Underestimated genomic footprint of kleptoplasty in complex algae evolution?

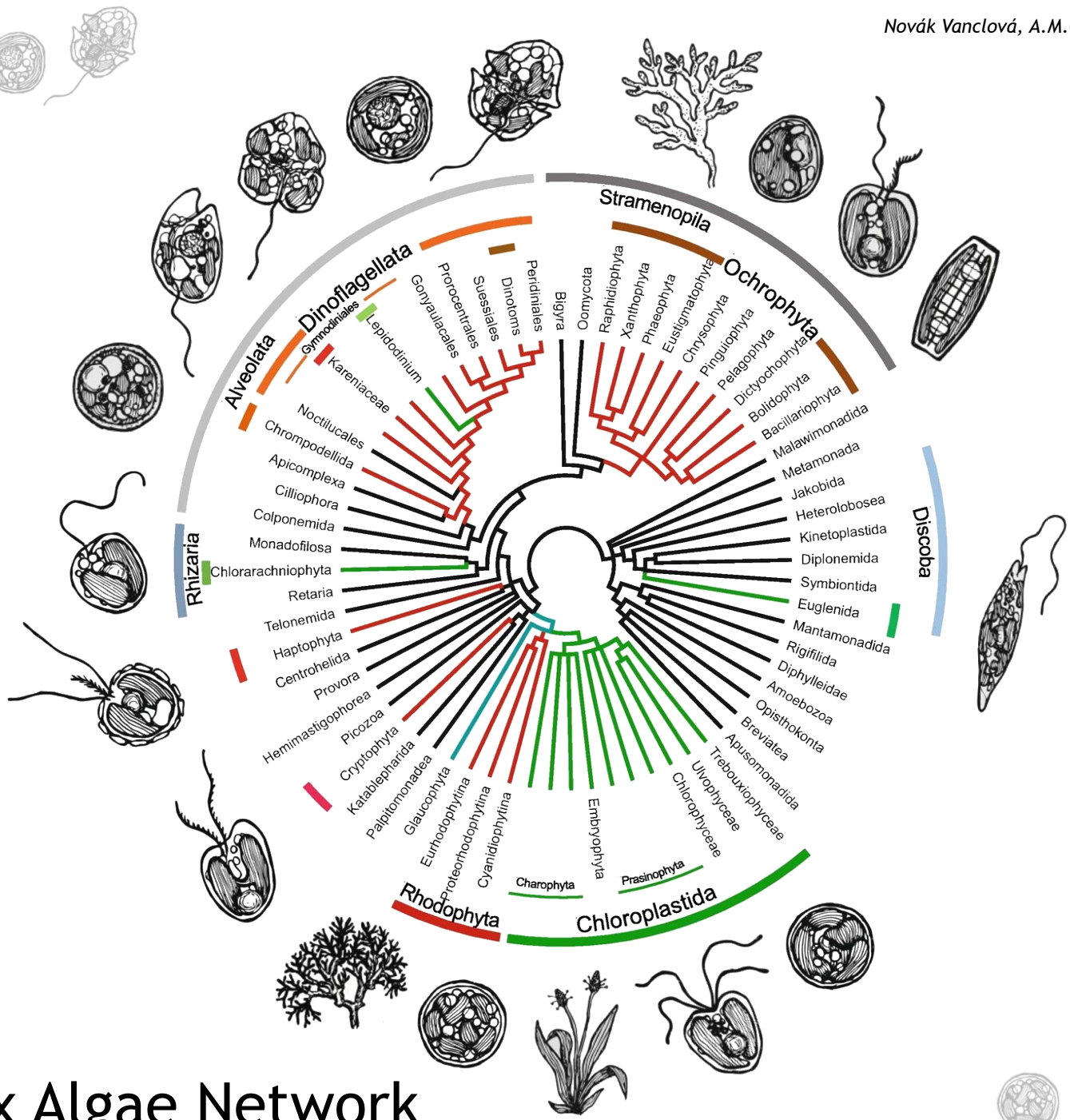
Mick Van Vlierberghe (mvanvlierberghe@doct.uliege.be) and Denis Baurain (denis.baurain@uliege.be)

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

ISE Symposium 2024
10-12 September 2024
Ruhr University Bochum

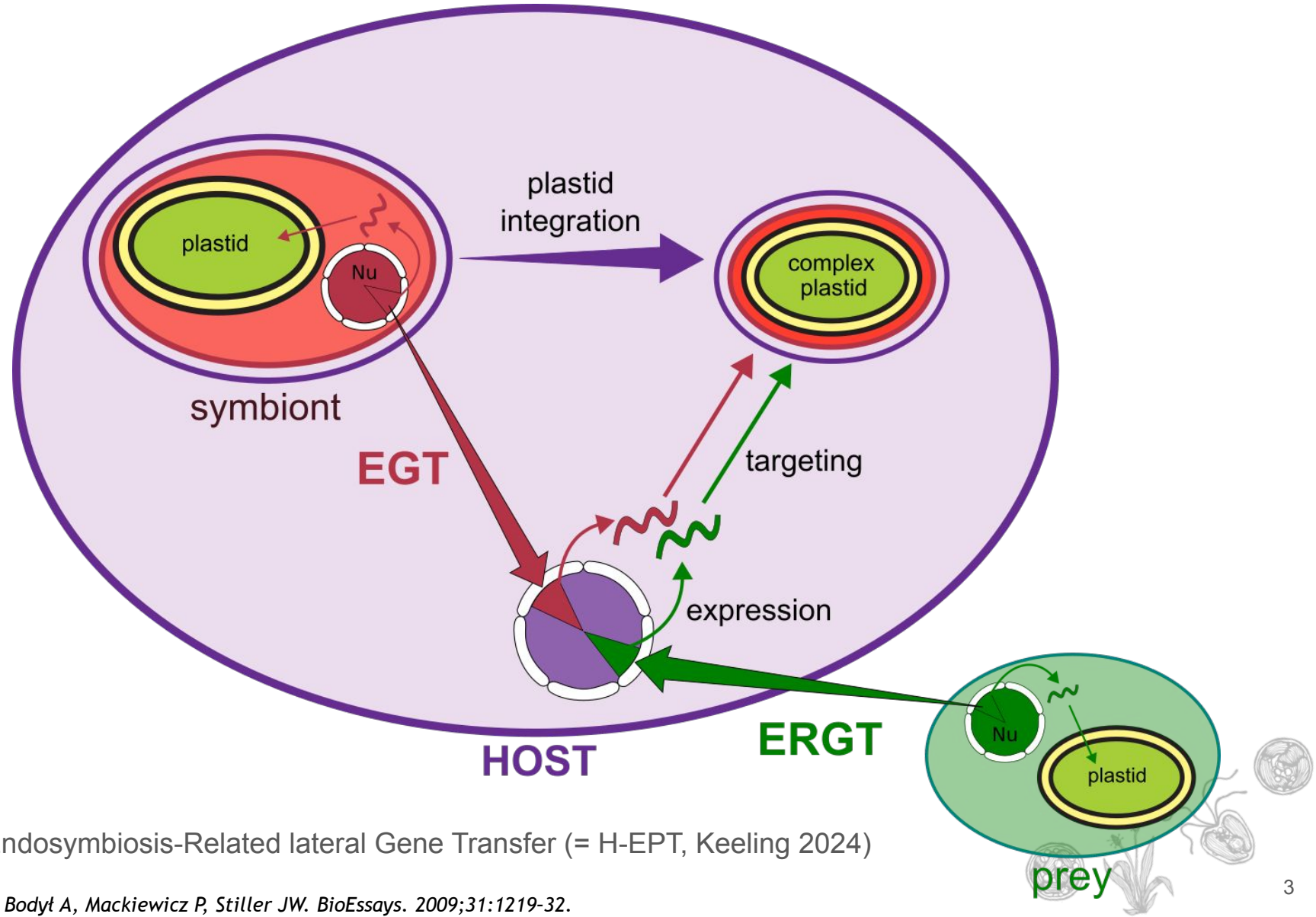
Novák Vanclová, A.M.G., Dorrell, R.G. (2024)

Novák Vanclová, A.M.G., Dorrell, R.G. (2024)



Complex Algae Network

Nucleus-encoded and plastid-targeted genes



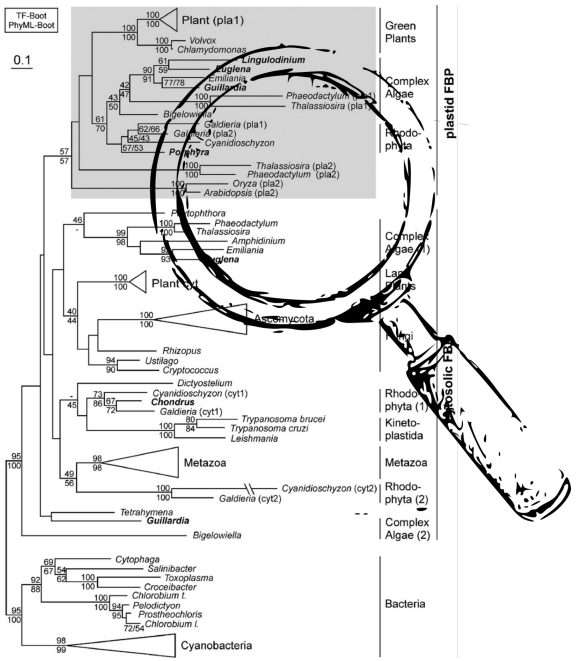
ERGT = Endosymbiosis-Related lateral Gene Transfer (= H-EPT, Keeling 2024)

Adapted from Bodył A, Mackiewicz P, Stiller JW. BioEssays. 2009;31:1219-32.

