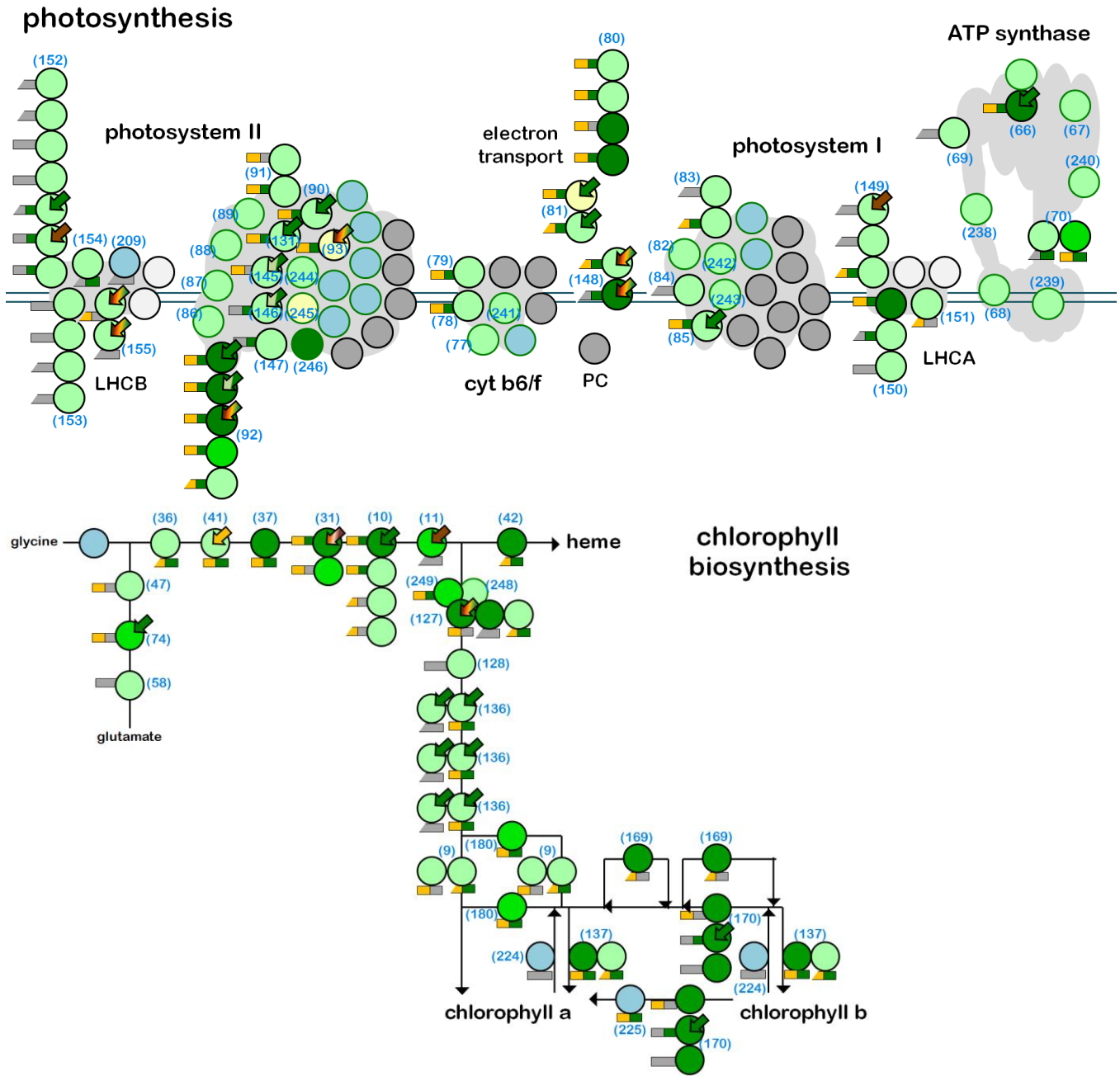


(#) protein identifier

## core metabolic pathways



S5.0: Legend for following (S5.1-11) detailed metabolic maps; proteins used in these maps can be searched in supplemental table **supplementary-dataset-1.xlsx** by the identifier in brackets; the identifier is also specified for some proteins with mere DNA/RNA-level evidence which could, however, localize to plastid and fill the gaps in the otherwise plastidial pathways or complexes.  
 S5.1: Metabolic map of *E. gracilis* chloroplast core metabolic pathways



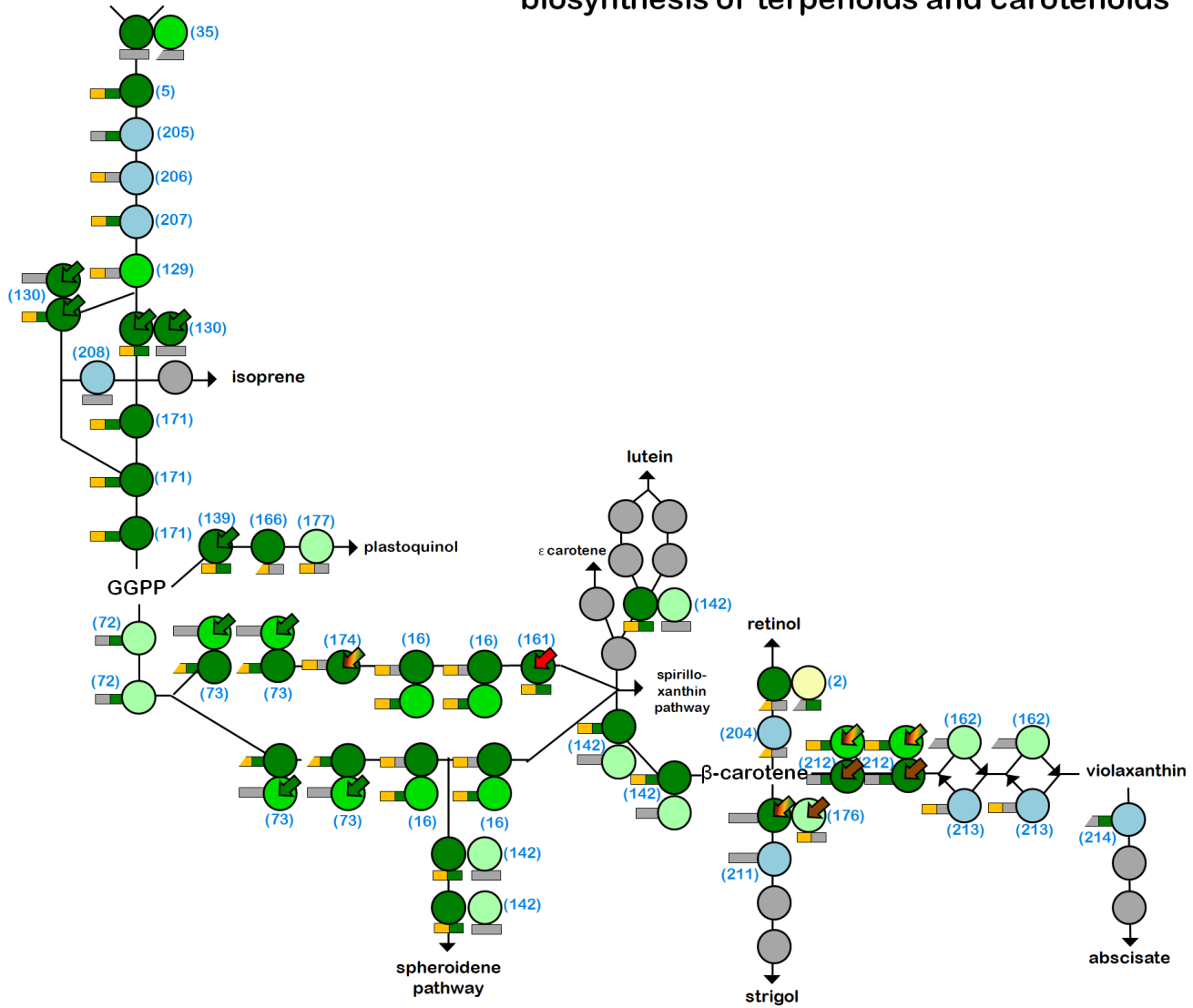
S5.2: Map of *E. gracilis* chloroplast photosynthetic apparatus

