AUTOMATED BIOINFORMATICS TO EVALUATE THE RELATIVE CONTRIBUTION OF ENDOSYMBIOSIS AND KLEPTOPLASTY IN THE EVOLUTION OF COMPLEX RED ALGAE

Mick Van Vlierberghe (mvanvlierberghe@doct.uliege.be) and Denis Baurain (denis.baurain@uliege.be)

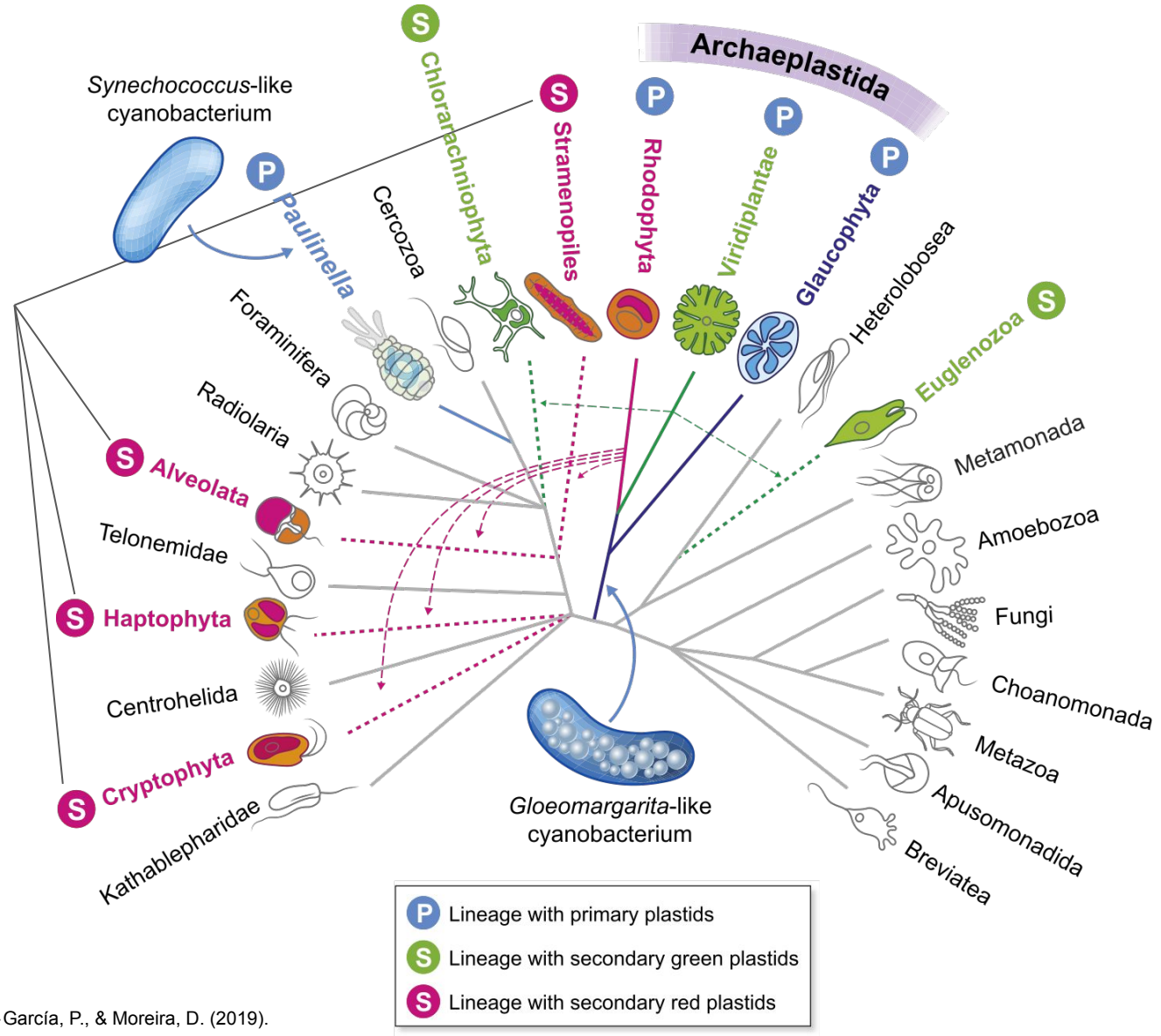
InBioS – PhytoSYSTEMS, Eukaryotic Phylogenomics, University of Liège, Liège, Belgium

7th European Phycological Congress
Tuesday, 27th August
Zagreb, Croatia

Ciccarelli et al., 2006

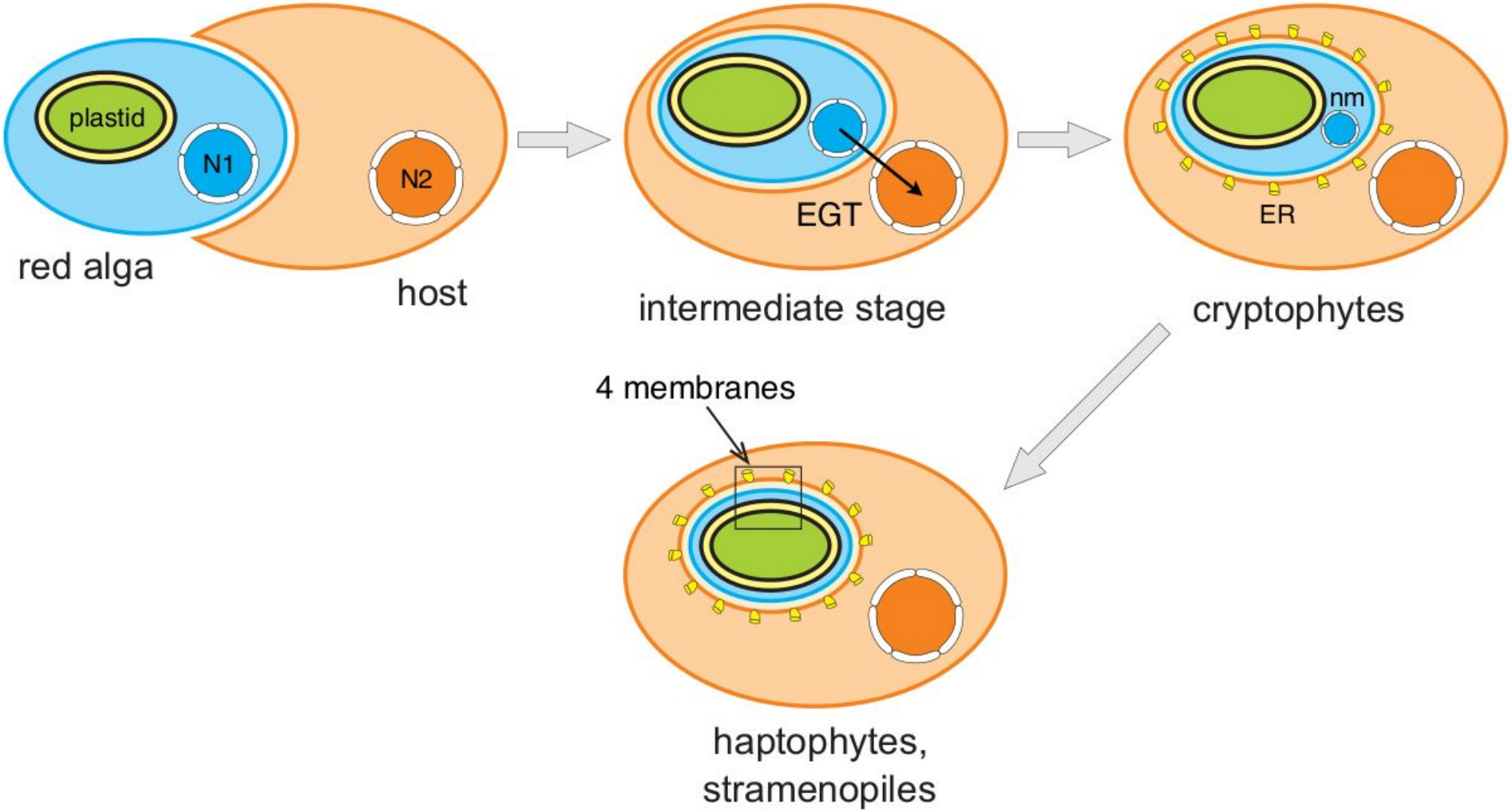
Complex Algae - Panorama

CASH lineages



Ponce-Toledo, R. I., López-García, P., & Moreira, D. (2019).

Endosymbiosis and EGT



Bodyl et al., 2009

Gene mosaicism

REPORTS

Genomic Footprints of a Cryptic Plastid Endosymbiosis in Diatoms

Ahmed Moustafa,[1]* Bánk Beszteri,[2]* Uwe G. Maier,[3] Chris Bowler,[4,5] Klaus Valentin,[2] Debashish Bhattacharya[1,6]†

ARTICLE

OPEN
doi:10.1038/nature11681

Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs

Bruce A. Curtis[1,2,3], Goro Tanifuji[1,2,3], Fabien Burki[3,4], Ansgar Gruber[5]†, Manuel Irimia[6], Shinichiro Maruyama[1,2,3], Maria C. Arias[7], Steven G. Ball[7], Gillian H. Gile[1,2,3], Yoshihisa Hirakawa[3,4], Julia F. Hopkins[1,2,3], Alan Kuo[8], Stefan A. Rensing[9]†, Jeremy Schmutz[8,10], Aikaterini Symeonidi[9], Marek Elias[11], Robert J. M. Eveleigh[1,2,12], Emily K. Herman[13], Mary J. Klute[13], Takuro Nakayama[1,2,3], Miroslav Obornik[14,15,16], Adrian Reyes-Prieto[3,17], E. Virginia Armbrust[18], Stephen J. Aves[19], Robert G. Beiko[20], Pedro Coutinho[21], Joel B. Dacks[13], Dion G. Durnford[17], Naomi M. Fast[4], Beverley R. Green[4], Cameron J. Grisdale[4], Franziska Hempel[22], Bernard Henrissat[21], Marc P. Höppner[23], Ken-Ichiro Ishida[24], Eunsoo Kim[25], Luděk Kořený[14,15], Peter G. Kroth[5], Yuan Liu[19,26], Shehre-Banoo Malik[1,2,3], Uwe G. Maier[22], Darcy McRose[27], Thomas Mock[28], Jonathan A. D. Neilson[17], Naoko T. Onodera[1,2,3], Anthony M. Poole[29], Ellen J. Pritham[30], Thomas A. Richards[26], Gabrielle Rocap[18], Scott W. Roy[31], Chihiro Sarai[24], Sarah Schaack[32], Shu Shirato[24], Claudio H. Slamovits[1,2,3], David F. Spencer[1,2,3], Shigekatsu Suzuki[24], Alexandra Z. Worden[27], Stefan Zauner[22], Kerrie Barry[8], Callum Bell[33], Arvind K. Bharti[33], John A. Crow[33], Jane Grimwood[8,10], Robin Kramer[33], Erika Lindquist[8], Susan Lucas[8], Asaf Salamov[8], Geoffrey I. McFadden[34], Christopher E. Lane[1,2,3,35], Patrick J. Keeling[3,4], Michael W. Gray[1,2,3], Igor V. Grigoriev[8] & John M. Archibald[1,2,3]

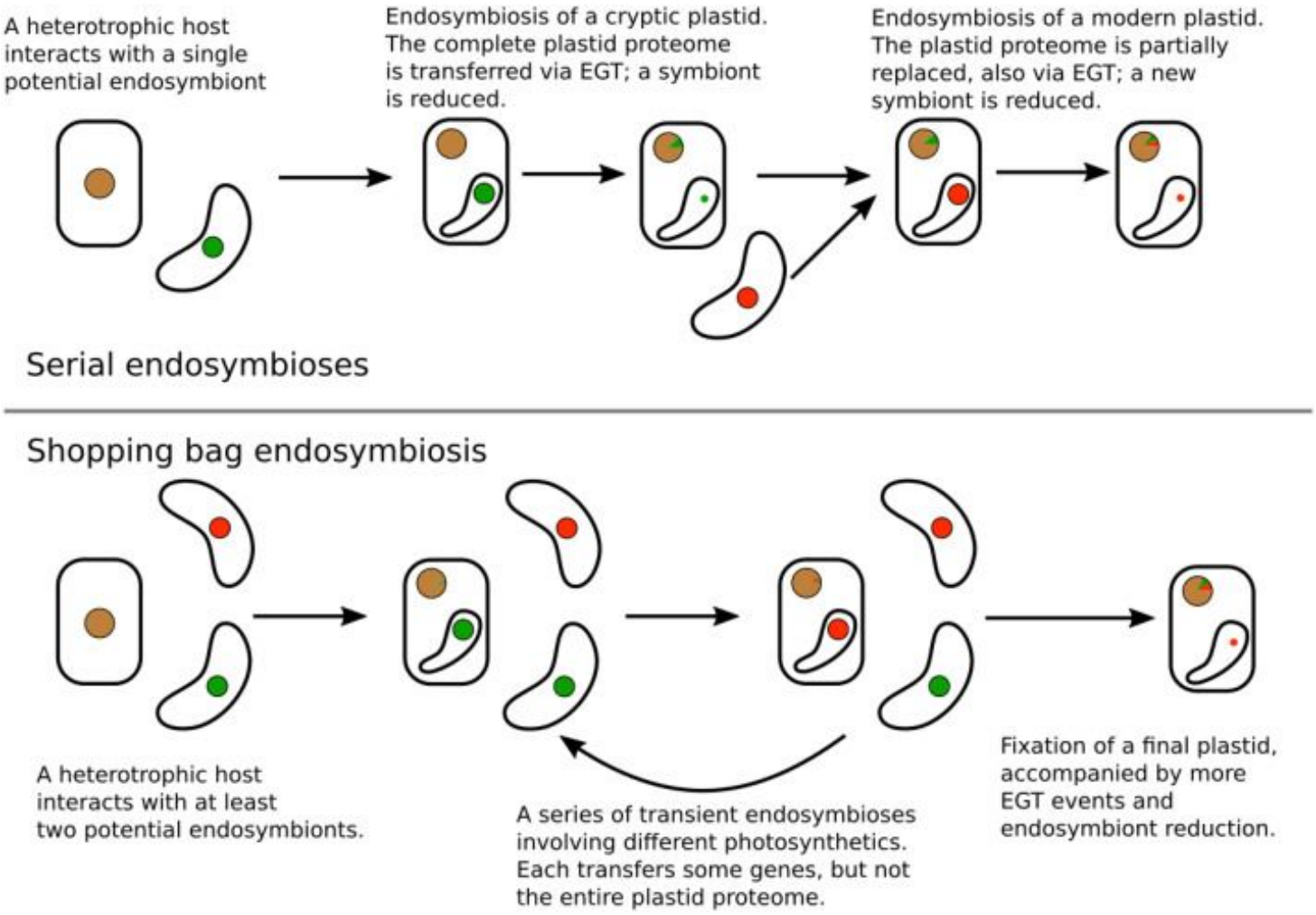
eLIFE
elifesciences.org

RESEARCH ARTICLE

Chimeric origins of ochrophytes and haptophytes revealed through an ancient plastid proteome

