Supplementary information to

A Unique LHCE Light-Harvesting protein Family is involved in Photosystem I and II Far-Red Absorption in *Euglena gracilis*

Héctor Miranda-Astudillo[1,2], Rameez Arshad[3], Félix Vega de Luna[1], Zhaida Aguilar-Gonzalez[2], Hadrien Forêt[1], Tom Feller[1], Alain Gervasi[1], Wojciech Nawrocki[1,4], Charles Counson[1,5], Pierre Morsomme[6], Hervé Degand[6], Denis Baurain[5], Roman Kouřil[3], Pierre Cardol[1]

[1] Genetics and Physiology of microalgae, InBioS/Phytosystems, University of Liège, Belgium.
[2] Present Address: Departamento de Biología Molecular y Biotecnología, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, Mexico City, Mexico.
[3] Faculty of Science, Department of Biophysics, Palacký University Olomouc, Czech Republic.
[4] Institute of Physical and Chemical Biology, French National Centre for Scientific Research, France.
[5] Eukaryotic Phylogenomics, InBioS–PhytoSYSTEMS, University of Liège, Belgium.
[6] Institut des Sciences de la Vie, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Table of Contents

ids	mick	emilie	koziol	uniprot
LHCA5-Nter_Koziol/LHCA5-Nter@130	EG_transcript_32126	comp60527_c0_seq1	LHCA5-Nter	
LHCB4-1_GEFR01021489.1/F-1/O1/P2@205	EG_transcript_21490			
LHCB4-2_GEFR01023973.1/F+2/O1/P1@144	EG_transcript_23979	comp58634_c0_seq1,comp58634_c0_seq2,comp58634_c0_seq3,com	LHC_CP29	A6
LHCB4-3_Koziol/LHC_CP29@373	EG_transcript_23979	comp58634_c0_seq1,comp58634_c0_seq2,comp58634_c0_seq3,com	LHC_CP29	A6
LHCB4-4_GEFR01069386.1/F+3/O1/P1@102	EG_transcript_69506			
LHCB4-5_GEFR01056955.1/F-3/O3/P1@54			LHC_CP29	A6
LHCB4-6_GEFR01056955.1/F-2/O2/P1@70	EG_transcript_23979	comp58634_c0_seq2,comp58634_c0_seq3,comp58634_c0_seq5	LHC_CP29	A6
LHCB7-1_GDJR01012385.1/F-2/O1/P1@432				
LHCB7-2_GDJR01012382.1+1/F-2/O1/P1@455				
LHCB7-3_GEFR01016328.1+1/F-1/O3/P1@214	EG_transcript_16325			
LHCBM1-1_GDJR01073855.1/F-1/O1/P2@226		comp55655_c0_seq3,comp55655_c0_seq5		
LHCBM1-2_HBDM01023705.1+2/F-1/O1/P2@226			LHCBM1-1	
LHCBM1-3_HBDM01068881.1+3/F+2/O1/P2@226				
LHCBM1-4_Koziol/LHCBM1-1@154			LHCBM1-1	A4
LHCBM1-5_Koziol/LHCBM1-3@236			LHCBM1-3	
LHCBM1-6_Koziol/LHCBM4-2@235	EG_transcript_19306		LHCBM4-2,LHCBM8-2	
LHCBM1-7_Koziol/LHCBM8-3@224			LHCBM8-3	
LHCBM1-8_GEFR01011081.1/F-2/O1/P1@250	EG_transcript_11081			
LHCBM1-9_HBDM01007123.1/F-2/O1/P1@431	EG_transcript_11551	comp64821_c0_seq1,comp64821_c0_seq3		
LHCBM1-10_GEFR01057898.1/F+3/O2/P2@60				
LHCBM1-11_GEFR01023953.1+1/F-2/O1/P2@58	EG_transcript_23978			
LHCBM1-12_GEFR01057898.1/F+3/O2/P1@28			LHCBM4-1	
LHCBM2-1_Koziol/LHCBM1-2@225	EG_transcript_18931	comp52658_c0_seq1,comp52658_c0_seq2,comp52658_c0_seq4	LHCBM1-2	A4
LHCBM2-2_GEFR01059358.1/F-3/O1/P1@103	EG_transcript_59393			
LHCBM3-1_HBDM01050735.1/F+2/O1/P2@182	EG_transcript_23978,EG_transcript_27290,EG_transcript_27357			
LHCBM3-2_GEFR01023953.1+1/F-2/O1/P1@132	EG_transcript_23978			
LHCBM3-3_GEFR01044404.1+1/F-2/O1/P1@122	EG_transcript_44360	comp16309_c0_seq1		
LHCBM3-4_GEFR01016314.1/F-2/O1/P2@225	EG_transcript_16331			
LHCBM3-5_GEFR01039900.1/F+2/O1/P1@88	EG_transcript_16331,EG_transcri	comp16309_c0_seq1		A2
LHCBM3-6_GEFR01057898.1/F-1/O2/P1@49				
LHCBM3-7_GDJR01073858.1+1/F-2/O1/P1@55	EG_transcript_19306			
LHCBM5-1_GDJR01073851.1/F-3/O1/P1@234		comp59498_c0_seq6		
LHCBM5-2_GDJR01073854.1/F-3/O1/P1@230		comp59498_c0_seq6,comp59498_c0_seq7		
LHCBM5-3_GDJR01073855.1/F-1/O1/P3@229		comp59498_c0_seq6,comp59498_c0_seq7		
LHCBM5-4_GDJR01073860.1/F-1/O1/P3@236		comp59498_c0_seq6,comp59498_c0_seq7		
LHCBM5-5_GDJR01073863.1/F-3/O1/P2@230		comp59498_c0_seq6,comp59498_c0_seq7		
LHCBM5-6_GDJR01073871.1/F-3/O1/P2@234		comp59498_c0_seq6		
LHCBM5-7_GDJR01073872.1/F-3/O1/P1@235				
LHCBM5-8_GDJR01073858.1+1/F-2/O1/P3@233		comp59498_c0_seq6		
LHCBM5-9_GDJR01073862.1+3/F-3/O1/P1@231				
LHCBM5-10_HBDM01068881.1+3/F+2/O1/P3@240		comp59498_c0_seq6		
LHCBM5-11_Koziol/LHCM6-2@240	EG_transcript_23156	comp59498_c0_seq6,comp59498_c0_seq7	LHCBM6-2,LHCBM9-2,L	A3
LHCBM5-12_Koziol/LHCM6-3@240		comp59498_c0_seq6,comp59498_c0_seq7	LHCBM6-3	A3
LHCBM5-13_Koziol/LHCBM6-4@234		comp59498_c0_seq6	LHCBM6-4	A3
LHCBM6-1_GDJR01073861.1/F-3/O1/P1@233				
LHCBM6-2_GDJR01073864.1/F-1/O1/P3@239				
LHCBM6-3_GDJR01073865.1/F-2/O1/P2@230				
LHCBM6-4_GDJR01073862.1+3/F-3/O1/P2@234			LHCBM5-1	
LHCBM6-5_Koziol/LHCM5-1@217			LHCBM5-1	A1
LHCBM6-6_Koziol/LHCM5-3@238			LHCBM5-3	A1
LHCBM6-7_Koziol/LHCBM7-b@239	EG_transcript_7055		LHCBM7-b	
LHCBM6-8_GEFR01049430.1/F+2/O1/P1@26				
LHCBM7-1_GDJR01073874.1/F-3/O1/P1@232		comp59498_c0_seq11		
LHCBM7-2_GDJR01073862.1+3/F-3/O1/P3@223				
LHCBM7-3_HBDM01023705.1+2/F-1/O1/P4@233		comp59498_c0_seq11		
LHCBM7-5_Koziol/LHCBM5-2@240	EG_transcript_7055	comp59498_c0_seq11	LHCBM5-2	A1
LHCBM7-6_Koziol/LHCBM7-a@229			LHCBM7-a	
LHCBM7-7_GEFR01023953.1+1/F-3/O4/P1@56				
LHCBM8-1_GDJR01021843.1/F+3/O1/P1@161	EG_transcript_19825,EG_transcript_60906		LHCBM3-1	A5
LHCBM8-2_HBDM01023680.1+4/F-1/O1/P2@223	EG_transcript_19825		LHCBM3-2	A5
LHCBM8-3_GEFR01041770.1/F-2/O1/P1@106	EG_transcript_41739			
LHCBM8-4_Koziol/LHCBM3-1@100	EG_transcript_19825		LHCBM3-1	A5
LHCBM8-5_GEFR01061045.1/F-2/O1/P1@112	EG_transcript_60906			
LHCBM8-6_GEFR01030289.1/F-2/O1/P1@70	EG_transcript_41739			
LHCBMX1-1_GEFR01040151.1/F-2/O1/P1@113	EG_transcript_40168			
LHCBMX1-2_GEFR01056651.1/F+3/O1/P1@86				A4
LHCBMX2-1_GDJR01073856.1/F-2/O1/P2@230		comp59498_c0_seq1		
LHCBMX2-2_GEFR01016314.1/F-2/O1/P3@32	EG_transcript_16331		LHCBM4-c,LHCBM4-2,LHCBM8-2,LHCBM8-3	
LHCBMX2-3_GDJR01073862.1+3/F-3/O1/P7@237		comp59498_c0_seq1		
LHCBMX3-1_GDJR01073865.1/F-2/O1/P1@185				
LHCBMX3-2_GDJR01073869.1/F-1/O1/P2@237				
LHCBMX4-1_GDJR01073867.1/F-2/O1/P1@184				
LHCBMX5-1_HBDM01020048.1+3/F-1/O1/P2@232		comp59498_c0_seq1		
LHCBMX5-2_GDJR01073870.1/F-3/O1/P1@231		comp59498_c0_seq1		
LHCBMX5-3_GEFR01071193.1/F-3/O1/P1@65				
LHCE1-1_GEFR01006499.1/F+2/O1/P1@153	EG_transcript_6500			
LHCE1-2_GEFR01025906.1/F+3/O2/P1@250	EG_transcript_25897			
LHCE1-3_HBDM01068858.1/F+1/O1/P1@342	EG_transcript_7606			
LHCE1-4_HBDM01071460.1/F-3/O1/P1@339				
LHCE1-5_HBDM01058159.1+1/F-2/O2/P1@332				
LHCE1-6_Koziol/LHCA1-1@347			LHCA1-1	
LHCE1-7_Koziol/LHCA2-1@344			LHCA2-1	A13
LHCE1-8_Koziol/LHCA4-1@85	EG_transcript_7606		LHCA4-1	A15
LHCE1-9_GEFR01060700.1/F+3/O1/P1@57				
LHCE2-1_Koziol/LHCA1-2@191		comp64492_c0_seq8,comp64492_c0_seq16,comp64492_c0_seq18	LHCA1-2	A12
LHCE2-2_Koziol/LHCA2-2@191			LHCA2-2	A13
LHCE2-3_GEFR01007608.1/F-2/O1/P2@151	EG_transcript_7606		LHCA4-2	A15
LHCE2-4_Koziol/LHCA4-2@190	EG_transcript_7606		LHCA4-2	A15
LHCE2-5_GEFR01006499.1/F+2/O1/P2@181	EG_transcript_6500			
LHCE2-6_GDJR01067079.1/F-2/O1/P3@181				
LHCE2-7_HBDM01056777.1/F-1/O1/P1@181				
LHCE2-8_HBDM01071460.1/F-3/O1/P2@181				
LHCE2-9_GEFR01007608.1/F-3/O8/P1@57	EG_transcript_6500	comp64492_c0_seq8,comp64492_c0_seq16,comp64492_c0_seq18	LHCA1-2	A12
LHCE3-1_GEFR01006499.1/F+2/O1/P3@174	EG_transcript_6500	comp63036_c0_seq6		
LHCE3-2_HBDM01046768.1/F-1/O1/P1@173				
LHCE3-3_GDJR01074337.1/F+1/O1/P2@172				
LHCE3-4_HBDM01058159.1+1/F-2/O2/P3@241				
LHCE3-5_Koziol/LHCA1-3@182			LHCA1-3	A12

LHCE3-6_Koziol/LHCA2-3@182			LHCA2-3	A13
LHCE3-7_Koziol/LHCA1-5@174			LHCA1-5	A12
LHCE3-8_Koziol/LHCA4-3@182			LHCA4-3	A15
LHCE3-9_Koziol/LHCA4-5@174			LHCA4-5	A15
LHCE3-10_GDJR01074336.1/F+1/O1/P2@174				
LHCE3-11_GDJR01074338.1/F+1/O1/P2@174				
LHCE3-12_GEFR01017658.1/F-1/O2/P1@178	EG_transcript_6500,EG_transcript	comp63036_c0_seq1,comp63036_c0_seq4,comp63036_c0_seq5,com	LHCA2-3,LHCA2-5	A13
LHCE4-1_Koziol/LHCA1-4@192			LHCA1-4	A12
LHCE4-2_Koziol/LHCA2-4@192		comp63036_c0_seq5	LHCA2-4	A13
LHCE4-3_GEFR01018054.1/F+1/O1/P2@182	EG_transcript_18045		LHCA1-4	A12
LHCE4-4_Koziol/LHCA4-4@192			LHCA4-4	A15
LHCE5-1_GEFR01002507.1/F-2/O1/P4@172	EG_transcript_2508			
LHCE5-2_HBDM01068688.1+1/F-3/O1/P4@172				
LHCE5-3_Koziol/LHCA6-3@182			LHCA6-3,LHCA8-5	A7
LHCE5-4_Koziol/LHCA8-3@182			LHCA8-3	A16
LHCE5-5_Koziol/LHCA6-5@174			LHCA6-5	A7
LHCE5-6_Koziol/LHCA5-6@174			LHCA5-5	A10
LHCE5-7_Koziol/LHCA5-3@182	EG_transcript_4401		LHCA5-3	A10
LHCE6-1_GEFR01002507.1/F-2/O1/P5@224	EG_transcript_2508	comp60656_c0_seq12	LHCA6-4	A7
LHCE6-2_GEFR01004399.1/F+3/O2/P4@181	EG_transcript_4401			
LHCE6-3_Koziol/LHCA8-4@189			LHCA8-4	A16
LHCE6-4_Koziol/LHCA6-4@191	EG_transcript_2508,EG_transcript	comp60656_c0_seq12	LHCA6-4	A7
LHCE6-5_Koziol/LHCA5-4@191			LHCA5-4	A10
LHCE7-1_HBDM01007325.1/F-3/O1/P2@175			LHCA7-1	A11
LHCE7-2_GEFR01002507.1/F-2/O1/P2@175	EG_transcript_2508		LHCA7-1	A11
LHCE7-3_Koziol/LHCA5-1@184	EG_transcript_32126	comp60527_c0_seq1	LHCA5-1	A10
LHCE7-4_Koziol/LHCA7-1@160	EG_transcript_2508		LHCA7-1	A11
LHCE7-5_GDJR01067079.1/F-2/O1/P2@175		comp64492_c0_seq1,comp64492_c0_seq12,comp64492_c0_seq18		
LHCE7-6_GEFR01004399.1/F+3/O1/P1@41				
LHCE8-1_GDJR01059671.1/F+2/O1/P2@178	EG_transcript_2508			
LHCE8-2_Koziol/LHCA5-2@187	EG_transcript_4401		LHCA5-2	A10
LHCE8-3_Koziol/LHCA7-2@182			LHCA7-2	A11
LHCE8-4_Koziol/LHCA8-2@66			LHCA8-2	A16
LHCE8-5_GEFR01044592.1/F+1/O1/P1@95	EG_transcript_2508	comp60527_c0_seq1		
LHCE9-1_Koziol/LHCA9-1@159			LHCA9-1	A14
LHCE9-2_Koziol/LHCA9-2@201			LHCA9-2	A14
LHCE9-3_Koziol/LHCA9-3@196			LHCA9-3	A14
LHCE9-4_HBDM01066990.1/F+1/O1/P1@58	EG_transcript_9691			
LHCE10-1_GEFR01014903.1+3/F+3/O1/P2@174	EG_transcript_14900			
LHCE10-2_Koziol/LHC10-2@184		comp60861_c0_seq1,comp60861_c0_seq2,comp60861_c0_seq3,com	LHCA11-b,LHCA10-2,LH	A9
LHCE10-3_GEFR01025204.1/F+2/O1/P2@93	EG_transcript_25199			
LHCE11-1_Koziol/LHC10-3@178			LHCA10-3	
LHCE11-2_GDJR01048454.1/F-2/O1/P2@171		comp60861_c0_seq1,comp60861_c0_seq2,comp60861_c0_seq4,com	LHCA11-c,LHCA10-5	A9
LHCE11-3_Koziol/LHCA11-a@70	EG_transcript_14900	comp60861_c0_seq5,comp60861_c0_seq6	LHCA11-a,LHCA11-c,LH	A9
LHCE12-1_HBDM01058828.1/F+3/O1/P1@372	EG_transcript_11021	comp41233_c0_seq1,comp52641_c0_seq1,comp52641_c0_seq2		
LHCE12-2_Koziol/LHCA3@398		comp41233_c0_seq1,comp52641_c0_seq1,comp52641_c0_seq2	LHCA3	A17
LHCE13-1_GDJR01039004.1/F+2/O1/P1@198				
LHCE13-2_GEFR01040018.1/F+3/O1/P1@150	EG_transcript_40006			
LHCE13-3_GDJR01039000.1+5/F+2/O1/P3@186				
LHCE13-4_GDJR01039000.1+5/F+2/O1/P2@199				
LHCE13-5_Koziol/LHCBM10-1@150			LHCBM10-1	A8
LHCE13-6_Koziol/LHCBM10-2@209			LHCBM10-2	A8
LHCE13-7_Koziol/LHCBM10-3@203			LHCBM10-3	A8
LHCE13-8_HBDM01032749.1/F-2/O1/P1@208				
LHCE13-9_HBDM01032749.1/F-2/O1/P2@148				
LHCE13-10_HBDM01070328.1/F-3/O1/P2@196				
LHCE13-11_GDJR01039003.1+1/F+2/O1/P1@198				
LHCSRL-1_GDJR01029061.1+2/F+2/O1/P1@327		comp60098_c0_seq3,comp60098_c0_seq5,comp60098_c0_seq6		
LHCSRL-2_GDJR01081559.1+2/F+3/O2/P1@338				
LHCSRL-3_HBDM01049636.1/F+2/O2/P1@338		comp3673_c0_seq1		

Supplementary Table S1. Corresponding nomenclature for *Euglena gracilis* LHC sequences. Column "protein name" corresponds to the new naming scheme proposed in this work, "protein identifier" provides the corresponding sequence identifier as appearing in the detailed phylogenetic tree of Suppl. Figure S48 (and built as follows: <name>_<accession>/<frame>/<orf#>/<segment#>@<length>; see Methods for details), "transcript identifier" gives the numeric part of the corresponding transcript accession(s) in the *Euglena gracilis* dataset used in Van Vlierberghe et al. (2021) (and originally downloaded from ftp://ftp.pride.ebi.ac.uk/pride/data/archive/2019/01/PXD009998/Euglena_translated_transcriptome.fasta), "Koziol annotation(s)" and "UniProt annotation(s)" are the closest BLAST matches to the protein sequences described in Koziol and Durnford (2008) and to those available in the UniProt database (https://www.uniprot.org/), respectively, using an identity threshold of 100% from the perspective of the final protein fragments. Moreover, since most LHC sequences encode polyproteins (Koziol and Durnford, 2008), annotations greedily applied to the full-length sequences, as in the UniProt database (Koziol et al., 2007), are less accurate than those applied to the final protein fragments (i.e.,

References to Supplementary Table S1

Koziol, A.G., Borza, T., Ishida, K.-I., Keeling, P., Lee, R.W., Durnford, D.G., 2007. Tracing the Evolution of the Light-Harvesting Antennae in Chlorophyll a/b-Containing Organisms. *Plant Physiol* **143**, 1802–1816.

Koziol, A.G., Durnford, D.G., 2008. Euglena light-harvesting complexes are encoded by multifarious polyprotein mRNAs that evolve in concert. *Mol Biol Evol* **25**, 92–100.

Van Vlierberghe M., Philippe H., Baurain D., 2021. Broadly sampled orthologous groups of eukaryotic proteins for the phylogenetic study of plastid-bearing lineages. *BMC Res Notes* **14**, 143.

Supplementary Table S2. Quantitative Proteomic Analysis of Representative *Euglena Gracilis* Samples Using LC-ESI-Q-TOF-MS

A1				sample 1	sample 2
Peptide count	Unique peptides	Confidence score	Description	Normalized abundance	
40	**20**	**153**	**LHCBM1-5**	**435530**	**435480**
27	**21**	**129**	**LHCBM6-6**	**292125**	**280648**
23	**23**	**147**	**LHCB4-CP29**	**281535**	**268402**
45	**43**	**200**	**PSBC-CP43**	**189103**	**187778**
63	**62**	**268**	**PSBB-CP47**	**179341**	**174607**
14	**8**	**97**	**LHCBM7-5**	**152378**	**154845**
53	**52**	**179**	**LHCE9-2**	**151339**	**152675**
20	**20**	**83**	**LHCBM5-11**	**108939**	**108990**
12	**12**	**74**	**PSBD-D2**	**93956**	**97203**
9	8	67	PSAB	93482	99937

A2				sample 1	sample 2
Peptide count	Unique peptides	Confidence score	Description	Normalized abundance	
22	**21**	**121**	**LHCBM1-5**	**216038**	**216158**
48	**48**	**239**	**PSBB-CP47**	**135110**	**141268**
15	**15**	**112**	**LHCB4-CP29**	**133801**	**132445**
31	**31**	**163**	**PSBC-CP43**	**97094**	**102436**
36	**23**	**163**	**LHCE9-2**	**71760**	**75994**
22	**22**	**63**	**PSBA-D1**	**71213**	**55354**
20	**18**	**116**	**LHCBM6-7**	**71174**	**74319**
8	**5**	**69**	**LHCBM7-5**	**43861**	**45614**
9	9	61	PSAB	43389	44902
11	11	68	PSAA	38899	41586
9	**9**	**56**	**PSBD-D2**	**38259**	**40266**
13	**13**	**57**	**LHCBM5-13**	**34833**	**35840**

A3				sample 1	sample 2
Peptide count	Unique peptides	Confidence score	Description	Normalized abundance	
34	10	118	**LHCE5-3**	217027	221040
34	34	130	**LHCBM8-2**	167901	169723
28	28	88	**LHCE11-1**	159113	164726
19	12	92	**LHCE10-2**	156839	155761
16	16	85	**LHCBM2-1**	154854	157320
38	37	132	**PSAA**	125847	130873
25	24	82	**LHCE6-5**	116950	123267
15	15	92	**PSAF**	112959	114694
15	5	83	**LHCE7-2**	96942	99359
16	16	116	LHCBM1-5	93079	93769
15	2	66	**LHCE13-7**	88524	90346

34	5	121	**LHCE5-7**	82747	82050
24	24	106	**PSAB**	81126	85385
20	20	112	**LHCBM8-4**	79330	78183
34	34	177	PSBB-CP47	75085	79888
22	8	108	**LHCE7-3**	73818	67562
13	13	88	LHCB4-CP29	70781	69436
27	27	156	PSBC-CP43	68453	71278
16	8	108	**LHCE8-2**	62973	65704
28	27	177	**PSAD**	61023	62046

A4				sample 1	sample 2
Peptide count	Unique peptides	Confidence score	Description	Normalized abundance	
11	**11**	**65**	**PSAA**	**73198**	**71418**
13	**6**	**70**	**LHCE6-5**	**66278**	**70903**
14	**14**	**69**	**LHCE11-1**	**52229**	**50918**
6	6	28	LHCB4-CP29	48849	49421
23	**4**	**104**	**LHCE5-7**	**44212**	**44547**
11	**11**	**68**	**PSAF**	**43935**	**43973**
24	**5**	**110**	**LHCE5-3**	**42958**	**43283**
11	**10**	**61**	**LHCBM8-2**	**40557**	**37873**
4	**4**	**32**	**PSAB**	**37786**	**39524**
12	12	80	PSBB-CP47	34879	33172
7	**7**	**44**	**LHCBM2-1**	**33974**	**32868**
11	11	70	PSBC-CP43	29274	28693
13	**8**	**71**	**LHCAE7-3**	**27336**	**28647**
23	**22**	**133**	**PSAD**	**26199**	**26528**
4	**3**	**33**	**LHCE7-1**	**19236**	**19259**
17	**7**	**80**	**LHCE8-2**	**16337**	**16501**
10	**6**	**55**	**LHE10-2**	**16148**	**16078**

A5				sample 1	sample 2
Peptide count	Unique peptides	Confidence score	Description	Normalized abundance	
7	7	27	**LHCBM1-4**	69267	75768
28	28	142	**PSBB-CP47**	61951	62876
12	12	88	**LHCBM1-5**	61207	63583
16	16	114	PSAD	59288	58060
25	25	136	**PSBC-CP43**	58589	58355
12	12	61	**PSBA-D1**	52719	56716
7	7	50	PSAA	32069	31767
21	21	127	**LHCE9-2**	25759	27856
5	5	31	LHCE3-5	18050	17379
6	6	37	**PSBD-D2**	16062	17721
10	9	49	PSAB	14592	15343

A6				sample 1	sample 2
Peptide count	Unique peptides	Confidence score	Description	Normalized abundance	

40	40	215	**PSBB-CP47**	133490	131562
27	26	162	**PSBC-CP43**	90215	88543
12	12	47	**PSBA-D1**	65284	60643
7	4	50	LHCE3-5	54338	52700
26	26	141	**LHCE9-2**	45416	44755
4	*4*	*23*	LHCE5-7	*41277*	*39864*
5	2	49	**LHCBM1-5**	26798	26786
10	10	32	**PSBD-D2**	26600	25868
6	1	58	LHCE1-7	25305	23898
10	10	52	**LHCB4-CP29**	18401	17622
5	5	45	LHCE4-1	16955	16290
11	11	65	LHCE12-2	12763	12555

A7

				sample 1	sample 2
Peptide count	Unique peptides	Confidence score	Description	Normalized abundance	
17	**10**	**90**	**LHCE3-5**	**197199**	**199363**
21	**2**	**139**	**LHCE1-7**	**129910**	**131640**
21	**7**	**121**	**LHCE2-1**	**87273**	**85212**
31	**12**	**155**	**LHCE1-6**	**86606**	**86393**
15	**15**	**107**	**LHCE12-2**	**45053**	**44618**
3	3	13	PSBB-CP47	1755	1836
8	8	51	LHCBM1-5	38390	38055
21	**1**	**125**	**LHCE2-2**	**34051**	**33150**
15	**1**	**102**	**LHCE4-1**	**32711**	**32717**

A8

				sample 1	sample 2
Peptide count	Unique peptides	Confidence score	Description	Normalized abundance	
55	**54**	**135**	**LHCBM1-5**	**1004668**	**1002999**
7	**7**	**15**	**LHCBM1-1**	**175163**	**48792**
18	**1**	**90**	**LHCBM5-11**	**140005**	**139821**
10	**9**	**97**	**LHCBM6-7**	**113955**	**118107**
9	**8**	**71**	**LHCBM7-5**	**102344**	**102144**
9	9	48	LHCA1-2	73668	72432
5	1	54	LHCA2-1	60381	60041

Supplemental Table S2 : Table A1–A8 present mass spectrometry results for eight representative *Euglena gracilis* samples (Sample 1 and Sample 2 per group) following hrCN-PAGE separation and LC-ESI-Q-TOF-MS analysis. Only proteins with a normalized abundance ≥2.5% of the total were included. For each protein, the peptide count, number of unique peptides, identification confidence score, protein description, and normalized abundance in both samples are shown. Protein identification was performed using Progenesis software against public (UniProt) and *in-house* databases.

Supplemental Table S3. Composition of the photosystem I and II cores complex in *Euglena gracilis*

Protein	Identified by	E: Score/Identified Peptides	S: Score/Identified Peptides	Accession	References
Photosystem I					
PsaA	E/S/P	153/16	250/12	EG_transcript_158	(Sobotka et al., 2017); This work
PsaB	E/S/P	142/15	29/3	NP_041907.1	(Sobotka et al., 2017); This work
PsaC	E/S/P	15/4	14/5	NP_041934.1	(Sobotka et al., 2017); This work
PsaD	E/S/P	199/18	44/4	EG_transcript_12295/EG_transcript_3688	(Sobotka et al., 2017); This work
PsaE	E/P	82/9	—	EG_transcript_31070	(Sobotka et al., 2017); This work
PsaF	E/P	76/12	—	EG_transcript_10340	(Sobotka et al., 2017); This work
PsaJ	P	—	—	YP_002720092.1	(Sobotka et al., 2017); (Bennett and Triemer, 2015); This work
PsaM	P	—	—	NP_041896.1	(Schlunegger and Stutz, 1984); This work
Photosystem II					
PsbA (D1)	E/P	123/20	—	EG_transcript_13288/NP_041895.1	(Sobotka et al., 2017); This work
PsbB(CP47)	E/S/P	309/58	378/19	EG_transcript_899	(Sobotka et al., 2017); This work
PsbC (CP43) (VII) (34kDa)	E/S/P	212/29	415/19	NP_041892.1	(Sobotka et al., 2017); This work
PsbD (D2)	E/S/P	140/16	274/14	EG_transcript_10715/NP_041889.2	(Sobotka et al., 2017); This work
PsbE (cyt559α)	E/S/P	14/4	122/8	NP_041908.1	(Cushman et al., 1988); This work
PsbF (cyt559β)	S/P	—	25/2	NP_041909.1	(Cushman et al., 1988); This work
PsbH (10kDa)	E/P	16/2	—	NP_041944.1	(Sobotka et al., 2017); This work
PsbI	E/S/P	14/2	43/3	NP_041904.1	(Sobotka et al., 2017); This work
PsbJ	P	—	—	NP_041911.1	(Jenkins et al., 1995); This work
PsbK	P	—	—	NP_041898.1	(Jenkins et al., 1995); This work
PsbL	P	—	—	NP_041910.1	(Jenkins et al., 1995); This work
PsbM	E/S/P	86/8	68/3	EG_transcript_7656	This work
PsbN	P	—	—	NP_041943.1	(Hallick et al., 1993); This work
PsbO (OEE1)	E/P	12/2	—	EG_transcript_15767/EG_transcript_8709	(Sobotka et al., 2017); This work
PsbP (OEE2)	E/P	88/10	—	EG_transcript_13851/EG_transcript_14464/EG_transcript_17799	(Sobotka et al., 2017); This work
PsbQ (OEE3)	E/P	146/24	—	EG_transcript_25934	(Sobotka et al., 2017); This work
PsbR	P	—	—	EG_transcript_22790	This work
PsbT (PsbTc) (ycf8)	P	—	—	NP_041945.1	(Jenkins et al., 1995); This work
PsbW	E/P	16/5	—	EG_transcript_30510	This work
PsbX	P			EG_transcript_39502	This work
PsbY	P	—	—	EG_transcript_32529	This work
PsbZ (ycf9)	S/P	—	15/2	NP_041931.1	(Sobotka et al., 2017); This work

Psb27	E/P	42/5	—	EG_transcript_16063/EG_transcript_16166/EG_transcript_16389	(Sobotka et al., 2017); This work
Psb28	P	—	—	EG_transcript_28462	This work
Psb29	P	—	—	EG_transcript_17545	This work
Psb30 (ycf12)	P	—	—	NP_041897.1	(Bennett and Triemer, 2015); This work
Psb32	P	—	—	EG_transcript_17812/EG_transcript_20226	This work
Psb33	P	—	—	EG_transcript_22691	This work
Ycf48	P	—	—	EG_transcript_12174	(Sobotka et al., 2017); This work

E: Liquid chromatography-Electrospray-ionisation quadrupole time-of-flight mass spectrometry from complete band from CN-PAGE. S: Tandem mass spectrometry from isolated spots from the 3D gels. P: Phylogenomic approach for conserved subunits.

“Accession” gives either the NCBI accession number, or the numeric part of the corresponding transcript accession(s) in the Euglena gracilis dataset used in Van Vlierberghe et al. (2021) (and originally downloaded from ftp://ftp.pride.ebi.ac.uk/pride/data/archive/2019/01/PXD009998/Euglena_translated_transcriptome.fasta),

Bennett, M.S., Triemer, R.E., 2015. Chloroplast Genome Evolution in the Euglenaceae. J. Eukaryot. Microbiol. 62, 773–785. https://doi.org/10.1111/jeu.12235

Cushman, C., Christopher, D.A., Little, M.C., Hailick, R.B., Price, C.A., 1988. Organization of the psbE, psbF, off38, and off42 gene loci on the Euglena gracilis chloroplast genome. Curr. Genet. 13, 173–180.

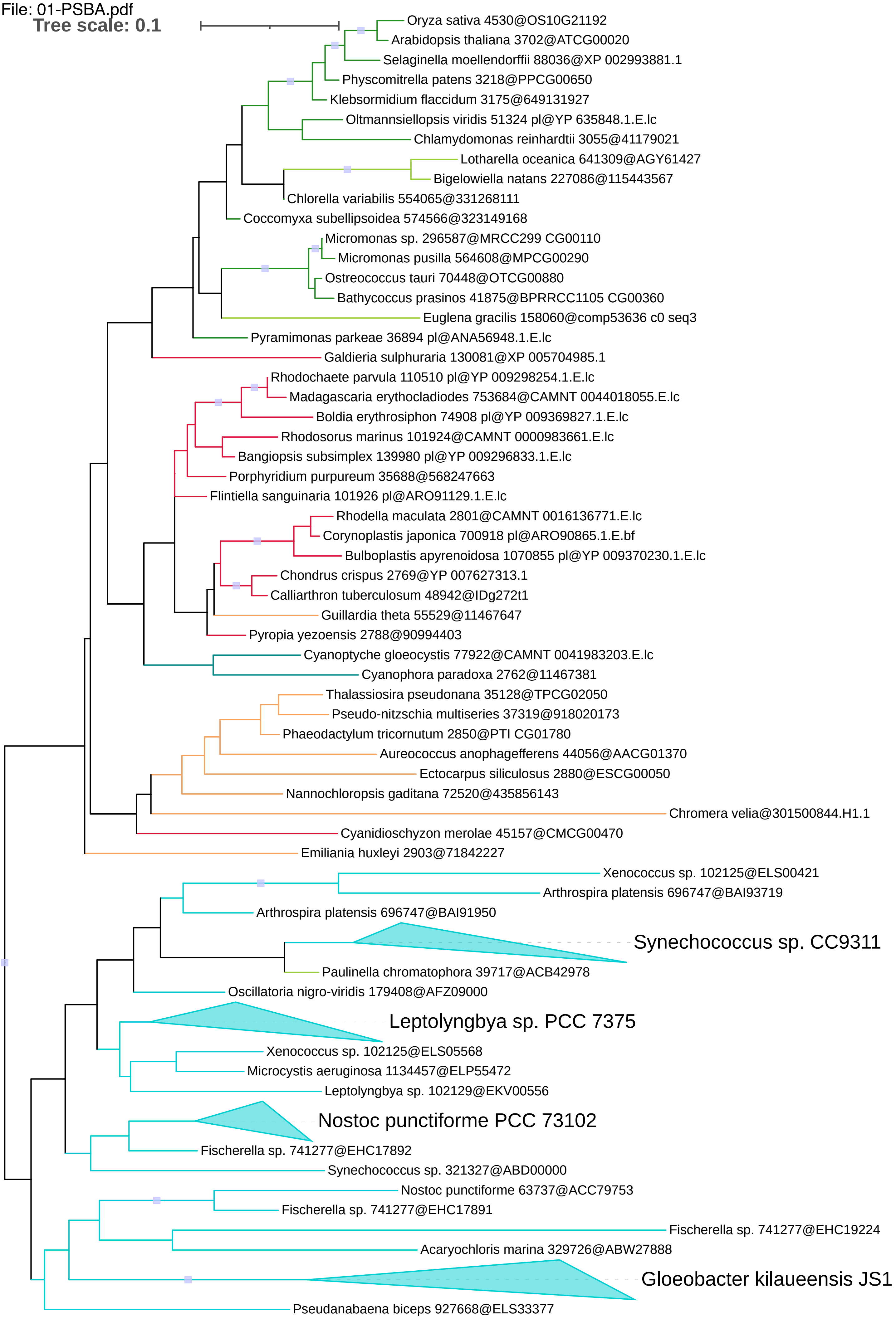
Hallick, R.B., Hong, L., Drager, R.G., Favreau, M.R., Monfort, A., Orsat, B., Spielmann, A., Stutz, E., 1993. Complete sequence of Euglena gracilis chloroplast DNA. Nucleic Acids Res. 21, 3537–3544. https://doi.org/10.1093/nar/21.15.3537

Jenkins, K.P., Hong, L., Hallick, R.B., 1995. Alternative splicing of the Euglena gracilis chloroplast roaA transcript. RNA 1, 624–633.

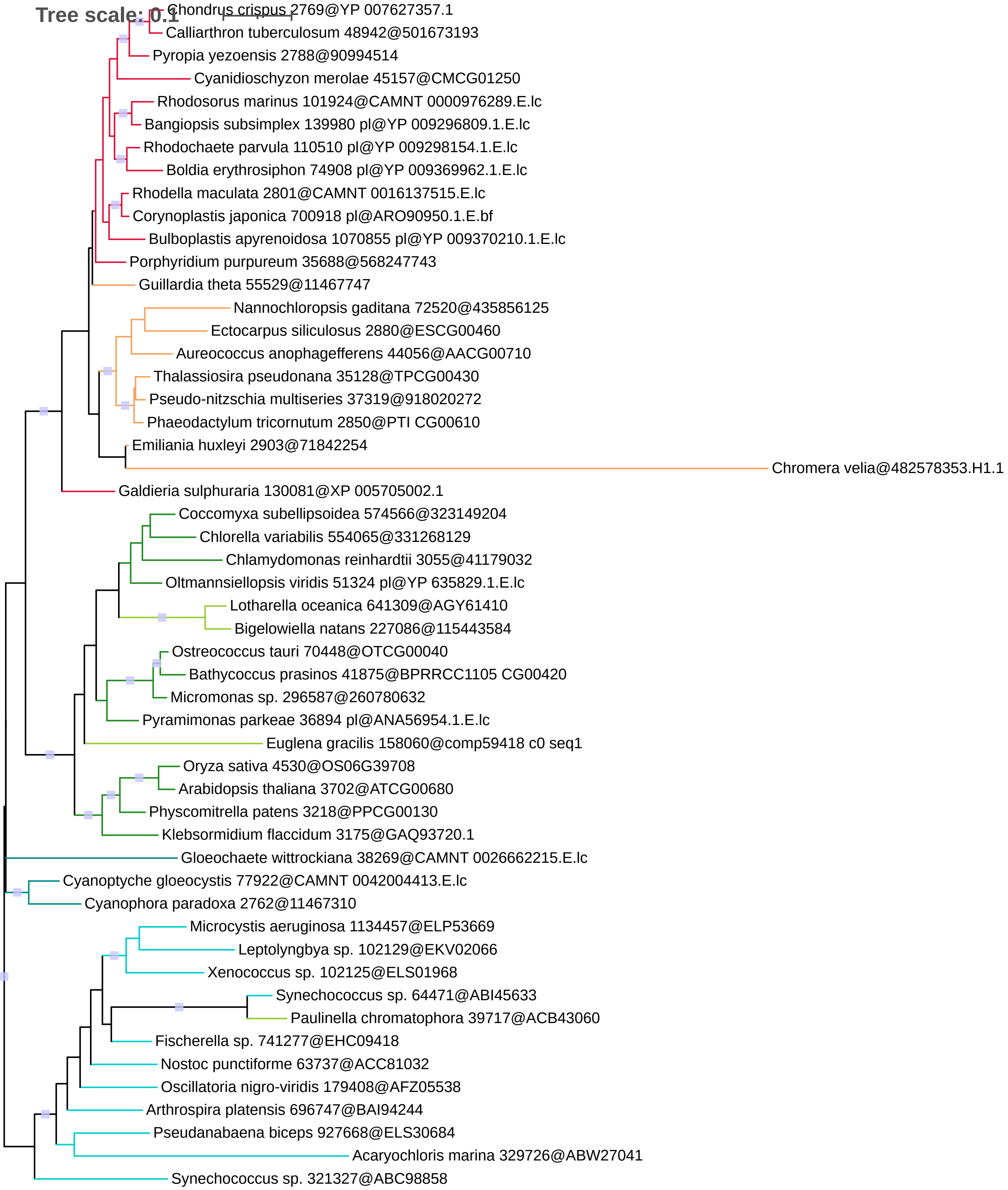
Schlunegger, B., Stutz, E., 1984. The Euglena gracilis chloroplast genome: structural features of a DNA region possibly carrying the single origin of DNA replication. Curr. Genet. 8, 629–634. https://doi.org/https://doi.org/10.1007/BF00395709

Sobotka, R., Esson, H.J., Koník, P., Trsková, E., Moravcová, L., Horák, A., Dufková, P., Oborník, M., 2017. Extensive gain and loss of photosystem i subunits in chromerid algae, photosynthetic relatives of apicomplexans. Sci. Rep. 7, 1–13. https://doi.org/10.1038/s41598-017-13575-x