S5.3: Metabolic map of *E. gracilis* chloroplast chlorophyll synthesis

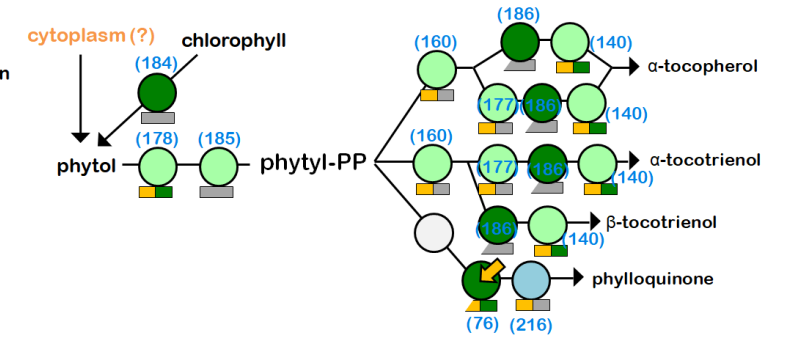


pyruvate + glyceraldehyde-3P

### biosynthesis of terpenoids and carotenoids



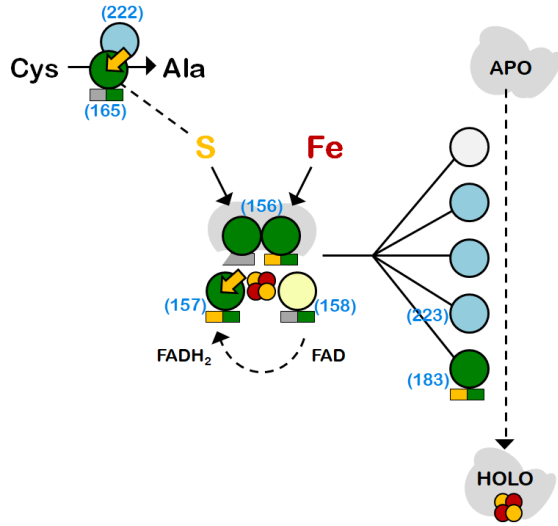
### biosynthesis of tocopherols



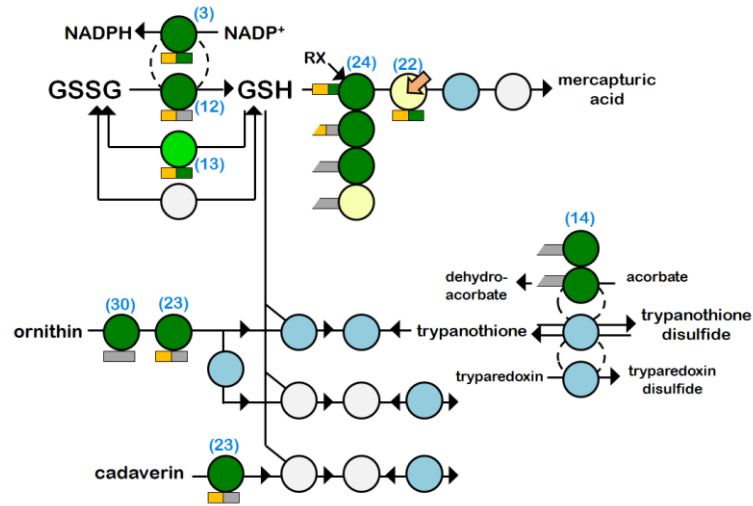
S5.4: Metabolic map of *E. gracilis* chloroplast terpenoid and carotenoid biosynthesis

S5.5: Metabolic map of *E. gracilis* chloroplast tocopherol biosynthesis

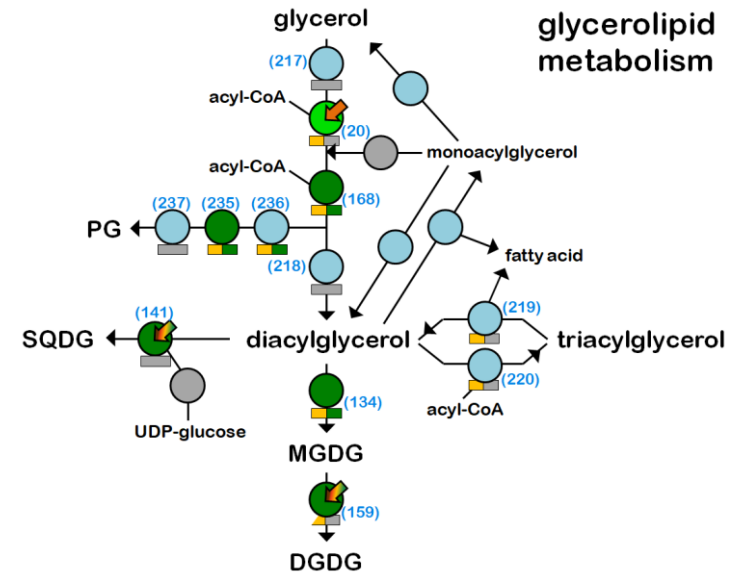
### iron-sulfur cluster assembly



### metabolism of glutathione and polyamines



### glycerolipid metabolism

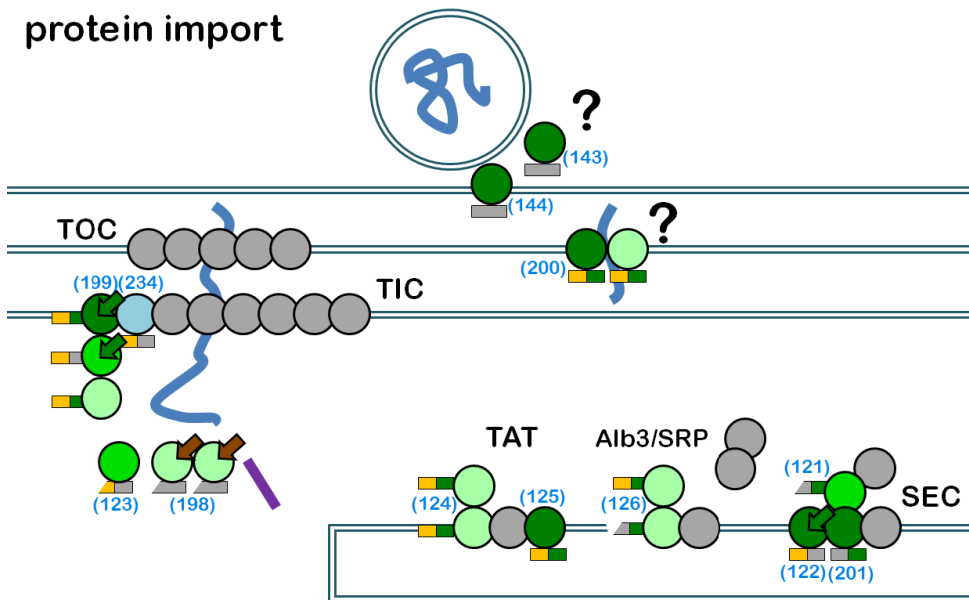


S5.6: Metabolic map of *E. gracilis* chloroplast SUF system

S5.7: Metabolic map of *E. gracilis* chloroplast part of metabolism of glutathione and polyamines

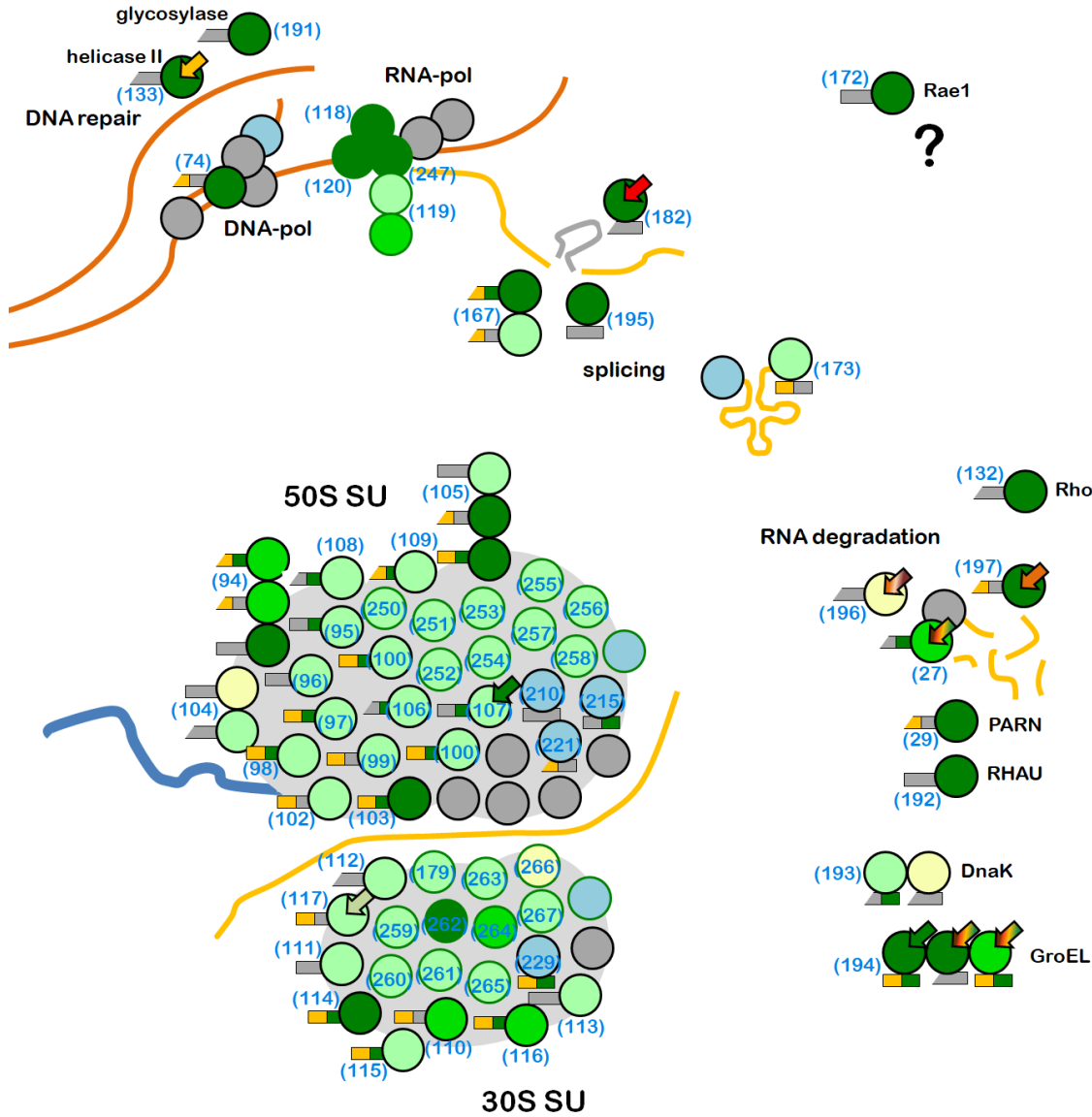
S5.8: Metabolic map of *E. gracilis* chloroplast metabolism of glycerolipids

