On the origin of
chlorophyll c-containing algae

A phylogenomic falsification of the Chromalveolate hypothesis

Denis BAURAIN

Université de Montréal / Université de Liège

Cologne – April 20, 2010

Acknowledgments

Mol Biol Evol. 2010 Mar 1. [Epub ahead of print]

Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes and stramenopiles.

Baurain D, Brinkmann H, Petersen J, Rodríguez-Ezpeleta N, Stechmann A, Demoulin V, Roger AJ, Burger G, Lang BF, Philippe H.

Département de Biochimie, Centre Robert-Cedergren, Université de Montréal, Montréal Québec, Canada.

Abstract

According to the chromalveolate hypothesis (Cavalier-Smith 1999, J Eukaryot Microbiol 46:347-366), the four eukaryotic groups with chlorophyll c-containing plastids originate from a single photosynthetic ancestor, which acquired its plastids by secondary endosymbiosis with a red alga. So far, molecular phylogenies have failed to either support or disprove this view. Here we devise a phylogenomic falsification of the chromalveolate hypothesis that estimates signal strength across the three genomic compartments: if the four chlorophyll c-containing lineages indeed derive from a single photosynthetic ancestor, then similar amounts of plastid, mitochondrial and nuclear sequences should allow to recover their monophyly. Our results refute this prediction, with statistical support levels too different to be explained by evolutionary rate variation, phylogenetic artifacts, or endosymbiotic gene transfer. Therefore, we reject the chromalveolate hypothesis as falsified in favor of more complex evolutionary scenarios involving multiple higher-order eukaryote-eukaryote endosymbioses.

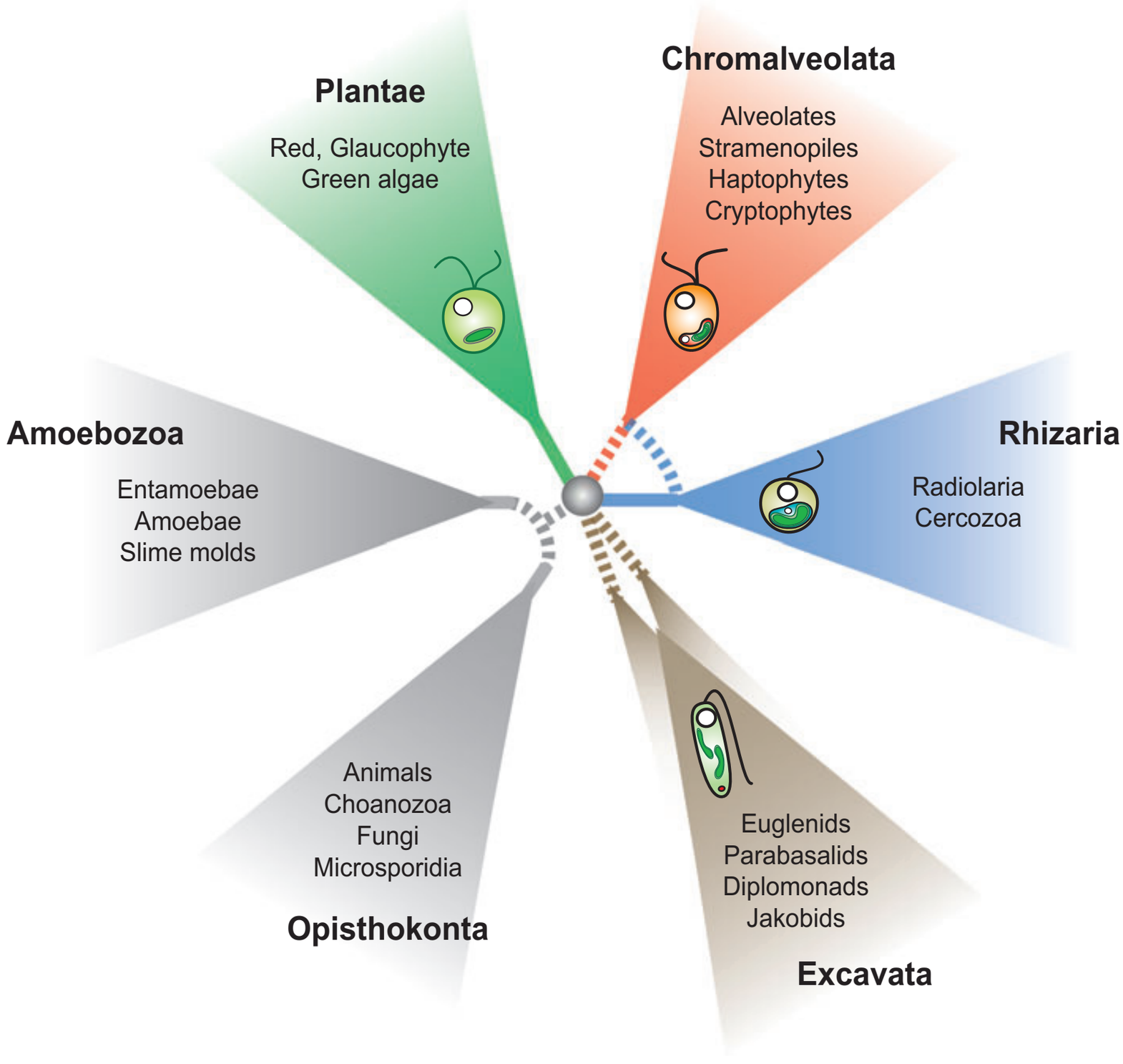
Do 'Chromalveolates' really exist?

Outline of the Talk

Do 'Chromalveolates' really exist?

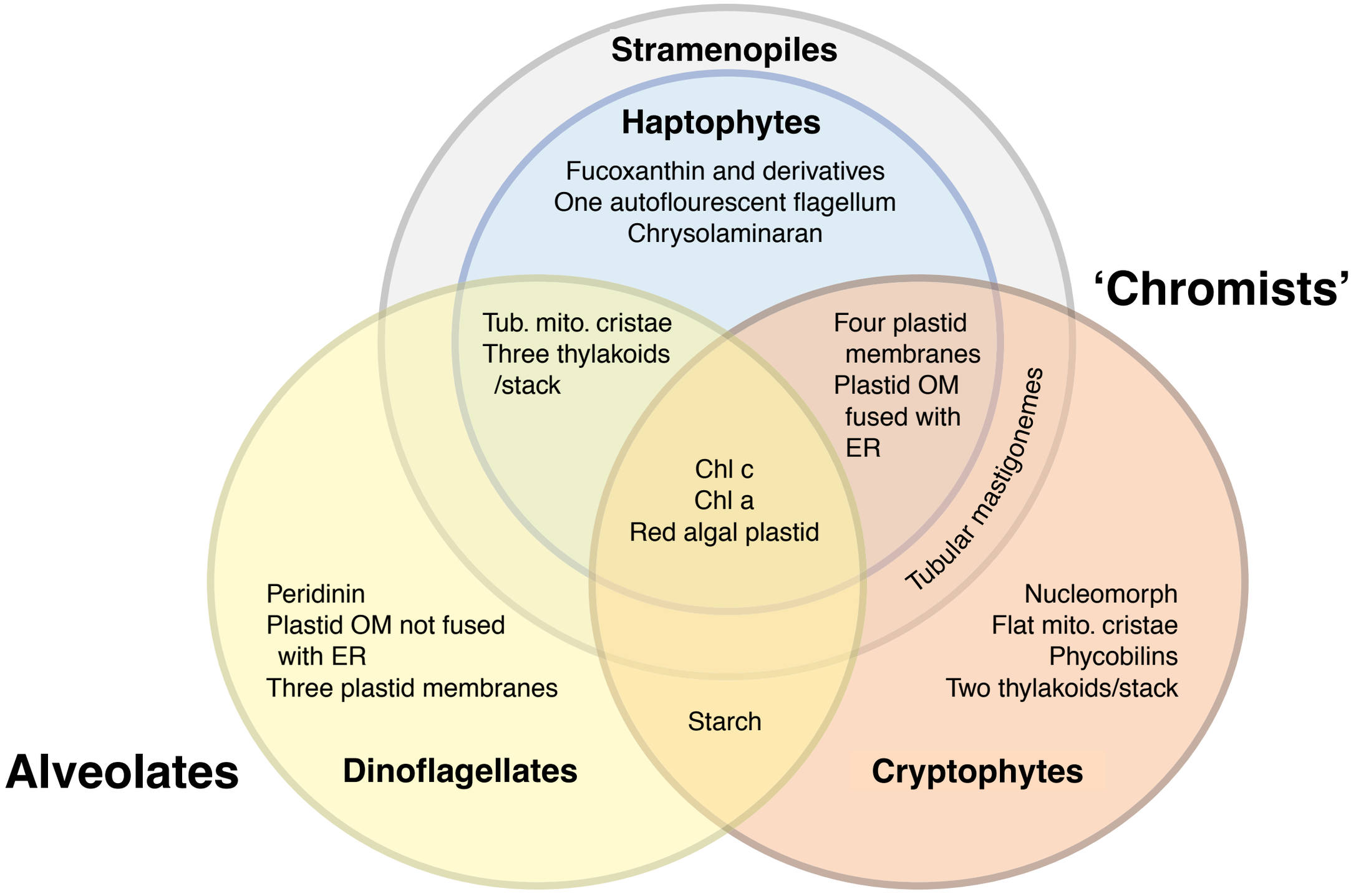
1. *What are 'Chromalveolates'?*
2. *How to test the Chromalveolate hypothesis?*
3. *How to check the assumptions of our test?*
4. *Can we specify an alternative hypothesis?*

What are ‘Chromalveolates’?



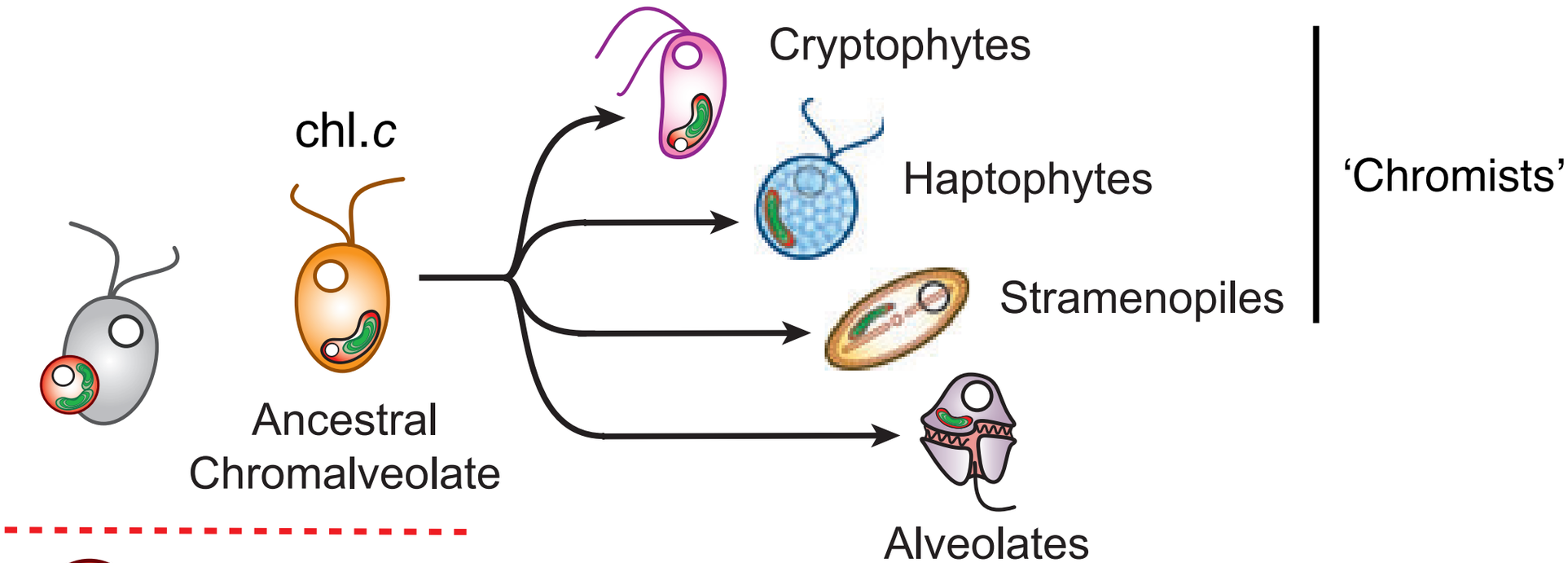
‘Chromalveolates’ are one of the six putative Eukaryotic supergroups.

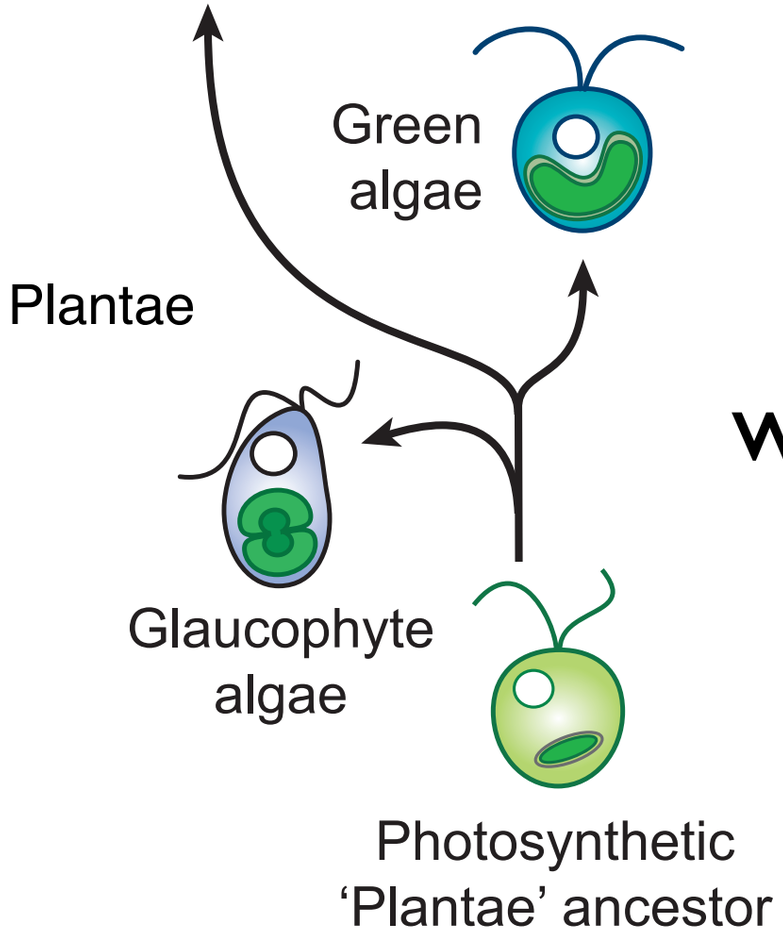
Reyes-Prieto et al. (2007) *Annu Rev Genet* 41:147-168



Chlorophyll *c*-containing algae share a number of biochemical and ultrastructural features.