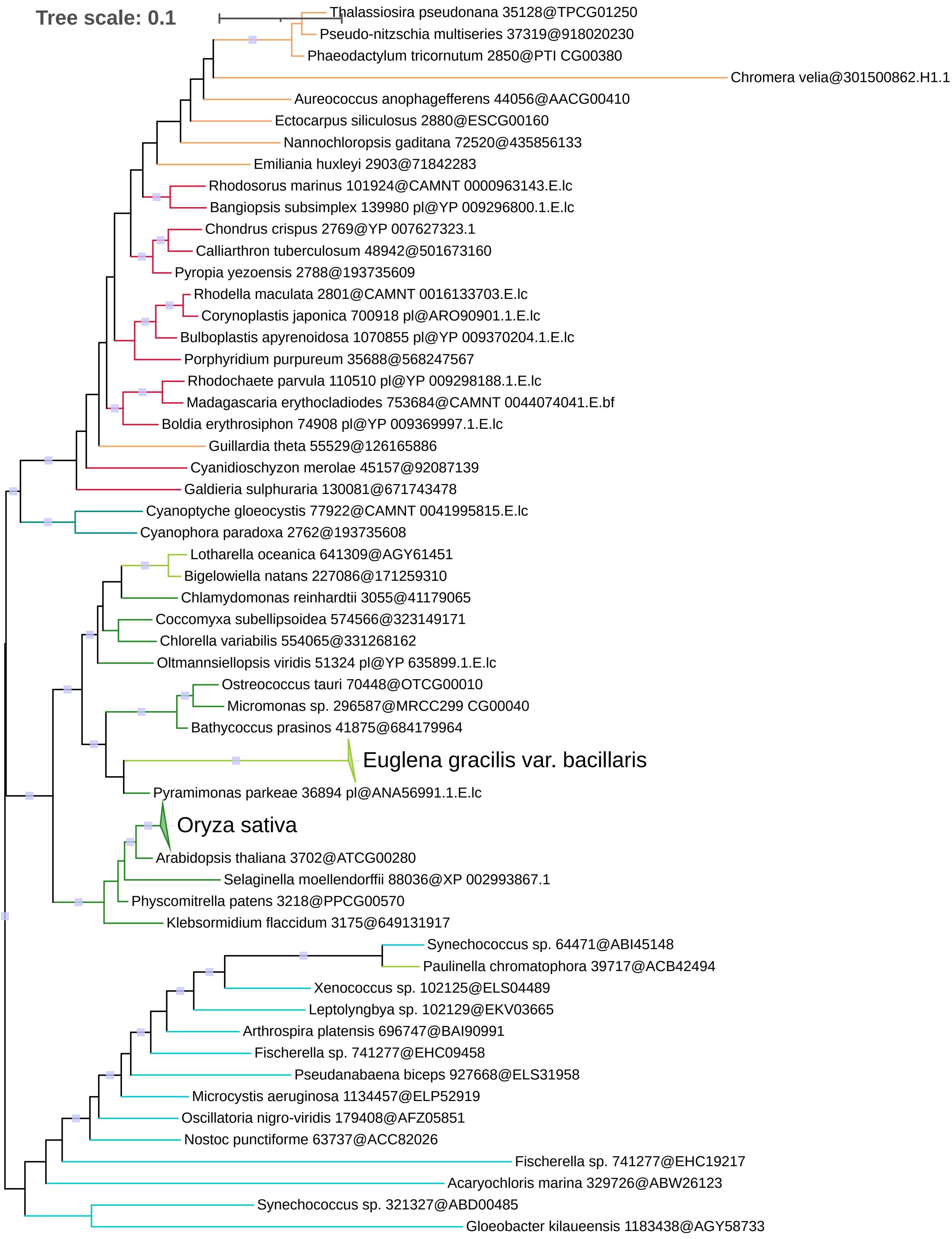
File: 01-PSBA.pdf



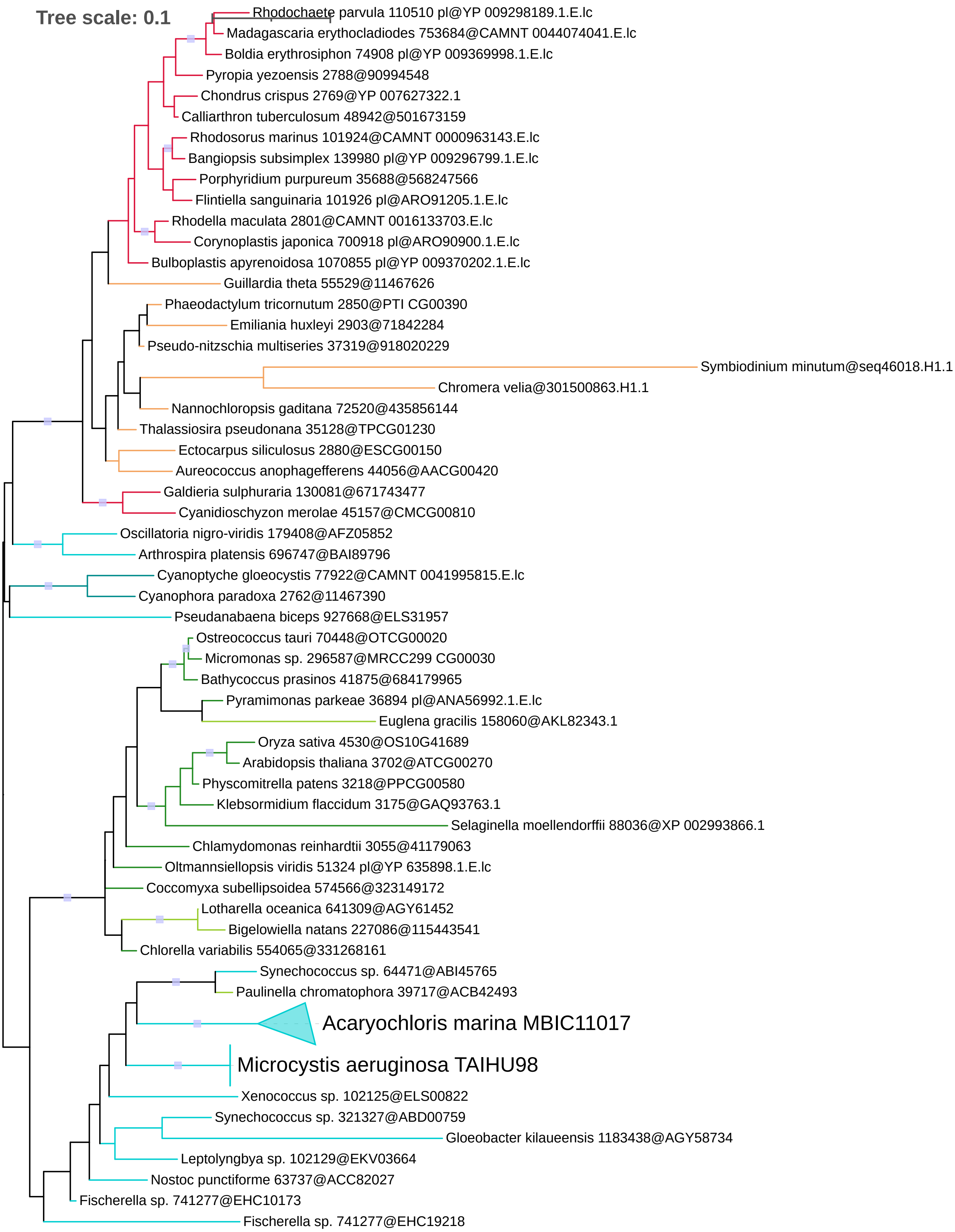
File: 02-PSBB.pdf



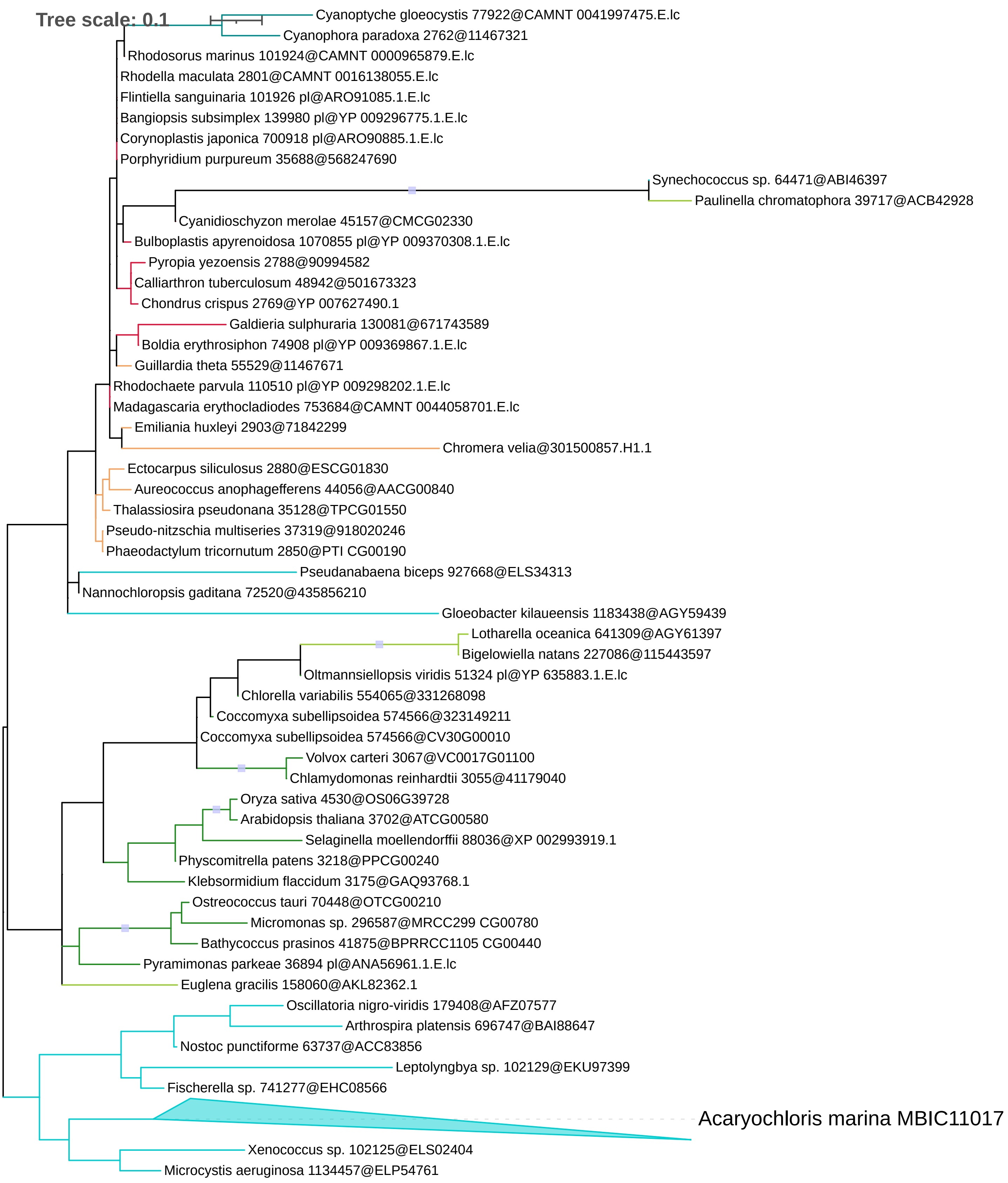
File: 03-PSBC.pdf



File: 04-PSBD.pdf

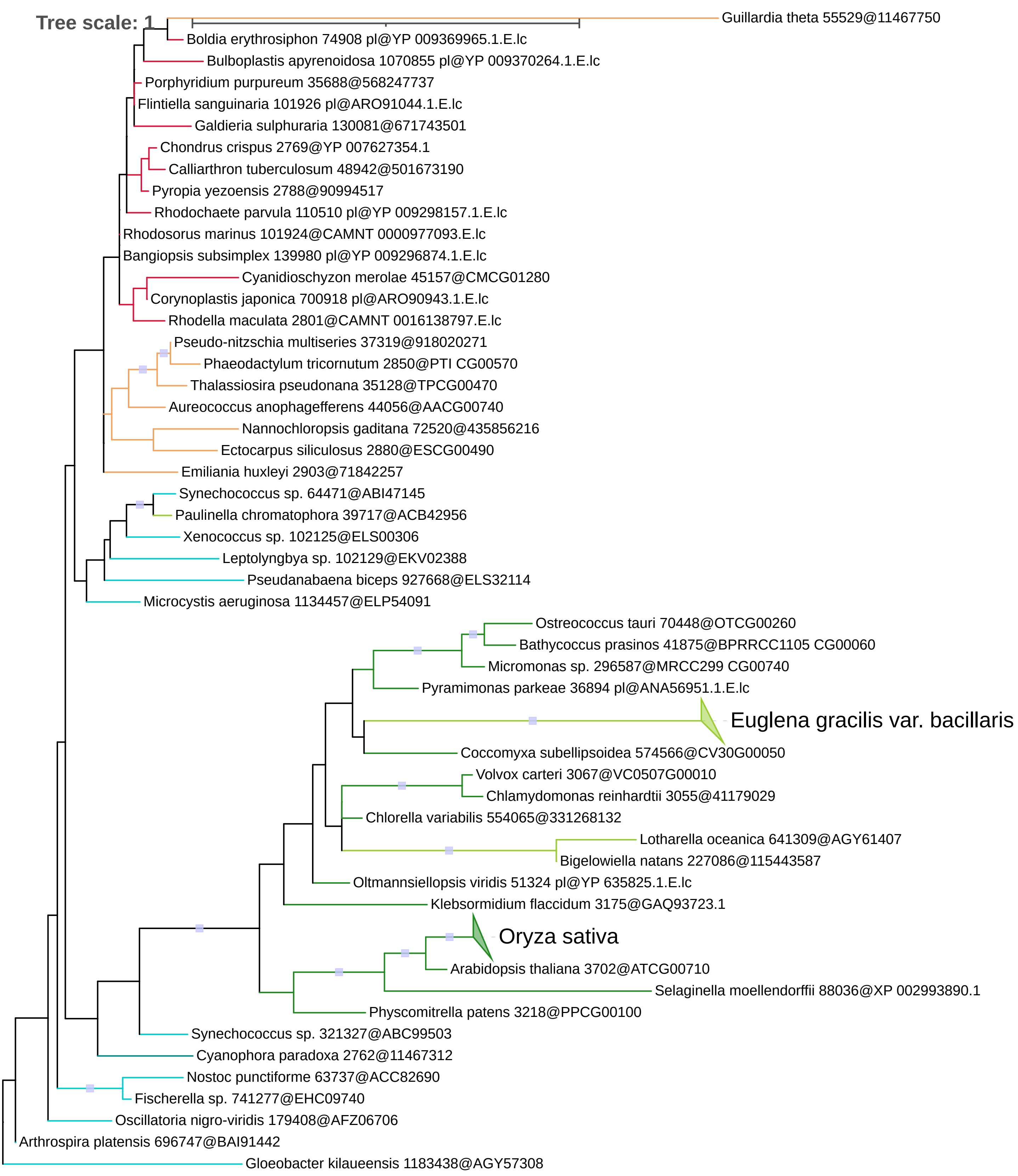


File: 05-PSBE.pdf

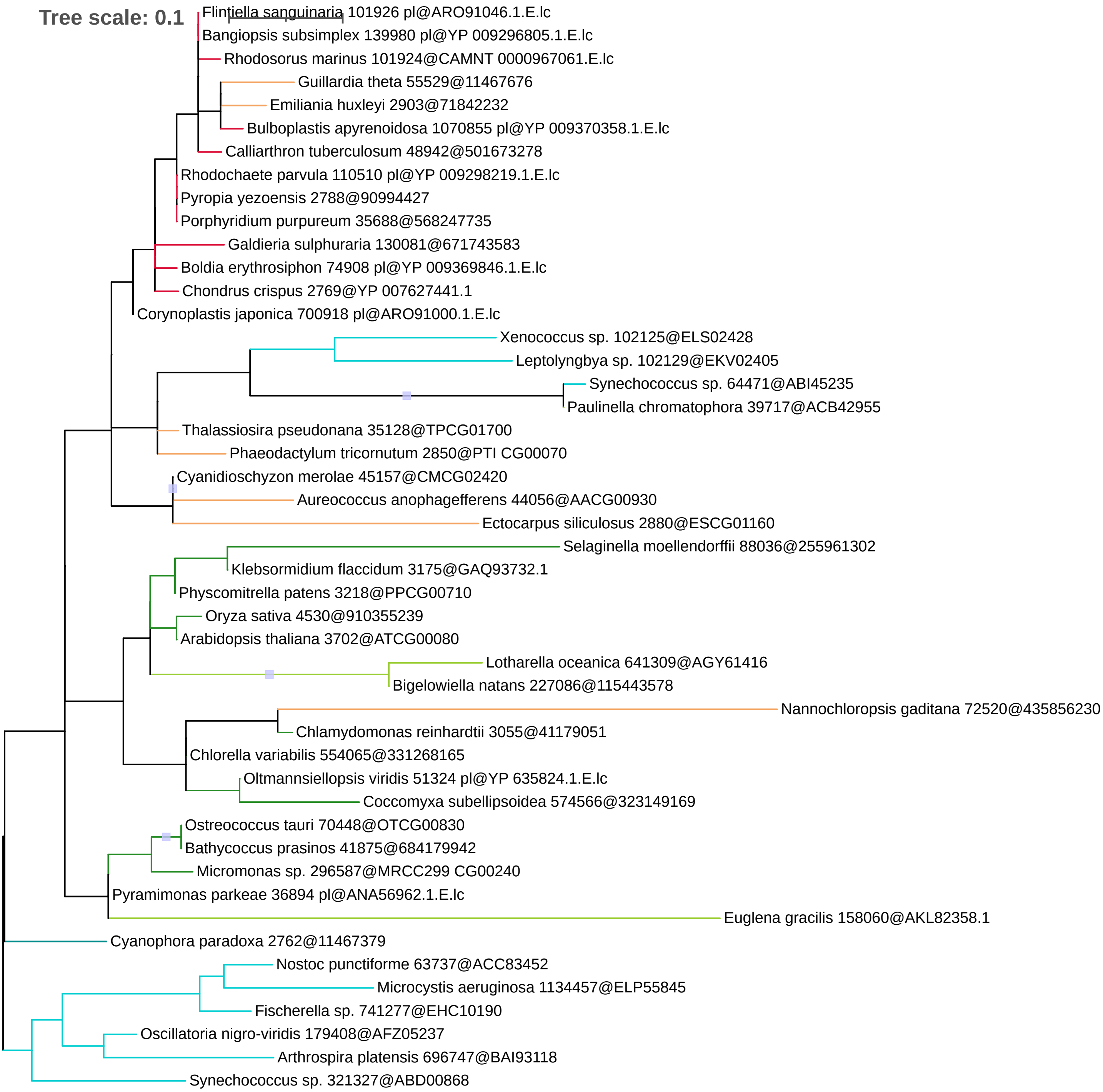


File: 06-PSBF.pdf

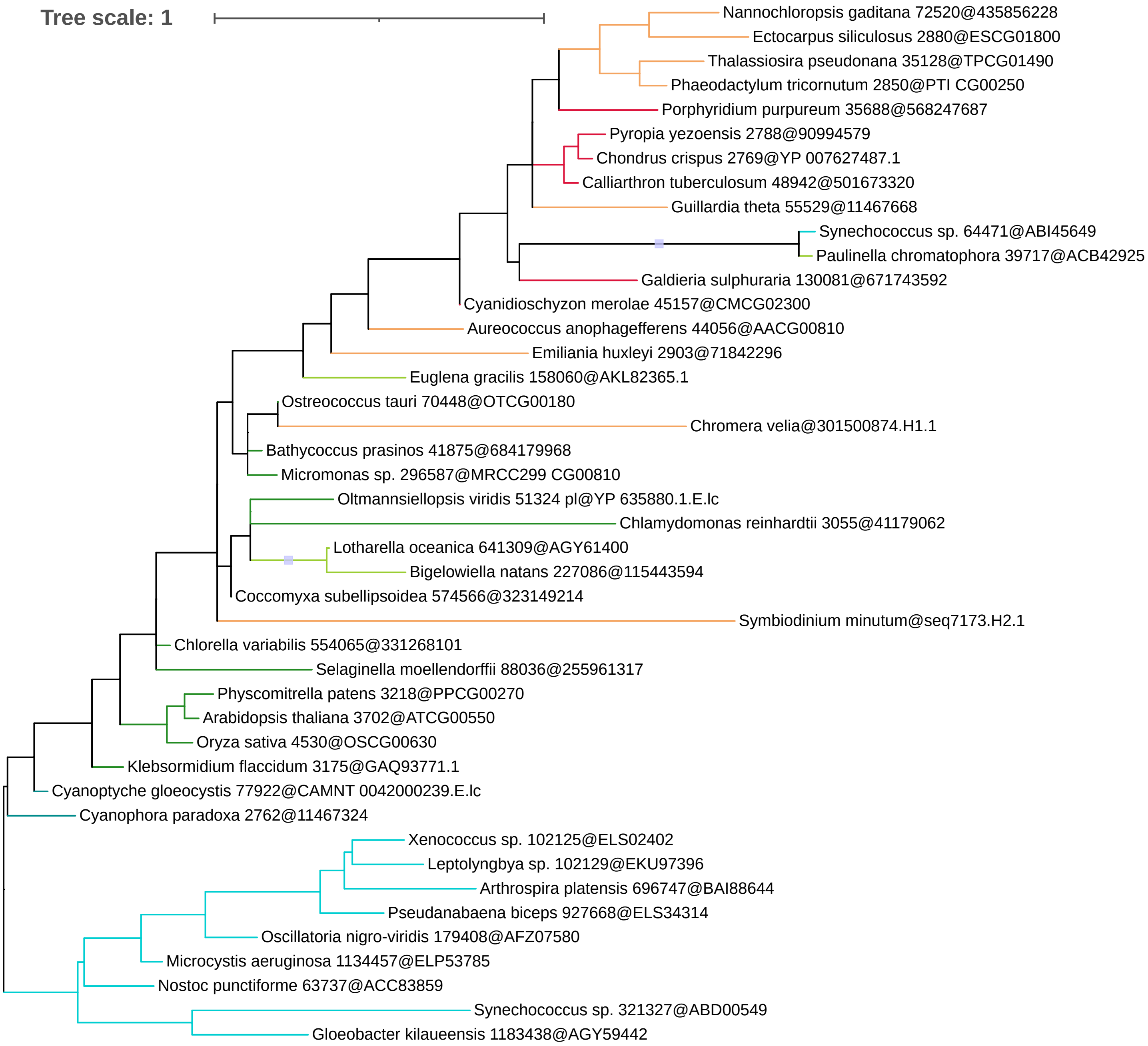
File: 07-PSBH.pdf



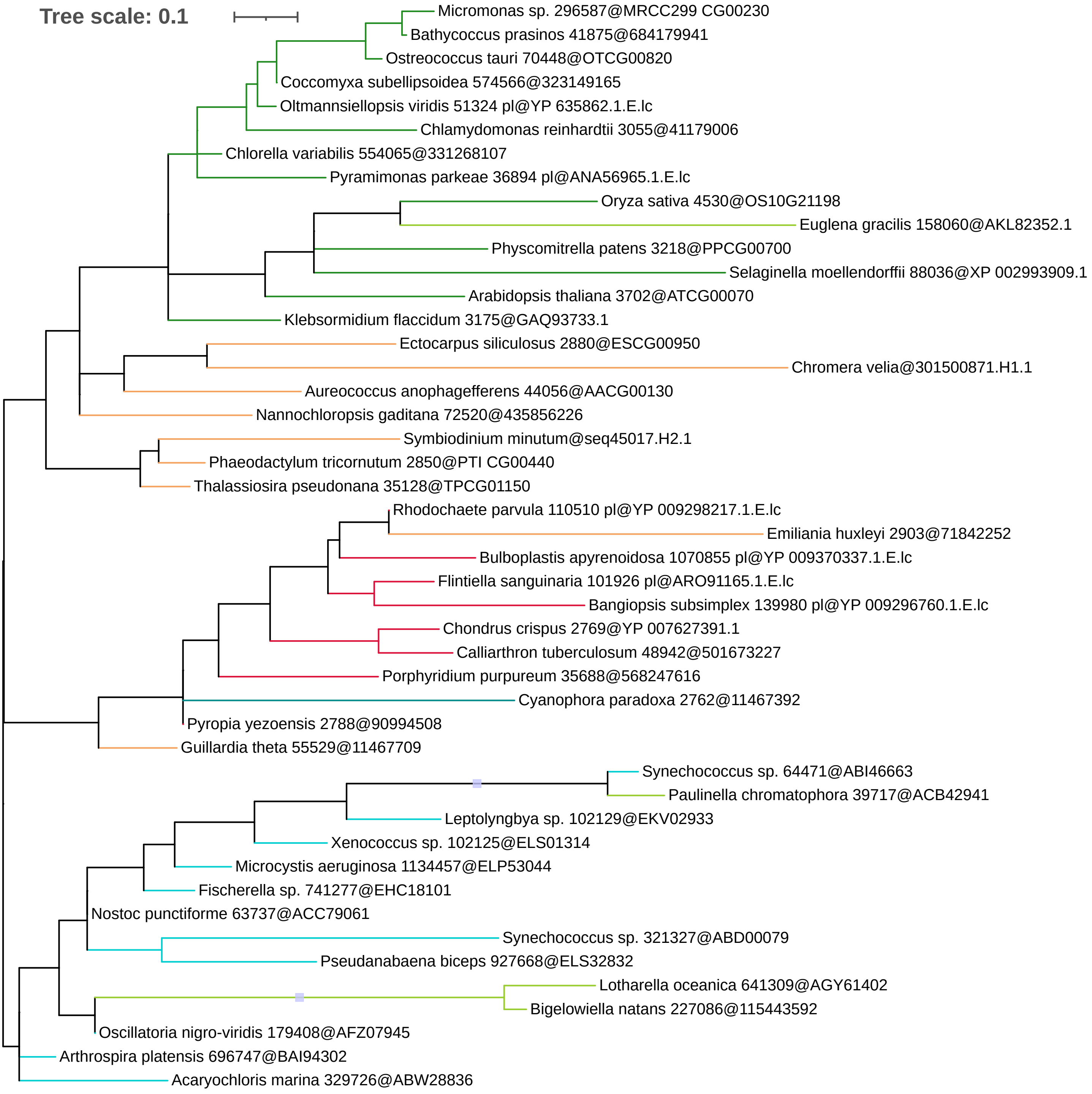
File: 08-PSBI.pdf



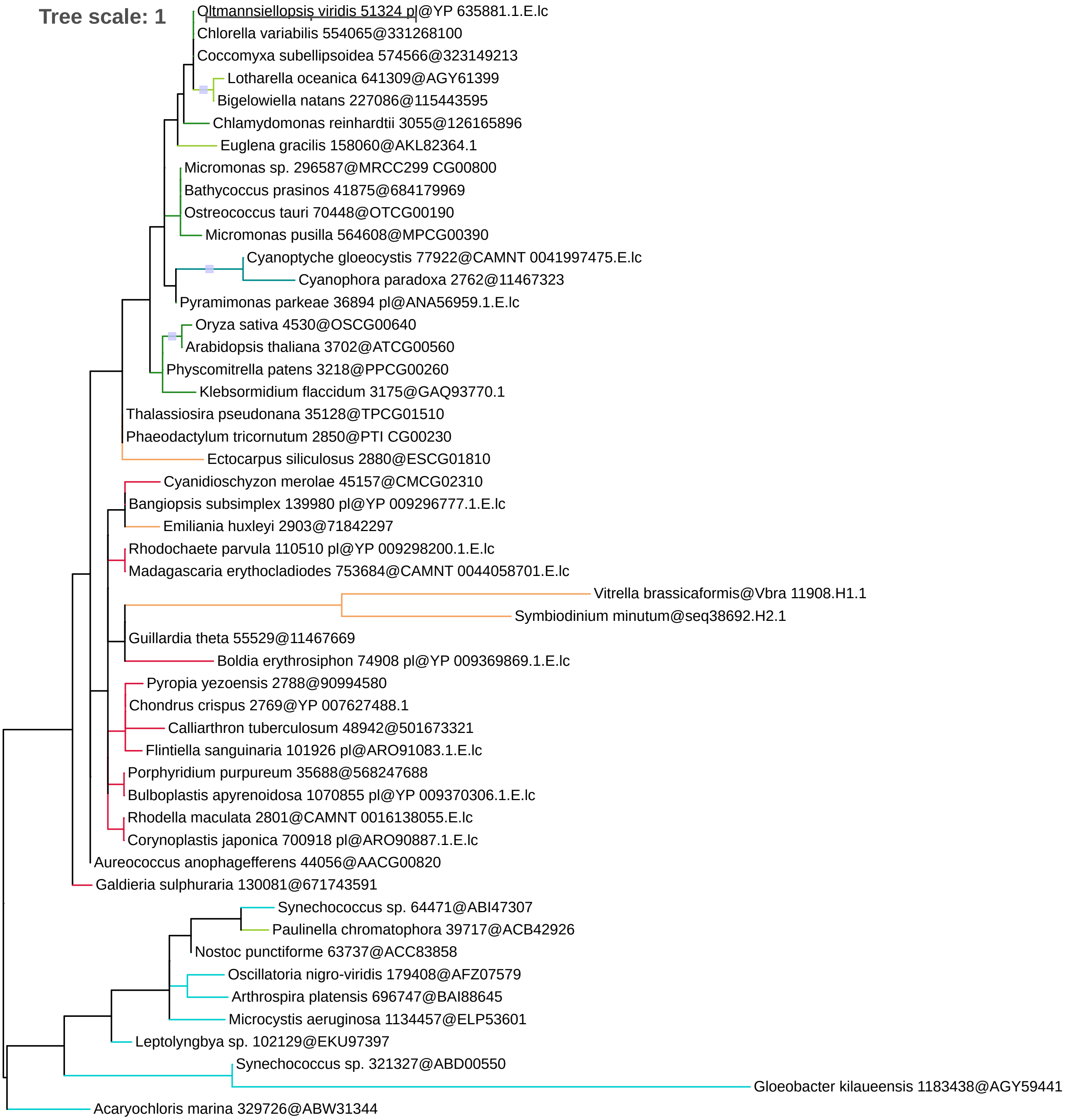
File: 09-PSBJ.pdf



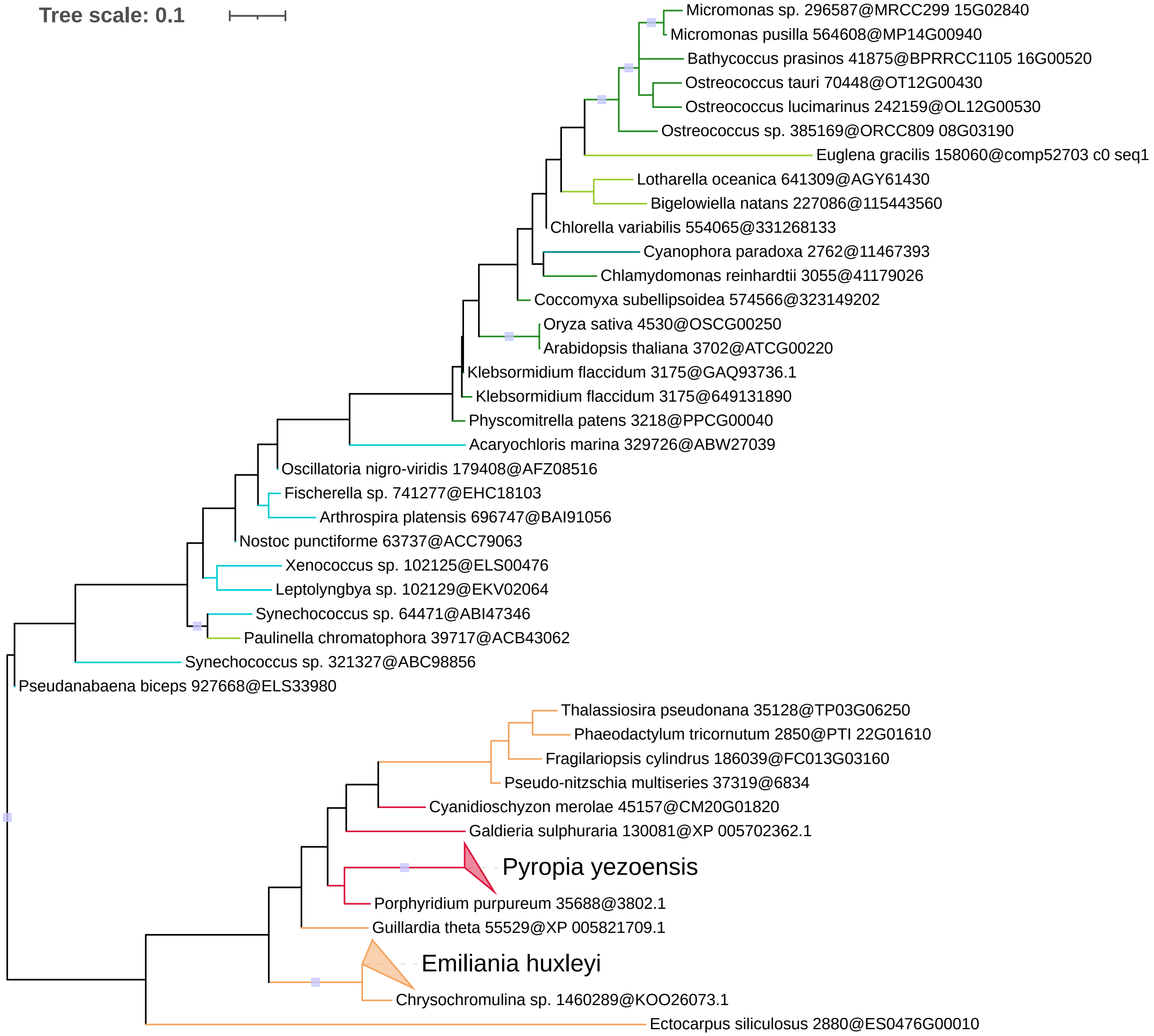
File: 10-PSBK.pdf



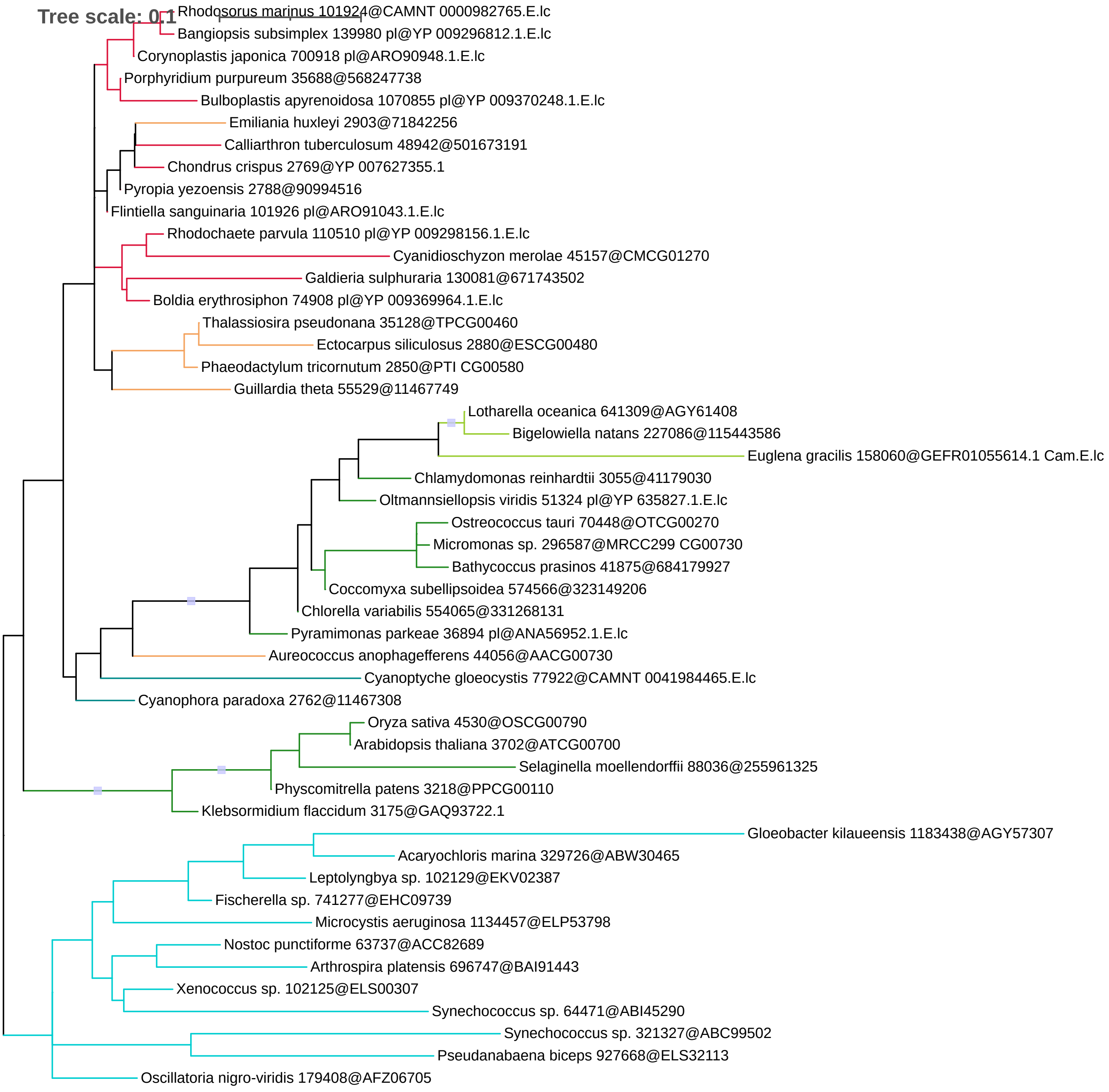
File: 11-PSBL.pdf



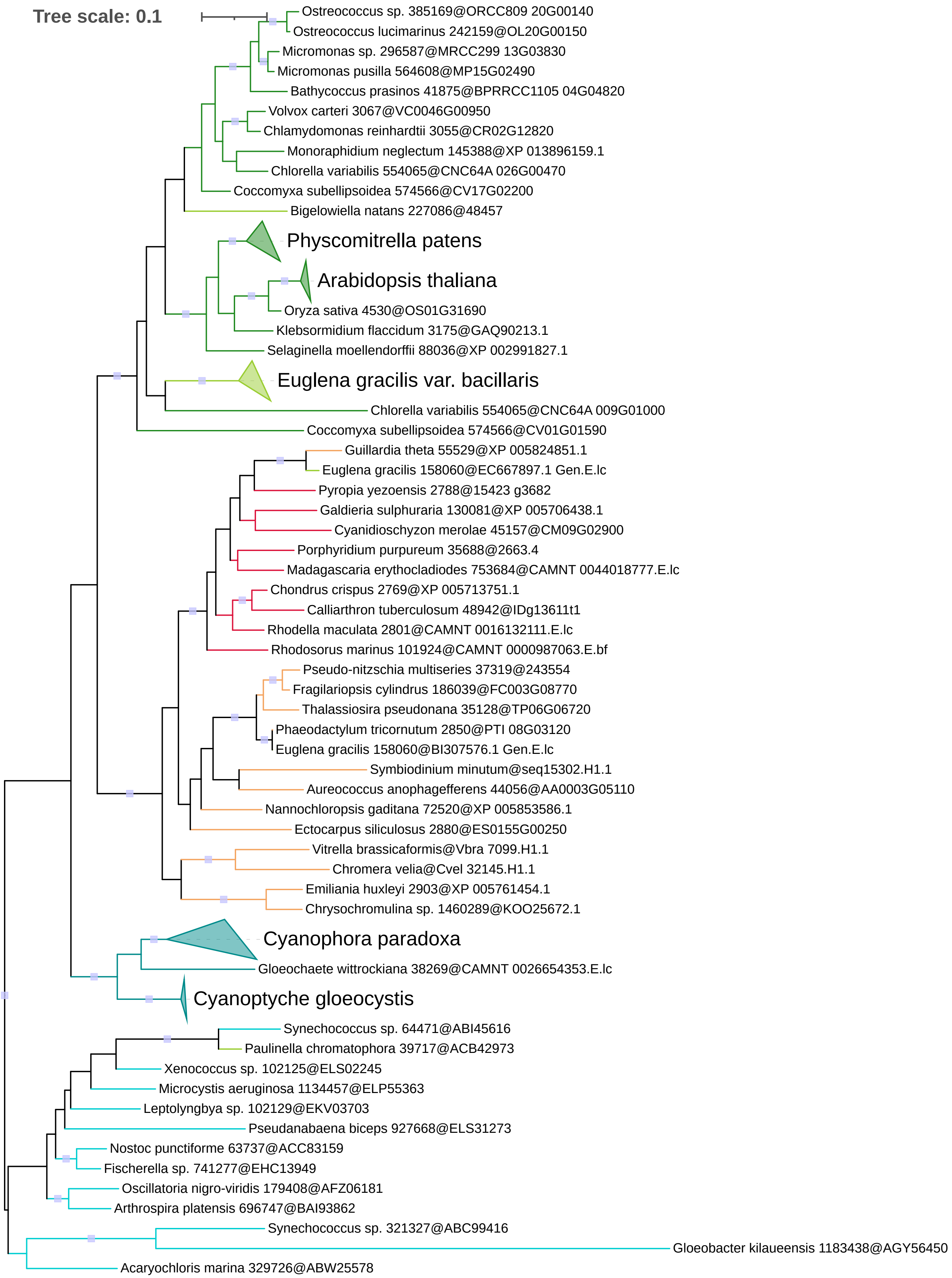
File: 12-PSBM.pdf



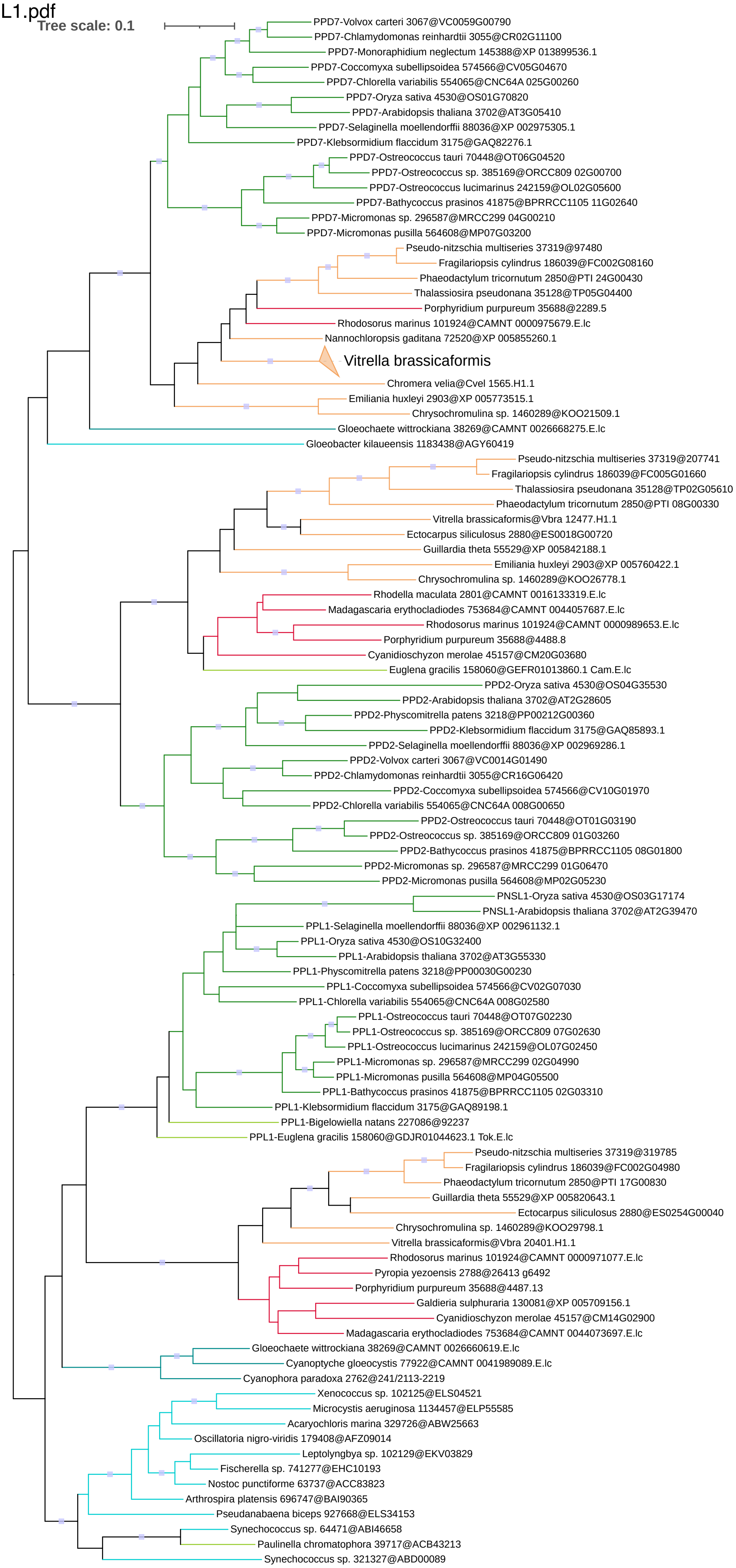
File: 13-PSBN.pdf



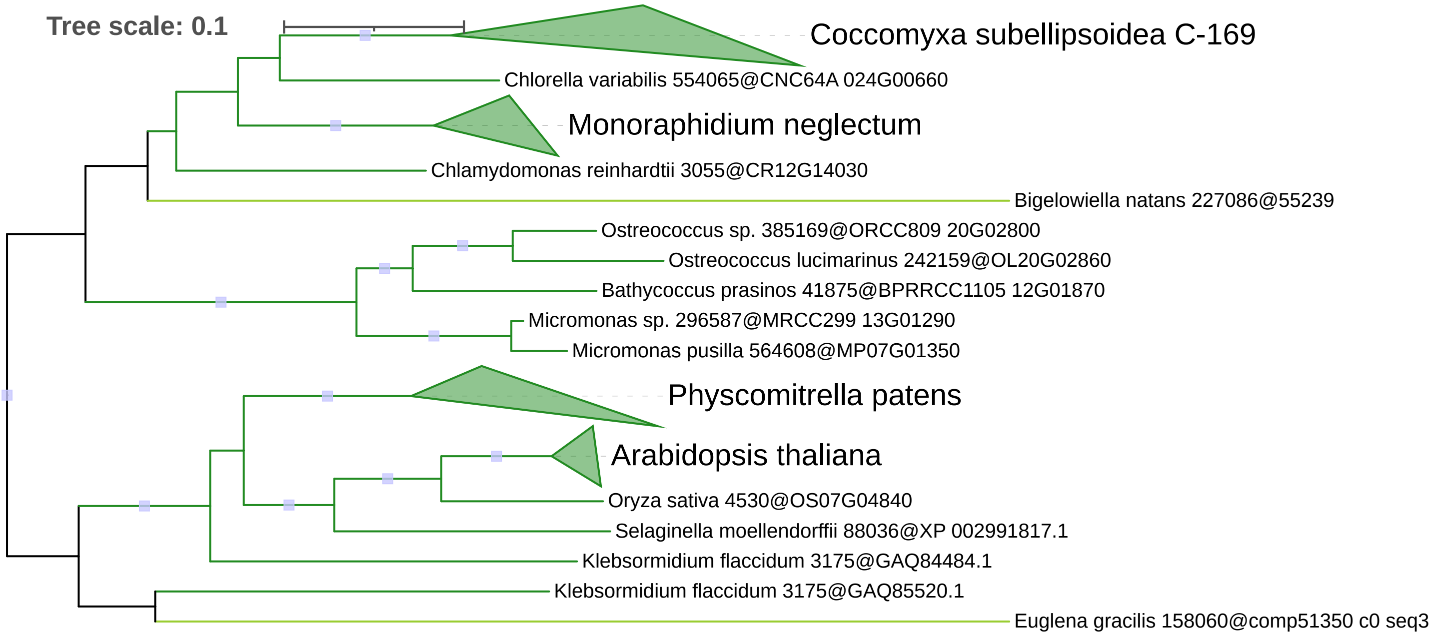
File: 14-PSBO.pdf



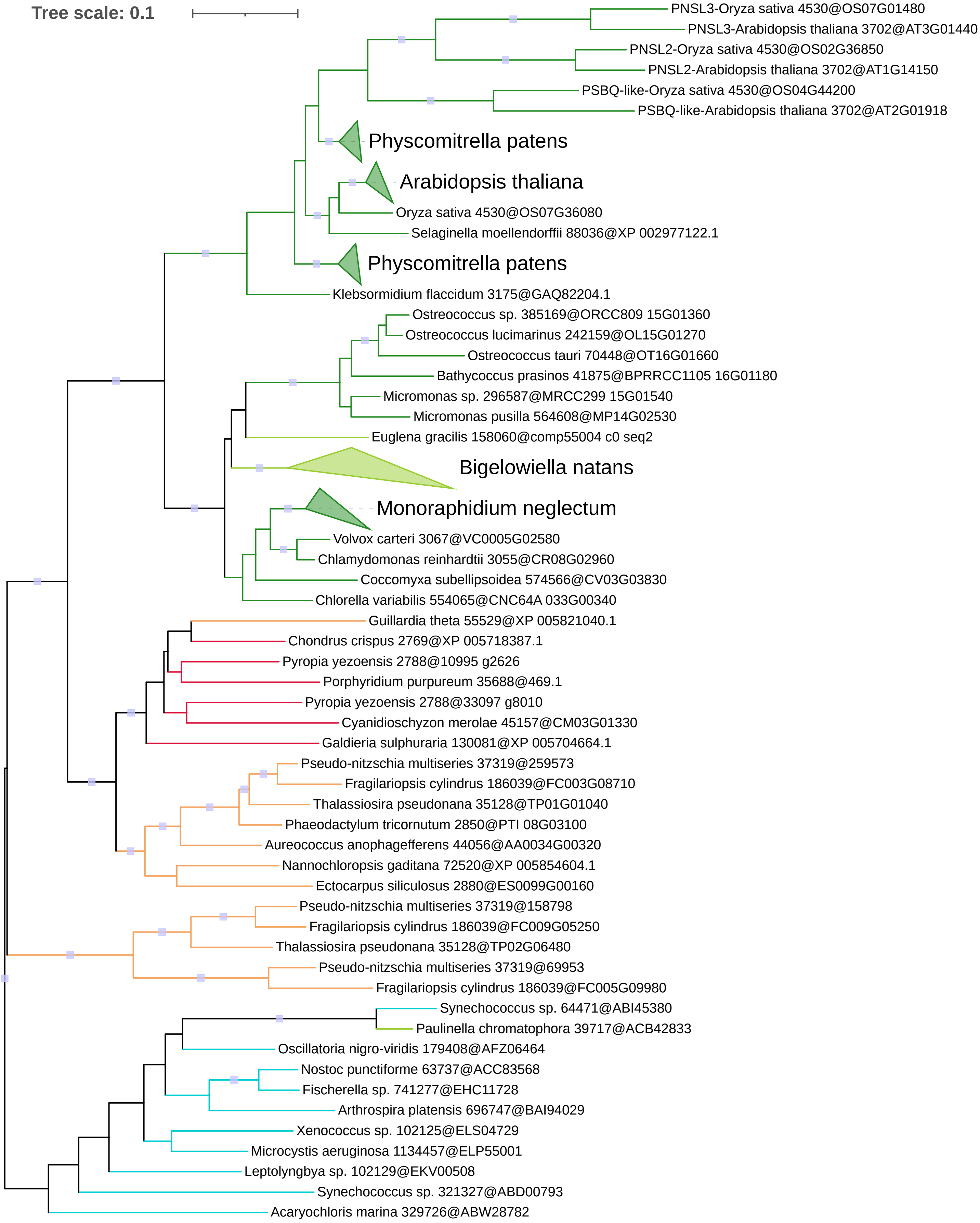
File: 15-PSBP-PPL1.pdf



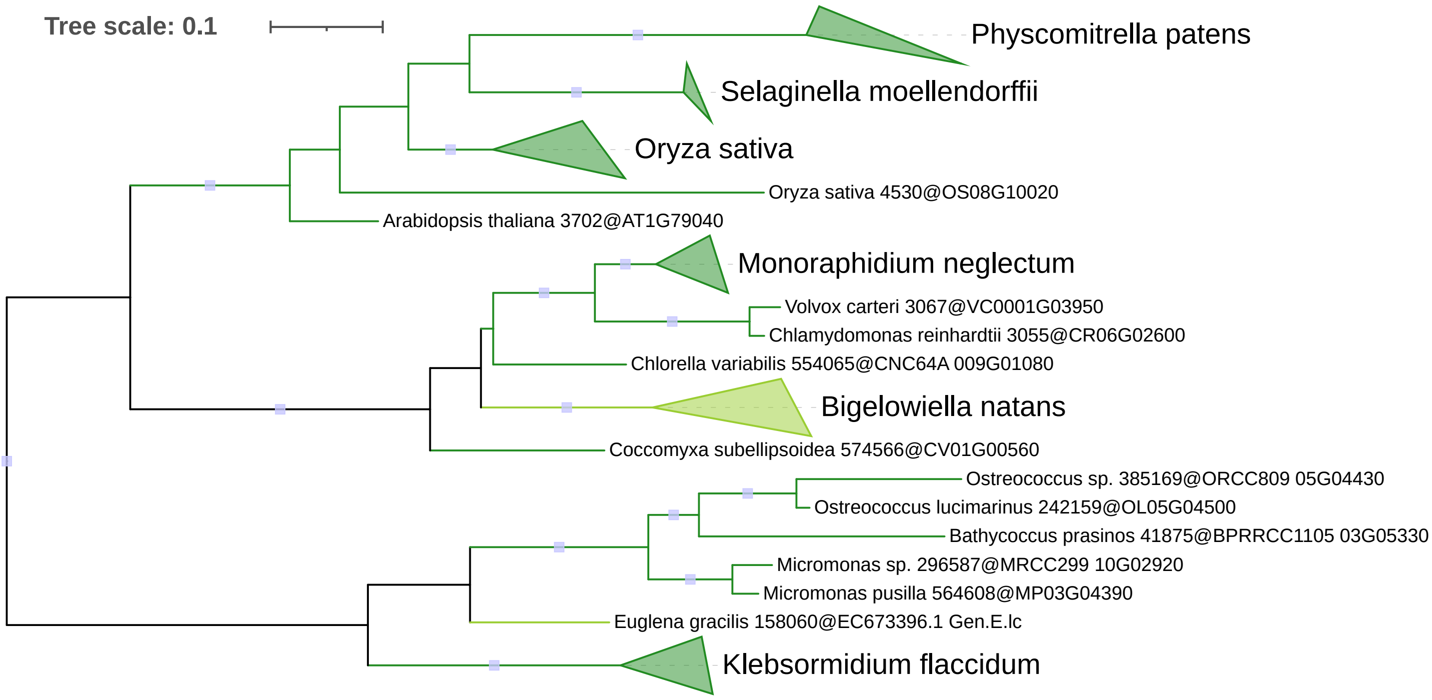
File: 16-PSBP2.pdf



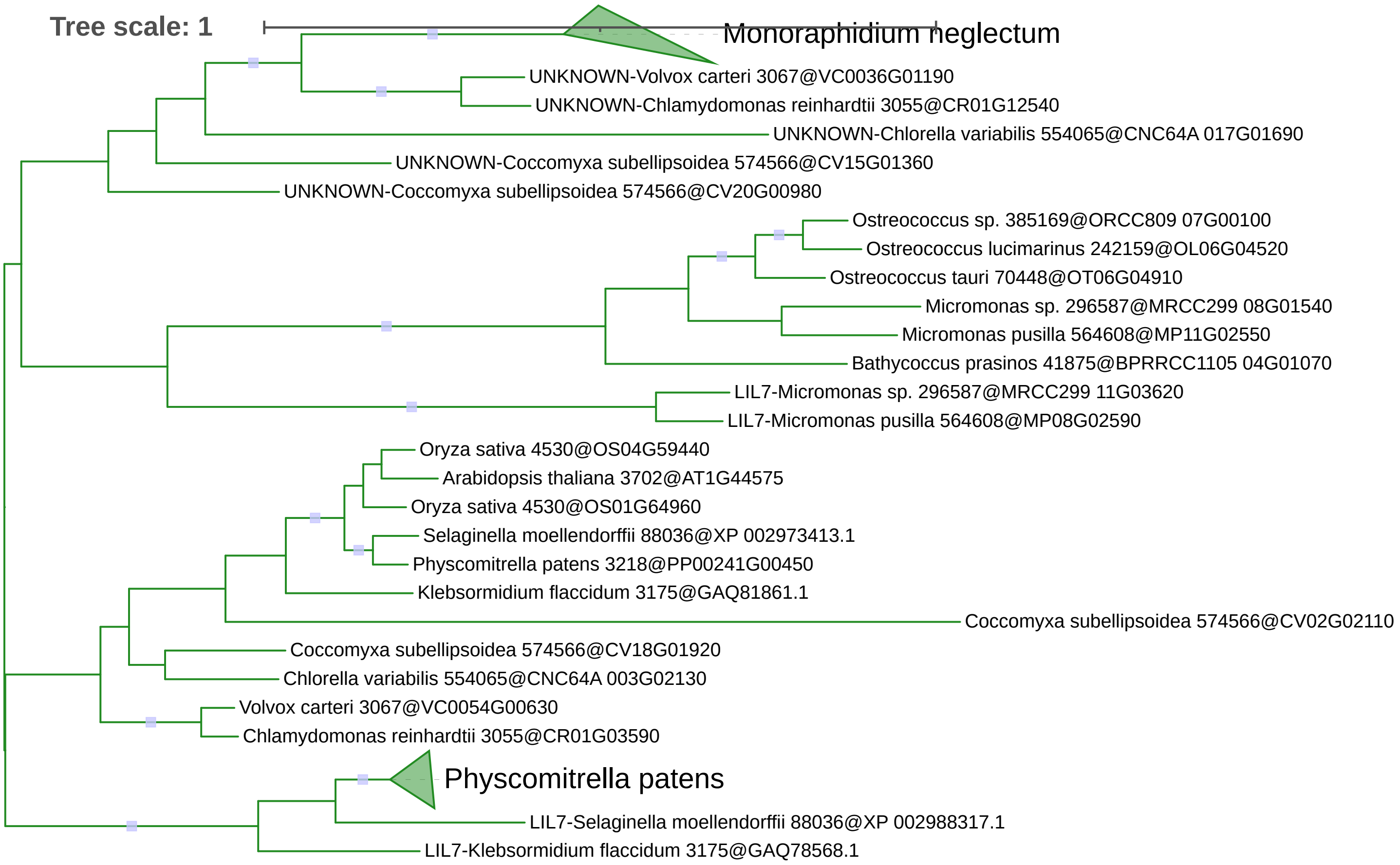
File: 17-PNSL23-PSBQ.pdf



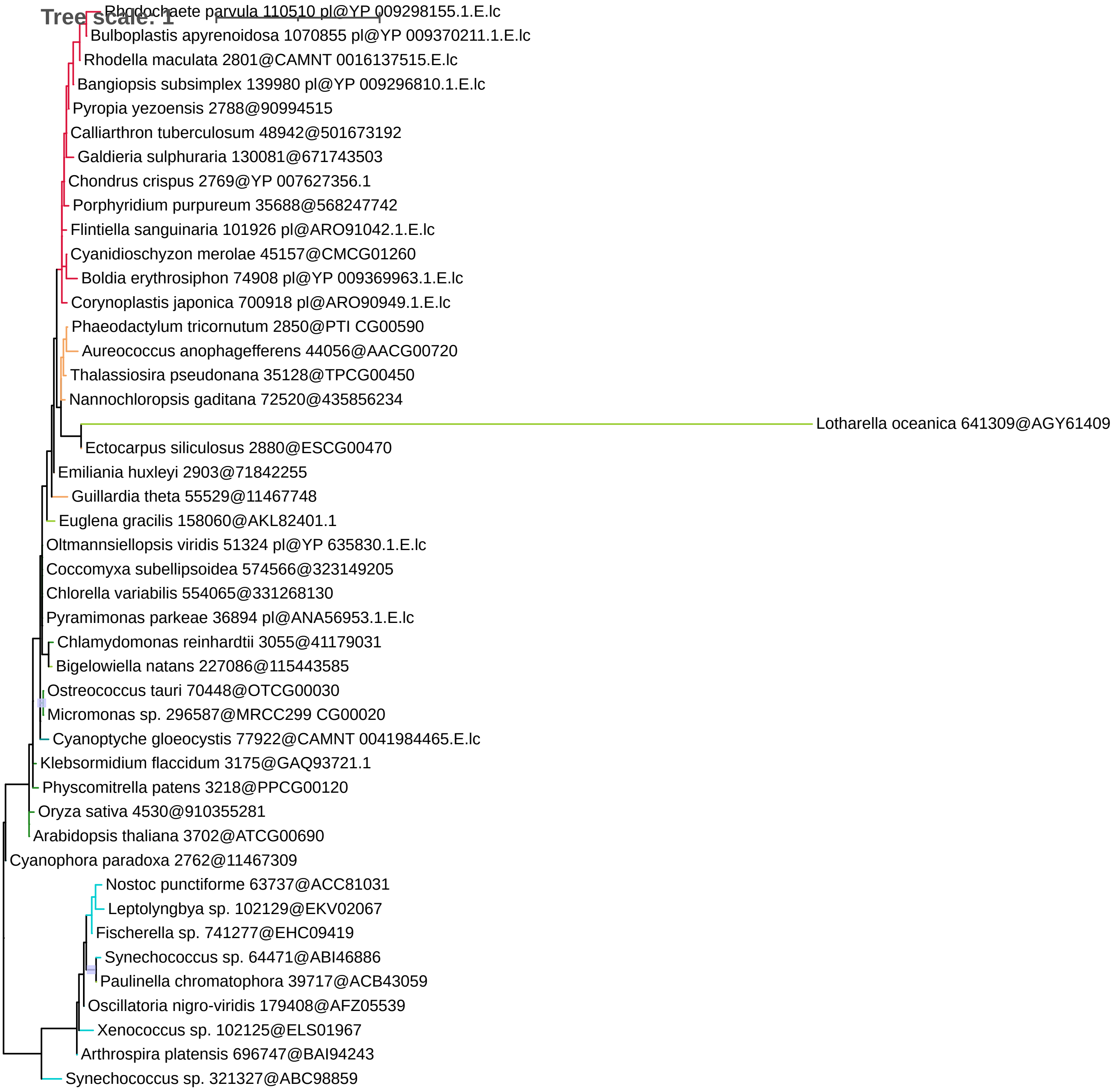
File: 18-PSBR.pdf



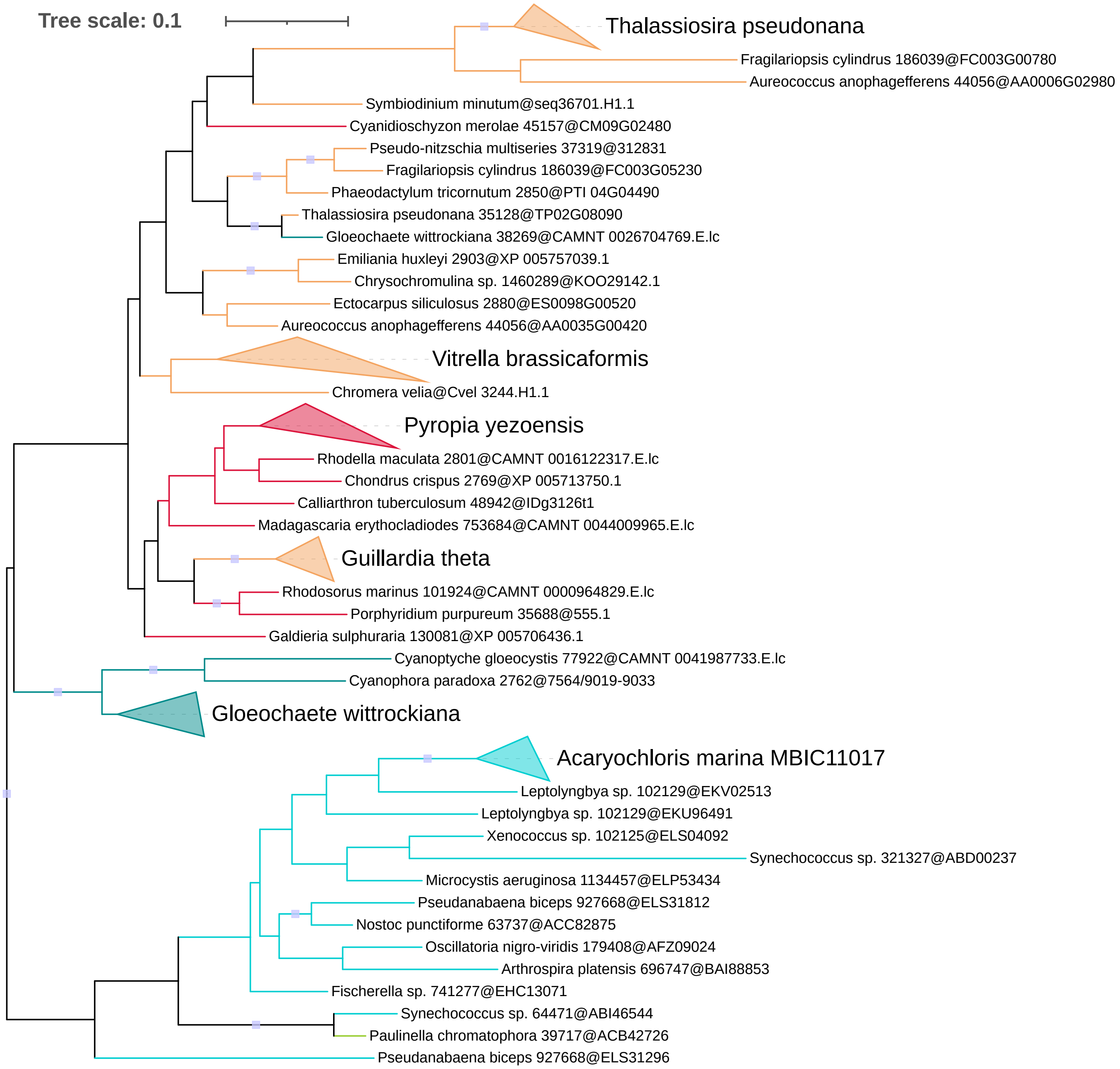
File: 19-PSBS.pdf



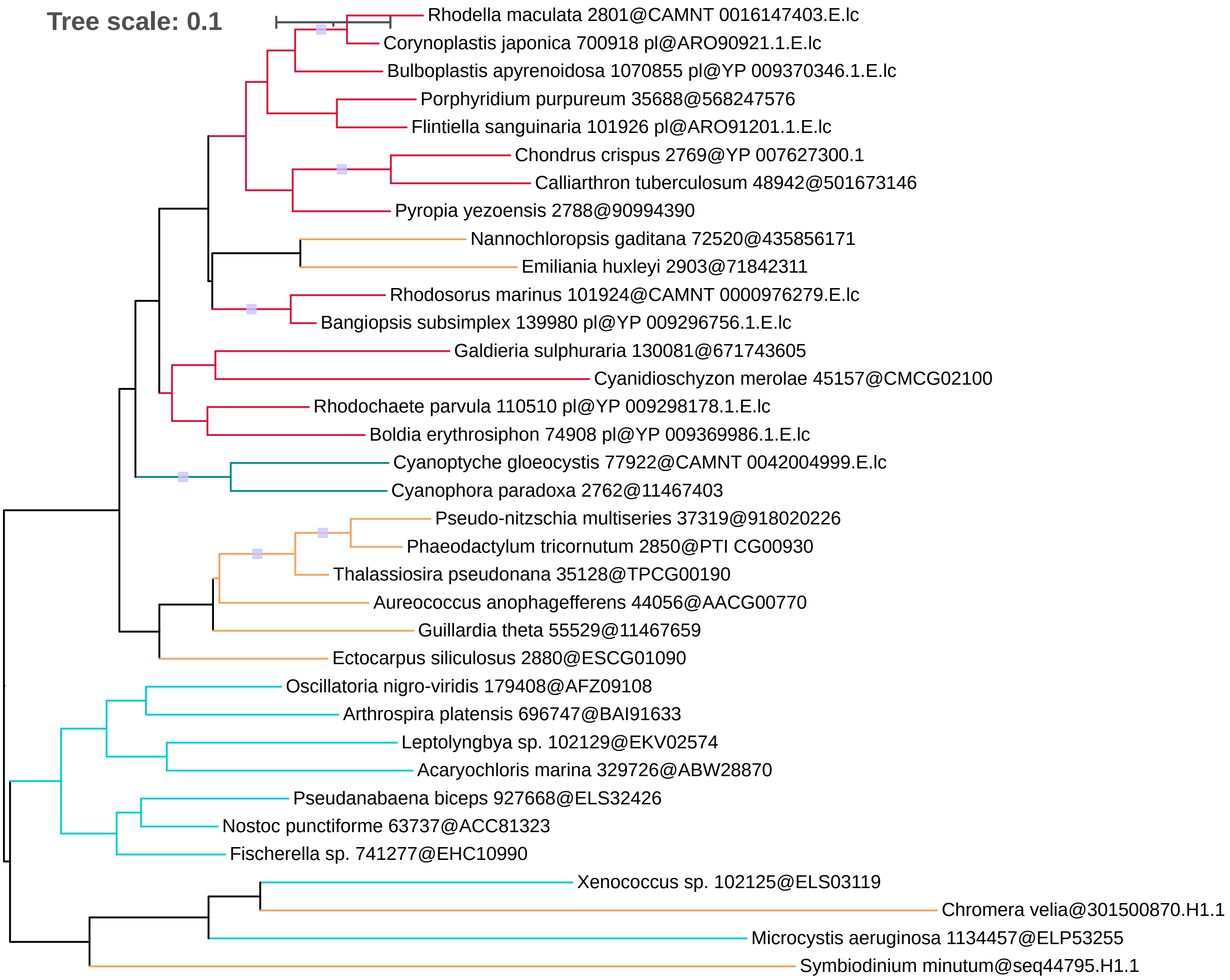
File: 20-PSBT.pdf



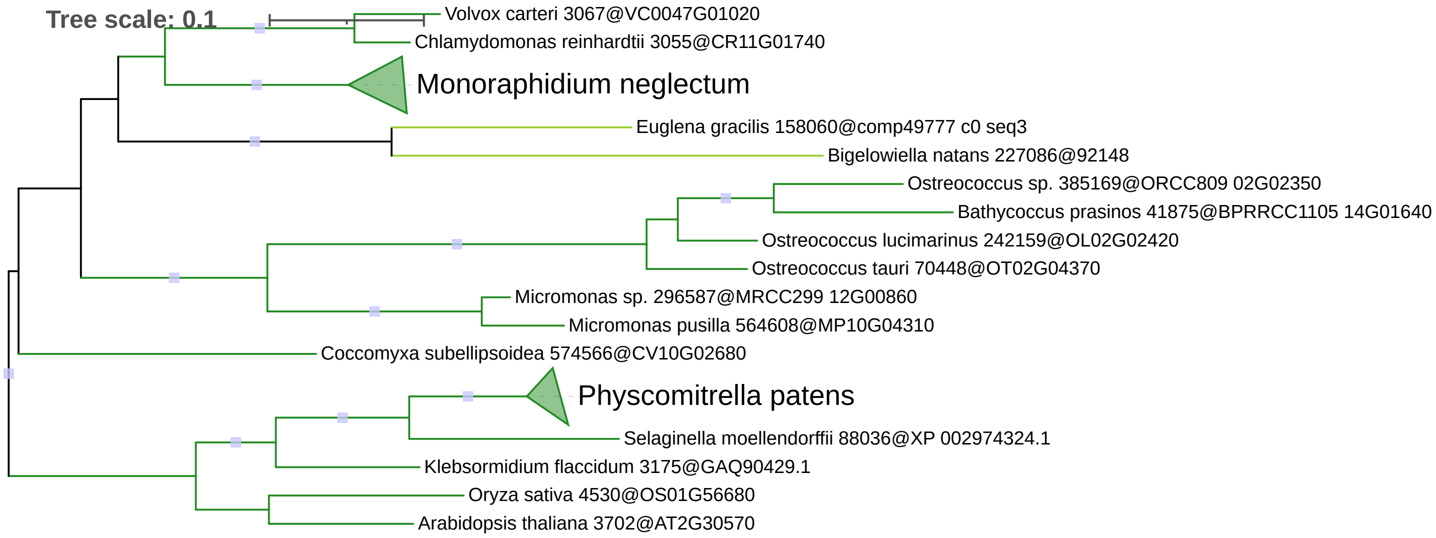
File: 21-PSBU.pdf



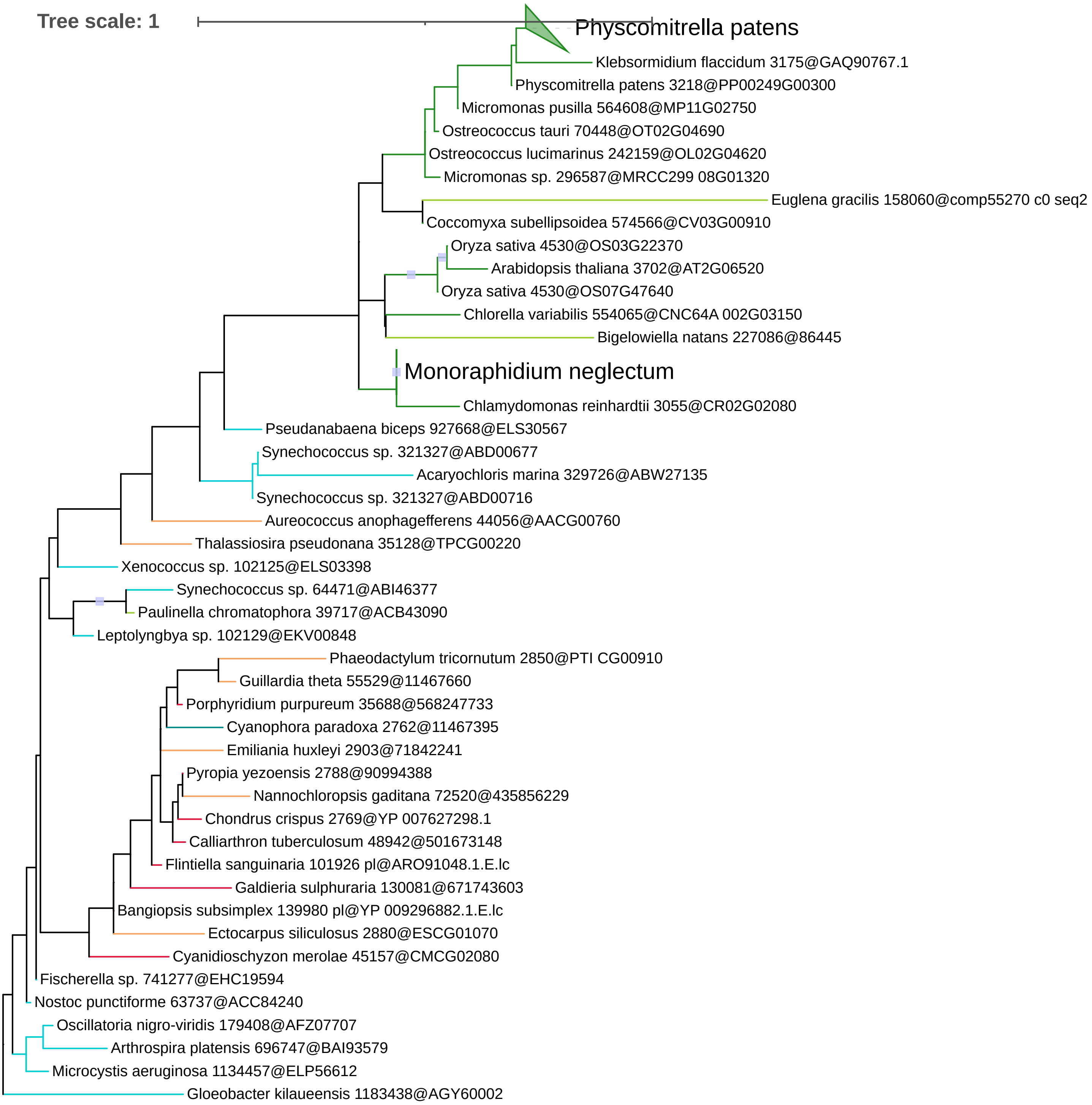
File: 22-PSBV.pdf



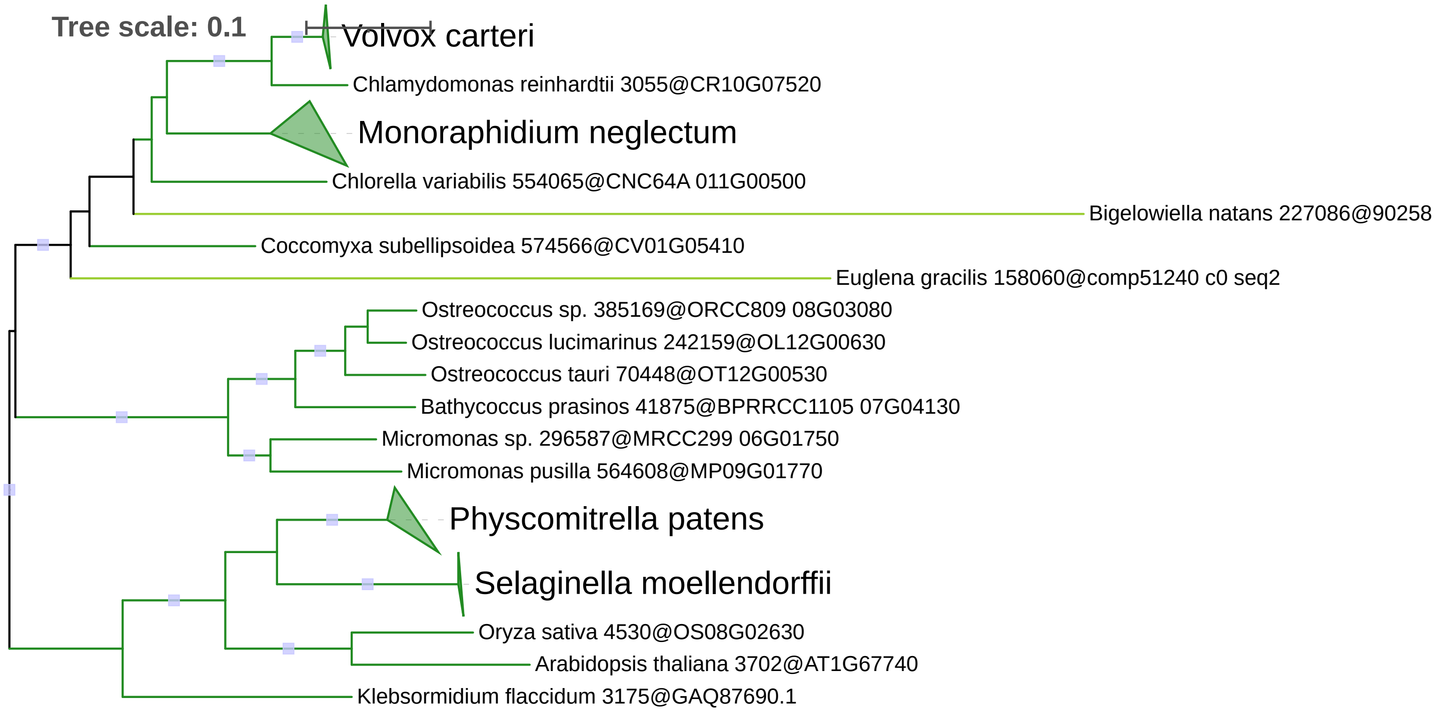
File: 23-PSBW.pdf



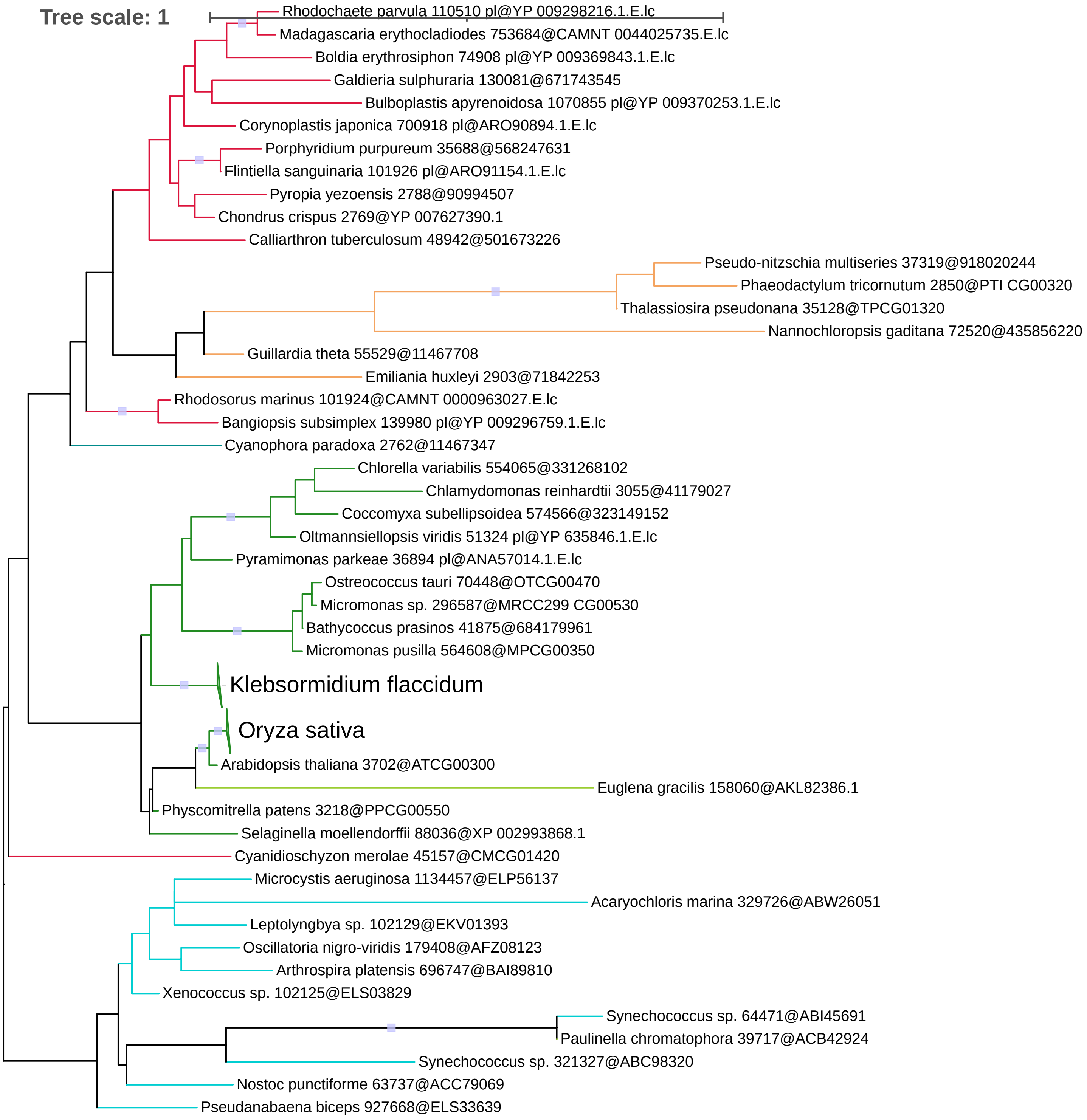
File: 24-PSBX.pdf



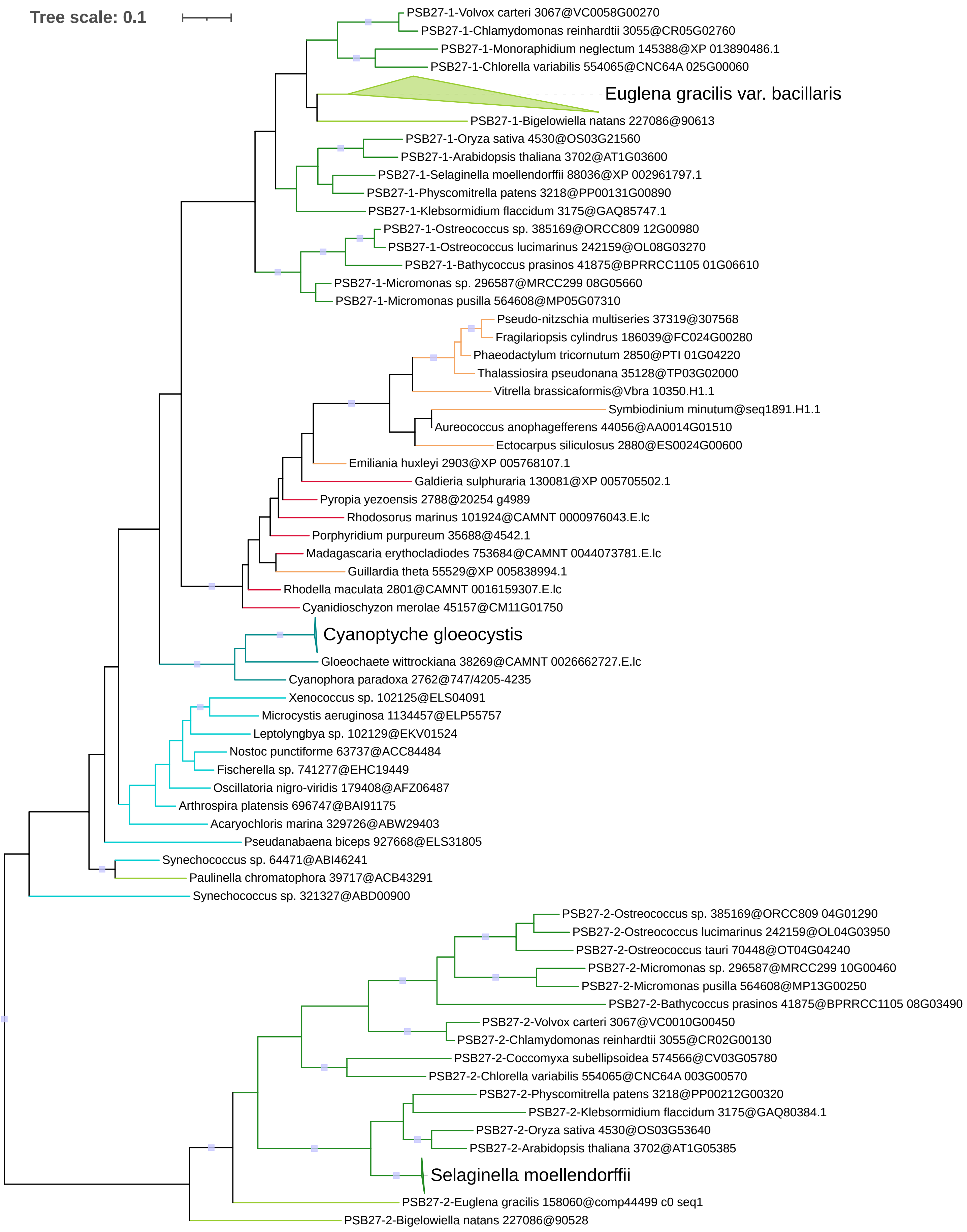
File: 25-PSBY.pdf



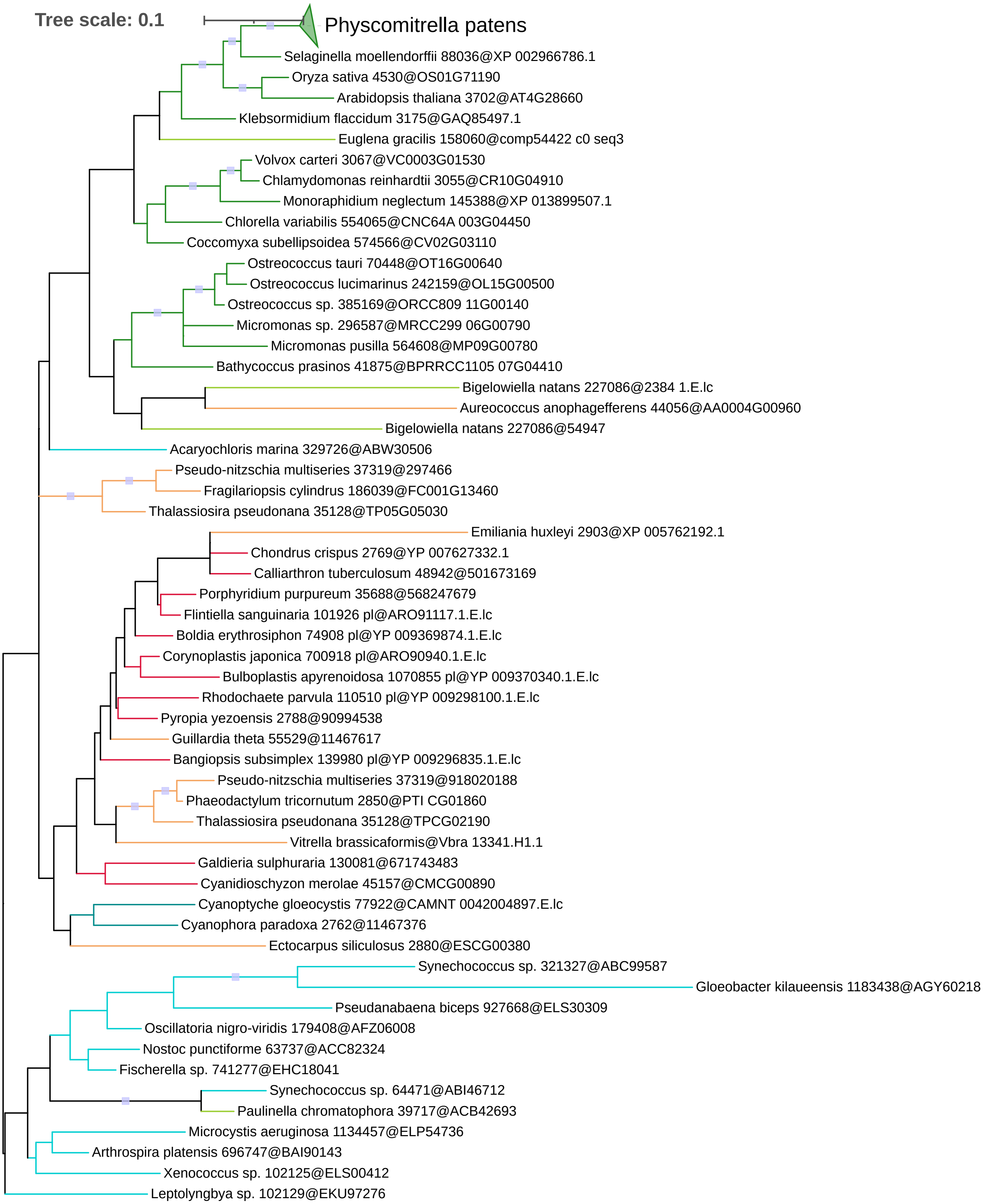
File: 26-PSBZ.pdf



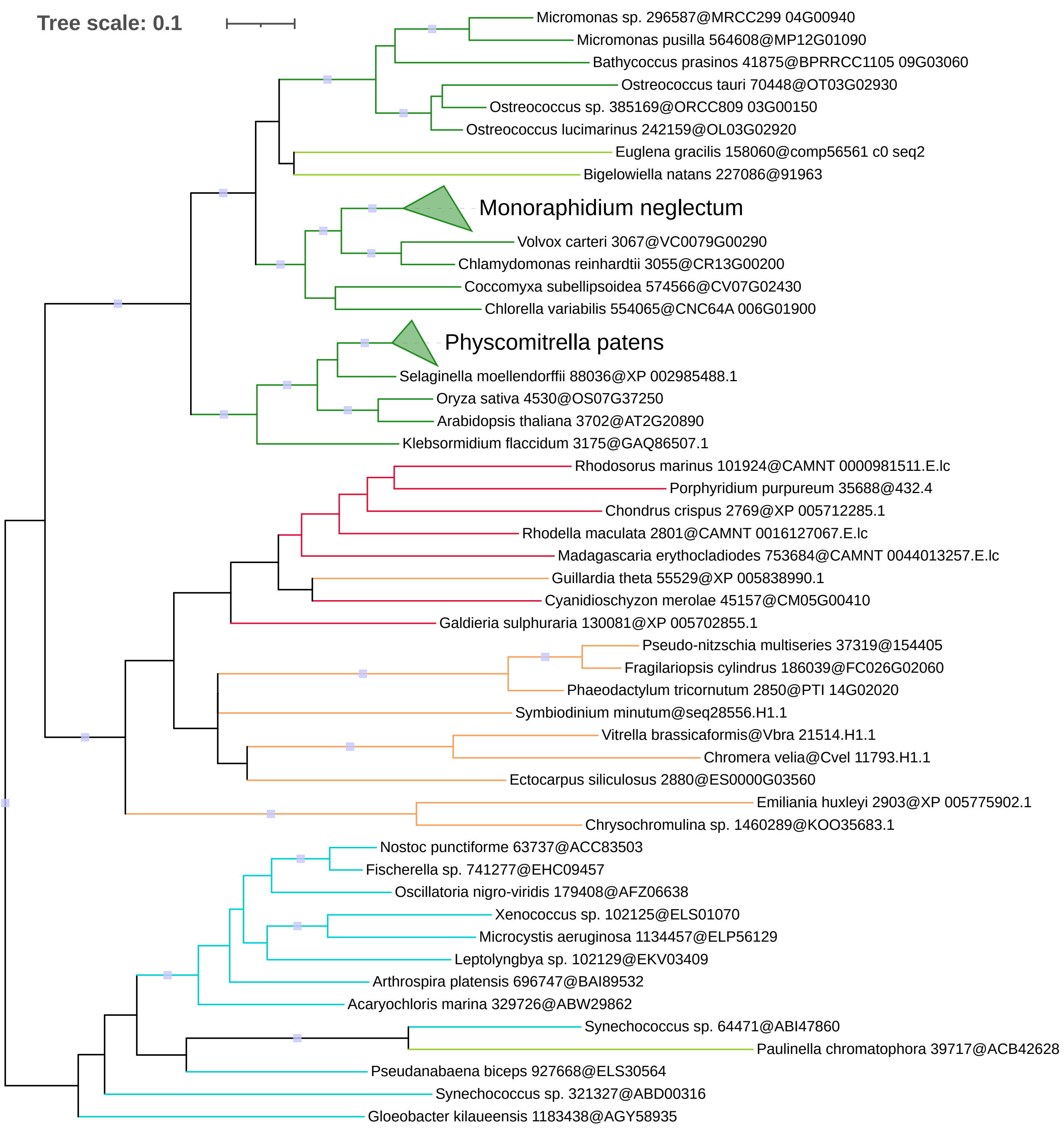
File: 27-PSB27-1-2.pdf



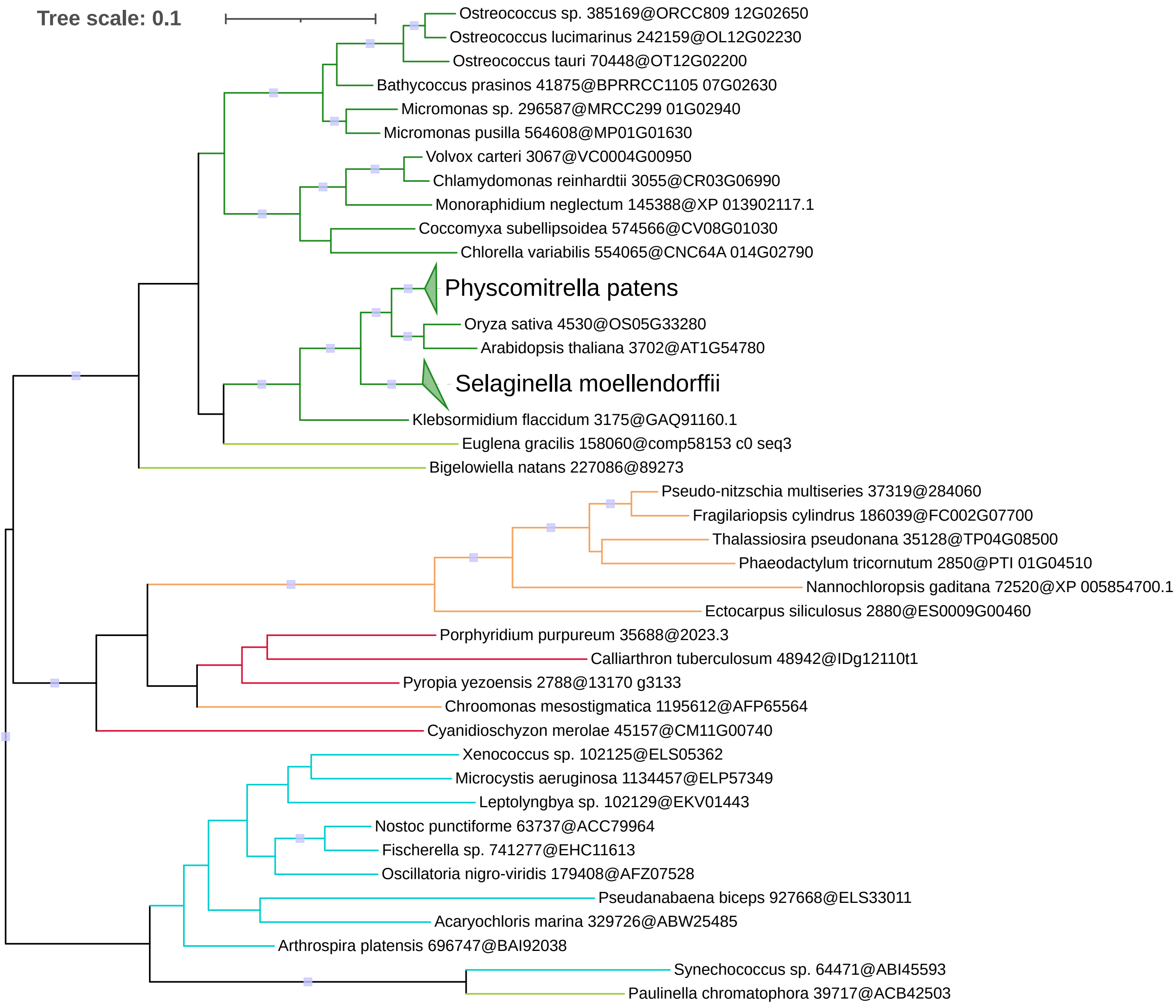
File: 28-PSB28.pdf



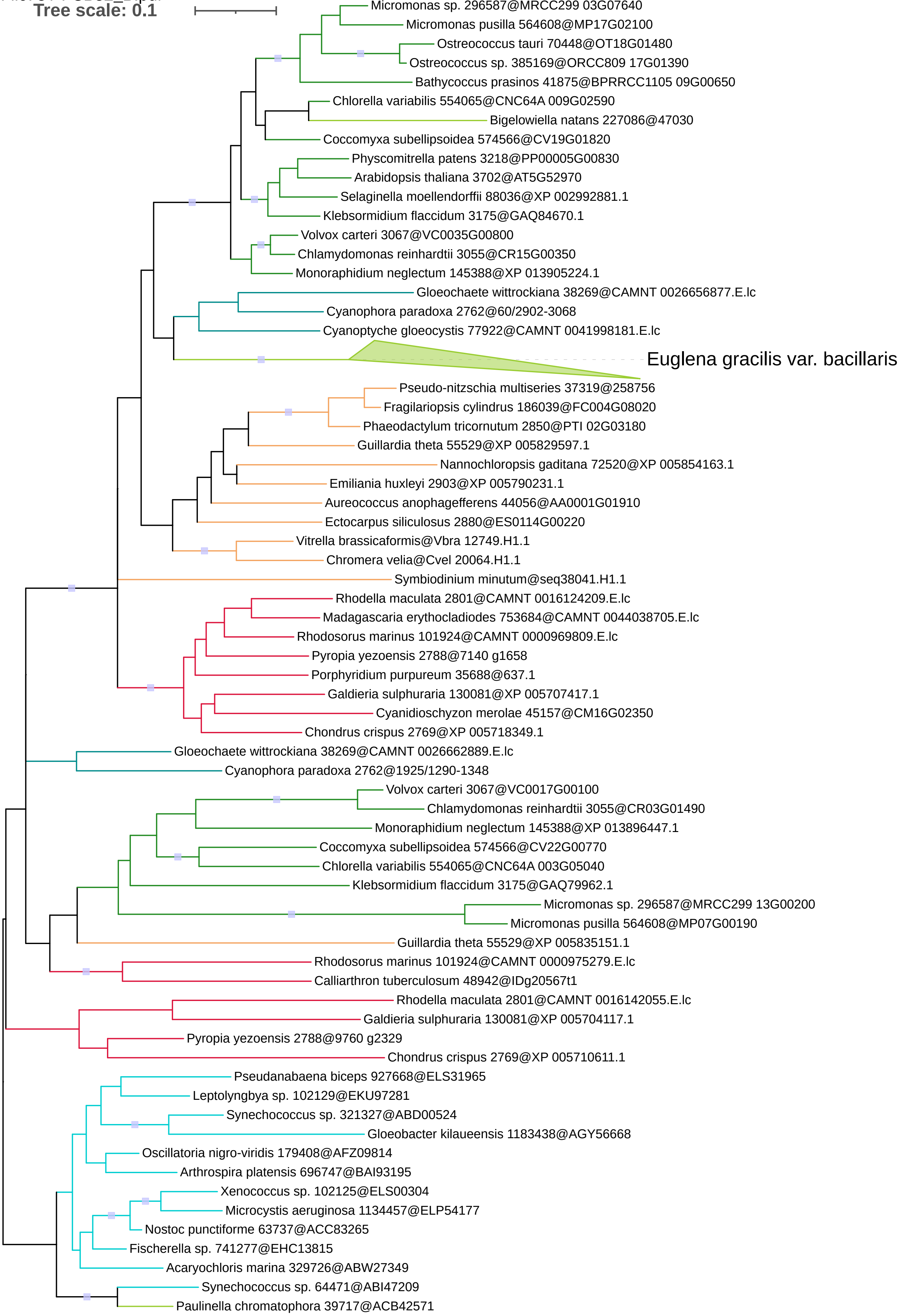
File: 29-PSB29.pdf



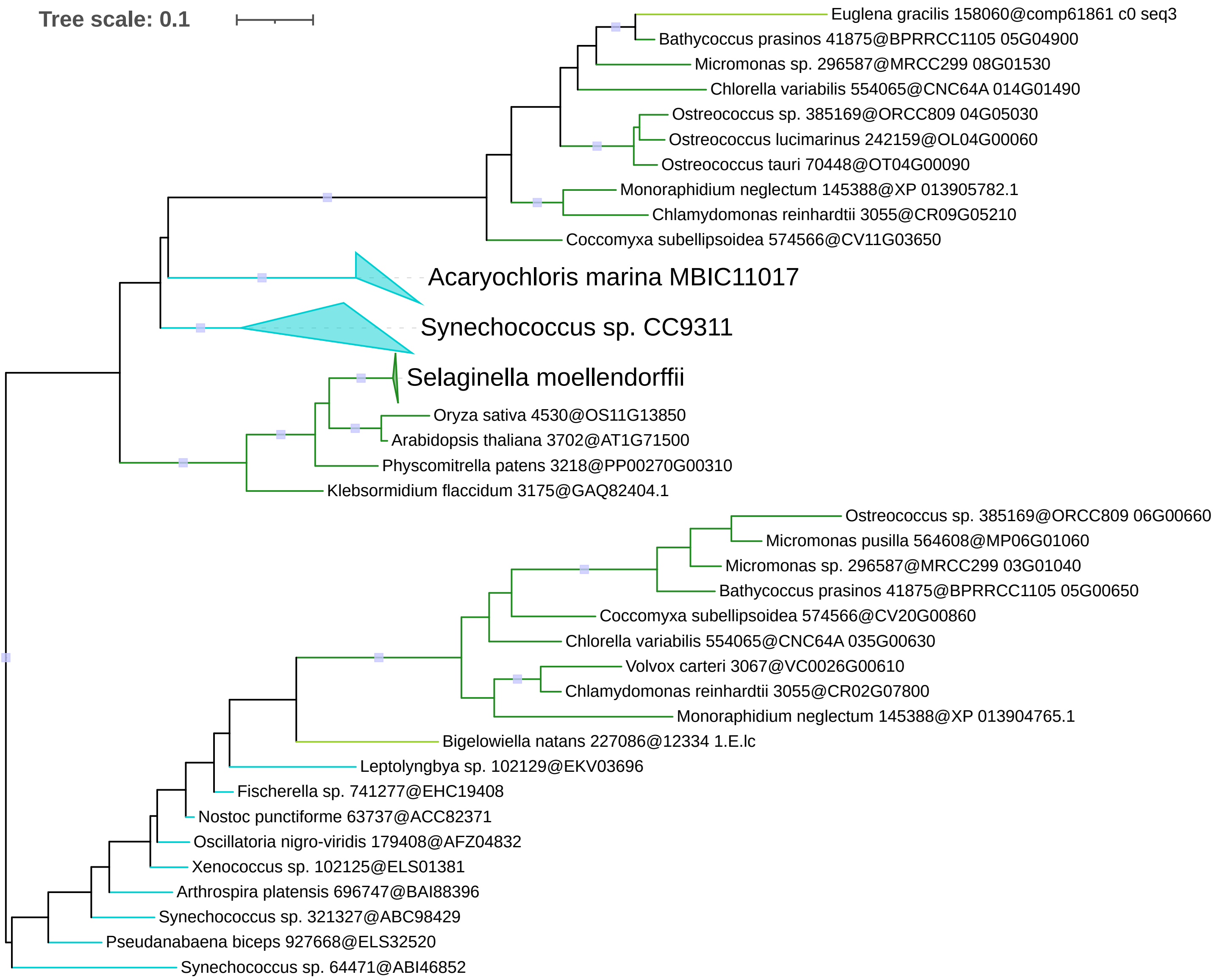
File: 30-PSB32_A.pdf



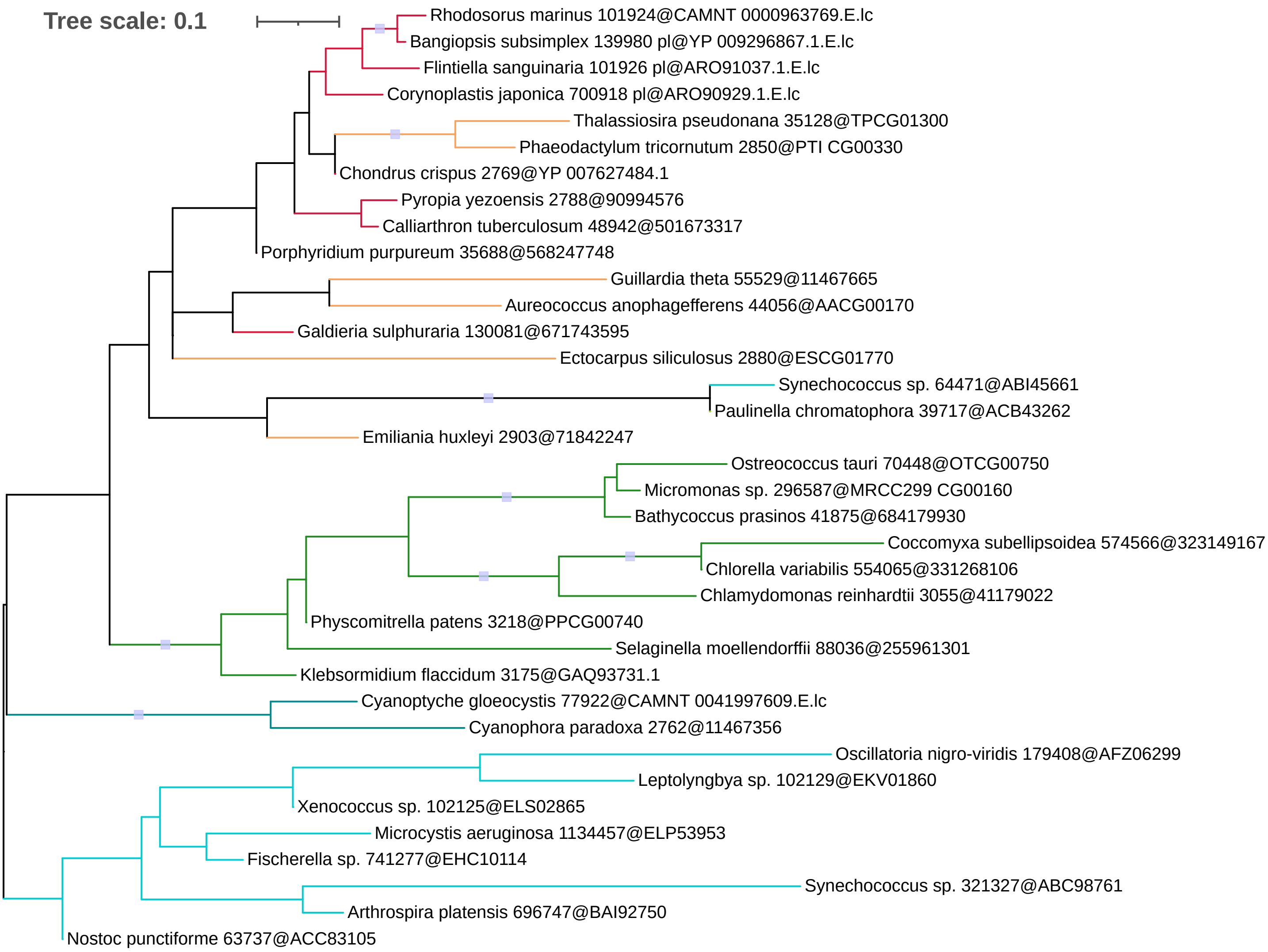
File: 31-PSB32_B.pdf



File: 32-PSB33.pdf

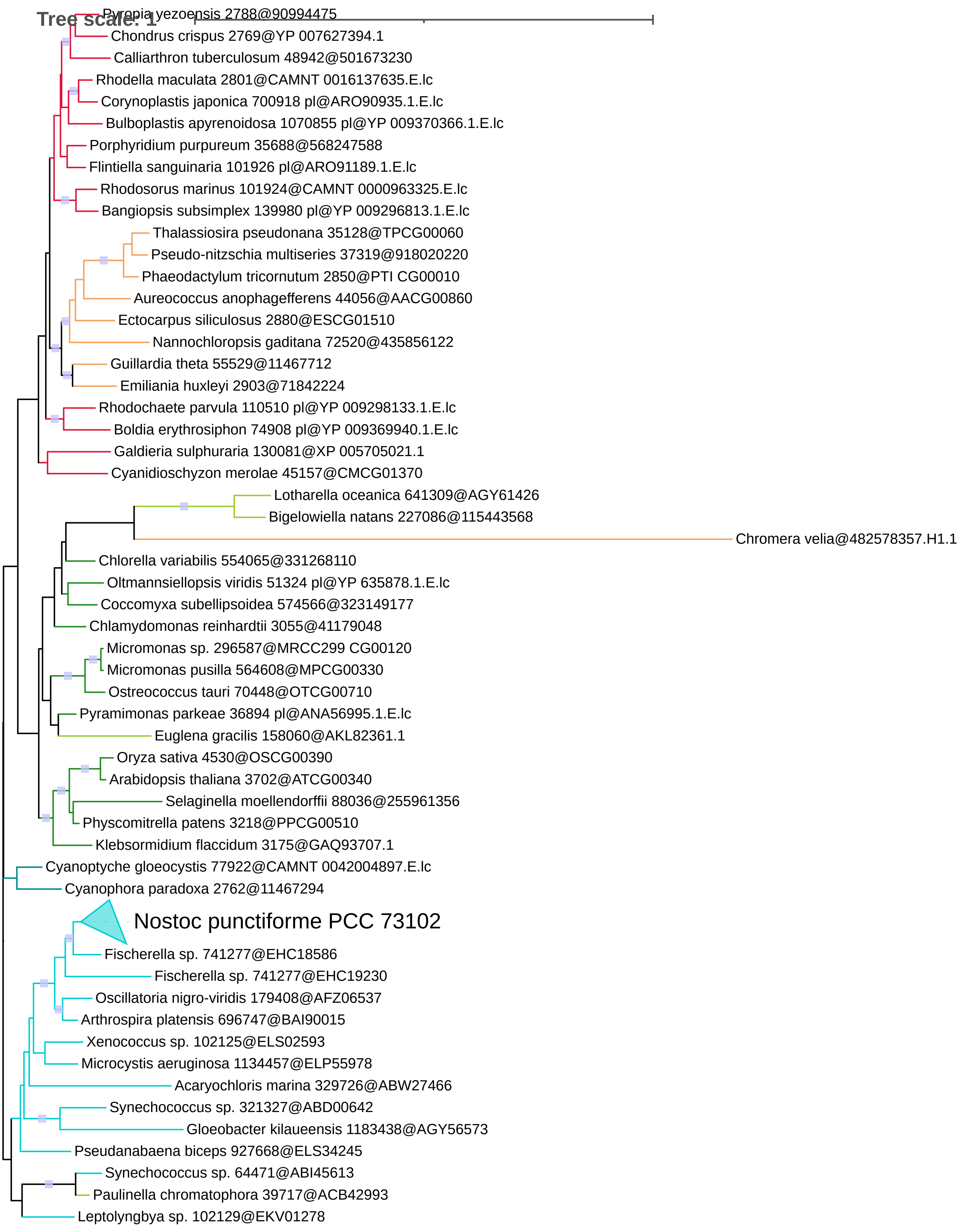


File: 33-YCF12.pdf

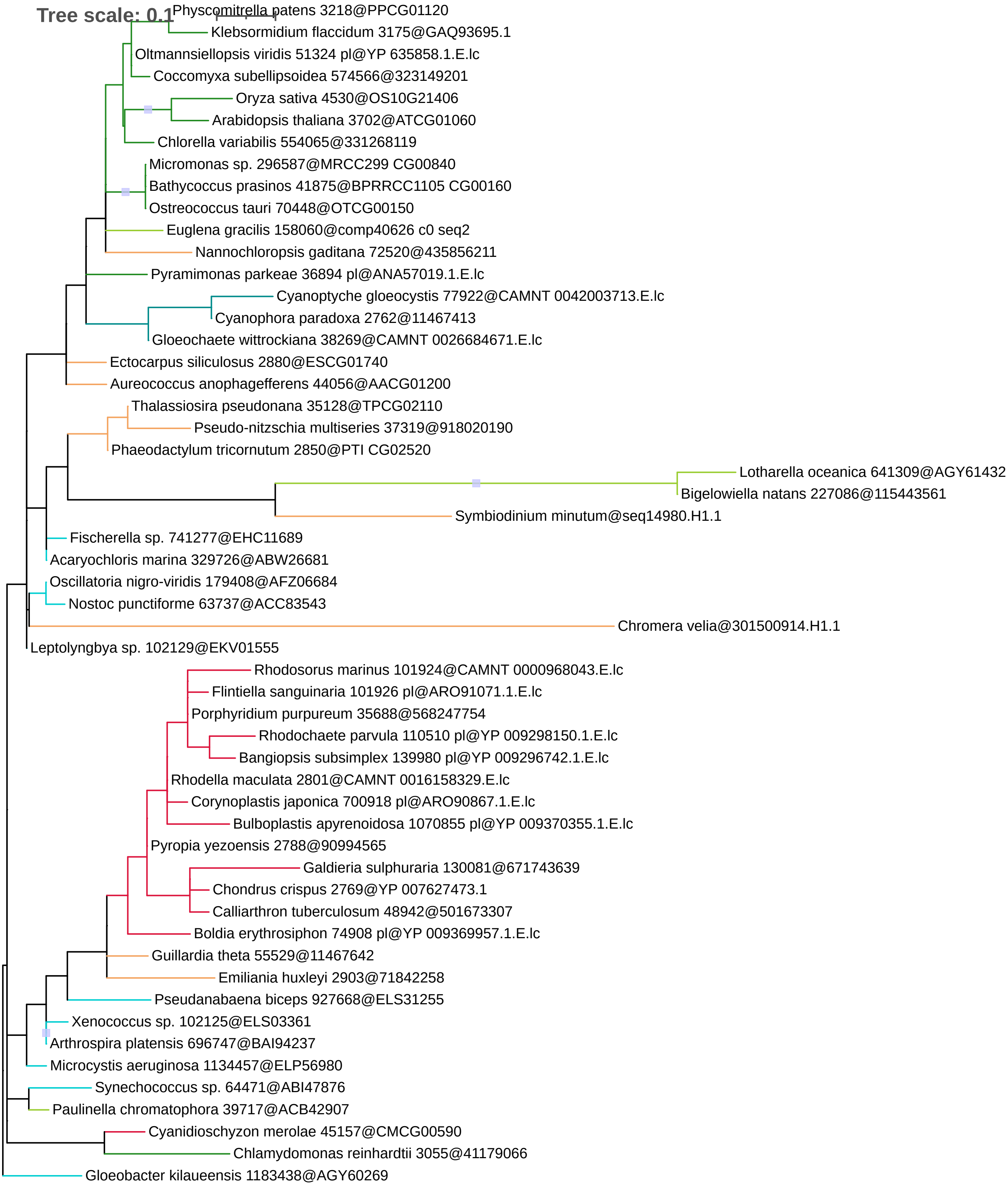


File: 34-PSAA.pdf

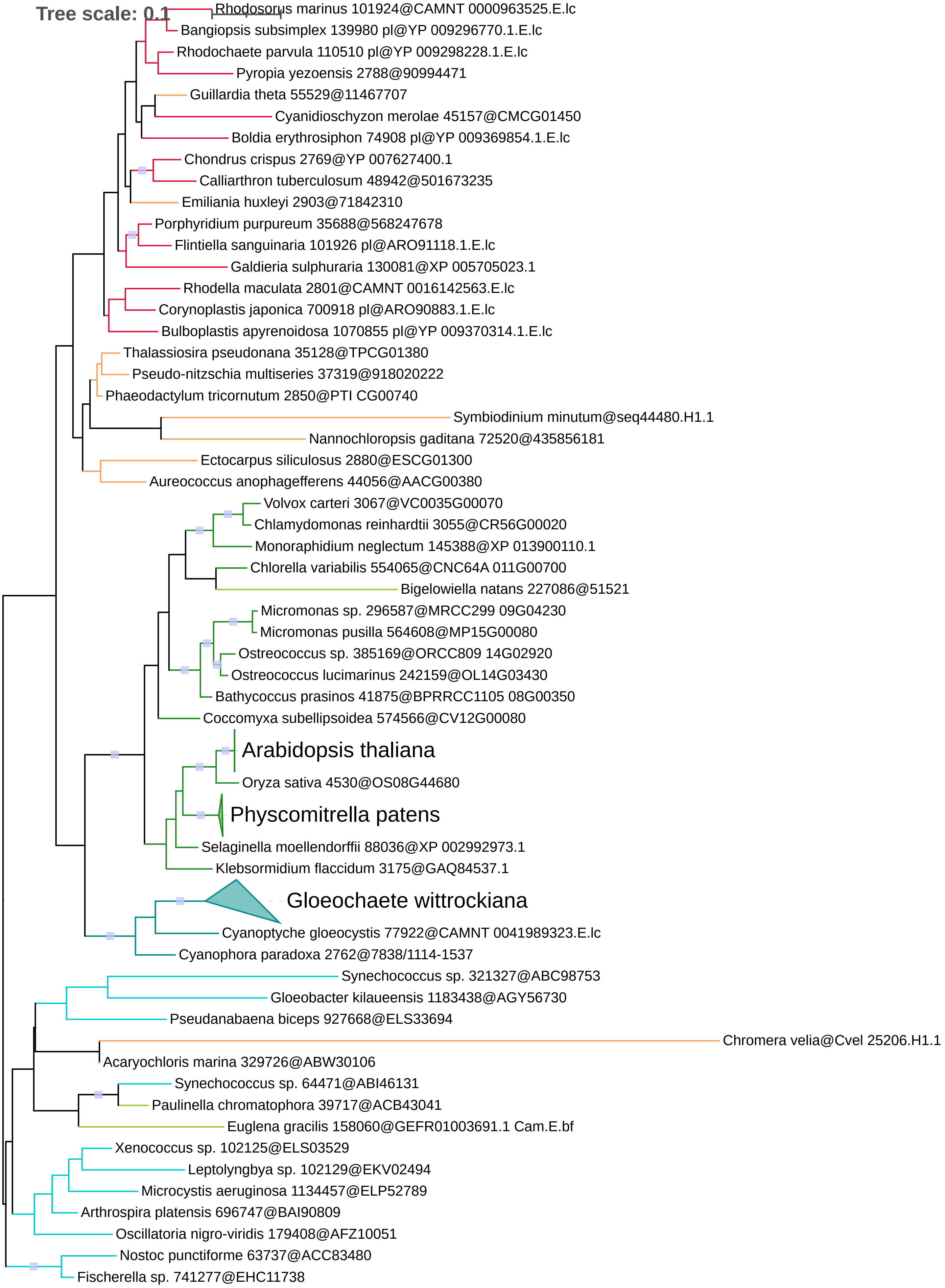
File: 35-PSAB.pdf



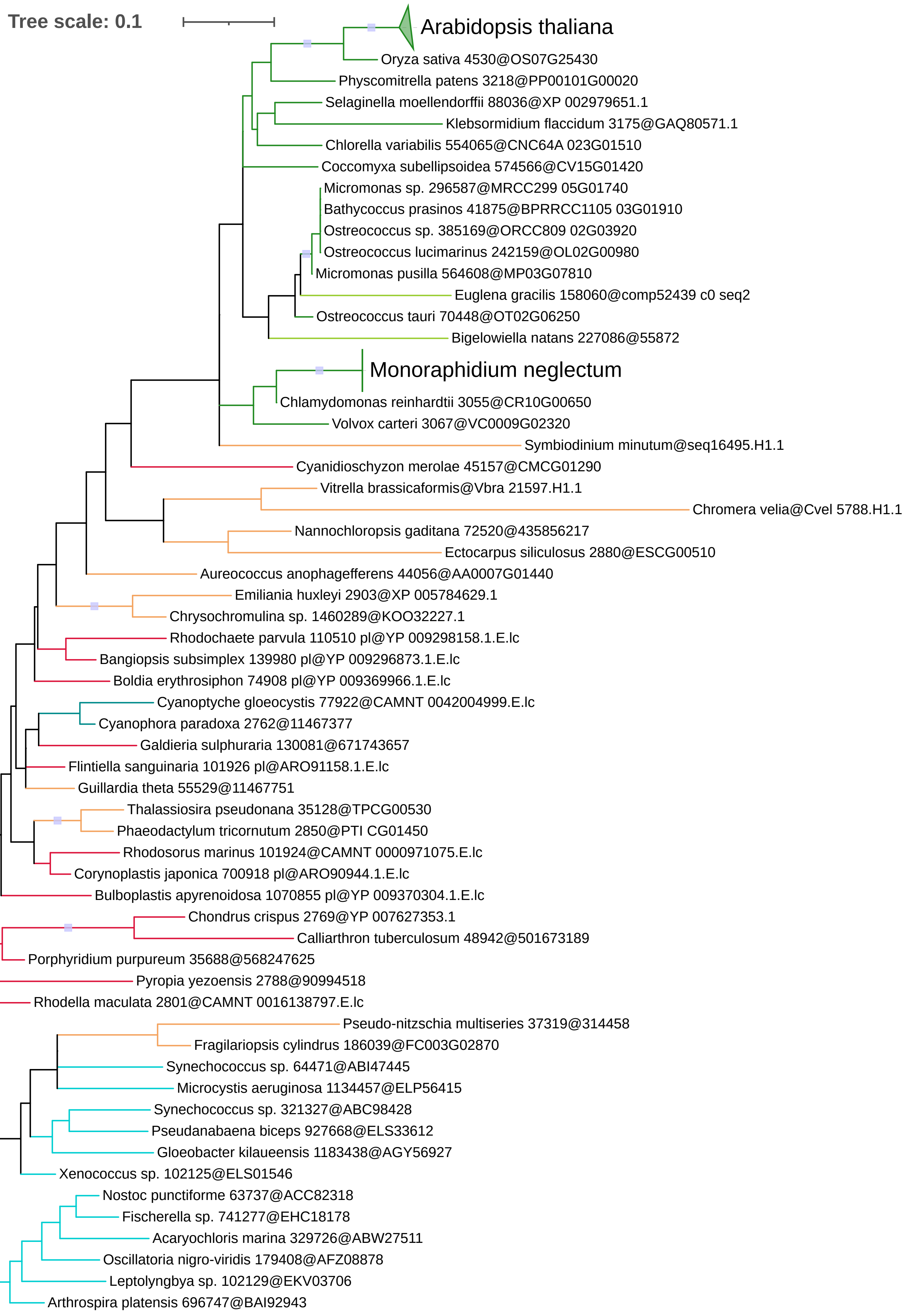
File: 36-PSAC.pdf



File: 37-PSAD.pdf

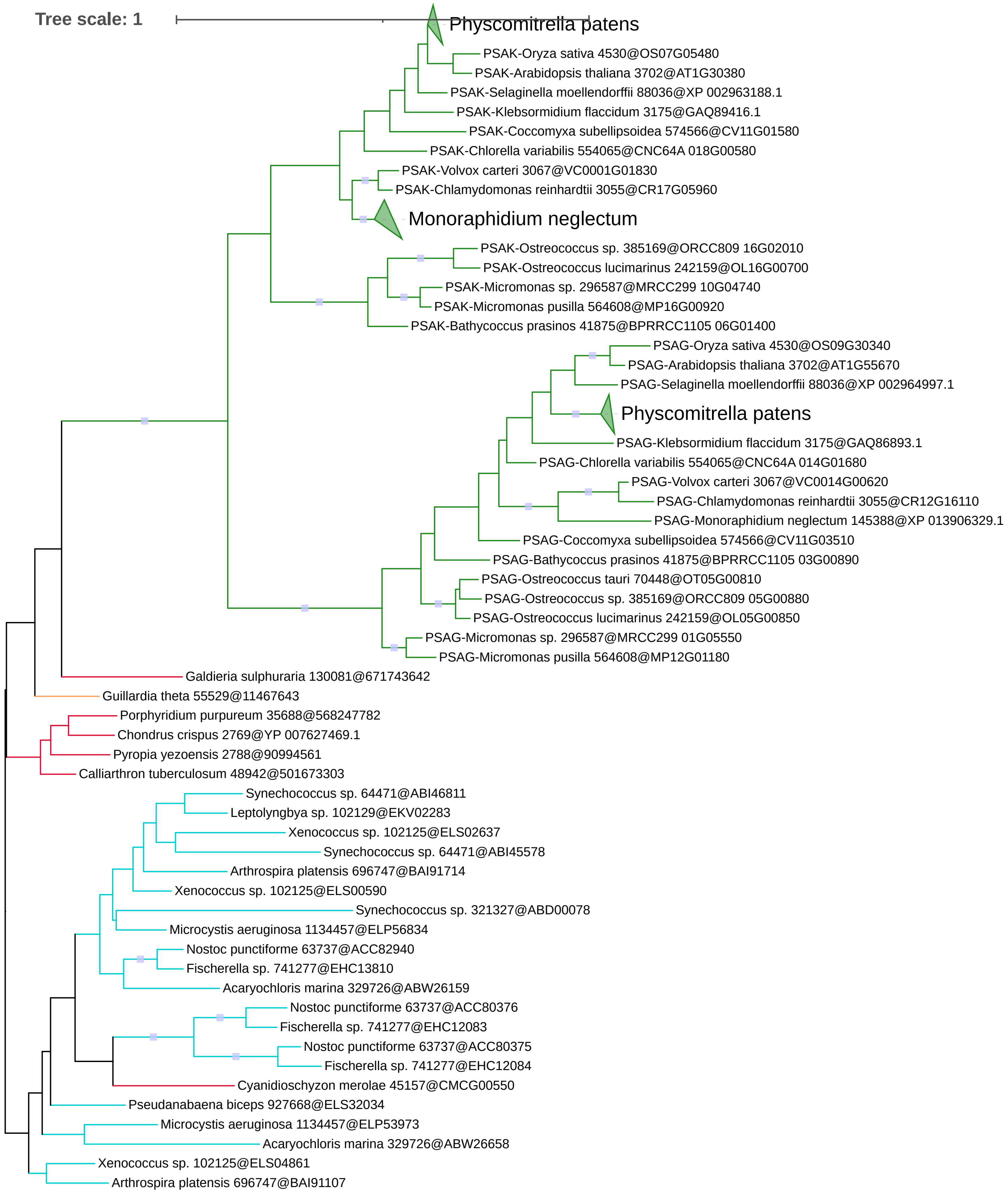


File: 38-PSAE.pdf

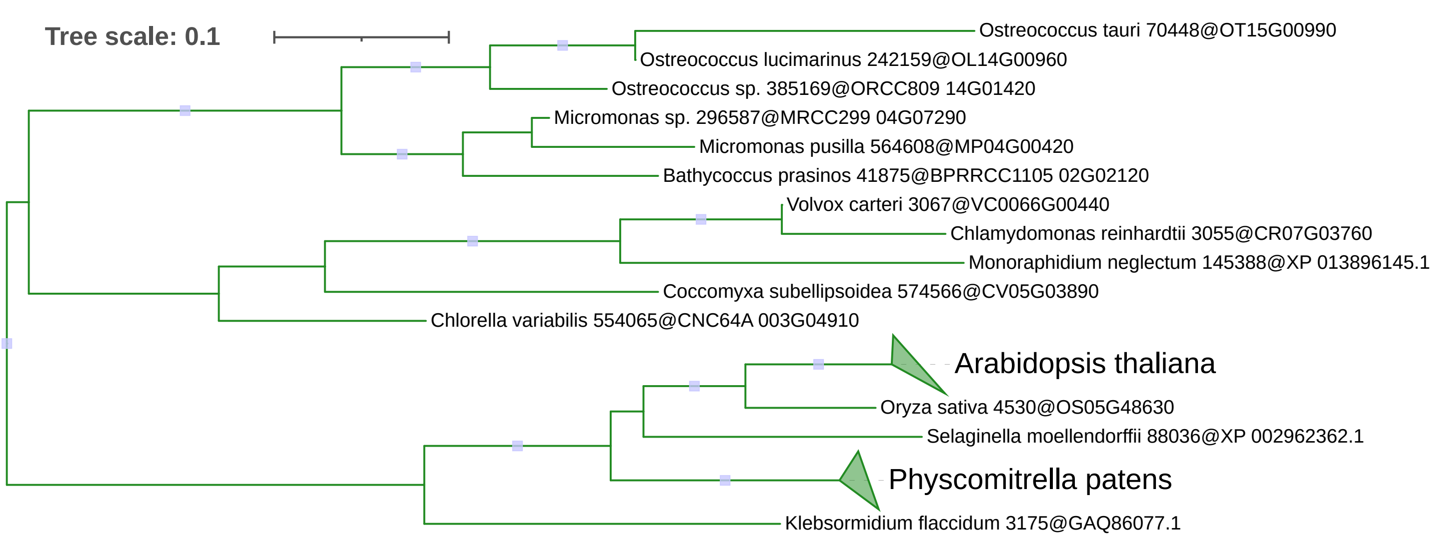


File: 39-PSAF.pdf

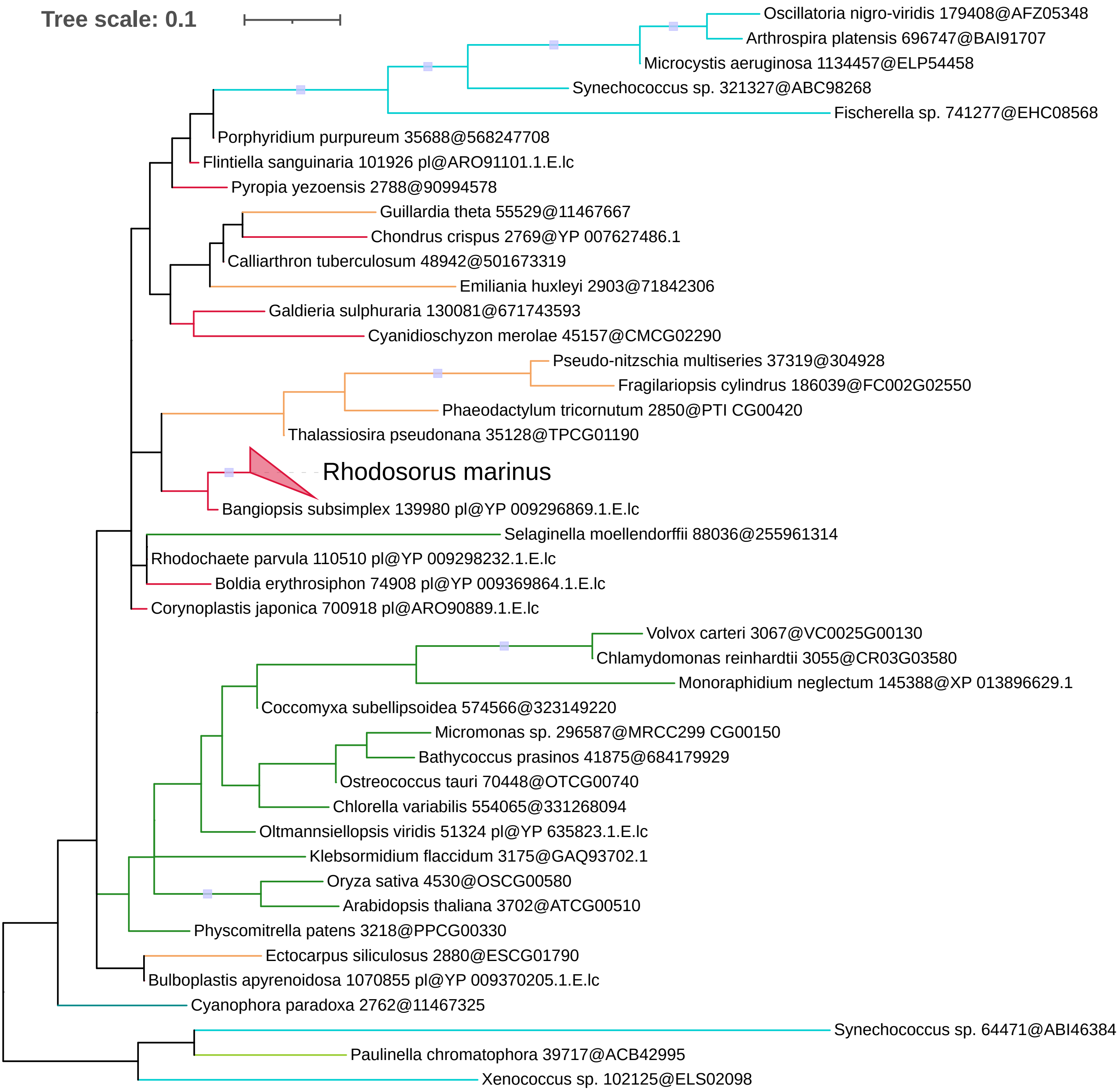
File: 40-PSAG-PSAK.pdf



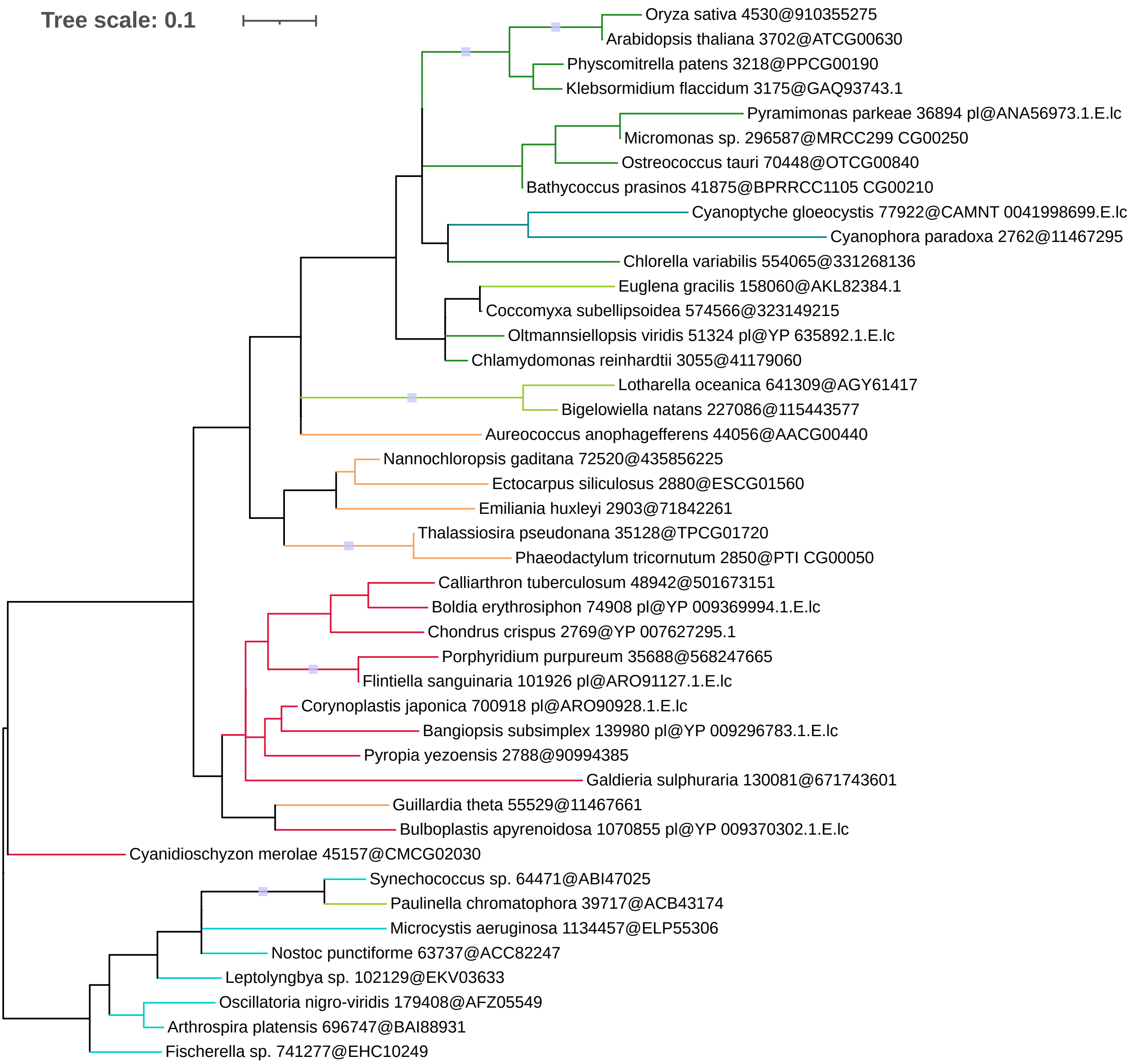
File: 41-PSAH.pdf



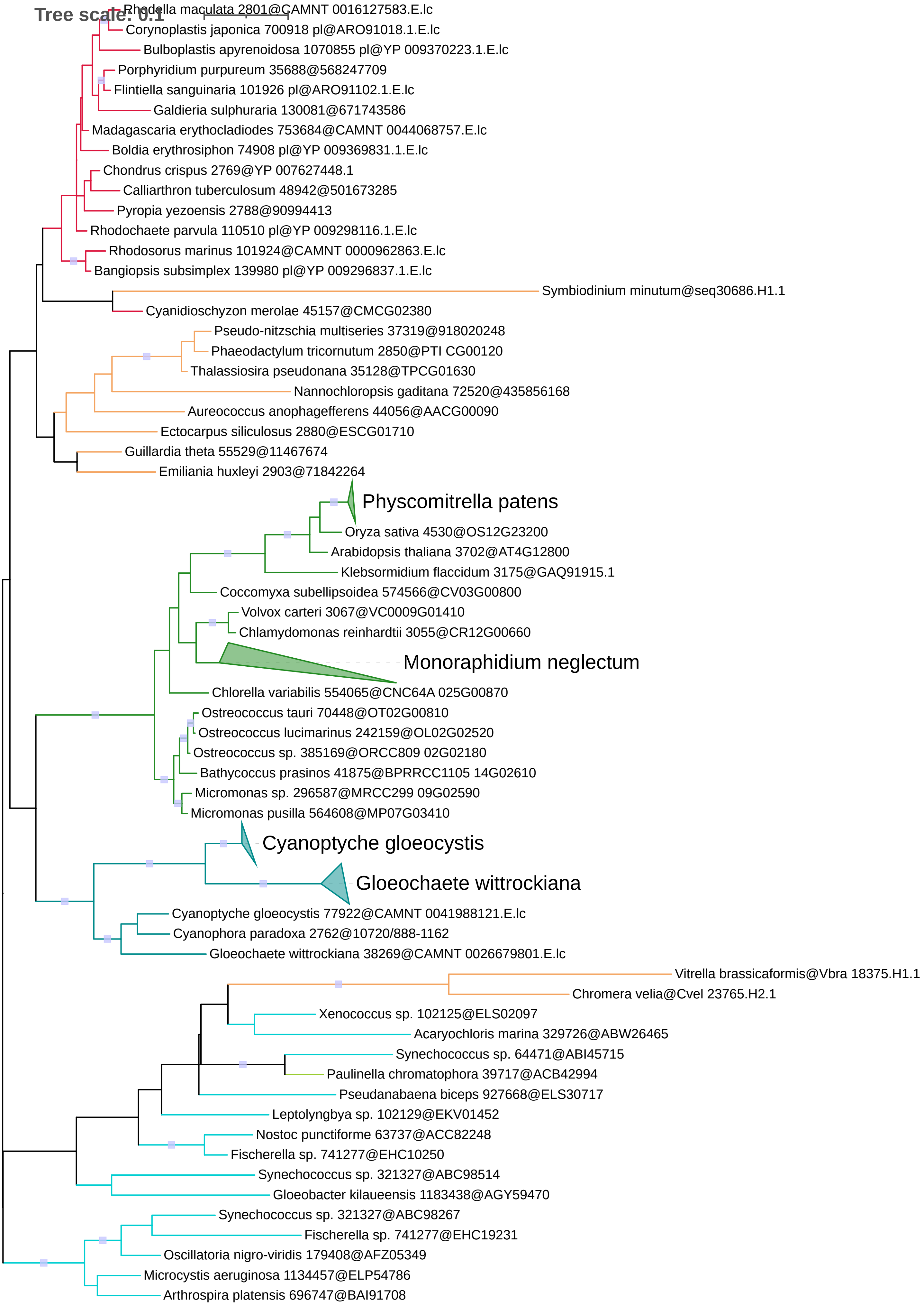
File: 42-PSAI.pdf



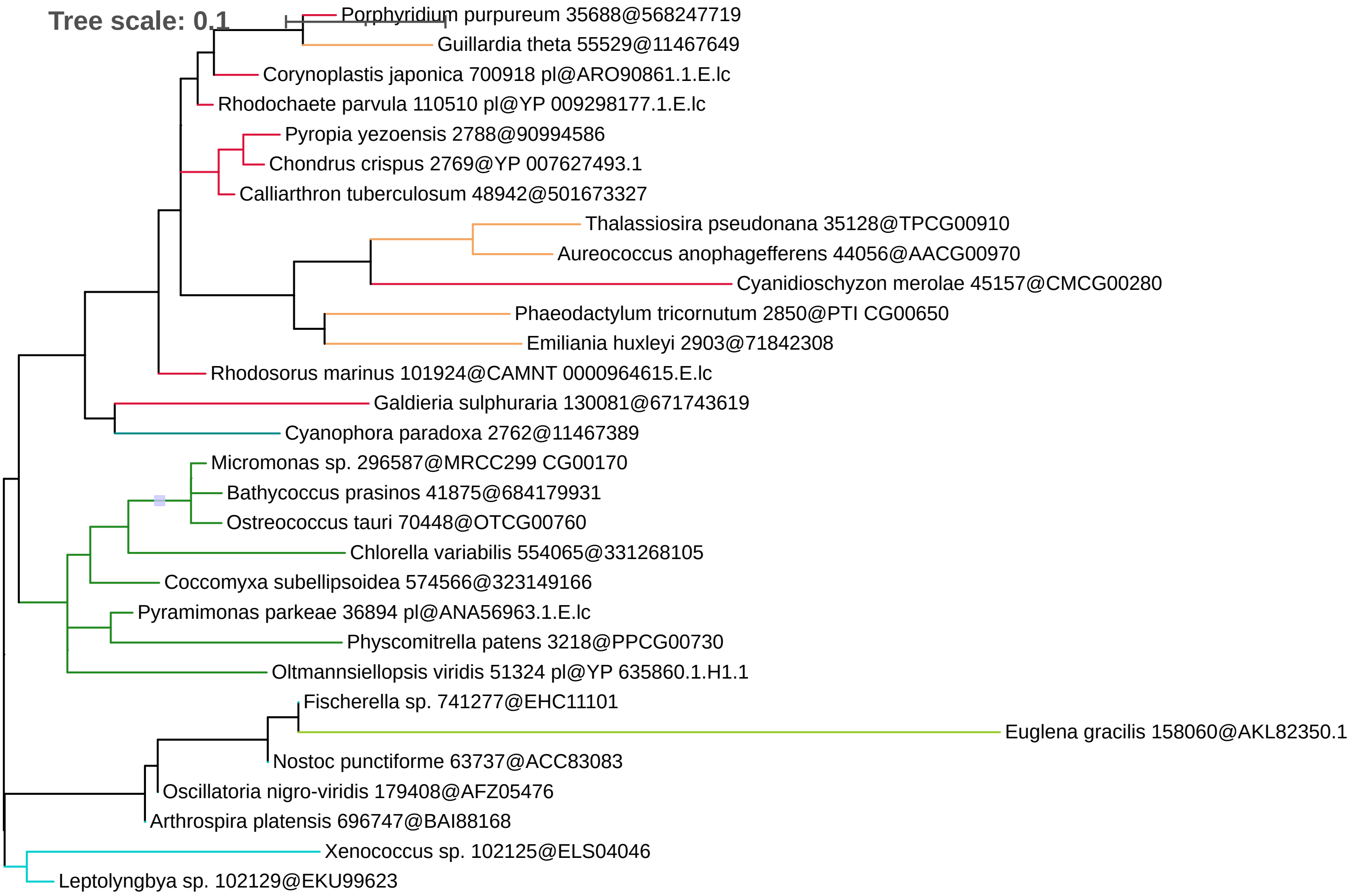
File: 43-PSAJ.pdf



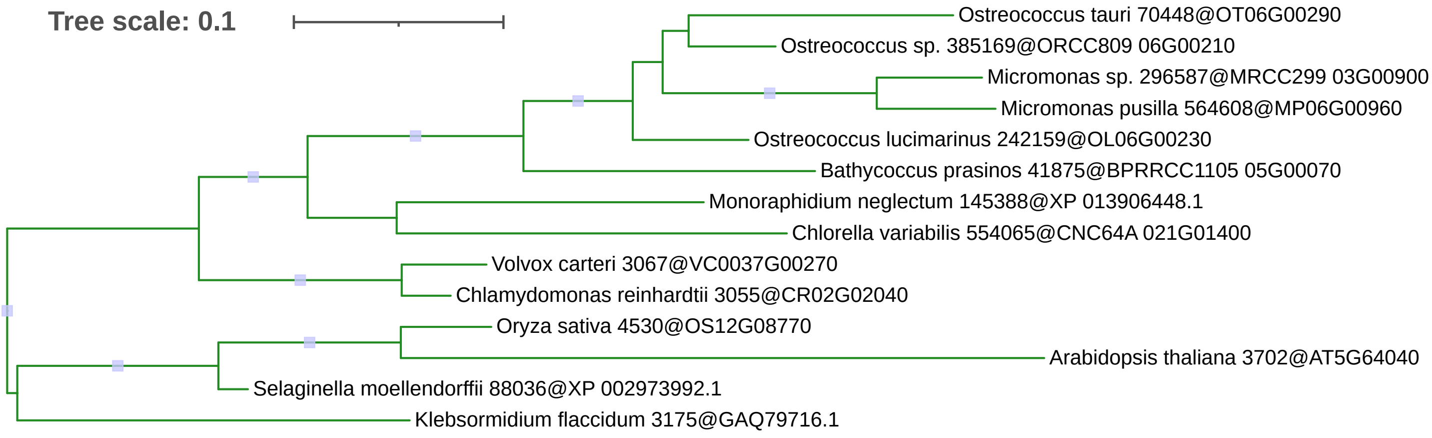
File: 44-PSAL.pdf



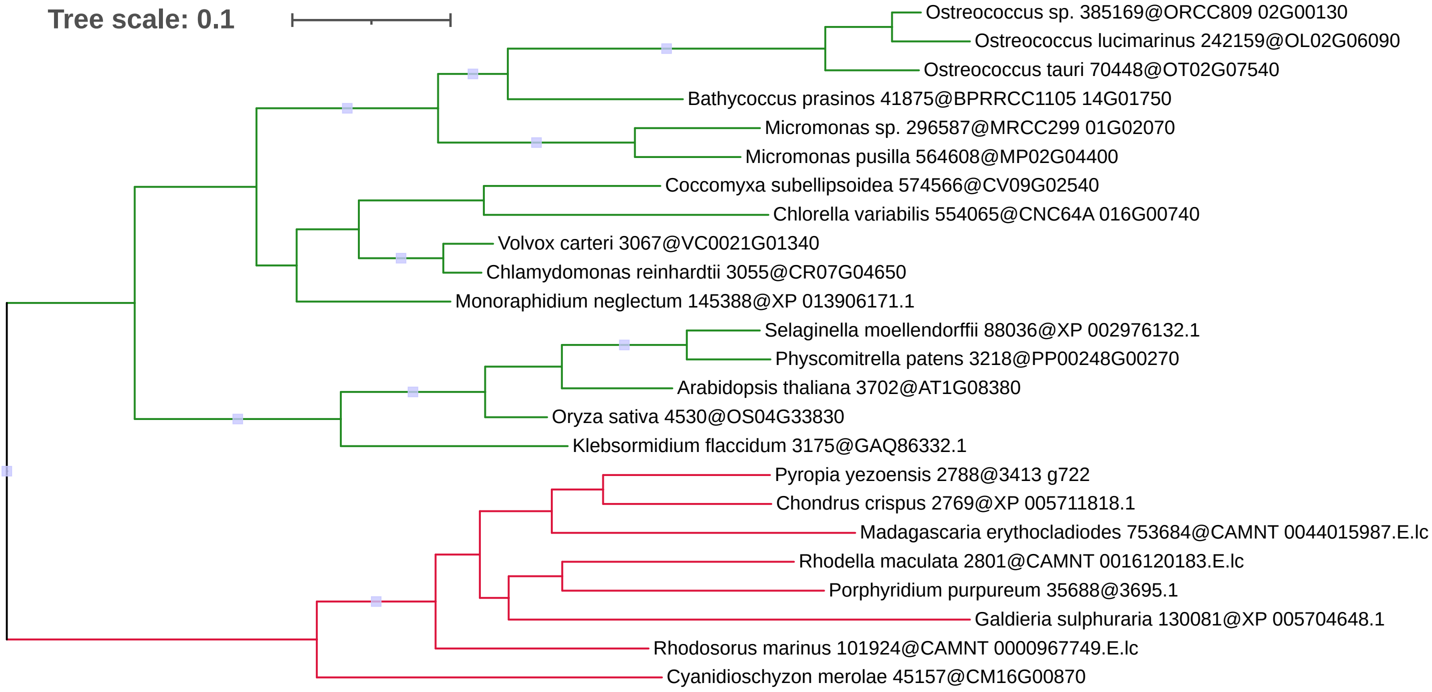
File: 45-PSAM.pdf



File: 46-PSAN.pdf

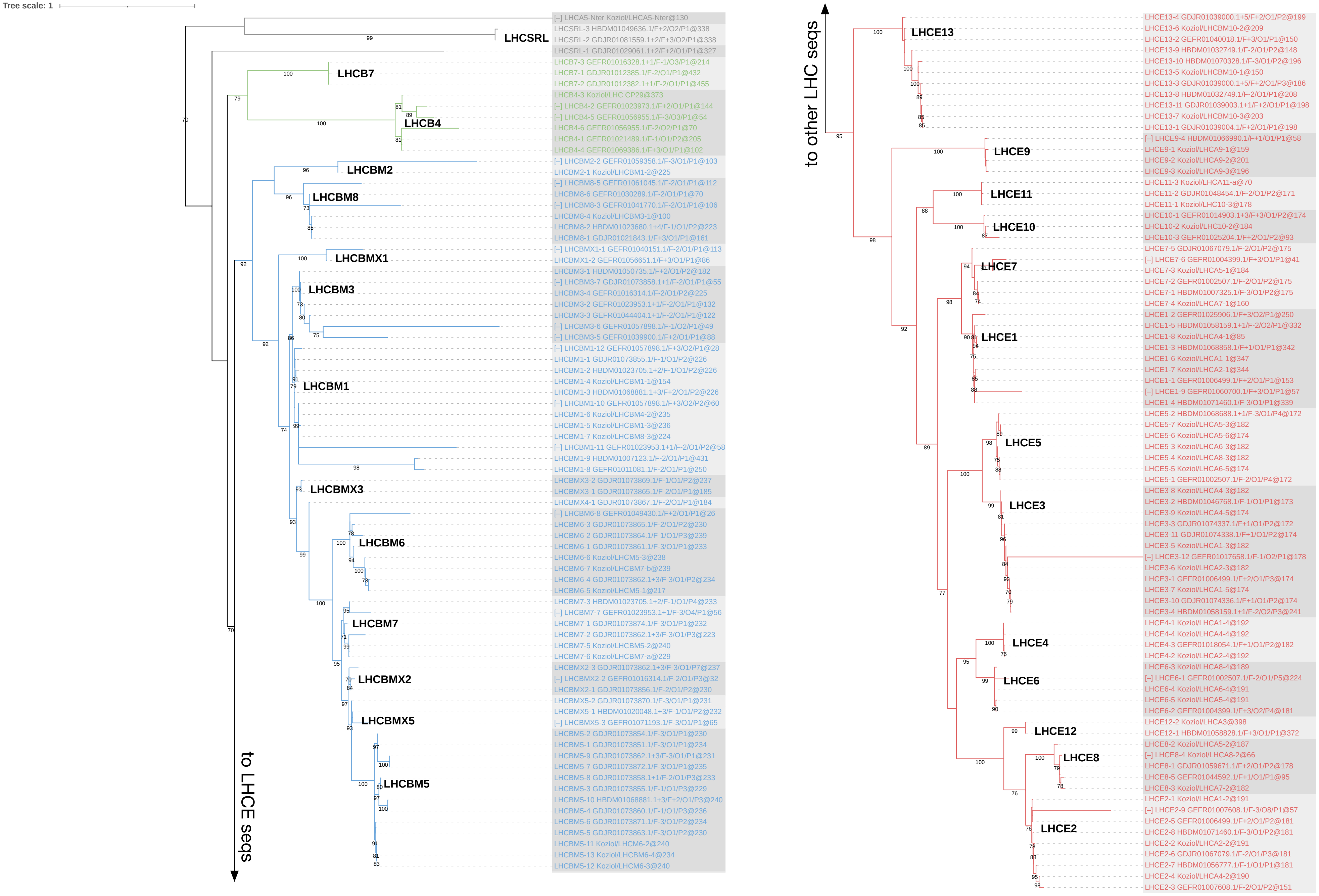


File: 47-PSAO.pdf

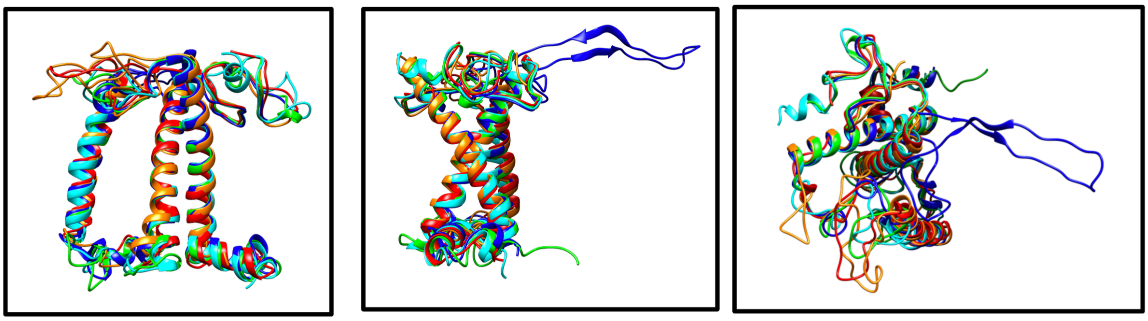


Supplementary Figures S1-47 : Phylogenetic trees of PSI/PSII subunits inferred with RAxML under the PROTGAMMALG4X model and 100 rapid bootstrap replicates. Branches are in substitutions per site and color-coded as follows: Cyanobacteria (light blue), Glaucophyta (turquoise), Rhodophyta (red), Viridiplantae (dark green), complex red algae (orange), complex green algae, including Euglena (light green). Nodes with bootstrap support values ≥70% are indicated by semi-transparent squares. When a species had multiple in-paralogous genes, the corresponding branches were collapsed in a triangle.

Supplementary Figure 1- PSBA
Supplementary Figure 2-PSBB
Supplementary Figure 3-PSBC
Supplementary Figure 4-PSBD
Supplementary Figure 5-PSBE
Supplementary Figure 6-PSBF
Supplementary Figure 7-PSBH
Supplementary Figure 8-PSBI
Supplementary Figure 9-PSBJ
Supplementary Figure 10-PSBK
Supplementary Figure 11-PSBL
Supplementary Figure 12-PSBM
Supplementary Figure 13-PSBN
Supplementary Figure 14-PSBO
Supplementary Figure 15-PSBP-PPL1
Supplementary Figure 16-PSBP2
Supplementary Figure 17-PNSL23-PSBQ
Supplementary Figure 18-PSBR
Supplementary Figure 19-PSBS
Supplementary Figure 20-PSBT
Supplementary Figure 21-PSBU
Supplementary Figure 22-PSBV
Supplementary Figure 23-PSBW
Supplementary Figure 24-PSBX
Supplementary Figure 25-PSBY
Supplementary Figure 26-PSBZ
Supplementary Figure 27-PSB27-1-2
Supplementary Figure 28-PSB28
Supplementary Figure 29-PSB29
Supplementary Figure 30-PSB32_A
Supplementary Figure 31-PSB32_B
Supplementary Figure 32-PSB33
Supplementary Figure 33-YCF12
Supplementary Figure 34-PSAA
Supplementary Figure 35-PSAB
Supplementary Figure 36-PSAC
Supplementary Figure 37-PSAD
Supplementary Figure 38-PSAE
Supplementary Figure 39-PSAF
Supplementary Figure 40-PSAG-PSAK
Supplementary Figure 41-PSAH
Supplementary Figure 42-PSAI
Supplementary Figure 43-PSAJ
Supplementary Figure 44-PSAL
Supplementary Figure 45-PSAM