### protein import

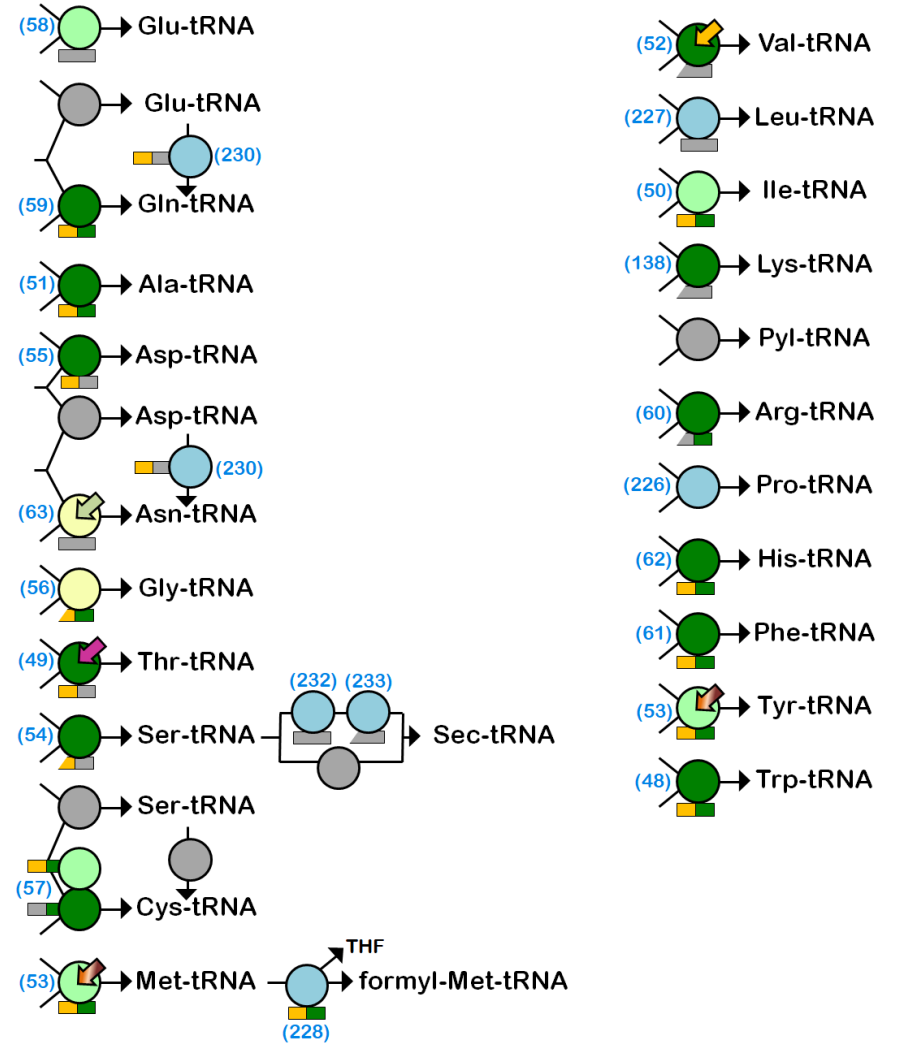


S5.9: Map of *E. gracilis* chloroplast protein importing machinery

# genetic information processing

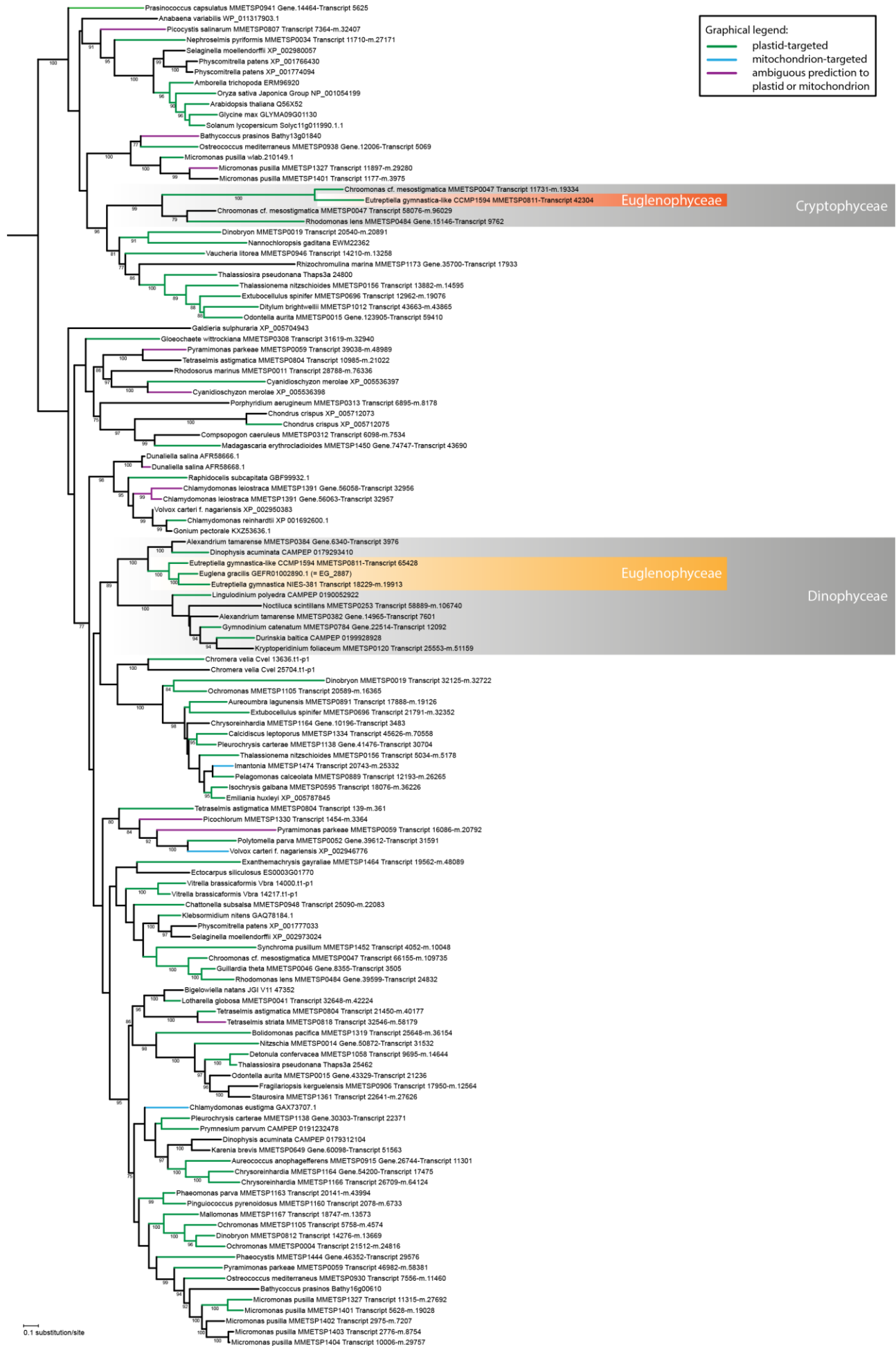


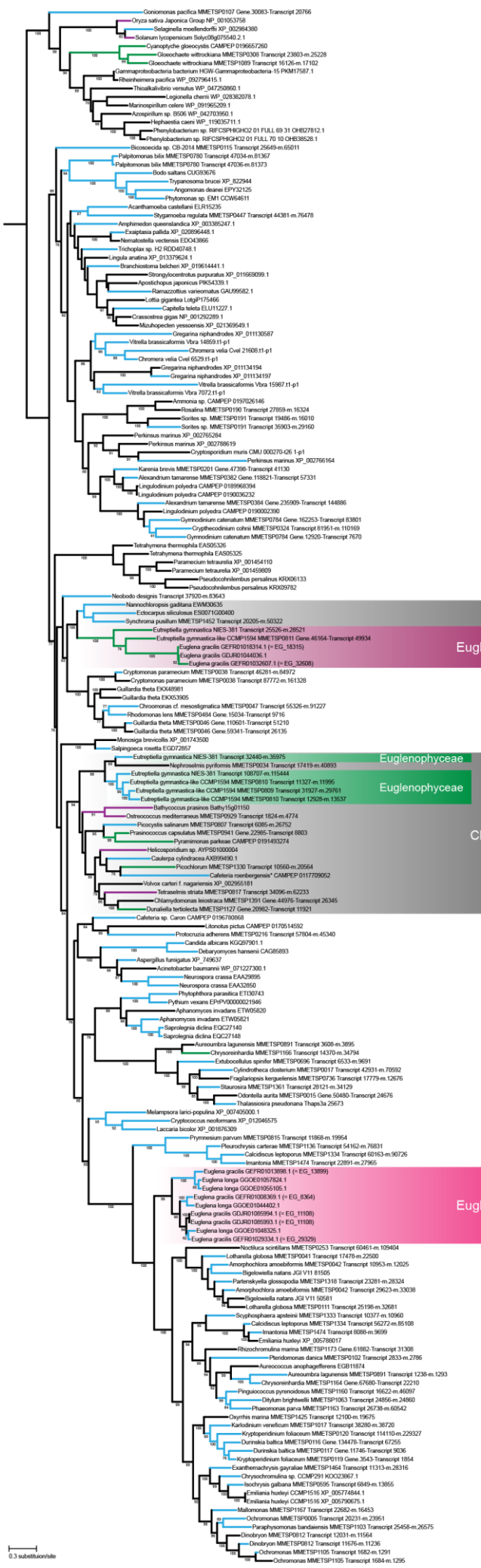
# aminoacyl-tRNA synthesis



S5.10: Map of *E. gracilis* chloroplast transcription, translation, RNA processing and degradation  
 S5.11: Map of *E. gracilis* chloroplast aminoacyl-tRNA synthesis

S6: Phylogenetic trees showing positions of the two plastid terminal oxidases (PTOX) identified in transcriptomic data of the three euglenophytes. While the PTOX1 position suggests conventional enzyme inherited from algae, the PTOX2 falls among mitochondrial alternative oxidases.