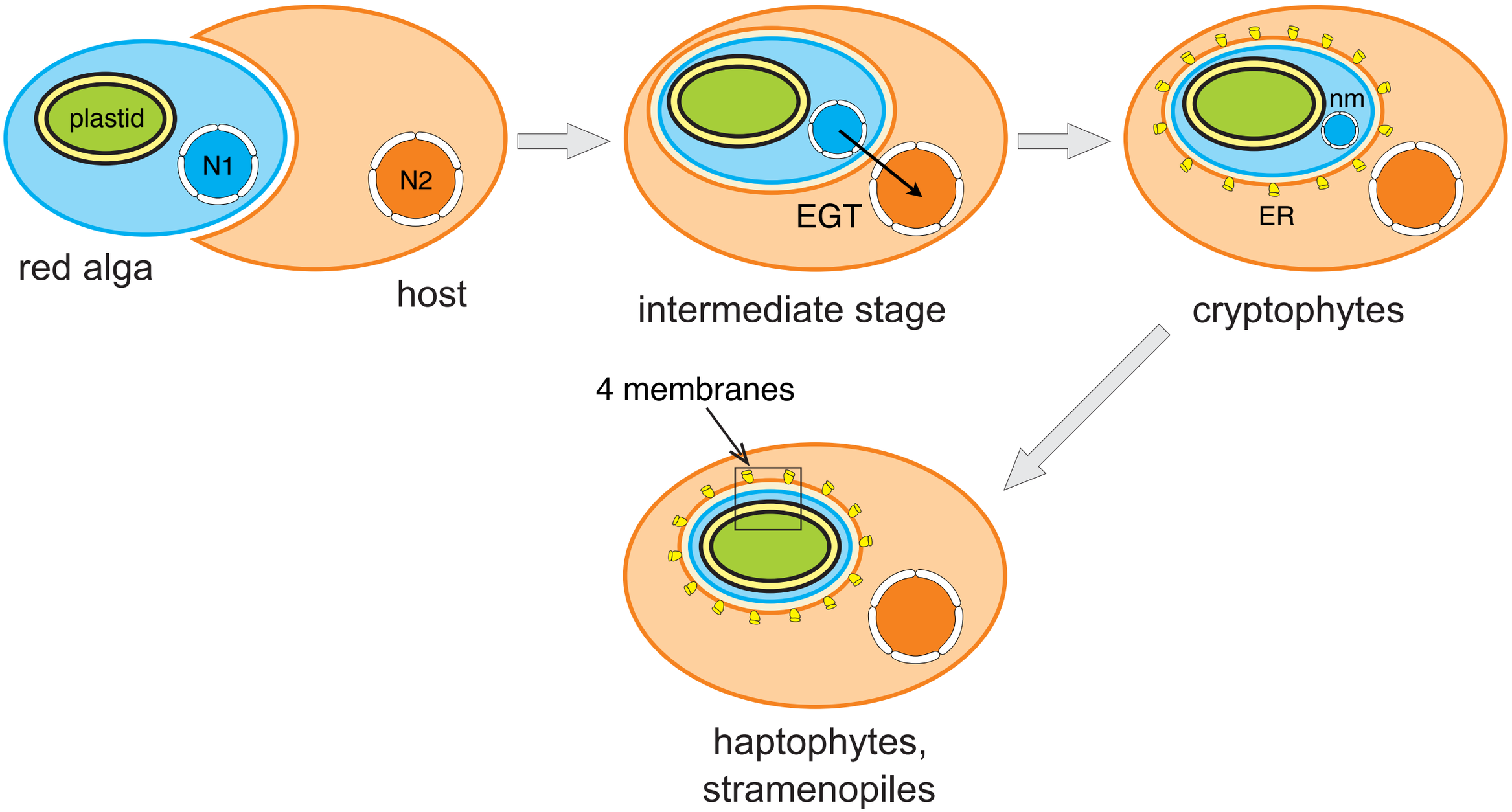
adapted from Archibald and Keeling (2002) *TiG* 18:577-584





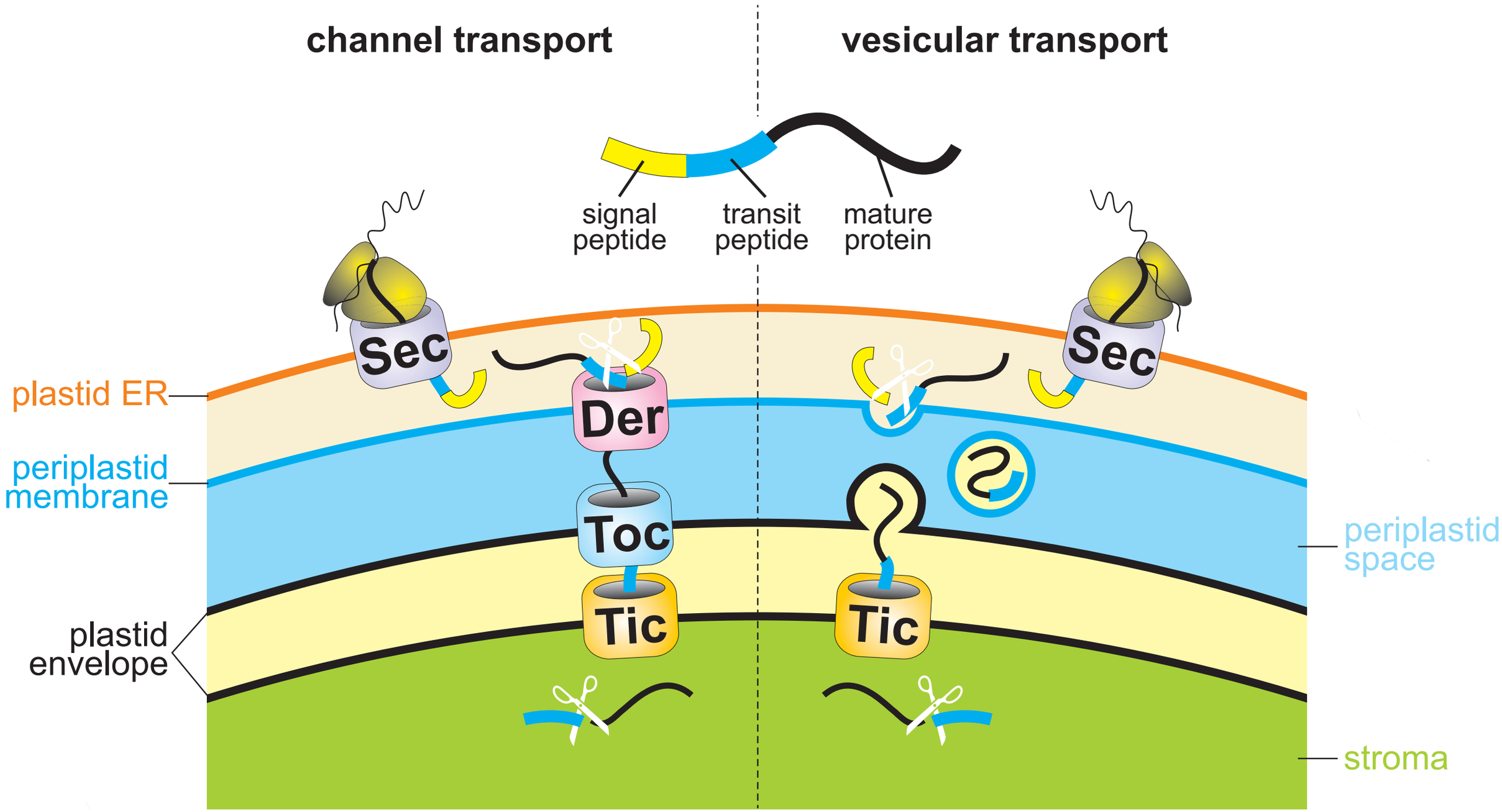
The Chromalveolate hypothesis posits a single secondary endosymbiosis with a red alga in the common ancestor of all chlorophyll *c*-containing algae.

adapted from Reyes-Prieto et al. (2007) *Annu Rev Genet* 41:147-168



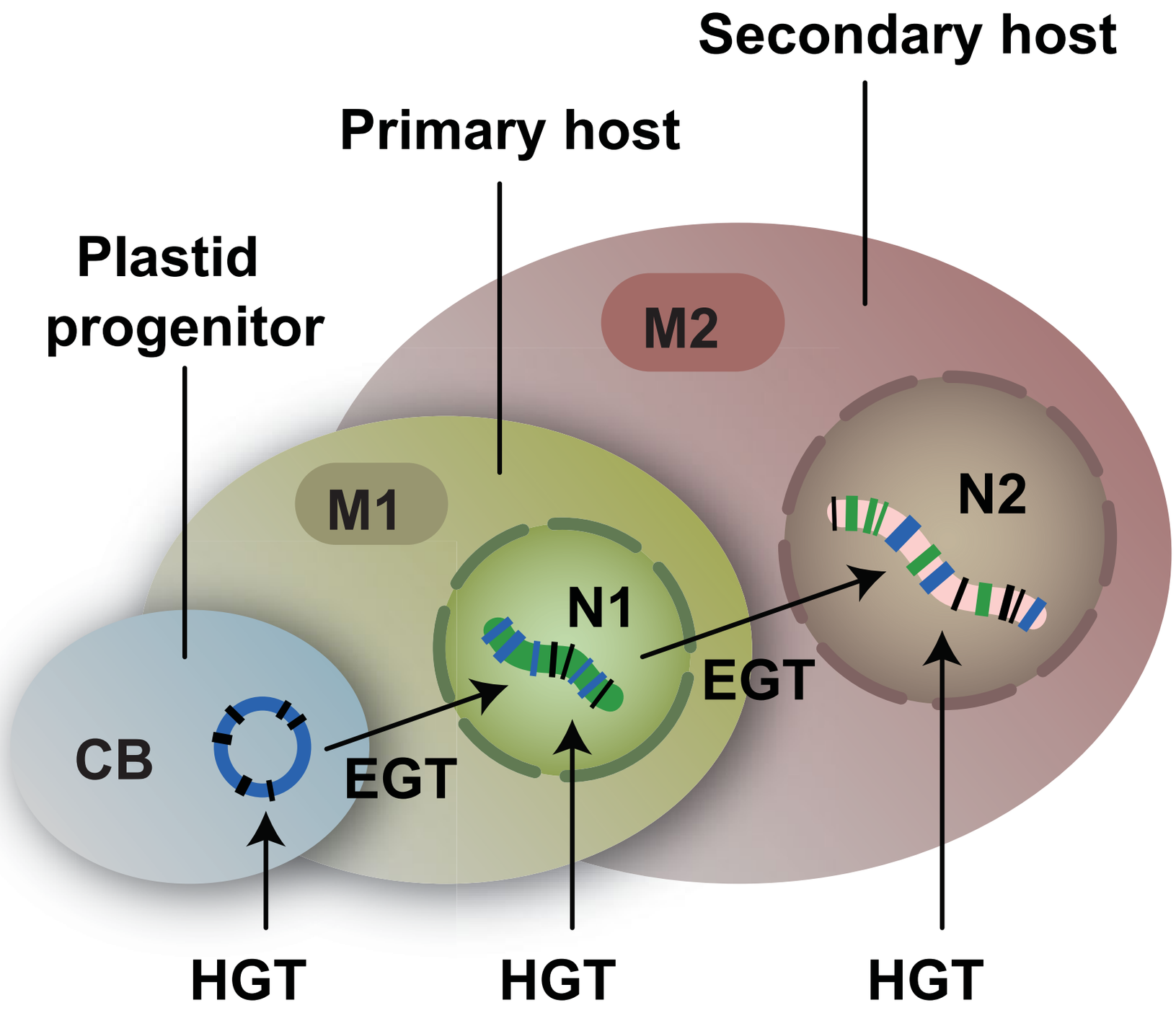
Plastid gain through endosymbiosis is considered very difficult and thus much rarer than plastid loss.

adapted from Bodyl et al. (2009) *BioEssays* 31:1219-1232



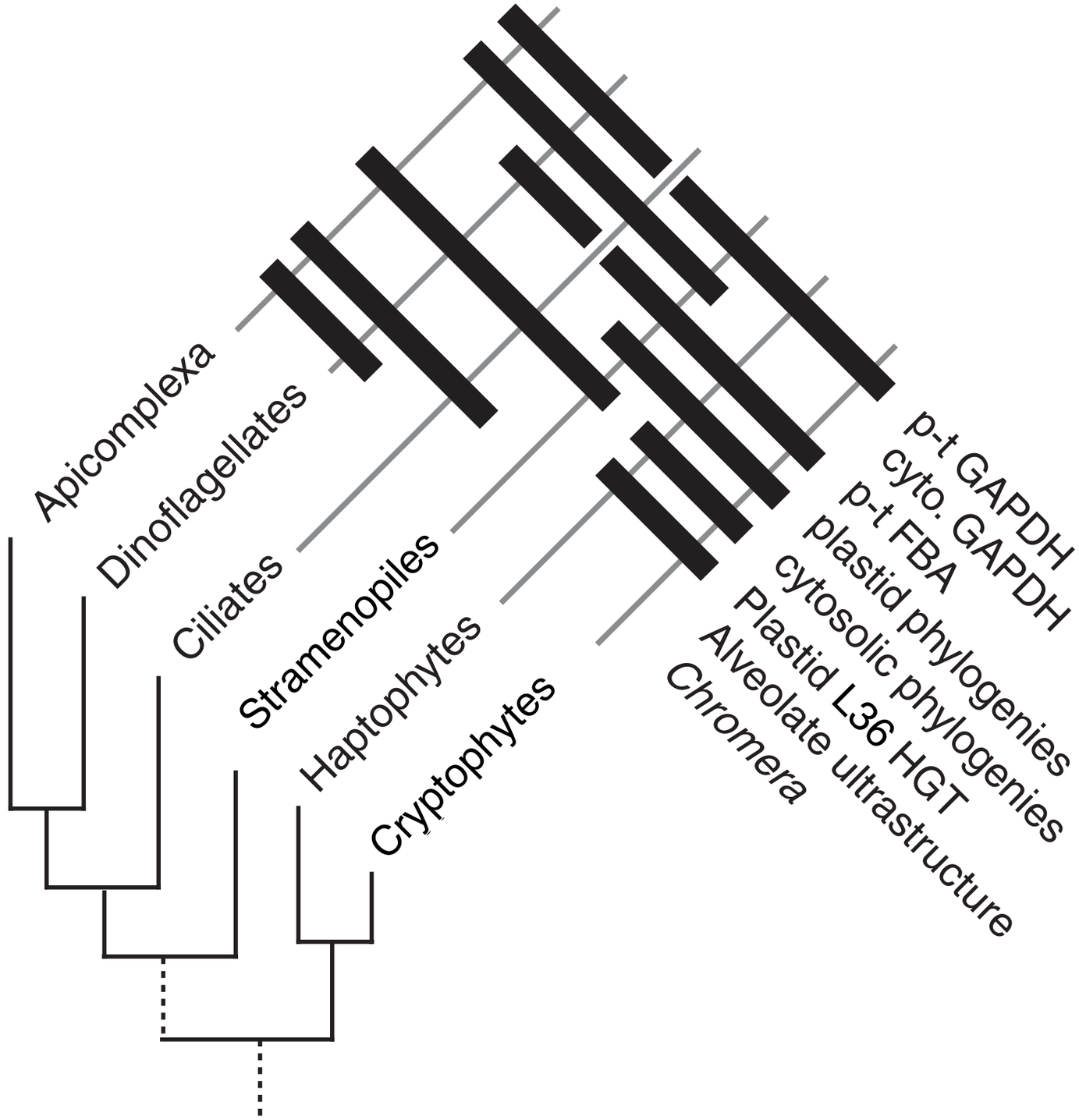
The acquisition of a new targeting signal by hundreds of genes is the step generally seen as limiting.

Bodyl et al. (2009) *BioEssays* 31:1219-1232



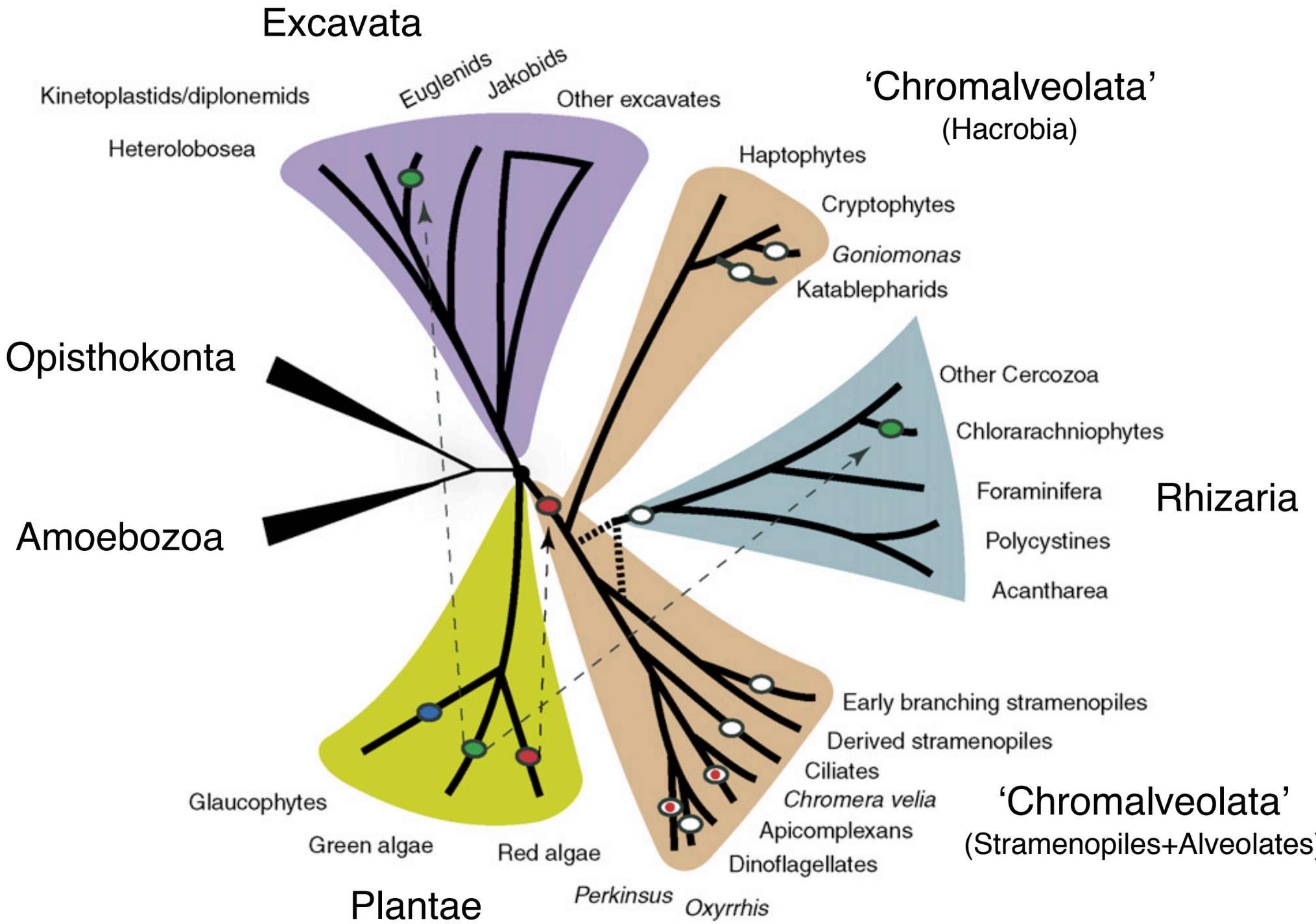
Eukaryotic algae have chimeric genomes that are expected to show evidence of their history.