Supplementary Figure 46-PSAN
Supplementary Figure 47-PSAO


Tree scale: 1
LHCSRL
LHCB7
LHCB4
LHCBM2
LHCBM8
LHCBMX1
LHCBM3
LHCBM1
LHCBMX3
LHCBM6
LHCBM7
LHCBMX2
LHCBMX5
LHCBM5
to LHCE seqs
[-] LHCA5-Nter Koziol/LHCA5-Nter@130
LHCSRL-3 HBDM01049636.1/F+2/O2/P1@338
LHCSRL-2 GDJR01081559.1+2/F+3/O2/P1@338
LHCSRL-1 GDJR01029061.1+2/F+2/O1/P1@327
LHCB7-3 GEFR01016328.1+1/F-1/O3/P1@214
LHCB7-1 GDJR01012385.1/F-2/O1/P1@432
LHCB7-2 GDJR01012382.1+1/F-2/O1/P1@455
LHCB4-3 Koziol/LHC CP29@373
[-] LHCB4-2 GEFR01023973.1/F+2/O1/P1@144
[-] LHCB4-5 GEFR01056955.1/F-3/O3/P1@54
LHCB4-6 GEFR01056955.1/F-2/O2/P1@70
LHCB4-1 GEFR01021489.1/F-1/O1/P2@205
LHCB4-4 GEFR01069386.1/F+3/O1/P1@102
[-] LHCBM2-2 GEFR01059358.1/F-3/O1/P1@103
LHCBM2-1 Koziol/LHCBM1-2@225
[-] LHCBM8-5 GEFR01061045.1/F-2/O1/P1@112
LHCBM8-6 GEFR01030289.1/F-2/O1/P1@70
[-] LHCBM8-3 GEFR01041770.1/F-2/O1/P1@106
LHCBM8-4 Koziol/LHCBM3-1@100
LHCBM8-2 HBDM01023680.1+4/F-1/O1/P2@223
LHCBM8-1 GDJR01021843.1/F+3/O1/P1@161
[-] LHCBMX1-1 GEFR01040151.1/F-2/O1/P1@113
LHCBMX1-2 GEFR01056651.1/F+3/O1/P1@86
LHCBM3-1 HBDM01050735.1/F+2/O1/P2@182
[-] LHCBM3-7 GDJR01073858.1+1/F-2/O1/P1@55
LHCBM3-4 GEFR01016314.1/F-2/O1/P2@225
LHCBM3-2 GEFR01023953.1+1/F-2/O1/P1@132
LHCBM3-3 GEFR01044404.1+1/F-2/O1/P1@122
[-] LHCBM3-6 GEFR01057898.1/F-1/O2/P1@49
[-] LHCBM3-5 GEFR01039900.1/F+2/O1/P1@88
[-] LHCBM1-12 GEFR01057898.1/F+3/O2/P1@28
LHCBM1-1 GDJR01073855.1/F-1/O1/P2@226
LHCBM1-2 HBDM01023705.1+2/F-1/O1/P2@226
LHCBM1-4 Koziol/LHCBM1-1@154
LHCBM1-3 HBDM01068881.1+3/F+2/O1/P2@226
[-] LHCBM1-10 GEFR01057898.1/F+3/O2/P2@60
LHCBM1-6 Koziol/LHCBM4-2@235
LHCBM1-5 Koziol/LHCBM1-3@236
LHCBM1-7 Koziol/LHCBM8-3@224
[-] LHCBM1-11 GEFR01023953.1+1/F-2/O1/P2@58
LHCBM1-9 HBDM01007123.1/F-2/O1/P1@431
LHCBM1-8 GEFR01011081.1/F-2/O1/P1@250
LHCBMX3-2 GDJR01073869.1/F-1/O1/P2@237
LHCBMX3-1 GDJR01073865.1/F-2/O1/P1@185
LHCBMX4-1 GDJR01073867.1/F-2/O1/P1@184
[-] LHCBM6-8 GEFR01049430.1/F+2/O1/P1@26
LHCBM6-3 GDJR01073865.1/F-2/O1/P2@230
LHCBM6-2 GDJR01073864.1/F-1/O1/P3@239
LHCBM6-1 GDJR01073861.1/F-3/O1/P1@233
LHCBM6-6 Koziol/LHCM5-3@238
LHCBM6-7 Koziol/LHCBM7-b@239
LHCBM6-4 GDJR01073862.1+3/F-3/O1/P2@234
LHCBM6-5 Koziol/LHCM5-1@217
LHCBM7-3 HBDM01023705.1+2/F-1/O1/P4@233
[-] LHCBM7-7 GEFR01023953.1+1/F-3/O4/P1@56
LHCBM7-1 GDJR01073874.1/F-3/O1/P1@232
LHCBM7-2 GDJR01073862.1+3/F-3/O1/P3@223
LHCBM7-5 Koziol/LHCBM5-2@240
LHCBM7-6 Koziol/LHCBM7-a@229
LHCBMX2-3 GDJR01073862.1+3/F-3/O1/P7@237
[-] LHCBMX2-2 GEFR01016314.1/F-2/O1/P3@32
LHCBMX2-1 GDJR01073856.1/F-2/O1/P2@230
LHCBMX5-2 GDJR01073870.1/F-3/O1/P1@231
LHCBMX5-1 HBDM01020048.1+3/F-1/O1/P2@232
[-] LHCBMX5-3 GEFR01071193.1/F-3/O1/P1@65
LHCBM5-2 GDJR01073854.1/F-3/O1/P1@230
LHCBM5-1 GDJR01073851.1/F-3/O1/P1@234
LHCBM5-9 GDJR01073862.1+3/F-3/O1/P1@231
LHCBM5-7 GDJR01073872.1/F-3/O1/P1@235
LHCBM5-8 GDJR01073858.1+1/F-2/O1/P3@233
LHCBM5-3 GDJR01073855.1/F-1/O1/P3@229
LHCBM5-10 HBDM01068881.1+3/F+2/O1/P3@240
LHCBM5-4 GDJR01073860.1/F-1/O1/P3@236
LHCBM5-6 GDJR01073871.1/F-3/O1/P2@234
LHCBM5-5 GDJR01073863.1/F-3/O1/P2@230
LHCBM5-11 Koziol/LHCM6-2@240
LHCBM5-13 Koziol/LHCBM6-4@234
LHCBM5-12 Koziol/LHCM6-3@240
to other LHC seqs
LHCE13
LHCE9
LHCE11
LHCE10
LHCE7
LHCE1
LHCE5
LHCE3
LHCE4
LHCE6
LHCE12
LHCE8
LHCE2
LHCE13-4 GDJR01039000.1+5/F+2/O1/P2@199
LHCE13-6 Koziol/LHCBM10-2@209
LHCE13-2 GEFR01040018.1/F+3/O1/P1@150
LHCE13-9 HBDM01032749.1/F-2/O1/P2@148
LHCE13-10 HBDM01070328.1/F-3/O1/P2@196
LHCE13-5 Koziol/LHCBM10-1@150
LHCE13-3 GDJR01039000.1+5/F+2/O1/P3@186
LHCE13-8 HBDM01032749.1/F-2/O1/P1@208
LHCE13-11 GDJR01039003.1+1/F+2/O1/P1@198
LHCE13-7 Koziol/LHCBM10-3@203
LHCE13-1 GDJR01039004.1/F+2/O1/P1@198
[-] LHCE9-4 HBDM01066990.1/F+1/O1/P1@58
LHCE9-1 Koziol/LHCA9-1@159
LHCE9-2 Koziol/LHCA9-2@201
LHCE9-3 Koziol/LHCA9-3@196
LHCE11-3 Koziol/LHCA11-a@70
LHCE11-2 GDJR01048454.1/F-2/O1/P2@171
LHCE11-1 Koziol/LHC10-3@178
LHCE10-1 GEFR01014903.1+3/F+3/O1/P2@174
LHCE10-2 Koziol/LHC10-2@184
LHCE10-3 GEFR01025204.1/F+2/O1/P2@93
LHCE7-5 GDJR01067079.1/F-2/O1/P2@175
[-] LHCE7-6 GEFR01004399.1/F+3/O1/P1@41
LHCE7-3 Koziol/LHCA5-1@184
LHCE7-2 GEFR01002507.1/F-2/O1/P2@175
LHCE7-1 HBDM01007325.1/F-3/O1/P2@175
LHCE7-4 Koziol/LHCA7-1@160
LHCE1-2 GEFR01025906.1/F+3/O2/P1@250
LHCE1-5 HBDM01058159.1+1/F-2/O2/P1@332
LHCE1-8 Koziol/LHCA4-1@85
LHCE1-3 HBDM01068858.1/F+1/O1/P1@342
LHCE1-6 Koziol/LHCA1-1@347
LHCE1-7 Koziol/LHCA2-1@344
LHCE1-1 GEFR01006499.1/F+2/O1/P1@153
[-] LHCE1-9 GEFR01060700.1/F+3/O1/P1@57
LHCE1-4 HBDM01071460.1/F-3/O1/P1@339
LHCE5-2 HBDM01068688.1+1/F-3/O1/P4@172
LHCE5-7 Koziol/LHCA5-3@182
LHCE5-6 Koziol/LHCA5-6@174
LHCE5-3 Koziol/LHCA6-3@182
LHCE5-4 Koziol/LHCA8-3@182
LHCE5-5 Koziol/LHCA6-5@174
LHCE5-1 GEFR01002507.1/F-2/O1/P4@172
LHCE3-8 Koziol/LHCA4-3@182
LHCE3-2 HBDM01046768.1/F-1/O1/P1@173
LHCE3-9 Koziol/LHCA4-5@174