Gene transfer analysis strategies

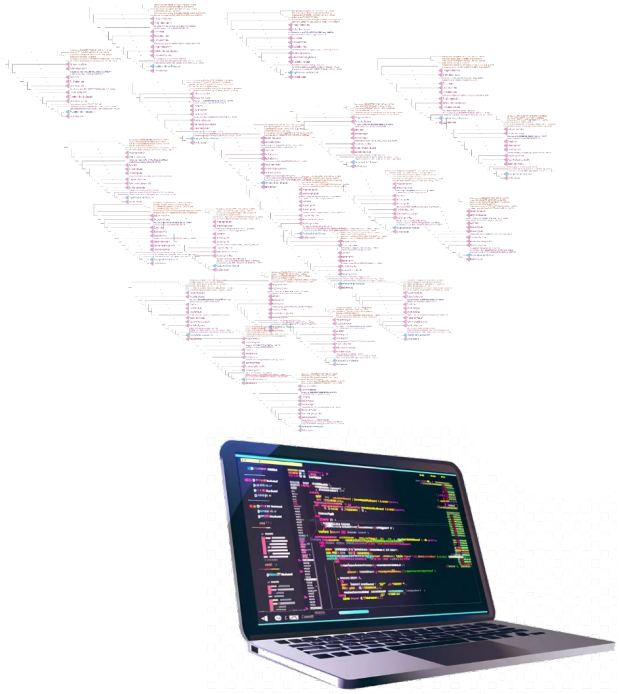
SINGLE-GENE

- scrutiny
- one/few gene(s) story
  - e.g. GAPDH, FBA, SBP/FBP



LARGE SCALE

- numerous trees
- sister relationship/not true GT



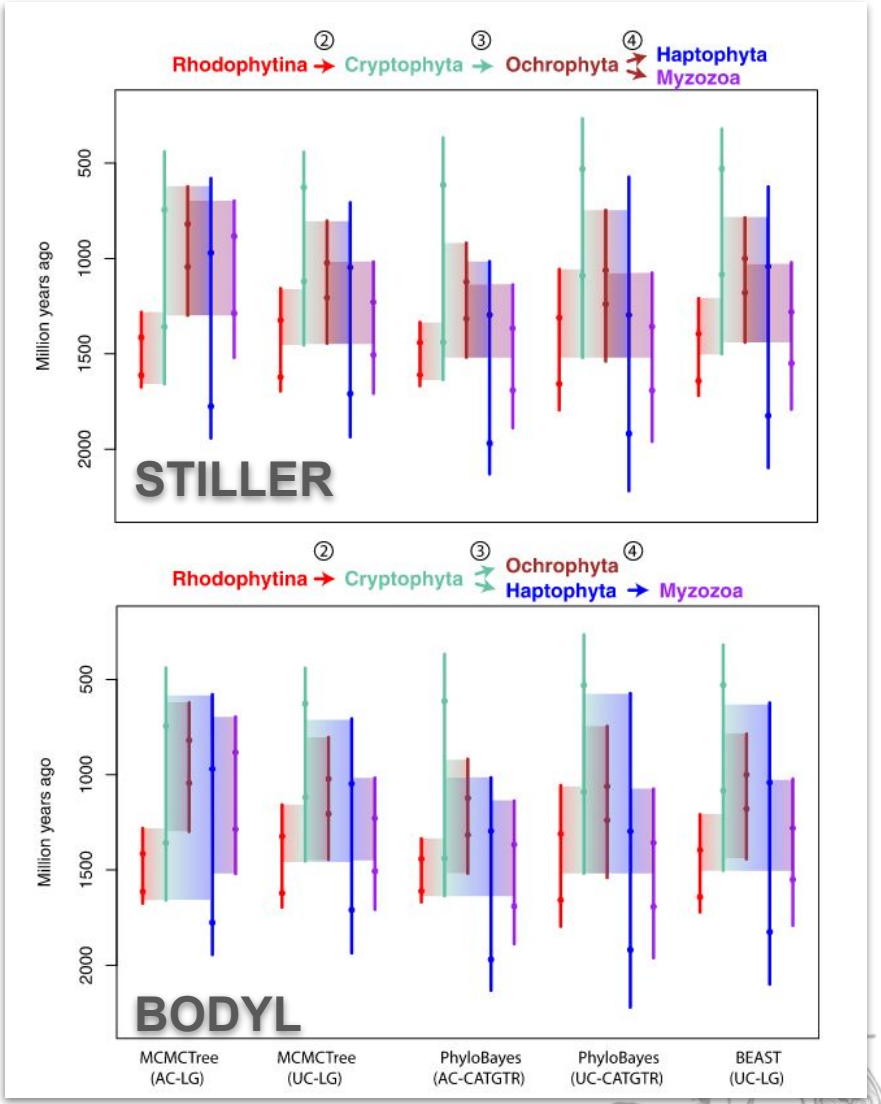
Combined approach for GT inference

 # Program

- tree parser
- true gene transfers
  - paraphyly detection

 # Core feature: age class

- molecular clock →
- avoid impossible transfers
- intruder handling
- nested transfers