Elias and Archibald (2009) *BioEssays* 31:1273-1279



There is no one piece of molecular evidence that unambiguously unites all ‘Chromalveolates’.

adapted from Keeling (2009) *J Eukaryot Microbiol* 56:1-8



Phylogenomics has led to dramatically expanded ‘Chromalveolates’ implying even more plastid losses.

adapted from Lane and Archibald (2008) *TREE* 23:268-275

What are 'Chromalveolates'?

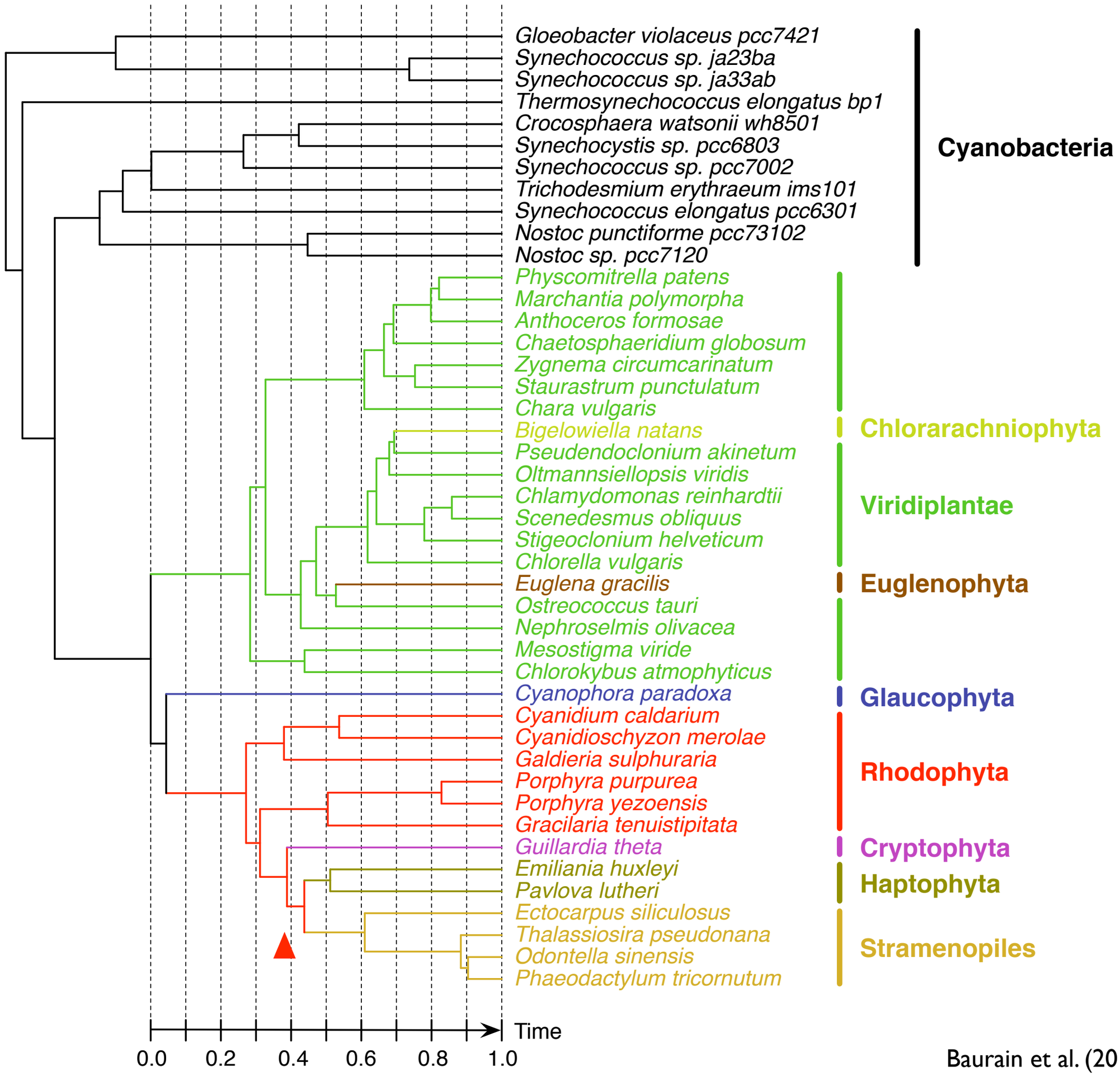
'Chromalveolates' are a putative, large and diverse supergroup of Eukaryotes considered to vertically descent from a single chlorophyll *c*-containing ancestor that acquired its plastid *via* a secondary ensymbiosis with a red alga.

The Chromalveolate hypothesis has always been parsimony-driven and regards plastid loss as much more common than plastid gain.

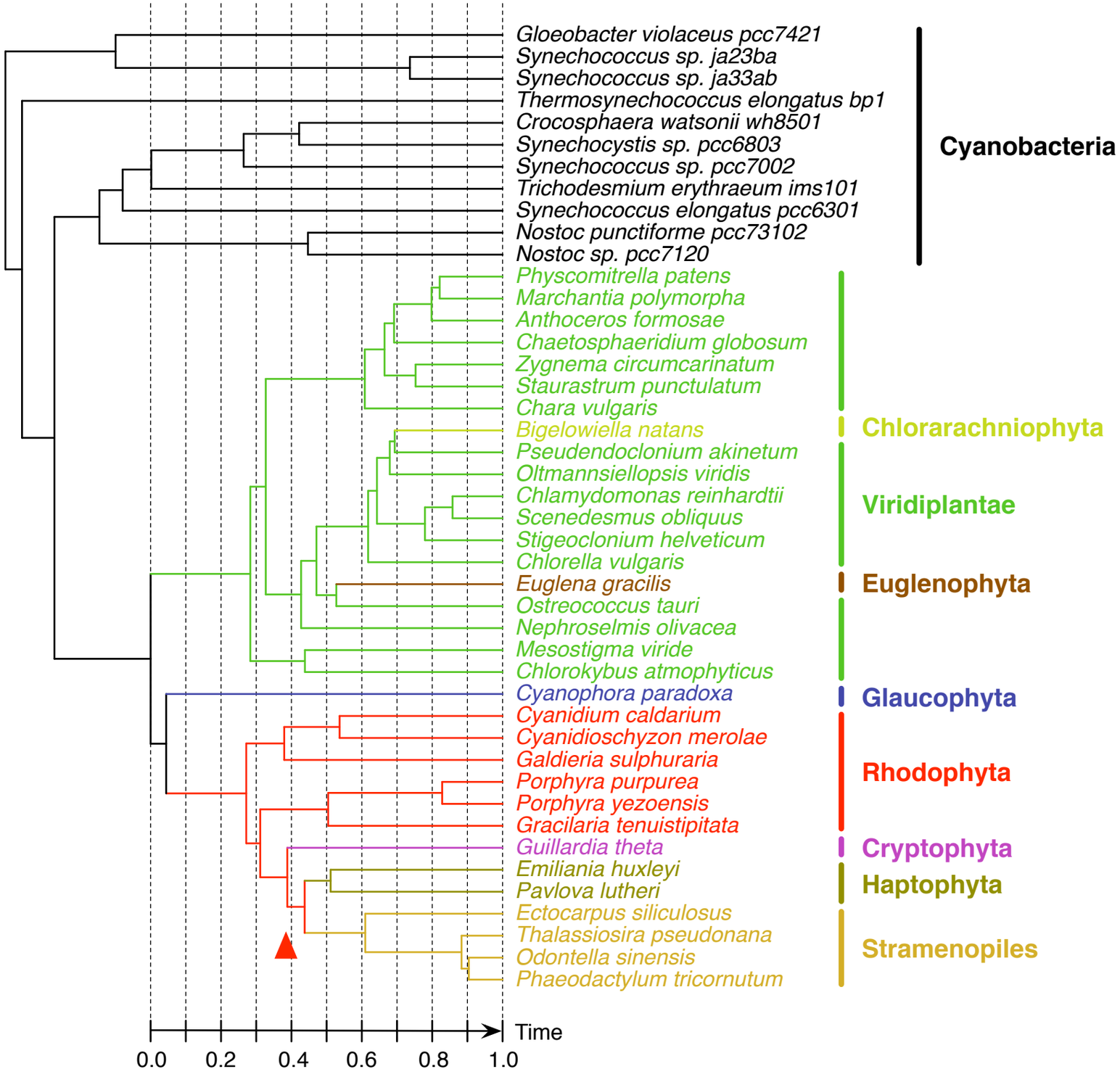
How to test the Chromalveolate hypothesis?

How to test the Chromalveolate hypothesis?

1. *How old are chl. c-containing plastids?*
2. *How to compare the plastid with mitochondrial and nuclear compartments?*
3. *How to estimate the strength of the phylogenetic signal?*
4. *How to validate our approach?*
5. *Now do 'Chromists' pass our test?*



Baurain et al. (2010) *Mol Biol Evol* (in press)



55 plastid-encoded proteins
44 species x 10,805 AA

CAT+Γ_4 model
relaxed-clock relative dating

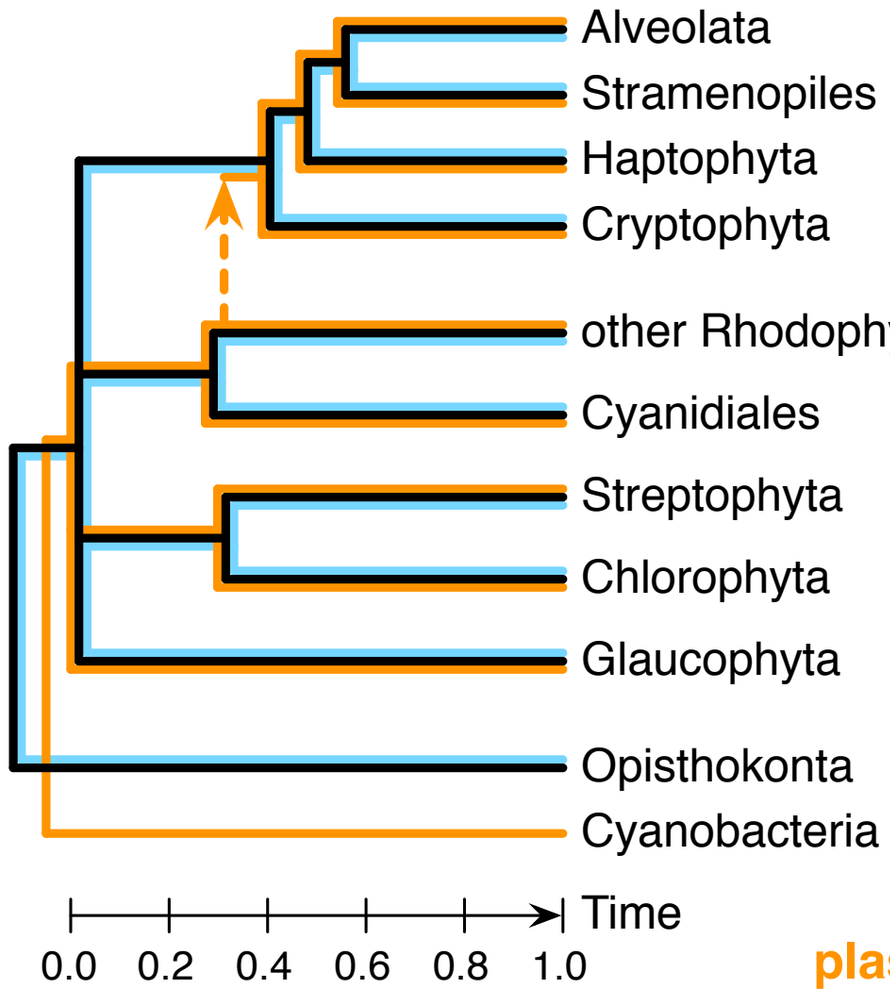
Chlorophyll *c*-containing plastids postdate the diversification of both red algae and green plants.

Baurain et al. (2010) *Mol Biol Evol* (in press)

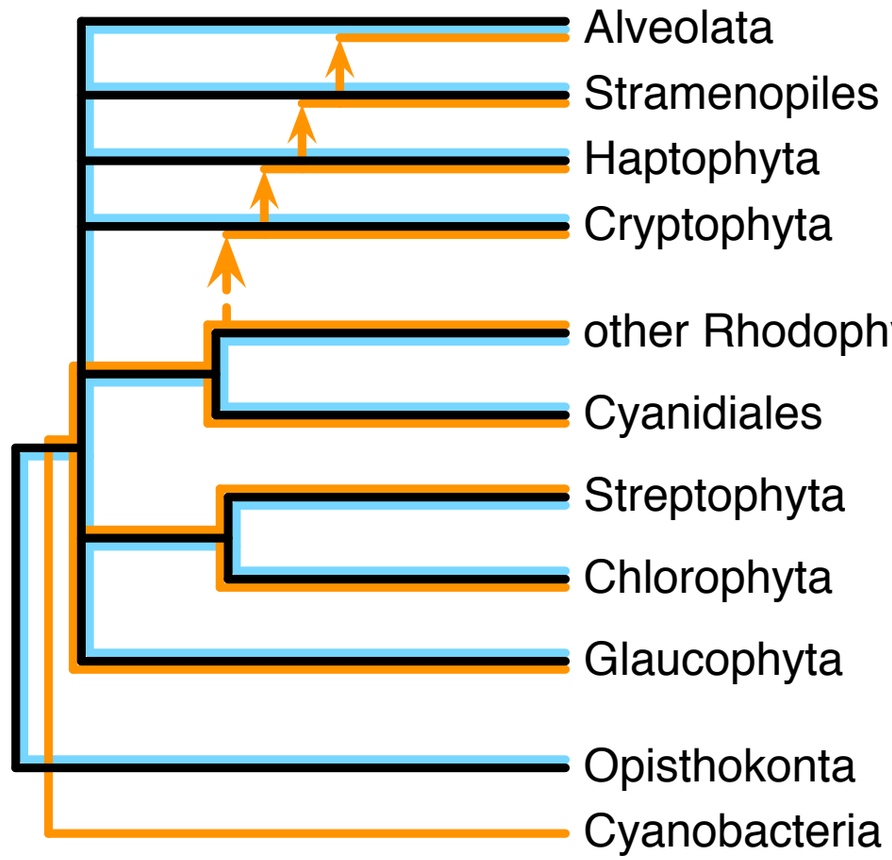
How to test the Chromalveolate hypothesis?