LHCE3-3 GDJR01074337.1/F+1/O1/P2@172
LHCE3-11 GDJR01074338.1/F+1/O1/P2@174
LHCE3-5 Koziol/LHCA1-3@182
[-] LHCE3-12 GEFR01017658.1/F-1/O2/P1@178
LHCE3-6 Koziol/LHCA2-3@182
LHCE3-1 GEFR01006499.1/F+2/O1/P3@174
LHCE3-7 Koziol/LHCA1-5@174
LHCE3-10 GDJR01074336.1/F+1/O1/P2@174
LHCE3-4 HBDM01058159.1+1/F-2/O2/P3@241
LHCE4-1 Koziol/LHCA1-4@192
LHCE4-4 Koziol/LHCA4-4@192
LHCE4-3 GEFR01018054.1/F+1/O1/P2@182
LHCE4-2 Koziol/LHCA2-4@192
LHCE6-3 Koziol/LHCA8-4@189
[-] LHCE6-1 GEFR01002507.1/F-2/O1/P5@224
LHCE6-4 Koziol/LHCA6-4@191
LHCE6-5 Koziol/LHCA5-4@191
LHCE6-2 GEFR01004399.1/F+3/O2/P4@181
LHCE12-2 Koziol/LHCA3@398
LHCE12-1 HBDM01058828.1/F+3/O1/P1@372
LHCE8-2 Koziol/LHCA5-2@187
[-] LHCE8-4 Koziol/LHCA8-2@66
LHCE8-1 GDJR01059671.1/F+2/O1/P2@178
LHCE8-5 GEFR01044592.1/F+1/O1/P1@95
LHCE8-3 Koziol/LHCA7-2@182
LHCE2-1 Koziol/LHCA1-2@191
[-] LHCE2-9 GEFR01007608.1/F-3/O8/P1@57
LHCE2-5 GEFR01006499.1/F+2/O1/P2@181
LHCE2-8 HBDM01071460.1/F-3/O1/P2@181
LHCE2-2 Koziol/LHCA2-2@191
LHCE2-6 GDJR01067079.1/F-2/O1/P3@181
LHCE2-7 HBDM01056777.1/F-1/O1/P1@181
LHCE2-4 Koziol/LHCA4-2@190
LHCE2-3 GEFR01007608.1/F-2/O1/P2@151

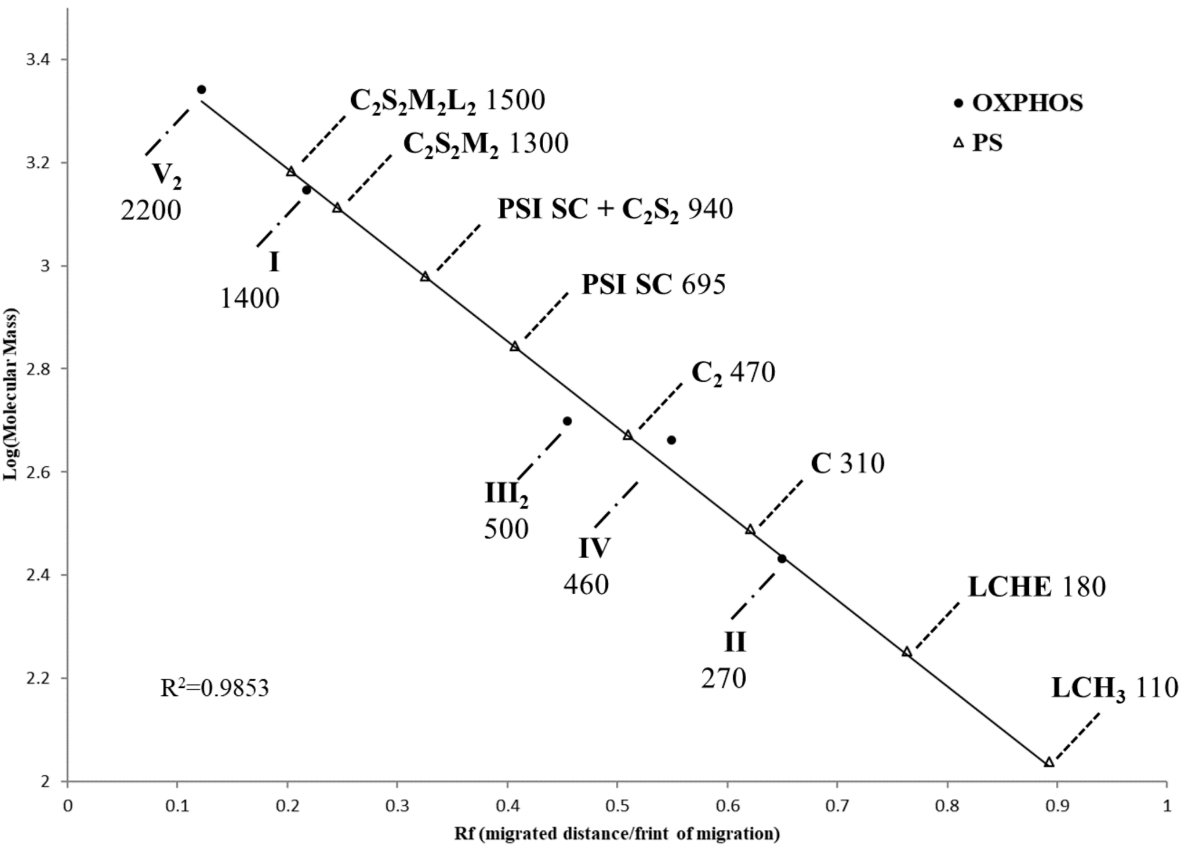
Supplementary Figure S48. Phylogenetic tree of *Euglena gracilis* Light-Harvesting Complex (Lhc) Proteins. (A) Phylogenetic tree of Lhc proteins in *Euglena gracilis*. The maximum likelihood phylogenetic tree was constructed using an expanded dataset of Euglena gracilis Lhc proteins. Bootstrap values are indicated at key nodes. Lhc proteins are grouped into distinct families, with major clades highlighted in different colors: LhcbM proteins (blue), LhcE proteins (red), Lhcb (green) other Lhc-related families (gray).



Supplementary Figure S49. Structural comparison of Light Harvesting Complexes proteins. Overlap of LHCII (green), CP24 (orange), CP26 (cyan), CP29 (blue), LhcE9 (red) structures. Left panel: side view, center panel: 90° rotation on the same plane view, right panel: upper view. Structures from LHCII, CP24, CP26 and CP29 were obtained from PSII-LHCII supercomplex from Pisum sativum (PDB: 5XNL (Su et al., 2017)). LhcE9 homology model were obtained with SWISS-MODEL (Waterhouse et al., 2018).

Su X, Ma J, Wei X, Cao P, Zhu D, Chang W, Liu Z, Zhang X, Li M. Structure and assembly mechanism of plant $C_2S_2M_2$-type PSII-LHCII supercomplex. Science. 2017 Aug 25;357(6353):815-820. doi: 10.1126/science.aan0327. PMID: 28839073.

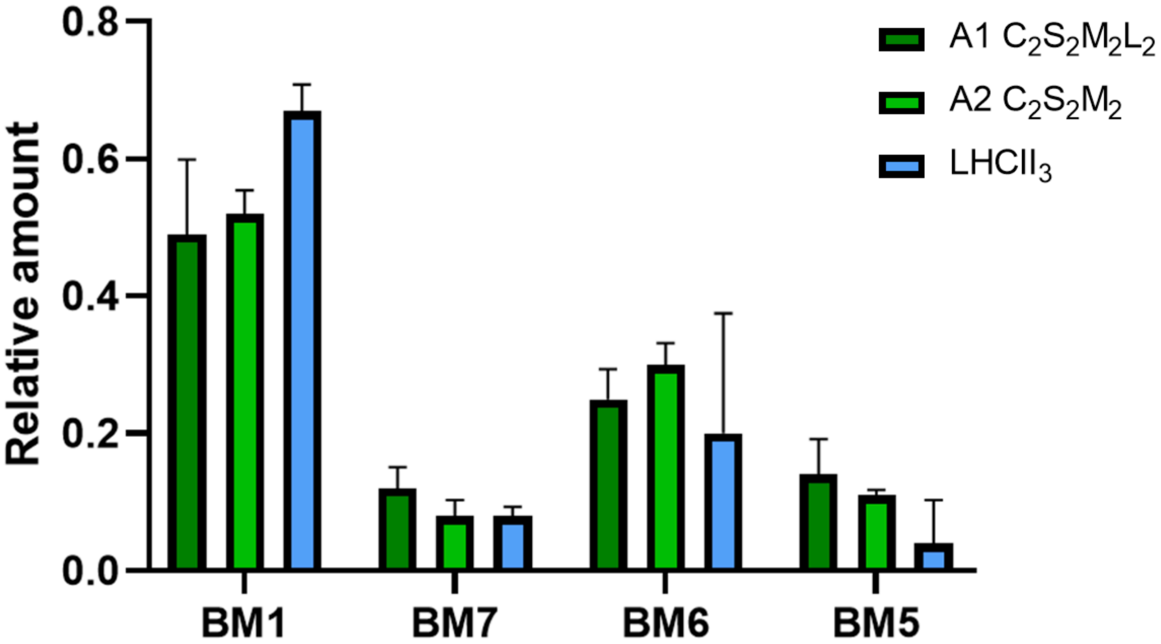
Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 2018 Jul 2;46(W1):W296-W303. doi: 10.1093/nar/gky427. PMID: 29788355; PMCID: PMC6030848.



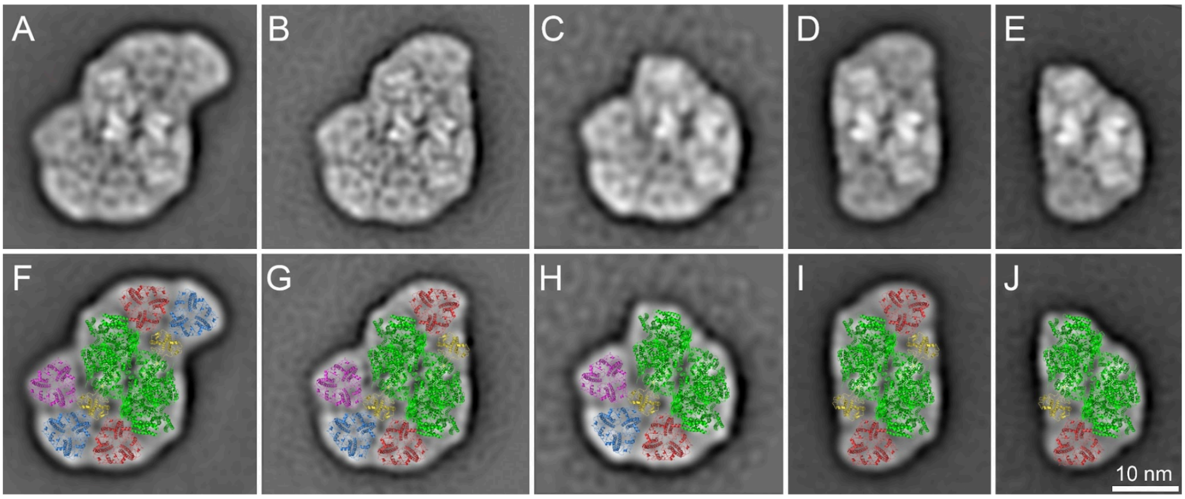
Supplementary Figure S50. Estimated molecular masses for the *Euglena gracilis* photosynthetic complexes. The logarithms of the molecular masses of previously characterized mitochondrial complexes (●) (V_2, I, III_2, IV) (Miranda-Astudillo et al., 2018; Yadav et al., 2017) were plotted against their migration distance in *hr*CN-PAGE ($R^2 = 0.9755$). Then, the migration distances of the *Euglena* photosynthetic complexes (Δ) were interpolated and their corresponding molecular masses inferred.

Yadav KNS, Miranda-Astudillo HV, Colina-Tenorio L, Bouillenne F, Degand H, Morsomme P, González-Halphen D, Boekema EJ, Cardol P. Atypical composition and structure of the mitochondrial dimeric ATP synthase from Euglena gracilis. Biochim Biophys Acta Bioenerg. 2017 Apr;1858(4):267-275. doi: 10.1016/j.bbabio.2017.01.007. Epub 2017 Jan 12. PMID: 28089911.

Miranda-Astudillo HV, Yadav KNS, Colina-Tenorio L, Bouillenne F, Degand H, Morsomme P, Boekema EJ, Cardol P. The atypical subunit composition of respiratory complexes I and IV is associated with original extra structural domains in Euglena gracilis. Sci Rep. 2018 Jun 26;8(1):9698. doi: 10.1038/s41598-018-28039-z. PMID: 29946152; PMCID: PMC6018760.

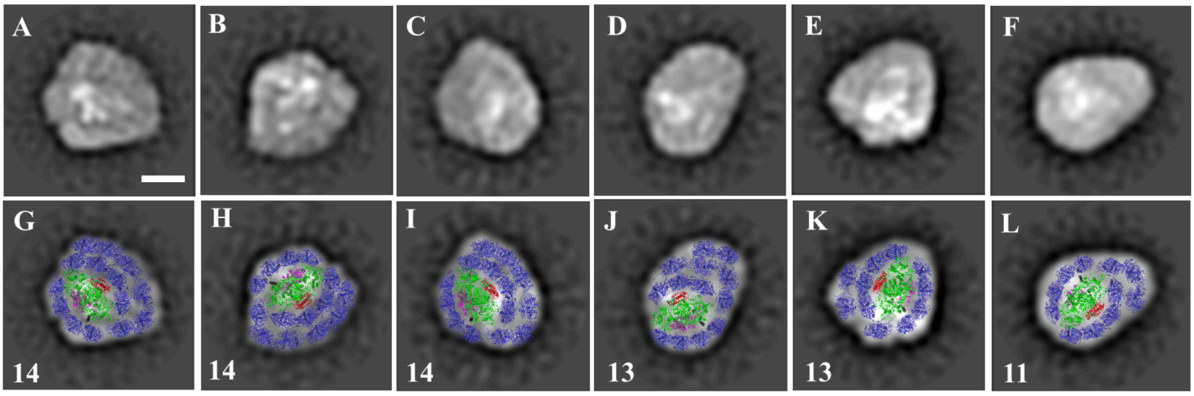


Supplementary Figure S51. Distribution of LhcbM in PSII-SC and Free LHCII Across Different PSII/LHCII Particles. The bar graph illustrates the relative amounts of LhcbM (BM-I, BM-VII, BM-VI, BM-V). The color-coded bars represent different PSII-LHCII compositions: $C_2S_2M_2N_2$ (dark green), $C_2S_2M_2$ (light green), and free $LHCII_3$ (blue). Error bars represent standard deviations.



Supplementary Figure S52. Projection maps and structural models of PSII supercomplexes from *Euglena gracilis* revealed by single particle electron microscopy. (A-E) Projection maps of several subforms of PSII supercomplex. (F-J) Structural models of different forms of the PSII supercomplexes obtained by fitting the high-resolution structure of the PSII supercomplex from green alga *Chlamydomonas reinhardtii* (PDB: 6AKD (Shen et al., 2019). The scale bar is 10 nm.