Richard G Dorrell[1]*, Gillian Gile[2], Giselle McCallum[1], Raphaël Méheust[3], Eric P Bapteste[3], Christen M Klinger[4], Loraine Brillet-Guéguen[5], Katalina D Freeman[2], Daniel J Richter[6,7], Chris Bowler[1]*

Models

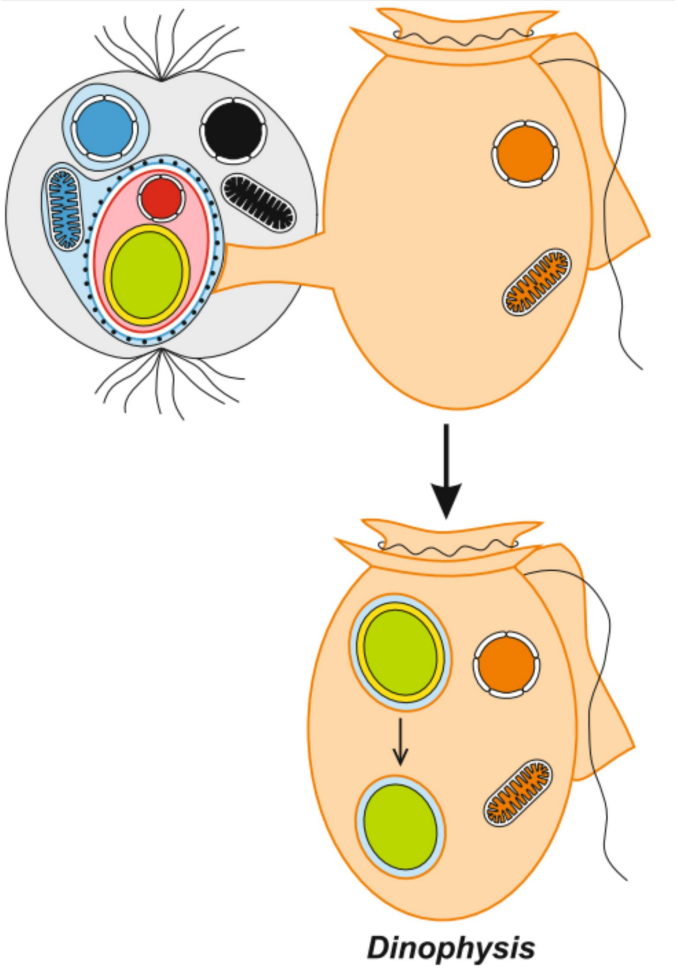


MorozovA.A.GalachyantsYuri P 2019.

“Shopping bag” by kleptoplasty

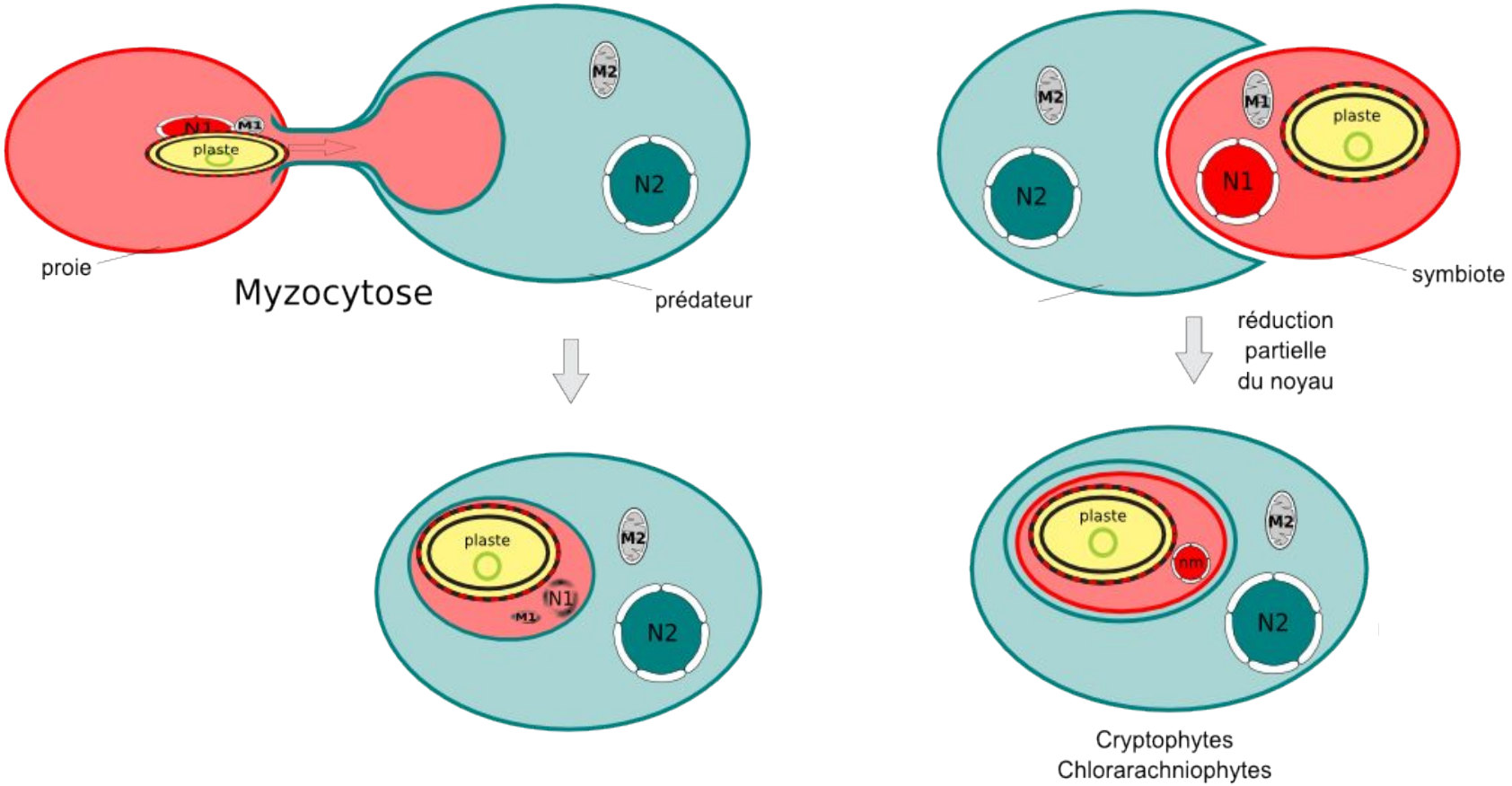
- real selective pressure

Kleptoplasty by myzocytosis

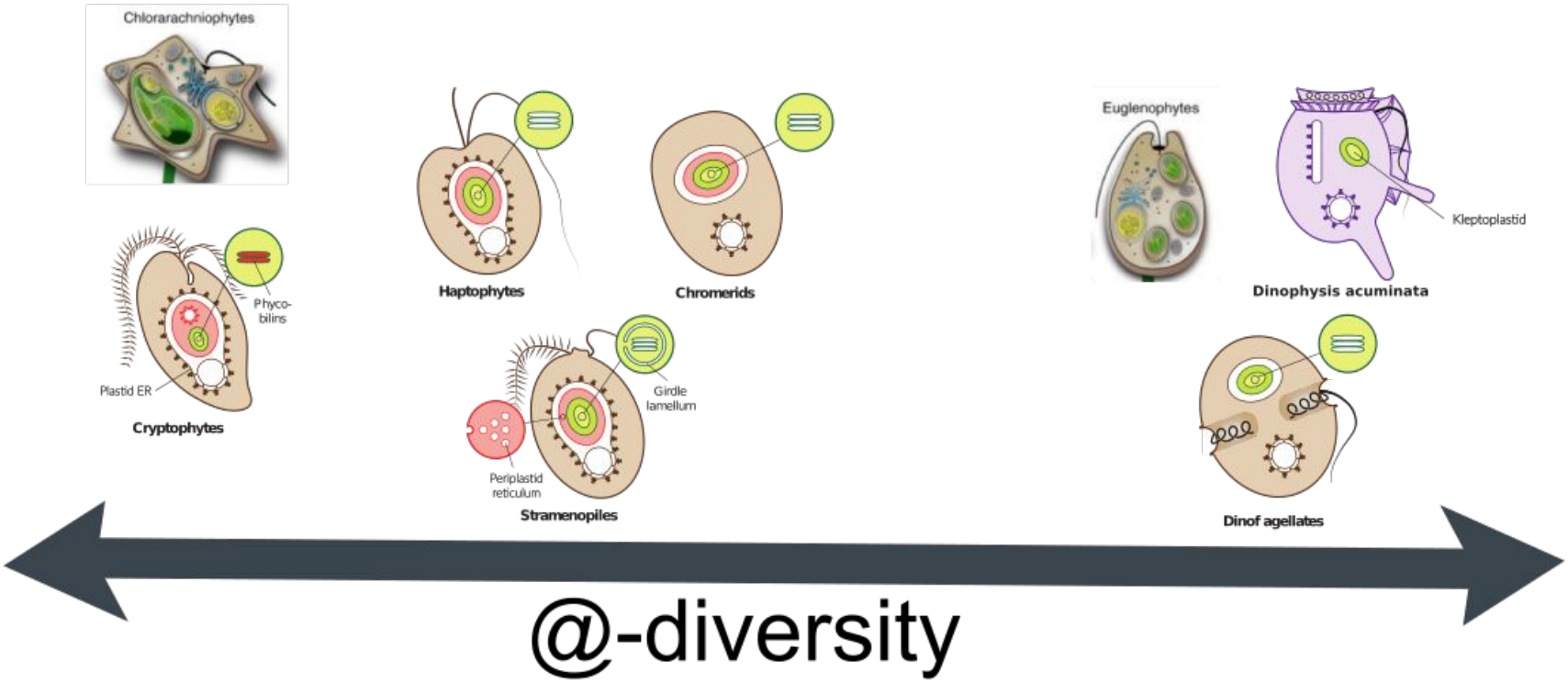


Gagat et al., 2014

Origin of 3 membranes plastids ?



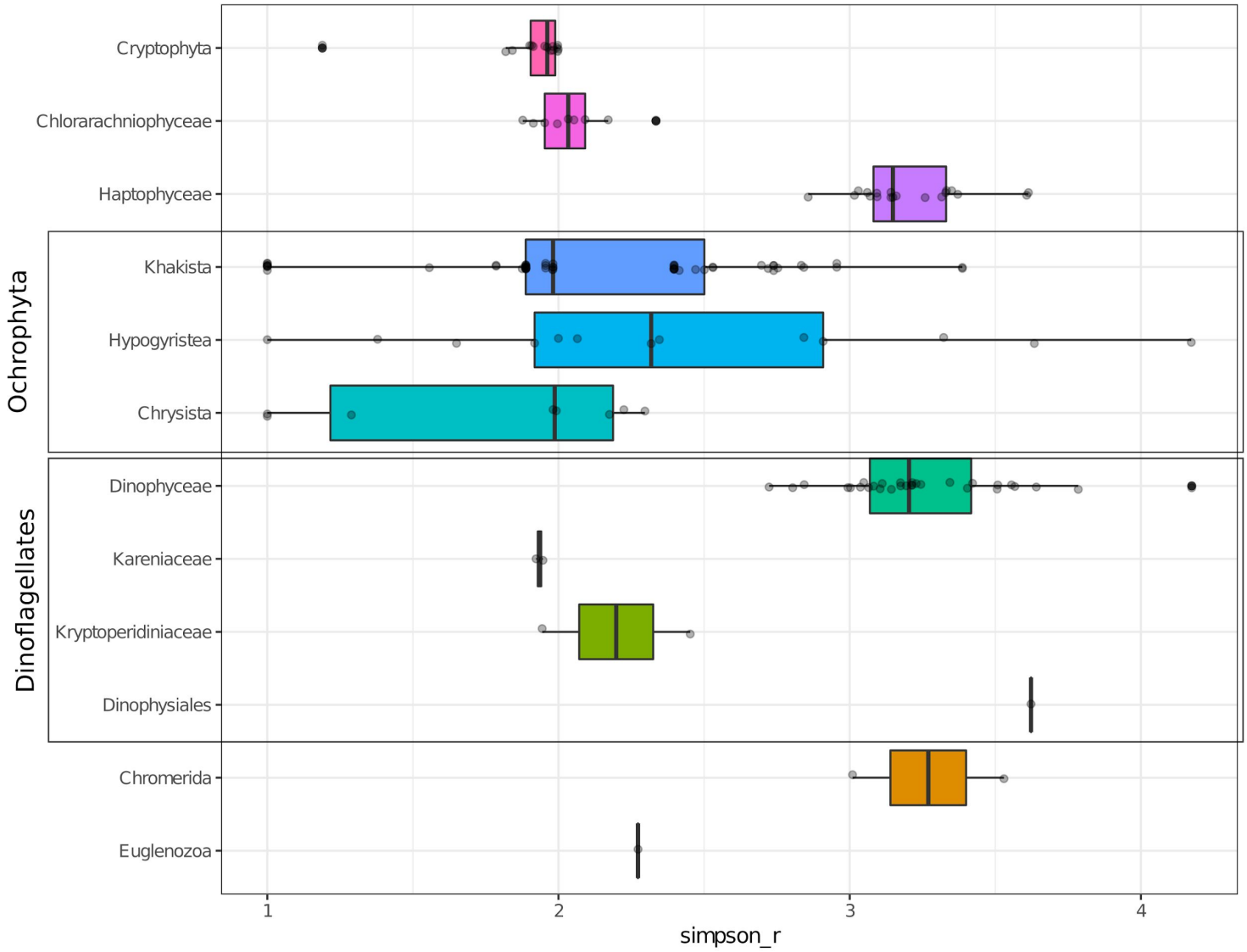
Testing the hypothesis



AIM

Provide additional (phylogenetic) data
to illuminate the debate…
in an automated fashion