1. *How old are chl. c-containing plastids?*
2. *How to compare the plastid with mitochondrial and nuclear compartments?*
3. *How to estimate the strength of the phylogenetic signal?*
4. *How to validate our approach?*
5. *Now do 'Chromists' pass our test?*

Chromalveolate Hypothesis



Alternative Hypothesis

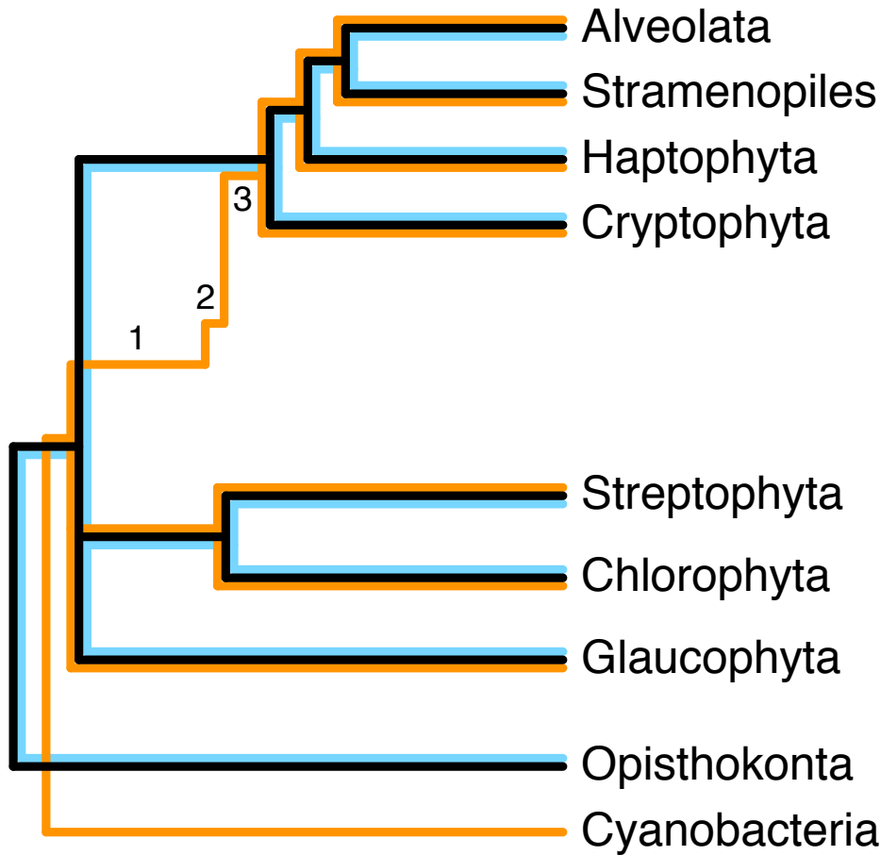


plastid / **mitochondrion** / **nucleus** histories

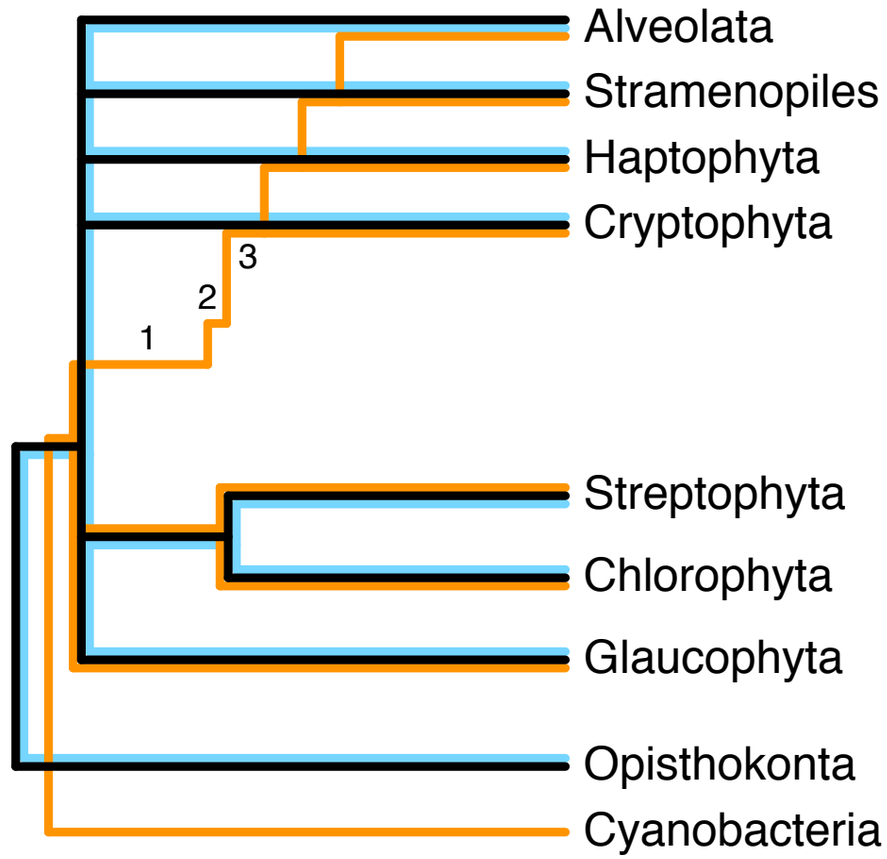
‘Chromalveolates’ are expected to be monophyletic whatever the genomic compartment considered.

Baurain et al. (2010) *Mol Biol Evol* (in press)

Chromalveolate Hypothesis



Alternative Hypothesis

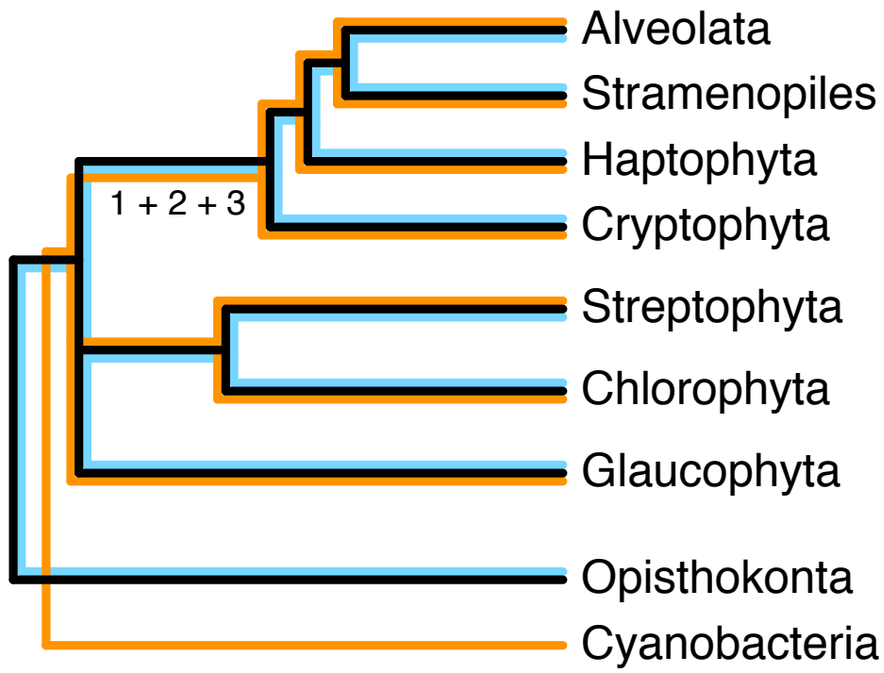


plastid / **mitochondrion** / **nucleus** histories

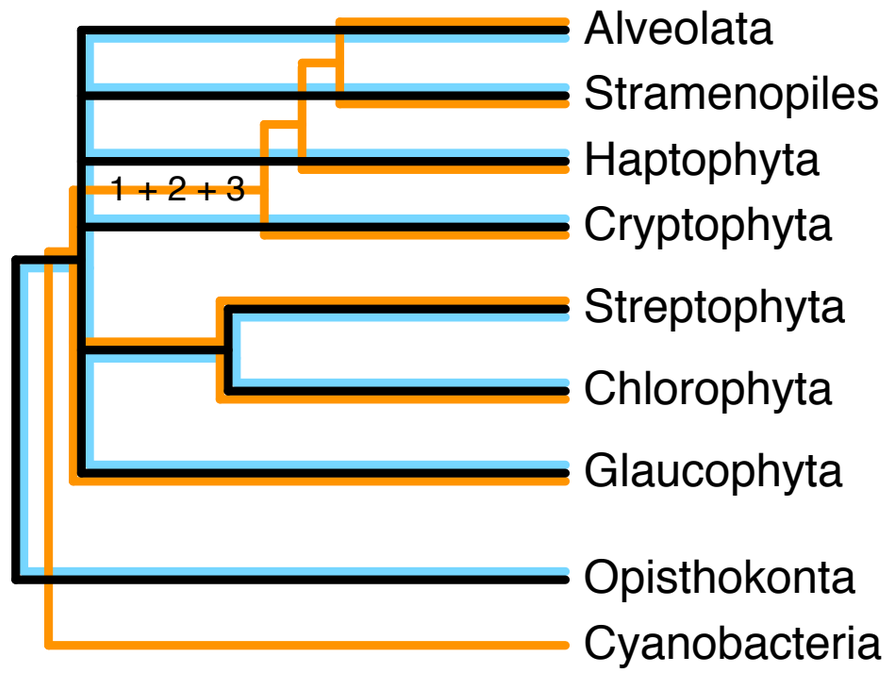
Removing red algae 'standardizes' the plastid history by creating a single branch out of three smaller ones.

Baurain et al. (2010) *Mol Biol Evol* (in press)

Chromalveolate Hypothesis



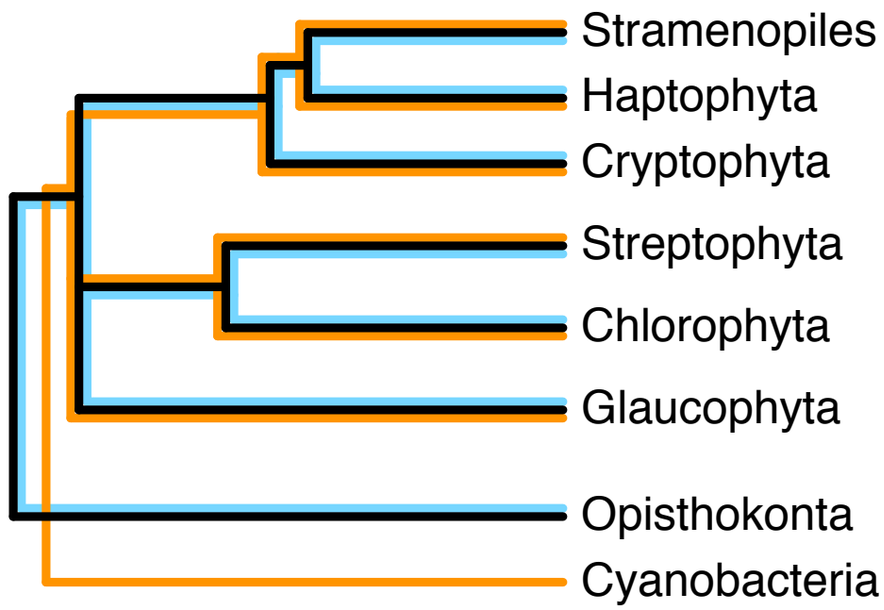
Alternative Hypothesis



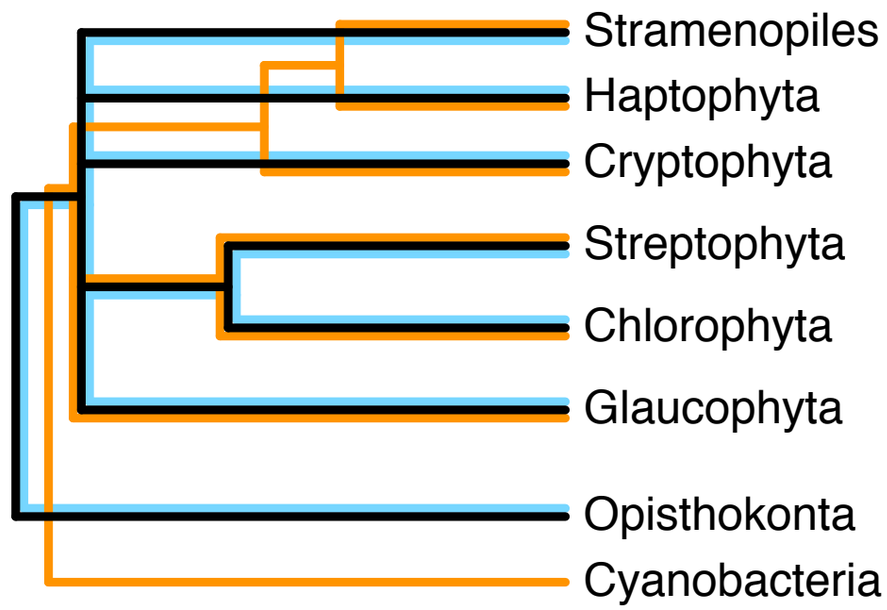
Without red algae, the signal for 'Chromalveolates' should be similarly strong across all compartments.

Baurain et al. (2010) *Mol Biol Evol* (in press)

Chromalveolate Hypothesis



Alternative Hypothesis



Alveolates (Dinoflagellates) have to be discarded because of their fast-evolving plastid genomes.

We test ‘Chromists’ as a proxy for ‘Chromalveolates’.

Baurain et al. (2010) *Mol Biol Evol* (in press)

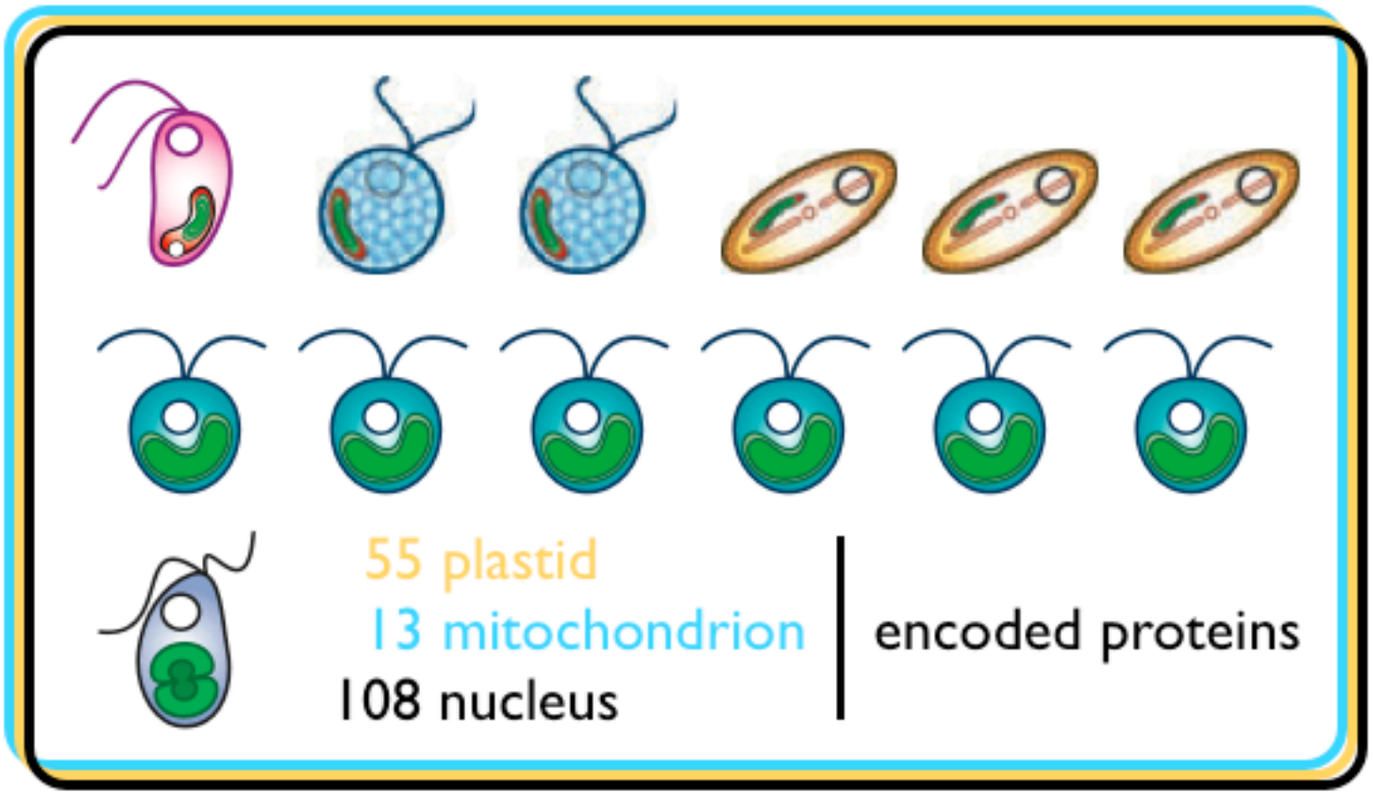
How to test the Chromalveolate hypothesis?