Shen L, Huang Z, Chang S, Wang W, Wang J, Kuang T, Han G, Shen JR, Zhang X. Structure of a $C_2S_2M_2N_2$-type PSII-LHCII supercomplex from the green alga *Chlamydomonas reinhardtii*. Proc Natl Acad Sci U S A. 2019 Oct 15;116(42):21246-21255. doi: 10.1073/pnas.1912462116. Epub 2019 Sep 30. PMID: 31570614; PMCID: PMC6800332.



Supplementary Figure S53. Structural models of PSI-LHC supercomplexes from Euglena gracilis revealed by single-particle electron microscopy. Structural models of different PSI-LHC supercomplex forms obtained through single-particle electron microscopy. Top rows: raw electron density projections. Bottom rows: corresponding fitted models with assigned structural components. Densities were assigned by fitting the high-resolution PSI structure from *Dunaliella salina* (PDB: 6RHZ, Perez-Boerema et al., 2020). Additional antenna proteins were modeled using copies of the Lhca1 subunit from *Dunaliella salina*. PSI core (green), PsaD subunit (violet); LHC (blue) belt at the PsaF side (red); Variable "flipper" region at the PsaM side (black). The scale bar represents 10 nm.

Perez-Boerema A, Klaiman D, Caspy I, Netzer-El SY, Amunts A, Nelson N. Structure of a minimal photosystem I from the green alga Dunaliella salina. Nat Plants. 2020 Mar;6(3):321-327. doi: 10.1038/s41477-020-0611-9. Epub 2020 Mar 2. PMID: 32123351.

Supplementary dataset 1. Reference sequences for LHH and PSI/PSII subunits (and a few other proteins) compiled from the literature for model organisms *C. reinhardtii*, *Arabidopsis thaliana*, *Ostreococcus tauri*, *Synechocystis* PCC 6803.

```
>LHCB5/CP26-Arabidopsis_thaliana_3702@AT4G10340
MASLGVSEMLGTPLNFRAVSRSSAPLASSPSTFKTVALFSKKKPAPAKSKAVSETSDELA
KWYGPDRRIFLPDGLLDRSEIPEYLNGEVAGDYGYDPFGLGKKPENFAKYQAFELIHARW
AMLGAAGFIIPEALNKYGANCGPEAVWFKTGALLLDGNTLNYFGKNIPINLVLAVVAEVV
LLGGAEYYRITNGLDFEDKLHPGGPFDPLGLAKDPEQGALLKVKEIKNGRLAMFAMLGFF
IQAYVTGEGPVENLAKHLSDPFGNNLLTVIAGTAERAPTL
>LHCB5/CP26-Chlamydomonas_reinhardtii_3055@CR16G05320
MQIQALFKKTGASAPAKKGTASTKVVKPSKAGGKATRGWLGGQGGAADLDKWYGPDRKLF
LPSGLYDRSEIPEYLNGELAGDYGYDPLGLGKDPETVAKYRENELLHARWAMLAAAGILI
PEGLQANGANIKGGTWFETGAEMLNGGTLNYFAVPWGIVSNPLPLFAVIAVEVGLMGAVE
FYRRNGTGPAGYSPGIGKFDSSVFDGLDPLYPGGPFDPLGLADDPEVLQELKVKEIKNGR
LAMVSVLGFAVQSYVTGEGPYANWTKHVADPFGYNLLTVLGAEERTPTL
>LHCB5/CP26-Ostreococcus_tauri_70448@OT14G01370
MRVTALFSKTKAAPAPAKGAKAPAGAKPGITLPSNPFAAKPATKATPAAKKPAAVKKTVV
KTTPVKKPTPATKRAPVKAARQPVDQKAAIAKWYGPNRKLYLPGGLLTKADLPSYLDGTL
AGDYGFDPLGLGADGKVKDYRVAEILHARWAMLAVPGVVIPEALNIPGGVWTETGKVFLD
GETGRDSILEFWPIYAAIQVVLMGGAEKFRQDGVGPAGFVPFKGKFEESAFNGLDPVNPG
GPLDFFDVAKNPQDLQILKVKEIKNGRLAMIAMLGVFVQGAVTGEGPAANWGKHVANPFG
YNFVTITSGIDRTPVL
>LHCBM-Chlamydomonas_reinhardtii_3055@CR12G13520
MAAIMKSAVRSSVRPTVSGRSARVVPRAAIEWYGPDRPKFLGPFSEGDTPAYLTGEFPGD
YGWDTAGLSADPETFKRYRELELIHARWAMLGALGCITPELLAKNGIPFGEAVWFKAGAQ
IFAEGGLNYLGNENLIHAQSIIATLAFQVVVMGLAEAYRANGGPLGEGLDPLHPGGAFDP
LGLADDPDTFAELKVKEIKNGRLAMFSMFGFFVQAIVTGKGPIQNLDDHLANPTAVNAFA
YATKFTPSA
>LHCBM-Chlamydomonas_reinhardtii_3055@CR12G13640
MAAIMKSSVRSSVRSTVSSRSARVVPRAAIEWYGPDRPKFLGPFSEGDTPAYLTGEFPGD
YGWDTAGLSADPETFKRYRELELIHARWAMLGALGCITPELLAKNGIPFGEAVWFKAGAQ
IFAEGGLNYLGNENLIHAQSIIATLAFQVVVMGLAEAYRANGGPLGEGLDPLHPGGAFDP
LGLADDPDTFAELKVKEIKNGRLAMFSMFGFFVQAIVTGKGPIQNLDDHLANPTAVNAFA
YATKFTPSA
>LHCBM-Chlamydomonas_reinhardtii_3055@CR06G07460
MAFALASSRKALQVTCKATGKKTAAKAAAPKGAGIEFYGPNRAKWLGPYSENSTPAYLTG
EFPGDYGWDTAGLSADPETFKRYRELELIHARWAMLGALGIVTPELLQKNGVQFGEAVWF
KAGAQIFQEGGLNYLGNPSLIHAQNIVATLAVQVVLLGLVEGYRVNGGPAGEGLDPLYPG
ESFDPLGLADDPDTFAELKVKEIKNGRLAMFSSFGFFVQAIVTGKGPIQNLDDHLSNPGV
NNAFAFATKYTPSA
>LHCBM-Chlamydomonas_reinhardtii_3055@CR06G07470
MAFALATSRKALQVTCKATGKKTAAKAAAPKSSGVEFYGPNRAKWLGPYSENSTPAYLTG
EFPGDYGWDTAGLSADPETFKRYRELELIHARWAMLGALGCLTPELLAKSGTKFGEAVWF
KAGAQIFSEGGLDYLGNPSLVHAQNIVATLAVQVILMGLIEGYRVNGGPAGEGLDPLYPG
ESFDPLGLADDPDTFAELKVKEIKNGRLAMFSMFGFFVQAIVTGKGPVQNLDDHLANPGV
NNAFAFATKFTPSA
>LHCBM-Chlamydomonas_reinhardtii_3055@CR32G00130
MAFALSFSRKALQVSAKATGKKGTGKTAAKQAPASSGIEFYGPNRAKWLGPYSENATPAY
LTGEFPGDYGWDTAGLSADPETFKRYRELELIHARWAMLGALGCITPELLAKSGTQFGEA
VWFKAGAQIFSEGGLDYLGNPSLVHAQNIVATLAVQVILMGLVEGYRVNGGPAGEGLDPL
YPGESFDPLGLADDPDTFAELKVKEIKNGRLAMFSMFGFFVQAIVTGKGPIQNLDDHLSN
PTVNNAFAFATKFTPSA
>LHCBM-Chlamydomonas_reinhardtii_3055@CR03G01970
MMLSRTVVNVQAKLTKKGGAPKKAAPASAQKKTIREKAGWWSNGGNEKLSAFYGPDRGLW
LGPLSGTTPAYLTGEFPGDYGWDSAGLSADPETFKRYRELELIHARWAMLGALGCITPEL
LAKNGTPIVEPVWFKAGAQIFAEGGLDYLGNPGLVHAQSILATLAVQVILMGAIEGYRVN
GGPAGEGLDKLHPGGQFFDPLGLAEDPDAFAELKVKEIKNGRLAMFSMFGFFVQAIVTGK
GPLANLDEHLASPFTSNAFTYAQKFTPQ
>LHCBM-Chlamydomonas_reinhardtii_3055@CR23G00170
MAFALAKSSARAAVSRRSTVKVEARRTVKPASKASTPDSFWYGPERPLFLGAFTGEPPSY
LTGEFPGDYGWDTAGLSADPETFKRYRELELIHARWAMLGALGCIFPELLGSYGVPFGEA
VWFKAGAQIFQEGGLDYLGNPNLVHAQSILAILGTQVLLMGAIEGYRVNGGPLGEGLDKL
YPGGSFDPLGLADDPDTFAELKVKEIKNGRLAMFSMFGFFVQAIVTGKGPLQNLTT
>LHCB2.3-Arabidopsis_thaliana_3702@AT3G27690
MATSAIQHSSFAGQTTLKPSNDLLRKIGASNGGGRIIMRRTVKSTPQSIWYGPDRPKYLG
```

```
PFSENTPSYLTGEYPGDYGWDTAGLSADPETFAKNRELEVIHSRWAMLGALGCTFPEILS
KNGVKFGEAVWFKAGSQIFSEGGLDYLGNPNLIHAQSILAIWACQVVLMGFIEGYRIGGG
PLGEGLDPLYPGGAFDPLNLAEDPEAFSELKVKELKNGRLAMFSMFGFFVQAIVTGKGPI
ENLFDHIADPVANNAWAYATNFVPGK
>LHCB2.2-Arabidopsis_thaliana_3702@AT2G05070
MATSAIQQSSFAGQTALKPSSDLIQKVGVLGGGRVTMRRTVKSTPQSIWYGPDRPKYLGP
FSENTPSYLTGEYPGDYGWDTAGLSADPETFAKNRELEVIHSRWAMLGALGCTFPEILSK
NGVKFGEAVWFKAGSQIFSEGGLDYLGNPNLIHAQSILAIWAVQVVLMGFIEGYRIGGGP
LGEGLDPLYPGGAFDPLNLAEDPEAFSELKVKELKNGRLAMFSMFGFFVQAIVTGKGPIE
NLFDHLADPVANNAWSYATNFVPGK
>LHCB2.1-Arabidopsis_thaliana_3702@AT2G05100
MATSAIQQSSFAGQTALKPSNELLRKVGVSGGGRVTMRRTVKSTPQSIWYGPDRPKYLGP
FSENTPSYLTGEYPGDYGWDTAGLSADPETFAKNRELEVIHSRWAMLGALGCTFPEILSK
NGVKFGEAVWFKAGSQIFSEGGLDYLGNPNLIHAQSILAIWAVQVVLMGFIEGYRIGGGP
LGEGLDPLYPGGAFDPLNLAEDPEAFSELKVKELKNGRLAMFSMFGFFVQAIVTGKGPIE
NLFDHLADPVANNAWSYATNFVPGN
>LHCB1.3-Arabidopsis_thaliana_3702@AT1G29930
MAASTMALSSPAFAGKAVKLSPAASEVLGSGRVTMRKTVAKPKGPSGSPWYGSDRVKYLG
PFSGESPSYLTGEFPGDYGWDTAGLSADPETFARNRELEVIHSRWAMLGALGCVFPELLA
RNGVKFGEAVWFKAGSQIFSDGGLDYLGNPSLVHAQSILAIWATQVILMGAVEGYRVAGN
GPLGEAEDLLYPGGSFDPLGLATDPEAFAELKVKELKNGRLAMFSMFGFFVQAIVTGKGP
IENLADHLADPVNNNAWAFATNFVPGK
>LHCB1.5-Arabidopsis_thaliana_3702@AT2G34420
MASSTMALSSPAFAGKAVKPAASDVLGSGRVTMRKTVAKPKGPSGSPWYGSDRVKYLGPF
SGEPPSYLTGEFPGDYGWDTAGLSADPETFARNRELEVIHSRWAMLGALGCVFPELLARN
GVKFGEAVWFKAGSQIFSDGGLDYLGNPSLVHAQSILAIWATQVILMGAVEGYRVAGDGP
LGEAEDLLYPGGSFDPLGLATDPEAFAELKVKELKNGRLAMFSMFGFFVQAIVTGKGPLE
NLADHLADPVNNNAWAFATNFVPGK
>LHCB1.4-Arabidopsis_thaliana_3702@AT2G34430
MAASTMALSSPALTGKAVKLSPAASEVFGTGRITMRKASKPTGPSGSPWYGSDRVKYLGP
FSGEPPSYLTGEFPGDYGWDTAGLSADPETFARNRELEVIHSRWAMLGALGCVFPELLAR
NGVKFGEAVWFKAGSQIFSDGGLDYLGNPSLVHAQSILAIWATQVILMGAVEGYRVAGDG
PLGEAEDLLYPGGSFDPLGLATDPEAFAELKVKELKNGRLAMFSMFGFFVQAIVTGKGPL
ENLADHLADPVNNNAWAFATNFVPGK
>LHCB3-Arabidopsis_thaliana_3702@AT5G54270
MASTFTSSSSVLTPTTFLGQTKASSFNPLRDVVSLGSPKYTMGNDLWYGPDRVKYLGPFS
VQTPSYLTGEFPGDYGWDTAGLSADPEAFAKNRALEVIHGRWAMLGAFGCITPEVLQKWV
RVDFKEPVWFKAGSQIFSEGGLDYLGNPNLVHAQSILAVLGFQVILMGLVEGFRINGLDG
VGEGNDLYPGGQYFDPLGLADDPVTFAELKVKEIKNGRLAMFSMFGFFVQAIVTGKGPLE
NLLDHLDNPVANNAWAFATKFAPGA
>LHCB7-Chlamydomonas_reinhardtii_3055@CR02G07970
MSLAMKTATPIMTGQTSALWQKSWVLRASQEPEGRSSEQPAPADAPETLTTRAAKLAAEI
VASPIFYLVAGLLAIKLVASTGEDGAAIFIFAALPITALTALSKSSVGKQVQEQLEAKLP
ELQAEADKTRAAHAEARRQSKWYGPSRPRLPGPLGSAEHLIGAIAGDAGFDPLGLSKEQA
SFAKYQEAELLHARWAMLGVVGCLVPELLALRGVDLGEPVWWKVGASKLNSDLTLNWGGI
EGFRIAGKQGIGLIAACQAVLMGGPEYARYVGIRSLEPVGVFLPGDQNYPGGGPFDPLNY
AADADGFVEQAVKEVKNGRLAMLAMLGFFVQAAVTRVGPVQNVLDFVADPAHNNIFWNLA
HLQL
>LHCB7-Arabidopsis_thaliana_3702@AT1G76570
MALFQEKLSSLSSSYSSIHSLPRILVSKPRNRIAVTKSRSICRASWQELAGVLVFSAIPF
TAVKAIANSSIGVSLRRRLEEKKKEAVENSSRFKSKAQEARNDSKWYGKERPRWFGPIPY
DYPPYLTGELPGDYGFDIAGLGKDRLTFDKYFNFEILHARWAMLAALGALIPEVFDLTGT
FHFAEPVWWRVGYSKLQGETLEYLGIPGLHVAGSQGVIVIAICQVLLMVGPEYARYCGIE
ALEPLGIYLPGDINYPGGTLFDPLNLSEDPVAFEDLKVKEIKNGRLAMVAWLGFYAQAAF
TGKGPVQNLVDHVSDPLHNNLIAMLQT
>LHCA5-Ostreococcus_tauri_70448@OT05G03910
MIAQRSALSARSIATPKATRRSAVVAAAADRKMWYAPAHLDGSIAGDYGFDPLGLGTNPD
RLKYYQEAELMNARWAMMAVAGIVGTEVAGIEPRWWEAGTEDYGFPPAALLAIQFPVMGY
LENKRIQGWMATGSSGVNETFPFDPMGMGSKDANMKLKEIKNGRAAMIAFVGIVVQAIVY
REGPVAALKDHISNPFGCNMATNIMNIPVNLA
>LHCA-Chlamydomonas_reinhardtii_3055@CR07G06850
MIAAKSQVALGRRAPVRGQRVVAAASARPTWLPGLNPPAHLKGALAGDNGFDPLGLGQDE
GRLKWYAEAEKTNGRWAMMAVAGILGQELLGVTPAWWEAGAKEYDIPAQALTPIEFIVMG
FLEIKRYQGFKQTGTSGFINSFPFDPAGMNSPSMATKEVKNGRLAMVAFIGFCVQALATR
TQPIEGLTAHLADPFGKNITYYLTHLPETLGSA
>LHCA2-Chlamydomonas_reinhardtii_3055@CR12G05150
MAMLLKSRVSAGVSRPSRATVRVSASTRPMWYPGATAPAHLDGSMLGDYGFDPLRLGVNK
```

```
DNLKWFREAELTNGRWAMAAVVGILFTDAVGLPKFWTAGAEKYALDNQTLALIEVAVFAV
LEGKRYEIYKKTGETGFLSFAPFDPMGMKSEEMKLKELKNGRLAMLAFLGFCSQAAVYGK
GPIETLQLHLADPGHNNIYTSSVGPETAVTVAVLCVLPMIIEATKTLNPGKESVPYFPWN
EPWNKV
>LHCA-Ostreococcus_tauri_70448@OT03G04770
MAGDYGFDPLRLGANVETLPYLQEAELMNGRWAMAATAGILFTDATGLPKWWEAGAADYG
YDFQTLVAFQVVVMGVLEAFRVRGLMKTGESGVLNMFPFDPAGLDAPDKRVKEVKNGRLA
MVAFLGMVSSYAVTGLSPLEALEAHMANPQAVNLFTSAVGGESVAFIAFLSCAPTFLLAQ
KTLGDGKEEFRPIPW
>LHCA-Ostreococcus_tauri_70448@OT02G07190
MFAATRSFTQTVAPKTSVKARARVTARASARPMWFPGAEAPAHLRGDLPCDYGFDPLNLG
EKPDNLARYREAELMHARWAMMGVAGAVGVEIAGQGDWASAQPTTWDATAKYLGNETHAP
LFAVIGVNGVLVAFAESQRQAATGEARLYPGGSFDPAGLSKGKDFETLKRKELANGRVAM
MAFFGIMAQHQADPSGPGPVKQLANHLADPWHVNVCTNPSAIPWL
>LHCA1-Arabidopsis_thaliana_3702@AT3G54890
MASNSLMSCGIAAVYPSLLSSSKSKFVSAGVPLPNAGNVGRIRMAAHWMPGEPRPAYLDG
SAPGDFGFDPLGLGEVPANLERYKESELIHCRWAMLAVPGILVPEALGYGNWVKAQEWAA
LPGGQATYLGNPVPWGTLPTILAIEFLAIAFVEHQRSMEKDPEKKKYPGGAFDPLGYSKD
PKKLEELKVKEIKNGRLALLAFVGFCVQQSAYPGTGPLENLATHLADPWHNNIGDIVIPF
N
>LHCA3-Chlamydomonas_reinhardtii_3055@CR10G08050
PPPDAAGRLLILPSVSSLSYLDGTLPGDMGFDPLGLFDPTNGSAGFMSQRWLHTAEVVHG
RWAMLGAAGCLAPEYLAHEKVIPKATGLLWFKTGFLPPVSAGFDYGLPVTALFLIQMGLM
GTVESLRWAELSNPGSLASNPLVPQPLARLIGTTRLGGGAYPGGLMFNPLGLGATRPAVM
REYQQSEIRHGRLAMVACLGYAAQAVVTGKGPYTNWVAHLEDPDIENVVVLLDKLFA
>LHCA3-Arabidopsis_thaliana_3702@AT1G61520
MAAQALVSSSLTSSVQTARQIFGSKPVASASQKKSSFVVKAAATPPVKQGANRPLWFASS
QSLSYLDGSLPGDYGFDPLGLSDPEGTGGFIEPRWLAYGEIINGRFAMLGAAGAIAPEIL
GKAGLIPAETALPWFQTGVIPPAGTYTYWADNYTLFVLEMALMGFAEHRRLQDWYNPGSM
GKQYFLGLEKGLAGSGNPAYPGGPFFNPLGFGKDEKSLKELKLKEVKNGRLAMLAILGYF
IQGLVTGVGPYQNLLDHLADPVNNNVLTSLKFH
>LHCA3-Chlamydomonas_reinhardtii_3055@CR18G01380
MMLTKSAQAAFSGKVARPAKANRARLVCRAEEKSIAKVDRSKDQLYVGASQSSLAYLDGS
LPGDFGFDPLGLLDPVNSGGFIEPKWLQYSEVIHARWAMLGAAGCIAPEVLGAAGLIPDA
TNIKWFESGVIPPAGSYNGYWADPYTIFFVEIVAMQFAELRRLQDFRYPGSMGQQYFLGL
EAIFKGSGDAAYPGGPFFNLFNLGKTEAAMKELKLKEIKNGRLAMLAMLGYGAQAVMTGK
GPFQNLVEHLADPVNNNILTNFAGRVSGSSQPWRPHGWWRRRYRSVAALALIRNRSVC
>LHCA3-Ostreococcus_tauri_70448@OT18G01370
MFATMNAIKITPFAKRQARRNVSVRAEEAEAAPAPVAKKTRNTEDLAPLYVLGNSEQSLS
YLDGSLPGDYGFDPLGLSDPEGAGGFVNPKWLAYSELIHGRWAMLGVAGMVAPEVLGGMG
IIPQETGLVWFKAGMIPAQGTYDYWASPFTIFWINAFLMNIAELRRAQDYWNPGSMGKQD
FAGLEKMLGGSGDPAYPGGFFNFMKQGEKDMAAMKTKEIKNGRLAMMACFGCGAQACMTG
EGPVKNLVDHVIDPFGHNLLVNFSQIGGVSPF
>LHCA2/prasino-Ostreococcus_tauri_70448@OT09G04370
MWVDRGGDIGAHRRLKARATTTRASRRSVVVSAEAGAERPVWYPGKAPAPHLDGSLPGDF
GFDPLSLSADPEMRKWMVQAELQHARWAMLGVAGAVAPELLTKIGVADLPNWVDAGTYQY
WAPAGPLFFIQMAMFNWAEVRRWQDMKNPGSMNTDPLFGYNSNDTNTDVGYPGGLFDKLG
YAKDPAKAKELKLKEIKNGRLAMVAFLGICAQYVQTGQGPVENLFSHIASPGSVGYFGSQ
GL
>LHC/prasino-Ostreococcus_tauri_70448@OT02G05720
MSAQFASTAVVGLRAPKTTKRSQRSVSVRASTQPMWFPGMDAPQHLKGELPGDYGFDPLN
LGKEPKDLEWYVQAELQHGRWAMLGVAGAAAPEILTKMGISDLPNWHDAPNYQGYFTDAT
TLFWVQMLMMNWAEVRRWQDMRKPGSVNEDPAFSGNKLPSGIVGYPGGIFDPLGYAKGDL
NKLKAKEIANGRLAMVAFAGIMVQYDHTGVGPVANLVAHMSDPAHNNVFQAKFIGF
>LHCA2-Arabidopsis_thaliana_3702@AT3G61470
MASSLCASSAIAAISSPSFLGGKKLRLKKKLTVPAVSRPDASVRAVAADPDRPIWFPGST
PPEWLDGSLPGDFGFDPLGLSSDPDSLKWNVQAEIVHCRWAMLGAAGIFIPEFLTKIGIL
NTPSWYTAGEQEYFTDKTTLFVVELILIGWAEGRRWADIIKPGSVNTDPVFPNNKLTGTD
VGYPGGLWFDPLGWGSGSPAKLKELRTKEIKNGRLAMLAVMGAWFQHIYTGTGPIDNLFA
HLADPGHATIFAAFTPK
>LHCA6-Arabidopsis_thaliana_3702@AT1G19150
MAFAIASALTSTLTLSTSRVQNPTQRRPHVASTSSTGGRLMRERLVVVRAGKEVSSVCEP
LPPDRPLWFPGSSPPEWLDGSLPGDFGFDPLGLGSDPDTLKWFAQAELIHSRWAMLAVTG
IIIPECLERLGFIENFSWYDAGSREYFADSTTLFVAQMVLMGWAEGRRWADLIKPGSVDI
EPKYPHKVNPKPDVGYPGGLWFDFMMWGRGSPEPVMVLRTKEIKNGRLAMLAFLGFCFQA
TYTSQDPIENLMAHLADPGHCNVFSAFTSH
>LHCA6-Chlamydomonas_reinhardtii_3055@CR13G07600
```

```
MMLVAKNAVAARPSARSARRSVVAKASSRPLWLPGSTPPAHLKGDLPGDFGFDPLGLGAN
AESLKWFKESELVHSRWAMAAVAGILVQEIVRPDVFWYNAGKEVESPLGPLGLLAVEFFL
MHWVEVRRWQDLRKPGSVDQDPIFSQYKLPPHEVGYPGGVFAPFIPGDLAELKVKEIKNG
RLAMLAFVGFVMAAQVTGKGPIAALQEHLADPWGTTIFSKAAVVPGQAVAPPCKIPASVS
YKGIEIPTPCFLQGLWP
>LHCA5-Chlamydomonas_reinhardtii_3055@CR10G01820
MAALMQKSALSRPACSTRSSRRAVVVRAAADRKLWAPGVVAPEYLKGDLAGDYGWDPLGL
GADPTALKWYRQSELQHARWAMLGVAGVLVQEIVKPDVYFYEAGLPQNLPEPFTNINMGG
LLAWEFILMHWVEVRRWQDYKNFGSVNEDPIFKGNKVPNPEMGYPGGIFDPFGFSKGNLK
ELQTKEIKNGRLAMIAYMAFILQAQATGKGPLAALSAHLSNPFGNNILKNIGTCTVPHSV
DVQGLTIPLTCLWPGSQ
>LHCA4-Arabidopsis_thaliana_3702@AT3G47470
MATVTTHASASIFRPCTSKPRFLTGSSGRLNRDLSFTSIGSSAKTSSFKVEAKKGEWLPG
LASPDYLTGSLAGDNGFDPLGLAEDPENLKWFVQAELVNGRWAMLGVAGMLLPEVFTKIG
IINVPEWYDAGKEQYFASSSTLFVIEFILFHYVEIRRWQDIKNPGSVNQDPIFKQYSLPK
GEVGYPGGIFNPLNFAPTQEAKEKELANGRLAMLAFLGFVVQHNVTGKGPFENLLQHLSD
PWHNTIVQTFN
>LHCA4-Chlamydomonas_reinhardtii_3055@CR10G07510
MAFVLAKSSAFGVAAKPVSRRSSVAVKASAVPENVKEAREWIDAWKSKSGGAKRDAALPS
WMPGADLPGYLNGTLPGDFGFDPLYLGQDPVKLKWYAQAELMNARFAMLAVAGILVPELL
SNIGFSWPGAGVAWYDAGKFEYFAPASSLFGVQMLLFAWVEIRRYQDFVKPGSANQDPIF
TNNKLPDGNEPGYPGGIFDPFGWSKGDIKSLKLKEIKNGRLAMLAFAGFIGQAYTTGTTP
LKNLSTHLADPWSTTVWQNDL
>LHCA5-Arabidopsis_thaliana_3702@AT1G45474
MAVVLRGGITGGFLHHRRDASSVITRRISSVKAAGGGINPTVAVERATWLPGLNPPPYLD
GNLAGDYGFDPLGLGEDPESLKWYVQAELVHSRFAMLGVAGILFTDLLRTTGIRNLPVWY
EAGAVKFDFASTKTLIVVQFLLMGFAETKRYMDFVSPGSQAKEGSFFFGLEAALEGLEPG
YPGGPLLNPLGLAKDVQNAHDWKLKEIKNGRLAMMAMLGFFVQASVTHTGPIDNLVEHLS
NPWHKTIIQTLFTSTS
>LHCA8-Chlamydomonas_reinhardtii_3055@CR06G05040
MALTMKRSGVAARSASSRKSVVTCVARQSWLPGSQIPAHLDTPAAQALAGNFGFDPLGLG
KDPVALRWYQQAELIHCRTAMAGVAGILIPGLLTKAGALNVPEWYDAGKVAIENSFAPWG
SLLAVQLFLCGFVEAKRWQDIRKPGSQGEPGSFLGFEASLKGTSELGYPGGPFDPLGLSK
EADKWADWKLKEVKNGRLAMLAFLGFVAQKYATGAGPVDNLAAHLKDPWHVNYATNGVSL
PFL
>LHCA7-Chlamydomonas_reinhardtii_3055@CR16G08380
MALSMIAQRRAGAFSARQAPRAVRAQAAVRPVWFPGNPPPAHLDGSLAGDYGFDPLFLGQ
EPQTLKWYVQAELVHGRFAMLGAAGIILTSIGAKVGLGFPEWYDAGKVVVEKNNIDFPTL
MVIQFYLMGWAETKRWYDFKNPGSQADGSFLGFTEEFKGLENGYPGGRFFDPMGLSRGDA
AKYQEYKQKEVKNGRLAMIACLGFAAQYAATGKGPLDNLV
>CP24-Arabidopsis_thaliana_3702@AT1G15820
MAMAVSGAVLSGLGSSFLTGGKRGATALASGVGTGAQRVGRKTLIVAAAAAQPKKSWIPA
VKGGGNFLDPEWLDGSLPGDFGFDPLGLGKDPAFLKWYREAELIHGRWAMAAVLGIFVGQ
AWSGVAWFEAGAQPDAIAPFSFGSLLGTQLLLMGWVESKRWVDFFNPDSQSVEWATPWSK
TAENFANYTGDQGYPGGRFFDPLGLAGKNRDGVYEPDFEKLERLKLAEIKHSRLAMVAML
IFYFEAGQGKTPLGALGL
>CP29-Ostreococcus_tauri_70448@OT16G02550
MFKSKAPTKQVAVKKAPVKTAVKKTARPVRKGPAPKAGDECLWRPGAVRPAWLDGSLPGD
RGFDPFGLGKPAEYLQFDLDALDGSAARNPSGNVIGKLKKADSKPKARTIVPYDEAFDIM
RFRECELLHSRWAMLGLVGAAAAEAATGITWADAGKIELEQPQYLGFPIQLTVTQLTIIE
VLAMGYLEIARSAELDQNKRMYPGGAFDPLGYVAKADAQEEFRLKTAELKHGRLAMAGML
IISIEAATGRGTLPGSW
>CP29-Chlamydomonas_reinhardtii_3055@CR17G05090
MVFKFPTPPGTQKKAGTTATKPAPKATTKKVATSTGTRSGGVGYRKYQGDALWLPNTTRP
EWLDGSLPGDRGFDPLGLSKPSEFVVIGVDENDQNAAKNNKGSVEAIVQATPDEVSSENR
LAPYSEVFGLARFRECELIHGRWAMLACLGALVAEATTGVSWVEAGKVELDGASYAGLSL
PFSITQLIWIEVILVGGAEFYRNSETNPEKRCYPGGVFDPLKLASEDEERAFRLKTAEIK
HARLAMVSFFGYGVQALSTGEGALGSLAKFADGLNNGKGL
>CP29.3-Arabidopsis_thaliana_3702@AT2G40100
MATTTAAAASGIFGIRIQDPRPGTGRVQARFGFSFGKKKPAPPPKKSRQVQDDGDRLVWF
PGANPPEWLDGSMIGDRGFDPFGLGKPAEYLQYDFDGLDQNLAKNVAGDIIGIIQESSEI
KPTPFQPYTEVFGIQRFRECELIHGRWAMLGTLGAIAVEALTGIAWQDAGKVELVEGSSY
LGQPLPFSLTTLIWIEVLVVGYIEFQRNSELDPEKRIYPGGYFDPLGLAADPEKLDTLKL
AEIKHSRLAMVAFLIFALQAAFTGKGPVSFLATFNN
>CP29.2-Arabidopsis_thaliana_3702@AT3G08940
MAATSTAAAASSIMGTRVVSDISSNSSRFTARFGFGTKKASPKKAKTVISDRPLWFPGAK
SPEYLDGSLVGDYGFDPFGLGKPAEYLQFDLDSLDQNLAKNLYGEVIGTRTEAVDPKSTP
```

FQPYSEVFGLQRFRECELIHGRWAMLATLGAITVEWLTGVTWQDAGKVELVDGSSYLGQP
LPFSISTLIWIEVLVIGYIEFQRNAELDSEKRLYPGGKFFDPLGLASDPVKKAQLQLAEI
KHARLAMVGFLGFAVQAAATGKGPLNNWATHLSDPLHTTIIDTFSSS
>CP29.1-Arabidopsis_thaliana_3702@AT5G01530
MAATSAAAAAASSIMGTRVAPGIHPGSGRFTAVFGFGKKKAAPKKSAKKTVTTDRPLWYP
GAISPDWLDGSLVGDYGFDPFGLGKPAEYLQFDIDSLDQNLAKNLAGDVIGTRTEAADAK
STPFQPYSEVFGIQRFRECELIHGRWAMLATLGALSVEWLTGVTWQDAGKVELVDGSSYL
GQPLPFSISTLIWIEVLVIGYIEFQRNAELDSEKRLYPGGKFFDPLGLAADPEKTAQLQL
AEIKHARLAMVAFLGFAVQAAATGKGPLNNWATHLSDPLHTTIIDTFSSS
>LHCP1/prasino-Ostreococcus_tauri_70448@OT01G04850
ANKFGSLAPPDLYPEFGTYPGGGESPVIPFGSEKNAEREIIHGRWAMLGVTGAWAAENGT
GIPWFTAGTLCTPDDCTAVADKFPGAVAPLAPAGSGYPNFWAVLAIEIFLVGSAECYRTG
LFENPFPELTQGDVTPGGRFDPLGFAEAGDLEELKIKELKHSRLAMFAWLGCIMQALATQ
EGPIANWTAHVADPIHANVLTNAAKGFKFY
>LHCP/prasino-Ostreococcus_tauri_70448@OT16G00400
MSALLASSFVSRVAAFKAQKVQNKSVSTTVKADIYPEFGTYPGGGESPIIPFGSEKNAER
EVIHGRWAMLGVTGAWAAENGTGIPWFTAGTLCTPDDCTAVADKFPGAVAPLAPEGSGYP
SFWNVLIIEIVLVGAAEAYRTGISDSPFDDGLTVGDVNPGGRFDPLGLAESGDLEELKIK
ELKHCRLSMFAWLGCIFQALATQEGPIANWQSHVADPVHSNVLTNAAKGFGFY
>LHCSR-Chlamydomonas_reinhardtii_3055@CR08G01590
MAMMMRKAAAVPASSRRSVAVNSVSGKRTVSGKAGAPVPEDVLAYAKTLPGVTAPFDNVF
DPAGFLATASVKDVRRWRESEITHGRVAMLAALGFIVGEQLQDFPLFFNFDGRVSGPAIY
HFQQIGQGFWEPLLIAIGVAESYRVAVGWATPTGTGFNSLKDDYEPGDLGFDPLGLKPTD
PEELKTLQTKELNNGRLAMIAIAAFVAQELVEQTEIFEHLVLRFEKEVILELEDVERDLG
LPLTPLPDNLKAI
>LHCSR-Chlamydomonas_reinhardtii_3055@CR08G01950
MLANVVSRKASGLRQTPARATVAVKSVSGRRTTAAEPQTAAPVAAEDVFAYTKNLPGVTA
PFEGVFDPAGFLATASIKDVRRWRESEITHGRVAMLAALGFVVGEQLQDFPLFFNWDGRV
SGPAIYHFQQIGQGFWEPLLIAIGVAESYRVAVGWATPTGTGFNSLKDDYEPGDLGFDPL
GLKPTDPEELKVMQTKELNNGRLAMIAIAAFVAQELVEQTEIFEHLALRFEKEAILELDD
IERDLGLPVTPLPDNLKSL
>LHCSR-Ostreococcus_tauri_70448@OT05G01790
MPMLVTMMAAKPQVRASGADVFQKGKNYANIQEATAAGLIGASAPFPEGLDILGFCNGVD
GATLQRYREAEVAHSRVCMLATVGFLVGEQVEGSSFLFDAQVTGPAVNHFQQVPLPFWFA
IGSVIAIAESARVQAGWQDPGQSDKLFLLKDGYTPGDLSFDPLGLGNGKSAEELDALRLK
ELNNGRLAMIAISGMVAQELVDGLNILPADIALELGNGDLQKMEQACAGKVDEAACAKAF
EASLEAASRM
Haut du formulaire
>XP_001700736.1 uncharacterized protein CHLRE_05g243800v5 [Chlamydomonas reinhardtii]
MASITCSSGKAAVVSRAKTARPVARRMSVVASAAPQGNASRRELLGLSAVAASLLLSSRPAHAIFGFADD
NTALFDTYTAETSAILDKVKVTLALDKDDPAKEDSVKGLRKDINNWVAKYRREPKVSGKPSFGNTYSALN
ALAGHFNSFGATAPIPKKRLERLQKELDDATLLLTRNR

>NP_563687.1 photosystem II family protein [Arabidopsis thaliana]
MASASATATLLKPNLPPHKPTIIASSVSPPLPPPRRNHLLRRDFLSLAATSTLLTQSIQFLAPAPVSAAE
DEEYIKDTSAVISKVRSTLSMQKTDPNVADAVAELREASNSWVAKYRKEKALLGKASFRDIYSALNAVSG
HYVSFGPTAPIPAKRKARILEEMETAEKALTRGR

>CAL57287.1 Photosystem II Pbs27 [Ostreococcus tauri]
MNKARATFNVSASALFKRVAEDASRAESAPSTSRRFVALSLALAAASTASAPGANAQKFKEVCDPTEDGA
ECRANILAQDGIDVGAYEVGGSRSFKAANPTGAKTSLTTYQSDTLEFVGEVEALLAMDVYDGSREKAVAA
FQKKSNNWSGKYAPGGSSKMASGRAFYNALNQLAGHYSFNGLAPIPKSRLDVVETNIAKTKELIAAGR

>WP_041425825.1 MULTISPECIES: photosystem II protein Psb27 [unclassified Synechocystis]
MSFLKNQLSRLLALILVVAIGLTACDSGTGLTGNYSQDTLTVIATLREAIDLPQDAPNRQEVQDTARGQI
NDYISRYRRKGDAGGLKSFTTMQTALNSLAGYYTSYGARPIPEKLKKRLQLEFTQAERSIERGV

>XP_001693657.1 uncharacterized protein CHLRE_13g562850v5 [Chlamydomonas reinhardtii]
MLVAKHFVAPSVAPRPKAAPAPRSSLARKAVATPVLAATFKPPTVAETKAKFLSGYNKPIASIYSTVLQE
LLVQQHFMRYSKNYQYNPIFALGFVSVYEQILESLSAEERGAIFKAYVDALGEDADKYKRDASALEQAAN
GLTPESLTPNADGNEVQKALASISSASAAGAFSYNKFVAIGLFRLLELSGAKEPSALEKLVKAVGVKPEA
VNRDLLMYKGVLSKLAAAKELMREFVEREKRKQAERDAAKAAKAAGTESSPAQVQA

>NP_565491.1 photosystem II reaction center PSB29 protein [Arabidopsis thaliana]
MAATAISSLSFPALGQSDKISNFASSRPLASAIRICTKFSRLSLNSRSTSKSLIHCMSNVTADVPPVSET
KSKFLKAYKRPIPSIYNTVLQELIVQQHLMRYKKTYRYDPVFALGFVTVYDQLMEGYPSDQDRDAIFKAY
IEALNEDPKQYRIDAQKMEEWARSQTSASLVDFSSKEGDIEAVLKDIAGRAGSKEGFSYSRFFAVGLFRL

```
LELASATDPTVLDKLCASLNINKKSVDRDLDVYRNLLSKLVQAKELLKEYVEREKKKQGERAQSQKANET
ISKCLGDTLYNPSFLVERKS

>CEF97146.1 Photosystem II Psp29, biogenesis [Ostreococcus tauri]
MRAQLARVTAPKALGASRKAPTSARARCAAITRAQSGIYGVTVADTKRKFLEAYPYPIPSVWATVVQELL
VQGHFQKYNKKSEYNELASLGFVSVYDQLFEGFPSEEEKGKIFNAFLGALDEDAVRTRADAETLGAFATS
ANGVEGLKENAIFAKLAAKSAEGTLLYTKYIAIGMFRMLELAKATDPAALEALVTAGGLSMSKVSGDLSM
YKGLLSKLAAAKELQEEFLEREKRKTAERLAKKAEAEQKAEAEA

>BAA18023.1 sll1414 [Synechocystis sp. PCC 6803]
MTKIRTVSDAKRKFFTHYSRPISSIYRRFVEELLVEMHLLSVNIDFTYDPIFALGIVTSFNSFMQGYQPA
EQLPAIFNALCHGVDQNPDQVRQDAKNVAASAHHIGLDAWVTAAASEQASGDNLLLNTLTGIHQRHKFKY
SRLFAIGLYTLLADQDPEVKDNDEKRQDYLTRLSELLDLSLDKVVKDLDLYRSNLEKVDQLLKVLEDAAE
AERKKKEKQAASTTPAIEEAPVTTAESSES

>NP_958413.1 photosystem II protein D1 (chloroplast) [Chlamydomonas reinhardtii]
MTAILERRENSSLWARFCEWITSTENRLYIGWFGVIMIPCLLTATSVFIIAFIAAPPVDIDGIREPVSGS
LLYGNNIITGAVIPTSNAIGLHFYPIWEAASLDEWLYNGGPYQLIVCHFLLGVYCYMGREWELSFRLGMR
PWIAVAYSAPVAAASAVFLVYPIGQGSFSDGMPLGISGTFNFMIVFQAEHNILMHPFHMLGVAGVFGGSL
FSAMHGSLVTSSLIRETTENESANEGYRFGQEEETYNIVAAHGYFGRLIFQYASFNNSRSLHFFLAAWPV
IGIWFTALGLSTMAFNLNGFNFNQSVVDSQGRVLNTWADIINRANLGMEVMHERNAHNFPLDLASTNSSS
NN

>NP_051039.1 photosystem II protein D1 (chloroplast) [Arabidopsis thaliana]
MTAILERRESESLWGRFCNWITSTENRLYIGWFGVLMIPTLLTATSVFIIAFIAAPPVDIDGIREPVSGS
LLYGNNIISGAIIPTSAAIGLHFYPIWEAASVDEWLYNGGPYELIVLHFLLGVACYMGREWELSFRLGMR
PWIAVAYSAPVAAATAVFLIYPIGQGSFSDGMPLGISGTFNFMIVFQAEHNILMHPFHMLGVAGVFGGSL
FSAMHGSLVTSSLIRETTENESANEGYRFGQEEETYNIVAAHGYFGRLIFQYASFNNSRSLHFFLAAWPV
VGIWFTALGISTMAFNLNGFNFNQSVVDSQGRVINTWADIINRANLGMEVMHERNAHNFPLDLAAVEAPS
TNG

>YP_717233.1 photosystem II protein D1 (chloroplast) [Ostreococcus tauri]
MTATLERRANATVWGRFCSWITSTENRLYIGWFGVLMIPTLLTATSVFIVAFIAAPPVDIDGIREPVSGS
LMYGNNIISGAVIPTSNAIGLHFYPIWEAASLDEWLYNGGPYQLIVCHFFIGICSYMGREWELSFRLGMR
PWIAVAYSAPVAAATAVFLIYPIGQGSFSDGMPLGISGTFNFMIVFQAEHNILMHPFHMLGVAGVFGGSL
FSAMHGSLVTSSLIRETTENESANAGYKFGQEEETYNIVAAHGYFGRLIFQYASFNNSRSLHFFLAIWPV
MGIWFTALGISTMAFNLNGFNFNQSVVDSNGRVINTWADIVNRANLGMEVMHERNAHNFPLDLASVEAPS
INA

>AIE73589.1 photosystem II protein D1 (PsbA) [Synechocystis sp. PCC 6714]
MTTTLQQRESASLWEQFCQWVTSTNNRIYVGWFGTLMIPTLLTATTCFIIAFIAAPPVDIDGIREPVAGS
LLYGNNIISGAVVPSSNAIGLHFYPIWEAASLDEWLYNGGPYQLVVFHFLIGIFCYMGRQWELSYRLGMR
PWICVAYSAPVSAATAVFLIYPIGQGSFSDGMPLGISGTFNFMIVFQAEHNILMHPFHMLGVAGVFGGSL
FSAMHGSLVTSSLVRETTEVESQNYGYKFGQEEETYNIVAAHGYFGRLIFQYASFNNSRSLHFFLGAWPV
IGIWFTAMGVSTMAFNLNGFNFNQSILDSQGRVIGTWADVLNRANIGFEVMHERNAHNFPLDLASGEQAP
VALTAPAING

>NP_958388.1 photosystem II 47 kDa protein (chloroplast) [Chlamydomonas reinhardtii]
MGLPWYRVHTVVINDPGRLISVHLMHTALVSGWAGSMALFEISVFDPSDPVLNPMWRQGMFVLPFMTRLG
ITQSWGGWTISGETATNPGIWSYEGVAAAHIILSGALFLASVWHWTYWDLELFRDPRTGKTALDLPKIFG
IHLFLSGLLCFGFGAFHVTGVFGPGIWVSDPYGLTGRVQPVAPSWGADGFDPYNPGGIASHHIAAGILGV
LAGLFHLCVRPSIRLYFGLSMGSIETVLSSSIAAVFWAAFVVAGTMWYGSAATPIELFGPTRYQWDQGFF
QQEIQKRVQASLAEGASLSDAWSRIPEKLAFYDYIGNNPAKGGLFRTGAMNSGDGIAVGWLGHASFKDQE
GRELFVRRMPTFFETFPVLLLDKDGIVRADVPFRKAESKYSIEQVGVSVTFYGGELDGLTFTDPATVKKY
ARKAQLGEIFEFDRSTLQSDGVFRSSPRGWFTFGHVCFALLFFFGHIWHGARTIFRDVFAGIDDDINDQV
EFGKYKKLGDTSSLREAF

>NP_051084.1 photosystem II 47 kDa protein (chloroplast) [Arabidopsis thaliana]
MGLPWYRVHTVVLNDPGRLLAVHIMHTALVAGWAGSMALYELAVFDPSDPVLDPMWRQGMFVIPFMTRLG
ITNSWGGWNITGGTITNPGLWSYEGVAGAHIVFSGLCFLAAIWHWVYWDLEIFCDERTGKPSLDLPKIFG
IHLFLSGVACFGFGAFHVTGLYGPGIWVSDPYGLTGKVQPVNPAWGVEGFDPFVPGGIASHHIAAGTLGI
LAGLFHLSVRPPQRLYKGLRMGNIETVLSSSIAAVFFAAFVVAGTMWYGSATTPIELFGPTRYQWDQGYF
QQEIYRRVSAGLAENQSLSEAWAKIPEKLAFYDYIGNNPAKGGLFRAGSMDNGDGIAVGWLGHPVFRNKE
GRELFVRRMPTFFETFPVVLVDGDGIVRADVPFRRAESKYSVEQVGVTVEFYGGELNGVSYSDPATVKKY
ARRAQLGEIFELDRATLKSDGVFRSSPRGWFTFGHASFALLFFFGHIWHGARTLFRDVFAGIDPDLDAQV
EFGAFQKLGDPTTKRQAV
```

```
>YP_717207.1 PsbB (chloroplast) [Ostreococcus tauri]
MTRLGVTNSWGGWTISGESTSNPGLWSYEGVAASHIILSGLLFLAAIWHWVFWDLELFRDPRTQQPALDL
PKIFGIHLFLSGVLCFGFGAFHVTGLFGPGIWVSDPYGLTGAVEPVAPAWGAEGFDPYNPGGIAAHHIAA
GIVGILAGLFHLSVRPPQRLYKALRMGNVETVLSSSIAAVFWAAFVVGGTMWYGCAATPIELFGPTRYQW
DQGFFQQEIEKRVQTSVAGGASLSTAWSTIPEKLAFYDYIGNNPAKGGLFRSGPMDNGDGIAAGWLGHAT
FTDKNGRELFVRRMPTFFETFPVILIDGDGVVRADVPFRRAESKYSIEQVGVNVTFYGGELDGLTFTDPA
TVKKYARRAQLGEVFEFDRATLQSDGVFRSSPRAWFTFAHVSFALLFFFGHIWHGARTIFRDVFAGIDPD
LDEQVEFGAFQKLGDVTTRRQAV

>BAA10458.1 photosystem II CP47 protein [Synechocystis sp. PCC 6803]
MGLPWYRVHTVVLNDPGRLISVHLMHTALVAGWAGSMALYELAIFDSSDAVLNPMWRQGMFVLPFMARLG
VTSSWNGWSVTGETGLDPGFWSFEGVAAAHIVLSGLLFLAAVWHWVFWDLELFVDPRTGESALDLPKMFG
IHLFLSGLLCFGFGAFHLTGVWGPGMWVSDPYGLTGHVQPVAPEWGPAGFNPFNPGGVVAHHIAAGIVGI
IAGLFHLTVRPPERLYKALRMGNIETVLSSSIAAVFFAAFVVAGTMWYGNATTPIELFGPTRYQWDKGYF
QEEIQRRVDSQLAEGASLSEAWSTIPEKLAFYDYVGNSPAKGGLFRTGAMNSGDGIAQEWIGHPIFKDKE
GRELEVRRMPNFFETFPVIMTDADGVVRADIPFRRSESKFSVEQTGVTVSFYGGALDGQTFSNPSDVKKF
ARKAQLGEGFDFDTETFNSDGVFRTSPRGWFTFGHAVFALLFFFGHIWHGSRTLFRDVFAGVDPGLEEQV
EFGVFAKVGDLSTRKEA

>NP_958422.1 photosystem II 44 kDa protein (chloroplast) [Chlamydomonas reinhardtii]
METLFNGTLTVGGRDQETTGFAWWSGNARLINLSGKLLGAHVAHAGLIVFWAGAMNLFEVSHFVPEKPMY
EQGLILLPHIATLGYGVGPGGEIIDTFPYFVSGVLHLISSAVLGFGGVYHSLIGPETLEESYPFFGYVWK
DKNKMTNILGYHLIMLGLGAWLLVWKAMYFGGVYDTWAPGGGDVRVITNPTTNAAVIFGYLVKSPFGGDG
WICSVDNMEDIIGGHIWIGTLEILGGIWHIYTTPWPWARRAFVWSGEAYLSYSLGAIGVMGFIACCMSWF
NNTAYPSEFYGPTGPEASQSQAFTFLVRDQRLGANVASAQGPTGLGKYLMRSPTGEIIFGGETMRFWDFR
GPWLEPLRGPNGLDLNKLKNDIQPWQERRAAEYMTHAPLGSLNSVGGVATEINAVNFVSPRSWLACSHFC
LGFFFFIGHLWHAGRARAAAAGFEKGIDRFDEPVLSMRPLD

>NP_051055.1 photosystem II 44 kDa protein (chloroplast) [Arabidopsis thaliana]
MKTLYSLRRFYHVETLFNGTLALAGRDQETTGFAWWAGNARLINLSGKLLGAHVAHAGLIVFWAGAMNLF
EVAHFVPEKPMYEQGLILLPHLATLGWGVGPGGEVIDTFPYFVSGVLHLISSAVLGFGGIYHALLGPETL
EESFPFFGYVWKDRNKMTTILGIHLILLGVGAFLLVFKALYFGGVYDTWAPGGGDVRKITNLTLSPSVIF
GYLLKSPFGGEGWIVSVDDLEDIIGGHVWLGSICIFGGIWHILTKPFAWARRALVWSGEAYLSYSLAALS
VCGFIACCFVWFNNTAYPSEFYGPTGPEASQAQAFTFLVRDQRLGANVGSAQGPTGLGKYLMRSPTGEVI
FGGETMRFWDLRAPWLEPLRGPNGLDLSRLKKDIQPWQERRSAEYMTHAPLGSLNSVGGVATEINAVNYV
SPRSWLSTSHFVLGFFLFVGHLWHAGRARAAAAGFEKGIDRDFEPVLSMTPLN

>YP_717204.2 photosystem II 44 kDa protein (chloroplast) [Ostreococcus tauri]
METLFNSTLVVGGRDQESTGFAWWAGNARLINLSGKLLGAHVAHAGLIVFWAGAMNLFEVAHFVPEKPMY
EQGLILLPHLAALGYGVGPGGEVLDTFPYFVSGVLHLISSAVLGFGGVYHSLIGPETLEESYPFFGYLWK
DKNKMTTILGIHLVLLGIGAWLLVWKAMYFGGVYDTWAPGGGDVRVISYPTYDPSVIFGYLLKSPFGGDG
WIISVDNMEDVIGGHIWIGTTLIIGGFFHIFTKPFAWARRAFVWSGEAYLSYSLASVSLMAFIAAVFVWF
NNTVYPSEFFGPTGPEASQAQAFTFLVRDQRLGANIASAQGPTGLGKYLMRSPTGEIIFGGETMRFWDMR
APWVEPLRGPNGLDLSKLKNDIQPWQERRSAEYMTHAPLGSLNSVGGVATEINSTNFVNPRSWLATSHYV
LGFFFFVGHLWHAGRARAAAAGFEKGIDRDTEPVLSMRPLD

>BAA17799.2 photosystem II CP43 protein [Synechocystis sp. PCC 6803]
MVTLSNTSMVGGRDLPSTGFAWWSGNARLINLSGKLLGAHVAHAGLIVFWAGAMTLFEVAHFIPEKPMYE
QGLILLPHIATLGWGVGPAGEVTDIFPFFVVGVLHLISSAVLGLGGIYHALRGPEVLEEYSSFFGYDWKD
KNQMTNIIGYHLILLGCGALLLVFKAMFFGGVYDTWAPGGGDVRVITNPTLNPAIIFGYLLKAPFGGEGW
IISVNNMEDIIGGHIWIGLICISGGIWHILTKPFGWARRALIWSGEAYLSYSLGALSLMGFIASVFVWFN
NTAYPSEFYGPTGMEASQSQAFTFLVRDQRLGANIASAQGPTGLGKYLMRSPSGEIIFGGETMRFWDFRG
PWLEPLRGPNGLDLDKLRNDIQPWQVRRAAEYMTHAPLGSLNSVGGVITDVNSFNYVSPRAWLATSHFVL
GFFFLVGHLWHAGRARAAAAGFEKGIDRETEPTLFMPDLD

>NP_958420.1 photosystem II protein D2 (chloroplast) [Chlamydomonas reinhardtii]
MTIAIGTYQEKRTWFDDADDWLRQDRFVFVGWSGLLLFPCAYFALGGWLTGTTFVTSWYTHGLATSYLEG
CNFLTAAVSTPANSMAHSLLFVWGPEAQGDFTRWCQLGGLWAFVALHGAFGLIGFMLRQFEIARSVNLRP
YNAIAFSAPIAVFVSVFLIYPLGQSGWFFAPSFGVAAIFRFILFFQGFHNWTLNPFHMMGVAGVLGAALL
CAIHGATVENTLFEDGDGANTFRAFNPTQAEETYSMVTANRFWSQIFGVAFSNKRWLHFFMLLVPVTGLW
MSAIGVVGLALNLRAYDFVSQEIRAAEDPEFETFYTKNILLNEGIRAWMAAQDQPHERLVFPEEVLPRGN
AL

>NP_051054.1 photosystem II protein D2 (chloroplast) [Arabidopsis thaliana]
MTIALGKFTKDEKDLFDIMDDWLRRDRFVFVGWSGLLLFPCAYFALGGWFTGTTFVTSWYTHGLASSYLE
GCNFLTAAVSTPANSLAHSLLLLWGPEAQGDFTRWCQLGGLWAFVALHGAFALIGFMLRQFELARSVQLR
PYNAIAFSGPIAVFVSVFLIYPLGQSGWFFAPSFGVAAIFRFILFFQGFHNWTLNPFHMMGVAGVLGAAL
```

```
LCAIHGATVENTLFEDGDGANTFRAFNPTQAEETYSMVTANRFWSQIFGVAFSNKRWLHFFMLFVPVTGL
WMSALGVVGLALNLRAYDFVSQEIRAAEDPEFETFYTKNILLNEGIRAWMAAQDQPHENLIFPEEVLPRG
NAL

>YP_717205.1 photosystem II protein D2 (chloroplast) [Ostreococcus tauri]
MTIALGKVDEKRGWFDNVDDWLKRDRFVFVGWSGILLFPCAYFALGGWLTGTTFVSSWYTHGLASSYLEG
CNFLTSAVSTPSNSMAHSLLLLWGPEAQGDFTRWCQLGGLWTFVALHGAFGLIGFMLRQFEIARAVQIRP
YNAIAFSAPIAVFVSVFLIYPLGQSGWFFAPSFGVAGIFRFILFFQGFHNWTLNPFHMMGVAGVLGAALL
CAIHGATVENTLFEDGDGSNTFRAFNPTQAEETYSMVTANRFWSQIFGVAFSNKRWLHFFMLFVPVTGLW
MSALGVVGLALNLRAYDFVSQEIRAAEDPEFETFYTKNILLNEGIRAWMAAQDQPHENLVFPEEVLPRGN
AL

>AAA85377.1 psbD [Synechocystis sp. PCC 6803]
MTIAVGRAPVERGWFDVLDDWLKRDRFVFIGWSGLLLFPCAFMALGGWLTGTTFVTSWYTHGLASSYLEG
ANFLTVAVSSPADAFGHSLLFLWGPEAQGNLTRWFQIGGLWPFVALHGAFGLIGFMLRQFEISRLVGIRP
YNAIAFSGPIAVFVSVFLMYPLGQSSWFFAPSFGVAGIFRFILFLQGFHNWTLNPFHMMGVAGILGGALL
CAIHGATVENTLFEDGEDSNTFRAFEPTQAEETYSMVTANRFWSQIFGIAFSNKRWLHFFMLFVPVTGLW
MSSVGIVGLALNLRAYDFVSQELRAAEDPEFETFYTKNILLNEGMRAWMAPQDQPHENFIFPEEVLPRGN
AL

>NP_958396.1 photosystem II protein V (chloroplast) [Chlamydomonas reinhardtii]
MAGKPVERPFSDILTSIRYWVIHSITVPALFIAGWLFVSTGLAYDVFGTPRPNEYFTEDRQEAPLITDRF
NALEQVKKLSGN

>NP_051076.1 photosystem II protein V (chloroplast) [Arabidopsis thaliana]
MSGSTGERSFADIITSIRYWVIHSITIPSLFIAGWLFVSTGLAYDVFGSPRPNEYFTESRQGIPLITGRF
DPLEQLDEFSRSF

>YP_717219.1 photosystem II protein V (chloroplast) [Ostreococcus tauri]
MSGSTGERPFSDIVTSIRYWVIHSVTIPSLFIAGWLFVSTGLAYDVFGTPRPNEYFTEERQELPLISDRY
GALAQLDDLVPN

>BAA17092.1 cytochrome b559 a subunit [Synechocystis sp. PCC 6803]
MSGTTGERPFSDIVTSIRYWVIHSITIPMLFIAGWLFVSTGLAYDAFGTPRPDEYFTQTRQELPILQERY
DINQEIQEFNQ

>NP_958399.1 photosystem II protein VI (chloroplast) [Chlamydomonas reinhardtii]
MTTKKSAEVLVYPIFTVRWLAIHGIAVPTIFFLGAITAMQFIQR

>NP_051075.1 photosystem II protein VI (chloroplast) [Arabidopsis thaliana]
MTIDRTYPIFTVRWLAVHGLAVPTVSFLGSISAMQFIQR

>YP_717218.1 photosystem II protein VI (chloroplast) [Ostreococcus tauri]
MSVKTYPIFTFRWLAVHALAVPTVFFLGSITAMQFIQR

>BAA17093.1 cytochrome b559 b subunit [Synechocystis sp. PCC 6803]
MATQNPNQPVTYPIFTVRWLAVHTLAVPSVFFVGAIAAMQFIQR

>NP_958385.1 photosystem II protein H (chloroplast) [Chlamydomonas reinhardtii]
MATGTSKAKPSKVNSDFQEPGLVTPLGTLLRPLNSEAGKVLPGWGTTVLMAVFILLFAAFLLIILEIYNS
SLILDDVSMSWETLAKVS

>NP_051087.1 photosystem II protein H (chloroplast) [Arabidopsis thaliana]
MATQTVEDSSRSGPRSTTVGKLLKPLNSEYGKVAPGWGTTPLMGVAMALFAVFLSIILEIYNSSVLLDGI
SVN

>YP_717221.1 photosystem II protein H (chloroplast) [Ostreococcus tauri]
MAAKDSKLAGTGKVTALGTALRPLNSEYGKVAPGWGTTILMSVFIGLFAVFIVILLEIYNKSLILDNVGV
NWLSSTGL

>BAA17629.1 photosystem II PsbH protein [Synechocystis sp. PCC 6803]
MAQRTRLGDILRPLNSEYGKVVPGWGTTPVMGVFMALFLVFLLIILQIYNSSLILEGFSVDWAG

>NP_958407.1 photosystem II protein I (chloroplast) [Chlamydomonas reinhardtii]
MLTLKIFVYTVVTFFVCLFIFGFLSNDPARNPGKNLD

>NP_051043.1 photosystem II protein I (chloroplast) [Arabidopsis thaliana]
```

```
MLTLKLFVYTVVIFFVSLFIFGFLSNDPGRNPGREE

>YP_717260.1 photosystem II protein I (chloroplast) [Ostreococcus tauri]
MLTLKILVYTVVSFFVSLFIFGFLSNDPGRNPNSKDFE

>BAA10085.1 photosystem II PsbI protein [Synechocystis sp. PCC 6803]
MLTLKIAVYIVVGLFISLFIFGFLSSDPTRNPGRKDFE

>NP_958419.1 photosystem II reaction center protein J (chloroplast) [Chlamydomonas reinhardtii]
MSNTGTTGRIPLWLVGTEEGTLAIGAISCFFYGFLCWFRFFSIIKFKFFV

>NP_051073.1 photosystem II protein J (chloroplast) [Arabidopsis thaliana]
MADTTGRIPLWVIGTVAGILVIGLIGIFFYGSYSGLGSSL

>YP_717216.1 photosystem II protein J (chloroplast) [Ostreococcus tauri]
MSNTTTGRIPLWFVGMVGGLAALGLLAIFFYGSYVGLGSSL

>BAA17095.1 photosystem II PsbJ protein [Synechocystis sp. PCC 6803]
MFAEGRIPLWVVGVVAGIGAIGVLGLFFYGAYAGLGSSM

>NP_958361.1 photosystem II protein K (chloroplast) [Chlamydomonas reinhardtii]
MTTLALVLAKLPEAYAPFAPIVDVLPVIPVFFILLAFVWQAAVSFR

>NP_051042.1 photosystem II protein K (chloroplast) [Arabidopsis thaliana]
MLNIFNLICIFFNSTLFSSTFLVAKLPEAYAFLNPIVDVMPVIPLFFLLLAFVWQAAVSFR

>YP_717259.1 photosystem II protein K (chloroplast) [Ostreococcus tauri]
MTQIFLIGRLPEAYAPFDPIVDVLPIIPVLFLLLAFVWQASVSFR

>BAA17075.1 photosystem II PsbK protein [Synechocystis sp. PCC 6803]
METIYLLAKLPEAYQIFDPLVDVLPVIPLFFLALAFVWQAAVGFK

>CAA46978.1 psbL (chloroplast) [Chlamydomonas reinhardtii]
MFKNFFMARPNPNKQVVELNRTSLYWGLLLIFVLAVLFSSYIFN

>NP_051074.1 photosystem II protein L (chloroplast) [Arabidopsis thaliana]
MTQSNPNEQSVELNRTSLYWGLLLIFVLAVLFSNYFFN

>YP_717217.1 photosystem II protein L (chloroplast) [Ostreococcus tauri]
MTNPNPNKQTVELNRTSLYWGLLLIFVLAVLFSSYIFN

>BAA17094.1 photosystem II PsbL protein [Synechocystis sp. PCC 6803]
MDRNSNPNRQPVELNRTSLYLGLLLVAVLGILFSSYFFN

>NP_958382.1 photosystem II protein M (chloroplast) [Chlamydomonas reinhardtii]
MEVNIYGLTATALFIIIPTSFLLILYVKTASTQD

>NP_051053.1 photosystem II protein M (chloroplast) [Arabidopsis thaliana]
MEVNILAFIATALFILVPTAFLLIIYVKTVSQND

>XP_003082196.1 Photosystem II PsbM [Ostreococcus tauri]
MPALASALGLIPLARARATPRARARVVAPRASAGVKLAQKVQKAVVAAPAIVAAPAFAAQEAIEQVALDN
VNILGVVATALFIIIPTSFLIILYVKSASEGNVSGYSQEYYDKSKAAGNKKTNLAAALKGKGLGMRPEK

>BAA16708.1 photosystem II PsbM protein [Synechocystis sp. PCC 6803]
MQVNNLGFIASILFVLVPTVFLLILFIQTGKQSES

>NP_958386.1 photosystem II protein N (chloroplast) [Chlamydomonas reinhardtii]
MESPAFFFTFFLWFLLLSVTGYSVYVSFGPPSKKLRDPFEEHED

>NP_051086.1 photosystem II protein N (chloroplast) [Arabidopsis thaliana]
METATLVAIFISGLLVSFTGYALYTAFGQPSQQLRDPFEEHGD

>YP_717222.1 PsbN (chloroplast) [Ostreococcus tauri]
MENSAFFYGVFLWCLLISVTGYSIYIGFGPPSKELRDPFEEHED

>BAA17630.1 photosystem II PsbN protein [Synechocystis sp. PCC 6803]
```

```
MESATVLSITFAVILIAITGLAVYTSFGPPSAELGDPFDDHED

>XP_001694699.1 uncharacterized protein CHLRE_09g396213v5 [Chlamydomonas reinhardtii]
MALRAAQSAKAGVRAARPNRATAVVCKAQKVGQAAAAAALATAMVAGSANALTFDEIQGLTYLQVKGSGI
ANTCPVLESGTTNLKELKAGSYKLENFCIEPTSFTVKEESQFKGGETEFVKTKLMTRLTYTLDAMSGSFK
VGSDGSAELKEDDGIDYAATTVQLPGGERVAFLFTIKQFDGKGTLDNIKGDFLVPSYRGSSFLDPKGRGG
STGYDNAVALPARADAEELLKENVKITKALKGSAVFSVAKVDPVTGEIAGVFESIQPSDTDLGAKPPKDI
KVTGLWYAQLK

>NP_201458.1 PS II oxygen-evolving complex 1 [Arabidopsis thaliana]
MAASLQSTATFLQSAKIATAPSRGSSHLRSTQAVGKSFGLETSSARLTCSFQSDFKDFTGKCSDAVKIAG
FALATSALVVSGASAEGAPKRLTYDEIQSKTYMEVKGTGTANQCPTIDGGSETFSFKPGKYAGKKFCFEP
TSFTVKADSVSKNAPPEFQNTKLMTRLTYTLDEIEGPFEVASDGSVNFKEEDGIDYAAVTVQLPGGERVP
FLFTVKQLDASGKPDSFTGKFLVPSYRGSSFLDPKGRGGSTGYDNAVALPAGGRGDEEELVKENVKNTAA
SVGEITLKVTKSKPETGEVIGVFESLQPSDTDLGAKVPKDVKIQGVWYGQLE

>CEG01950.1 Photosystem II PsbO, manganese-stabilising [Ostreococcus tauri]
MASIARSAVFGNTAFAVRAKSAKRSAKAQAPKASLNEFAEKVSKATVSLGVAGAVLAQSASAVTYDELQG
LTYLQVKGTGLANTCPVVEGGVSGKDIKAGDYKLERFCMEPTSFTVKEESQFKAGESEFQKTRLMTRLTY
TLDGMTGDFKVGSDGSVQIKENDGLDYAPVTVQLPGGERVPFLFTLKEFTGKGNTSQFGGDFVVPSYRGS
SFLDPKGRGGSTGYDNAVALPARADSEELQKENNKSAAALKGSAVFNIAKYDPATGEIAGVFESIQPSDT
DLGAKVPKDVKITGLWYGNIAK

>BAA18474.1 photosystem II manganese-stabilizing polypeptide [Synechocystis sp. PCC 6803]
MRFRPSIVALLSVCFGLLTFLYSGSAFAVDKSQLTYDDIVNTGLANVCPEISSFTRGTIEVEPNTKYFVS
DFCMEPQEYFVKEEPVNKRQKAEYVKGKVLTRQTTSLEQIRGSIAVGADGTLTFKEKDGIDFQPITVLLP
GGEEVPFFFTVKNFTGTTEPGFTSINSSTDFVGDFNVPSYRGAGFLDPKARGLYTGYDNAVALPSAADKF
RTNKKETPLGKGTLSLQVTQVDGSTGEIAGIFESEQPSDTDLGAKEPLDVKVRGIFYGRVDTDV

>sp|P11471.1|PSBP_CHLRE RecName: Full=Oxygen-evolving enhancer protein 2, chloroplastic; Short=OEE2; Flags:
Precursor
MATALCNKAFAAAPVARPASRRSAVVVRASGSDVSRRAALAGFAGAAALVSSSPANAAYGDSANVFGKVT
NKSGFVPYAGDGFALLLPAKWNPSKENDFPGVILRYEDNFDAVNNLVVIAQDTDKKAIADFGSQDKFLES
VSYLLGKQAYSGETQSEGGFAPNRVSAASLLDVSTTTDKKGKTYYKYELLVRSADGDEGGRHQLIGATVG
SDNKLYIIKIQIGDKRWFKGAKKEAMGAFDSFTVV

>NP_191224.1 thylakoid lumenal protein (Mog1/PsbP/DUF1795-like photosystem II reaction center PsbP family protein)
[Arabidopsis thaliana]
MATASLVPTSKIFSVSPKSSASIKARSRVVVASSQQQQPRRRELLLKSAVAIPAILQLKEAPISAAREV
EVGSYLPLSPSDPSFVLFKAKPSDTPALRAGNVQPYQFVLPPNWKQLRIANILSGNYCQPKCAEPWIEVK
FENEKQGKVQVVASPLIRLTNKPNATIEDLGEPEKVIASLGPFVTGNSYDSDELLKTSIEKIGDQTYYKY
VLETPFALTGSHNLAKATAKGSTVVLFVVSATEKQWQSSQKTLEAILDSFQL

>XP_003083092.1 photosystem II oxygen-evolving complex 23 kDa protein, chloropla (IC) [Ostreococcus tauri]
METLHVILTLSLDARRRLRDGEGFARRAQVALLTLEPTLVTGLKLHDVQVAVAHGGGDQLVASAFVAIRG
AREDFILVVLLALVLAGTDFEHGGGGNLVRREPTFGSRFAGVRLLTEQVRDLFQERLRLAVVRDLALRGL
FDDDGDVRHRVKVILPVRVLVRERALLGRVVLRREQERVSIARVRDETYVFANTSTATPRGPTTSPNPIR
ERIFAKPIALGHPPVLHRARRTGRVRDTAPNVRGITIAPASVARLHRAIASRSRAYPSHASVSIAHPRRH
RR

>ALJ67700.1 PsbP [Synechocystis sp. PCC 6803]
MLKKSLSTAVVLVTLLLSFTLTACGGVGIASLQRYSDTKDGYEFLYPNGWIGVDVKGASPGVDVVFRDLI
ERDENLSVIISEIPSDKTLTDLGTATDVGYRFMKTVNDASQGDRQAELINAEARDEDGQVYYTLEYRVLV
GDNVERHDLASVTTNRGKLITFDLSTAEDRWDTVKSLFDTVASSFHVY

>XP_001701331.1 oxygen evolving enhancer protein 3 [Chlamydomonas reinhardtii]
MALASKVATRPAVASRRGAVVVRASGESRRAVLGGLLASAVAAVAPKAALALTPVDLFDDRSVRDRGFDL
IYEARDLDLPQNVREGFTQARASLDETKKRVKESEARIDADLDVFIQKSYWTEAREQLRRQVGTLRFDLN
TLASTKEKEAKKAALGLRKEFIQAVEDLDFALREKDQASAAKKLEITKAKLDSVLAAVL

>NP_001031687.1 photosystem II subunit QA [Arabidopsis thaliana]
MASMGGLHGASPAVLEGSLKINGSSRLNGSGRVAVAQRSRLVVRAQQSEETSRRSVIGLVAAGLAGGSFV
QAVLADAISIKVGPPPAPSGGLPAGTDNSDQARDFALALKDRFYLQPLPPTEAAARAKESAKDIINVKPL
IDRKAWPYVQNDLRSKASYLRYDLNTIISSKPKDEKKSLKDLTTKLFDTIDNLDYAAKKKSPSQAEKYYA
ETVSALNEVLAKLG

>CEG00370.1 Photosystem II PsbQ, oxygen evolving complex [Ostreococcus tauri]
```