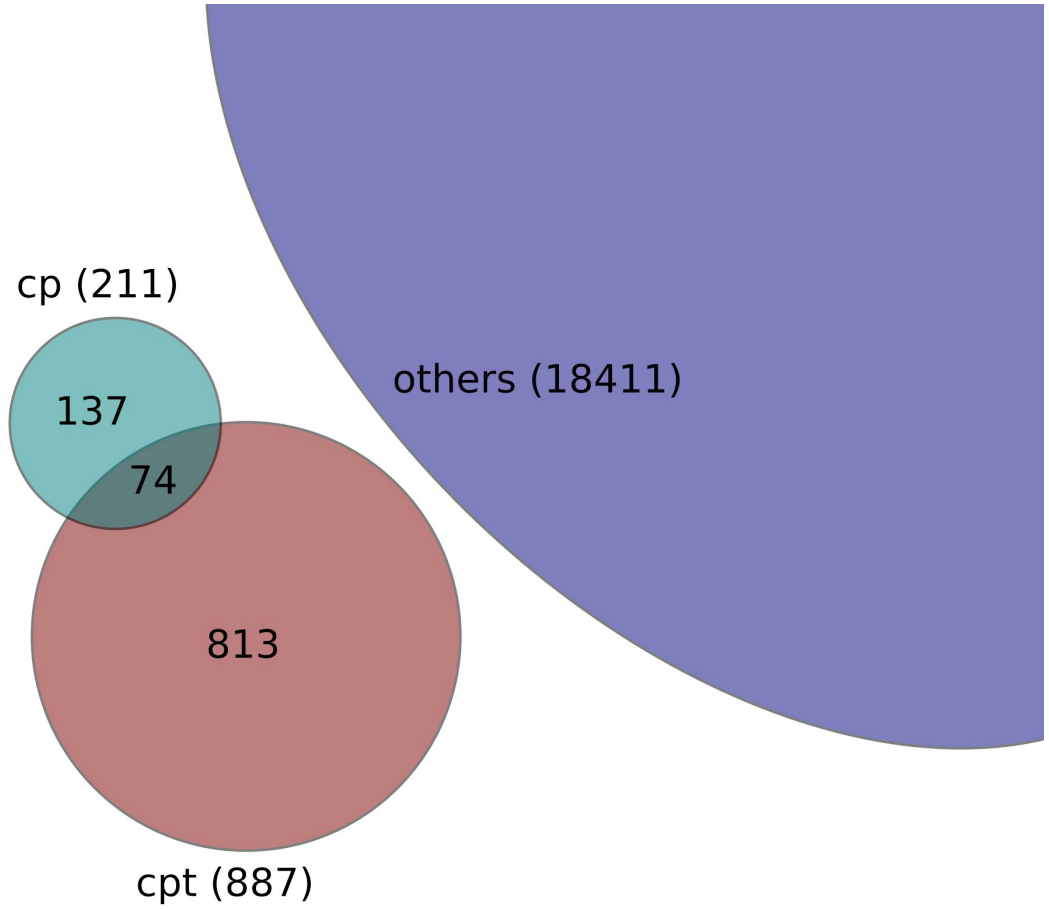
Gene diversity

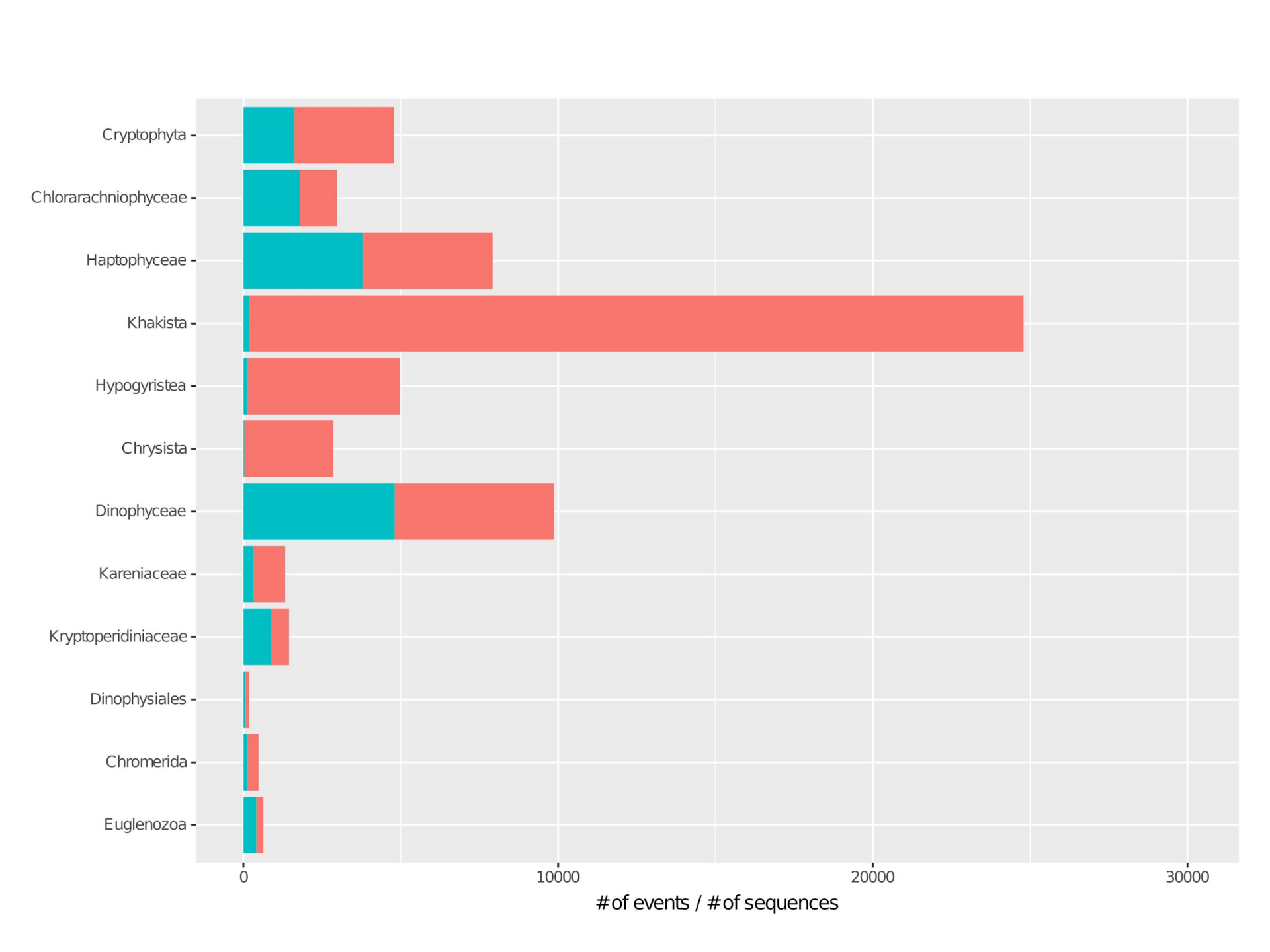


Gene set

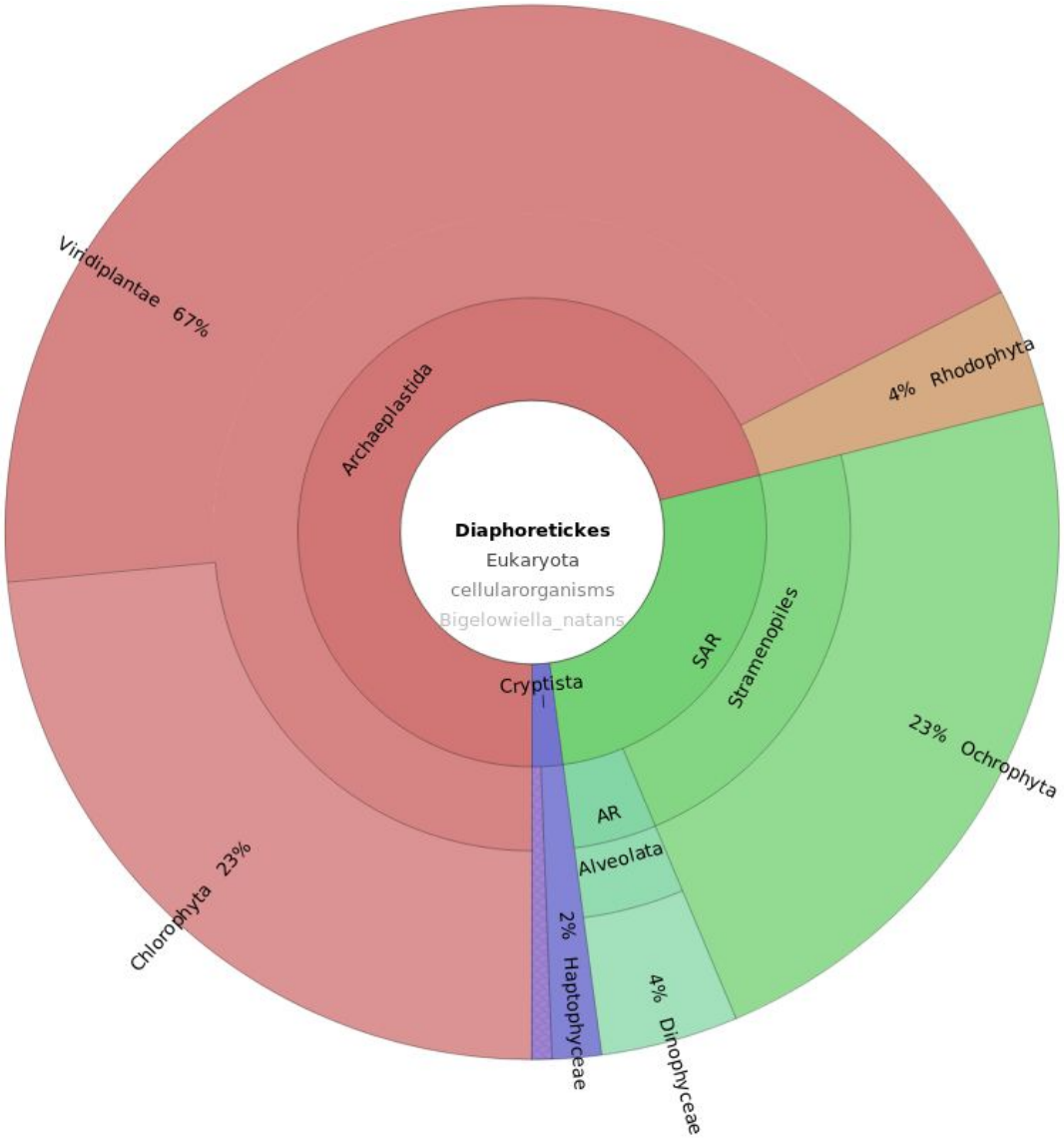
- plastid targeted genes

  (experimental)

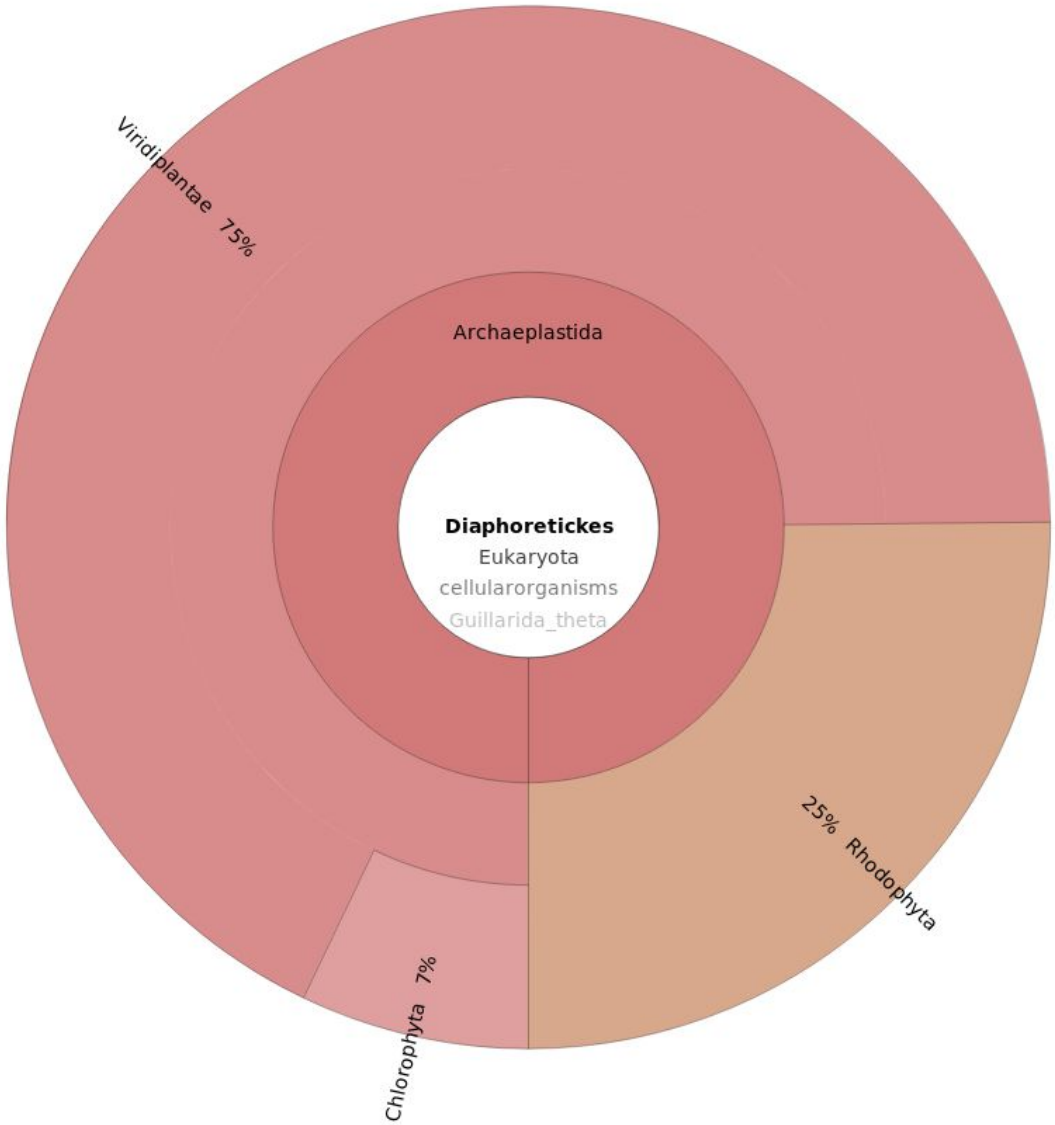



Cryptophyta
Chlorarachniophyceae
Haptophyceae
Khakista
Hypogyristea
Chrysista
Dinophyceae
Kareniaceae
Kryptoperidiniaceae
Dinophysiales
Chromerida
Euglenozoa
0
10000
20000
30000
#of events / #of sequences