**Graphical legend:**

- plastid-targeted
- mitochondrion-targeted
- ambiguous prediction to plastid or mitochondrion

Ochrophyta  
Euglenophyceae plastid terminal oxidase

Euglenophyceae  
Euglenophyceae  
Chloroplastida

Euglenophyceae

0.3 substitutions

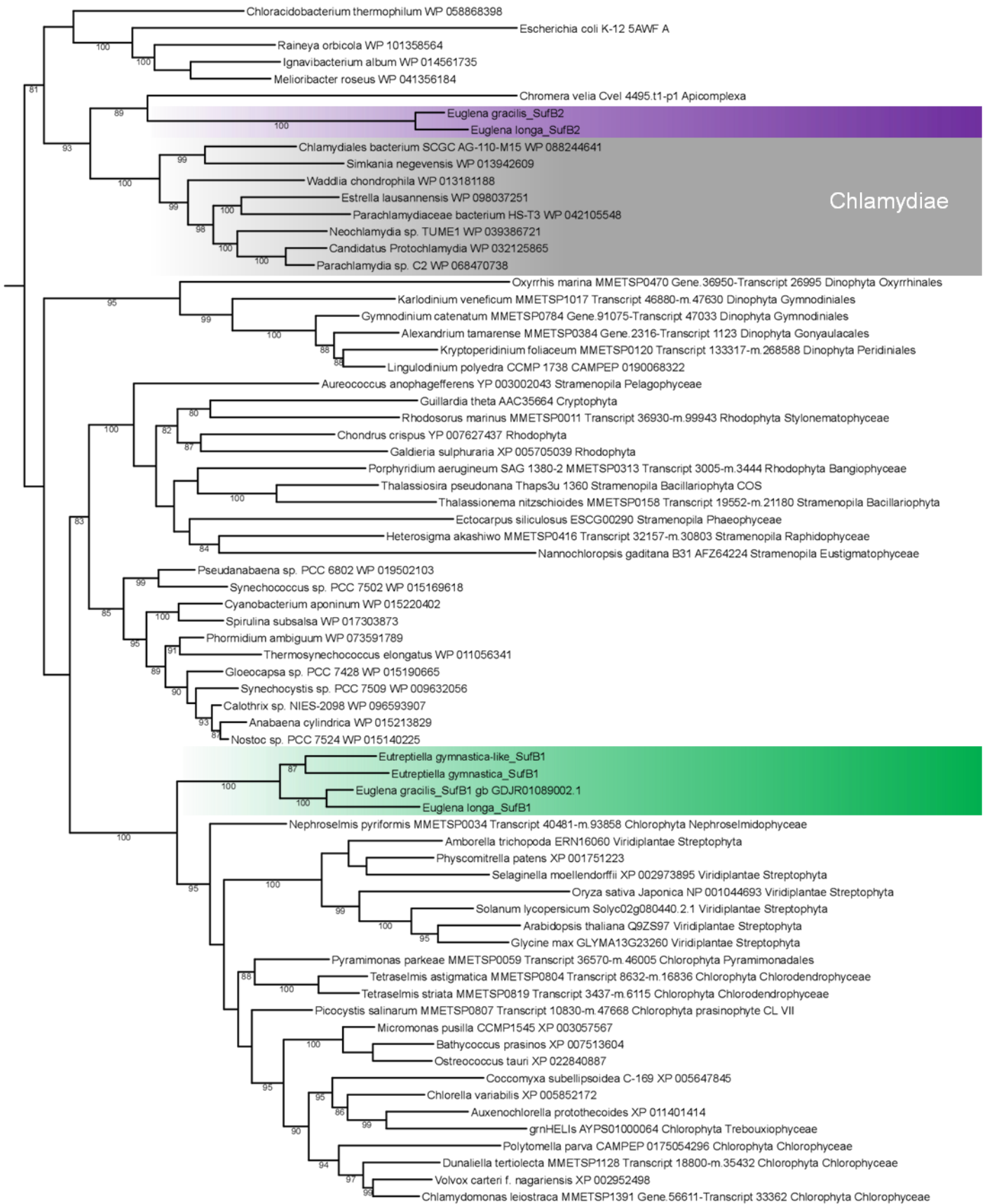
S7: Overview of SUF subunits identified in transcriptomes of *Euglena gracilis*, *Euglena longa* and *Eutreptiella gymnastica* with their sources and accession numbers and proteins corresponding to the *E. gracilis* transcripts with their respective log10 CP/MT ratios representing the credibility of plastidal localization. In case of SufB1, SufE1 and SufS1, the protein was captured in one replicate of mass spectrometry analysis of the plastidal fraction only, suggesting the protein is of lower abundance but plastid-localized. In case of SufD1 and SufS2, the protein was not captured by mass spectrometry and its localization was inferred based on the N-terminal signals and/or localizations of their putative interaction partners.

pathway	annotation	<i>E. gracilis</i> protein ID	log10 CP/MT	<i>E. gracilis</i> (GenBank)	<i>Eut. gymnastica</i> NIES-381 (MMETSP)	<i>Eut. gymnastica-like</i> CCMP1594 (MMETSP)	<i>E. longa</i> (Záhonová et Füßy et al. 2018)
SUF1	SufB1	6397	>0, one replicate only	GDJR01089002	CAMNT_0000679585 + reads	CAMNT_0046511287	Contig18694 + Contig38168 + PCR
	SufC1	13141	3+	GDJR01041264	CAMNT_0000683911	CAMNT_0046510015	Contig4493 + SL-PCR
	SufD1	6034, 33281	not captured by MS	GDJR01038170	MMETSP0039-Transcript_4862	CAMNT_0046444101	Contig15035
	SufE1	23255	>0, one replicate only	GDJR01007756	MMETSP0039-Transcript_116548	CAMNT_0046518931	Contig14296
	SufS1	12732	>0, one replicate only	GDJR01047871	CAMNT_0000708793 corrected + reads	CAMNT_0046452723	Contig3452
SUF2	SufB2	8044	3+	GEFR01008046	-	-	Contig15117
	SufC2	24338	3+	GDJR01028245	-	-	Contig63589
	SufD2	13381	0,750	GDJR01049034	MMETSP0039-Transcript_74532	MMETSP0809-Transcript_58667 + MMETSP0810-Transcript_89787 + MMETSP0811-Transcript_48154 + MMETSP0811-Transcript_48153 + MMETSP0811-Transcript_48158 + reads	Contig36957
	SufE2	32911	not captured by MS	GDJR01012039	MMETSP0039-Transcript_109175	-	Contig52519 + reads
	SufS2	9032	3+	GDJR01072295	MMETSP0039-Transcript_75235	CAMNT_0046485773 + CAMNT_0046488253 + CAMNT_0046492131 + reads	Contig31 + Contig43468

S8: Phylogenetic trees showing positions of SUF subunits identified in transcriptomic data of the three euglenophytes; the algae-related genes are highlighted in green while the genes of prokaryotic origin are highlighted in purple.