MSVAASTSTFAGKAIAQKAVVKTSKVSTVVKASANKALSLAAATAVVLSTGPAFANTAIELTDKRAEQQN
GLQLIYEARDLDLSNKPNAEGNSRFSFQKLGSTETAARARESVARINGDVGTYVEKKYWTQAANELRRQV
GTLRFDINNLVNEKGANAAEAKAFYKKLENLDFSIRQKDQAAATALLAEVQSTANTLISSLA

>WP_028947222.1 MULTISPECIES: photosystem II protein PsbQ [Cyanophyceae]
MSRLRSLLSLVLVLVTTVLVSCSSPQVEIPTTYSPEKIAQLQVYVNPIAVARDGMEKRLQGLIADQNWVD
TQTYIHGPLGQLRRDMLGLSSSLLPKDQDKAKTLAKEVFGHLERLDVAAKDRNSSQAKIQYQEALADFDA
FLNLLPQAS

>XP_001696588.2 uncharacterized protein CHLRE_06g261000v5 [Chlamydomonas reinhardtii]
MATMQISAKGLAPLRPRVSSRRVVKPVASGGGKTDITKVGLNSIEDPVVKQNLMGKSRFMNKKDWKDASG
RKGKGYGVYRYEDKYGANVDGYSPIYTPDLWTESGDSYTLGTKGLIAWAGLVLVLLAVGVNLIISTSQLG
A

>NP_178025.1 photosystem II subunit R [Arabidopsis thaliana]
MAASVMLSSVTLKPAGFTVEKTAARGLPSLTRARPSFKIVASGVKKIKTDKPFGINGSMDLRDGVDASGR
KGKGYGVYKYVDKYGANVDGYSPIYNENEWSASGDVYKGGVTGLAIWAVTLAGILAGGALLVYNTSALAQ

>CEF98147.1 Photosystem II PsbR [Ostreococcus tauri]
MSALTSSTFVTKSVAATATKSSRARRGALVVMNNSSGPKRGTKAAPAQRSVGYKGSTDAGSAPKTRSGKS
GYVYKLGLKNGKANVDEYSPIWTPQDFKSDGDKYEGNLTYWAITLAGIIGTGALAIILTSQL

>XP_001689923.1 uncharacterized protein CHLRE_01g016750v5 [Chlamydomonas reinhardtii]
MAMTLSTKAFAQRGVSARKNTVRVYAATTKVNPKLASKTEVERFKQATGLPAPAINGKQFPLKLGFTKTN
ELFVGRLAMVGFSASLIGEILTGKGALAQFGYETGLNGIEVDGLVIGLIAFNLIAAVLPTSQTFVPEEQD
TISERPAGPLQDPRITLLEPKKFFGVQGFGFTKENELFVGRAAQLGFAFSLIGEAVTGKGALAQFDIETG
LSLRDTEFGLVVFILFLLFAAINEGSGKFVDEESA

>NP_175092.1 Chlorophyll A-B binding family protein [Arabidopsis thaliana]
MAQTMLLTSGVTAGHFLRNKSPLAQPKVHHLFLSGNSPVALPSRRQSFVPLALFKPKTKAAPKKVEKPKS
KVEDGIFGTSGGIGFTKANELFVGRVAMIGFAASLLGEALTGKGILAQLNLETGIPIYEAEPLLLFFILF
TLLGAIGALGDRGKFVDDPPTGLEKAVIPPGKNVRSALGLKEQGPLFGFTKANELFVGRLAQLGIAFSLI
GEIITGKGALAQLNIETGIPIQDIEPLVLLNVAFFFFAAINPGNGKFITDDGEES

>XP_003080104.1 photosystem II 22 kda polypeptide (ISS) [Ostreococcus tauri]
MHARAIASSSVAHARPGSTRHARGGNGESVTSTSRPAARVRGKRAVARAVAPADEGERDVTDIEINLGRA
AMLGFLGTTLGDVITRGEGPIEQLEDEAAYVINHVNPVDLTRDALEVAGIYVESVILVWLLLGCALLLGV
SQGMRNPIRTVSGRTTKQPYETNVKEQKPYELFNGRLAMLGSAFALVGDVRTGGLGPLEQVQNEVGIPVI
DEEIFAVIFLGGVALNVVSTGITAFRRAYAKGRE

>NP_958387.1 photosystem II protein T (chloroplast) [Chlamydomonas reinhardtii]
MEALVYTFLLVGTLGIIFFSIFFRDPPRMIK

>NP_051085.1 photosystem II protein T (chloroplast) [Arabidopsis thaliana]
MEALVYTFLLVSTLGIIFFAIFFREPPKISTKK

>YP_717206.1 photosystem II protein T (chloroplast) [Ostreococcus tauri]
MEAIVYTFLLVGTLGIIFFAIFFREPPRIAK

>BAA10133.1 photosystem II PsbT protein [Synechocystis sp. PCC 6803]
MESVAYILVLTMALAVLFFAIAFREPPRIEK

>BAA16847.1 photosystem II 12 kD extrinsic protein [Synechocystis sp. PCC 6803]
MKFISRLLVACSLLIGLMGFLGADLAQALTPNPILAELNAVDAKLTTDFGQKIDLNNSDIRDFRGLRGFY
PNLASEIIKNAPYDTVEEVLDIPGLSETQKSRLEANLGSFTVTEPSIELTSGDDRINPGVY

>BAA18512.1 cytochrome c550 [Synechocystis sp. PCC 6803]
MKRFFLVAIASVLFFFNTMVGSANAVELTESTRTIPLDEAGGTTTLTARQFTNGQKIFVDTCTQCHLQGK
TKTNNNVSLGLADLAGAEPRRDNVLALVEFLKNPKSYDGEDDYSELHPNISRPDIYPEMRNYTEDDIFDV
AGYTLIAPKLDERWGGTIYF

>XP_001702016.1 photosystem II reaction center W protein [Chlamydomonas reinhardtii]
MQALSARAPRVAAKPVSRSGARSAVTVVCKATTVRSEVAKKVAMLSTLPATLAAHPAFALVDERMNGDGT
GRPFGVNDPVLGWVLLGVFGTMWAIWFIGQKDLGDFEDADDGLKL

>NP_180615.1 photosystem II reaction center W [Arabidopsis thaliana]
MASFTASASTVSAARPALLLKPTVAISAPVLGLPPMGKKKGGVRCSMETKQGNVSVMGAGVSAAATAALT

AVMSNPAMALVDERMSTEGTGLPFGLSNNLLGWILFGVFGLIWTFFFVYTSSLEEDEESGLSL

>CEG01031.1 Photosystem II PsbW, class 2 [Ostreococcus tauri]
MLTILSSLSPICNLSPRRAGPRQASNQRHGVTVRATPQEAGKQVFGTSMLISSMNTVLPALAIVDDRLNG
DGTGLPLGINDGALGIILASVPLAIFSLYFIAGKQDAGIENGGKNDDSGLSL

>CAO86551.1 psbW [Microcystis aeruginosa PCC 7806]
MAEIQFSRGLKEPVVPDIRLSRTKTGEKGSAKFYFDQPTILGDQQTEEITGMYLIDEEGEIVTREVKGKF
INGKPMAIEATVIFNSPAEWDRFMRFMERYGSENGLEFTKSS

>XP_001701704.1 4.1 kDa photosystem II subunit [Chlamydomonas reinhardtii]
MAAVCVSSKAVAVRPATRVASRPAARMVVRASAKPAQISEKAIAAGVSAAAAASLALAPIAEAAVTPSLK
NFLGSLIAGATVLGGIALAITAVSKFDKTIR

>NP_565335.1 photosystem II subunit X [Arabidopsis thaliana]
MASTSAMSLVTPLNQTRSSPFLKPLPLKPSKALVATGGRAQRLQVKALKMDKALTGISAAALTASMVIPE
IAEAAGSGISPSLKNFLLSIASGGLVLTVIIGVVVGVSNFDPVKRT

>CAL52344.1 Photosystem II PsbX [Ostreococcus tauri]
MVVYTSKLRSLAIPKAQVRNARKLGKTSAVSAKLAPAEGTFQFDGIESAKAKVGAAVSAAAVATISSPLA
AEASVTPSLRNTLLSVVAGGVVLAAIAVAVIGVSSFDKVSRK

>BAA10311.1 photosystem II PsbX protein [Synechocystis sp. PCC 6803]
MTPSLANFLWSLVLGAAIVLIPATVGLIFISQKDKITRS

>XP_001692627.1 ycf32-related polyprotein of photosystem II, partial [Chlamydomonas reinhardtii]
VGTVATLSSVAASWMMAGNAQAATELASLAASDNRAGILATLLVPVLGWVGFNIFGSLQAQLNQMDAKNK
RAVPAAVGMGAAASLLFAQSAEASTEIATLAASDNRVAILATLLVPVIGWVGFNIFGSLQAQLNQMDAKN
KRAVPAAVGMGAAASLLFAQSAEASTEIATLAASDNRVAILATLLVPVIGWVGFNIFGSLQAQLNQMDAK
NKRAVPAAVGMGAAASLLFAQSAEASTEIATLAASDNRVAILATLLVPVIGWVGFNIFGSLQAQLNQMDA
KNKRAVPAAVGMGAAASLLFAQSAEASTEIATLAASDNRVAILATLLVPVIGWVGFNIFGSLQAQLRQMD
AKNK

>NP_176940.1 photosystem II BY [Arabidopsis thaliana]
MAAAMATATKCMSLNPSPPKLQNQTKSKPFISLPTPPKPNVSLAVTSTALAGAVFSSLSYSEPALAIQQI
AQLAAANASSDNRGLALLLPIVPAIAWVLYNILQPAINQVNKMRESKGIVVGLGIGGGLAASGLLTPPPE
AYAAAEAAAASSDSRGQLLLIVVTPALLWVLYNILQPALNQINKMRSGD

>XP_003082206.1 manganese-binding protein PsbY precursor, photosystem II-associated-s (ISS) [Ostreococcus tauri]
MLCTVNVPALRASRAAPLAKRERRAKRTAIASKVATKVATIESDEIVTRGQTAAMSLAATMLFADQAHAA
QEVGDLAGIDGRIAIFFVFAPVLGWVAYNILGPGLRQFEDMQKKNAKSKGVLAGAGLSAAALLGMPEASD
AAEKLGDLAGIDGRIAIFFVFAPVLGWVAYNILGPGLRQFEDMQKKNAKSKGVLAGAGLSAAALLGMPEA
SDAAEKLGDLAGIDGRIAIFFVFAPVLGWVAYNILGPERGQEVDGLRRRSLQTPP

>BAA17722.1 ycf32 gene product [Synechocystis sp. PCC 6803]
MDWRVIVVVSPLLIAATWAAINIGAAAIRQLQDVLGREA

>NP_958383.1 photosystem II protein Z (chloroplast) [Chlamydomonas reinhardtii]
MTSILQVALLALIFVSFALVVGVPVVFATPNGWTDNKGAVFSGLSLWLLLVFVVGILNSFVV

>NP_051056.1 photosystem II protein Z (chloroplast) [Arabidopsis thaliana]
MTIAFQLAVFALIITSSILLISVPVVFASPDGWSSNKNVVFSGTSLWIGLVFLVGILNSLIS

>YP_717234.1 photosystem II protein Z (chloroplast) [Ostreococcus tauri]
MLFIFQLTLLAFIGLSLALVIGVPVLLASPEGWAQSKGLVFSGSALWMLLVFVVGALNSFVS

>BAA18462.1 photosystem II 11 kD protein [Synechocystis sp. PCC 6803]
MMSFLKNQLSRLLALILVVAIGLTACDSGTGLTGNYSQDTLTVIATLREAIDLPQDAPNRQEVQDTARGQ
INDYISRYRRKGDAGGLKSFTTMQTALNSLAGYYTSYGARPIPEKLKKRLQLEFTQAERSIERGV

>sp|P50370.1|PSB30_CHLRE RecName: Full=Photosystem II reaction center protein Psb30; AltName: Full=Photosystem II reaction center protein Ycf12
MNIELALTLVSLVLVVSAGPLVVVLLSARGGNL

>YP_717256.1 Ycf12 (chloroplast) [Ostreococcus tauri]
MNLELIGQLVTVALVVGAGPIIIGALFARGGNL

```
>WP_009631668.1 photosystem II reaction center protein Ycf12/Psb30 [Synechocystis sp. PCC 7509]
MESLFDGLGSINWLVLAQLVSVGLIMISGPVVIFILAFRNGNL

>XP_001696125.1 uncharacterized protein CHLRE_08g365900v5 [Chlamydomonas reinhardtii]
MAMMMRKAAAVPASSRRSVAVNSVSGKRTVSGKAGAPVPEDVLAYAKTLPGVTAPFDNVFDPAGFLATAS
VKDVRRWRESEITHGRVAMLAALGFIVGEQLQDFPLFFNFDGRVSGPAIYHFQQIGQGFWEPLLIAIGVA
ESYRVAVGWATPTGTGFNSLKDDYEPGDLGFDPLGLKPTDPEELKTLQTKELNNGRLAMIAIAAFVAQEL
VEQTEIFEHLVLRFEKEVILELEDVERDLGLPLTPLPDNLKAI

>XP_001696064.1 uncharacterized protein CHLRE_08g367500v5 [Chlamydomonas reinhardtii]
MLANVVSRKASGLRQTPARATVAVKSVSGRRTTAAEPQTAAPVAAEDVFAYTKNLPGVTAPFEGVFDPAG
FLATASIKDVRRWRESEITHGRVAMLAALGFVVGEQLQDFPLFFNWDGRVSGPAIYHFQQIGQGFWEPLL
IAIGVAESYRVAVGWATPTGTGFNSLKDDYEPGDLGFDPLGLKPTDPEELKVMQTKELNNGRLAMIAIAA
FVAQELVEQTEIFEHLALRFEKEAILELDDIERDLGLPVTPLPDNLKSL

>NP_958375.1 photosystem I P700 chlorophyll a apoprotein A1 (chloroplast) [Chlamydomonas reinhardtii]
MTISTPEREAKKVKIAVDRNPVETSFEKWAKPGHFSRTLSKGPNTTTWIWNLHADAHDFDSHTSDLEEIS
RKVFSAHFGQLGIIFIWLSGMYFHGARFSNYEAWLSDPTHIKPSAQVVWPIVGQEILNGDVGGGFQGIQI
TSGFFQLWRASGITSELQLYTTAIGGLVMAAAMFFAGWFHYHKAAPKLEWFQNVESMLNHHLGGLLGLGS
LAWAGHQIHVSLPVNKLLDAGVDPKEIPLPHDLLLNRAIMADLYPSFAKGIAPFFTLNWSEYSDFLTFKG
GLNPVTGGLWLSDTAHHHVAIAVLFLVAGHMYRTNWGIGHSMKEILEAHRGPFTGEGHVGLYEILTTSWH
AQLAINLALFGSLSIIVAHHMYAMPPYPYLATDYGTQLSLFTHHTWIGGFCIVGAGAHAAIFMVRDYDPT
NNYNNLLDRVIRHRDAIISHLNWVCIFLGFHSFGLYIHNDTMSALGRPQDMFSDTAIQLQPVFAQWIQNT
HFLAPQLTAPNALAATSLTWGGELGAHGGKVAMMPISLGTSDFMVHHIHAFTIHVTVLILLKGVLFARSS
RLIPDKANLGFRFPCDGPGRGGTCQVSAWDHVFLGLFWMYNSLSIVIFHFSWKMQSDVWGTVTASGVSHI
TGGNFAQSANTINGWLRDFLWAQSSQVIQSYGSALSAYGLIFLGAHFVWAFSLMFLFSGRGYWQELIESI
VWAHNKLKVAPAIQPRALSITQGRAVGVAHYLLGGIATTWSFFLARIISVG

>NP_051059.1 photosystem I P700 chlorophyll a apoprotein A1 (chloroplast) [Arabidopsis thaliana]
MIIRSPEPEVKILVDRDPIKTSFEEWAKPGHFSRTIAKGPDTTTWIWNLHADAHDFDSHTSDLEEISRKV
FSAHFGQLSIIFLWLSGMYFHGARFSNYEAWLSDPTHIGPSAQVVWPIVGQEILNGDVGGGFRGIQITSG
FFQIWRASGITSELQLYCTAIGALVFAALMLFAGWFHYHKAAPKLAWFQDVESMLNHHLAGLLGLGSLSW
AGHQVHVSLPINQFLNAGVDPKEIPLPHEFILNRDLLAQLYPSFAEGATPFFTLNWSKYSEFLTFRGGLD
PVTGGLWLTDIAHHHLAIAILFLIAGHMYRTNWGIGHGIKDILEAHKGPFTGQGHKGLYEILTTSWHAQL
SLNLAMLGSLTIIVAHHMYSMPPYPYLATDYATQLSLFTHHMWIGGFLIVGAAAHAAIFMVRDYDPTNRY
NDLLDRVLRHRDAIISHLNWVCIFLGFHSFGLYIHNDTMSALGRPQDMFSDTAIQLQPVFAQWIQNTHAL
APGVTAPGETASTSLTWGGGELVAVGGKVALLPIPLGTADFLVHHIHAFTIHVTVLILLKGVLFARSSRL
IPDKANLGFRFPCDGPGRGGTCQVSAWDHVFLGLFWMYNAISVVIFHFSWKMQSDVWGSISDQGVVTHIT
GGNFAQSSITINGWLRDFLWAQASQVIQSYGSSLSAYGLFFLGAHFVWAFSLMFLFSGRGYWQELIESIV
WAHNKLKVAPATQPRALSIIQGRAVGVTHYLLGGIATTWAFFLARIIAVG

>YP_717254.1 photosystem I P700 chlorophyll a apoprotein A1 (chloroplast) [Ostreococcus tauri]
MTISPPEREIKTVKIVVDRDPVPTSFEKWAKPGHFSRTLAKGPATTTWVWDLHADAHDFDSHTTDLEDIS
RKVFSAHFGQLGVIFIWLSGMYFHGARFSNYEAWLSDPTHIKPSAQVVWPIVGQEILNGDVGGGFQGVQI
TSGFFQLWRGSGITSELQLYSTAIGALICAGLMFFAGWWHYHKAAPKLEWFQNVESMMNHHLAGLLGLGS
LAWAGHQIHIALPVNTLLDAGVDPKEIPLPHEFMLNRALMAELYPSFAKGLTPFFTFNWTEYSDFLTFRG
GLNPVTGGLWLTDMAHHHLAIAVLFLVAGHQYRTNWGIGHSMKEILEAHKGPFTGEGHKGLYEILTTSWH
AQLAINLALMGSLSIIVSHHMYSMPPYPYLATDYGTQLSLFTHHMWIGGFCICGAAAHAAIFMVRDYDPA
TNYNNVLDRVLRHRDAIISHLNWVCIFLGFHSFGLYIHNDTMSALGRPQDMFSDTAIQLQPIFAQFVQHT
HALAPELTAPTAAGSTSASWGGDIVAVGGKIAMMPISLGTSDFMVHHIHAFTIHVTVLILLKGVLFARSS
RLIPDKANLGFRFPCDGPGRGGTCQVSAWDHVFLGLFWMYNSISIVIFHFSWKMQSDVWGSVTGNGVSHI
TGGNFAQSANTINGWLRDFLWAQSSQVIQSYGSALSAYGLIFLGAHFIWAFSLMFLFSGRGYWQELIESI
VWAHNKLKVAPAIQPRALSITQGRAVGVAHYLLGGIATTWSFFLARIIAVG

>BAA17437.1 P700 apoprotein subunit Ia [Synechocystis sp. PCC 6803]
MTISPPEREAKAKVSVDNNPVPTSFEKWGKPGHFDRTLARGPKTTTWIWNLHANAHDFDSQTSDLEDVSR
KIFSAHFGHLAVVFVWLSGMYFHGAKFSNYEGWLADPTHIKPSAQVVWPIVGQGILNGDVGGGFHGIQIT
SGLFYLWRASGFTDSYQLYCTAIGGLVMAALMLFAGWFHYHVKAPKLEWFQNVESMMNHHLAGLLGLGSL
GWAGHQIHVSMPINKLLDAGVAPKDIPLPHEFILEPSKMAELYPSFAQGLTPFFTLNWGVYSDFLTFKGG
LNPVTGGLWLSDTAHHHLAIAVLFIIAGHMYRTNWGIGHSMKEILEAHKGPFTGEGHKGLYEILTTSWHA
QLAINLALLGSLTIIVAQHMYAMPPYPYQAIDYATQLSLFTHHMWIGGFLIVGAGAHGAIFMVRDYDPAK
NVNNLLDRMLRHRDAIISHLNWVCIFLGFHSFGLYIHNDTMRALGRPQDMFSDTAIQLQPIFAQWVQHLH
TLAPGATAPNALATASYAFGGETIAVAGKVAMMPITLGTADFMVHHIHAFTIHVTALILLKGVLYARSSR
LVPDKANLGFRFPCDGPGRGGTCQVSGWDHVFLGLFWMYNSLSVVIFHFSWKMQSDVWGTVSPDGSVTHV
TLGNFAQSAITINGWLRDFLWAQAANVINSYGSALSAYGIMFLAGHFVFAFSLMFLFSGRGYWQELIESI
VWAHNKLNVAPAIQPRALSIIQGRAVGVAHYLLGGIVTTWAFFLARSLSIG
```

```
>NP_958404.1 photosystem I P700 chlorophyll a apoprotein A2 (chloroplast) [Chlamydomonas reinhardtii]
MATKLFPKFSQGLAQDPTTRRIWYGLAMAHDFESHDGMTEENLYQKIFASHFGQLSIIFLWTSGNLFHVA
WQGNFEQWVTDPVHIRPIAHAIWDPHFGQPAVEAFTRGGASGPVNISTSGVYQWWYTIGMRTNQDLYVGS
VFLALVSAIFLFAGWLHLQPNFQPSLSWFKDAESRLNHHLSGLFGVSSLAWTGHLVHVAIPESRGQHVGW
DNFLSVLPHPQGLTPFFTGNWAAYAQSPDTASHVFGTAQGSGQAILTFLGGFHPQTQSLWLTDMAHHHLA
IAVIFIVAGHMYRTNFGIGHRMQAILEAHTPPSGSLGAGHKGLFDTVNNSLHFQLGLALASVGTITSLVA
QHMYSLPPYAFQAIDFTTQAALYTHHQYIAGFIMCGAFAHGAIFFIRDYDPEQNKGNVLARMLDHKEALI
SHLSWVSLFLGFHTLGLYVHNDVMQAFGTPEKQILIEPVFAQWIQAAHGKALYGFDFLLSSKTSAAFANG
QSLWLPGWLDAINNNQNSLFLTIGPGDFLVHHAIALGLHTTTLILVKGALDARGSKLMPDKKDFGYSFPC
DGPGRGGTCDISAYDAFYLAVFWMLNTIGWVTFYWHWKHLTLWQGNVAQFDESSTYLMGWLRDYLWLNSS
QLINGYNPFGMNSLSVWAWTFLFGHLIYATGFMFLISWRGYWQELIETLVWAHEKTPLANLVYWKDKPVA
LSIVQARLVGLAHFSVGYIFTYAAFLIASTSGRFG

>NP_051058.1 photosystem I P700 chlorophyll a apoprotein A2 (chloroplast) [Arabidopsis thaliana]
MALRFPRFSQGLAQDPTTRRIWFGIATAHDFESHDDITEERLYQNIFASHFGQLAIIFLWTSGNLFHVAW
QGNFETWVQDPLHVRPIAHAIWDPHFGQPAVEAFTRGGALGPVNIAYSGVYQWWYTIGLRTNEDLYTGAL
FLLFLSALSLIGGWLHLQPKWKPRVSWFKNAESRLNHHLSGLFGVSSLAWTGHLVHVAIPASRGEYVRWN
NFLNVLPHPQGLGPLFTGQWNLYAQNPDSSSHLFGTSQGSGTAILTLLGGFHPQTQSLWLTDMAHHHLAI
AILFLIAGHMYRTNFGIGHSIKDLLEAHIPPGGRLGRGHKGLYDTINNSIHFQLGLALASLGVITSLVAQ
HMYSLPAYAFIAQDFTTQAALYTHHQYIAGFIMTGAFAHGAIFFIRDYNPEQNEDNVLARMLDHKEAIIS
HLSWASLFLGFHTLGLYVHNDVMLAFGTPEKQILIEPIFAQWIQSAHGKTSYGFDVLLSSTSGPAFNAGR
SIWLPGWLNAINENSNSLFLTIGPGDFLVHHAIALGLHTTTLILVKGALDARGSKLMPDKKDFGYSFPCD
GPGRGGTCDISAWDAFYLAVFWMLNTIGWVTFYWHWKHITLWQGNVSQFNESSTYLMGWLRDYLWLNSSQ
LINGYNPFGMNSLSVWAWMFLFGHLVWATGFMFLISWRGYWQELIETLAWAHERTPLANLIRWKDKPVAL
SIVQARLVGLAHFSVGYIFTYAAFLIASTSGKFG

>YP_717253.1 photosystem I P700 chlorophyll a apoprotein A2 (chloroplast) [Ostreococcus tauri]
MATKFPKFSQGLASDPTTRRIWFGIATAHDFETHDGMTEEKLYQKIFASHFGQLAIIFLWTSGNLFHVAW
QGNFEKWGEDPLHVRPIAHTIWDPHFGQPAVEAFTRGGASAPVNIAYSGVYQWWYTIGMRTNVDLYNGSL
FLLFVAGLFLFAGWLHLQPTFAPAVSWFKNAESRLNHHLSGLFGVSSLAWTGHLVHVAIPASRGETVRWD
NFLTTLPHPAGLAPFFTGQWAVYAQNPDTAGHIFGTSEGAGTAILTFLGGFHPQTQSLWLTDMAHHHLAI
AVVFIIAGHQYRTNFGIGHSMKEILEAHTAPSGRLGAGHTGLFDTVNNSLHFQLGLALASVGVLCSLTAQ
HMYSMPPYAFLAQDFTTQAALYSHHQYIAGFIMCGAFAHGAIFFIRDYDPEANKGNVLARMLEHKEAIIS
HLSWVSLFLGFHTLGLYVHNDVMQAFGTPEKQILIEPVFAQWIQAAQGKALYGFDILLSGDNAATAAGNS
IYLPGWLAGINSSTNSLFLPIGPGDFLVHHAIALGLHTTTLILVKGALDARGSKLMPDKKDFGYSFPCDG
PGRGGTCDISAWDAFYLAVFWELNTVSWTVFYFHWKHLALWQGNSAQFDESSTYIMGWLRDYLWLNSSQL
INGYNPFGMNSLSVWAWMFLFGHLIYATGFMFLISWRGYWQELIETLVWAHERTPLANFVKWNDKPVALS
IVQARLVGLTHFAVGFVLTYAAFVIASTSGKFG

>BAA17438.1 P700 apoprotein subunit Ib [Synechocystis sp. PCC 6803]
MATKFPKFSQDLAQDPTTRRIWYGIATAHDFETHDGMTEENLYQKIFASHFGHIAIIFLWTSGTLFHVAW
QGNFEQWIKDPLNIRPIAHAIWDPHFGEGAVNAFTQAGASNPVNIAYSGVYHWFYTIGMTTNQELYSGAV
FLLVLASLFLFAGWLHLQPKFRPSLAWFKNAESRLNHHLAGLFGVSSLAWAGHLVHVAIPEARGQHVGWD
NFLSTPPHPAGLMPFFTGNWGVYAADPDTAGHIFGTSEGAGTAILTFLGGFHPQTESLWLTDIAHHHLAI
AVIFIIAGHMYRTNWGIGHSIKEILNAHKGPLTGAGHTNLYDTINNSLHFQLGLALASLGVITSLVAQHM
YSLPSYAFIAQDHTTQAALYTHHQYIAGFLMVGAFAHGAIFFVRDYDPVANKDNVLARMLEHKEALISHL
SWVSLFLGFHTLGLYVHNDVVVAFGTPEKQILIEPVFAQWIQATSGKALYGFDVLLSNPDSIASTTGAAW
LPGWLDAINSGTNSLFLTIGPGDFLVHHAIALGLHTTALILIKGALDARGSKLMPDKKDFGYSFPCDGPG
RGGTCDISAWDAFYLAMFWMLNTLGWLTFYWHWKHLGVWSGNVAQFNENSTYLMGWFRDYLWANSAQLIN
GYNPYGVNNLSVWAWMFLFGHLVWATGFMFLISWRGYWQELIETIVWAHERTPLANLVRWKDKPVALSIV
QARLVGLAHFTVGYVLTYAAFLIASTAGKFG

>NP_958423.1 photosystem I subunit VII (chloroplast) [Chlamydomonas reinhardtii]
MAHIVKIYDTCIGCTQCVRACPLDVLEMVPWDGCKASQMASAPRTEDCVGCKRCETACPTDFLSVRVYLG
SESTRSMGLSY

>NP_051110.1 photosystem I subunit VII (chloroplast) [Arabidopsis thaliana]
MSHSVKIYDTCIGCTQCVRACPTDVLEMIPWDGCKAKQIASAPRTEDCVGCKRCESACPTDFLSVRVYLW
HETTRSMGLAY

>YP_717214.1 photosystem I subunit VII (chloroplast) [Ostreococcus tauri]
MSHAVKIYDTCIGCTQCVRACPTDVLEMVPWDGCKAGQIASAPRTEDCVGCKRCEAACPTDFLSVRVYLG
SETTRSMGLAY

>BAA10036.1 photosystem I subunit VII [Synechocystis sp. PCC 6803]
MSHSVKIYDTCIGCTQCVRACPLDVLEMVPWDGCKAAQIASSPRTEDCVGCKRCETACPTDFLSIRVYLG
AETTRSMGLAY
```

>XP_001697722.1 uncharacterized protein CHLRE_05g238332v5 [Chlamydomonas reinhardtii]
MAVMMRTQAPAATRASSRVAVAARPAARRAVVVRAEAEAAPAAAKKAAEKPAWTVPTLNPDTPSPIFGGS
TGGLLRKAQTEEFYVITWEAKKEQIFEMPTGGAAIMRQGPNLLKFGKKEQCLALTTQLRNKFKLTPCFYR
VFPDGKVQYLHPADGVYPEKVNAGRVGANQNMRRIGQNVNPIKVKFSGRMMSPAEI

>NP_192186.1 photosystem I subunit D-1 [Arabidopsis thaliana]
MATQAAGIFNSAITTAATSGVKKLHFFSTTHRPKSLSFTKTAIRAEKTDSSAAAAAAPATKEAPVGFTPP
QLDPNTPSPIFAGSTGGLLRKAQVEEFYVITWNSPKEQIFEMPTGGAAIMREGPNLLKLARKEQCLALGT
RLRSKYKITYQFYRVFPNGEVQYLHPKDGVYPEKANPGREGVGLNMRSIGKNVSPIEVKFTGKQSYDL

>CEG01861.1 Photosystem I PsaD [Ostreococcus tauri]
MSVTHAPARAVAKPTAMRTRKSARSATIVRAEEAKAKTESAAPAVWTAPKLDPNTPSPIFGGSTGGLLRK
AQVEEFYVLTWEAKKEAIFEMPTGGAAIMRKGPNLLKLARKEQCLALLNTFRSKMKLDGCIYRVFPSGEV
QYLHPKDGVYPEKVNKGRVGANQNMRSIGKNTNPAKIKFQGKLGPFEV

>BAA16688.1 photosystem I subunit II [Synechocystis sp. PCC 6803]
MTELSGQPPKFGGSTGGLLSKANREEKYAITWTSASEQVFEMPTGGAAIMNEGENLLYLARKEQCLALGT
QLRTKFKPKIQDYKIYRVYPSGEVQYLHPADGVFPEKVNEGREAQGTKTRRIGQNPEPVTIKFSGKAPYE
V

>XP_001702611.1 uncharacterized protein CHLRE_10g420350v5 [Chlamydomonas reinhardtii]
MQALSSRVNIAAKPQRAQRLVVRAEEVKAAPKKEVGPKRGSLVKILRPESYWFNQVGKVVSVDQSGVRYP
VVVRFENQNYAGVTTNNYALDEVVAAK

>NP_567818.2 Photosystem I reaction centre subunit IV / PsaE protein [Arabidopsis thaliana]
MAMTTASTVFVLPANVTSVAGASSSRSSVSFLPMRNAGSRLVVRAAEDPAPASSSSKDSPAAAAAPDGAT
ATKPKPPPIGPKRGSKVKILRRESYWFKNVGSVVAVDQDPKTRYPVVVRFAKVNYANISTNNYALDEVEE
VAA

>CAL52500.2 Photosystem I PsaE, reaction centre subunit IV [Ostreococcus tauri]
MFACTKVQANVFAVKTTKTVRRSTVTRAEGEAAKPAAKPQVGPPRGTMVKILRPESYWFNDYGKVISVDQ
TGVRYPVVVRFEKVNYAGVSTNNYALDEVEY

>BAA18383.1 photosystem I subunit IV [Synechocystis sp. PCC 6803]
MALNRGDKVRIKRTESYWYGDVGTVASVEKSGILYPVIVRFDRVNYNGFSGSASGVNTNNFAENELELVQ
AAAK

>XP_001696798.1 uncharacterized protein CHLRE_09g412100v5 [Chlamydomonas reinhardtii]
MALTMRNPAVKASSRVAPSSRRALRVACQAQKNETASKVGTALAASALAAAVSLSAPSAAMADIAGLTPC
SESKAYAKLEKKELKTLEKRLKQYEADSAPAVALKATMERTKARFANYAKAGLLCGNDGLPHLIADPGLA
LKYGHAGEVFIPTFGFLYVAGYIGYVGRQYLIAVKGEAKPTDKEIIIDVPLATKLAWQGAGWPLAAVQEL
QRGTLLEKEENITVSPR

>NP_174418.1 photosystem I subunit F [Arabidopsis thaliana]
MSLTIPANLVLNPRSNKSLTQSVPKSSARFVCSDDKSSSSTPQSMKAFSAAVALSSILLSAPMPAVADIS
GLTPCKDSKQFAKREKQQIKKLESSLKLYAPESAPALALNAQIEKTKRRFDNYGKYGLLCGSDGLPHLIV
NGDQRHWGEFITPGILFLYIAGWIGWVGRSYLIAISGEKKPAMKEIIIDVPLASRIIFRGFIWPVAAYRE
FLNGDLIAKDV

>XP_003078706.1 Photosystem I PsaF, reaction centre subunit III [Ostreococcus tauri]
MHAHAQAQLTSKVTVSGLKRSARVQTRAVKTVAHCSAVDFSKKAASLAVAAAVAAAPLVAVEEAFARDVQ
PYAGLTPCKTSKAFAKREKTELKALEKRLKKYDPESAPALALQATMEKTKTRFANYGESGLLCGKDGLPH
LIVDGNLEHLGEFAIPGLGFLYVAGWIGYAGRSYIQENKTASKPTEGEIIIDVPKALGLMFQAGAWPLLA
GLELKNGTLTAPESEITVSPR

>BAA18108.1 photosystem I subunit III [Synechocystis sp. PCC 6803]
MKHLLALLLAFTLWFNFAPSASADDFANLTPCSENPAYLAKSKNFLNTTNDPNSGKIRAERYASALCGPE
GYPHLIVDGRFTHAGDFLIPSILFLYIAGWIGWVGRSYLIEIRESKNPEMQEVVINVPLAIKKMLGGFLW
PLAAVGEYTSGKLVMKDSEIPTSPR

>CAA33257.1 polypeptide 35 precursor [Chlamydomonas reinhardtii]
MQTLASRPSLRASARVAPRRAPRVAVVTKAALDPQIVISGSTAAFLAIGRFVFLGYQRREANFDSTVGPK
TTGATYFDDLQKNSTIFATNDPAGFNIIDVAGWGALGHAVGFAVLAINSLQGANLS

>NP_175963.1 photosystem I subunit G [Arabidopsis thaliana]
MATSASALLSPTTFSTAISHKNPNSISFHGLRPLRLGGSSSALPKLSTTGRKSSSAVVRAELSPSIVISL

STGLSLFLGRFVFFNFQRENVAKQGLPEQNGKTHFEAGDDRAKEYVSLLKSNDPIGFNIVDVLAWGSIGH
IVAYYILATSSNGYDPSFFG

>CAL53836.1 Photosystem I PsaG/PsaK protein [Ostreococcus tauri]
MLARTALPTKTTASFNGRVAKQTRTAKRGAKVAARAFDDVNVVMSASNALALYLGRFVFLPYQRAQVERV
GMPTQNGQTHFAAGDTRAEEASFITSTNDPAGFNLVDVFAWGSIGHAIGFLALASQSAQQML

>XP_001690629.1 uncharacterized protein CHLRE_07g330250v5 [Chlamydomonas reinhardtii]
MALVARPVLSARVAASRPRVAARKAVRVSAKYGENSRYFDLQDMENTTGSWDMYGVDEKKRYPDNQAKFF
TQATDIISRRESLRALVALSGIAAIVTYGLKGAKDADLPITKGPQTTGENGKGGSVRSRL

>NP_188235.1 photosystem I subunit H-1 [Arabidopsis thaliana]
MASLATVAAVKPSAAIKGLGGSSLAGAKLSIKPSRLSFKPKSIRANGVVAKYGDKSVYFDLEDLGNTTGQ
WDVYGSDAPSPYNPLQSKFFETFAAPFTKRGLLLKFLILGGGSLLTYVSATSTGEVLPIKRGPQEPPKLG
PRGKL

>CEG01722.1 Photosystem I PsaH, reaction centre subunit VI [Ostreococcus tauri]
MAHVKVHTWTQIARNTTTPPTTTTTTTRERPARQSSMVAHAKTTAFATGAALARSTPTTRRARATVMTRA
KYGDESVYFDLSDVEATTGSWDVYGVDSASRYPEQQAAFFEAAAQGLGRREAVYSVLAVSAGLLTVAYGV
KGAKDAKLPITVGPQQPAQVGPRDRI

>XP_001703367.1 uncharacterized protein CHLRE_03g165100v5 [Chlamydomonas reinhardtii]
MALRAVSAKSAVRPTVARASVKPVAALKPAQKMALAGAASVALLAASSSSAEASQVIATVASAAQGYPFV
PPSWAPSVFVPLTGLVLPAIAMATLFVYIEKEAPSS

>NP_051069.1 photosystem I subunit VIII (chloroplast) [Arabidopsis thaliana]
MTTFNNLPSIFVPLVGLVFPAIAMASLFLHIQKNKIF

>YP_717255.1 photosystem I subunit VIII (chloroplast) [Ostreococcus tauri]
MAASLLPSIFVPLVGLVFPAVAMASLFLYIEKEQVS

>BAA19609.1 PsaI [Synechocystis sp. PCC 6803]
MDGSYAASYLPWILIPMVGWLFPAVTMGLLFIHIESEGEG

>NP_958417.1 photosystem I subunit IX (chloroplast) [Chlamydomonas reinhardtii]
MKDFTTYLSTAPVIATIWFTFTAGLLIEINRYFPDPLVFSF

>NP_051079.1 photosystem I subunit IX (chloroplast) [Arabidopsis thaliana]
MRDLKTYLSVAPVLSTLWFGSLAGLLIEINRLFPDALTFPFFSF

>YP_717261.1 PsaJ (chloroplast) [Ostreococcus tauri]
MKNFQIYLSTAPVLAAVWFTVLAGILIELNRFFPDALSFPLT

>BAA18107.1 photosystem I subunit IX [Synechocystis sp. PCC 6803]
MDGLKSFLSTAPVMIMALLTFTAGILIEFNRFYPDLLFHP

>XP_001697230.1 uncharacterized protein CHLRE_17g724300v5 [Chlamydomonas reinhardtii]
MQALATRPSAIRPTKAARRSSVVVRADGFIGSSTNLIMVASTTATLAAARFGLAPTVKKNTTAGLKLVDS
KNSAGVISNDPAGFTIVDVLAMGAAGHGLGVGIVLGLKGIGAL