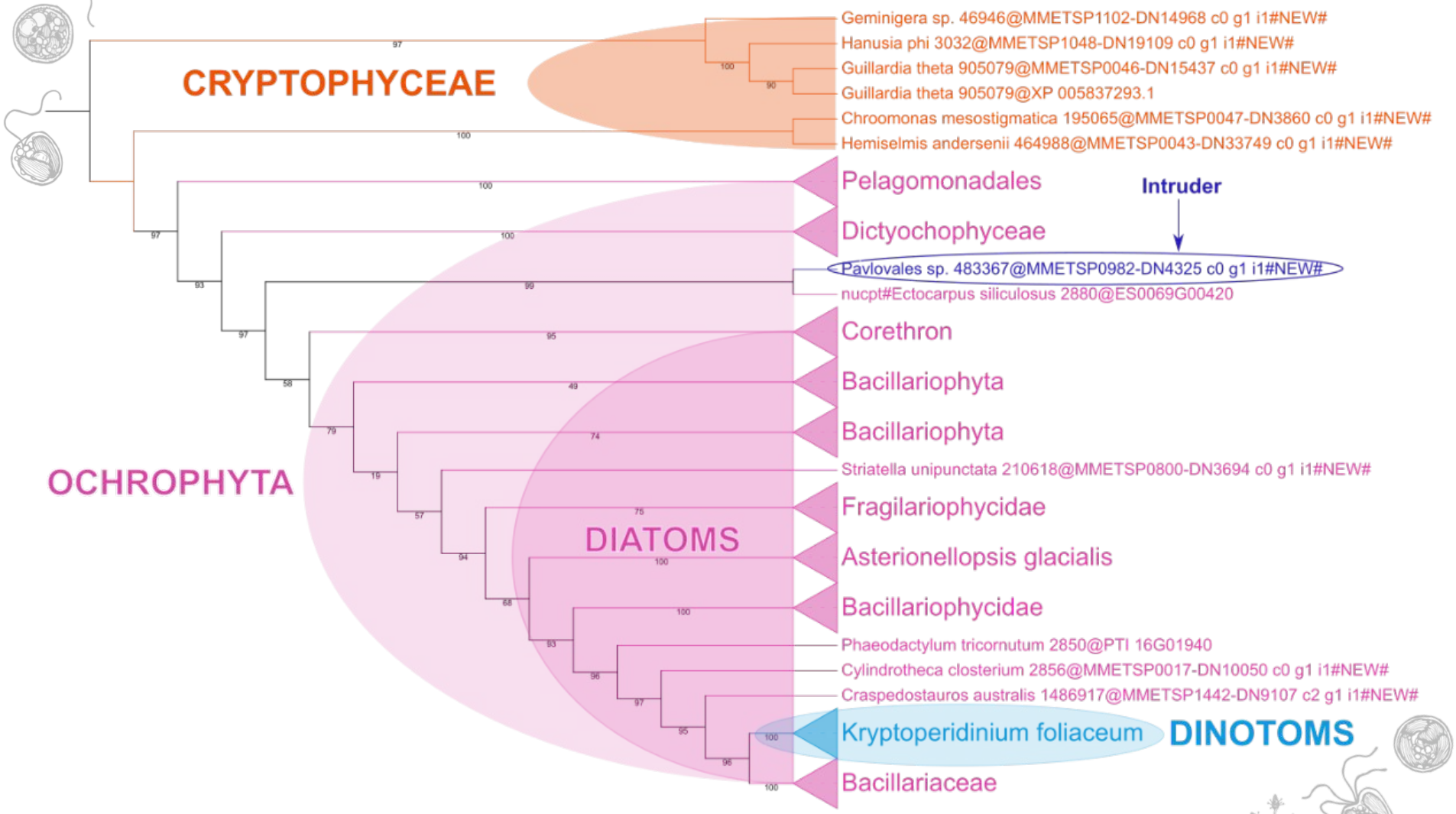


RHODOPLEX-like scenarios →

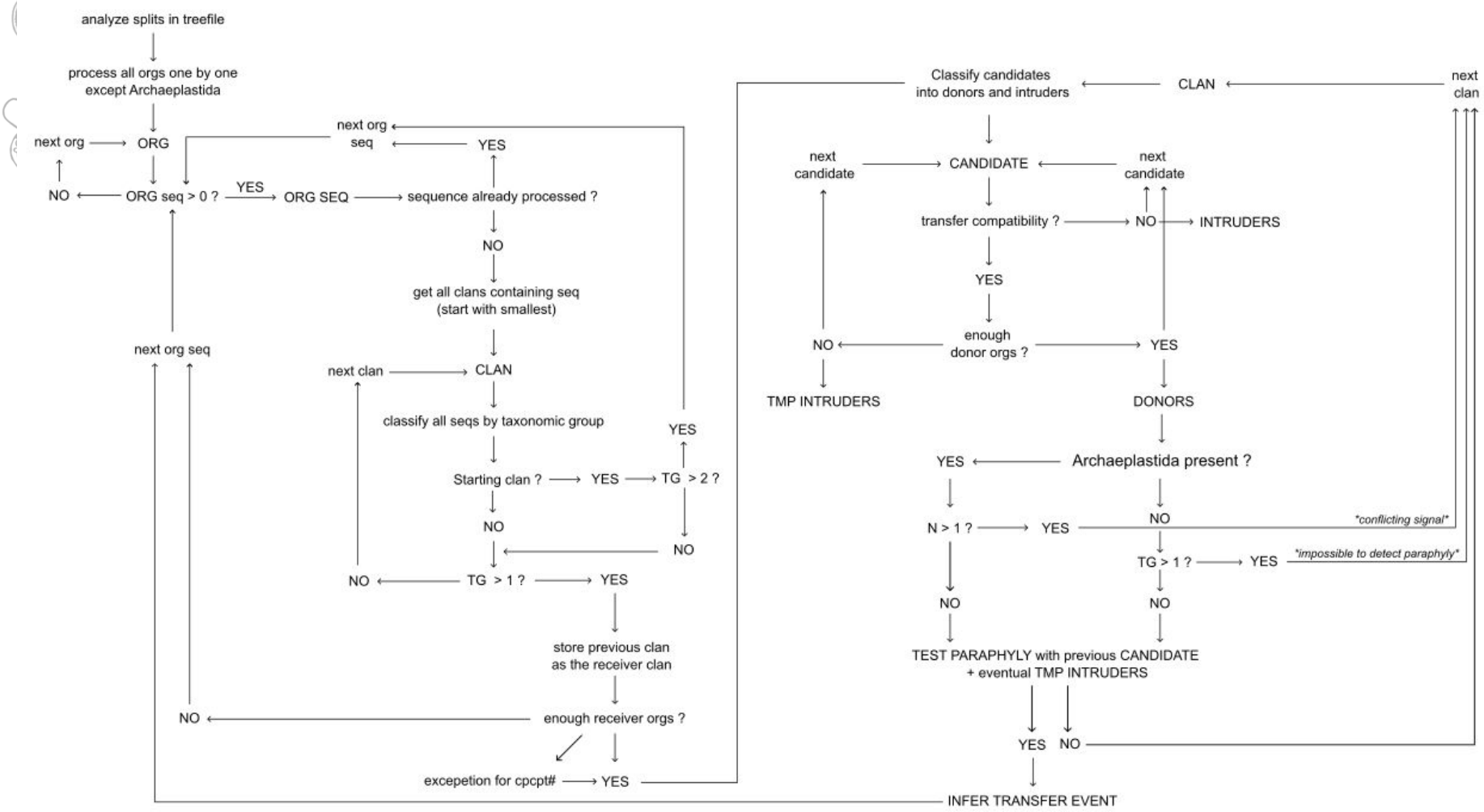
Petersen, J., ..., Baurain, D., & Brinkmann, H. (2014).GBE. 6, 666-684.

Strassert JFH, Irisarri I, Williams TA, Burki F. Nat Commun. 2021;12.

GT inference by paraphyly detection

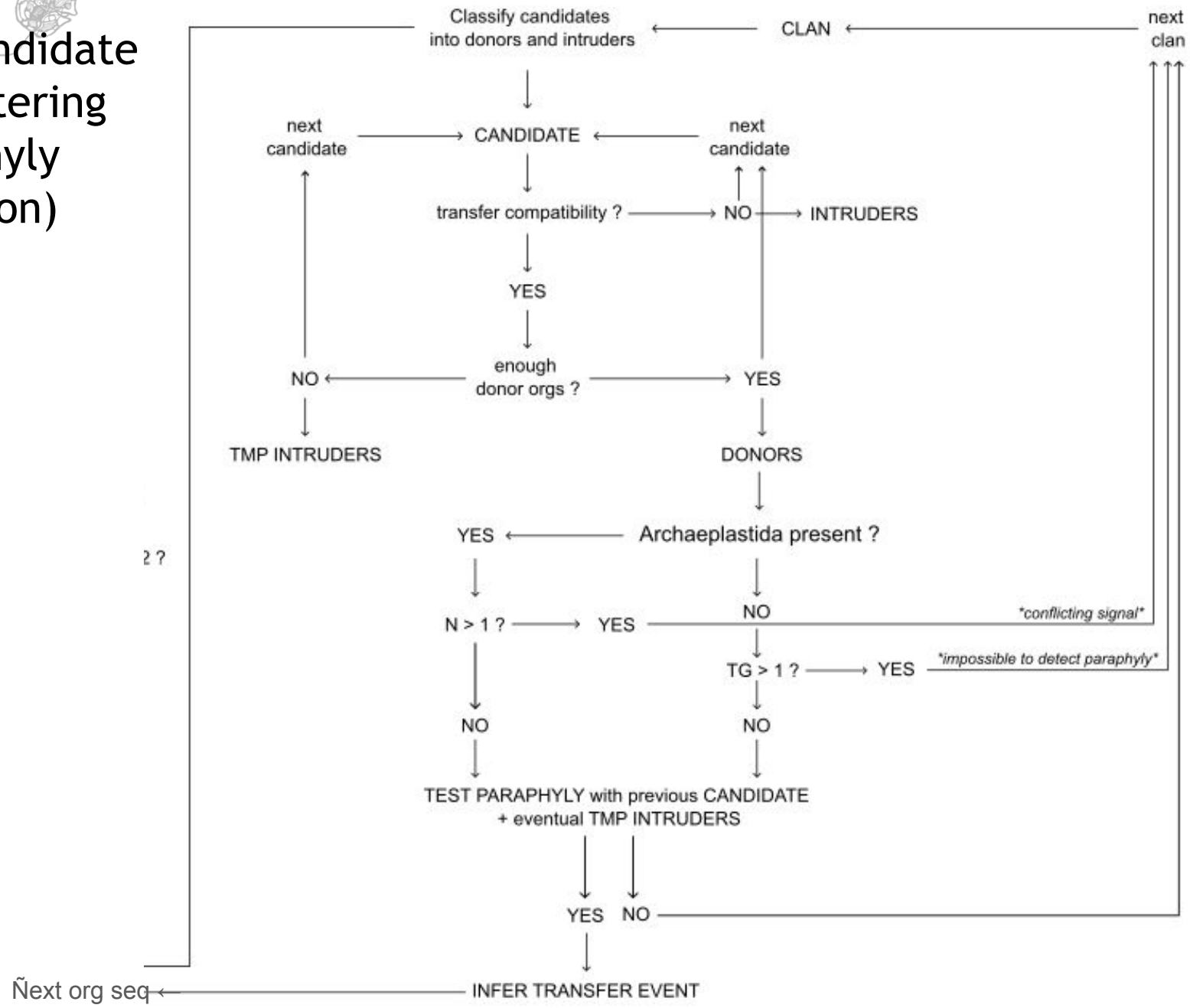


Flowchart GT-detector.pl



Step 1 : Receiver clan identification

Step 2 : Candidate donors filtering (paraphyly detection)