Tree scale: 0.1

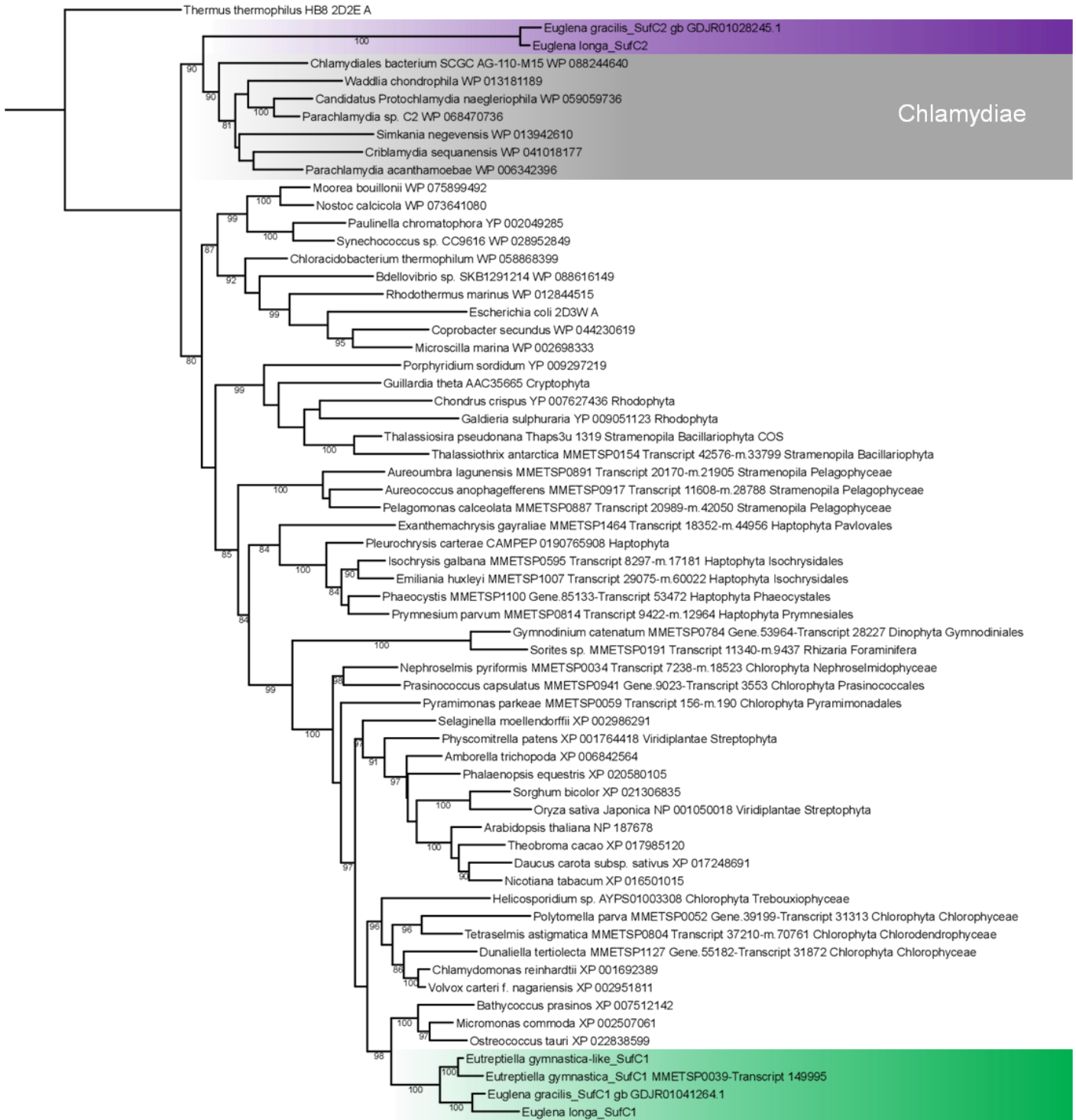
SufB



Tree scale: 0.1



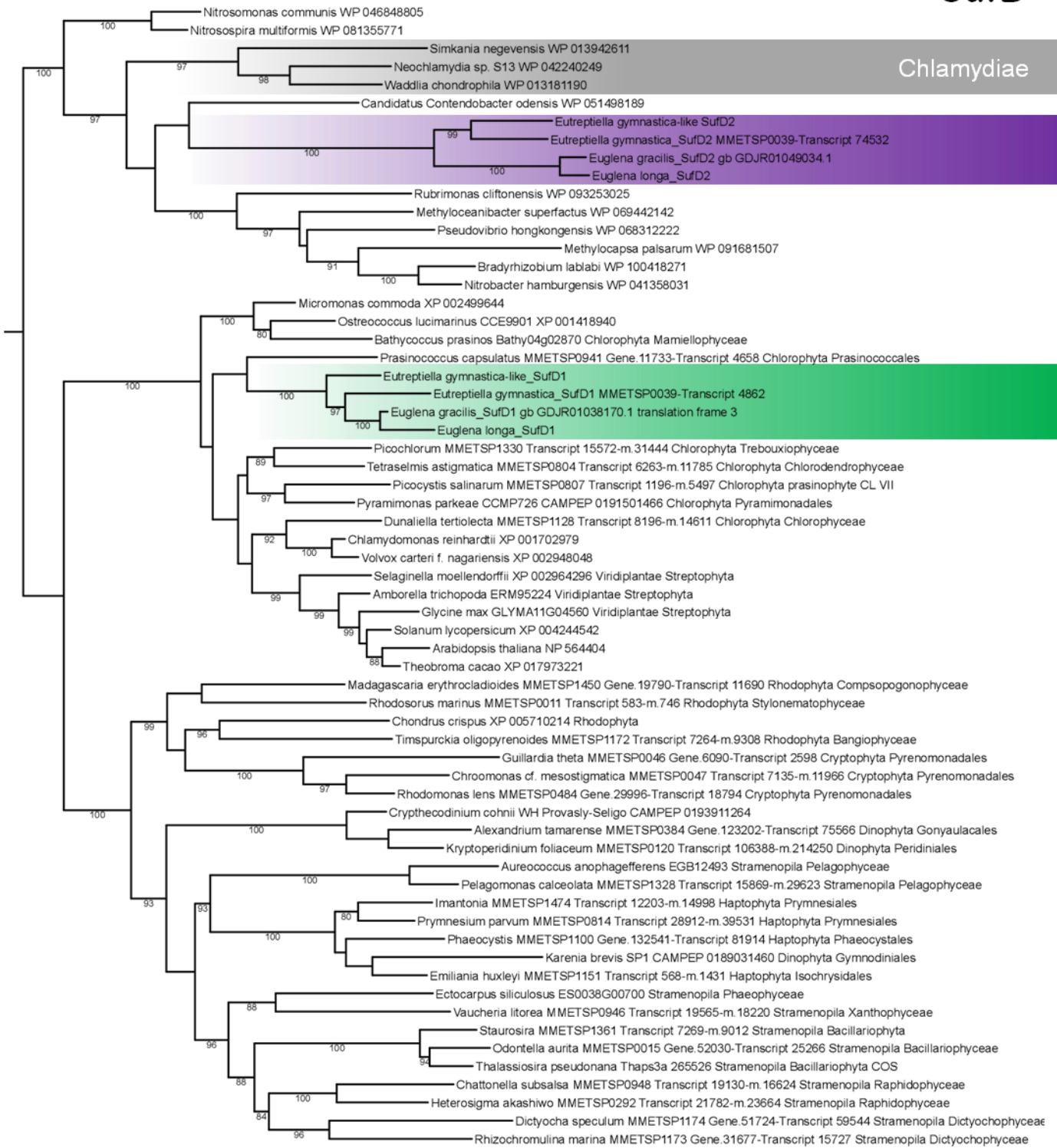
SufC





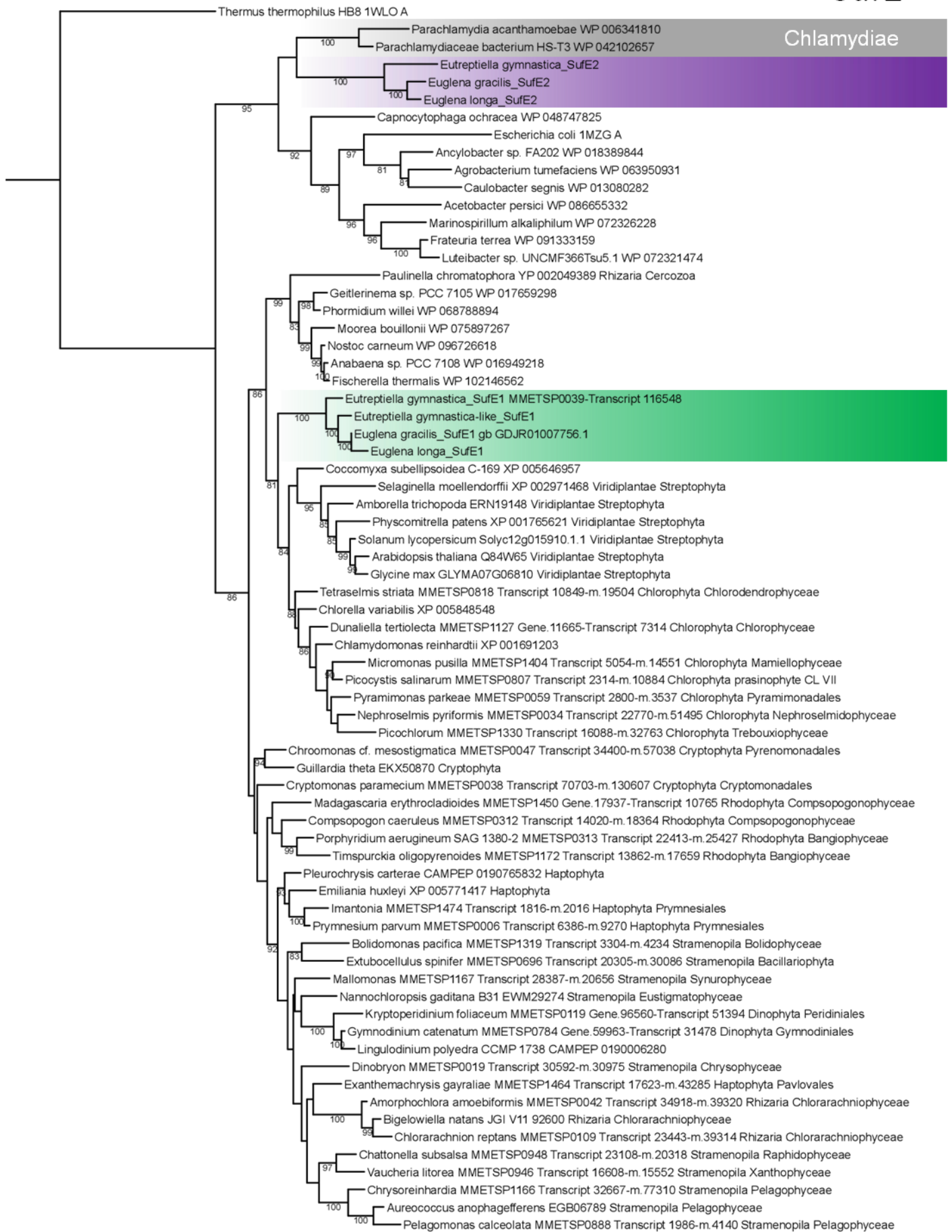
Tree scale: 0.1

SufD



Tree scale: 1

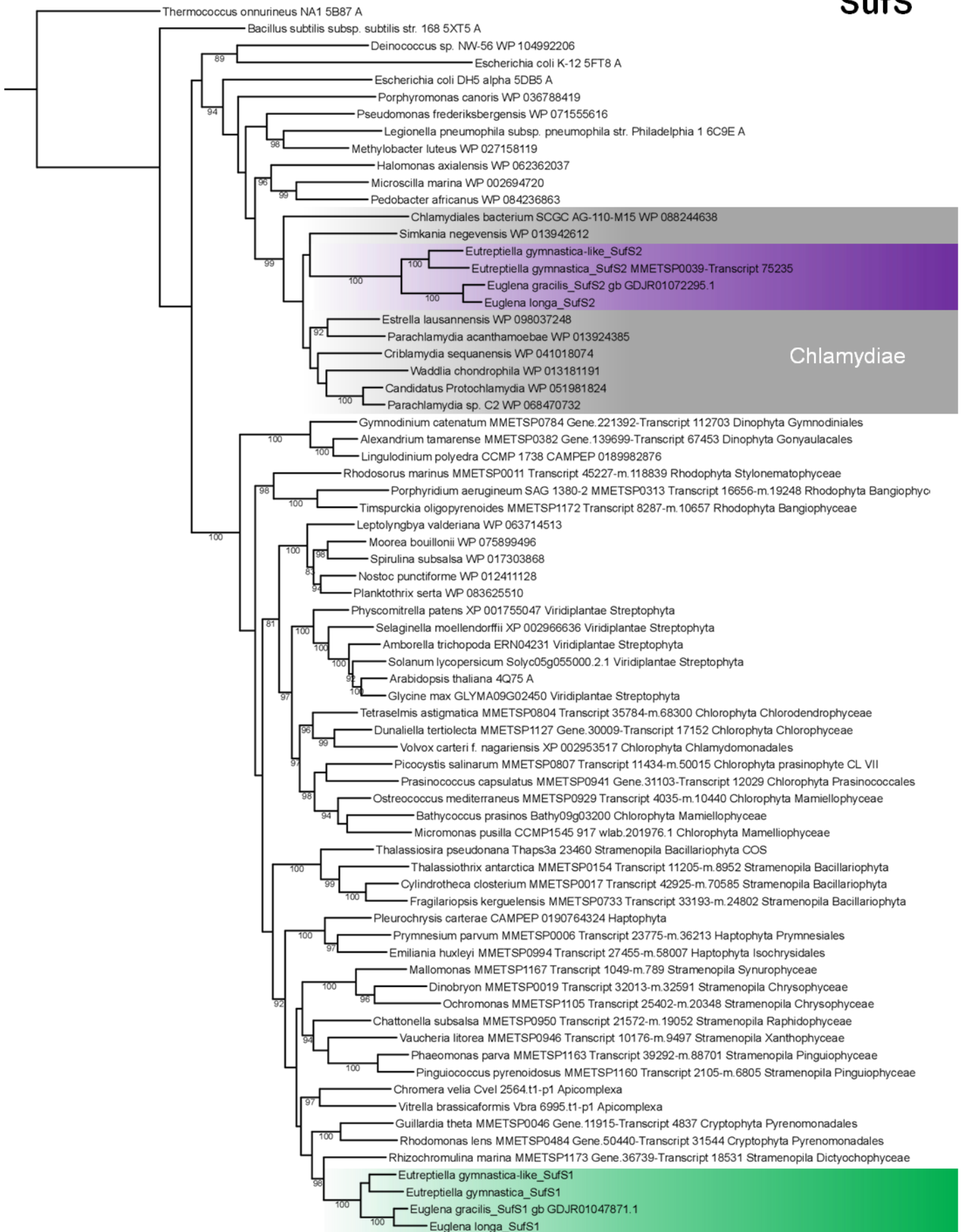