1. *How old are chl. c-containing plastids?*
2. *How to compare the plastid with mitochondrial and nuclear compartments?*
3. *How to estimate the strength of the phylogenetic signal?*
4. *How to validate our approach?*
5. *Now do 'Chromists' pass our test?*

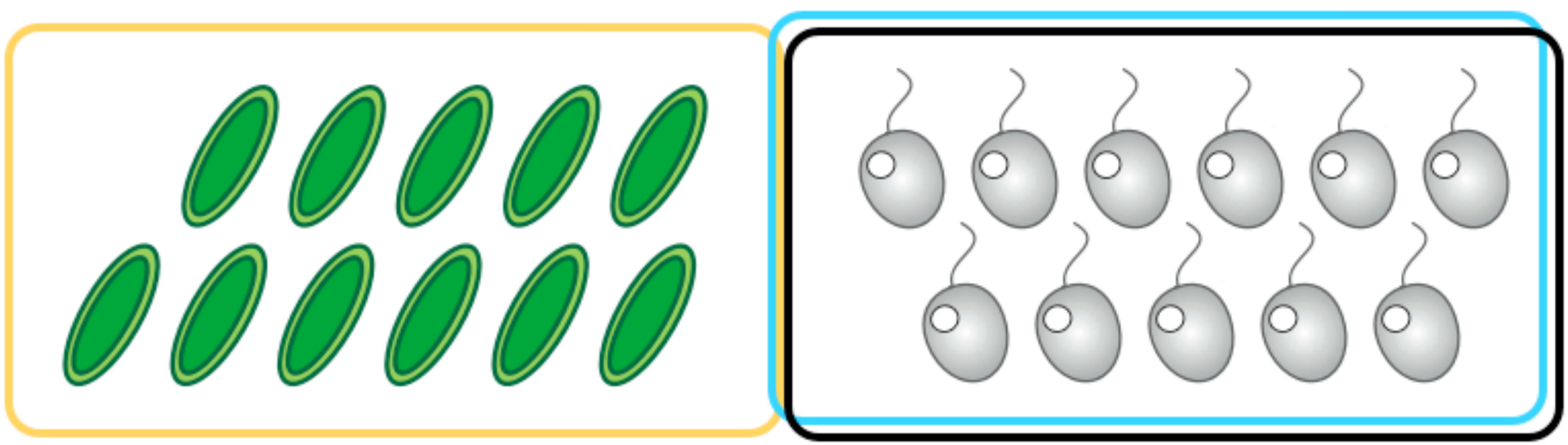
6 'Chromists'

6 green plants

1 glaucophyte



ingroup



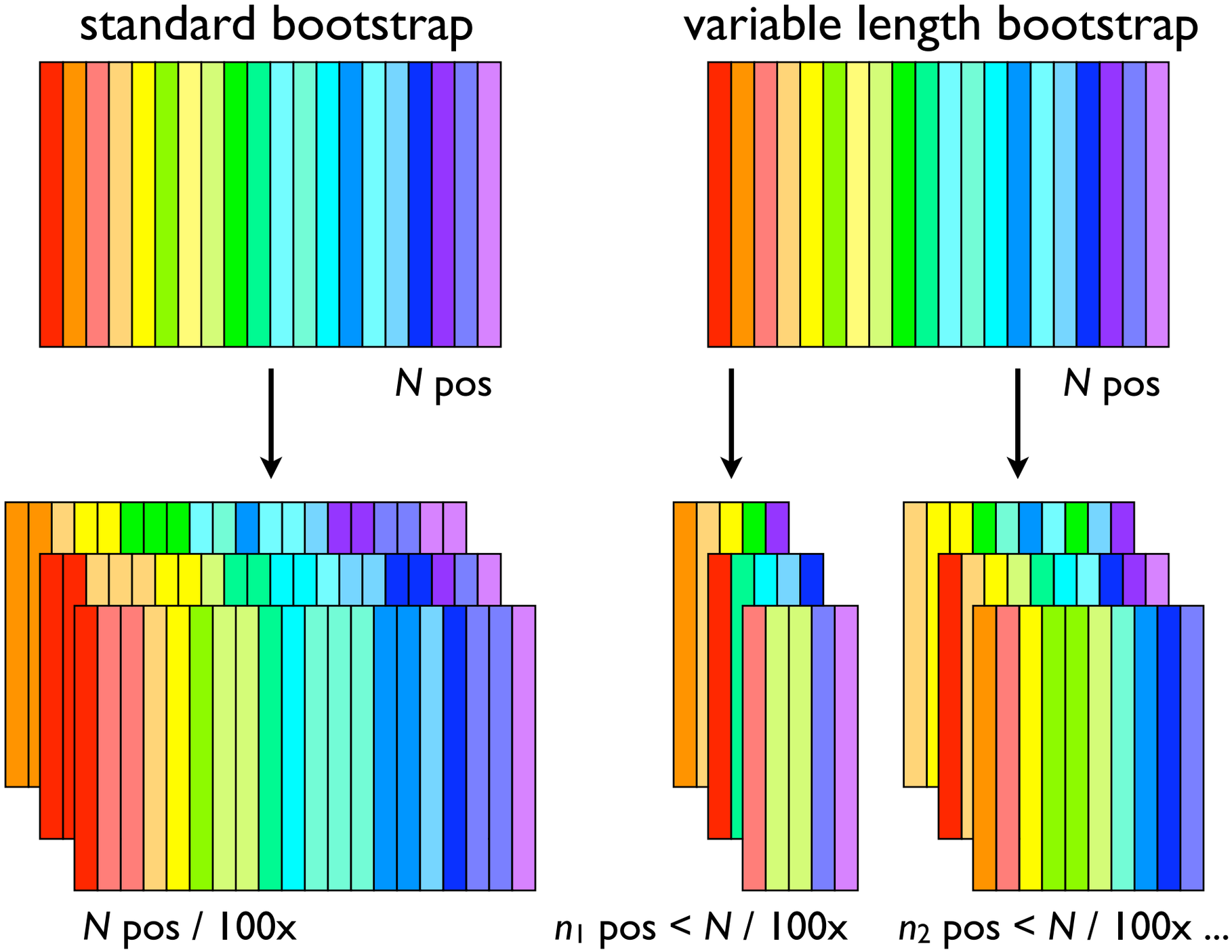
11 cyanobacteria

11 opisthokonts

outgroup

Using the same species sampling, we assemble one concatenated protein data set per compartment.

algal icons from Reyes-Prieto et al. (2007) *Annu Rev Genet* 41:147-168



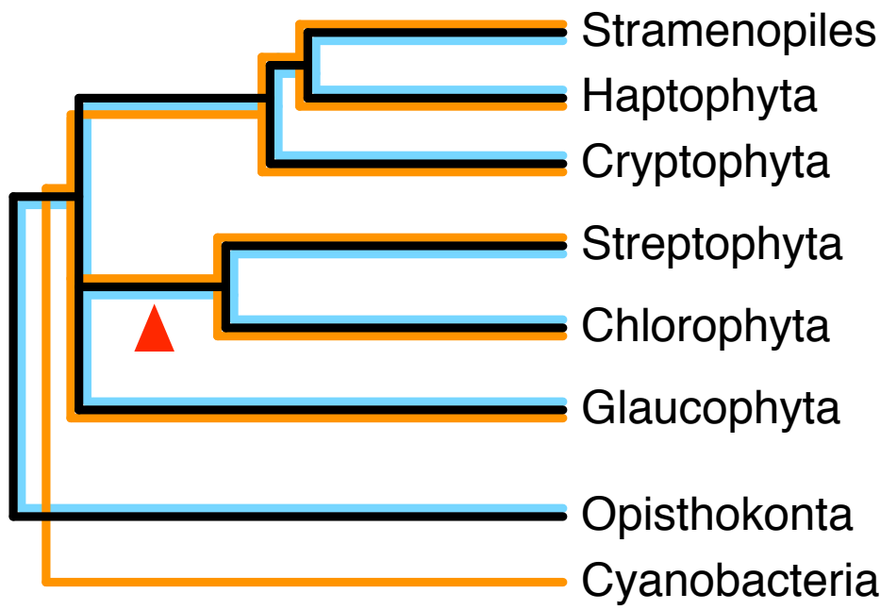
We will analyze these three phylogenomic data sets with a variable length bootstrap approach.

inspired from Springer et al. (2001) *Mol Biol Evol* 18:132-143

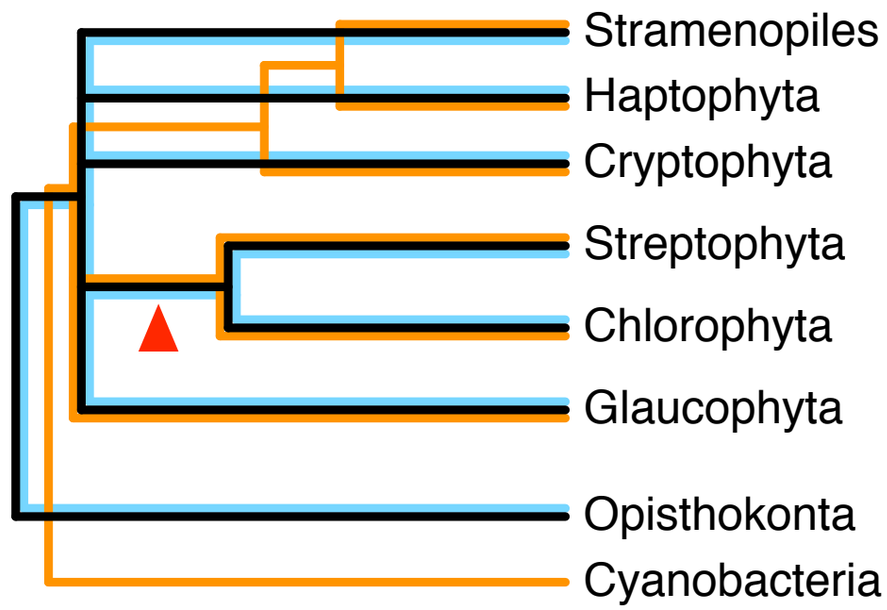
How to test the Chromalveolate hypothesis?

1. *How old are chl. c-containing plastids?*
2. *How to compare the plastid with mitochondrial and nuclear compartments?*
3. *How to estimate the strength of the phylogenetic signal?*
4. *How to validate our approach?*
5. *Now do 'Chromists' pass our test?*

Chromalveolate Hypothesis



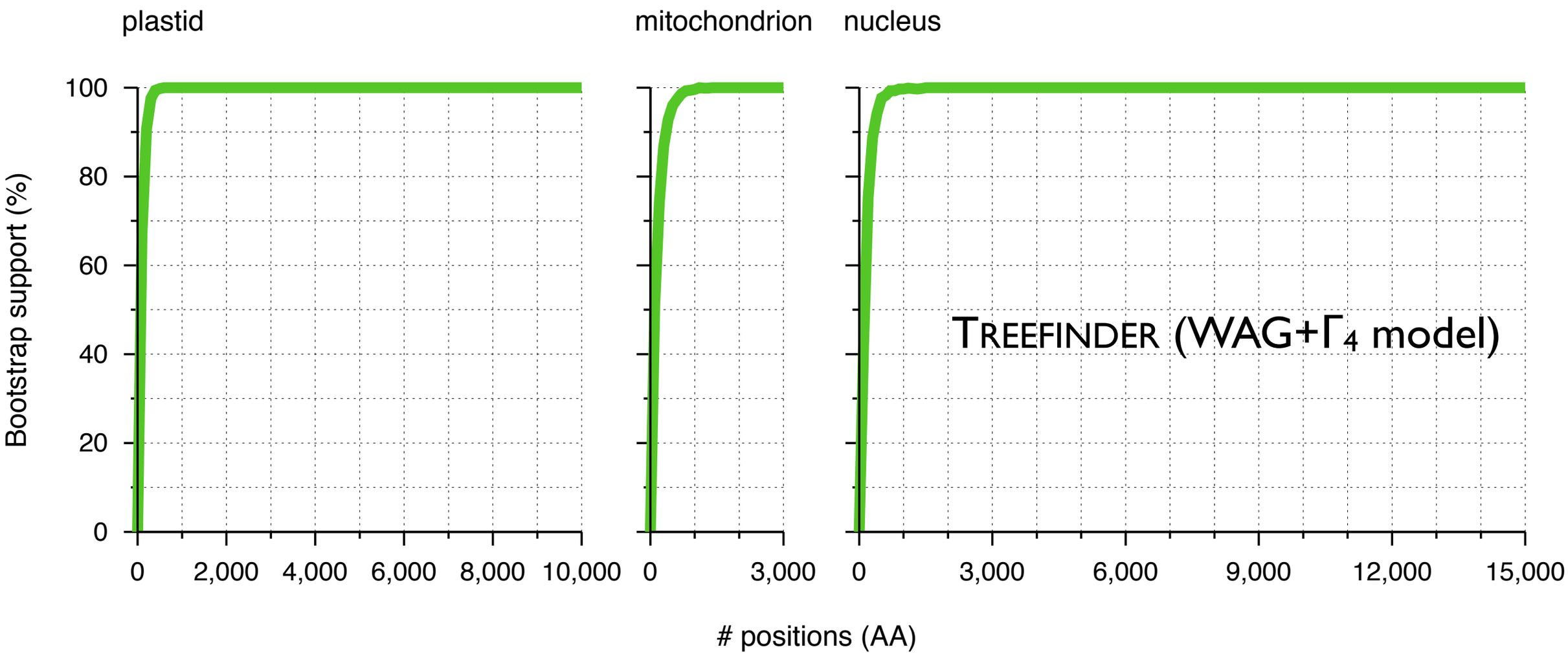
Alternative Hypothesis



We will use green plants as a test case.

Whatever the hypothesis or the compartment, their monophyly is expected to be easy to recover.

Baurain et al. (2010) *Mol Biol Evol* (in press)



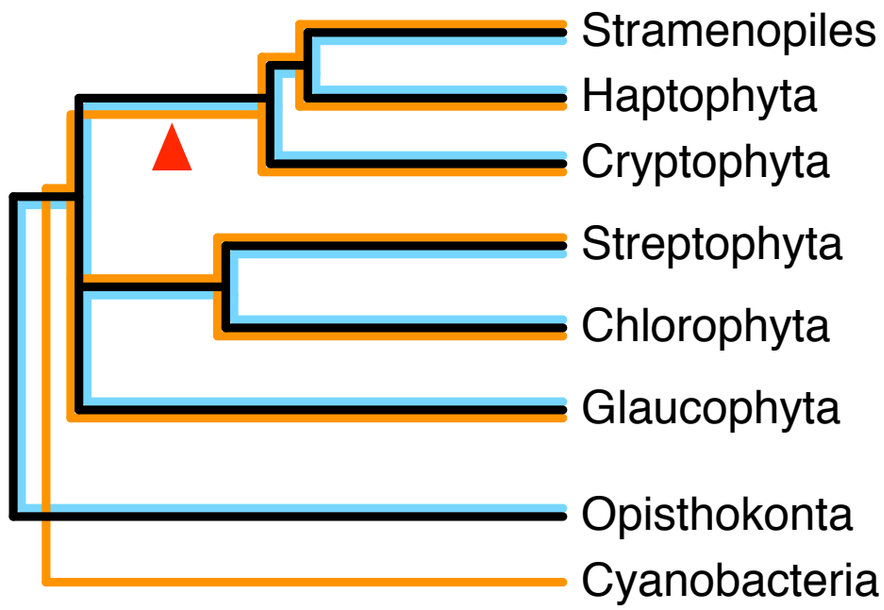
Whatever the genomic compartment considered, the monophyly of green plants is easily recovered.

Baurain et al. (2010) *Mol Biol Evol* (in press)

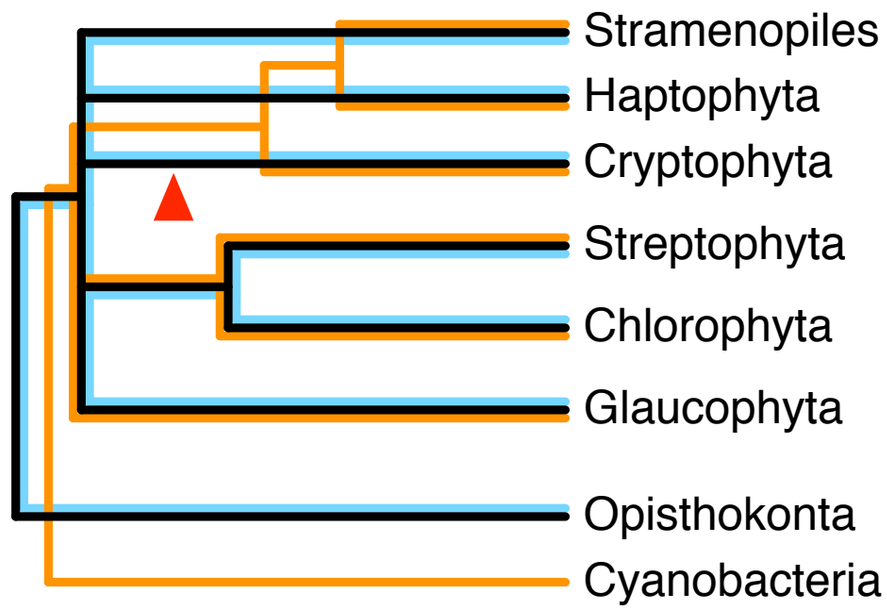
How to test the Chromalveolate hypothesis?