Dataset

Common issues & solutions

- taxon sampling
  - proteomes (public databases)
  - transcriptomes (**MMETSP**)
- contamination & redundancy
  - *dedicated pipelines (BMC Res. Notes)*
- gene families & paralogs
  - Orthogroups → trees → PS sub-trees

Data note | Open access | Published: 17 April 2021

Broadly sampled orthologous groups of eukaryotic proteins for the phylogenetic study of plastid-bearing lineages

Mick Van Vlierberghe, Hervé Philippe & Denis Baurain

BMC Research Notes **14**, Article number: 143 (2021) | Cite this article

1577 Accesses | **5** Citations | **2** Altmetric | Metrics

Data note | Open access | Published: 09 August 2021

Decontamination, pooling and dereplication of the 678 samples of the Marine Microbial Eukaryote Transcriptome Sequencing Project

Mick Van Vlierberghe, Arnaud Di Franco, Hervé Philippe & Denis Baurain

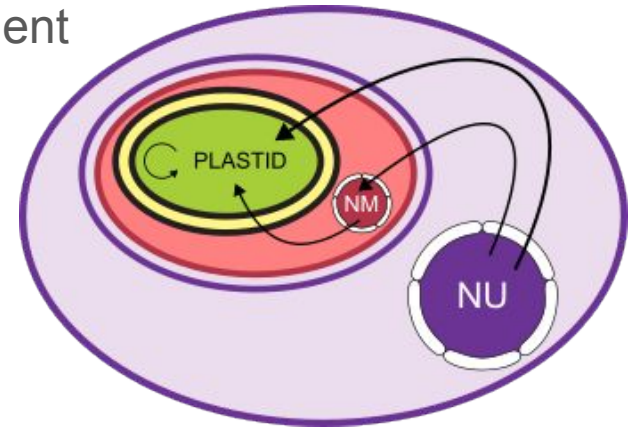
BMC Research Notes **14**, Article number: 306 (2021) | Cite this article

1862 Accesses | **7** Citations | **2** Altmetric | Metrics

Plastid-targeted proteins tagging

Curtis 2012, Dorrell 2017, Novak Vanclova 2020, …

- encoding compartment



Datasets

Data note | Open access | Published: 17 April 2021

Broadly sampled orthologous groups of eukaryotic proteins for the phylogenetic study of plastid-bearing lineages

Mick Van Vlierberghe, Hervé Philippe & Denis Baurain

BMC Research Notes **14**, Article number: 143 (2021) | Cite this article

1577 Accesses | **5** Citations | **2** Altmetric | Metrics

- ★ 73 high quality proteomes
- ★ ~ 40 000 photosynthetic clans

Data note | Open access | Published: 09 August 2021

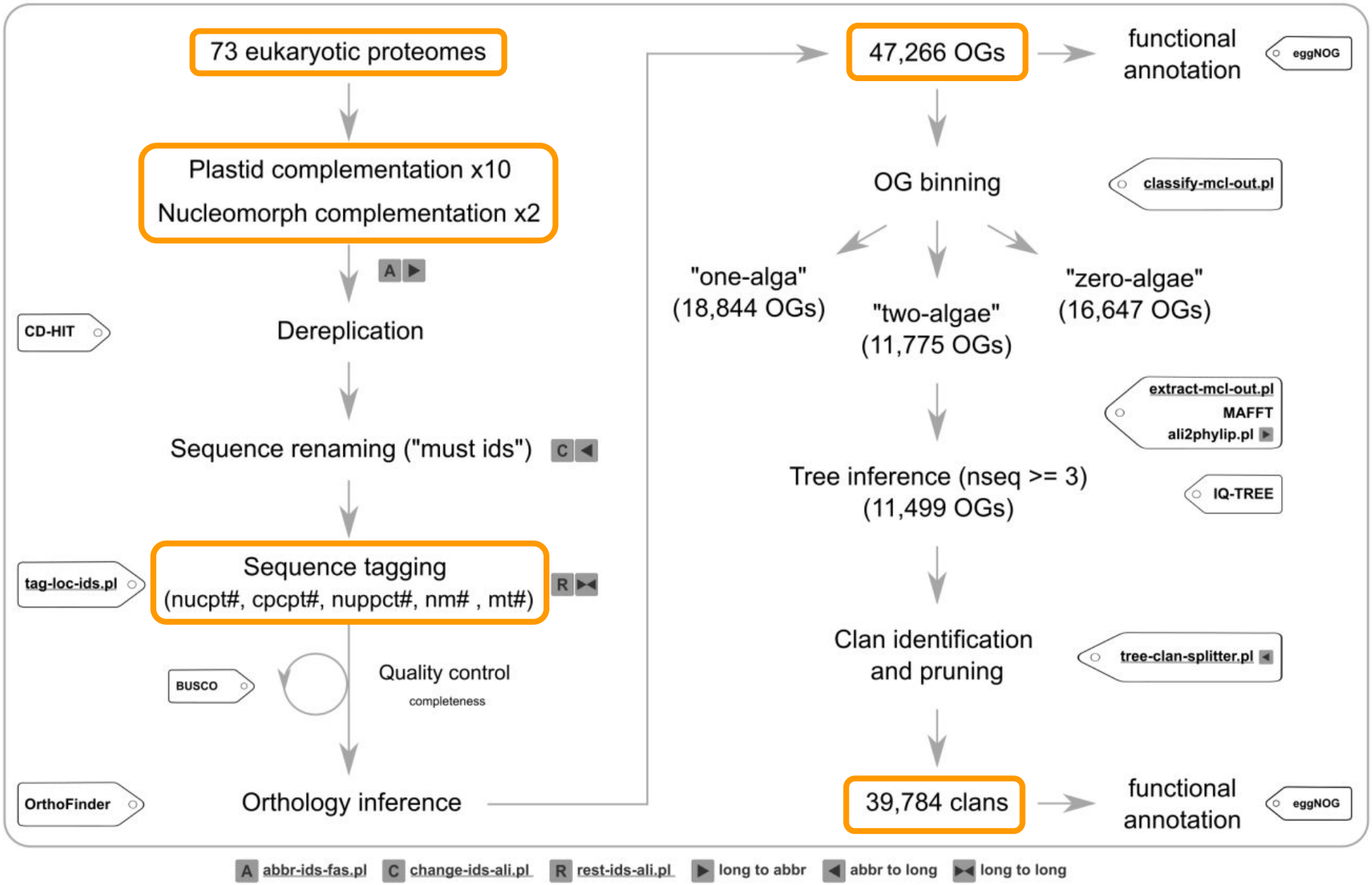
Decontamination, pooling and dereplication of the 678 samples of the Marine Microbial Eukaryote Transcriptome Sequencing Project

Mick Van Vlierberghe, Arnaud Di Franco, Hervé Philippe & Denis Baurain

BMC Research Notes **14**, Article number: 306 (2021) | Cite this article

1862 Accesses | **7** Citations | **2** Altmetric | Metrics

- ★ redundancy [678 -> 260]
- ★ contaminations [260 -> 224]


73 eukaryotic proteomes
Plastid complementation x10
Nucleomorph complementation x2
A
CD-HIT
Dereplication
Sequence renaming ("must ids")
C
tag-loc-ids.pl
Sequence tagging
(nucpt#, cpcpt#, nuppct#, nm# , mt#)
R
BUSCO
Quality control
completeness
OrthoFinder
Orthology inference
47,266 OGs
functional annotation
eggNOG
OG binning
classify-mcl-out.pl
"one-alga" (18,844 OGs)
"two-algae" (11,775 OGs)
"zero-algae" (16,647 OGs)
extract-mcl-out.pl
MAFFT
ali2phylip.pl
Tree inference (nseq >= 3)
(11,499 OGs)
IQ-TREE
Clan identification and pruning
tree-clan-splitter.pl
39,784 clans
functional annotation
eggNOG
A abbr-ids-fas.pl
C change-ids-ali.pl
R rest-ids-ali.pl
long to abbr
abbr to long
long to long