>NP_174327.1 photosystem I subunit K [Arabidopsis thaliana]
MASTMMTTLPQFNGLRATKISAAPVQGLASVQPMRRKGNGALGAKCDFIGSSTNLIMVTSTTLMLFAGRF
GLAPSANRKATAGLRLEARDSGLQTGDPAGFTLADTLACGTVGHIIGVGVVLGLKNIGAI

>BAA16719.1 photosystem I subunit X [Synechocystis sp. PCC 6803]
MHSFLLATAVPATLSWSPKVAGVMIACNILAIAFGKLTIKQQNVGTPMPSSNFFGGFGLGAVLGTASFGH
ILGAGVILGLANMGVL

>XP_001691084.1 uncharacterized protein CHLRE_12g486300v5 [Chlamydomonas reinhardtii]
MAVAMRSSTGLRATAARRQMPLGLGRVSTVRVCAADTKKAQVISPVNGDPFVGMLETPVTSAPIVATYLS
NLPAYRTGVAPVLRGVEIGLAHGFLLAGPFIKLGPLRNVPETAEIAGSLSAAGLVLILALCLSIYGSAQF
QSTPSIGVKTLSGRSVARDPLFSADGWSEFAAGFLVGGEAGVAWAYVCTQILPYYS

>NP_193016.1 photosystem I subunit l [Arabidopsis thaliana]
MAASASPMASQLRSSFSSASLSQRLAVPKGISGAPFGVSPTKRVSSFTVRAVKSDKTTFQVVQPINGDPF
IGSLETPVTSSPLIAWYLSNLPGYRTAVNPLLRGVEVGLAHGFFLVGPFVKAGPLRNTAYAGSAGSLAAA
GLVVILSMCLTIYGISSFKEGEPSIAPSLTLTGRKKQPDQLQTADGWAKFTGGFFFGGISGVTWAYFLLY

```
VLDLPYFVK

>CEG00928.1 Photosystem I PsaL, reaction centre subunit XI [Ostreococcus tauri]
MVLASKLTLDVRSCIGAKSLRTERPRNSAIGSNKTSRRQTKVSASSDKVQIVKPINGDPFIGMLETPVTS
SPDVANFLSNLPAYRTGVAPLLRGVEVGLAHGFFVTGPFIKLGPLRSTDAAELAGCLSGAGLVLILTACL
SIYGATAFQREEVVGLKTLSGRTIAKDPLQSSEGWASFTSGWLVGGLSGVAWSYVLTQVLPYYS

>BAA18773.1 photosystem I subunit XI [Synechocystis sp. PCC 6803]
MAESNQVVQAYNGDPFVGHLSTPISDSAFTRTFIGNLPAYRKGLSPILRGLEVGMAHGYFLIGPWTLLGP
LRDSEYQYIGGLIGALALILVATAALSSYGLVTFQGEQGSGDTLQTADGWSQFAAGFFVGGMGGAFVAYF
LLENLSVVDGIFRGLFN

>YP_717257.1 photosystem I subunit XII (chloroplast) [Ostreococcus tauri]
MPLSDTQVFTALFVALITGILSLRLGTQLYK

>BAA17005.1 photosystem I PsaM subunit [Synechocystis sp. PCC 6803]
MALSDTQILAALVVALLPAFLAFRLSTELYK

>XP_001701700.1 photosystem I reaction center subunit N [Chlamydomonas reinhardtii]
MAISARSSIKMQAARPARATTVVVRASAGRRELLGLGALALGLTMTLAPVANAGVVSDLQAKSAANKALN
DKKRLATSYANLARSRTVYDGTCTFPENFFGCEELAFNKGVKFIAEDIKIECEGKTAKECGSKFTLRSN

>NP_201209.1 photosystem I reaction center subunit PSI-N, chloroplast, putative / PSI-N, putative (PSAN) [Arabidopsis
thaliana]
MAAMNSSVLTCSYAIAGSGSVELNQKVGLVNSSVGFGQKKQMIMPVIKAQRVVGDDVDGSNGRRSAMVFL
AATLFSTAAVSASANAGVIDEYLERSKTNKELNDKKRLATSGANFARAFTVQFGSCKFPENFTGCQDLAK
QKKVPFISEDIALECEGKDKYKCGSNVFWKW

>CEF98203.1 Photosystem I PsaN, reaction centre subunit N [Ostreococcus tauri]
MYTVCGANIIAAKAPTAAKKQDKVKRAVSLGLAGLAAATVSVAPAHADLTADLLARTEANKSLNDQKRAA
TSSANFERSRLVTDGFCQFPQNIFGCQNAAEKGSVKFLSDDMAIECEGTADGKTCSSKAPGAYPSFLGL

>XP_001700109.1 uncharacterized protein CHLRE_07g334550v5 [Chlamydomonas reinhardtii]
MAVAMRSAAMPSLASRPRVSSRRSVVVRAEASNKSFPRDWVKTDPLVPVLGFAGWTIPANIGVSAFGGQS
LFGLFTQSIGENLAHFPTGPALDDKFWLYLITYHLGLFLTITLGQIGVQGRKQGYW

>NP_563815.1 photosystem I subunit O [Arabidopsis thaliana]
MAATFATPSTVIGLGGSSITTKPFSSSFLKPTLSAKNPLRLAGASGGRVTCFERNWLRRDLNVVGFGLIG
WLAPSSIPAINGKSLTGLFFDSIGTELAHFPTPPALTSQFWLWLVTWHLGLFLCLTFGQIGFKGRTEDYF
```

Supplementary dataset 2. Comprehensive set of 158 predicted Lhc proteins in *E. gracilis.*

```
>LHCA5-Nter_Koziol/LHCA5-Nter@130
FFSDKKMYGTVSASSSNKATSTVAVCACAAVFAVAGFVAVQPRSLYAPAAVRPVVQATMT
TVPVAHSVQSNTVAAQIQSRPLNVESNVFAQVGSTFAPIALGVGALAAFVAALIHKGQQP
VAMAASSGER
>LHCB4-1_GEFR01021489.1/F-1/O1/P2@205
GKAPAAAKGGRPLWFPSAVAPDYLDGSLPGDRGFDPWGFAKPEELLQYTTGTADITLSKN
KDEEYVGKVKEERDVLTGEPLVPWSSAFGLKRLRECELIHGRWAMLFVLGALVGEAVTGV
AWQNAGLVEAREGYTYFGFSLPFSGAQITWFEVLTMGFVEIFRNTEMNIEKRVYPGGAFD
PLGLASKDAETRDISLLPTSETFNS
>LHCB4-2_GEFR01023973.1/F+2/O1/P1@144
LQLPHEIQKFMKIFELVSEVTCIHGTSPNPGVGLVEAREGYTYFGFSLPFSGSQITWFEV
LTMGFVEIFRNTEMNIEKRVYPGGAFDPLGLASKDAETTFRLKEAEIKHARLAMVAFLGF
ATAAAKTGAGATTAFSIWAASFGN
>LHCB4-3_Koziol/LHC_CP29@373
FRSIFFRFTKMYSEQTSYSRVAAAALLAVSAGACVGVAIAPAASSFYAPAAVRPISSQVA
VMNVPRMQLGAVNAPYYAENAATEVQAVDGTSNVNAWTILSGGLFAAAAAMAAFAFKGSR
TETSEQIAMLAAAGKKGKAPAAAKGGRPLWFPSAVAPDYLDGSLPGDRGFDPWGFAKPEE
LLQYTTGTADITLSKNKDEEYVGKVKEERDVLTGEPLVPWSSAFGLKRLRECELIHGRWA
MLFVLGALVGEAVTGVAWQNAGLVEAREGYTYFGFSLPFSGSQITWFEVLTMGFVEIFRN
TEMNIEKRVYPGGAFDPLGLASKDAETTFRLKEAEIKHARLAMVAFLGFATAAAKTGAGA
TTAFSIWAASFGN
>LHCB4-4_GEFR01069386.1/F+3/O1/P1@102
GKKGKAPAAAKGGRPLWFPSAVAPDYLDGSLPGDRGFDPWGVDLVFMLELSFSLGFCFIS
LPFLSGFAKPEELLQYTTGTADITLSKNKDEEYVGKVKEERD
>LHCB4-5_GEFR01056955.1/F-3/O3/P1@54
MQKPGIFLCCQLPKQLIAVQLSLPQITLKRTTFRLKEAEIKHARLAMVAFLGFA
>LHCB4-6_GEFR01056955.1/F-2/O2/P1@70
MGFVEIFRNTEMNIEKRVYPGGAFDPLGLASKDAETRDISLLPTSETVNCSAAFPPTNNP
EEDHFPPEGG
>LHCB7-1_GDJR01012385.1/F-2/O1/P1@432
SIFFREMKLFCFAVVAFVICLSFSSSLMHPFPQLICTSVDGKLCFRISFRKKTTPLHQLT
SLSRTPRESVWLSPRSRTTVGQASKPNDPADPPSSSTSRRWASFLTGVGASPLFYVSIGA
LAAAAAVKGSADSSAVVLTLAGLPVVGLTILKNLSVGESLRATLDQQLPQLEEAVALRLQ
AQAEAQRGSRYYGPSRVGRLFPAPPHLDGTLPGDAGFDPLGLARSDSRPKLPDTLASCRL
LMHKRRVALGLEAAPGAADTRPTLERLLDCEVLHGRWAQLAVVGCLVPSKLAEEAATLDY
LGIPGLHIAGGQGIAIIAACQFFLMGGPEYARYVGLRGLEPVGIFLDGTDVTYPGGQLFD
PLGLAKDPVAFQEQRVAEVKHGRLAMVAMAGMAAQALATRAGPVENLADFLADPGHNNVL
RSLHTFSHAFAS
>LHCB7-2_GDJR01012382.1+1/F-2/O1/P1@455
SIFFREMKLFCFAVVAFVICLSFSSSLMHPFPQLICTSVDGKLCFRISFRKKTTPLHQLT
SLSRTPRESVWLSPRSRATVGQASKPNDPADPPSSSTSRRWASFLTGVGASPLFYVSIGA
LAAAAAVKGSADSSAVVLTLAGLPVVGLTILKNLSVGESLRATLDQQLPQLEEAVALRLQ
AQAEAQRGSRYYGPSRVGRLFPAPPHLDGTLPGDAGFDPLGLARSDSRPKLPDTLASCRL
LMHKRRVALGLEAAPGAADTRPTLERLLDCEVLHGRWAQLAVVGCLVPEALGRALGSSVQ
FGETVWWKVGASKLAEEAAALDYLGIPGLHIAGGQGIAIIAACQFFLMGGPEYARYVGLR
GLEPVGIFLDGTDVTYPGGQLFDPLGLAKDPVAFQEQRVAEVKHGRLAMVAMAGMAAQAL
ATRAGPVENLADFLADPGHNNVLRSLHTFSHAFAS
>LHCB7-3_GEFR01016328.1+1/F-1/O3/P1@214
MYKRRVALGLEAAPGAADTRPTLERLLDCEVLHGRWAQLAVVGCLVPEALGRALGSSVQF
GETVWWKVGASKLAEEAAALDYLGIPGLHIAGGQGIAIIAACQFFLMGGPEYARYVGLRG
LEPVGIFLDGTDVTYPGGQLFDPLGLAKDPVAFQEQRVAEVKHGRLAMVAMAGMAAQALA
TRAGPVENLADFLADPGHNNVLRSLHTFSHAFAS
>LHCBM1-1_GDJR01073855.1/F-1/O1/P2@226
SKAPAADNLSQWYGPDRAKWLGPLTGQVPAYLTGELPGDYGWDTAGLGADPVTLARYREA
EVIHARWAMLGALGVVTPELLAGNGVPFGEGAVWYKAGAQIFSADGLNYLGNPSLIHAQS
VVLTFLSTLAIMGAVEAYRYGGGVGDFGRELDTLYPGGPFDPLGLANDPDALAELKVKEL
KNGRLAMVAMLGFYVQPLVTKAGPVENLTFHLADPSSNNIFSFTSG
>LHCBM1-2_HBDM01023705.1+2/F-1/O1/P2@226
```

```
SKAPAADNLSQWYGPDRAKWLGPLTGQVPAYLTGELPGDYGWDTAGLGADPVTLARYREA
EVIHARWAMLGALGVVTPELLAGNGVPFGEGAVWYKAGAQIFSADGLNYLGNPSLIHAQS
VVLTFLSTLAIMGAVEAYRYGGGVGDFGRELDTLYPGGPFDPLGLANDPDALAELKVKEL
KNGRLAMVAMLGFYVQPLVTKAGPVENLTFHLADPSANNIFSFTSG
>LHCBM1-3_HBDM01068881.1+3/F+2/O1/P2@226
SKAPAADNLSQWYGPDRAKWLGPLTGQVPAYLTGELPGDYGWDTAGLGADPTTLARYREA
EVIHARWAMLGALGVVTPELLAGNGVPFGEGAVWYKAGAQIFSADGLNYLGNPSLIHAQS
VVLTFLSTLAIMGAVEAYRYGGGVGDFGRELDTLYPGGPFDPLGLANDPDALAELKVKEL
KNGRLAMVAMLGFYVQPLVTKAGPVENLTFHLADPSSNNIFSFTSG
>LHCBM1-4_Koziol/LHCBM1-1@154
XGNGVPFGEGAVWYKAGAQIFSADGLNYLGNPSLIHAQSVVLTFLSTLAIMGAVEAYRYG
GGVGDFGRELDTLYPGGPFDPLGLANDPDALAELKVKELKNGRLAMVAMLGFYVQPLVTK
AGPVENLTFHLADPSANNIFSFTSGFAMFAATGSK
>LHCBM1-5_Koziol/LHCBM1-3@236
SAAPKSDNLSQWYGPDRAKWLGPLTGEVPSYLTGELPGDYGWDTAGLGSDPTTLARYREA
EVIHARWAMLGALGVVTPELLAGNGVPFGEGAVWYKAGAQIFSADGLNYLGNPSLIHAQS
VVLTFLSTLAIMGAVEAYRYGGGVGDFGRELDTLYPGGPFDPLGLANDPDALAELKVKEL
KNGRLAMVAMLGFYVQPLVTKAGPVENLTFHLADPSANNIFSFTSGFAMFAAAGKK
>LHCBM1-6_Koziol/LHCBM4-2@235
SAPKSDNLSQWYGPDRAKWLGPLTGEVPSYLTGELPGDYGWDTAGLGSDPTTLARYREAE
VIHARWAMLGALGVVTPELLAGNGVPFGEGAVWYKAGAQIFSADGLNYLGNPSLIHAQSV
VLTFLSTLAIMGAVEAYRYGGGVGDFGRELDTLYPGGPFDPLGLANDPDALAELKVKELK
NGRLAMVAMLGFYVQPLVTKAGPVENLTFHLADPSSNNIFSFTSGFAMFAASGRK
>LHCBM1-7_Koziol/LHCBM8-3@224
SAPKSDNLSQWYGPDRAKWLGPLTGEVPSYLTGELPGDYGWDTAGLGSDPTTLARYREAE
VIHARWAMLVTPELLAGNGVPFGEGAVWYKAGAQIFSADGLNYLGNPSLIHAQSVVLTFL
STLAIMGAVEAYRYGGGVGDFGRELDTLYPGGPFDPLGLANDPDALAELKVKELKNGRLA
MVAMLGFYVQPLVTKAGPVENLTFHLADPSSNNIFSFTSGFAMF
>LHCBM1-8_GEFR01011081.1/F-2/O1/P1@250
VTLLNGGLWFGYGLGGTNDPFIYVPNGIGMVAAIVQIVTVLIFCGKRCTKSQAANDTPIP
GAELHSVSTFPLQADESKSLRLSFSELPDDLSYSPTIDSHHGAEALHGRWAMLSALGVVT
PELLSSNSVPFSGGAVWSMQSIQMVSRIALNYLGNDRLIHAQSVGLTFLSALAIVHAVEF
SKREEVRELQRYTGDKHAAFDHSSVAVALVAMLGSHIYPLVDKAGSVENLVFHLADSSKD
NIFSFTSLHR
>LHCBM1-9_HBDM01007123.1/F-2/O1/P1@431
FFFAFDLSPMADIVTAYVFPGIGVGISCALFGSSALSVWRAFQSRNIEAIDSPTLAMQFI
NTLMWMRYGVIKDDIFVWVSNAPGSLITGFYMGVVFFLCGHKGDMRTLRLIAGILALGLV
YLMLDISVIAFANVDKSVTNLMAGFSTNIWLGGMYLAPIMNVIRAFREGNADFIYFPLVI
VTLLNGGLWFGYGLGGANDPFIYVPNGIGMVAAIVQIVTVLIFCGKRCASKGQAASDTPV
PGADLHSISTFPLQADESKSLCLSFSELPDDLSYSPTIDSFRGAEALHGRWAMLSALGVV
TPELLSSNSVPFSGGAVWSMQSIQLVSRVALNYLGNDRLIHAQSVGLTFLSALSIAHAVE
SSKREEVRELQRYSGDKHAAFDHRSVAVALVAMLGSHIYPLVDKAGSVENLVFHLADSSK
NNIFSFPSSHH
>LHCBM1-10_GEFR01057898.1/F+3/O2/P2@60
SAPKSDNLSDRAKWLGPLTGEVPSYLTGELPGDYGWDTAGLGSDPTTLARYREAEVIHAR
>LHCBM1-11_GEFR01023953.1+1/F-2/O1/P2@58
SAPKSDNLSQWYGPGQACPPVFSPSLTHSPPPPKHHPPQQGHHLQRPLQRSVQPARSV
>LHCBM1-12_GEFR01057898.1/F+3/O2/P1@28
MAKAGPVENLTFHLADPSSNNIFSFTSG
>LHCBM2-1_Koziol/LHCBM1-2@225
WYGPNRPKWLGPLSGGAVPEYLKGEYAGDYGFDTAGLAADPKLFQRYRDAELQNGRWAML
GVLGCLAPEVLSNVFGVPYPEPVWFKTGATILNGGSIDYLANPKLIHASNLLLTLVLELV
FFFAAETWREAGEGPLGKAQDKSYPGGVFDPLGLSKDPAAFAEAKVKEVKNGRLAMLAML
GLFVQAGVTGQSPLENLSAHLANPGVNFWTSYAPTLAMFAASGRK
>LHCBM2-2_GEFR01059358.1/F-3/O1/P1@103
QRLLVIGHRGGVVWYFCCKLSKFHARGTGIVWRLWCHTGYVSVFFQTDHPAQRYRDAELQ
NGRWAMLGVLGCLAPEVLSNVFGVPYPEPVWFKATPSLRGSSV
>LHCBM3-1_HBDM01050735.1/F+2/O1/P2@182
SKAPAADNLSQWYGPDRAKWLGPLTGQVPAYLTGELPGDYGWDTAGLGADPTTLARYREA
EVIHARWAMLGALGVVTPELLAGNGVPFGEGAVWYKAGAQIFSADGLNYLGNPSLIHAQS
```

```
VVLTFLSTLAIMGAVEGYRYGGGWPLGDDLDRLYPGGPFDPLGLANDPDAFAELKVKELK
NG
>LHCBM3-2_GEFR01023953.1+1/F-2/O1/P1@132
WYKAGAQIFSADGLNYLGNPSLIHAQSVVLTFLSTLAIMGAVEGYRYGGGWPLGDDLDRL
YPGGPFDPLGLANDPDAFAELKVKELKNGRLAMVAMLGFYVQPLVTKSGPLANLSLHLAD
PSSNNIFSFTSG
>LHCBM3-3_GEFR01044404.1+1/F-2/O1/P1@122
WMPTPSPPPRASKKPVSVVLTFLSTLAIMGAVEGYRYGGGWPLGDDLDRLYPGGPFDPLG
LANDPDAFAELKVKELKNGRLAMVAMLGFYVQPLVTKSGPLANLSLHLADPSNNNIFSFT
SG
>LHCBM3-4_GEFR01016314.1/F-2/O1/P2@225
SKAPAADNLSQWYGPDRAKWLGPLTGQVPAYLTGELPGDYGWDTAGLGADPVTLARYREA
EVIHARWAMLGALGVVTPELLAGNGVPFGEGAVWYKAGAQIFSADGLNYLGNPSLIHAQS
VVLTFLSTLAIMGAVEGYRYGGGWPLGDDLDRLYPGGPFDPLGLANDPDAFAELKVKELK
NGRLAMVAMLGFYVQPLVTKSGPLANLSYHLADPSANNIFSFTSG
>LHCBM3-5_GEFR01039900.1/F+2/O1/P1@88
GNPSLIHAQSVVLTFLSTLAIMGAVEGYRYGGGWPLGDDLDRLYPGEHDLVQGKFSRCPR
ERAVAKRILGLRGPGLQAITTFGQNGGN
>LHCBM3-6_GEFR01057898.1/F-1/O2/P1@49
MAKPEVKLKMLLLEGSAKWKVRFSTGPALAMLGFYVQPLVTKAGPVENL
>LHCBM3-7_GDJR01073858.1+1/F-2/O1/P1@55
FAELKVKELKNGRLAMVAMLGFYVQPLVTKSGPLANLSYHLADPSSNNIFSFTSG
>LHCBM5-1_GDJR01073851.1/F-3/O1/P1@234
AKAAPANLSAWYGPNRNKWLGPFSEGSTPSYLTGEFPGDYGWDTAGLAADPITFAKYREA
ELIHARWAMLGALGVVFPEALAKAGVPLAEPVWFKAGAQIFSEGGLDYLGNSNLVHAQSI
LAILAFQVVLMGLAEAYRVNGGPLGEDSDALYPGGAFDPLGLADDPETFAELKVKELKNG
RLAMFAMFGFYVQALVTGEGPVENWAAHVADPFNVNGLTAVYVTQFAPSPVAMF
>LHCBM5-2_GDJR01073854.1/F-3/O1/P1@230
AKAAPANLSAWYGPNRNKWLGPFSEGSTPSYLTGEFPGDYGWDTAGLAADPITFAKYREA
ELIHARWAMLGALGVVFPEALAKAGVPLAEPVWFKAGAQIFSEGGLDYLGNSNLVHAQSI
LAILAFQVVLMGLAEAYRVNGGPLGEDSDALYPGGAFDPLGLADDPETFAELKVKELKNG
RLAMFAMFGFYVQALVTGEGPVENWAAHVADPFNVNGLTAVYVTQFAPTP
>LHCBM5-3_GDJR01073855.1/F-1/O1/P3@229
SAPKSDNLSQWYGPDRAKWLGPLTGEVPSYLTGELPGDYGWDTAGLAADPITFAKYREAE
LIHARWAMLGALGVVFPEALAKAGVPLAEPVWFKAGAQIFSEGGLDYLGNSNLVHAQSIL
AILAFQVVLMGLAEAYRVNGGPLGEDSDALYPGGAFDPLGLADDPETFAELKVKELKNGR
LAMFAMFGFYVQALVTGEGPVENWAAHVADPFNVNGLTAVYVTQFAPTP
>LHCBM5-4_GDJR01073860.1/F-1/O1/P3@236
KPAAGAVSKDPFSPLNKLYGPDRVKWLGGFTGFVPEYLTGELPGDYGWDTAGLAADPITF
AKYREAELIHARWAMLGALGVVFPEALAKAGVPLAEPVWFKAGAQIFSEGGLDYLGNSNL
VHAQSILAILAFQVVLMGLAEAYRVNGGPLGEDSDALYPGGAFDPLGLADDPETFAELKV
KELKNGRLAMFAMFGFYVQALVTGEGPVENWAAHVADPFNVNGLTAVYVTQFAPTP
>LHCBM5-5_GDJR01073863.1/F-3/O1/P2@230
SAPASDKLAAWYGPNRNKWLGPFSEGSTPSYLTGEFPGDYGWDTAGLAADPITFAKYREA
ELIHARWAMLGALGVVFPEALAKAGVPLAEPVWFKAGAQIFSEGGLDYLGNSNLVHAQSI
LAILAFQVVLMGLAEAYRVNGGPLGEDSDALYPGGAFDPLGLADDPETFAELKVKELKNG
RLAMFAMFGFYVQALVTGEGPVENWAAHVADPFNVNGLTAVYVTQFAPTP
>LHCBM5-6_GDJR01073871.1/F-3/O1/P2@234
SAPASDKLAAWYGPNRNKWLGPFSEGSTPSYLTGEFPGDYGWDTAGLAADPITFAKYREA
ELIHARWAMLGALGVVFPEALAKAGVPLAEPVWFKAGAQIFSEGGLDYLGNSNLVHAQSI
LAILAFQVVLMGLAEAYRVNGGPLGEDSDALYPGGAFDPLGLADDPETFAELKVKELKNG
RLAMFAMFGFYVQALVTGEGPVENWAAHVADPFNVNGLTAVYVTQFAPSPVAMF
>LHCBM5-7_GDJR01073872.1/F-3/O1/P1@235
AKAAPANLSAWYGPNRNKWLGPYSDRSVPSYLTGELPGDYGWDTAGLGADPFTLQRYREA
ELIHGRWAMLGALGVVFPEYLSRNFDIPIAEPVWFKAGAQIFSEGGLDYLGNSNLVHAQS
ILAILAFQVVLMGLAEAYRVNGGPLGEDSDALYPGGAFDPLGLADDPETFAELKVKELKN
GRLAMFAMFGFYVQALVTGEGPVENWAAHVADPFNVNGLTAVYVTQFAPSPVAMF
>LHCBM5-8_GDJR01073858.1+1/F-2/O1/P3@233
SAPKSDNLSQWYGPDRAKWLGPLTGEVPSYLTGELPGDYGWDTAGLAADPITFAKYREAE
LIHARWAMLGALGVVFPEALAKAGVPLAEPVWFKAGAQIFSEGGLDYLGNSNLVHAQSIL
```

AILAFQVVLMGLAEAYRVNGGPLGEDSDALYPGGAFDPLGLADDPETFAELKVKELKNGR
LAMFAMFGFYVQALVTGEGPVENWAAHVADPFNVNGLTAVYVTQFAPSPVAMF
>LHCBM5-9_GDJR01073862.1+3/F-3/O1/P1@231
AKAAPANLSAWYGPNRNKWLGPYSDRSVPSYLTGELPGDYGWDTAGLGADPFTLQRYREA
ELIHGRWAMLGALGVVFPEYLSRNFDIPIAEPVWFKAGAQIFSEGGLDYLGNSNLVHAQS
ILAILAFQVVLMGLAEAYRVNGGPLGEDSDALYPGGAFDPLGLADDPETFAELKVKELKN
GRLAMFAMFGFYVQALVTGEGPVENWAAHVADPFNVNGLTAVYVTQFAPTP
>LHCBM5-10_HBDM01068881.1+3/F+2/O1/P3@240
KPAAGAVSKDPFSPLNKLYGPDRVKWLGGFTGFVPEYLTGELPGDYGWDTAGLAADPITF
AKYREAELIHARWAMLGALGVVFPEALAKAGVPLAEPVWFKAGAQIFSEGGLDYLGNSNL
VHAQSILAILAFQVVLMGLAEAYRVNGGPLGEDSDALYPGGAFDPLGLADDPETFAELKV
KELKNGRLAMFAMFGFYVQALVTGEGPVENWAAHVADPFNVNGLTAVYVTQFAPSPVAMF
>LHCBM5-11_Koziol/LHCM6-2@240
STPASDKLAAWYGPNRNKWLGPFSEGSTPSYLTGEFPGDYGWDTAGLAADPITFAKYREA
ELIHARWAMLGALGVVFPEALAKAGVPLAEPVWFKAGAQIFSEGGLDYLGNSNLVHAQSI
LAILAFQVVLMGLAEAYRVNGGPLGEDSDALYPGGAFDPLGLADDPETFAELKVKELKNG
RLAMFAMFGFYVQALVTGEGPVENWAAHVADPFNVNGLTAVYVTQFAPTPVAMFATAGRK
>LHCBM5-12_Koziol/LHCM6-3@240
STPASDKLAAWYGPNRNKWLGPFSENSTPSYLTGEFPGDYGWDTAGLAADPITFAKYREA
ELIHARWAMLGALGVVFPEALAKAGVPLAEPVWFKAGAQIFSEGGLDYLGNSNLVHAQSI
LAILAFQVVLMGLAEAYRVNGGPLGEDSDALYPGGAFDPLGLADDPETFAELKVKELKNG
RLAMFAMFGFYVQALVTGEGPVENWAAHVADPFNVNGLTAVYVTQFAPTPVAMFATAGRK
>LHCBM5-13_Koziol/LHCBM6-4@234
STPASDKLAAWYGPNRNKWLGPFSENSTPSYLTGEFPGDYGWDTAGLAADPITFAKYREA
ELIHARWAMLGALGVVFPEALAKAGVPLAEPVWFKAGAQIFSEGGLDYLGNSNLVHAQSI
LAILAFQVVLMGLAEAYRVNGGPLGEDSDALYPGGAFDPLGLADDPETFAELKVKELKNG
RLAMFAMFGFYVQALVTGEGPVENWAAHVADPFNVNGLTAVYVTQFAPSPVAMF
>LHCBM6-1_GDJR01073861.1/F-3/O1/P1@233
AKAAPANLSAWYGPNRNKWLGPYSDRSVPSYLTGELPGDYGWDTAGLGADPFTLQRYREA
ELIHGRWAMLGAVGCFLPEYLNRFFGVQFGEPIWFKAGAQIFSAEGLNYFGQPALIHAKS
LLAILATQVLLMGAVEAYRAGNTAPGQFGEDLDRLYPGGPFDPLGLADDPDTFAELKVKE
IKNGRLAMFAMFGFYVQAIVTGESPLENLSAHLASPFTVNGLNAAYMTQFTPT
>LHCBM6-2_GDJR01073864.1/F-1/O1/P3@239
KPAAGAVSKDPFSPLNKLYGPDRVKWLGGFTGFVPEYLTGELPGDYGWDSAGLGADPITL
QRYREAELIHGRWAMLGAVGCFLPEYLNRFFGVQFGEPIWFKAGAQIFSAEGLNYFGQPA
LIHAKSLLAILATQVLLMGAVEAYRAGNTAPGQFGEDLDRLYPGGPFDPLGLADDPDTFA
ELKVKEIKNGRLAMFAMFGFYVQAIVTGESPLENLSAHLASPFTVNGLNAAYMTQFTPT
>LHCBM6-3_GDJR01073865.1/F-2/O1/P2@230
APAGLSQWYGPDRSKWLGPYSDALTPPYLNGELPGDYGWDSAGLGADPITLQRYREAELI
HGRWAMLGAVGCFLPEYLNRFFGVQFGEPIWFKAGAQIFSAEGLNYFGQPALIHAKSLLA
ILATQVLLMGAVEAYRAGNTAPGQFGEDLDRLYPGGPFDPLGLADDPDTFAELKVKEIKN
GRLAMFAMFGFYVQAIVTGESPLENLSAHLASPFTVNGLNAAYMTQFTPT
>LHCBM6-4_GDJR01073862.1+3/F-3/O1/P2@234
STPASDKLAAWYGPNRNKWLGPYSDRSVPSYLTGELPGDYGWDTAGLGADPFTLQRYREA
ELIHGRWAMLGAVGCFLPEYLNRFFGVQFGEPIWFKAGAQIFSAEGLNYFGQPALIHAKS
LLAILATQVLLMGAVEAYRAGNTAPGQFGEDLDRLYPGGPFDPLGLADDPDTFAELKVKE
IKNGRLAMVAMLGFYAQAIVTGEGPVENWLYHLQDPSAHNGLTAALVTQFAPTP
>LHCBM6-5_Koziol/LHCM5-1@217
VPSYLTGELPGDYGWDTAGLGADPFTLQRYREAELIHGRWAMLGAVGCFLPEYLNRFFGV
QFGEPIWFKAGAQIFSAEGLNYFGQPALIHAKSLLAILATQVLLMGAVEAYRAGNTAPGQ
FGEDLDRLYPGGPFDPLGLADDPDTFAELKVKEIKNGRLAMVAMLGFYAQAIVTGEGPVE
NWLYHLQDPSAHNGLTAALVTQFAPTPVAMFAAGGSK
>LHCBM6-6_Koziol/LHCM5-3@238
AKAAPANLSAWYGPNRNKWLGPYSDRSVPSYLTGELPGDYGWDTAGLGADPFTLQRYREA
ELIHGRWAMLGAVGCFLPEYLNRFFGVQFGEPIWFKAGAQIFSAEGLNYFGQPALIHAKS
LLAILATQVLLMGAVEAYRAGNTAPGQFGEDLDRLYPGGPFDPLGLADDPDTFAELKVKE
IKNGRLAMMAMLGFYAQAIVTGEGPVENWLYHLQDPSAHNGLTAALVTQFAPTPVALL
>LHCBM6-7_Koziol/LHCBM7-b@239
KAKAAPANLSAWYGPNRNKWLGPYSDRSVPSYLTGELPGDYGWDTAGLGADPFTLQRYRE
AELIHGRWAMLGAVGCFLPEYLNRFFGVQFGEPIWFKAGAQIFSAEGLNYFGQPALIHAK

```
SLLAILATQVLLMGAVEAYRAGNTAPGQFGEDLDRLYPGGPFDPLGLADDPDTFAELKVK
EIKNGRLAMVAMLGFYAQAIVTGEGPVENWLYHLQDPSAHNGLTAALVTQFAPTPVALL
>LHCBM6-8_GEFR01049430.1/F+2/O1/P1@26
LTGELPGDYGWDTAGLGAGQTHIWLL
>LHCBM7-1_GDJR01073874.1/F-3/O1/P1@232
AKAAPANLSAWYGPNRNKWLGPYSDRSVPSYLTGELPGDYGWDTAGLGADPFTLQRYREA
ELIHGRWAMLGALGVVFPEYLSRNFDIPIAEPVWFKAGAQIFSEGGLDYLGDARLVHAQS
ILLILGTQVLLMGAAEAYRTGNGLGNFGKDLDSLYPGGPFDPLGLANDPDTLAELKVKEI
KNGRLAMFAMFGFYVQAIVTGESPLENLSAHLASPFTVNGLNAAYMTQFTPT
>LHCBM7-2_GDJR01073862.1+3/F-3/O1/P3@223
APAGLSQWYGPDRSKWLGPYSDALTPPYLNGELPGDYGWDSAGLGADPITLQRYREAELI
HGRWAMLGALGVVFPEYLSRNFDIPIAEPVWFKAGAQIFSEGGLDYLGDARLVHAQSILL
ILGTQVLLMGAAEAYRTGNGLGNFGKDLDSLYPGGPFDPLGLANDPDALAELKVKELKNG
RLAMVAMLGFYVQPLVTKAGPVENLTFHLADPSSNNIFSFTSG
>LHCBM7-3_HBDM01023705.1+2/F-1/O1/P4@233
SAAPKSDNLSQWYGPDRAKWLGPLTGEVPSYLTGELPGDYGWDTAGLGSDPATLARYRVA
EVIHGRWAMLGALGVVFPEYLSRNFDIPIAEPVWFKAGAQIFSEGGLDYLGDARLVHAQS
ILLILGTQVLLMGAAEAYRTGNGLGNFGKDLDSLYPGGPFDPLGLANDPDTLAELKVKEI
KNGRLAMFAMFGFYVQAIVTGESPLENLSAHLASPFTVNGLNAAYMTQFTPTP
>LHCBM7-5_Koziol/LHCBM5-2@240
APAGLSQWYGPDRSKWLGPYSDALTPPYLNGELPGDYGWDSAGLGADPITLQRYREAELI
HGRWAMLGALGVVFPEYLSRNFDIPIAEPVWFKAGAQIFSEGGLDYLGDARLVHAQSILL
ILGTQVLLMGAAEAYRTGNGLGNFGKDLDSLYPGGPFDPLGLANDPDTLAELKVKEIKNG
RLAMFAMFGFYVQAIVTGESPLENLSAHLASPFTVNGLNAAYMTQFTPTPVAMFATAGKK
>LHCBM7-6_Koziol/LHCBM7-a@229
PDRSKWLGPYSDALTPPYLNGELPGDYGWDSAGLGADPITLQRYREAELIHGRWAMLGAL
GVVFPEYLSRNFDIPIAEPVWFKAGAQIFSEGGLDYLGDARLVHAQSILLILGTQVLLMG
AAEAYRTGNGLGNFGKDLDSLYPGGPFDPLGLANDPDTLAELKVKEIKNGRLAMFAMFGF
YVQAIVTGESPLENLSAHLASPFTVNGLNAAYMTQFMPTPVAMFATAGK
>LHCBM7-7_GEFR01023953.1+1/F-3/O4/P1@56
MPADAFSGAVDGAAPGGRLGADRAKWLGPLTGEVPSYLTGELPGDYGWDTAGLGSD
>LHCBM8-1_GDJR01021843.1/F+3/O1/P1@161
CIVPELLARNYGVPFPEPVWFKTGATVFSEEGLNYLGNPSLIHAKSIAAILVTEILFVGA
AEAFRVSGGPLGPATDLVYPGKAFDPLGLSKDETAFAELKVKEVKNGRLAMLGMLGLFMQ
GFATGKGPLQNLADHLADPTGANIITFHQRVLSDLGGVAMF
>LHCBM8-2_HBDM01023680.1+4/F-1/O1/P2@223
LSDWYGPNRKLFLGPLSDGAPEHLKGELPGDYGFDVLGLATQPTRLERYRQGEIINGRWA
MLGIVGCIVPELLARNYGVPFPEPVWFKTGATVFSEEGLNYLGNPSLIHAKSIAAILVTE
ILFVGAAEAFRVSGGPLGPATDLVYPGKAFDPLGLSKDETAFAELKVKEVKNGRLAMLGM
LGLFMQGFATGKGPLQNLADHLADPTGANIITFHQRVLSDLGG
>LHCBM8-3_GEFR01041770.1/F-2/O1/P1@106
WVGGLTPSPPRVGFRKELEGPEHIFWVFSPMFADISPLAPSKWRSQWKPQVALAAMLGML
GLFMQGFATGKGPLENLADHLADPTGANIITFHQRVLSDLGGVAMF
>LHCBM8-4_Koziol/LHCBM3-1@100
GGPLGPATDLVYPGKAFDPLGLSKDETAFAELKVKEVKNGRLAMLGMLGLFMQGFATGKG
PLQNLADHLADPTGANIITFHQRVLSDLGGVAMFAATGEK
>LHCBM8-5_GEFR01061045.1/F-2/O1/P1@112
TEILFVGAAEAFRVSGGPLGPATDLVYPGKAFDPLGLSKDETAFAELKVKEVKNGRLAIR
AALLALVICLQQMPAGQILQILGKVELHVQPTGKFDPSWQLIVTRLGLRPDF
>LHCBM8-6_GEFR01030289.1/F-2/O1/P1@70
DETAFAELKVKEVKNGRLAMLGMLGLFMQGFATGKGPLENLADHLADPTGANIITFHQRV
LSDLGGVAMF
>LHCBMX1-1_GEFR01040151.1/F-2/O1/P1@113
DDPEAFLELKVKEVKNGRLAMVAIFGFFVQGILTGKGPVENWVDHLTDPFENAQRHLYPG
VFDYRLAFNFWCGGVFRLFILSVPLCLVWWCFGCSLHLVRAFLWSPAISRRIN
>LHCBMX1-2_GEFR01056651.1/F+3/O1/P1@86
GDFGADLDPLYPGGPFDPLGLADDPEAFLELKVKEVKNGRLAMVAIFGFFVQGILTGKGP
VENWVDHLTDPFVNNIFQLTPGFAMF
>LHCBMX2-1_GDJR01073856.1/F-2/O1/P2@230
SAPKSDNLSQWYGPDRAKWLGPLTGEVPSYLTGELPGDYGWDTAGLAADPITFAKYREAE
```

```
LIHARWAMLGALGVVFPEALAKAGVPLAEPVWFKAGAQIFSEGGLDYLGDARLVHAQSIL
LILGTQVLLMGAAEAYRTGNGLGNFGKDLDSLYPGGPFDPLGLANDPDTLAELKVKEIKN
GRLAMFAMFGFYVQAIVTGESPLENLSAHLASPFTVNGLNAAYMTQFTPT
>LHCBMX2-2_GEFR01016314.1/F-2/O1/P3@32
SAPKSAPKSDNLSQWYGPDRAKWLGPLTGEVP
>LHCBMX2-3_GDJR01073862.1+3/F-3/O1/P7@237
KPAAGAVSKDPFSPLNKLYGPDRVKWLGGFTGFVPEYLTGELPGDYGWDTAGLAADPITF
AKYREAELIHARWAMLGALGVVFPEALAKAGVPLAEPVWFKAGAQIFSEGGLDYLGDARL
VHAQSILLILGTQVLLMGAAEAYRTGNGLGNFGKDLDSLYPGGPFDPLGLANDPDTLAEL
KVKEIKNGRLAMFAMFGFYVQAIVTGESPLENLSAHLASPFTVNGLNAAYMTQFTPT
>LHCBMX3-1_GDJR01073865.1/F-2/O1/P1@185
SDPTTLARYREAEVIHARWAMLGALGVVTPELLAGNGVPFGEGAVWYKAGAQIFSADGLN
YLGNPSLIHAQSVVLTFLSTLAIMGAVEAYRYGGGVGDFGRELDTLYPGGPFDPLGLAND
PDALAELKVKELKNGRLAMVAMLGFYAQAIVTGEGPVENWLYHLQDPSAHNGLTAALVTQ
FAPTP
>LHCBMX3-2_GDJR01073869.1/F-1/O1/P2@237
SKAPAADNLSQWYGPDRAKWLGPLTGQVPAYLTGELPGDYGWDTAGLGADPVTLARYREA
EVIHARWAMLGALGVVTPELLAGNGVPFGEGAVWYKAGAQIFSADGLNYLGNPSLIHAQS
VVLTFLSTLAIMGAVEAYRYGGGVGDFGRELDTLYPGGPFDPLGLANDPDALAELKVKEL
KNGRLAMVAMLGFYAQAIVTGEGPVENWLYHLQDPSAHNGLTAALVTQFAPTPVALL
>LHCBMX4-1_GDJR01073867.1/F-2/O1/P1@184
SDPTTLARYREAEVIHARWAMLGALGVVTPELLAGNGVPFGEGAVWYKAGAQIFSADGLN
YLGNPSLIHAQSVVLTFLSTLAIMGAVEAYRYGGGVGDFGRELDTLYPGGPFDPLGLADD
PDTFAELKVKEIKNGRLAMFAMFGFYVQAIVTGESPLENLSAHLASPFTVNGLNAAYMTQ
FTPT
>LHCBMX5-1_HBDM01020048.1+3/F-1/O1/P2@232
SAPASDKLAAWYGPNRNRWLGPFSEGSTPSYLTGEFPGDYGWDTAGLAADPITFAKYREA
ELIHARWAMLGALGVVFPEALAKAGVPLAEPVWFKAGAQIFSEGGLDYLGDARLVHAQSI
LLILGTQVLLMGAAEAYRTGNGLGNFGKDLDSLYPGGPFDPLGLANDPDTLAELKVKEIK
NGRLAMFAMFGFYVQAIVTGESPLENLSAHLASPFTVNGLNAAYMTQFTPTP
>LHCBMX5-2_GDJR01073870.1/F-3/O1/P1@231
AKAAPANLSAWYGPNRNKWLGPFSEGSTPSYLTGEFPGDYGWDTAGLAADPITFAKYREA
ELIHARWAMLGALGVVFPEALAKAGVPLAEPVWFKAGAQIFSEGGLDYLGDARLVHAQSI
LLILGTQVLLMGAAEAYRTGNGLGNFGKDLDSLYPGGPFDPLGLANDPDTLAELKVKEIK
NGRLAMFAMFGFYVQAIVTGESPLENLSAHLASPFTVNGLNAAYMTQFTPT
>LHCBMX5-3_GEFR01071193.1/F-3/O1/P1@65
LGNFGKDLDSLYPGGPFDPLGLANDPDTLAELKVKEIKNGRLAMFAMFGFYVQAIVTGEQ
CWTGA
>LHCE1-1_GEFR01006499.1/F+2/O1/P1@153
FDPLGLSTDPVVFERMRISEVFHGRLAMLGIVGCIVPELFFSQGPWFDLSSFNLTRLGLV
ALQVIAPLEYWRGNRGFSWNGDDGPDQSYPGFDPLRLTDDETKLKEIKNGRLAMTAMLGL
EVQSHITGKSPLVNLSEHLSNPASVNILTGGGS
>LHCE1-2_GEFR01025906.1/F+3/O2/P1@250
MYGAVQGPSSSKTTSVVALCSCAALCAVAGFVAAQPGYLYAPTAVRPAVQSSVATTPLTT
VVRPSFATAAQRESASWSTERSGLPQIDSTVPLALGVGALTALITTILYKSKQSIALAAT
ASTRGQSTRGVSTRGSTRGQSTRGQRSTRGNAPEEERLVWFPNATPPDHLTGEFPGDRGF
DPLGLSKDPVVFERMRISEVFHGRLAMLGIVGCGVPEPLPRQAAATKVIAPLEYWRGNRG
FSWNGDDGPD
>LHCE1-3_HBDM01068858.1/F+1/O1/P1@342
LFFFVGNNMYRAVQGPSSSKTTSAVALCSCAALCAVAGFVAAQPGYLYAPTTVRPAVQSS
VATAPLNTVVRPSFATTAQREPAVWSPERSGLPQIGANFVPIALGVGALTALITTILFKS
KQSIALAATAGQSTRGQSTRGGSARGQRFGSTRGQKSTRGNAPAEERLVWFPNATPPDHL
TGEFPGDRGFDPLGLSKDPVVFERMRISEVFHGRLAMLGIVGCIVPELFFSQGPWFDLSG
FNLTRLGLVALQVIAPLEYWRGNRGFSWDGDDGPDQSYPGFDPLRLTDEETKLKEIKNGR
LAMTAMLGLEVQSHLTGKSPLVNLSEHLSNPASVNILTGGGS
>LHCE1-4_HBDM01071460.1/F-3/O1/P1@339
LFFFVDSMYGAVQGPSSSKTTSVVALCSCAALCAVAGFVAAQPGYLYAPTAVRPAVQSSA
STTPLTTVVRPSFATAAQRESASWSTERSGLPQIDSTVPLALGVGALTALITTILYKSKQ
SIALAATASTRGQSTRGVSTRGSTRGQSTRGQRSTRGNAPEEERLVWFPNATPPDHLTGE
FPGDRGFDPLGLSTDPVVFERMRISEVFHGRLAMLGIVGCIVPELFFSQGPWFDLSSFNL
```

```
TRLGLVALQVIAPLEYWRGNRGFSWNGDDGPDQSYPGFDPLRLTDEETKLKEIKNGRLAM
TAMLGLEVQSHITGKSPLVNLSEHLSNPASVNILTGGGS
>LHCE1-5_HBDM01058159.1+1/F-2/O2/P1@332
MYGAVQGPSSSKTTSVVALCSCAALCAVAGFVAAQPGYLYAPTAVRPAVQSSVATTPLTT
VVRPSFATAAQRESASWSTERSGLPQIDSTVPFALGVGALTALITTILYKSKQSIALAAT
ASTRGQSTRGVSTRGSTRGQSTRGQRSTRGNAPEEERLVWFPNATPPDHLTGEFPGDRGF
DPLGLSKDPVVFERMRISEVFHGRLAMLGIVGCIVPELFFSQGPWFDLSGFNLTRLGLVA
LQVIAPLEYWRGNRGFSWNGDDGPDQSYPGFDPLRLTDEETKLKEIKNGRLAMTAMLGLE
VQSHITGKSPLVNLSEHLSNPASVNILTGGGS
>LHCE1-6_Koziol/LHCA1-1@347
FFFVDSMYGAVQGPSSSKTTSVVALCSCAALCAVAGFVAAQPGYLYAPTAVRPAVQSSAS
TTPLTTVVRPSFATAAQRESASWSTERSGLPQIDSTVPLALGVGALTALITTILYKSKSI
ALAATASTRGQSTRGVSTRGSTRGQSTRGQRSTRGSAPEEERLVWFPNATPPDHLTGEFP
GDRGFDPLGLSKDPVVFERMRISEVFHGRLAMLGIVGCIVPELFFSQGPWFDLSSFNLTR
LGLVALQVIAPLEYWRGNRGFSWNGDDGPDQSYPGFDPLRLTDEETKLKEIKNGRLAMTA
MLGLEVQSHITGKSPLVNLSEHLSNPASVNILTGGGSVAMFAGSGRK
>LHCE1-7_Koziol/LHCA2-1@344
NNMYGAVQGPSSSKTTSVVALCSCAALCAVAGFVAAQPGYLYAPTAVRPAVQSSVATTPL
TTVVRPSFATAAQRESASWSTERSGLPQIDSTVPFALGVGALTALITTILYKSKQSIALA
ATASTRGQSTRGVSTRGSTRGQSTRGQRSTRGNAPEEERLVWFPNATPPDHLTGEFPGDR
GFDPLGLSTDPVVFERMRISEVFHGRLAMLGIVGCIVPELFFSQGPWFDLSGFNLTRLGL
VALQVIAPLEYWRGNRGFSWNGDDGPDQSYPGFDPLRLTDDETKLKEIKNGRLAMTAMLG
LEVQSHITGKSPLVNLSEHLSNPASVNILTGGGSVAMFAASGRK
>LHCE1-8_Koziol/LHCA4-1@85
WDGDDGPDQSYPGFDPLRLTDEETKLKEIKNGRLAMTAMLGLEVQSHLTGKSPLVNLSEH
LSNPASVNILTGGGSVAMFYASGRK
>LHCE1-9_GEFR01060700.1/F+3/O1/P1@57
GFDPLGLSTDPVVFERMRISEVFHGRLAMLGIVGCTQALHFFPHLLTSQGPRHISVR
>LHCE2-1_Koziol/LHCA1-2@191
SQKSDERPLWFPNTLAPSWLTGEYYGDRGFDPAGLAADPDTFERMRIAEVYHGRLAMLAI
VGALVPDYLGKGIWYEAAQNAGIGLKEVAIFTAAYGVFEVARGVKENSDPTTIYPGFDPL
NLTTDYTKEAEIKNGRLALTALLGFEVQRHVVGGSPLANLADHLQQPLQHNIADSIMHQW
PVAMFASSGHK
>LHCE2-2_Koziol/LHCA2-2@191
SQKSDERPLWFPNTLAPSWLNGEYYGDRGFDPAGLAADPDTFERMRIAEVYHGRLAMLAV
VGALVPDYLGKGIWYEAAQNAGIGLKEVAIFTAAYGVFEVARGVKENSDPTTIYPGFDPL
NLTTDYTKEAEIKNGRLALTALLGFEVQRHVVGGSPLANLAEHLQQPLQRNIADSIMHQW
PVAMFASSGHK
>LHCE2-3_GEFR01007608.1/F-2/O1/P2@151
TQSEDRPLWFPNTLAPSWLTGEYFGDRGFDPAGLAADPETFERMRIAEVYHGRLAMLAVV
GALVPDYLGKGIWYEAAQNAGIGVKEVAIFTAAYGVFEVARGVKENSDPTTVYPGFDPLN
LTTDYTKEAEIKNGRLALTALLGFIHFIPTR
>LHCE2-4_Koziol/LHCA4-2@190
TQSEDRPLWFPNTLAPSWLTGEYFGDRGFDPAGLAADPETFERMRIAEVYHGRLAMLAVV
GALVPDYLGKGIWYEAAQNAGIGVKEVAIFTAAYGVFEVARGVKENSDPTTVYPGFDPLN
LTTDYTKEAEIKNGRLALTALLGFEVQRHVVGGSPLANLADHLQQPLQRNIADSIMHQWP
VAMFASSGHK
>LHCE2-5_GEFR01006499.1/F+2/O1/P2@181
SQKSDERPLWFPNTLAPSWLNGEYYGDRGFDPAGLAADPDTFERMRIAEVYHGRLAMLAI
VGALVPDYLGKGIWYEAAQNAGIGLKEVAIFTAAYGVFEVARGVKENSDPTTIYPGFDPL
NLTTDYTKEAEIKNGRLALTALLGFEVQRHVVGGSPLANLADHLQQPLQHNIADSIMHQW
P
>LHCE2-6_GDJR01067079.1/F-2/O1/P3@181
SQKSDERPLWFPNTLAPSWLTGEYYGDRGFDPAGLAADPDTFERMRIAEVYHGRLAMLAV
VGALVPDYLGKGIWYEAAQNAGIGLKEVAIFTAAYGVFEVARGVKENSDPTTIYPGFDPL
NLTTDYTKEAEIKNGRLALTALLGFEVQRHVVGGSPLANLAEHLQQPLQRNIADSIMHQW
P
>LHCE2-7_HBDM01056777.1/F-1/O1/P1@181
SQKSDERPLWFPNTLAPSWLTGEYYGDRGFDPAGLAADPDTFERMRIAEVYHGRLAMLAV
VGALVPDYLGKGIWYEAAQNAGIGVKEVAIFTAAYGVFEVARGVKENSDPTTVYPGFDPL
```

```
NLTTDYTKEAEIKNGRLALTALLGFEVQRHVVGGSPLANLADHLQQPLQRNIADSIMHQW
P
>LHCE2-8_HBDM01071460.1/F-3/O1/P2@181
SQKSDERPLWFPNTLAPSWLTGEYYGDRGFDPAGLAADPDTFERMRIAEVYHGRLAMLAI
VGALVPDYLGKGIWYEAAQNAGIGLKEVAIFTAAYGVFEVARGVKENSDPTTIYPGFDPL
NLTTDYTKEAEIKNGRLALTALLGFEVQRHVVGGSPLANLAEHLQQPLQRNIADSIMHQW
P
>LHCE2-9_GEFR01007608.1/F-3/O8/P1@57
MMEWWKINSCHTVSAASWGWNDCEVQRHVVGGSPLANLADHLQQPLQHNIADSIMHQ
>LHCE3-1_GEFR01006499.1/F+2/O1/P3@174
DGLWFPNAEPPAHLTGEYPADRGFDPLNLAADPDVYARMRVAEVFHGRLAMLCMVGCIVP
ELLGKGAWFEAGDSVDGLKLGFITMAIAAPTEYWRGQGGFNWQKGELDRSYPGFDPLNLT
TDYTRAAEVKNGRLALTAVAGLTAQYLATGESPLANLAAHLANPLGANITTNLA
>LHCE3-2_HBDM01046768.1/F-1/O1/P1@173
KDGLWFPDAEPPAHLTGEYPADRGFDPLNLAADPDVYARMRVAEVFHGRLAMLCMVGCIV
PELLGKGAWFEAGDSVDGLKLGFIAMAIAAPTEYWRGHGGFNWQKGDLDRSYPGFDPLNL
TTDYTKAAEVKNGRLALTAVAGLTAQYLATGESPLANLSAHLANPIGANIATN
>LHCE3-3_GDJR01074337.1/F+1/O1/P2@172
DGLWFPNAEPPAHLTGEYPADRGFDPLNLAADPDVYARMRVAEVFHGRLAMLCMVGCIVP
ELLGKGAWFEAGDSVDGLKLGFIAMAIAAPTEYWRGQGGFNWQKGDLDRSYPGFDPLNLT
TDYTRAAEVKNGRLALTAVAGLTAQYLATGESPLANLSAHLANPIGANIATN
>LHCE3-4_HBDM01058159.1+1/F-2/O2/P3@241
DGLWFPNAEPPAHLTGEYPADRGFDPLNLAADPDVYARMRVAEVFHGRLAMLCMVGCIVP
ELLGKGAWFEAGGQPSNENIGSVRVRDTEGLNMILPSSGFWQREEAPVCCSVCKKSSGKR
DVFLECAALTLQERSPAGWHFVFLLVPLFAGCGFDSVDGLKLGFITMAIAAPTEYWRGHG
GFNWQKGDLDRSYPGFDPLNLTTDYTRAAEVKNGRLALTAVAGLTAQYLATGESPLANLA
A
>LHCE3-5_Koziol/LHCA1-3@182
DGLWFPNAEPPAHLTGEYPADRGFDPLNLAADPDVYARMRVAEVFHGRLAMLCMVGCIVP
ELLGKGAWFEAGDSVDGLKLGFITMAIAAPTEYWRGQGGFNWQKGDLDRSYPGFDPLNLT
TDYTRAAEVKNGRLALTAVAGLTAQYLATGESPLANLSAHLANPIGANIATNLAMFASSG
AK
>LHCE3-6_Koziol/LHCA2-3@182
DGLWFPNAEPPAHLTGEYPADRGFDPLNLAADPDVYARMRVAEVFHGRLAMLCMVGCIVP
ELLGKGAWFEAGDSVDGLKLGFITMAIAAPTEYWRGQGGFNWQKGELDRSYPGFDPLNLT
TDYTRAAEVKNGRLALTAVAGLTAQYLATGESPLANLSAHLANPIGANITTNLAMFASSG
AK
>LHCE3-7_Koziol/LHCA1-5@174
DGLWFPNAEPPAHLTGEYPADRGFDPLNLAADPDVYARMRVAEVFHGRLAMLCMVGCIVP
ELLGKGAWFEAGDSVDGLKLGFITMAIAAPTEYWRGQGGFNWQKGDLDRSYPGFDPLNLT
TDYTRAAEVKNGRLALTAVAGLTAQYLATGESPLANLTAHLANPLGANITTNLA
>LHCE3-8_Koziol/LHCA4-3@182
DGLWFPDAEPPAHLTGEYPADRGFDPLNLAADPDVYARMRVAEVFHGRLAMLCMVGCIVP
ELLGKGAWFEAGDSVDGLKLGFIAMAIAAPTEYWRGHGGFNWQKGDLDRSYPGFDPLNLT
TDYTKAAEVKNGRLALTAVAGLTAQYLATGESPLANLSAHLANPIGANITTNLAMFASSG
VK
>LHCE3-9_Koziol/LHCA4-5@174
DGLWFPDAEPPAHLTGEYPADRGFDPLNLAADPDVYARMRVAEVFHGRLAMLCMVGCIVP
ELLGKGAWFEAGDSVDGLKLGFIAMAIAAPTEYWRGHGGFNWQKGDLDRSYPGFDPLNLT
TDYTKAAEVKNGRLALTAVAGLTAQYLATGESPLANLAAHLANPVGANIATNLV
>LHCE3-10_GDJR01074336.1/F+1/O1/P2@174
DGLWFPNAEPPAHLTGEYPADRGFDPLNLAADPDVYARMRVAEVFHGRLAMLCMVGCIVP
ELLGKGAWFEAGDSVDGLKLGFIAMAIAAPTEYWRGQGGFNWQKGDLDRSYPGFDPLNLT
TDYTRAAEVKNGRLALTAVAGLTAQYLATGESPLANLAAHLANPLGANITTNLA
>LHCE3-11_GDJR01074338.1/F+1/O1/P2@174
DGLWFPNAEPPAHLTGEYPADRGFDPLNLAADPDVYARMRVAEVFHGRLAMLCMVGCIVP
ELLGKGAWFEAGDSVDGLKLGFIAMAIAAPTEYWRGQGGFNWQKGDLDRSYPGFDPLNLT
TDYTRAAEVKNGRLALTAVAGLTAQYLATGESPLANLAAHLANPVGANIATNLV
>LHCE3-12_GEFR01017658.1/F-1/O2/P1@178
MAIAAPTEYWRGQGGFNWQKGELDRSYPGFDPLNLTTDAGLHRDGHRRPHRVLARPWRLQ
```

```
LAEGGSGPVVPWLRPPEPHHGLHQGRGGQERPTGLDGGCGPDGAVPGHGGEPAGQPGCPP
RQPCGRQHCHQLGLRNRRLFYSVLPGDSVTRVPSGSLCSITEPFPSHLLLYRHSSQLR
>LHCE4-1_Koziol/LHCA1-4@192
EERELWFPNIVPPRYLTGEAYGDKGFDPAGLAADPVTFERMQVAEVFHCRLSMLALVGCL
VPELLGNGAWFQIWDKVDFNRFAVVALQVVAPLEYWRGNGGFLWNDEETIDQSYPGFDPL
NLTTEYTKEAEIKNGRLAMIGMFGLEVQSHVTGKSPITNLIDHLSHPLTANIGTNLSQPW
PPVAMFASSGHK
>LHCE4-2_Koziol/LHCA2-4@192
EERELWFPNIVPPRYLTGEAYGDKGFDPAGLAADPVTFERMQVAEVFHCRLSMLALVGCL
VPELLGNGAWFQIWDKVDFNRFAVVALQVVAPLEYWRGNGGFLWNDEETVDQSYPGFDPL
NLTTEYTKEAEIKNGRLAMIGMFGLEVQSHVTGKSPITNLIDHLSHPLTANIGTNLSQPW
PQVAMFASSGHK
>LHCE4-3_GEFR01018054.1/F+1/O1/P2@182
EERELWFPNIVPPRYLTGEAYGDKGFDPAGLAADPVTFERMQVAEVFHCRLSMLALVGCL
VPELLGNGAWFQIWDKVDFNRFAVVALQVVAPLEYWRGNGGFLWNDEETIDQSYPGFDPL
NLTTEYTKEAEIKNGRLAMIGMFGLEVQSHVTGKSPITNLIDHLSHPLTANIGTNLSQPW
PQ
>LHCE4-4_Koziol/LHCA4-4@192
EERELWFPNIVPPRYLTGEAFGDKGFDPAGLAADPVTFERMQVAEVFHCRLSMLALVGCL
VPELLGNGAWFQIWDKVDFNRFAVVALQVVAPLEYWRGNGGFLWNDEETVDQSYPGFDPL
NLTTEYTKEAEIKNGRLAMIGMFGLEVQSHVTGKSPITNLIDHLSSPLTANIGTNLSQPW
PPVAMFASSGHK
>LHCE5-1_GEFR01002507.1/F-2/O1/P4@172
DGVWFPGATPPAHLTGEYPADRGFDPLSLAADPTVYARMRVSEVFHARLCMLAIVGAIVP
EVLGKGAWFEAGNSVDGIKLGFIAMAIAAPTEYWRGNGGFNWDKGAADRSYPGFDPLKLT
TDYTKAAEIKNGRLAMAGLLGLTFQYLATGESPLANLAAHLANPVGTNITTT
>LHCE5-2_HBDM01068688.1+1/F-3/O1/P4@172
DGVWFPGATPPAHLTGEYPADRGFDPLSLAADPTVYARMRVSEVFHARLSMLAIVGSIVP
EVLGKGAWFEVGNSVDGIKLGFILMAIAAPTEYWRGNGGFNWDKGTADRSYPGFDPLKLT
TDYTKAAEIKNGRLAMTGLLGLTFQYLATGESPLANLAAHLANPVGANITTT
>LHCE5-3_Koziol/LHCA6-3@182
DGVWFPGATPPAHLTGEYPADRGFDPLSLAADPTVYARMRVSEVFHARLSMLAIVGAIVP
EVLGKGAWFEAGNSVDGIKLGFIAMAIAAPTEYWRGNGGFNWDKGAADRSYPGFDPLKLT
TDYTKAAEIKNGRLAMAGLLGLTFQYLATGESPLANLAAHLANPVGANITTTLAMYSTSG
EK
>LHCE5-4_Koziol/LHCA8-3@182
DGVWFPGATPPAHLTGEYPADRGFDPLSLAADPTVYARMRVSEVFHARLSMLAIVGAIVP
EVLGKGAWFEAGNSVDGIKLGFIAMAIAAPTEYWRGNGGFNWDKGAADRSYPGFDPLKLT
TDYTKAAEIKNGRLAMAGLLGLTFQYLATGESPLANLAAHLANPVGTNITTTLAMYSTSG
DK
>LHCE5-5_Koziol/LHCA6-5@174
DGVWFPGATPPAHLTGEYPADRGFDPLSLAADPTVYARMRVSEVFHARLCMLAIVGAIVP
EVLGKGAWFEAGNSVDGIKLGFIAMAIAAPTEYWRGNGGFNWDKGAADRSYPGFDPLKLT
TDYTKAAEIKNGRLAMAGLLGLTFQYLATGESPLANLAAHLANPVGANITTTLA
>LHCE5-6_Koziol/LHCA5-6@174
DGVWFPGAQPPAHLTGEYPADRGFDPLSLAADPTVYARMRVSEVFHARLSMLAIVGSILP
ELLGKGAWFEVGNSVDGIKLGFILMAIAAPTEYWRGNGGFNWDKGTADRSYPGFDPLKLT
TDYTKAAEIKNGRLAMTGLLGLTFQYLATGESPLANLAAHLANPVGANITTTLA
>LHCE5-7_Koziol/LHCA5-3@182
DGVWFPGAQPPAHLTGEYPADRGFDPLSLAADPTVYARMRVSEVFHARLSMLAIVGSIVP
ELLGKGAWFEVGNSVDGIKLGFILMAIAAPTEYWRGNGGFNWDKGTADRSYPGFDPLKLT
TDYTKAAEIKNGRLAMTGLLGLTFQYLATGESPLANLAAHLANPVGANITTTLAMYSTSG
EK
>LHCE6-1_GEFR01002507.1/F-2/O1/P5@224
YRGLWFPNITPPAYLTGEFPADRGFDPAGLAADPKVYERMRIAEVFNGRLAMLAIVGCVY
PELLGNGVWFEVWNKVDFYRFALISLQVVAPLEYWRGNGGFGWDGEEKYDRSYPGFDPCN
LTTEYTKAAEIKNGRLAMIGMFGLEVQSHVTAQGPVANLIEHLRPCSWSVRLESYTPVCW
HAPGFFQNFQRMGQRAPFQWGVTEPPSFIHTSATVNEGWWGSES
>LHCE6-2_GEFR01004399.1/F+3/O2/P4@181
YRGLWFPNITPPAYLTGEFPADRGFDPAGLAADPKVYERMRVAEVFNGRLAMLAIVGCVY
```

```
PELLGNGVWFEVWNKVDFYRFALISLQVVAPLEYWRGNGGFGWDGEEKYDRSYPGFDPCN
LTTEYTKAAEIKNGRLAMIGMFGLEVQNHVTAQGPVANLIEHLRHPLAANIGANLAHPWP
P
>LHCE6-3_Koziol/LHCA8-4@189
YRGLWFPNITPPAYLTGEFPADRGFDPAGLAADPKVYERMRIAEVFNGRLAMLAIVGCVY
PELLGNGVWFEVWNKVDFYRFALISLQVVAPLEYWRGNGGFGWDGEEKYDRSYPGFDPCN
LTTEYTKAAEIKNGRLAMIGMFGLEVQSHVTAQGPVANLMEHLRHPLAANIGANHDQWPV
AMFATTGHK
>LHCE6-4_Koziol/LHCA6-4@191
YRGLWFPNITPPAYLTGEFPADRGFDPAGLAADPKVYERMRIAEVFNGRLAMLAIVGCVY
PELLGNGVWFEVWNKVDFYRFALISLQVVAPLEYWRGNGGFGWDGEEKYDRSYPGFDPCN
LTTEYTKAAEIKNGRLAMIGMFGLEVQSHVTAQGPVANLIEHLRHPLAANIGANLAHPWP
PVAMFATTGHK
>LHCE6-5_Koziol/LHCA5-4@191
YRGLWFPNITPPAYLTGEFPADRGFDPAGLAADPKVYERMRVAEVFNGRLAMLAIVGCVY
PELLGNGVWFEVWNKVDFYRFALISLQVVAPLEYWRGNGGFGWDGEEKYDRSYPGFDPCN
LTTEYTKAAEIKNGRLAMIGMFGLEVQNHVTAQGPVANLIEHLRHPLAANIGASLAHPWP
PVAMFATTGHK
>LHCE7-1_HBDM01007325.1/F-3/O1/P2@175
LLWVPNATPPAHLTGEFPGDRGFDPLGLAKDPKVYQRMRISEVFHGRLAMLGIVGCVAPE
LFFSKGAWFDYSDYDLNRLGLIALQVIAPLEYWRGNGGFSWDGNDGPDRSYPGFDPLGLT
NEETKLQEIKNGRLAMTAMLGLEVQSHITGKSPLTNLSEHLSHPLSANLLTGGGS
>LHCE7-2_GEFR01002507.1/F-2/O1/P2@175
LLWIPNATPPAHLTGEFPGDRGFDPLGLAKDPKVYQRMRISEVFHGRLAMLGIVGCVAPE
LFFSKGAWFDYSDYDLNRLGLIALQVIAPLEYWRGNGGFSWDGNDGPDRSYPGFDPLGLT
NEETKLQEIKNGRLAMTAMLGLEVQSHITGKSPLTNLSEHLSHPLSANLLTGGGS
>LHCE7-3_Koziol/LHCA5-1@184
LLWIPNATPPAHLTGEFPGDRGFDPLGLSKDPKVFARMRISEVFHGRLAMLGIVGCVGQE
WLFNKGAWFDYSDFDLPRLGLIALQVIAPLEYWRGNGGFSWNGNDGPDRSYPGFDPLGLT
TEDTKLREIKNGRLAMSAMLGLEVQSHITGKSPLTNLGDHLSSPFTANILTGGGSVAMFS
TSGK
>LHCE7-4_Koziol/LHCA7-1@160
LGLAKDPKVYQRMRISEVFHGRLAMLGIVGCVAPELFFSKGAWFDYSDYDLNRLGLIALQ
VIAPLEYWRGNGGFSWDGNDGPDRSYPGFDPLGLTNEETKLQEIKNGRLAMTAMLGLEVQ
SHITGKSPLTNLSEHLSHPLSANLLTGGGSLAMFSTAGQK
>LHCE7-5_GDJR01067079.1/F-2/O1/P2@175
LLWIPNATPPAHLTGEFPGDRGFDPLGLAKDPKVYQRMRISEVFHGRLAMLGIVGCVAPE
LFFSKGAWFDYSDYDLNRLGLIALQVIAPLEYWRGNGGFSWDGNDGPDRSYPGFDPLGLT
NEETKLQEIKNGRLAMTAMLGLEVQSHITGKSPLVNLSEHLSNPASVNILTGGGS
>LHCE7-6_GEFR01004399.1/F+3/O1/P1@41
RMRISEVFHGRLAMLGIVGCVGQEWLFNKGAWFDYSEQGSL
>LHCE8-1_GDJR01059671.1/F+2/O1/P2@178
TDRRPLWFPGGYAPSWLTGQHFGDRGFDPAGLAADPKVFERMRVSEVYHGRLAMLAIVGA
VVPDIQGKGAWYEAAQSAGIGVNEVAVFTAAYGIAEVARGLKANSDPTSRYPGFDPLNLT
TDYTKEAEIKNGRLALTAMLGLEVQRHVTGVSPLVNLVEHVKHPLGHNIAESVMHQWP
>LHCE8-2_Koziol/LHCA5-2@187
VDRPLWFPGGYAPSYLTGQYFGDRGFDPAGLAADPKVFERMRVSEVYHGRLAMLAIVGAV
VPDILGKGAWYEAAQNAGIGVNEVAVFTAAYGVVEVARGLKANSDPTKNYPGFDPLKLTT
DYTKEAEIKNGRLALTGMLGLEVQRHVTGVSPLVNLVEHVKHPLNHNIAESVMHQWPVAM
FAATGHK
>LHCE8-3_Koziol/LHCA7-2@182
TDRRPLWFPGGYAPSWLTGQHFGDRGFDPAGLAADPKVFERMRVSEVYHGRLAMLAIVGA
VVPDIQGKGAWYEAAQSAGIGVNEVAVFTAAYGIAEVARGLKANSDPTSRYSGFDSLNLT
TDYTKEAEIKNGRLALTAMLGLQVQRHVTGVSPLLNLVEHVKHPLVHTIAESVMHQWPVA
MF
>LHCE8-4_Koziol/LHCA8-2@66
HTKEAEIKNGRLALTAMLGLEVQRHVTGVSPLVNLVEHVKHPLGHNIAESVMHQWPVAMF
AATGHK
>LHCE8-5_GEFR01044592.1/F+1/O1/P1@95
AMLAIVGAVVPDIQGKGAWYEAAQSAGIGVNEVAVFTAAYGIAEVARGLKANSDPTSRYP
```

```
GFDPLNLTTDYTKEAEIKNGRLALTAMLGLGHTFL
>LHCE9-1_Koziol/LHCA9-1@159
LGKDPNALERFRVSEVFHGRLAMLAFAGALVPELLRGVAWYDAVHTPDPITLTTFLLVLA
GATGPIELWRYDGRFPAWEEGEEPEPKVADIAYPGADPFEWSSEEAKLKELKNGRLAMIA
VLGFYFQHAKTGASPIENLVAFINDGANITTAMFATAGK
>LHCE9-2_Koziol/LHCA9-2@201
KVRPTPAPAAASTERLLWIPGVKAPEHLNNTLVGDRGFDPLNLGKDPNALERFRVSEVFH
GRLAMLAFAGALVPELLRGVAWYDAVKTPDPITLTTFLLALAVATGPIELWRYDGRFPAW
EEGEEPEPKVADIAYPGADPFEWSSEEAKLKELKNGRLAMIAVLGFYFQHAKTGASPIEN
LVAFINDGANITTAMFATAGK
>LHCE9-3_Koziol/LHCA9-3@196
KVRPTPAPAAASTERLLWIPGVKAPEHLNNTLVGDRGFDPLNLGKNPNALERFRVSEVFH
GRLAMLAFAGALVPELLRGVAWYDAVKTPDPITLTTFLLALAVATGPIELWRYDGRFPAW
EEGEEPEPKVANIAYPGADPFEWSSEEAKLKELKNGRLAMIAVLGFYFQHAKTGASPIEN
LVAFINDGANITTAMF
>LHCE9-4_HBDM01066990.1/F+1/O1/P1@58
DPFEWSSEEAKLKELKNGRLAMIAVLGFYFQHVKTGASPIENLVAFINDGANITTAMF
>LHCE10-1_GEFR01014903.1+3/F+3/O1/P2@174
RPLWFPGAKPPAYLTGEFPADRGFDPLGLAKDPATYERMRSSEVFHSRLAMLGVVGSLIP
ELQGQGAWYTLSEKQFGPGPNGEVIGFTELALIGMLFSYPLEYWRGNGGFGWFEEKGDRT
YPGFDPLQLTSEYTKTAEIKNGRLAMSALLGFAVQHATTGASPLENWAATTGGS
>LHCE10-2_Koziol/LHC10-2@184
RPVWFPGATPPAYLTGEFPADRGFDPLGLAKDPATYERMRSSEVFHSRLAMLGVVGSLIP
ELQGQGAWYTLSEKQFGPGPNGEVIGFTELALIGMLFSYPLEYWRGNGGFGWFEEKGDRT
YPGFDPLQLTSEYTKTAEIKNGRLAMSALLGFAVQHATTGASPLENWAATTGGSVAMFAA
AGER
>LHCE10-3_GEFR01025204.1/F+2/O1/P2@93
RPVWFPGATPPAYLTGEFPADRGFDPLGLAKDPATYERMRSSEVFHSRLAMLGVVGSLIP
ELQGQGAWYTLSEKQFGPGPNGRRTPTALDSFQ
>LHCE11-1_Koziol/LHC10-3@178
ATWFPGAKAPPHLTGEFPGDRGFDPVDFAANPESFERMRASEVFHGRMAMLGVAGCLIPE
LLGRGVWFQAGESVDAAQLGLYFMFLAAPSEYWRGNGGFGWDQKEKKAGGNRIYPGFDPL
KMTSDETKTKEIKNGRLAMIAFLGLVSQANTTGVSPFQNLAAVFGASPVAMFTTTGEK
>LHCE11-2_GDJR01048454.1/F-2/O1/P2@171
ATWFPGAKAPPHLTGEFPGDRGFDPVDFAANPESFERMRASEVFHGRMAMLGVAGCLIPE
LLGRGVWFQAGESVDAAQLGLYFMFLAAPSEYWRGNGGFGWDQKEKKAGNRIYPGFDPLK
MTSDETKTKEIKNGRLAMIAFLGLVSQANTTGVSPFQNLAAVFGASPVAML
>LHCE11-3_Koziol/LHCA11-a@70
HGNRIYPGFDPLKMTSDETKTKEIKNGRLAMIAFLGLVSQANTTGVSPFQNLAAVFGASP
VAMFTTTGEK
>LHCE12-1_HBDM01058828.1/F+3/O1/P1@372
DLFFFGEIFFGNMYGAVETPSSTKSTGFVAICSCAALFSVVGFIAAQPASLYAPTALRPP
VLRSVMVTPSSSSTIAPPIFATAAQREAQDWNIQRSGLAQVGATMMPIVLGVGALTTLIT
TLLFKSKNSIALAATAGQSTRGQSTRGQSTRGVPGSTRGQVSGSTRGQVSGSTRGQVSGS
TRGQRSASTRGSASVEERLLWFPNVTPPAYLTGEFPGDRGFDPAGLSADPETFDRMRIAE
IIHGRVAMLAVVGALVPDFLGKGNWFEAPAIAGIDIGDVQQFILGYGVFEFGRGLNKSSN
PAAIYPGFDPLNLTSDYTKEAEIKNGRLALTAMLGFEVQRHVVGGSPLANLSDHLQQPLK
FNIADSVMHPWP
>LHCE12-2_Koziol/LHCA3@398
FFFGEIFFGNMYGAVETPSSTKSTGFVAICSCAALFSVVGFIAAQPASLYAPTALRPPVL
RSVMVTPSSSSTIAPPIFATAAQREAQDWNIQRSGLAQVGATMMPIVLGVGALTTLITTL
LFKSKNSIALAATAGQSTRGQSTRGQSTRGVPGSTRGQVSGSTRGQVSGSTRGQRSASTR
GSASVEERLLWFPNVTPPAYLTGEFPGDRGFDPAGLSADPETFDRMRIAEIIHGRVAMLA
VVGALVPDFLGKGNWFEAPAIAGIDIGDVQQFILGYGVFEFGRGLNKSSNPAAIYPGFDP
LNLTSDYTKEAEIKNGRLALTAMLGFEVQRHVVGGSPLANLSDHLQQPLKFNIADSVMHP
WPVAMFASTGSKDGLWFPNSKPPAHLISEYSTTATLVP
>LHCE13-1_GDJR01039004.1/F+2/O1/P1@198
PGAEAPAWLTGEYPGDRGFDPLGLAKDPEDFAKNRDSEVYHGRWAMLGLVGCLVPEVFGN
LGIAQLPAWYEAGAVANTGSLDYLGNPNLVHASNVPLIFFTTLLLFLPIEAWRWNGQIAP
EAKSEERTTYPGGPFDPLKLGASPELKLKEIKNGRLAMVGMFGFWAQSYVTGEGPLANLA
```

```
AHLADPAHNNLLNTVAVF
>LHCE13-2_GEFR01040018.1/F+3/O1/P1@150
RPTWFPSAEAPAWLTGEYPGDRGFDPFGLAKDPADFAKFRDSEVFHGRWAMLGLVGCLVP
EVFGNLGIAQLPAWYDAATVANTSNLDYLGNPNLVHASNVSFIALSTLLLMGPVEAWRWN
GALASEAKSAERTTYPGGPFDPLKLGASPE
>LHCE13-3_GDJR01039000.1+5/F+2/O1/P3@186
RPTWFPGAEAPAWLTGEYPGDRGFDPLGLAKDPEDFAKNRDSEVYHGRWAMLGLVGCLVP
EVFGNLGIAQLPAWYEAGAVANTGSLDYLGNPNLVHASNVPLIFFTTLLLFLPIEAWRWN
GQIAPEAKSKEWTTYPGGPFDPLKLGASPELKLKEIKNGRLAMVGMFGFWAQSYVTGEGP
LANLAA
>LHCE13-4_GDJR01039000.1+5/F+2/O1/P2@199
RPTWFPGAEAPAWLTGEFPGDRGFDPCGLARDPVDFAKFRDSEVFHGRWAMLGLVGCLVP
EVFANLGIAQLPAWYDAATVANTSNLDYLGNPNLIHASNVSFIALSTLLLMGPVEAWRWN
GALASEAKSAERTTYPGGPFDPLKLGASPELKLKEIKNGRLAMVGMFGFWAQSYVTGEGP
LANLAAHLADPAHNNLLNT
>LHCE13-5_Koziol/LHCBM10-1@150
PEVFGNLGIAQLPAWYEAGAVANTGSLDYLGNPNLVHASNVPLIFFTTLLLFLPIEAWRW
NGQIAPDAKSKEWTTYPGGPFDPLKLGASPELKLKEIKNGRLAMVGMFGFWAQSYVTGEG
PLANLAAHLADPAHNNLLNTVAMFAASGEK
>LHCE13-6_Koziol/LHCBM10-2@209
RPTWFPGAEAPTWLTGEYPGDRGFDPFGLAKDPADFAKFRDSEVFHGRWAMLGLVGCLVP
EVFGNLGIAQLPAWYDAATVANTSNLDYLGNPNLVHASNVSFIALSTLLLMGPVEAWRWN
GALASEAKSAERTTYPGGPFDPLKLGASPELKLKEIKNGRLAMVGMFGFWAQSYVTGEGP
LANLAAHLADPAHNNLLNTVAMFAASGEK
>LHCE13-7_Koziol/LHCBM10-3@203
RPTWFPGAEAPAWLTGEYPGDRGFDPLGLAKDPEDFAKNRDSEVYHGRWAMLGLVGCLVP
EVFGNLGIAQLPAWYEAGAVANTGSLDYLGNPNLVHASNVPLIFFTTLLLFLPIEAWRWN
GQIAPEAKSAERTTYPGGPFDPLKLGASPELKLKEIKNGRLAMVGMFGFWAQSYVTGEGP
LANLAAHLADPAHNNLLNTVAVF
>LHCE13-8_HBDM01032749.1/F-2/O1/P1@208
AMFAASGGKRPTWFPGAEAPAWLTGEYPGDRGFDPLGLAKDPEDFAKNRDSEVYHGRWAM
LGLVGCLVPEVFGNLGIAQLPAWYEAGAVANTGSLDYLGNPNLVHASNVPLIFFTTLLLF
LPIEAWRWNGQIAPEAKSKEWTTYPGGLFDPLKLGASPELKLKEIKNGRLAMVGMFGFWA
QSYVTGEGPLANLAAHLADPAHNNLLNT
>LHCE13-9_HBDM01032749.1/F-2/O1/P2@148
RPTWFPGAEAPAWLTGEYPGDRGFDPFGLAKDPADFAKFRDSEVFHGRWAMLGLVGCLVP
EVFGNLGIAQLPAWYDAATVANTSNLDYLGNPNLVHASNVSLIFFTTLLLFLPIEAWRWN
GQIAPEAKSAERTTYPGGPFDPLKLGAS
>LHCE13-10_HBDM01070328.1/F-3/O1/P2@196
RPTWFPGAEAPAWLTGEFPGDRGFDPCGLARDPVDFAKFRDSEVFHGRWAMLGLVGCLVP
EVFGNLGIAQLPAWYEAGAVANTGSLDYLGNPNLVHASNVPLIFFTTLLLFLPIEAWRWN
GQIAPEAKSKEWTTYPGGLFDPLKLGASPELKLKEIKNGRLAMVGMFGFWAQSYVTGEGP
LANLAAHLADPAHNNL
>LHCE13-11_GDJR01039003.1+1/F+2/O1/P1@198
PGAEAPAWLTGEYPGDRGFDPLGLAKDPEDFAKNRDSEVYHGRWAMLGLVGCLVPEVFGN
LGIAQLPAWYEAGAVANTGSLDYLGNPNLVHASNVPLIFFTTLLLFLPIEAWRWNGQIAP
EAKSEERTTYPGGPFDPLKLGASPELKLKEIKNGRLAMVGMFGFWAQSYVTGEGPLANLA
AHLADPAHNNLLNTVAMF
>LHCSRL-1_GDJR01029061.1+2/F+2/O1/P1@327
SIFFRRLAPLPAMSGSMSTLAFATAVAAAVFCLSSVVVAGSPGVGALYATPTVRPAPTVA
QRPTLTAAPPQPMTVLGAITTEAALRSGEAPASPAGGTPNTFTLALASAITGATAALTYT
LGRLVRGTVRNRSVEAQPLVMLAMGAGDDDDNDLFGLLGLAETAGLPGAIQPLGYFDPLK
LSQGKDFNELKKVRESELKHGRVAMLAVVGLPLAEQFHPLFGGAVDGPGVQAVLRPELAV
FWLLLLPLFYFLEQPTLAAYDTTGKLPGLFKEAHKPGDYGFDPLGLKPATDTEYEAMQCK
ELSNGRLAMIAILGMLAQEVVTGRPLF
>LHCSRL-2_GDJR01081559.1+2/F+3/O2/P1@338
MHGRLFVFRGALWVPWLLCTAALVLFCSIDTPAMLTAYGRITRSGPGVSAGMREKPFGKP
LLTRGGSGLQRLDGAFGALWGEQSGMKRFFISETITSGASILFGPLIAVVATLLAFLYDE
RRKPKGAGALRPLDAIGGSPLLPLTVPLSGLRTPPQTPTLDSRASTPSPWSVTKRPCWLG
RYAESSIPDYLRGQLPGLLGFDPAGVKAEGVVAPAKPLHAMEVARARVALLLAAGQVAPH
```

```
GAGVPAAVACLPLLQSAAEGPNLLQDPSPRSPTGIVSLRAPGHRVRLLAALAVLAAHLHR
LAAREGGGAAGRAAWAADLARVDPALLEALLGLPANRR
>LHCSRL-3_HBDM01049636.1/F+2/O2/P1@338
MHGRLFVFRGALWVPWLLCTAALVLFCSIDTPAMLTAYGRITRSGPGVSAGMREKPFGKP
LLTRGGSGLQRLDGAFGALWGEQSGVKRFFISETITSGASILFGPLIAVMATLLAFLYDE
RRKPKGAGALRPLDAVGGSPLLPLTVPLSGLRTPPQTPTLDSRASTPSPWSVTKRPCWLG
RYAESSIPDYLRGQLPGLLGFDPAGVKAEGVVVPAKPLHAMEVARARVALLLAAGQVAPH
GAGVPAAVACLPLLQSAAEGPNLLQDPSPRSPSGIVSLRAPGHRVWLLAALAVLAAHLHR
LAAREGGGAAGRAAWAADLARVDPALLEALLGLPANRR
```