SufE



Tree scale: 0.1

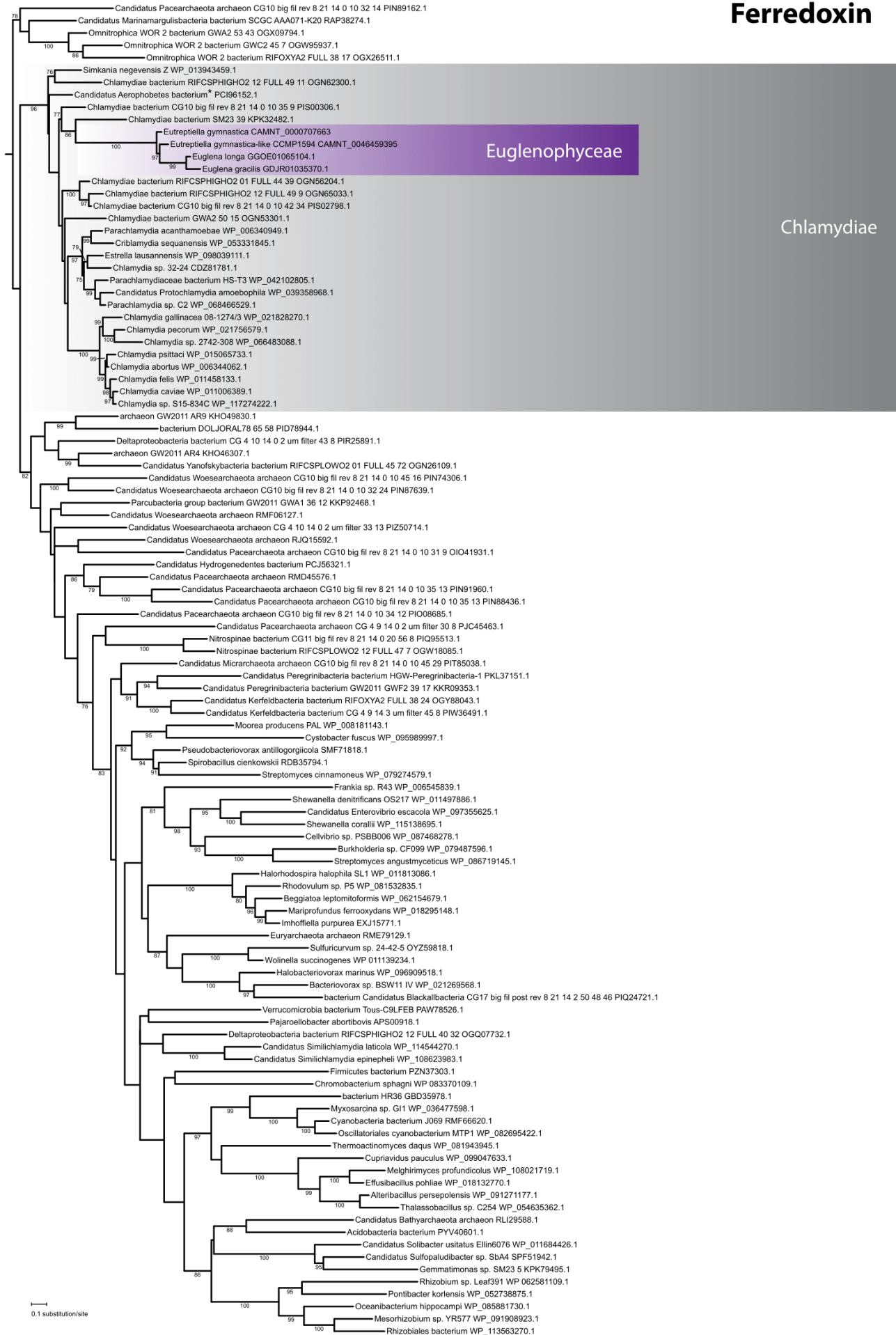


# SufS

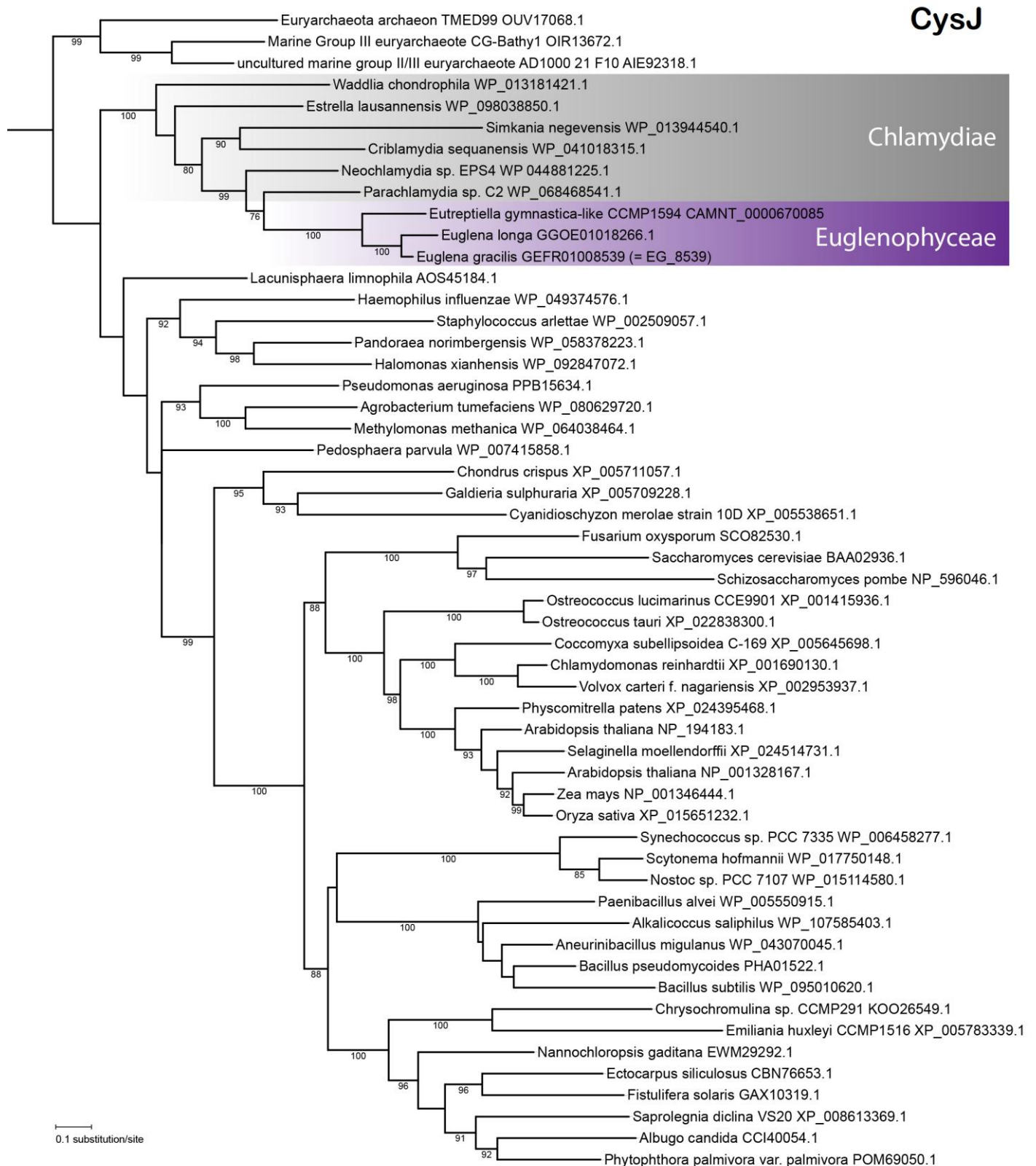


S9: Phylogenetic tree showing position of plastidial ferredoxin among chlamydiae.

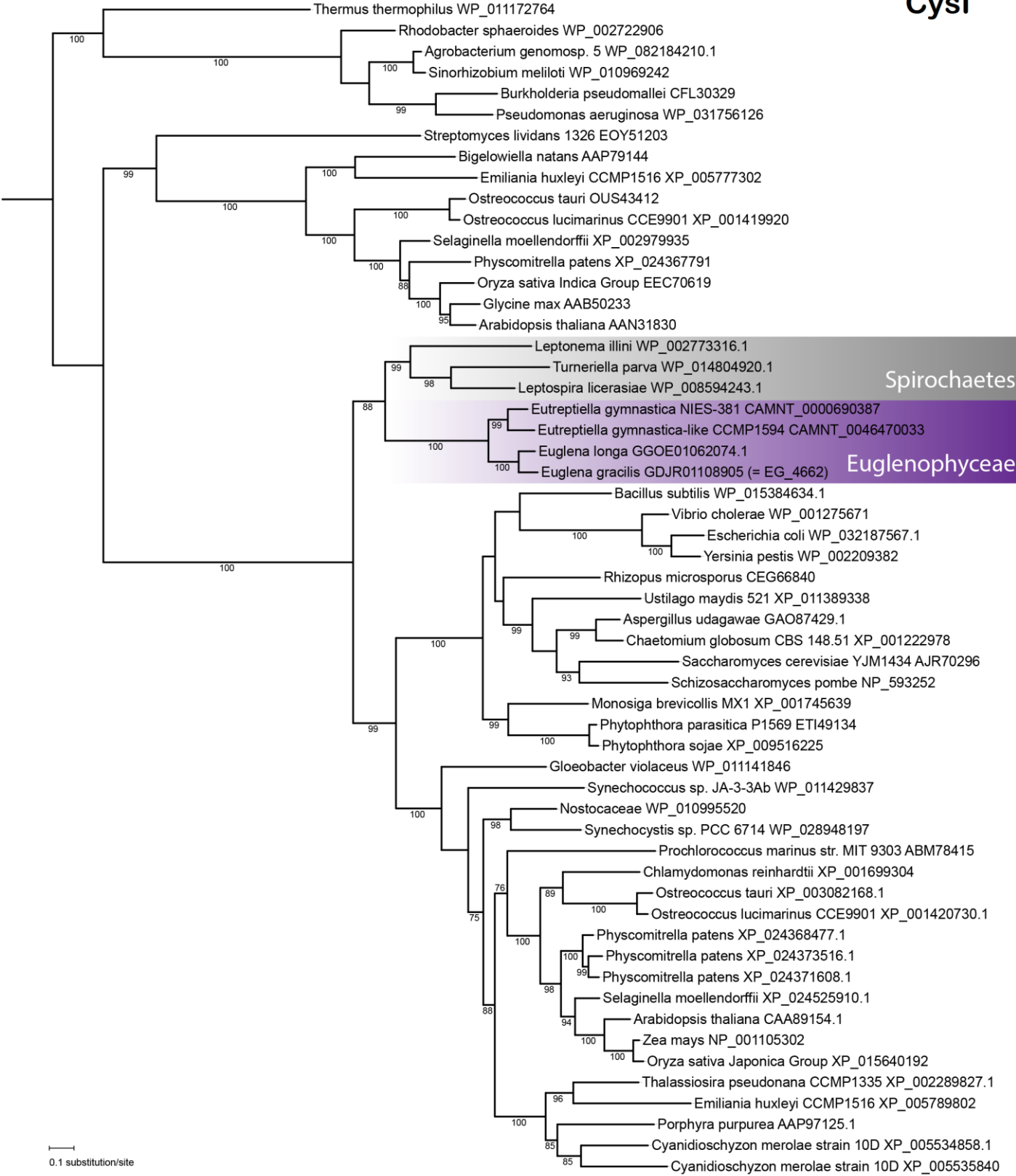
# Ferredoxin



S10: Phylogenetic trees showing positions of sulfite reductase subunits. While the subunit alpha (CysJ) sits inside chlamydiae, the subunit beta (CysI), which is also of putative prokaryotic evolutionary origin, is sister to spirochaetes.