Target: nucleus-encoded and plastid-targeted genes

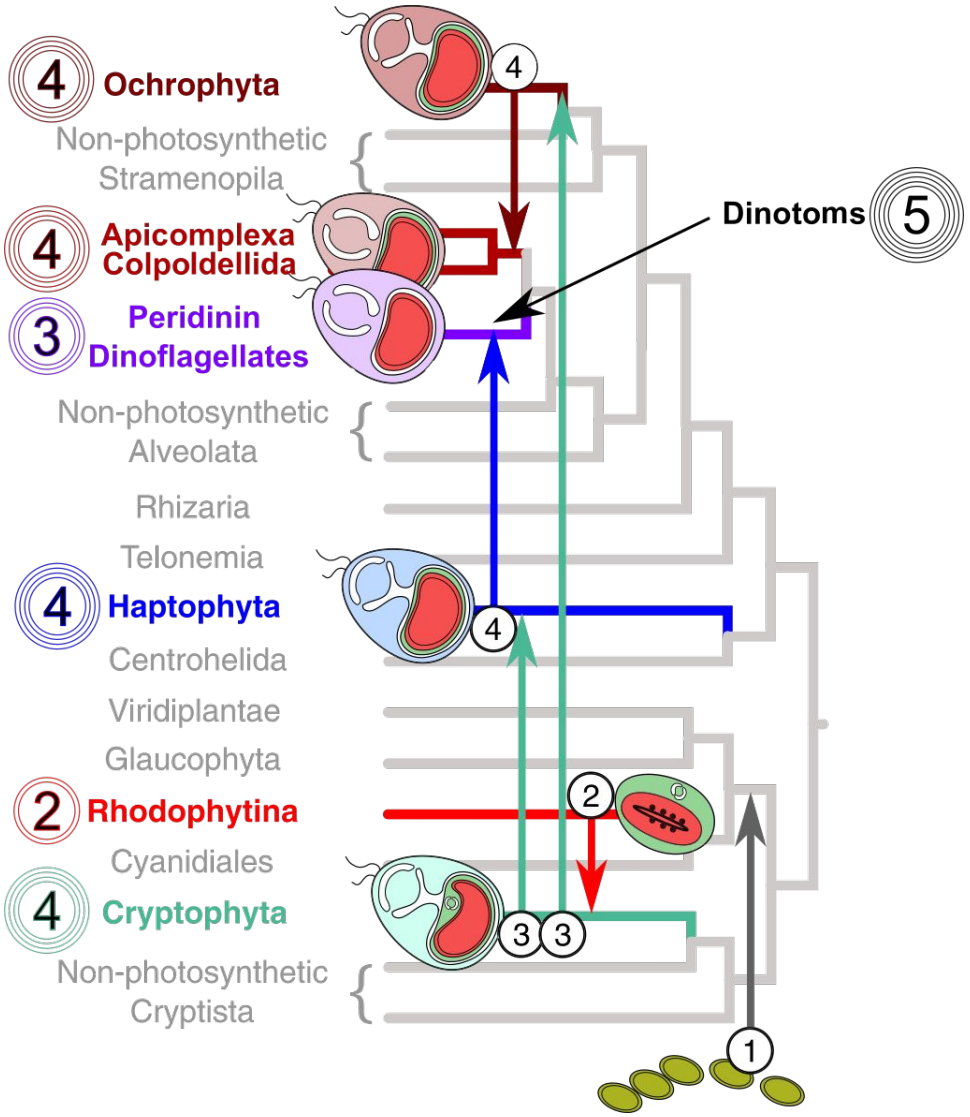
Tool: tree parser

Data: published dataset

Working Hypothesis:

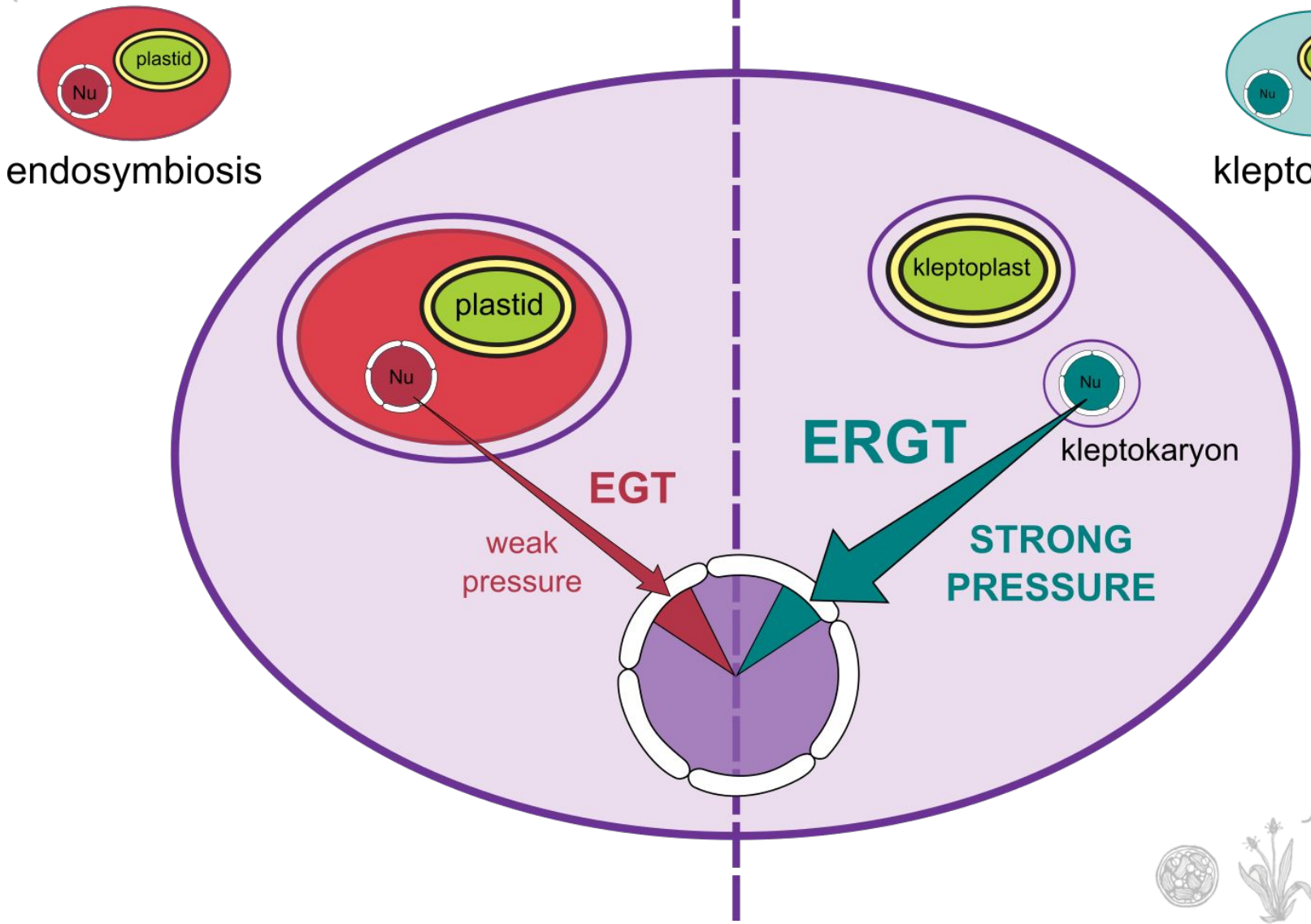
Kleptoplasty is a major mechanism in complex plastid acquisition

[1] Kleptoplasty? Lower membrane accumulation



Adapted from Strassert JFH, Irisarri I, Williams TA, Burki F. Nat Commun. 2021;12.