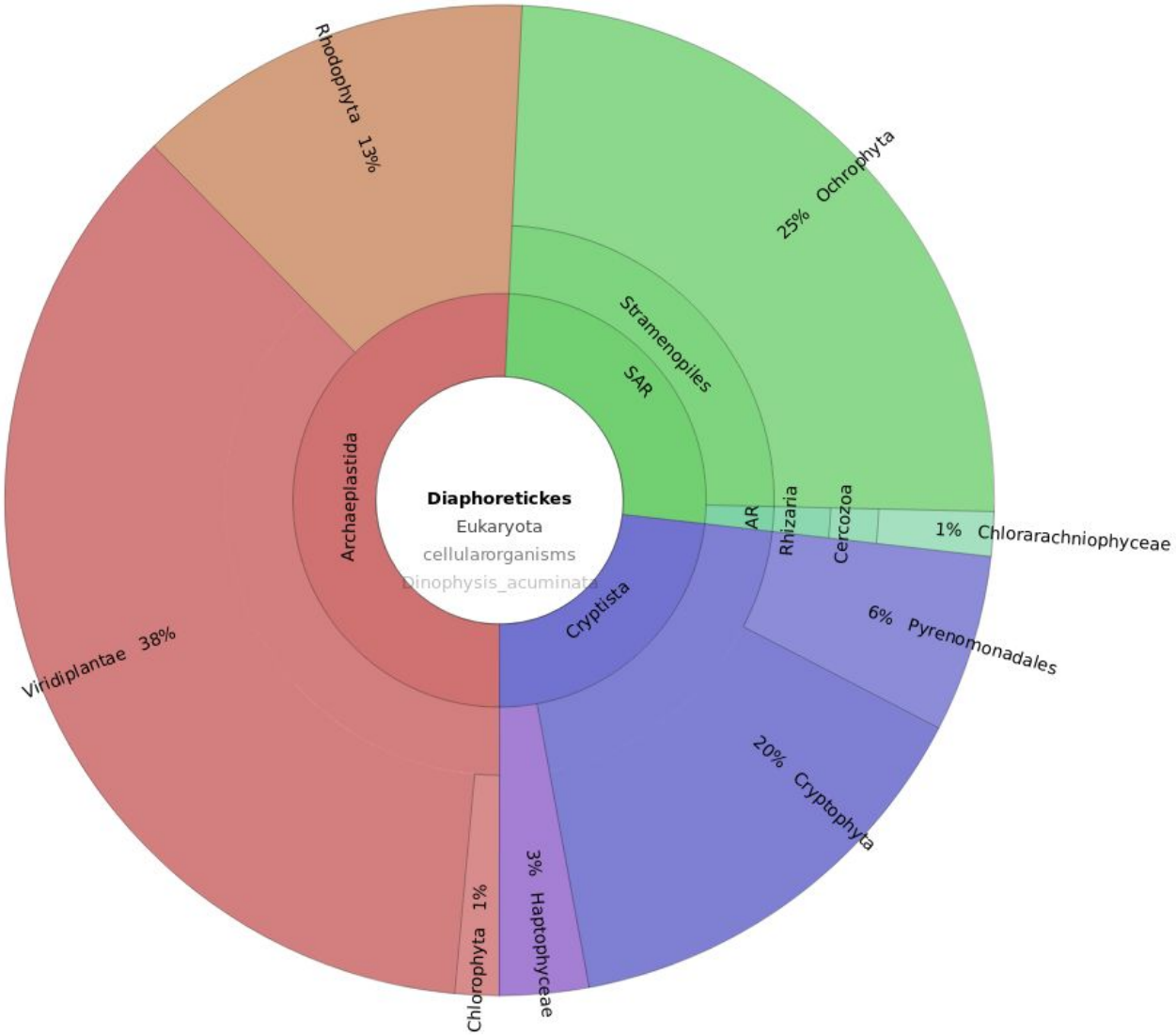
Bigelowiella natans



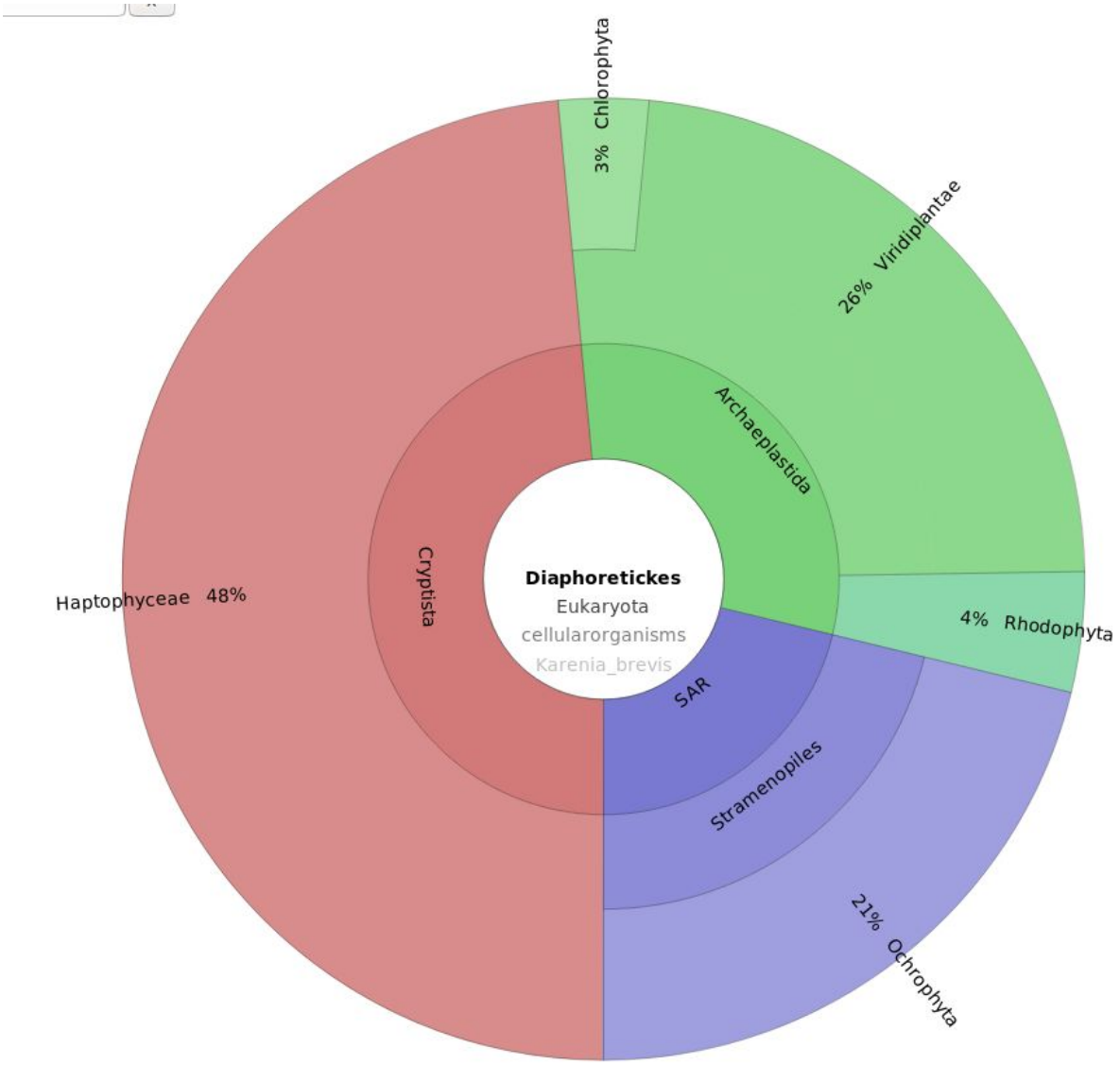
Guillardia theta



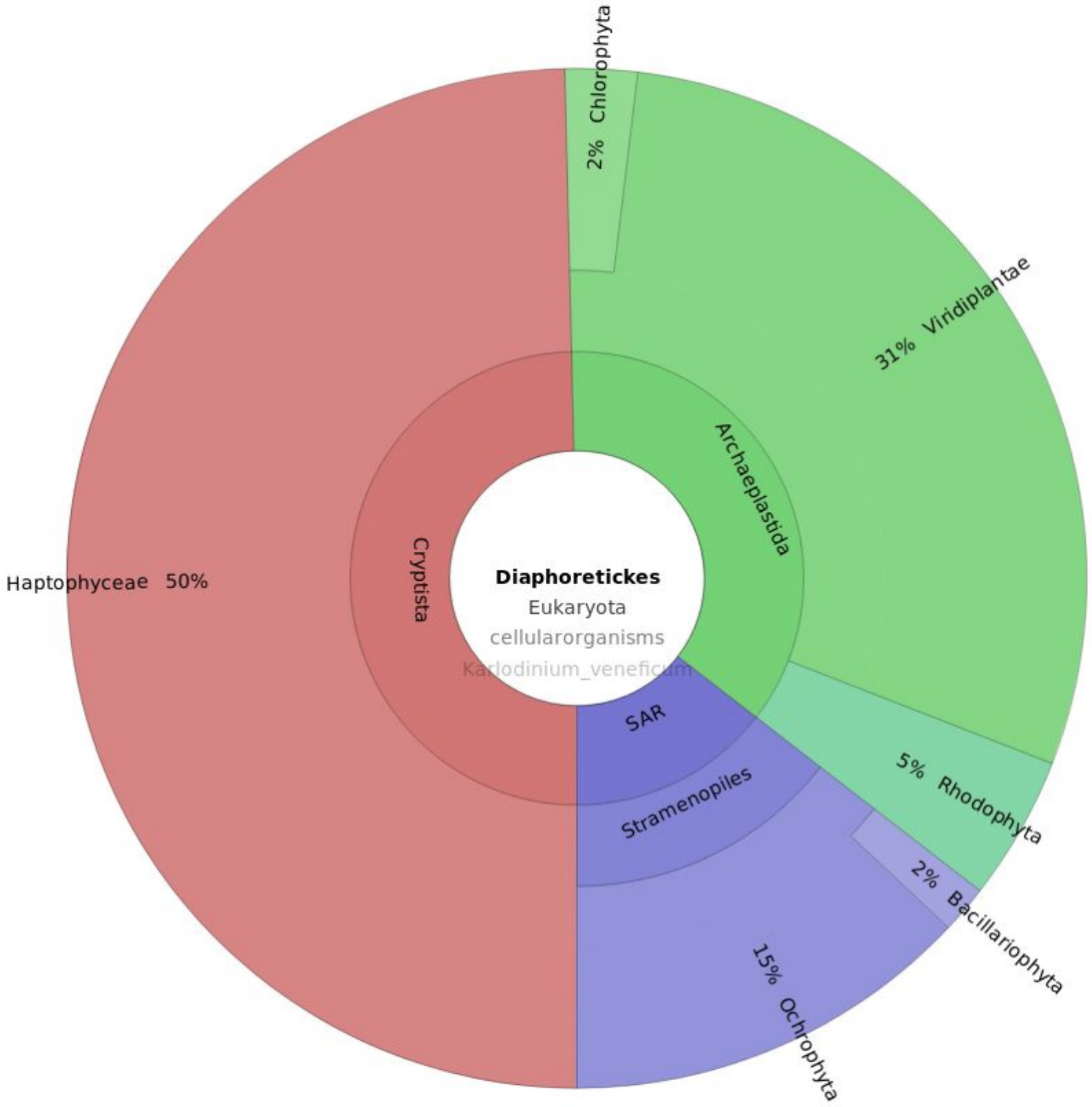
Dinophysis acuminata



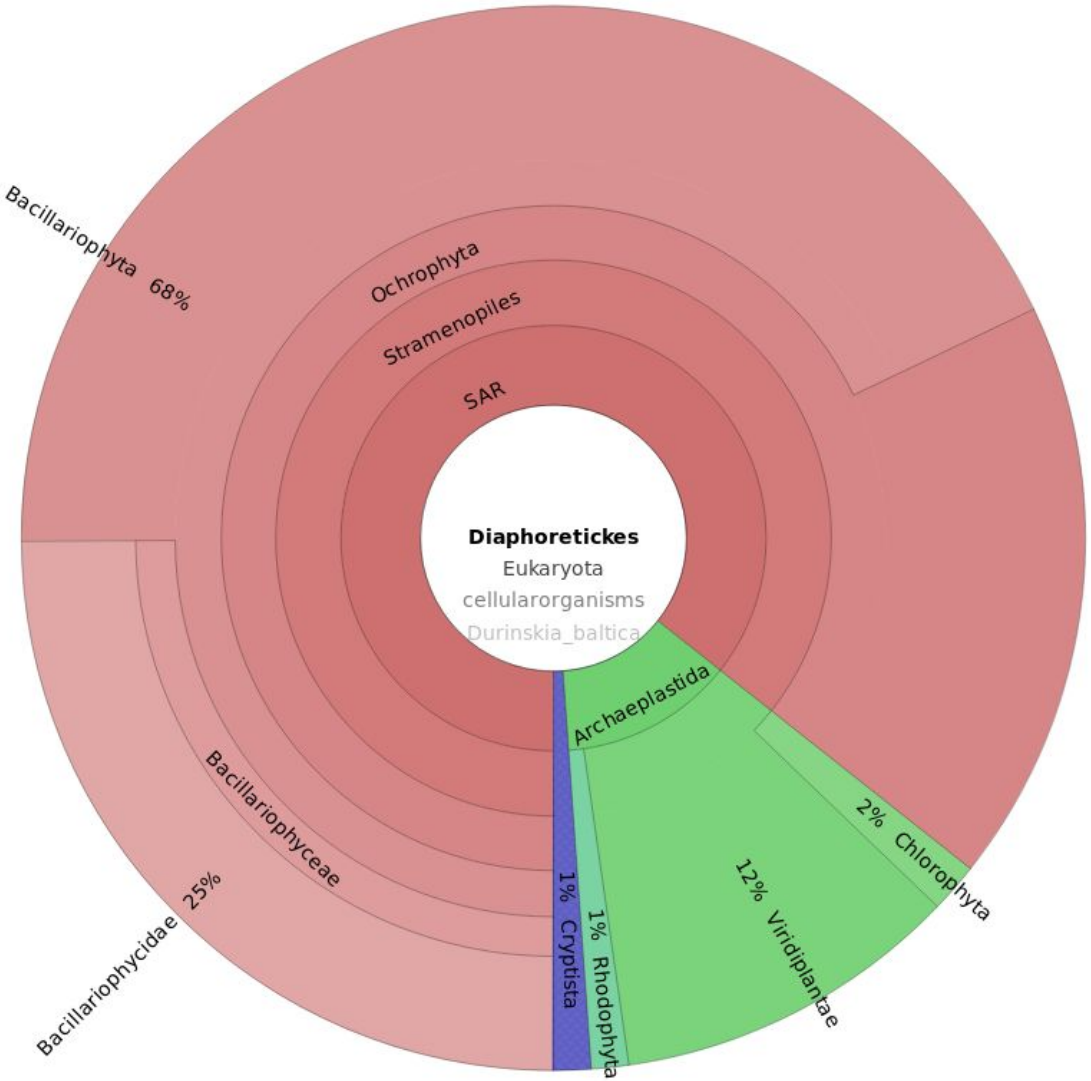
Karenia brevis



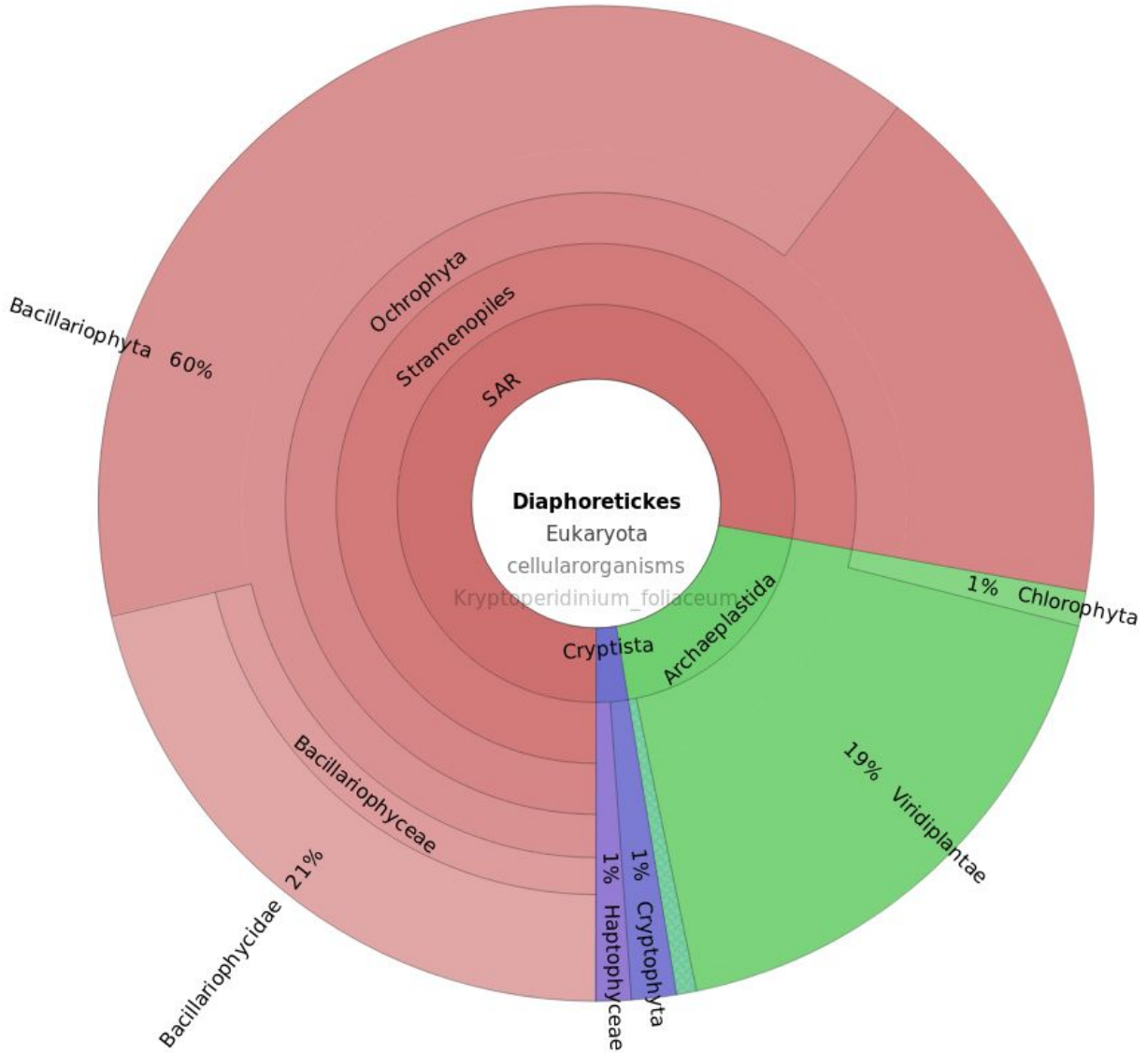
Karlodinium venificum



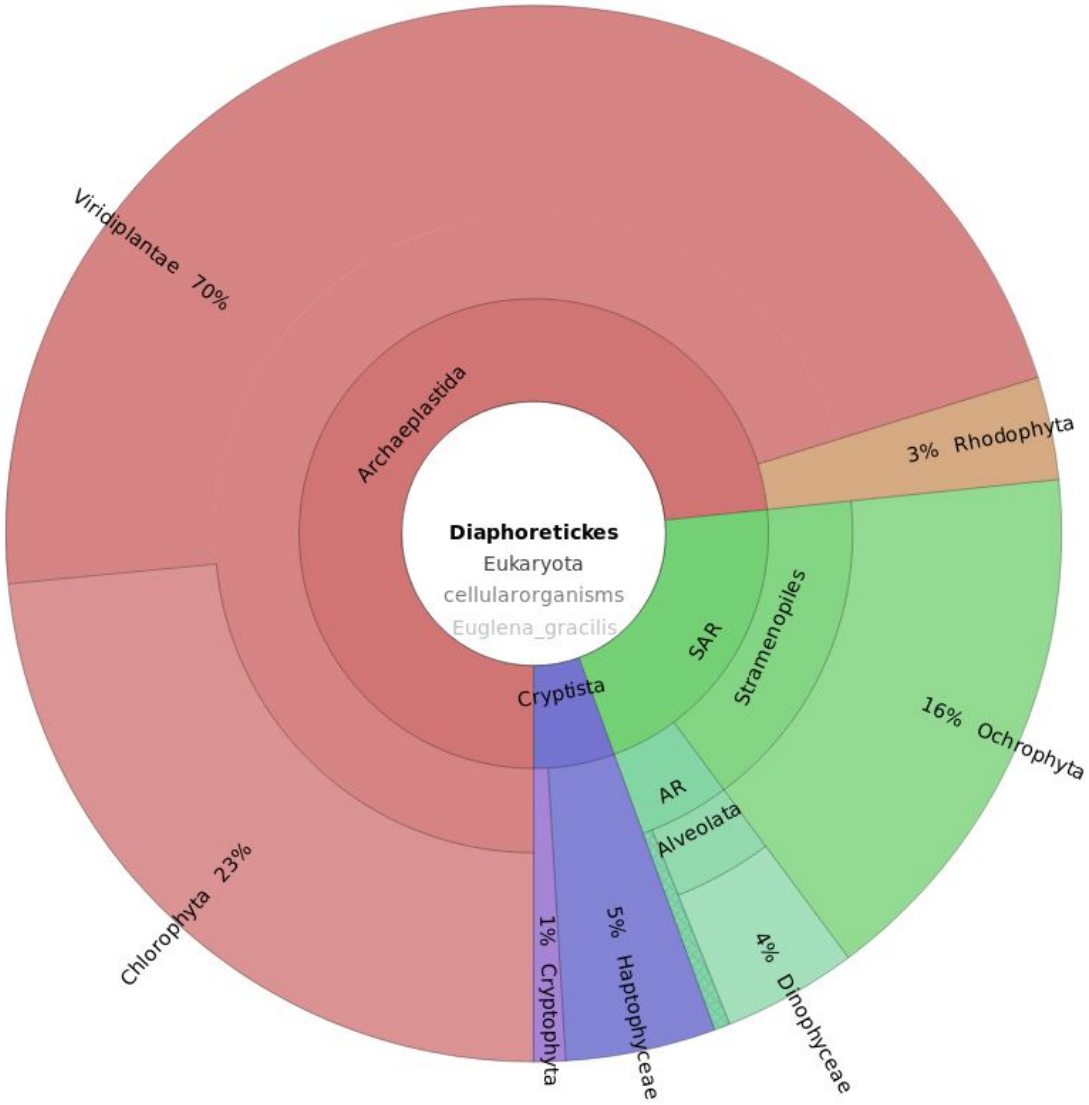
Durinskia baltica



Kryptoperidinium foliaceum



Euglena gracilis



MAKING-OF

PROTEOMES (100+)

NCBI

GenBank
RefSeq

nucleus

plastid

nucleomorph

QUALITY ASSESSMENT AND FILTERING

BUSCO v3

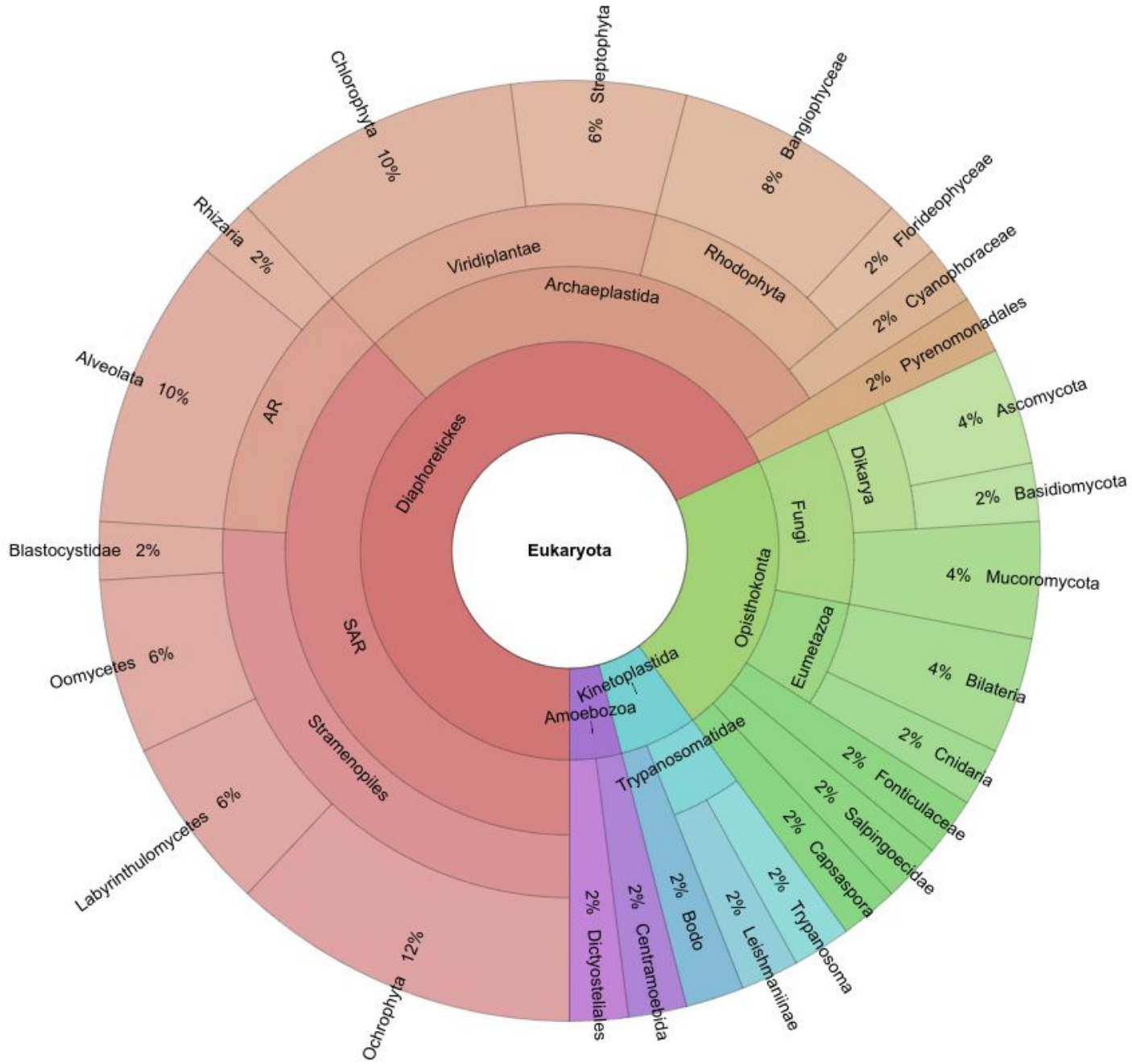
303 universal markers

78 ribosomal markers

H. Philippe, manually curated

50 contam-free and complete proteomes

CD-HIT Representative Sequences.


Eukaryota
Archaeplastida
Viridiplantae
Chlorophyta 10%
Streptophyta 6%