Supplementary Dataset S3. PSI and PSII subunit amino acid sequences in *E. gracilis*.

```
>PSBP2-Euglena gracilis_158060@comp51350_c0_seq3
SLHDASDLFFFGHMSYSTMNISYEAPARNSYAAPLIGAAVGAVIGVACFATASSINLYSASAVRPAVST
MAAPVAQTYQTSQVATAQFAAPRASYAEEAAQQNTAVYGLAFGMFAAVAGAVNGLFLWNNRRASM
AATTGRREALMAGAAAAGFAAAPAFAAYGEGANIFGKRKEVDQFFEISGDGWSGKLPAKYNASSEQE
AVGQVARWEDNFDAVNTTIVIVRPVGKNSVTEVGSIDDVRNNYVIPLLGQQAFDGPSISEGGFAPGRFG
SAAILDQKEEQQNGKTYYFYELLTRTADGNEGGRHHLFKVAASNGKLTILKSQAGDKRWFKGLERQL
RKVTDEFKVL
>PSBP-Euglena gracilis_158060@comp38210_c0_seq3
ALPISARKPLKTHFEELEVNSPTLKKCSVGLAVDPIKLPSLEAFGDPAFVGERVVGVEKTRDGVLSAQLL
GATNEPRDGLAYYELEYAVAGSRGEYHHLARVAVAKGYLFVLTAKVKEATWSEAEPIVRQILRSFRVQ
A
>PSB33-Euglena gracilis_158060@comp61861_c0_seq3
PAARWTAAVLRSLWAAPVVLLAGWAASRRVPALGMASTTGEQVSAEATDADYVFATHIAEIPKGTRK
LLSLRGASVLLFWYRNELYAIENRSPAEGAFNQGFEQARFTQNFGILCPSTDTEFSLKTGEVLNWMPNN
SVLRILTPPCPPLEVFPVQVRGEDVFVQVMEGERYYFDGGARSSFDRNNVFGIEPRMSLEDGTYVDEMG
TMKVNVDPSTVLITVIAVAIIAVAGTATCLYFENLPALVLFWLTGFGITAKLVLDLTGAAKE
>PSBA-Euglena gracilis_158060@comp53636_c0_seq3
KNRLYVGWFGVLMIPTLLTAATVFIIAFIAAPPVDIDGIREPVSGSLFYGNNIITGAVVPTSNAIGLHFYPI
WEATSLDEWLYNGGPYQLIVCHFFIGICSYMGREWELSFRLGMRPWIAVAYSAPVAAASAVFIVYPLG
QGSFSDGMPLGISGTFNFMIVFQAEHNILMHPFHMLGVAGVFGGSLFSAMHGSLVTSSLLRETTENESIN
VGYKFGQEEETYNIIAAHAYFGRLIFQYASFNNSRSLHFFLAVWPVVGIWFTALGVSTMAFNLNGFNFN
QSVIDSQGRVINTWADIINRANLGMEVMHERNAHNFPLDLADRKSG
>PSBD-Euglena gracilis_158060@AKL82343.1
MTFTDLNTENKNGWFDVADDWLKKDRFIFIGWSGLLLFPCSYLALGGWLTGITFVTSWYTHGLASSFL
EGCNALTAAVSTPPNSMGHSLLLLWGPEAQWDFTRWLQIGGLWPFIALHGAFGLIGFMLRQFEIAKAV
QIRPYNAIAFSAPISVFVSVFLIYPLGQSGWFFAPSFGVAAIFRFILFFQGFHNWTLNPFHMMGVAGVLGA
ALLCAIHGATVENTIFEDGDSPNTFRAFNPLQSEETYSMVTANRFWSQIFGVAFSNKRWLHFFMLFVPV
TGLWMSALGIVGLALNLRAYDFVSQEIRAAEDPEFETFYTKNILLNEGIRAWMAAQDQPHEQFIFPEEV
LPRGNAL
>PSBC-Euglena gracilis_158060@GDJR01004816.1_Tok.E.lc
MNNLYSRRRFFHVETLFNKKLTVGGRDQESTGFAWWAGNARLINVSGKLLGAHVAHAGLIVFWAGA
MNLFEVAHFVPQKPMYEQGLILLPHLATLGFGVGFDGIVLDTYPFFVCGVLHLISSAVLGFGGVFHSLA
GPDTLEESFRFFGYTWKRKKKLAAILGIHLCFLGLGALLLAWKAMYGGGVYDTWCPGGGDVRSITNPT
LNPFIIFGYLVKSPFGGEGWIVSVDSMEDVIGGHIWIGILLVSGGFWHICSRPSPWVVRTFVWSGEAYLS
YSLGAVATMGFIAVPMVWFNNTVYPSEFYGPTGPEASQAQAFTFLIRDQRLGTNIASAQGPTGLGKYL
MKSPTGEIIFGGETMRFWDFRGPWLEPLRGPNGLDLNKLKNDVQPWQERRAAEYMTHAPLGSLNSVG
GVATEINAVNFVNPRSWLATSHFVLAFFFFVGHLWHAGRARAAAIGFEKGIDRSREIARKLKPLD
>PSBC-Euglena gracilis_158060@comp61278_c0_seq2
MKVLELGWLHKISLMNNLYSRRRFFHVETLFNKKLTVGGRDQESTGFAWWAGNARLINVSGKLLGAH
VAHAGLIVFWAGAMNLFEVAHFVPQKPMYEQGLILLPHLATLGFGVGFDGIVLDTYPFFVCGVLHLISS
AVLGFGGVFHSLAGPDTLEESFRFFGYTWKRKKKLAAILGIHLCFLGLGALLLAWKAMYGGGVYDTW
CPGGGDVRSITNPTLNPFIIFGYLVKSPFGGEGWIVSVDSMEDVIGGHIWIGILLVSGGFWHICSRPSPWV
VRTFVWSGEAYLSYSLGAVATMGFIAVPMVWFNNTVYPSEFYGPTGPEASQAQAFTFLIRDQRLGTNI
ASAQGPTGLGKYLMKSPTGEIIFGGETMRFWDFRGPWLEPLRGPNGLDLNKLKNDVQPWQERRAAEY
MTHAPLGSLNSVGGVATEINAVNFVNPRSWLATSHFVLAFFFFVGHLWHAGRARAAAIGFEKGIDRSR
EIARKLKRSEER
>PSBP-PPL1-Euglena gracilis_158060@GEFR01013860.1_Cam.E.lc
```

```
QETSSAPLQTASGPTQSWSFEYPADWETARKPLKTHFEELEVNSPTLKKCSVGLAVDPIKLPSLEAFGDP
AFVGERVVGVEKTRDGVLSAQLLGATNEPRDGLAYYELEYAVAGSRGEYHHLARVAVAKGYLFVLT
AKVKEATWSEAEPIVRQILRSFRVQA
>PSBP-PPL1-PPL1-Euglena gracilis_158060@GDJR01044623.1_Tok.E.lc
AAIPSGFALHKDQYDGYEFLYPFGWQPLEVAGQDVLYKDIIEPLETMSVTIIPTDKESITELGDLERVCFT
FADKFLTSSDQAVTLLDTKQEVDDGITYYSMEFLAAAPNYTRHALAVVAICNNKLYTCLTGASERRWP
KMKDRLQTAVDSF
>PSAA-Euglena gracilis_158060@comp64894_c0_seq1
LTITPPEQQVKRVRVAFVSNPVETSFEKWSRPGHFSRLLSKGPNTTTWIWNLHADAHDFDNHTTDLEDI
SRKVFSAHFGQLAIIEIWLSGMFFHGARFSNYEAWLLDPIHVKPSAQIVWPIVGQEILNGDVGGNFQGIQI
TSGLFQLWRSCGITSEFQLYITALTGLIFSAVLFFAGWFHYHKAAPKLEWFQNVESMLNHHLSGLLGLG
CLSWAGHQIHVSLPINKLLDSGVNPAELPLPHDFILDKSLISQLYPSFSKGLAPFFTFHWSEYSDFLTFRG
GLNNVTGGLWLTDVAHHHLALAVLFILAGHMYKTNWKIGHDIKGLLESHTGPFTGQGHKGLYEIFTNS
WHAQLSLNLAMMGSLSIIVAQHMYSMPPYPYIAIDYGTELSLFTHHYWIGGFCIVGAAAHAAIFMVRD
YDPALNFNNLLDRVLLHRDAIISHLNWVCIFLGLHSFGLYIHNDTLSALGRPQDMFSDSAIQLQPVFAQ
WIQNTHYLAPTLTAFNLVSPTTPVWGGDVVSISGKVAMMPIKLGTADFLVHHIHAFTIHVTVLILLKGV
LFSRSSRLIPDKASLGFRFPCDGPGRGGTCQVSAWDHVFLGLFWMYNSISVAIFHFSWKMQSDVWGTV
LANKVSHITGGNFSQGSLTINGWLRDFLWAQSSQVIQSYGSPLSAYGLMFLGAHFVWAFSLMFLFSGR
GYWQELIESIVWAHNKLKVAPNIQPRALSITQGRAVGVAHYLLGGIATTWSFFLARIISVG
>PSAB-Euglena gracilis_158060@AKL82361.1
MATKFPKFSQGLAQDPTTRRIWFGIATSHDFESHDGMTENNLYQKIFASHFGQLAIIFLWTSGNLFHVA
WQGNFEQWIKDPLHIRPIAHAISDPHFGQPAIEAFTRSQFPGPVNIAYSGVYQWWYTIGLRTNVDLYNG
SMFLLFIATLALFAGWLHLEPKYSPKVSWFKDAESRLNHHLSALFGLSSLAWSGHLIHVAIPESRGIHVR
WDNFLSQLPHPSGLEPFFKGNWSLYSENPDSLTHVFGTASGAGTAVLTFLGGFHPETKSLWLTDIAHHH
LAIAVLFIVAGHMYRTNFAIGHRIDDILNAHKAPSGKLGLGHFGLYETINNSLHFQLGLALASLGVITSL
VAQHMYSLSPYAFLIQDRTTMAALYTHHQYIAGFIMTGAFAHGAIFFIRDFDEEKNKGNVLSRILDHKE
AIISHLSWVTLFLGFHTLGLYVHNDVMQAFGTPEKQILIEPVFAQWIQSAHGKNIYELNILLSSDSSNAFS
ASQAIWLPGWLNGINDKSTSLFLQIGPGDFLVHHAIALGLHTTTLILVKGALDARGSRLMPDKKDFGYS
FPCDGPGRGGTCDISAWDAFYLAVFWMLNTIGWTTFYWHWKHITLWQGNVGQFNESSTYLMGWLRD
YLWLNSSQLINGYNPFGMNSLAVWGWMFLFGHLVWATGFMFLISWRGYWQELIETLVWAHERTPITN
VFKWTDKPVALSIVQARLVGLAHFSVGYVFTYAAFLIASTSAKFG
>PSB28-Euglena gracilis_158060@comp54422_c0_seq3
TTAIQFVRGTDELTIPDVKLSRSKDGAAGRAVFRFQKPNIFEISSDIGQISGLYLIDEEGTLSSLQVTAKFIN
GSPVSIEATYITRSPWEWDRFMRWMERYAKENGLNFKKA
>PSBO-Euglena gracilis_158060@BI307576.1_Gen.E.lc
QSGEEGDDDLFKENNKTFDITTGRIEMEVNKVNAEEQEIGGVFVATQLSDTDMGSKVPKKVLTKGIFY
ARVE
>PSBO-Euglena gracilis_158060@EC667897.1_Gen.E.lc
TLAGVEGPLFYDKDGSLTFLEKEGIDYAATTVQLPGGERVPFMFTVKDLEAKSTDPSGISAGTEFGGTF
KVPSYRSGLFLDPKGRGGTTGYDMAQALPGMEADGVEGQEELFKETNKVFQVTDGTIEFAVNKVDAA
TGEFSGVFVSEQLSDTDMGSKAPKKVLLKGIIFGKVGAATGEESY
>PSBO-Euglena gracilis_158060@GEFR01015765.1_Cam.E.lc
LTFDELQSLSYLEVKGSGVAGQCPFIADGASGKLNLKGGKYEINNMCLEPSSFQVKLPPTAKQPVPGFE
KTKLLTRMTYTLDPISADLNVGSDGSWTIQEKDGMDYAATTVQLGGGERVPFLFTVKNLQAKGDGGQ
FLGKFDVPSYRGATFQDPKGRGGATGYDTAVALPASGDDAQYTKENNKSTESGVGTIAFKVAKVNAE
TGEIAGVFESIQPSDTDLGAKLPKDIKTSGVWYAQIS
>PSBO-Euglena gracilis_158060@EC674035.1_Gen.E.lc
AALIGAAVAGSANAASLTYDELQSLSYLEVKSSDLACTCPVLADGVSSKLSLKAGKYEINNWCLEPSSF
QVKLPPTEKQQVPEFERTKLMTRLTYTLDAISADLNVGGDGSWTIQEKDGLDYAATTVQLAGGERVPF
LFTIKNLLAKGDAGQFLGQFDVPSYRGATFLDPKGRGGASGYDTAVALPASGDDEEYAKENSKSTAAS
VGTIAFKVAKVNAETGEIAGVFESIQPSDTDLGAKVPKDIKTSGVWYAQISPSK
>PSB32_A-Euglena gracilis_158060@comp58153_c0_seq3
SVYFFSIHWWSGLMAPANTYSTAADVESLILSSAPTTRTSRPWQICLATLLGMTCGALFVLHPGPTATTL
PVLPLQPQAATGIVQKPVPSKAQDTRSLVQKVGNGLAALSLATMANLGTPAIAPASELDVLYSPTPTTS
YLYDDGGILSKITVFDLQSKLPEIEKKTGYRINVVTLRKLNSFDDAFQFADKLIENWYPTAELGDKKAVF
VIVKYSKEGGLVGGPGFSAAVPDSLLESIFNDNLGALLPEEKFNQAVESSIDRLAAVLEGKEDPGAPRLL
YTKGEKQTATYKTKAETDEKRGVYVGVFGALMFIAFVAPMVQFFGYTKDD
>PSB32_B-Euglena gracilis_158060@GDJR01037673.1_Tok.E.lc
DDLDAASLLPPGPPMNVVDKVQLLTNSQMKEATRLIDGVEARYGVKLRVLIQKSPLPPSEAVSEYWQA
DRNTVILIVDDSFPNLLHFSYWRPGRFWIHLASDYGTLKYVRENGEDQAVLNALRAIDEGL
```

```
>PSB32_B-Euglena gracilis_158060@comp60094_c0_seq3
LSFAIGFCGTVLVLNRSTFLQVPNTTAGRLVAQLRTSASNLPTIGRRDVLATGIAAAVAMTPQVALARK
PGVNKPELLPAGDVVNVIDLAGMLTPSQIKELERNIEAIQRLYGVKLRVLCQRYPETPGEAIVDYWKVD
DSTIVLVADDTFPNLLNFNVGVDADLLAPNGRFWTQLRANFGTQTFVRKNGEDEAILEAVEAILDGMT
NVDQRGKSMLLGL
>YCF48-Euglena gracilis_158060@EC672176.1_Gen.E.lc
AEEKNSSKWKQISLPVDTVLFDLDFVKDSPKHGWLVGSKGTFLETIDGGENWIPRTFANLDPDEELTYR
FETISFSGNDGWVIGKPAIMLHTKDGGKSWLRVPVSPKLPGEPISITALGPDSAEMTTSSGAVYITDNGG
RNWKAQVKETIDSTLNRTISSGVSGASYFTGSIIGVIRDLTGKYLAISSRG
>YCF48-Euglena gracilis_158060@comp62730_c0_seq2
GSRRQVLAGGVAGMAAVTTLAADVRPASAGLFGGGVEEPADWEQVFIPDVSGQDFTFYDVAMLDDK
RGFLIGSRQLFLETNDGGKTWLPRSIDTQDDLPYRFTSISFNGDEGWMVGKPPILLHTTNAGKDWRRIG
LSANLPGNPILIQALAEPGCAEMCTDQGGIYVTNNTAQTWKAAVEETVDATLNRTVSSGISGASYYTGS
FGNVTRNDAGQYIGVSSRGNFYMTWEPGQTFWMPHNRSGARRIQNMGFKPDGGVWEIGRGGQLAFA
KGTNLPTDADNEEFEQLKVNSRGFGLLDVAWRTDNEVWTVGGSGTLIKSSDGGQKWARQRSLDDIAA
NLYEVRFFNNGKIGFVLGNSGVLLRYLSK
>PSAC-Euglena gracilis_158060@comp40626_c0_seq2
MSHSVKIYNTCIGCTQCVRACPTDVLEMVPWDGCKAGQIASSPRTEDCVGCKRCESACPTDFLSVRVY
LGSETSRSMGLAY
>PSAE-Euglena gracilis_158060@comp52439_c0_seq2
SDLFFFGQITMAFNNAMADPWFKFTAVSSATLMAIGCALMAIAELTAPQQQALYAAPAVQTRMITRLA
DDAAARVADRTQFIEEYKERRAEAPATPKAKAGPARGSTVKVTRPESYWYNQYGKVISVDESGAIRYP
VVVRFEKENYQGVTTNNFALDELEY
>PSB29-Euglena gracilis_158060@comp56561_c0_seq2
RSIFFRPHQAPMSSTMEFKHNSFTTSVCTAIAVSTVVFATTFATKHINQYATSTSSSLATKAVPIAPLPRT
GQVLHTSFPRTSRLSVKETGFQSTPSLTVNFVQFALAGSLAIVSFFVVLWRKNSIRSPFSMAAVAGTRPT
LADTKLRFLTLCDLSIPNMYSIPLQELLAQHHLIRFSSKYQYDEVFALGIFTVINQLLEGLPEADSERIQTA
YFSALDDNFAKFKADATELELWASSKGKELNLSGSDDVTTKLKAVATRAAGGDFQYSKFFSIGLFQLL
QQSGQTDPESFTQLVNGMNLDKERVTKDLKIYKDLLKRLEMAKDMKEKFLERERKKQEERQAQKAKP
AEV
>PSAF-Euglena gracilis_158060@comp59715_c0_seq1
LATAGVAASAVLASGAMTAMPANATEAILGLTPCAESAKFQKVLTKEVKALEKREKLYEPGSAPFQAL
EATKAKTQARFAKYAESGLLCGADGLPHLISDPGLAVKYGHAGETLIPTIGFLYVAGYIGTAGRTYLQEI
STRQKPTQSEIIIDVPLATSIAFKSWDWPNKAVKELASGTLLEDAKNVPVSRR
>PSAD-Euglena gracilis_158060@GEFR01003691.1_Cam.E.bf
SYKGSTGTLLSAAEKEEKYVITWTSPKKQIFELPTGGAAEMEEGENIFFFARKEQCLALGAQLRTAFKPR
IENFQIFRVFPNGEVQYLHPKDGVFPEKVNQGRTKANHNPRSIGENAEPASVKFTGTTPKDV
>PNSL23-PSBQ-Euglena gracilis_158060@comp55004_c0_seq2
MSTLKQVAIVAPCFLLGLVGALVMTAPNAAVQHNVRAQLVGASVAAAAIANPVFAAPTPIDVADYRN
QREGYEYIYQARDLDLDEDVRAGYAQARDPSVAKARISESVSRLINLGSLVKKAYWTAGREELRGQVG
TLRFDIDTVASFKSSNAEKKALLAAKADLFTAIDKLDFAMRKKDQASALKFYDEVAAKAAAIANA
>PSBH-Euglena gracilis_158060@AKL82400.1
MTTISKNKTSNSKGKTTTLGTILKPLNSKYGKVLPGWGTAGIMLIFMTLFAIFLTIILEIYNSSVILDIK
>PSBH-Euglena gracilis_158060@comp45461_c0_seq1
FRSFHPEFNIMTTISKNKTSNSKGKTTTLGTILKPLNSKYGKVLPGWGTAGIMLIFMTLFAIFLTIILEIYNS
SVILDRSEER
>PSBM-Euglena gracilis_158060@comp52703_c0_seq1
SDLFFFDHLKMQRAGYQNLEGKAWSTTAVVALVATASLAVGVTAGIAFSSSAETTLYAPVATTLQPAV
VQAPISQMAATNAGIAAATGALVGAIAYKAKAAVQATMLAAQPAWAMQTNSQAPIATALFIALPTFFL
VILYVKTAGNPDIGSGGNNQEYYDRSKAAGIKRANMVVEMTGEGVEMYSKRRDQK
>PSBB-Euglena gracilis_158060@comp59418_c0_seq1
MGLPWYRVHTVVLNDPGRLISVHLMHTALVSGWAGSMALYELAIFDPSDVALNPMWRQGMFVLPFM
TRLGVTKSWGAWSVTGESFSDPGIWSYEGVAVAHIILSGLLFLAAIWHWVYWDLDLFRDPASGELKLD
LPRVFGVHLFLSGALCLAFGVFHVTGVFGPGIWVSDPYGLSGKIEPVIPSWGAEGFDPYNVGGIASHHIA
AGLLGLIAGGFHVLVRPSQRLFVLLRMGNIETVLSSSIAAVFWSAFVVSGTMWYGSASTPIELFGPTRYQ
WDKGYFQEEIERRVQASLSDGCSLSEAWGAISPKLAFYDYIGNNPAKGGLFRSGPMNNGDGIATAWLG
HAVFIDKEGNSLFVRRMPTFFETFPVILLDQNGVVRADIPFRRAESKYSIEQVGVTVRFFGGSFDTLSFND
PATVKRYARRAQLGEIFDFNRSILQSDGVFRSSPRGWFTFGHLSFALIFFFGHIWHGARTLFKYLLAGIDP
DLEEEIEFGTFEKLGDDTTKKELV
>PSBX-Euglena gracilis_158060@comp55270_c0_seq2
```

```
AVAACVMMGIWTASQPSVSFVQAPVRPATTTSAAVPVEARGVDFRQATVAAGAAMSTLAPMAAQAA
EYSSAKALGDSVFGGALILGLLLGAVVLVGGFDKIERRGRR
>PSBK-Euglena gracilis_158060@AKL82352.1
MSFITLNNFFNTNTFFGLLPEQYAPFDPLIDVLPIIPVLFFLLAFVWQASVKFR
>PSBJ-Euglena gracilis_158060@AKL82365.1
MSNSENTGRIPLWLVLTIIGLAAIALLALFFYGSYSGLGSSL
>PSBZ-Euglena gracilis_158060@AKL82386.1
MLLFTFTFQALVLALIIFSFILVLTLPVIFASPKGWENNKSRIWLACRFWFFLVFLIGILDGIFL
>PSBI-Euglena gracilis_158060@AKL82358.1
MLILKVFVYALILIFVSLFVFGLLSNDPGRNPYDDTGE
>PSAJ-Euglena gracilis_158060@AKL82384.1
MKYFTTYLSTAPVVAVLWFTLTASLLIEINRFFPDIL
>PSBN-Euglena gracilis_158060@GEFR01055614.1_Cam.E.lc
IFLGFFIISTTIYSIYMGFGPMSRKLKDPFEEHEN
>PSBT-Euglena gracilis_158060@AKL82401.1
MEALVYTFLLIGTLGVIFFAIFFRESPRINK
>PSAM-Euglena gracilis_158060@AKL82350.1
MEITTNQVYIALLASLIPAFFAFKLGKSLNQ
>PSBR-Euglena gracilis_158060@EC673396.1_Gen.E.lc
KGGKARGGQVGVGYKGSTEPGSAPPTRDGKPGYVYKLGLKNGLGNVDEYSPIYTPETWKKDGDVYEP
GATGLALWAAGLLALLGVSAYLIYATSQL
>PSBY-Euglena gracilis_158060@comp51240_c0_seq2
LPIYFFSLFHLLTMYGESQSWSVGMLAVAAVLGFVGGASLGHLYAGANTNLYVATNQPVANRAMAV
RAMPRNVEAVIPQTVAAASILAASEPAMAAQQVAAVADNRLALVGLILLPALGWAAFNILKPALRQLE
NMQKKAAGRR
>PSBW-Euglena gracilis_158060@comp49777_c0_seq3
RAGLALWAAIGAALCVMRVASPATTNLKVQTRPIVQPAITQIRNVGAAVTSAATAAAVQAPAAMALV
DERLNGEGTGLPLGINDPILFWVMLTVFTTIWALWATSVKEISNNDDPDSGVGLD
>PSB27-1-2-PSB27-1-Euglena gracilis_158060@GDJR01026382.1_Tok.E.lc
ESVYDADSIRILEEIRKVLALKANDPTTASAVDALCDQLGEWLAKYRREFEFEGEASYIKVYHAVRIMA
GHYAAEGNTVPLDPDRSAKVLVLIDEAELLLKQGK
>PSB27-1-2-PSB27-1-Euglena gracilis_158060@comp56040_c0_seq1
AATLTSFTTAFFLVFSRHRNVSLAAASGQKDEEIGLSRREALLGLSAAGALATSAPAYAFLGLGEDAGY
PGDTQNIIAEIKSVLVLAADDPTKPDRITTLRRDMNSYVAKYRRDTRYGGRASFGNVYSVINAVAGHY
NTFGVKTPIPKKRATQIVSEIGTAEQLLAKGR
>PSB27-1-2-PSB27-2-Euglena gracilis_158060@comp44499_c0_seq1
SSDLAAGVASLGMPASAGALDLNLGNLFGKPEEEEGPLPKEYRTGTLKLVAAIRKTLEAEAAGAKESE
VRKQGSDVRPLIKDFLDPKWEKTSAVVADPSFSAMKSAIRLLGQFYGQNGPRATMTEEARGPILALLAE
AEVALEAAAPPAAQK
>PSBE-Euglena gracilis_158060@AKL82362.1
MAGSTGERPFSDIITSIRYWVIHSVTIPSLFVGGWIFVSTGIVYDIFGTPRPSEYFTETRQQAPLISDRFNAL
EQMDQFTK
>PSBF-Euglena gracilis_158060@AKL82363.1
MTTNKDTRYPIFTFRWLAVHALAIPTVFFLGSISAMQFIQR
>PSBL-Euglena gracilis_158060@AKL82364.1
MAQTNSNKKTVELNRTSLYWGLLLIFVLAVLFSSYIFN
```