[2] Kleptoplasty? Stronger selective pressure



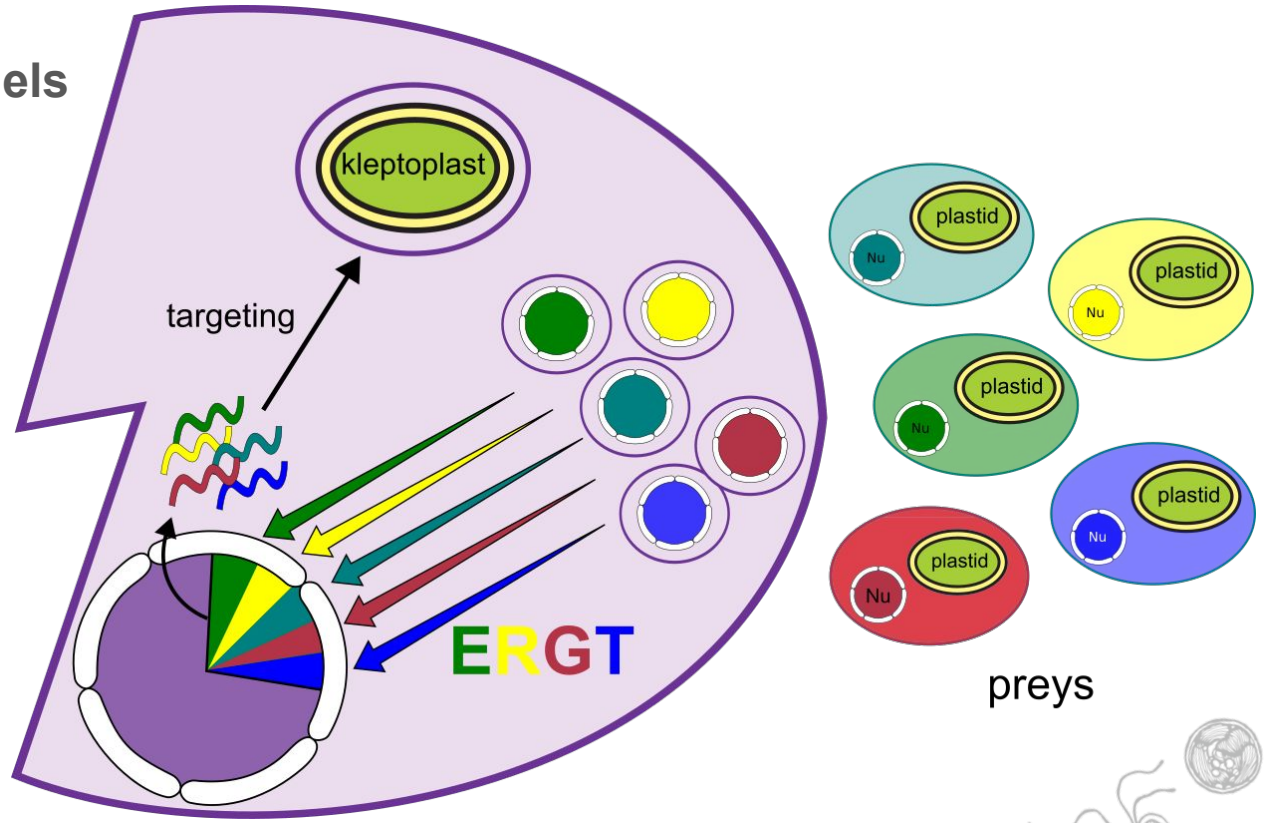
Prediction: Higher gene mosaicism

 # Diversity

- preys → NUPTs

 # Pre-adaptation models

- Shopping bag[1]
- Red carpet[2]



1. Morozov AA, Galachyants YP. Mar Genomics. 2019;45 February:72–8.
2. Ponce-Toledo RI, López-García P, Moreira D. New Phytol. 2019.

Framework to compare gene mosaicism

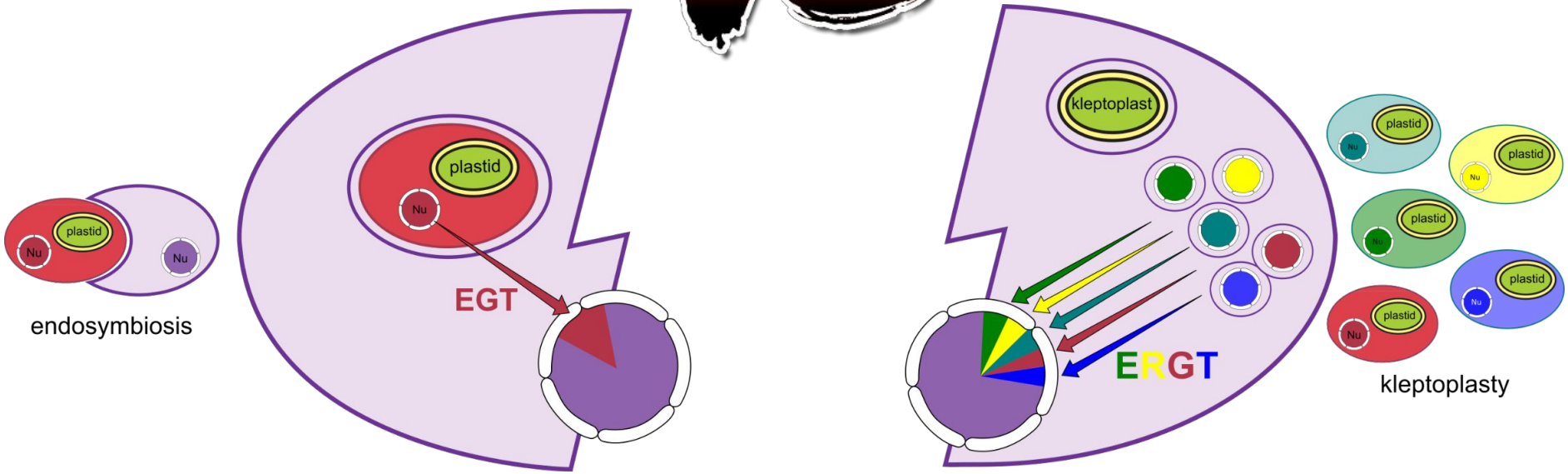
Endosymbiosis

Single origin (symbiont)

VS

Kleptoplasty

Multiple origins (preys)