Rhodophyta
8% Bangiophyceae
2% Florideophyceae
2% Cyanophoraceae
2% Pyrenomonadales
Diaphoretickes
SAR
Alveolata 10%
Rhizaria 2%
AR
Stramenopiles
Ochrophyta 12%
Labyrinthulomycetes 6%
Oomycetes 6%
Blastocystidae 2%
Amoebozoa
Kinetoplastida
Opisthokonta
Fungi
Dikarya
4% Ascomycota
2% Basidiomycota
4% Mucoromycota
Eumetazoa
4% Bilateria
2% Cnidaria
2% Fonticulaceae
2% Salpingoecidae
2% Capsaspora
Trypanosomatidae
2% Trypanosoma
2% Leishmaniinae
2% Bodo
2% Centramoebida
2% Dictyostelia

TRANSCRIPTOMES (678)

Simion P et al. 2018

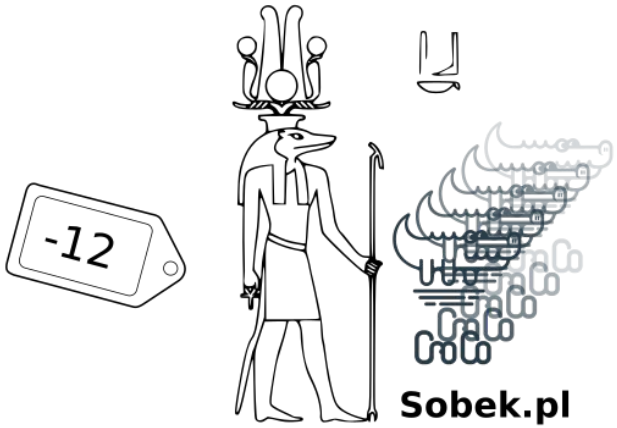
MMETSP re-assembly

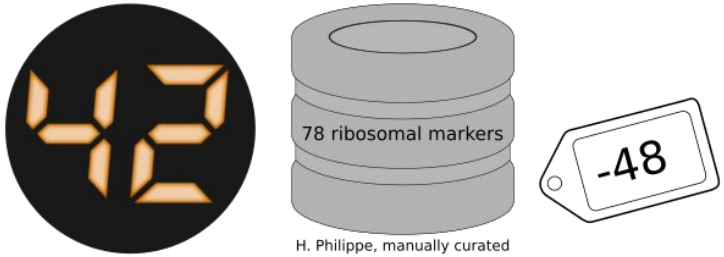
Lisa K Johnson et al. 2018

CONSOLIDATION

- samples pooled by genus (106 runs, 564 samples)
- 469 samples into 114 mega samples
*- total of 323 samples (**256** complex algae)*