# CysI





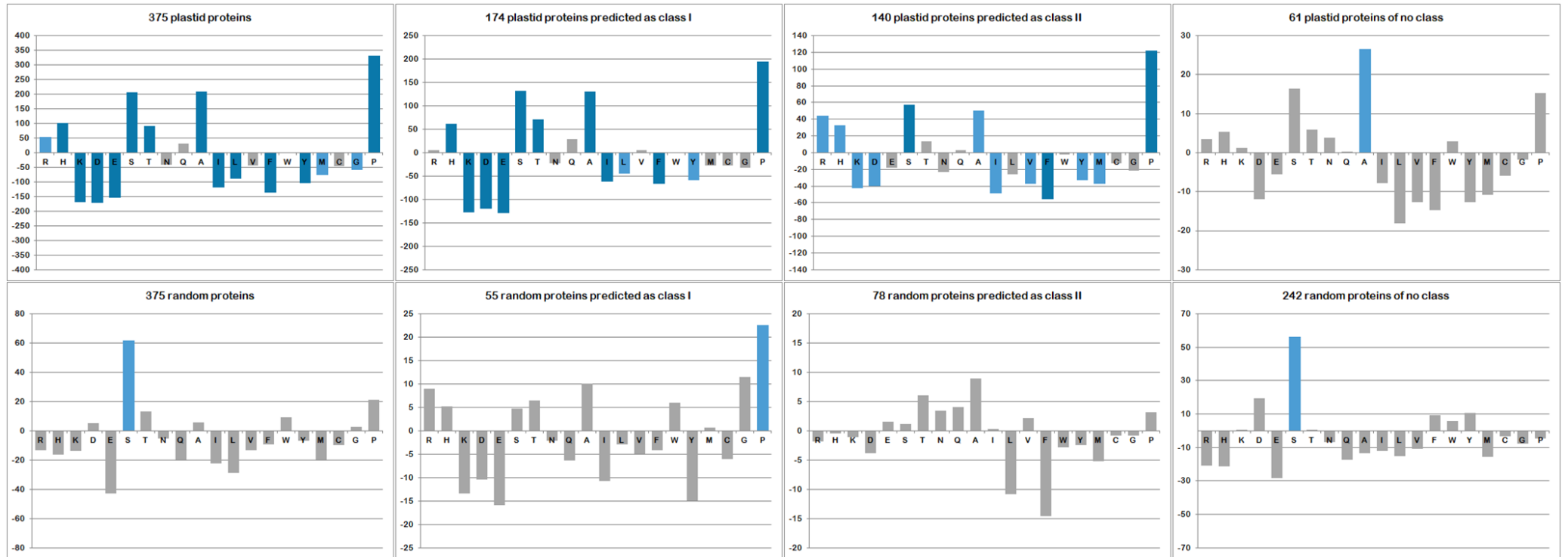
SAMPLE

CONTROL

ALL					CLASS I				CLASS II				no class						
	chisq sum	p-val chisq sum	norm chisq sum	p-val norm chisq sum		chisq sum	p-val chisq sum	norm chisq sum	p-val norm chisq sum		chisq sum	p-val chisq sum	norm chisq sum	p-val norm chisq sum		chisq sum	p-val chisq sum	norm chisq sum	p-val norm chisq sum
R	175,086	1	53,1541	0,00605	R	68,4805	1	5,5466	0,67413	R	76,4798	1	44,0918	0,00019	R	30,1258	0,99969	3,51573	0,65261
H	196,786	1	100,417	1,59E-07	H	103,726	0,99999	62,2853	2,04E-06	H	68,4743	1	32,7265	0,00485	H	24,5851	0,99999	5,40481	0,48533
K	295,634	0,99906	-165,47	0	K	165,277	0,66994	-125,249	0	K	97,0751	0,99779	-41,4843	0,00045	K	33,2821	0,99854	1,26311	0,87152
D	270,683	0,99999	-168,934	0	D	146,398	0,93696	-118,104	0	D	74,2336	1	-38,8899	0,00101	D	50,0515	0,84044	-11,9406	0,1263
E	331,85	0,94694	-150,778	6,88E-15	E	181,6	0,33097	-127,026	0	E	91,39	0,9995	-18,2782	0,1224	E	58,8598	0,55389	-5,47392	0,48339
S	401,66	0,16465	206,083	0	S	232,689	0,00196	132,049	0	S	124,561	0,82088	57,6567	1,10E-06	S	44,4105	0,94553	16,377	0,03601
T	220,009	1	90,9184	2,67E-06	T	139,829	0,97327	71,7089	5,44E-08	T	49,0246	1	13,2613	0,26238	T	31,1553	0,99947	5,94825	0,4463
N	127,755	1	-42,4131	0,0283	N	62,6668	1	-23,0973	0,07908	N	42,1703	1	-23,2282	0,04963	N	22,9175	1	3,91239	0,61642
Q	171,625	1	31,1712	0,10747	Q	109,334	0,99997	28,1758	0,03268	Q	44,4741	1	2,79381	0,81334	Q	17,8174	1	0,20158	0,97941
A	550,223	8,93E-09	207,823	0	A	322,218	6,09E-11	130,78	0	A	141,913	0,43895	50,5491	1,94E-05	A	86,0922	0,01889	26,4941	0,00069
I	198,963	1	-116,336	1,79E-09	I	94,6478	1	-60,7521	4,11E-06	I	75,0626	1	-47,7654	5,09E-05	I	29,2522	0,99981	-7,81817	0,31682
L	193,851	1	-87,2043	6,69E-06	L	84,6792	1	-43,2203	0,00105	L	69,5416	1	-25,8399	0,02897	L	39,6304	0,98461	-18,1441	0,02017
V	197,749	1	-44,6475	0,02113	V	77,5955	1	4,64804	0,72456	V	80,1055	0,99999	-36,5342	0,00202	V	40,0481	0,98254	-12,7614	0,10227
F	208,752	1	-134,126	4,32E-12	F	101,429	1	-64,428	1,04E-06	F	84,0705	0,99995	-54,9889	3,36E-06	F	23,253	1	-14,7091	0,05966
W	78,9335	1	-0,15093	0,9937	W	33,1091	1	-0,69202	0,95755	W	32,2305	1	-2,38569	0,83849	W	13,5939	1	2,92677	0,70318
Y	131,868	1	-101,205	1,43E-07	Y	60,5727	1	-56,6235	1,58E-05	Y	41,1337	1	-31,9114	0,0066	Y	30,1611	0,99954	-12,6704	0,10189
M	106,69	1	-74,8029	9,86E-05	M	37,7605	1	-27,7899	0,03357	M	45,2825	1	-36,2452	0,00196	M	23,6471	1	-10,7678	0,168
C	69,7508	1	-45,7416	0,01298	C	22,9606	1	-27,2205	0,03354	C	27,8388	1	-12,5845	0,24666	C	18,9514	1	-5,93654	0,43168
G	203,731	1	-55,8315	0,00394	G	97,7821	1	-32,365	0,01414	G	64,0528	1	-21,6138	0,06775	G	41,8965	0,97065	-1,85269	0,81249
P	794,957	0	331,381	0	P	425,55	0	193,94	0	P	300,299	1,00E-13	122,179	0	P	69,1068	0,22258	15,2626	0,05068
ALL					CLASS I				CLASS II				no class						
	chisq sum	p-val chisq sum	norm chisq sum	p-val norm chisq sum		chisq sum	p-val chisq sum	norm chisq sum	p-val norm chisq sum		chisq sum	p-val chisq sum	norm chisq sum	p-val norm chisq sum		chisq sum	p-val chisq sum	norm chisq sum	p-val norm chisq sum
R	177,641	1	-13,522	0,48501	R	27,4672	0,99931	9,04872	0,22242	R	42,351	0,99967	-1,7648	0,84162	R	107,823	1	-20,8059	0,18107
H	104,238	1	-16,265	0,40095	H	30,0401	0,99756	5,29145	0,47554	H	11,4542	1	-0,46625	0,9579	H	62,7435	1	-21,0902	0,17519
K	150,159	1	-13,8913	0,47079	K	16,3777	1	-13,4226	0,06776	K	22,1346	1	-1,06981	0,90297	K	111,647	1	0,60112	0,96905
D	139,253	1	5,40623	0,78011	D	25,0269	0,99983	-10,3959	0,16098	D	24,2952	1	-3,77318	0,66921	D	89,9305	1	19,5753	0,20827
E	207,541	1	-42,7289	0,02735	E	23,9834	0,99991	-15,9465	0,03154	E	25,6805	1	1,61052	0,8553	E	157,877	0,99999	-28,3929	0,06798
S	278,938	0,99994	62,0444	0,00136	S	29,437	0,99815	4,80842	0,51675	S	40,2188	0,99988	1,16952	0,89465	S	209,282	0,93691	56,0665	0,00031
T	145,228	1	13,1402	0,49742	T	25,8714	0,99972	6,54869	0,37722	T	16,4064	1	6,0564	0,49287	T	102,95	1	0,53511	0,97256
N	101,854	1	-5,51324	0,7747	N	16,039	1	-2,09914	0,77514	N	14,2222	1	3,48531	0,69123	N	71,5924	1	-6,89941	0,65606
Q	127,543	1	-19,6212	0,31095	Q	12,2343	1	-6,35625	0,3914	Q	24,9884	1	4,05136	0,64643	Q	90,3204	1	-17,3163	0,26565
A	275,145	0,99997	5,67527	0,76947	A	56,5231	0,41786	10,0478	0,17547	A	36,8012	0,99998	8,98436	0,30902	A	181,821	0,99852	-13,3569	0,39055
I	131,641	1	-22,2844	0,2492	I	24,5822	0,99987	-10,6552	0,15079	I	21,1974	1	0,31489	0,97156	I	85,8615	1	-11,9441	0,44166
L	197,538	1	-28,5685	0,14014	L	22,4132	0,99997	-2,84671	0,70109	L	34,1793	1	-10,7669	0,2228	L	140,945	1	-14,9549	0,33638
V	164,256	1	-13,4159	0,48844	V	23,582	0,99993	-4,91324	0,50765	V	35,4584	0,99999	2,2388	0,79989	V	105,216	1	-10,7415	0,48989
F	145,28	1	-9,19004	0,63464	F	20,2589	1	-4,13733	0,57693	F	24,7953	1	-14,6199	0,09785	F	100,225	1	9,5672	0,53771
W	81,3488	1	9,13389	0,63305	W	14,2621	1	6,06124	0,41376	W	11,9919	1	-2,84992	0,74535	W	55,0948	1	5,92257	0,69863
Y	137,628	1	-6,54392	0,73405	Y	19,4967	1	-14,9079	0,04441	Y	19,1168	1	-2,46614	0,77868	Y	99,0148	1	10,8301	0,48359
M	85,777	1	-19,9739	0,29778	M	17,7007	1	0,66401	0,928	M	14,7746	1	-5,19602	0,55376	M	53,3018	1	-15,4419	0,31583
C	104,193	1	-10,0069	0,59993	C	12,039	1	-6,06273	0,40935	C	16,4919	1	-0,77839	0,92932	C	75,6619	1	-3,16583	0,8357
G	178,568	1	2,84623	0,88315	G	27,4145	0,99933	11,4667	0,12206	G	21,0289	1	-0,84	0,92423	G	130,125	1	-7,78046	0,61697
P	427,888	0,03065	21,1201	0,27543	P	67,3212	0,12313	22,6067	0,0023	P	63,2568	0,88685	3,17688	0,71906	P	297,31	0,00879	-4,66352	0,76434

S12: Putative transit peptide region amino acid composition analysis results: The predicted “TP” region was compared to the predicted mature chain of the same sequence. This was performed a) for both experimental and control protein sets (two rows of tables), b) for the whole sets of 375 proteins regardless of their classification as well as for each subset representing a class of pre-proteins (four columns of tables), and c) for each amino acid (twenty rows in each table).  $\chi^2$  sums and their respective  $p$ -values, as well as normalized  $\chi^2$  sums (sum of  $\chi^2$  with plus or minus sign depending on the positive or negative value of its residual) and their  $p$ -values are shown in each table.  $P$ -values lower than 0.01 are colored in yellow,  $p$ -values lower than 0.00001 are colored in light orange.





S13: Graphical representation of normalized  $\chi^2$  sums for each amino acid in the TP vs. mature chain comparison in the two sets: positive or negative value reflects whether the amino acid frequency in the “TP” region is higher or lower than expected, colored bars represent statistically significant differences with  $p < 0.01$  (lighter blue) and  $p < 0.00001$  (darker blue).