Low diversity

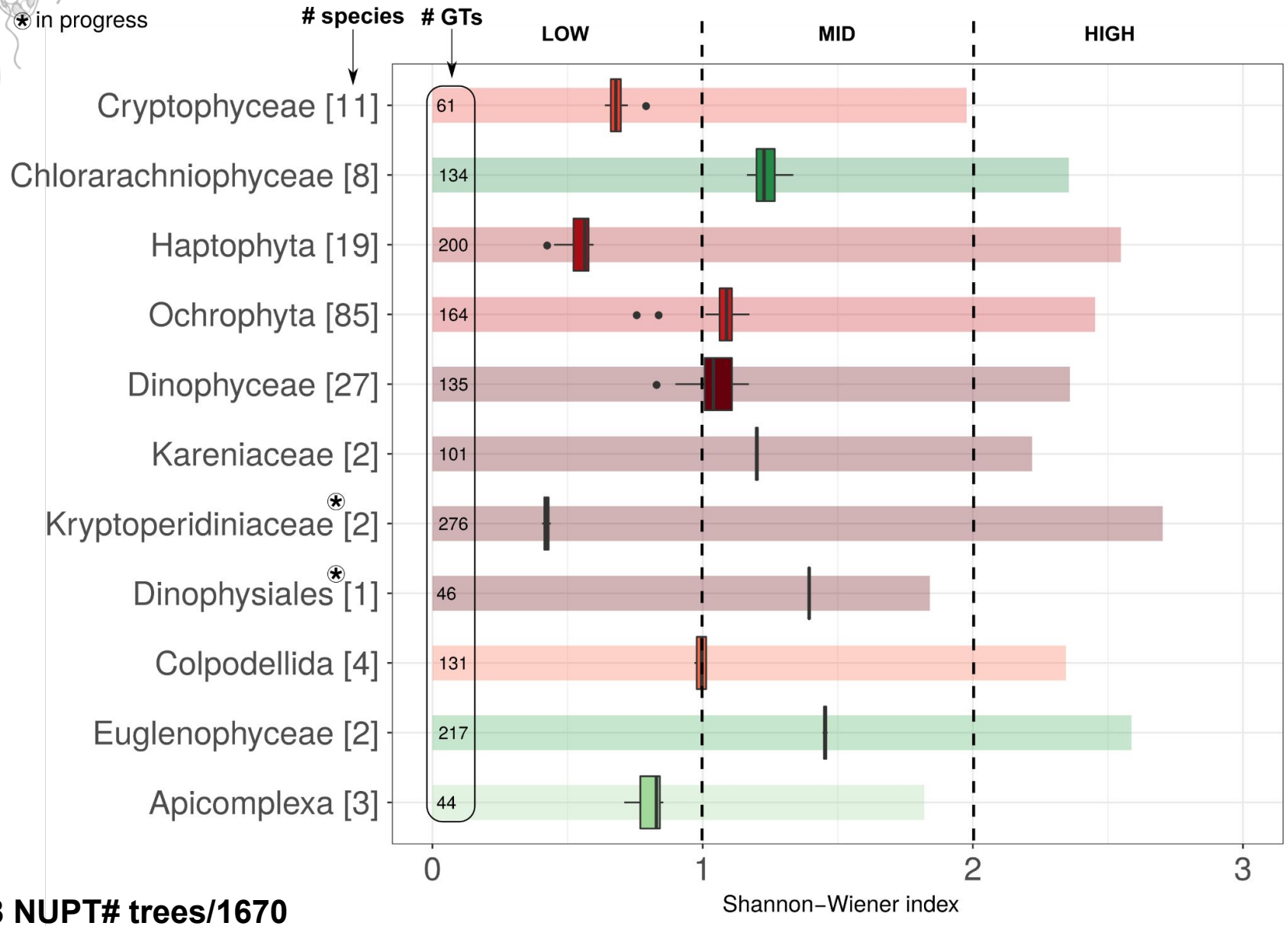
α-diversity metric

High diversity

Shannon-Wiener index

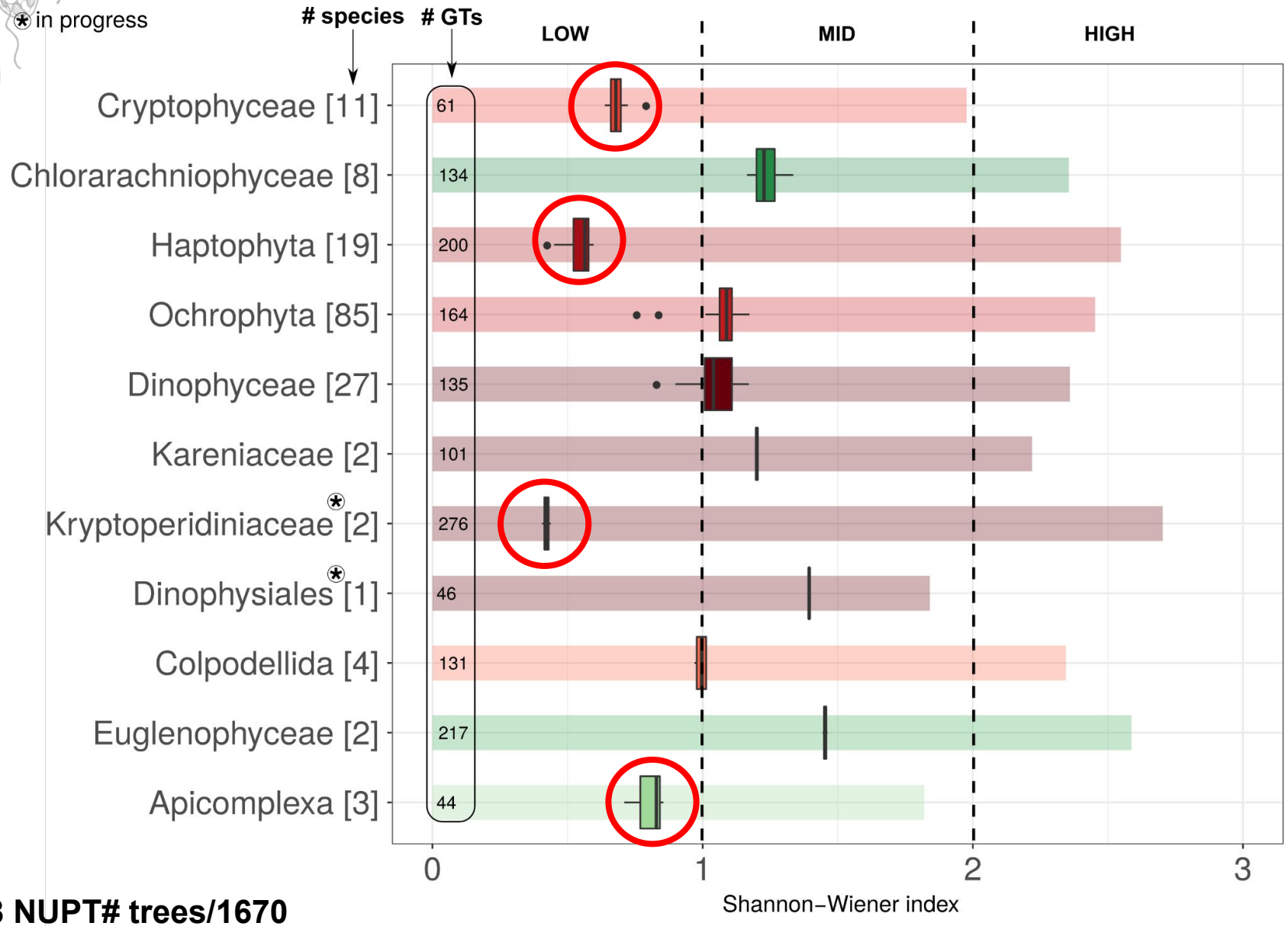
Results

⊛ in progress



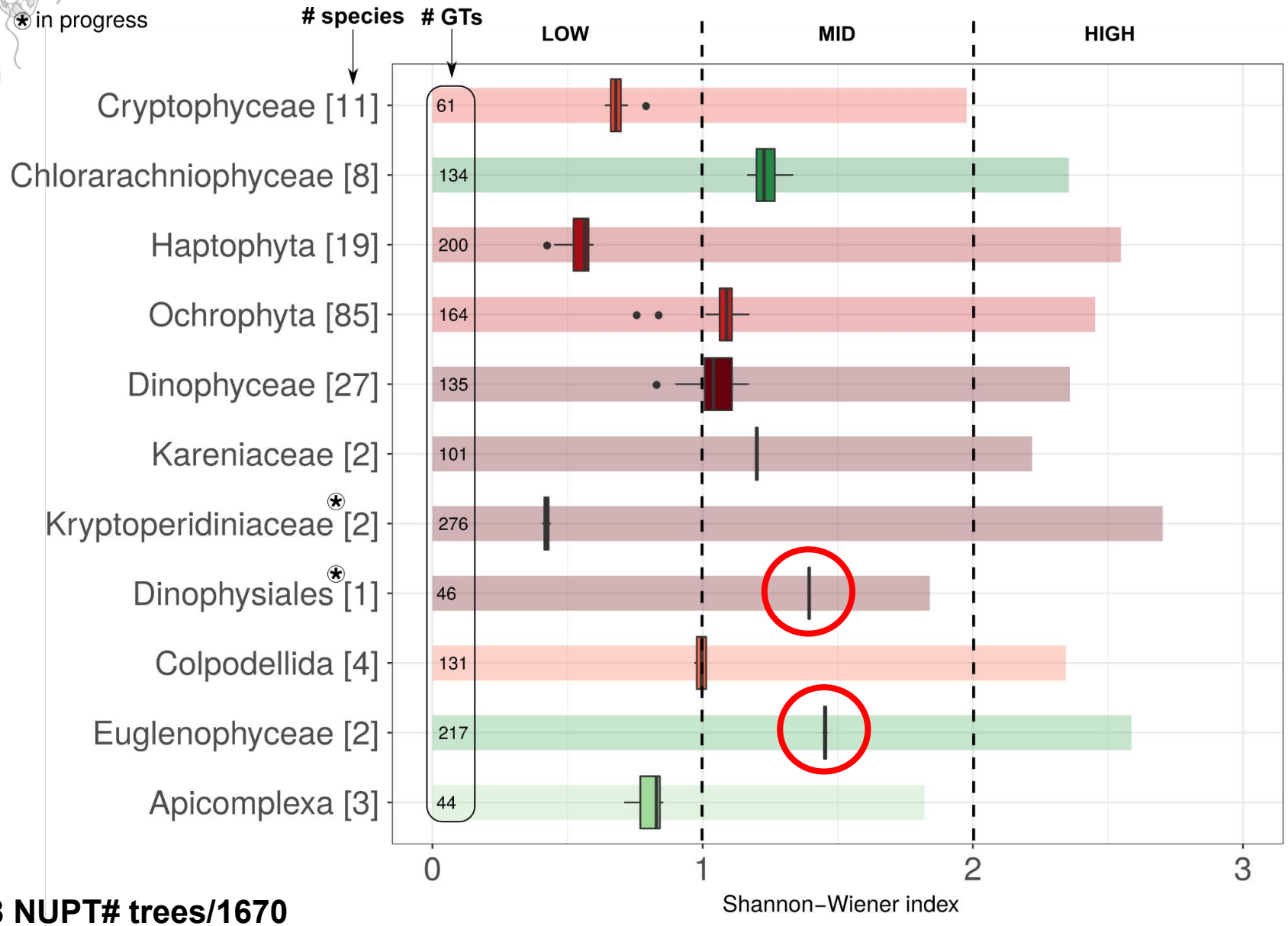
1178 NUPT# trees/1670

Results



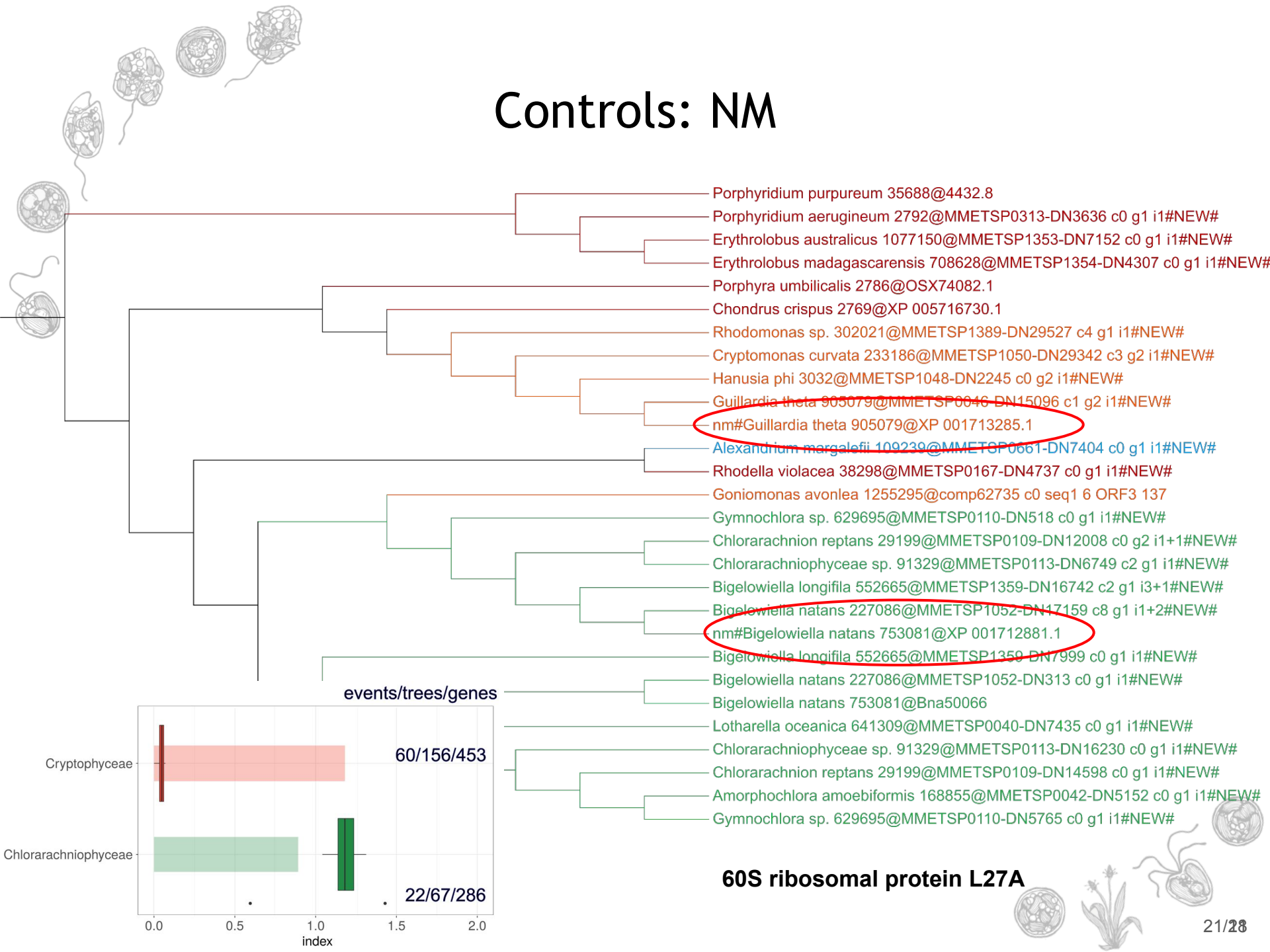
1178 NUPT# trees/1670

Results



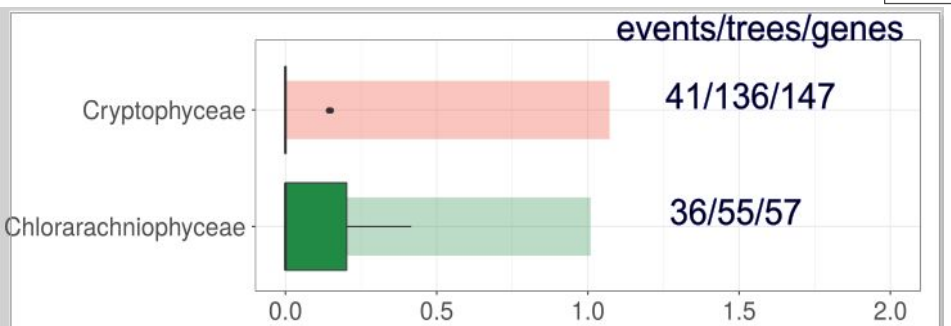
1178 NUPT# trees/1670

Controls: NM



60S ribosomal protein L27A

Controls: CPCPT



Karenia brevis 156230@MMETSP0574-DN32289 c0 g1 i1+1#NEW#
Isochrysidales
Prymnesium parvum 97485@MMETSP1083-DN11425 c2 g1 i1#NEW#
Karlodinium veneficum 342587@MMETSP1017-DN40206 c3 g1 i1#NEW#
cpcpt#Chrysochromulina sp. 1460289@AHY04333.1
cpcpt#Chromera velia 505693@YP 003795319.1
Corethron hystrix 216773@MMETSP0010-DN1581 c0 g1 i1#NEW#
Proboscia sp. 216776@MMETSP0176-DN16944 c13 g1 i1#NEW#
Bacillariophyta
cpcpt#Nannochloropsis gaditana 72520@AFZ64260.1
Dinophyceae
Florenciella parvula 236787@MMETSP1344-DN10580 c0 g1 i1#NEW#
cpcpt#Aureococcus anophagefferens 44056@AACG00550
Ochrophyta
cpcpt#Galdieria sulphuraria 130081@671743522
cpcpt#Chondrus crispus 2769@YP 007627376.1
Hemiselmis tepida 464990@MMETSP1355-DN13414 c3 g1 i1#NEW#
cpcpt#Guillardia theta 905079@NP 050781.1
Rhodella violacea 38298@MMETSP0167-DN7605 c0 g1 i1#NEW#
cpcpt#Porphyra umbilicalis 2786@ASN78775.1