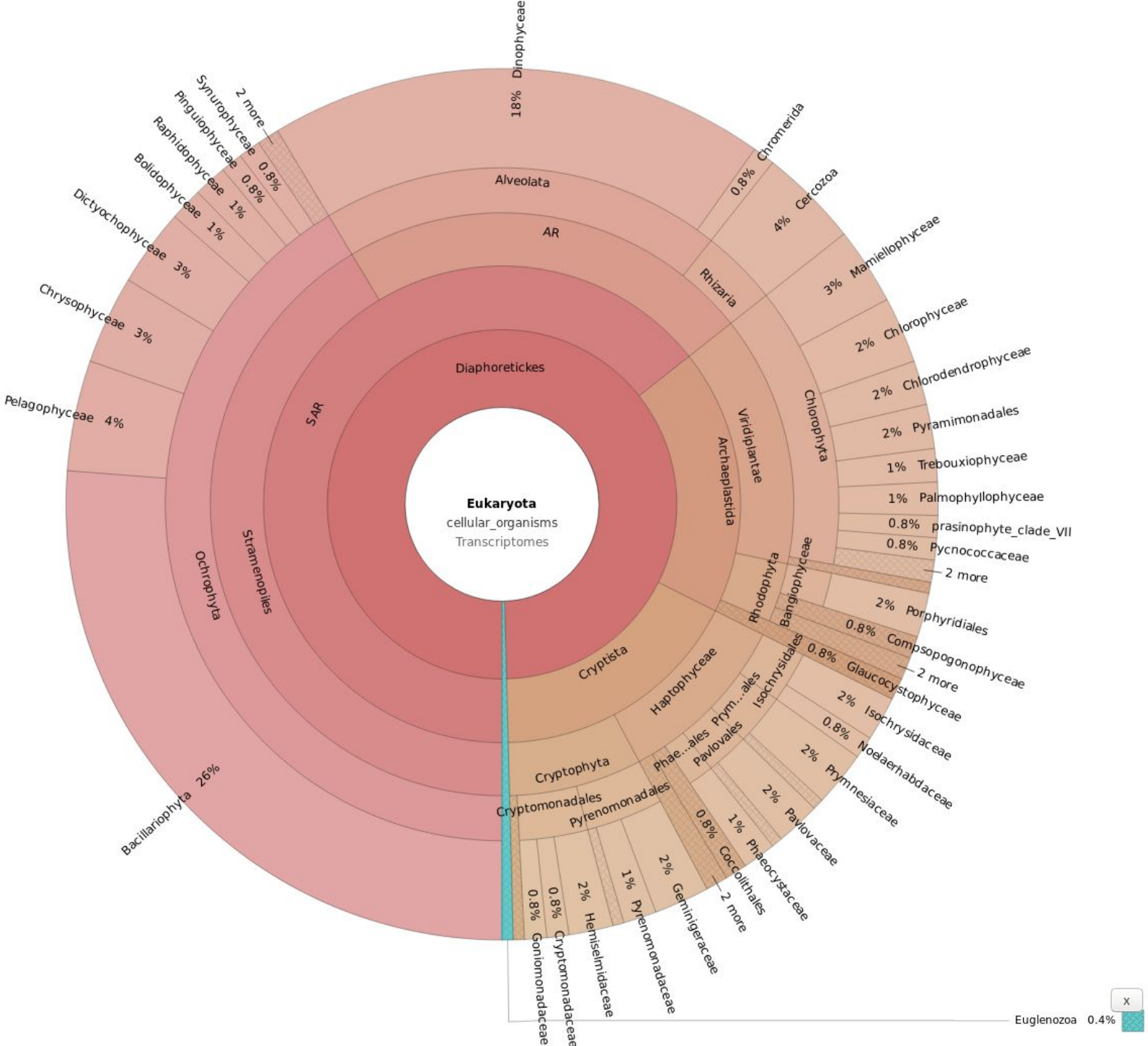
QUALITY ASSESSMENT AND FILTERING





196 complex algae transcriptomes

CD-HIT Representative Sequences.

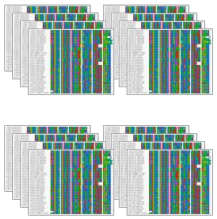

Eukaryota
cellular_organisms
Transcriptomes
Diaphoretickes
AR
SAR
Alveolata
Rhizaria
Stramenopiles
Archaeplastida
Cryptista
Viridiplantae
Ochrophyta
Chlorophyta
Rhodophyta
Cryptophyta
Bangiophyceae
Haptophyceae
Cryptomonadales
Isochrysidales
Pyrenomonadales
Pavlovales
Coccolithales
Dinophyceae 18%
Chromerida 0.8%
Cercozoa 4%
Mamiellophyceae 3%
Chlorophyceae 2%
Chlorodendrophyceae 2%
Pyramimonadales 2%
Trebouxiophyceae 1%
Palmophyllophyceae 1%
prasinophyte_clade_VII 0.8%
Pycnococcaceae 0.8%
2 more
Porphyridiales 2%
Compsopogonophyceae 0.8%
Glaucocystophyceae 0.8%
2 more
Isochrysidaceae 2%
Noelaerhabdaceae 0.8%
Prymnesiaceae 2%
Pavlovaceae 2%
Phaeocystaceae 1%
2 more
Geminigeraceae 2%
Pyrenomonadaceae 1%
Hemiselmidaceae 2%
Cryptomonadaceae 0.8%
Goniomonadaceae 0.8%
Euglenozoa 0.4%
Bacillariophyta 26%
Pelagophyceae 4%
Chrysophyceae 3%
Dictyochophyceae 3%
Bolidophyceae 1%
Raphidophyceae 0.8%
Pinguiophyceae 0.8%
Synurophyceae 0.8%
2 more

Pipeline

ORTHOLOGY

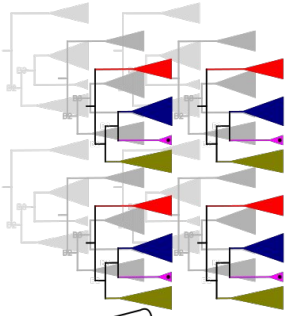
Orthology Inference

OrthoFinder



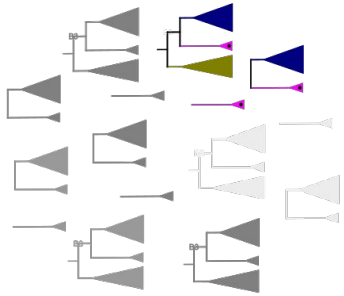
photosynthetics orthogroups

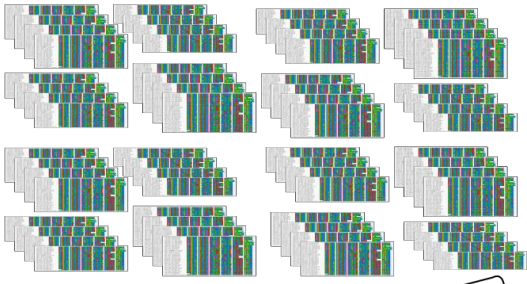
RAxML trees [fast]



10579

Tree Splitting [photosynthetics]





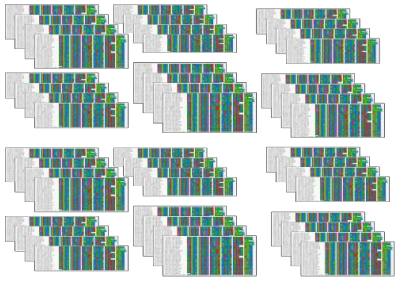
photosynthetic clans

21298

ENRICHMENT

Subtrees enrichment

196 Complex Algae de-contaminated transcriptomes [MMETSP]



enriched photosynthetic clans

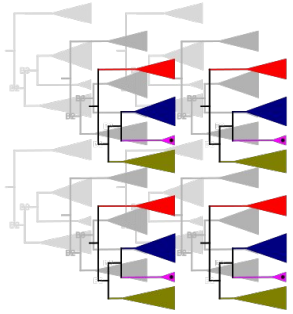
MSAs filtering

Hmmcleaner.pl
A. Di Franco et al. 2019

↓

ali2phylip.pl

IQ-trees [LG4X]

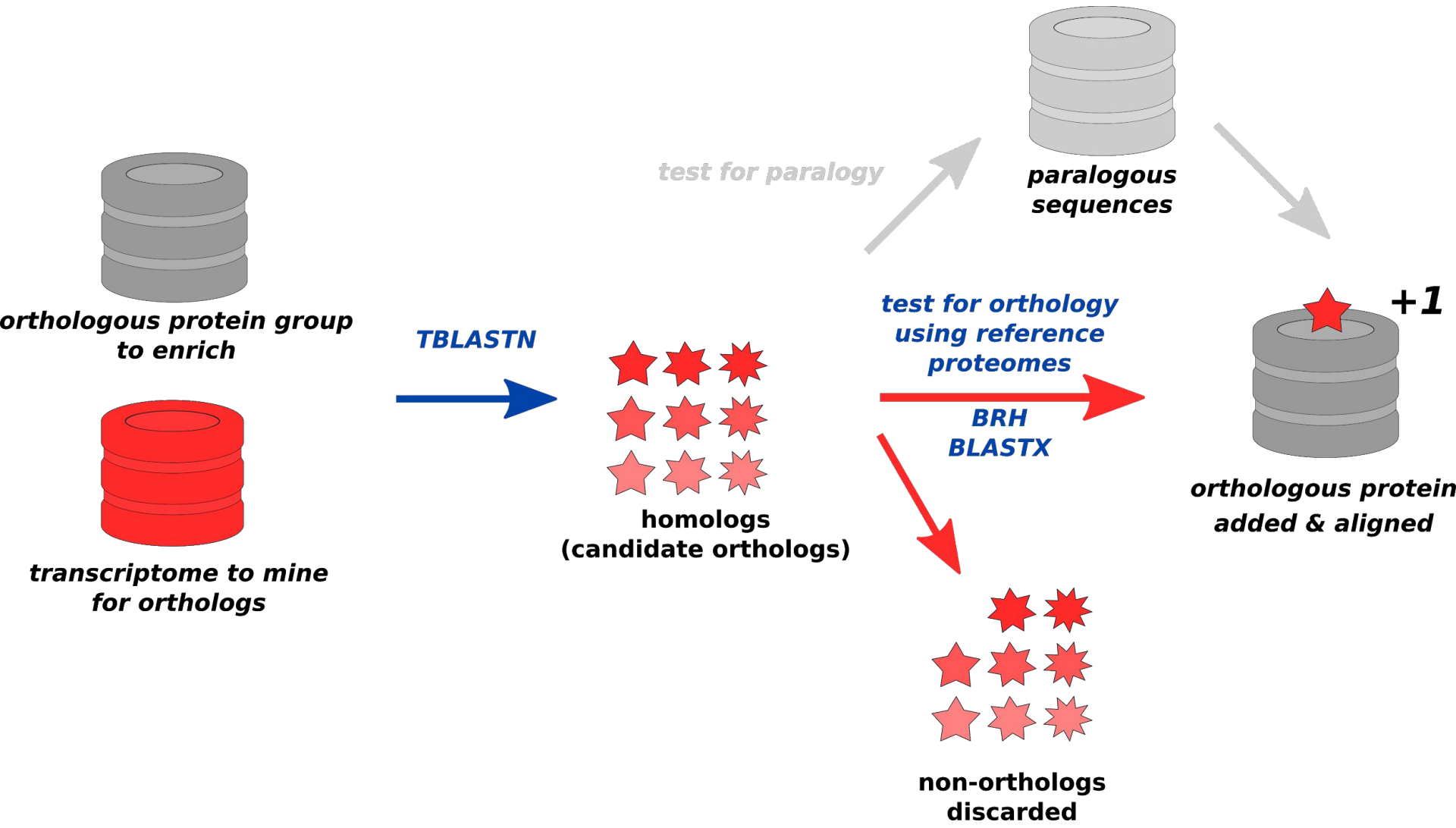


Tree Parsing

@-DIVERSITY

19509

Forty-Two

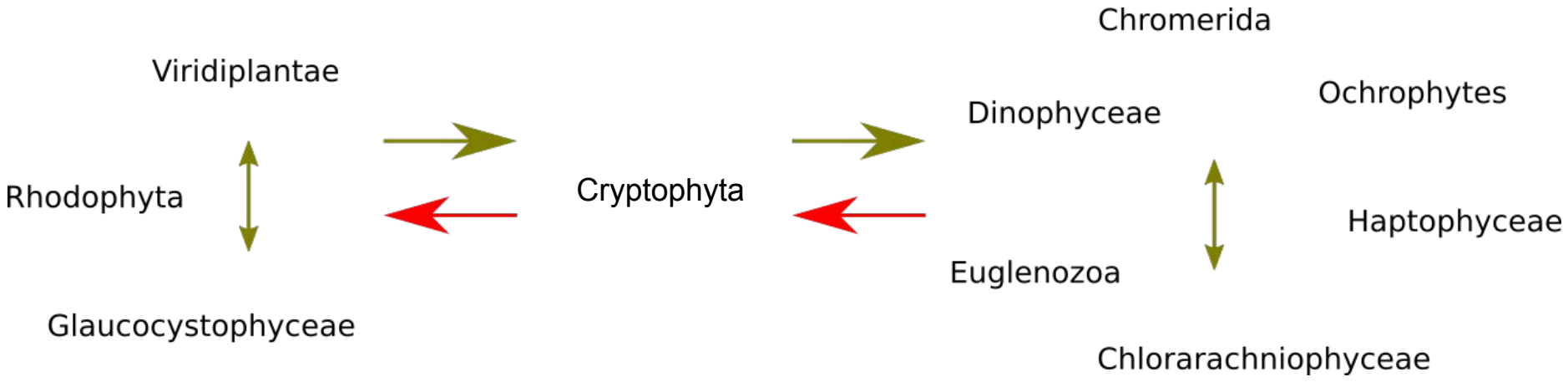

test for paralogy
paralogous sequences
orthologous protein group to enrich
TBLASTN
test for orthology using reference proteomes
BRH
BLASTX
+1
homologs (candidate orthologs)
orthologous protein added & aligned
transcriptome to mine for orthologs
non-orthologs discarded