1. *How old are chl. c-containing plastids?*
2. *How to compare the plastid with mitochondrial and nuclear compartments?*
3. *How to estimate the strength of the phylogenetic signal?*
4. *How to validate our approach?*
5. *Now do 'Chromists' pass our test?*

Chromalveolate Hypothesis



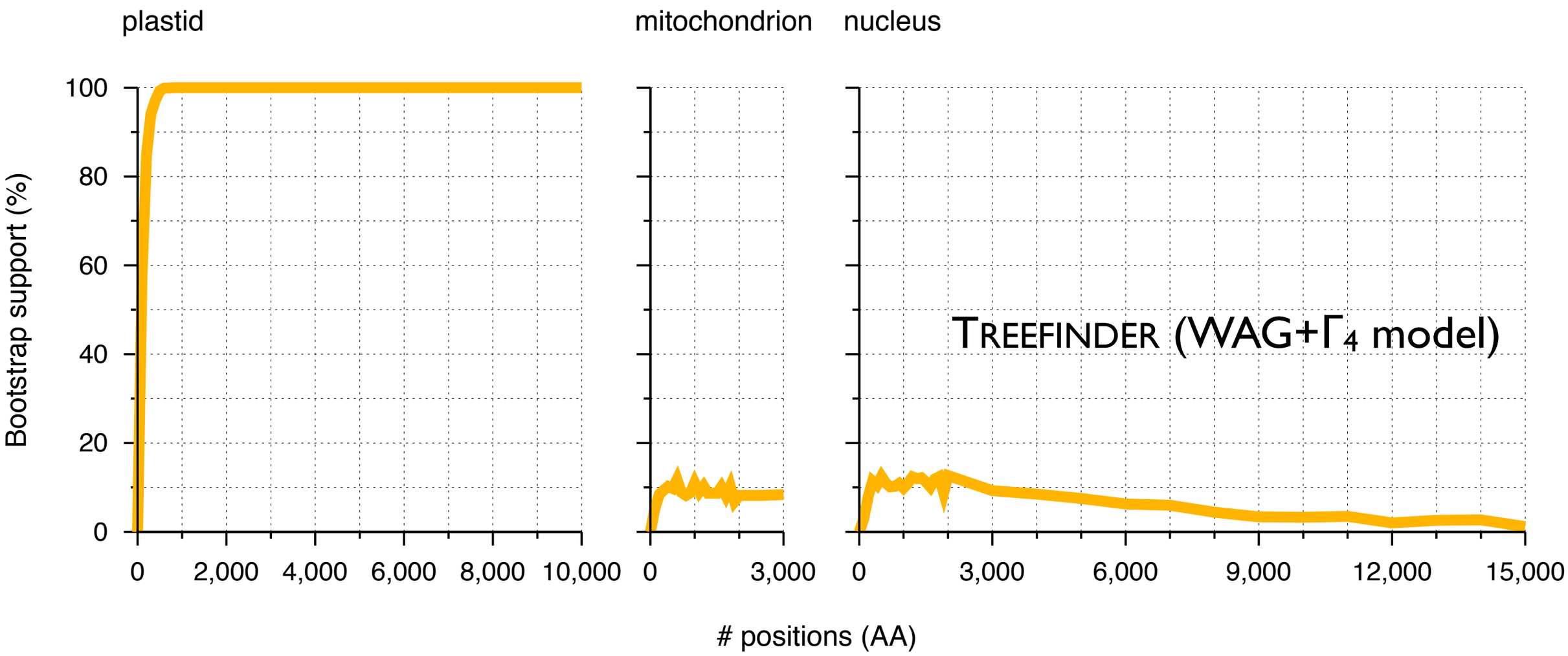
Alternative Hypothesis



If 'Chromists' exist, their monophyly is expected to be easy to recover whatever the compartment.

Alternatively, their monophyly is expected to be easy to recover only with the plastid.

Baurain et al. (2010) *Mol Biol Evol* (in press)



The monophyly of 'Chromists' is only recovered with plastid genomes.

Baurain et al. (2010) *Mol Biol Evol* (in press)

How to test the Chromalveolate hypothesis?

When not considering red algae, the phylogenetic signal for the monophyly of 'Chromalveolates' is expected to be similarly strong across plastid, mitochondrial and nuclear compartments.

Our variable length bootstrap analyses show that this prediction of the 'Chromalveolate hypothesis' is not fulfilled whereas it is for green plants.

How to check the assumptions of our test?

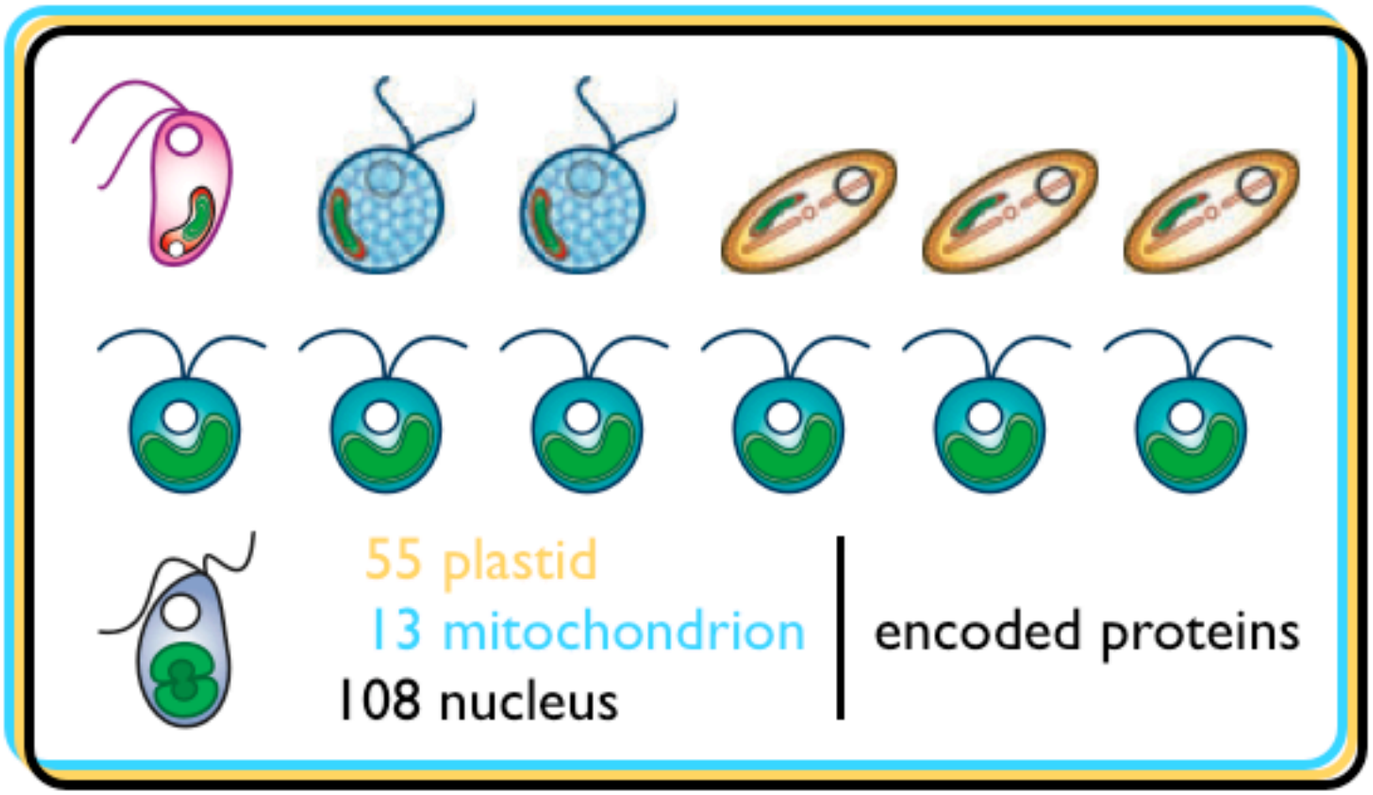
How to check the assumptions of our test?

1. *What if Plantae are actually paraphyletic?*
2. *How to deal with heterogeneous rates?*
3. *Do we have enough phylogenetic power?*
4. *Are we misled by phylogenetic artifacts?*
5. *Are we misled by undetected EGT?*

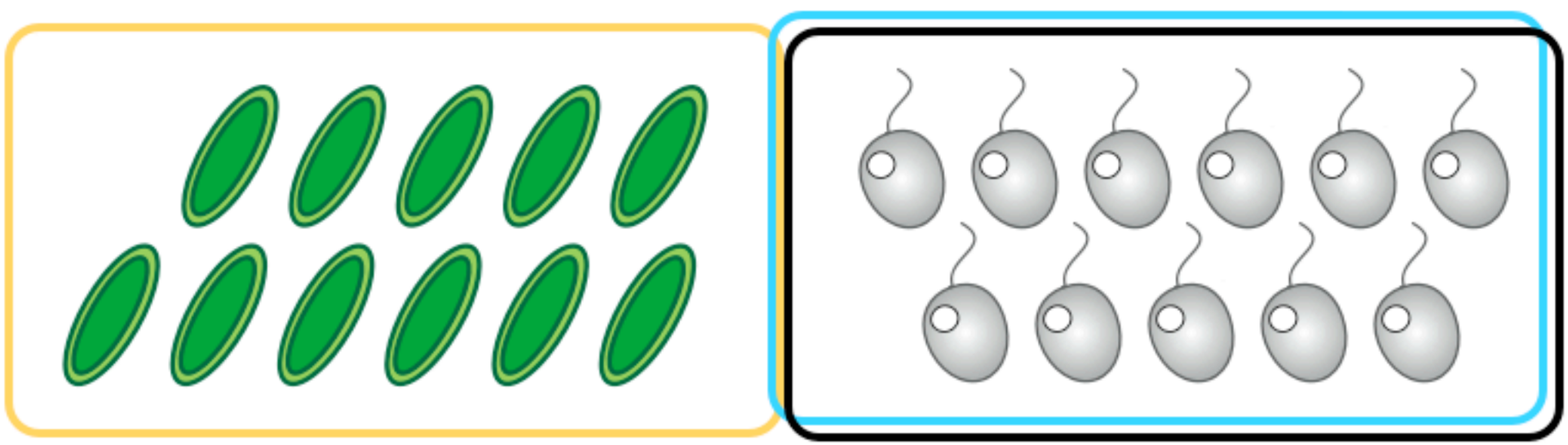
6 'Chromists'

6 green plants

1 glaucophyte



ingroup



11 cyanobacteria

11 opisthokonts

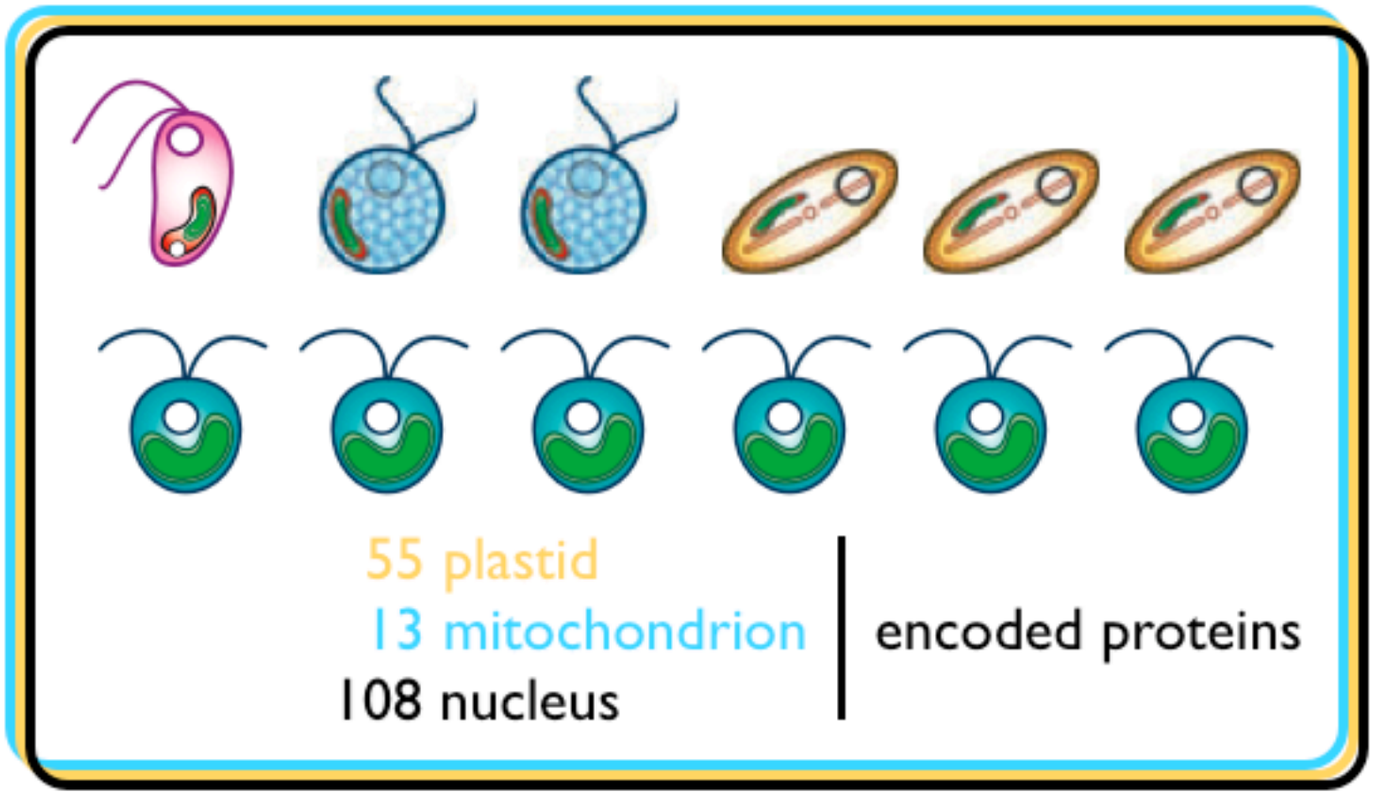
outgroup

Removing glaucophytes reduces Plantae to green plants.