cpcpt#Pyropia yezoensis 2788@YP 536982.1
Cyanoptyche sp. 77921@MMETSP1086-DN14043 c3 g1 i1#NEW#
cpcpt#Cyanophora paradoxa 2762@NP 043192.1
cpcpt#Cyanidioschyzon merolae 45157@CMCG01770
cpcpt#Porphyridium purpureum 35688@568247648
cpcpt#Euglena gracilis 3039@NP 041920.1
Selaginella moellendorffii 88036@XP 002981185.1
nucpt#Arabidopsis thaliana 3702@AT4G01310
Selaginella moellendorffii 88036@XP 002963746.1
Selaginella moellendorffii 88036@XP 002974829.1
Dolichomastix tenuilepis 195969@MMETSP0033-DN9275 c0 g1 i1#NEW#
Pyramimonas sp. 36882@MMETSP1169-DN12802 c4 g1 i1#NEW#
Viridiplantae
Stichococcus sp. 1481547@MMETSP1473-DN4861 c2 g1 i1#NEW#
cpcpt#Bigelowiella natans 753081@YP 778571.1
Chlorophyta

50S ribosomal protein L5 (plastid)

Concluding remarks

Tree parser

- true gene transfer ✔
- controls ✔
- artifacts ❌

Data

- unexploited trees (+30,000 PS sub-trees) ✔❌
- lack of protein targeting data (tagging) ✔❌
- uneven sampling of algae ❌

Kleptoplasty?

- comparative study ✔
- *Euglenids*
  - highest diversity ✔
  - *Dinophysis acuminata* ✔
  - 3 membranes ✔
  - *Rapaza viridis (Karnkowska 2023)* ✔
    - loss of membrane by phagocytosis ✔
    - shared NUPTs with Euglenids ✔

Durandal

Denis Baurain

ULg, B22

Institut de Botanique

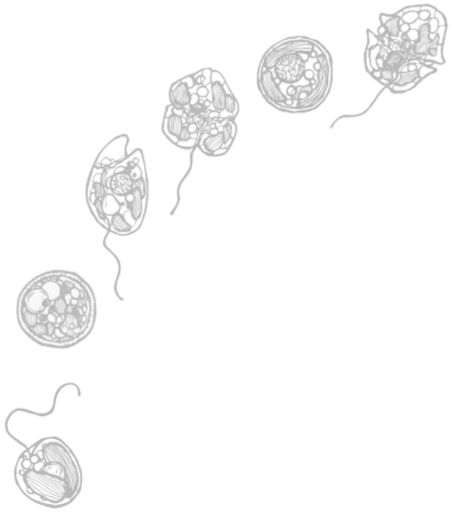
Me

Google

LIÈGE université

fnrs

LA LIBERTÉ DE CHERCHER



Thank you!