Forty-Two - additional features

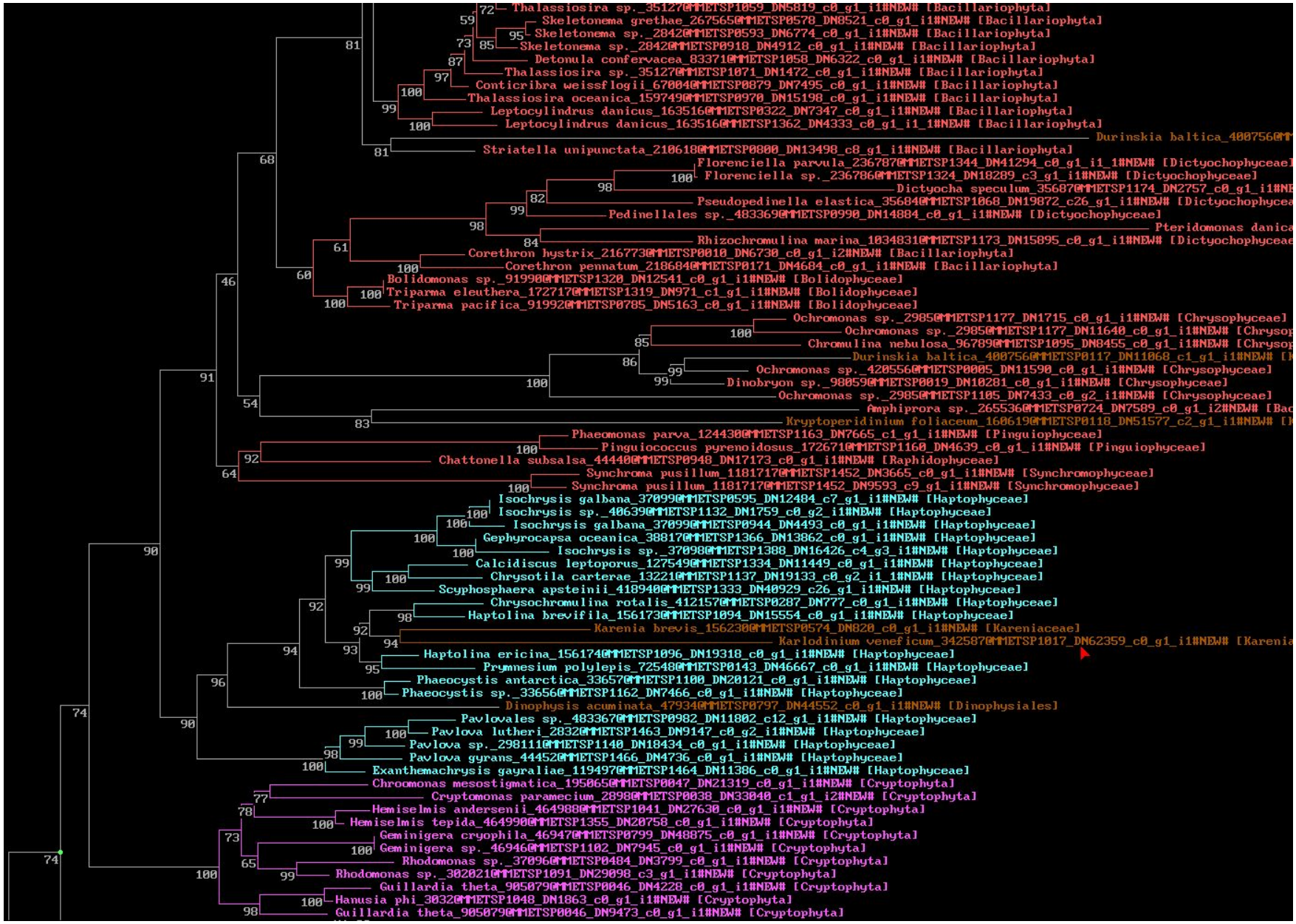
- ★ CAP3 (nt)
  - contiguing overlapping orthologuous transcripts
- ★ Alignment (BLAST/Exonerate)
  - closest relative
- ★ Redundancy check
  - merging of fragments from the same transcript
  - No chimeras
- ★ Taxonomic affiliation
  - best-hit/lca
  - contamination detection with ribosomal markers
- ★ Forty-Two publicly available @ https://metacpan.org/pod/Bio::MUST::Apps::FortyTwo

Tree parsing and inference

- ★ clan-level parsing
- ★ polarising unrooted trees
- ★ minimum species sampling for giving/receving
- ★ time constraints



Paraphyly - Intruders - Taxonomy


Thalassiosira sp._35127@MMETSP1059_DN5819_c0_g1_i1#NEW# [Bacillariophyta]
Skeletonema grethae_267565@MMETSP0578_DN8521_c0_g1_i1#NEW# [Bacillariophyta]
Skeletonema sp._2842@MMETSP0593_DN6774_c0_g1_i1#NEW# [Bacillariophyta]
Skeletonema sp._2842@MMETSP0918_DN4912_c0_g1_i1#NEW# [Bacillariophyta]
Detonula confervacea_83371@MMETSP1058_DN6322_c0_g1_i1#NEW# [Bacillariophyta]
Thalassiosira sp._35127@MMETSP1071_DN1472_c0_g1_i1#NEW# [Bacillariophyta]
Conticribra weissflogii_67004@MMETSP0879_DN7495_c0_g1_i1#NEW# [Bacillariophyta]
Thalassiosira oceanica_159749@MMETSP0970_DN15198_c0_g1_i1#NEW# [Bacillariophyta]
Leptocylindrus danicus_163516@MMETSP0322_DN7347_c0_g1_i1#NEW# [Bacillariophyta]
Leptocylindrus danicus_163516@MMETSP1362_DN4333_c0_g1_i1_1#NEW# [Bacillariophyta]
Durinskia baltica_400756@M
Striatella unipunctata_210618@MMETSP0800_DN13498_c8_g1_i1#NEW# [Bacillariophyta]
Florenciella parvula_236787@MMETSP1344_DN41294_c0_g1_i1_1#NEW# [Dictyochophyceae]
Florenciella sp._236786@MMETSP1324_DN18289_c3_g1_i1#NEW# [Dictyochophyceae]
Dictyocha speculum_35687@MMETSP1174_DN2757_c0_g1_i1#NE
Pseudopedinella elastica_35684@MMETSP1068_DN19872_c26_g1_i1#NEW# [Dictyochophycea
Pedinellales sp._483369@MMETSP0990_DN14884_c0_g1_i1#NEW# [Dictyochophyceae]
Pteridomonas danica
Rhizochromulina marina_10348310@MMETSP1173_DN15895_c0_g1_i1#NEW# [Dictyochophyceae
Corethron hystrix_216773@MMETSP0010_DN6730_c0_g1_i2#NEW# [Bacillariophyta]
Corethron pennatum_218684@MMETSP0171_DN4684_c0_g1_i1#NEW# [Bacillariophyta]
Bolidomonas sp._91990@MMETSP1320_DN12541_c0_g1_i1#NEW# [Bolidophyceae]
Triparma eleuthera_172717@MMETSP1319_DN971_c1_g1_i1#NEW# [Bolidophyceae]
Triparma pacifica_91992@MMETSP0785_DN5163_c0_g1_i1#NEW# [Bolidophyceae]
Ochromonas sp._2985@MMETSP1177_DN1715_c0_g1_i1#NEW# [Chrysophyceae]
Ochromonas sp._2985@MMETSP1177_DN11640_c0_g1_i1#NEW# [Chryso
Chromulina nebulosa_96789@MMETSP1095_DN8455_c0_g1_i1#NEW# [Chryso
Durinskia baltica_400756@MMETSP0117_DN11068_c1_g1_i1#NEW# [
Ochromonas sp._420556@MMETSP0005_DN11590_c0_g1_i1#NEW# [Chrysophyceae]
Dinobryon sp._98059@MMETSP0019_DN10281_c0_g1_i1#NEW# [Chrysophyceae]
Ochromonas sp._2985@MMETSP1105_DN7433_c0_g2_i1#NEW# [Chrysophyceae]
Amphiprora sp._265536@MMETSP0724_DN7589_c0_g1_i2#NEW# [Bac
Kryptoperidinium foliaceum_160619@MMETSP0118_DN51577_c2_g1_i1#NEW# [
Phaeomonas parva_124430@MMETSP1163_DN7665_c1_g1_i1#NEW# [Pinguiophyceae]
Pinguiococcus pyrenoidosus_172671@MMETSP1160_DN4639_c0_g1_i1#NEW# [Pinguiophyceae]
Chattonella subsalsa_44440@MMETSP0948_DN17173_c0_g1_i1#NEW# [Raphidophyceae]
Synchroma pusillum_1181717@MMETSP1452_DN3665_c0_g1_i1#NEW# [Synchromophyceae]
Synchroma pusillum_1181717@MMETSP1452_DN9593_c9_g1_i1#NEW# [Synchromophyceae]
Isochrysis galbana_37099@MMETSP0595_DN12484_c7_g1_i1#NEW# [Haptophyceae]
Isochrysis sp._40639@MMETSP1132_DN1759_c0_g2_i1#NEW# [Haptophyceae]
Isochrysis galbana_37099@MMETSP0944_DN4493_c0_g1_i1#NEW# [Haptophyceae]
Gephyrocapsa oceanica_38817@MMETSP1366_DN13862_c0_g1_i1#NEW# [Haptophyceae]
Isochrysis sp._37098@MMETSP1388_DN16426_c4_g3_i1#NEW# [Haptophyceae]
Calcidiscus leptoporus_127549@MMETSP1334_DN11449_c0_g1_i1#NEW# [Haptophyceae]
Chrysotila carterae_13221@MMETSP1137_DN19133_c0_g2_i1_1#NEW# [Haptophyceae]
Scyphosphaera apsteinii_418940@MMETSP1333_DN40929_c26_g1_i1#NEW# [Haptophyceae]
Chrysochromulina rotalis_412157@MMETSP0287_DN777_c0_g1_i1#NEW# [Haptophyceae]
Haptolina brevifila_156173@MMETSP1094_DN15554_c0_g1_i1#NEW# [Haptophyceae]
Karenia brevis_156230@MMETSP0574_DN820_c0_g1_i1#NEW# [Kareniaceae]
Karlodinium veneficum_342587@MMETSP1017_DN62359_c0_g1_i1#NEW# [Karenia
Haptolina ericina_156174@MMETSP1096_DN19318_c0_g1_i1#NEW# [Haptophyceae]
Prymnesium polylepis_72548@MMETSP0143_DN46667_c0_g1_i1#NEW# [Haptophyceae]
Phaeocystis antarctica_33657@MMETSP1100_DN20121_c0_g1_i1#NEW# [Haptophyceae]
Phaeocystis sp._33656@MMETSP1162_DN7466_c0_g1_i1#NEW# [Haptophyceae]
Dinophysis acuminata_47934@MMETSP0797_DN44552_c0_g1_i1#NEW# [Dinophysiales]
Pavlovales sp._483367@MMETSP0982_DN11802_c12_g1_i1#NEW# [Haptophyceae]
Pavlova lutheri_2832@MMETSP1463_DN9147_c0_g2_i1#NEW# [Haptophyceae]
Pavlova sp._298111@MMETSP1140_DN18434_c0_g1_i1#NEW# [Haptophyceae]
Pavlova gyrans_44452@MMETSP1466_DN4736_c0_g1_i1#NEW# [Haptophyceae]
Exanthemachrysis gayraliae_119497@MMETSP1464_DN11386_c0_g1_i1#NEW# [Haptophyceae]
Chroomonas mesostigmatica_195065@MMETSP0047_DN21319_c0_g1_i1#NEW# [Cryptophyta]
Cryptomonas paramecium_2898@MMETSP0038_DN33040_c1_g1_i2#NEW# [Cryptophyta]
Hemiselmis andersenii_464988@MMETSP1041_DN27630_c0_g1_i1#NEW# [Cryptophyta]
Hemiselmis tepida_464990@MMETSP1355_DN20758_c0_g1_i1#NEW# [Cryptophyta]
Geminigera cryophila_46947@MMETSP0799_DN48875_c0_g1_i1#NEW# [Cryptophyta]
Geminigera sp._46946@MMETSP1102_DN7945_c0_g1_i1#NEW# [Cryptophyta]
Rhodomonas sp._37096@MMETSP0484_DN3799_c0_g1_i1#NEW# [Cryptophyta]
Rhodomonas sp._302021@MMETSP1091_DN29098_c3_g1_i1#NEW# [Cryptophyta]
Guillardia theta_905079@MMETSP0046_DN4228_c0_g1_i1#NEW# [Cryptophyta]
Hanusia phi_3032@MMETSP1048_DN1863_c0_g1_i1#NEW# [Cryptophyta]
Guillardia theta_905079@MMETSP0046_DN9473_c0_g1_i1#NEW# [Cryptophyta]

Take-home message and improvements

- ★ Data quality
  - conscientious contamination handling
  - completeness
- ★ Subcellular localisation tagging
- ★ Proficient orthogroups enrichment
  - out-paralogy handling
- ★ Sophisticated tree splitting and parsing
  - taxonomy aware
  - paraphyly detection
  - intruders tolerance
- ★ Biological insight
  - nm bearing algae
  - *Dinophysis*

- ★ sampling++?
- ★ 

- ★ accurate & broad scale plastid dating
- ★ adjacent group inference ?
- ★ other measures ?

Denis Baurain

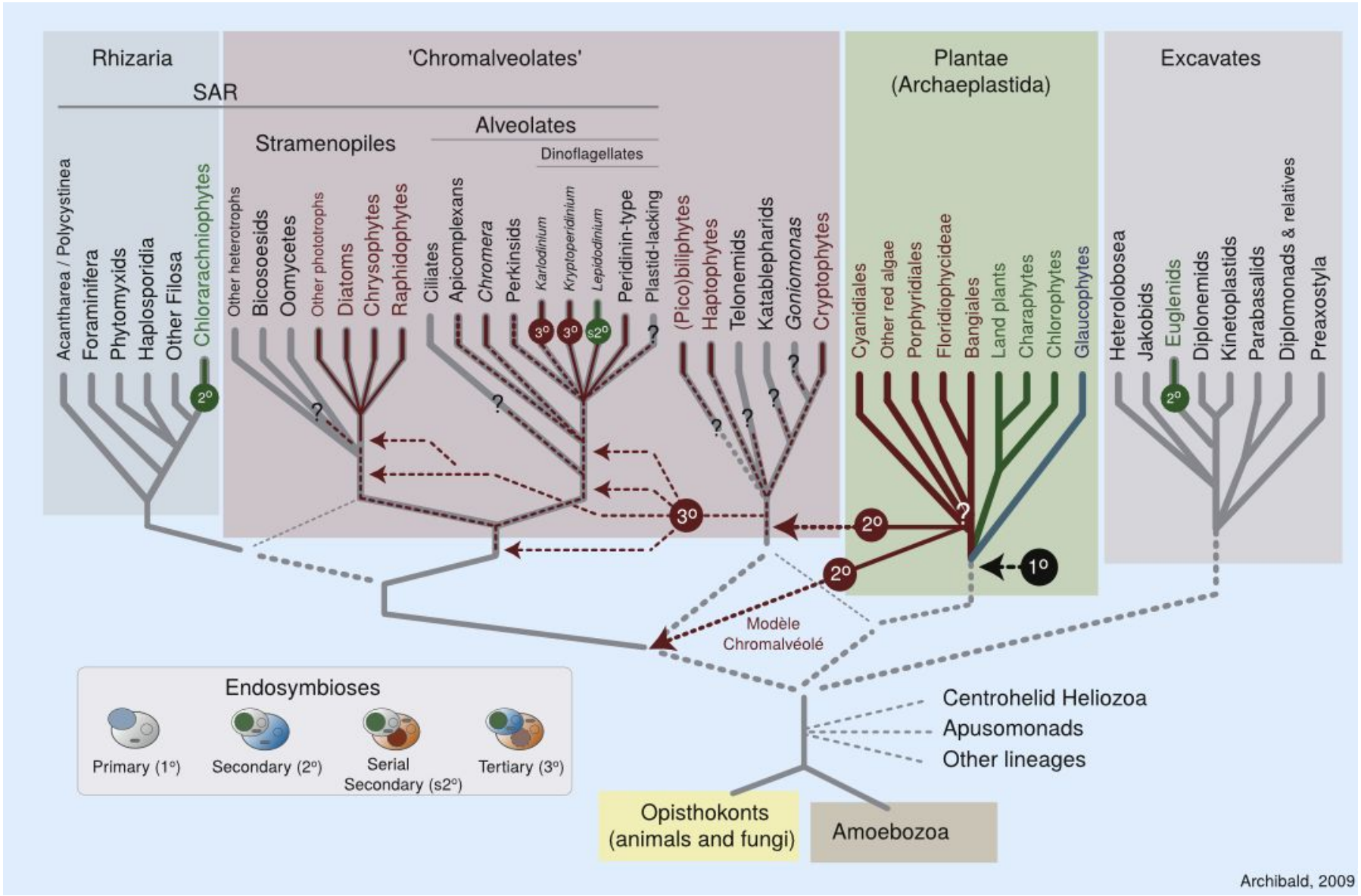
Durandal

Damien Sirjacobs

LIÈGE université

fnrs
LA LIBERTÉ DE CHERCHER

Complex Algae - Panorama



Why Kleptoplasty ?