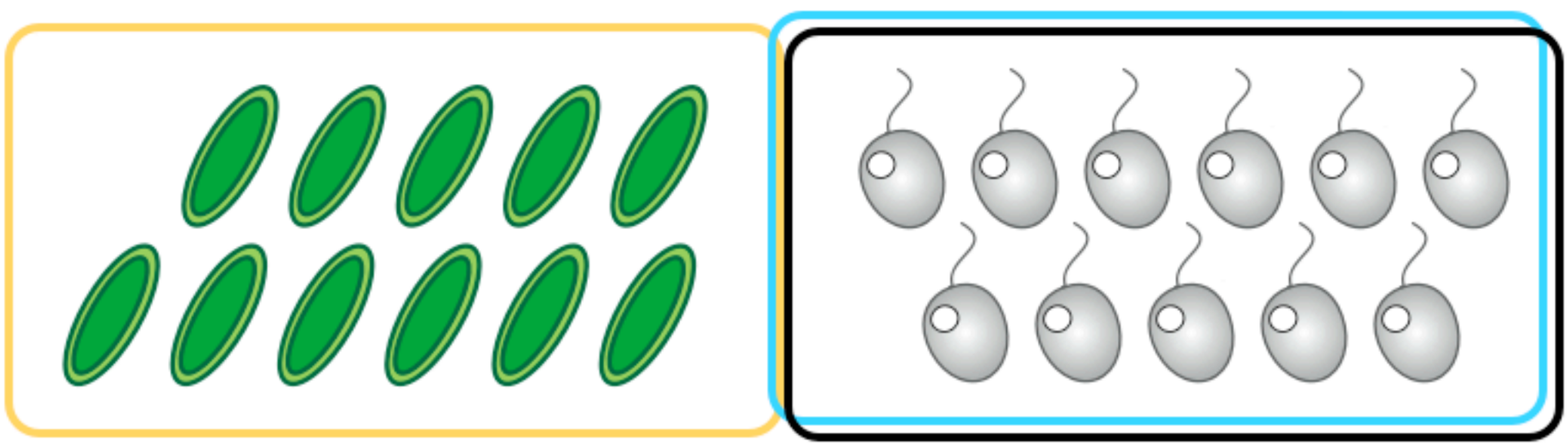
algal icons from Reyes-Prieto et al. (2007) *Annu Rev Genet* 41:147-168

6 'Chromists'

6 green plants



ingroup



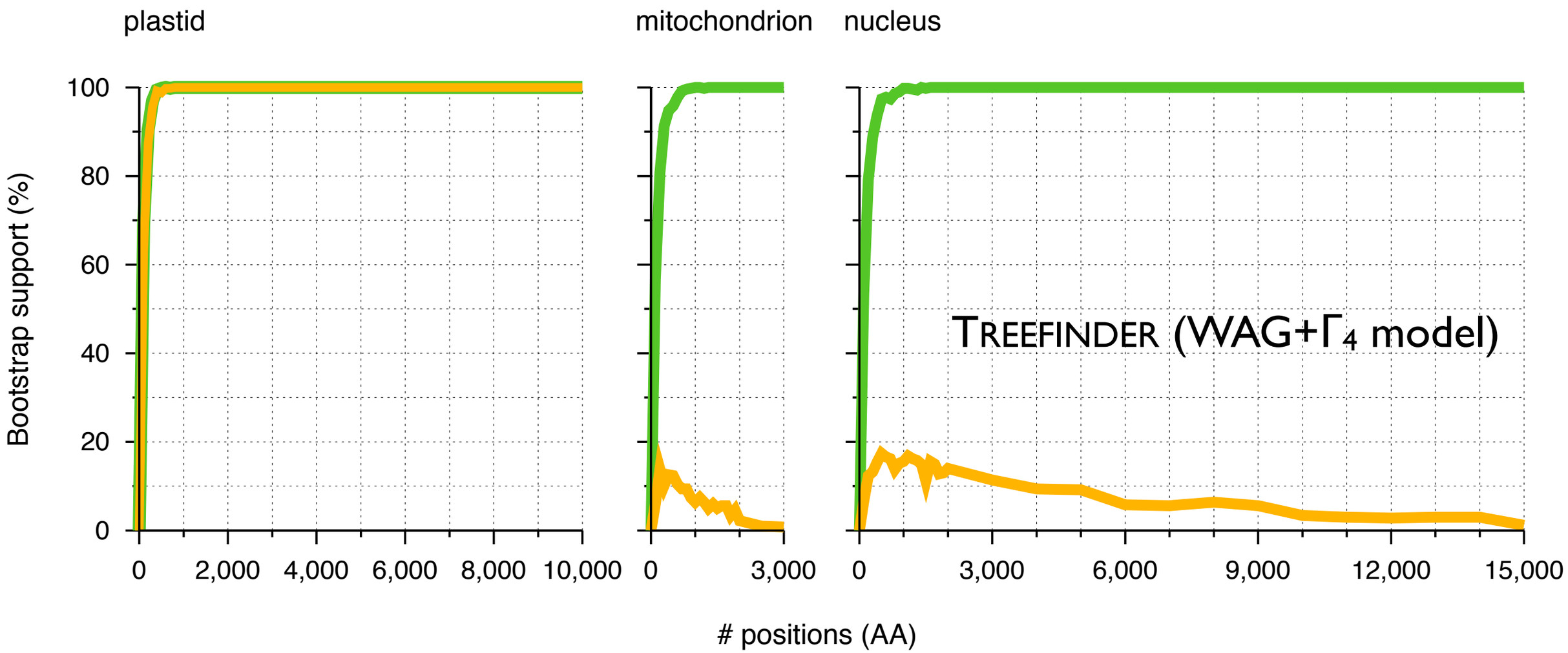
11 cyanobacteria

11 opisthokonts

outgroup

Removing glaucophytes reduces Plantae to green plants.

algal icons from Reyes-Prieto et al. (2007) *Annu Rev Genet* 41:147-168



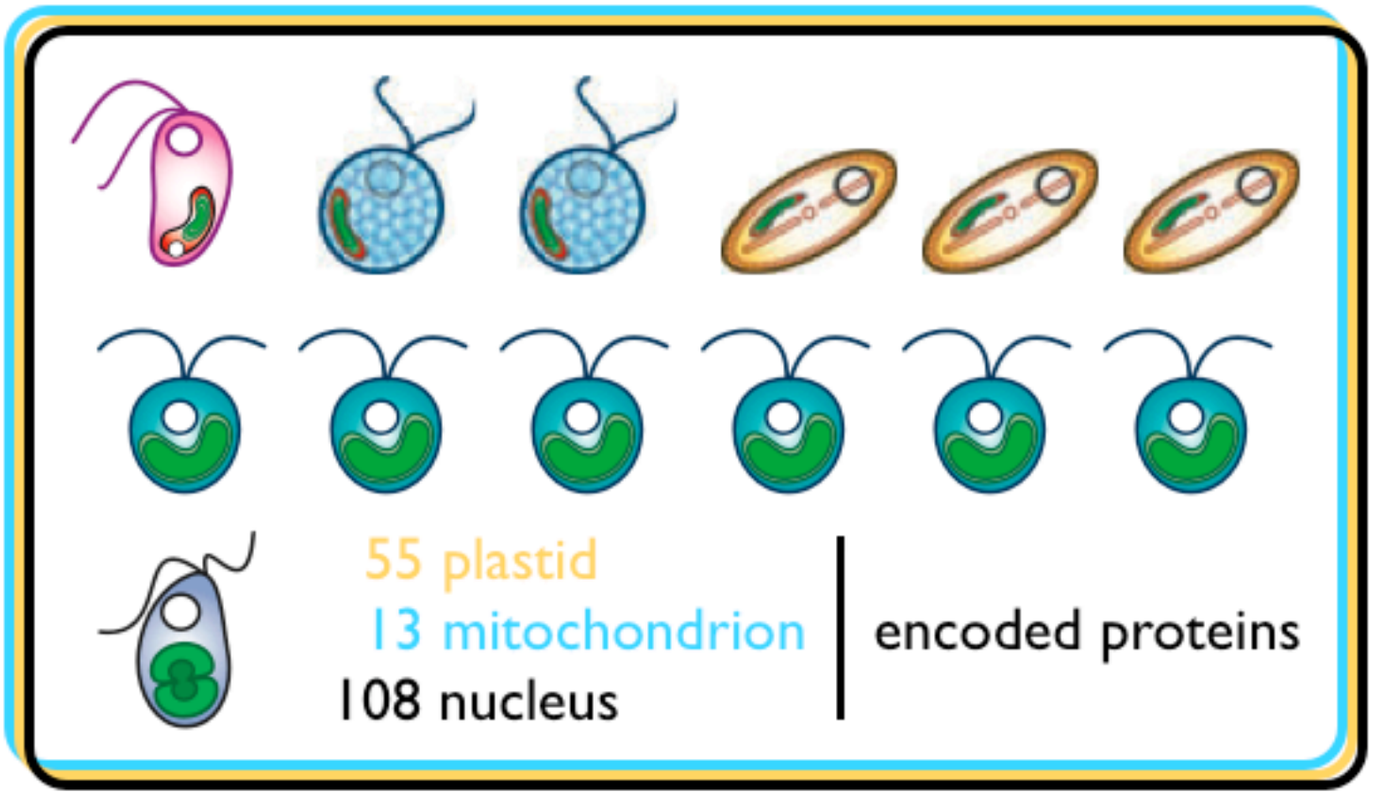
Using only green plants as Plantae does not improve the recovery of the monophyly of 'Chromists'.

Baurain et al. (2010) *Mol Biol Evol* (in press)

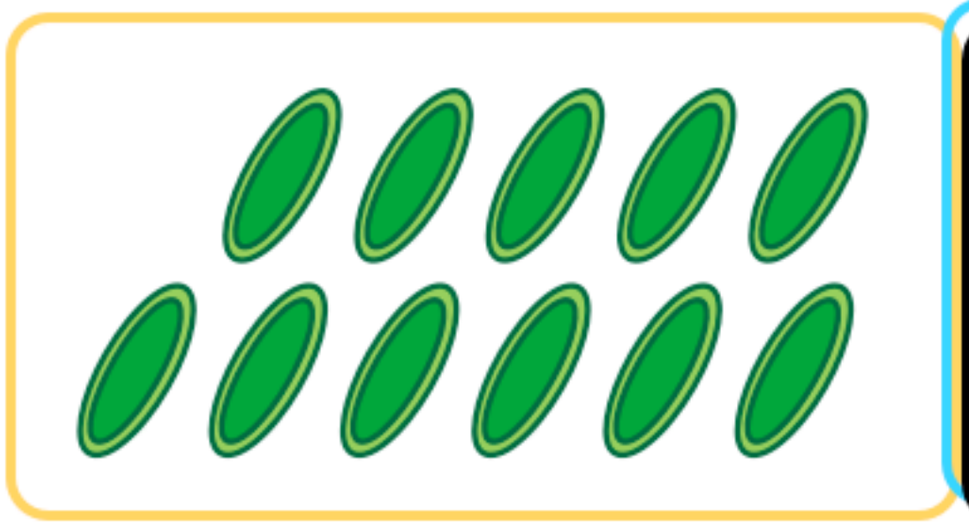
6 'Chromists'

6 green plants

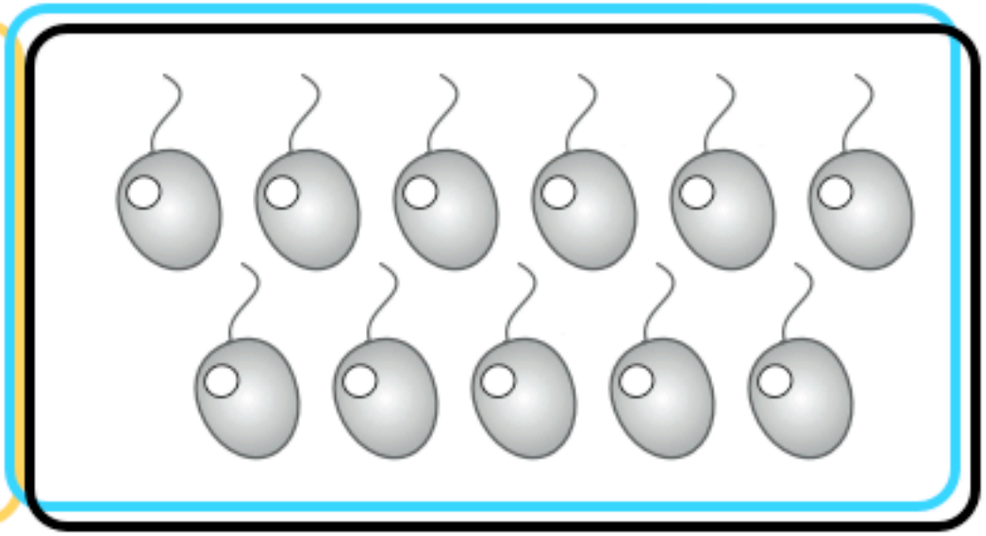
1 glaucophyte



ingroup



11 cyanobacteria



11 opisthokonts

outgroup

Removing green plants reduces Plantae to glaucophytes.

algal icons from Reyes-Prieto et al. (2007) *Annu Rev Genet* 41:147-168

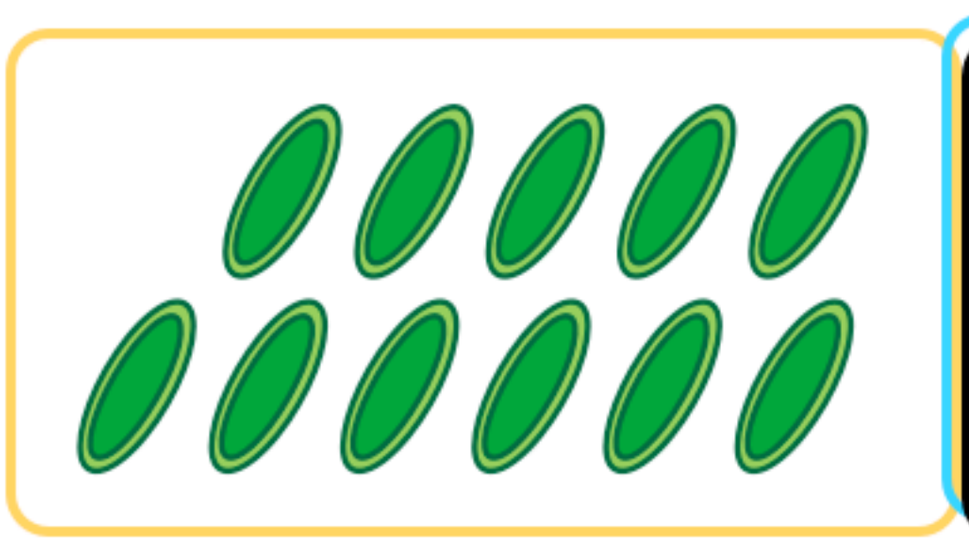
6 ‘Chromists’

1 glaucophyte

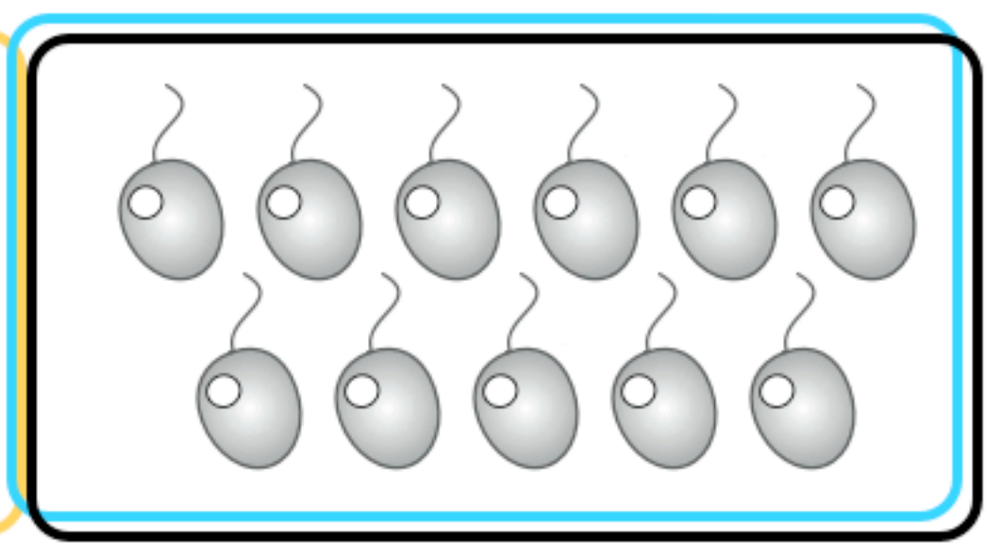
55 plastid
13 mitochondrion
108 nucleus

encoded proteins

ingroup



11 cyanobacteria

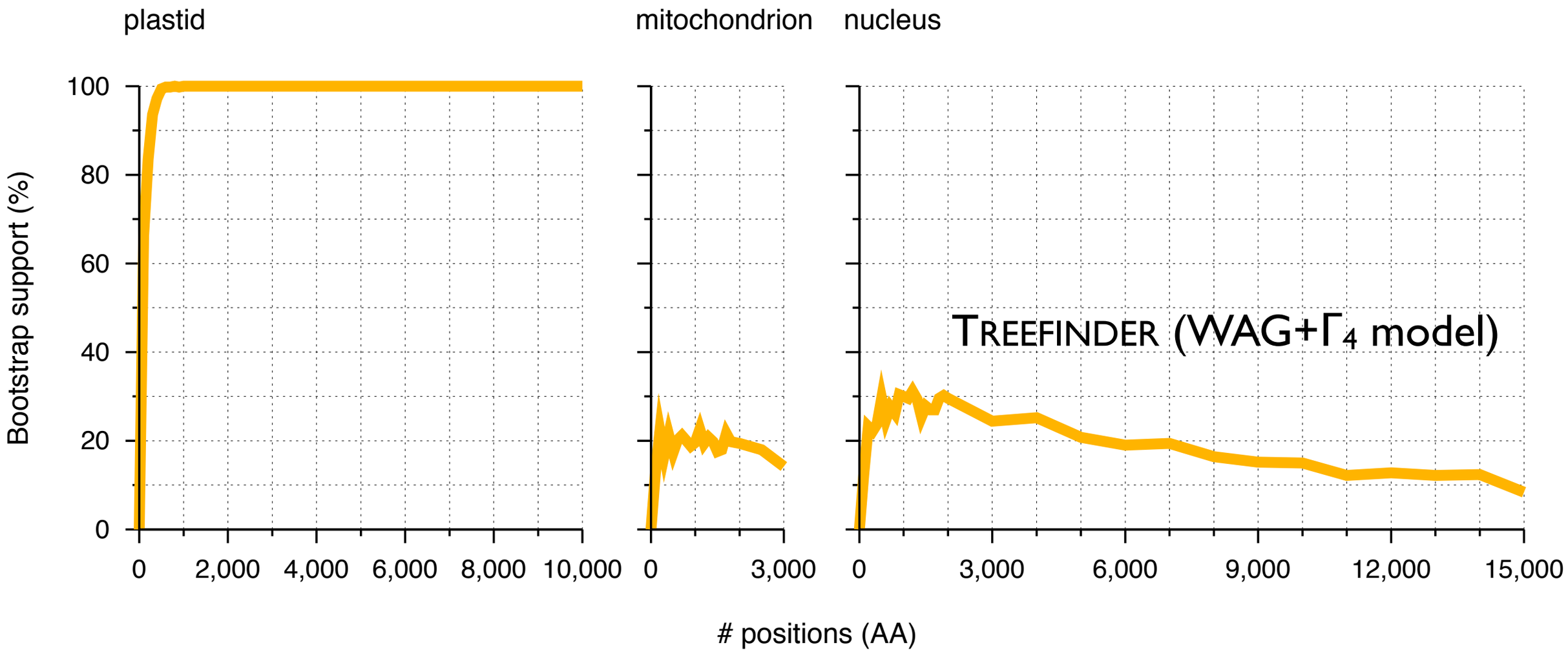


11 opisthokonts

outgroup

Removing green plants reduces Plantae to glaucophytes.