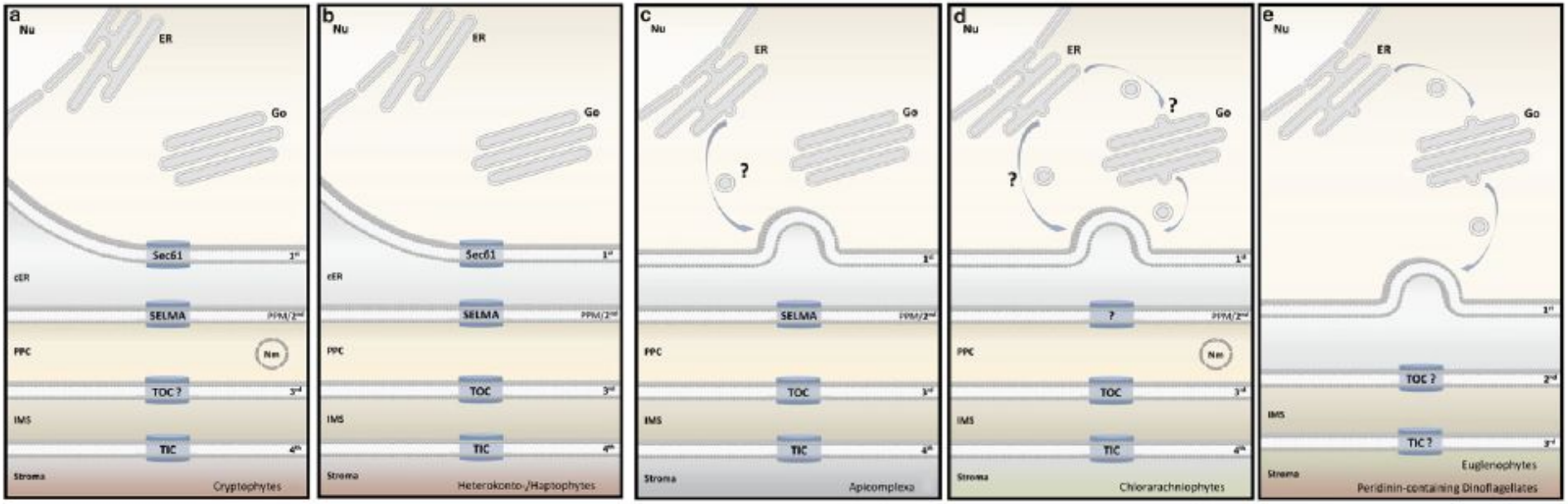
Cryptophyta
- 4 membranes
- nucleomorph

Haptophyta
Ochrophyta
- 4 membranes

Apicomplexa
Chromerida
- 4 membranes

Chlorarachniophyta
- 4 membranes
- nucleomorph

Euglena
Dinophyceae
- 3 membranes
- nucleomorph



Guillardia theta

Vaucheria litorea
Emiliania huxleyi

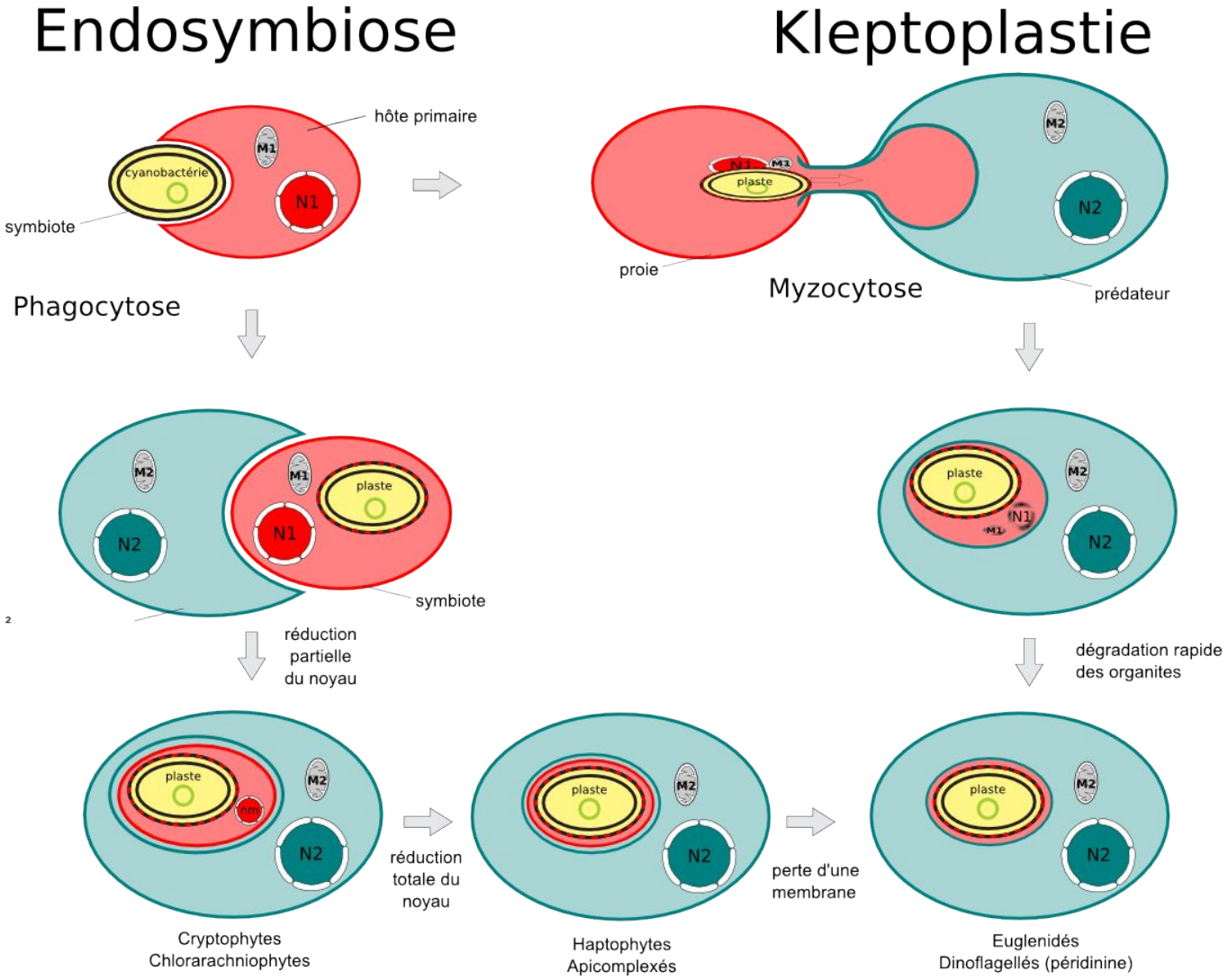
Plasmodium falciparum

Bigelowiella natans

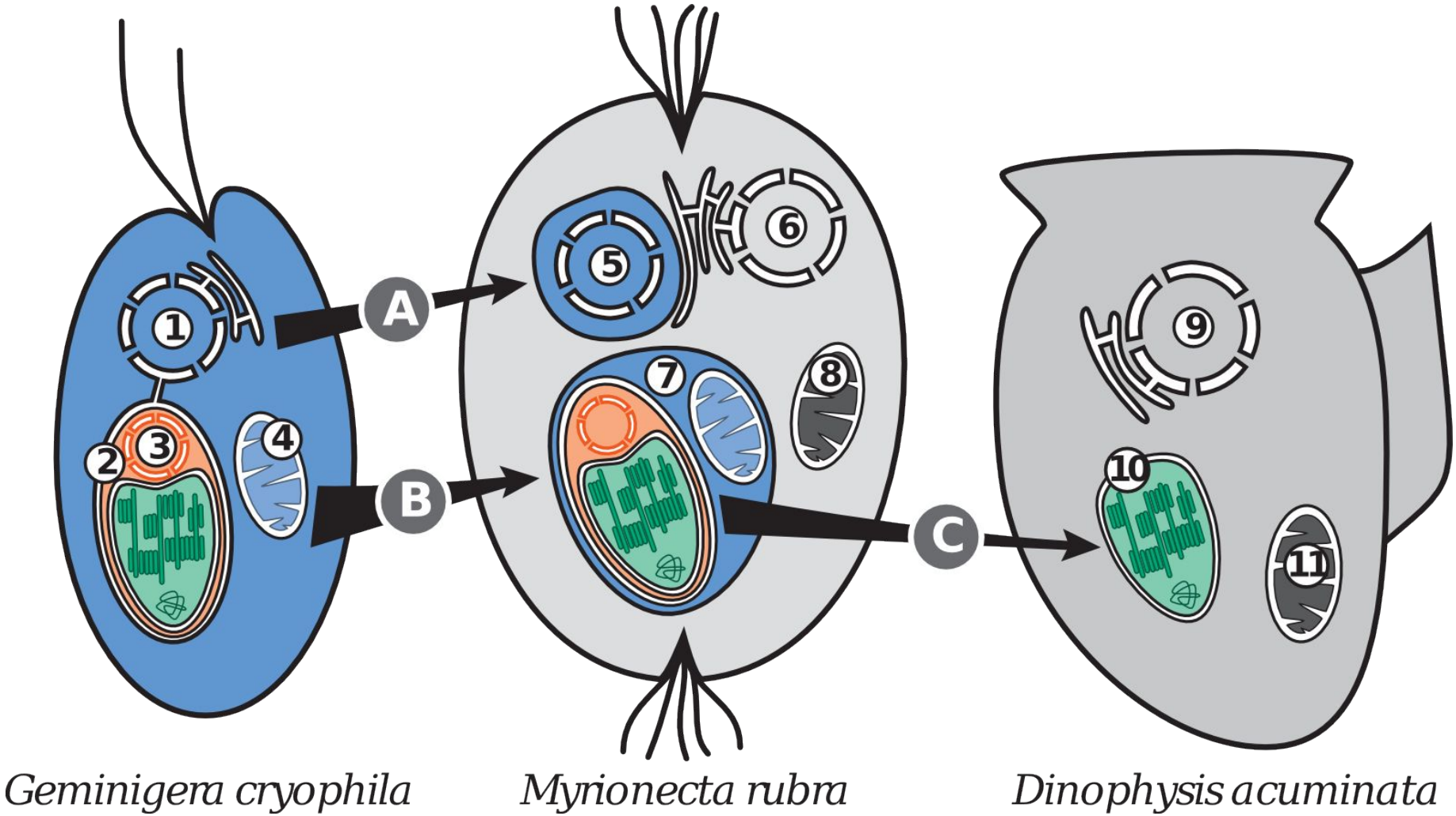
Euglena gracilis
Symbiodinium minutum

Grosche et al., 2014

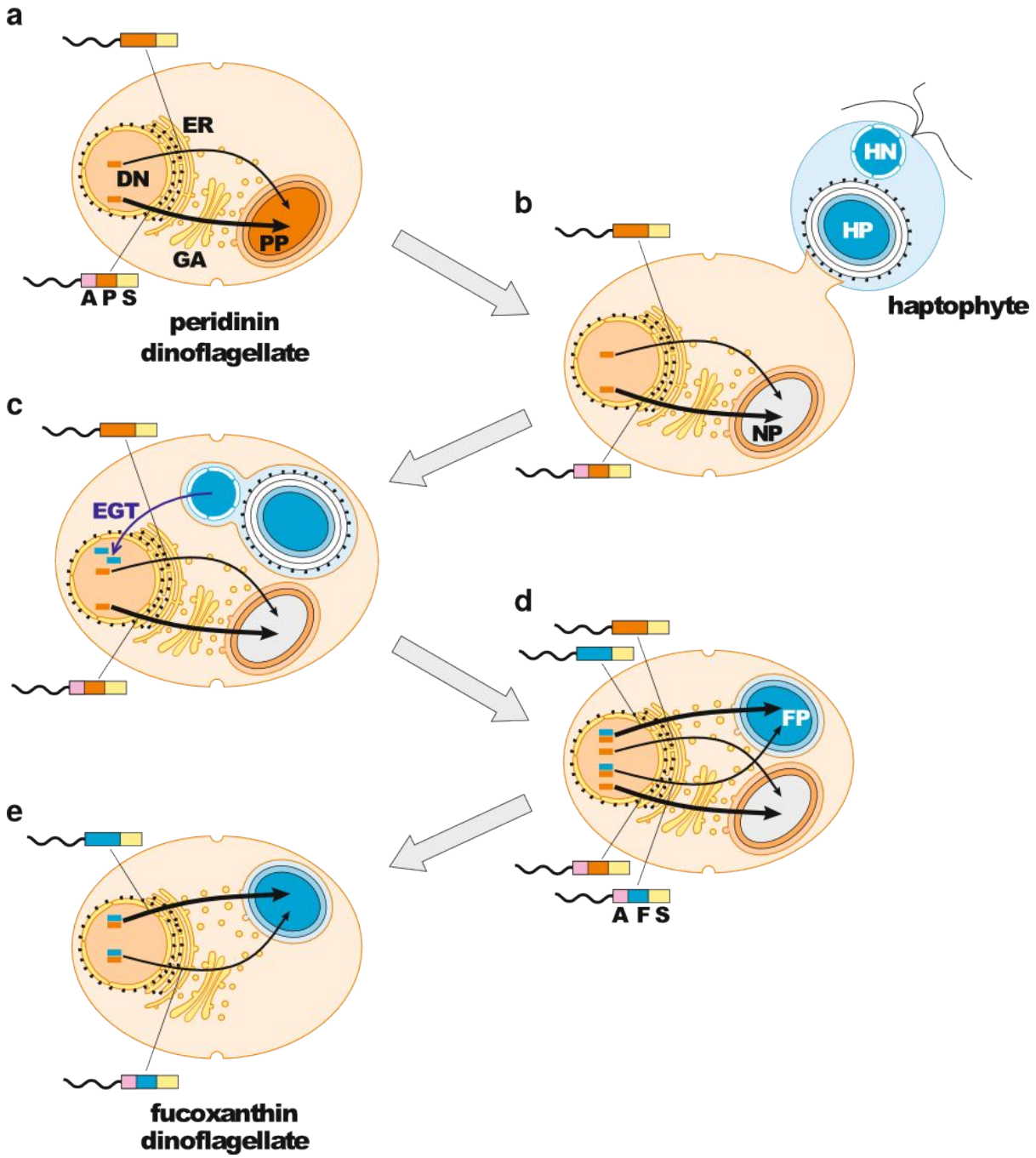
1 - Mechanisms



Actual Kleptoplastids



Wisecarver et al., 2010


a
ER
DN
PP
GA
A P S
peridinin
dinoflagellate
b
HN
HP
haptophyte
NP
c
EGT
d
FP
A F S
e
fucoxanthin
dinoflagellate