algal icons from Reyes-Prieto et al. (2007) *Annu Rev Genet* 41:147-168

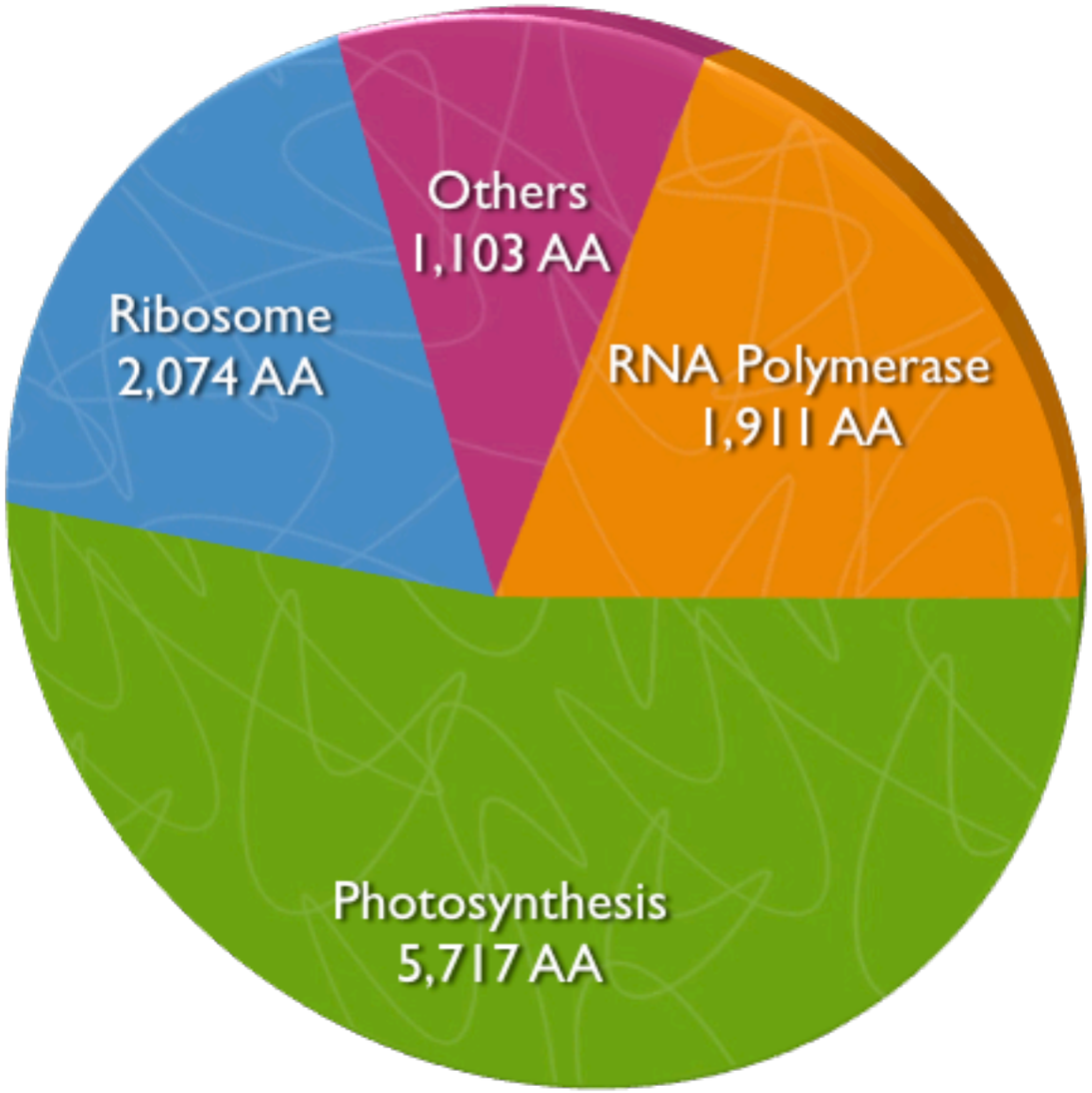


Using only glaucophytes as Plantae does not improve the recovery of the monophyly of 'Chromists'.

Baurain et al. (2010) *Mol Biol Evol* (in press)

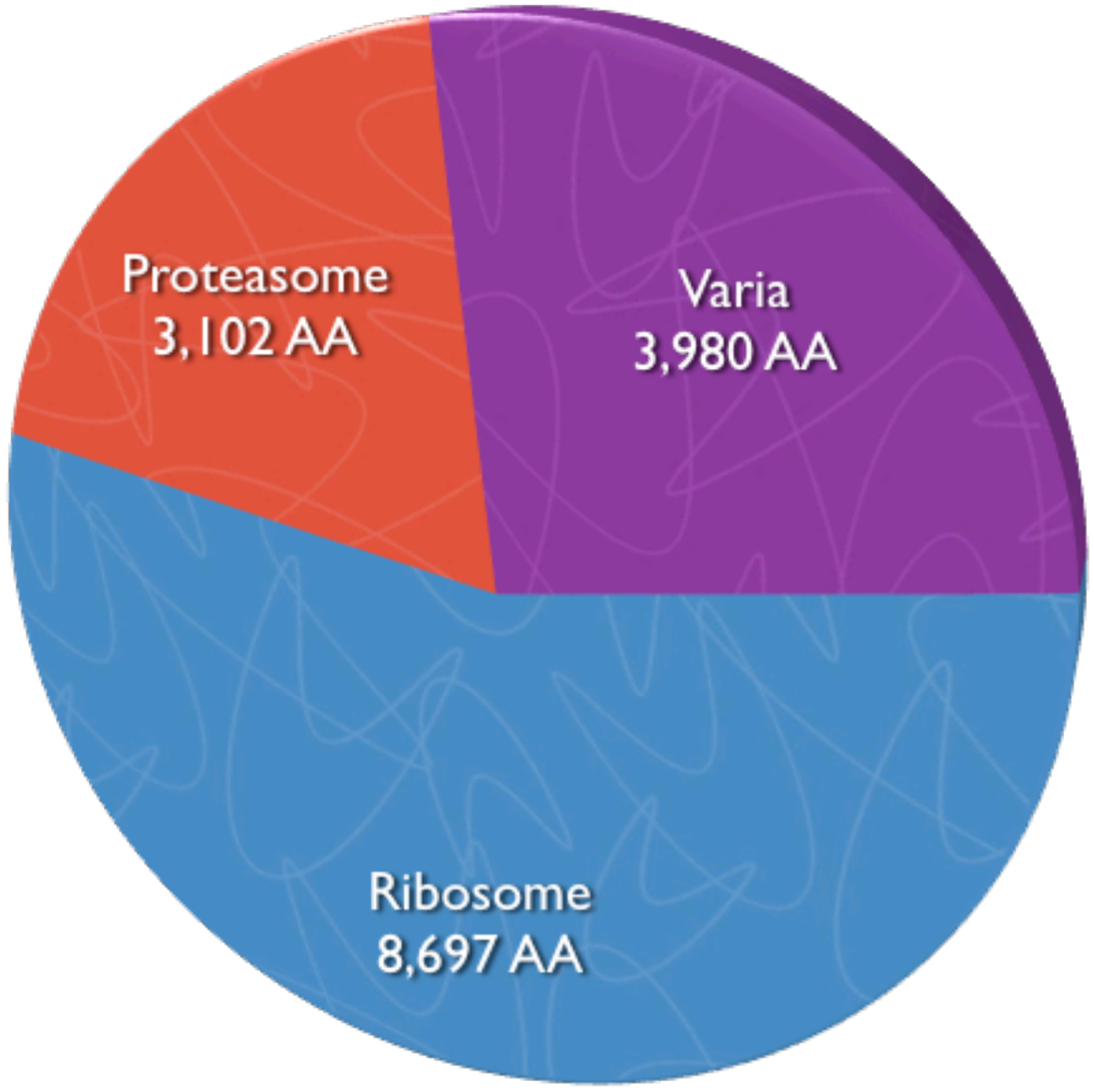
How to check the assumptions of our test?

1. *What if Plantae are actually paraphyletic?*
2. *How to deal with heterogeneous rates?*
3. *Do we have enough phylogenetic power?*
4. *Are we misled by phylogenetic artifacts?*
5. *Are we misled by undetected EGT?*



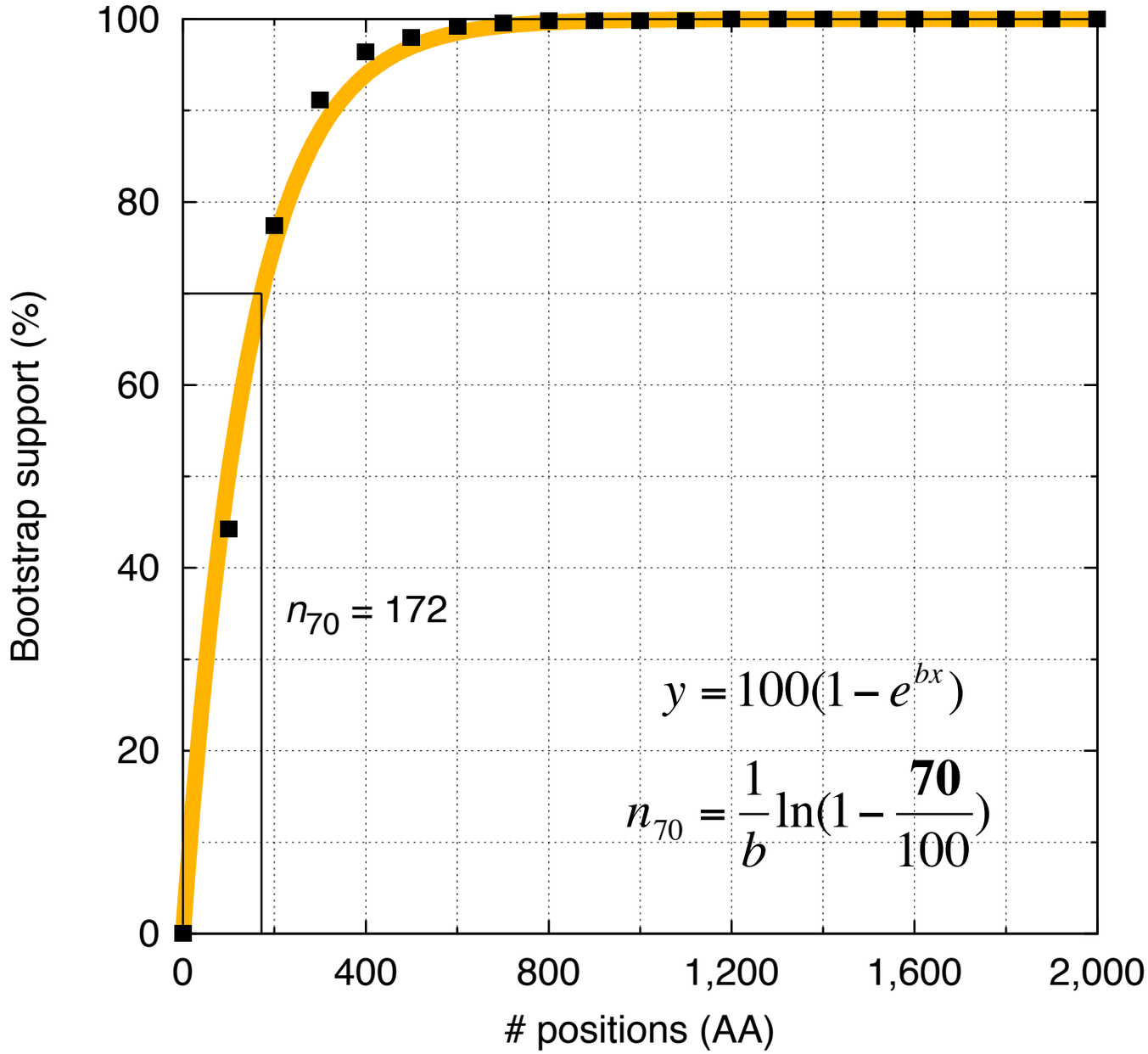
Plastid (10,805 AA)

Nucleus (15,392 AA)



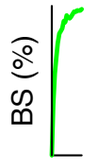
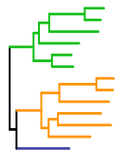
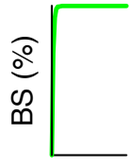
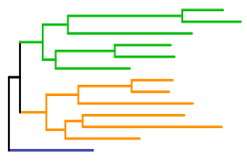
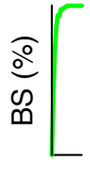
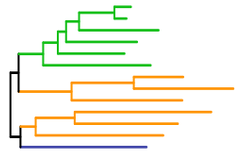
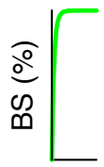
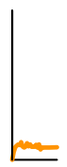
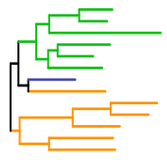
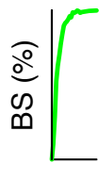
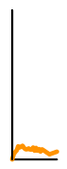
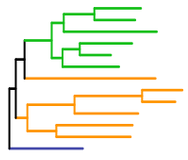
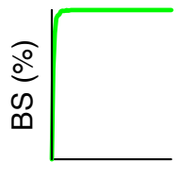
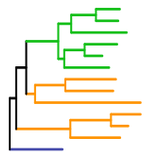
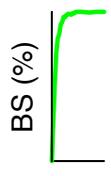
We split each of our plastid and nuclear data sets into smaller data sets according to functional class.

Baurain et al. (2010) *Mol Biol Evol* (in press)

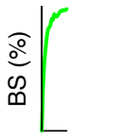
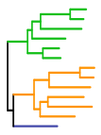
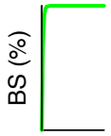
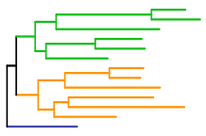
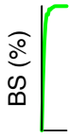
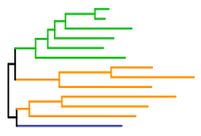
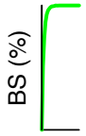
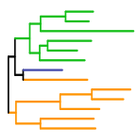
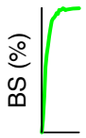
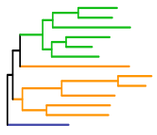
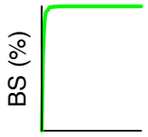
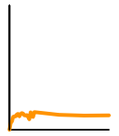
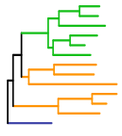
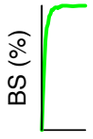


n_{70} is the number of positions required to reach a support ≥ 70% for the monophyly of the tested group.

inspired from Lecointre et al. (1994) *Mol Phylo Evol* 3:292-309

	# Pos	f(Sub)	Trees	VLBs		n_{70} Values	
				Greens	'Chromists'	Gre	'Chr'
Plastid							
Polymerase	**1,911**	**4.79**				**365**	**1,802**
Photosynthesis	**5,717**	**1.63**				**101**	**95**
Ribosome	**2,074**	**3.56**				**208**	**172**
Mitochondrion	**3,106**	**3.51**				**176**	**n.c.**
Nucleus							
Proteasome	**3,102**	**2.34**				**501**	**n.c.**
Ribosome	**8,697**	**2.72**				**130**	**n.c.**
Varia	**3,980**	**1.99**				**305**	**n.c.**

Baurain et al. (2010) *Mol Biol Evol* (in press)

	# Pos	f(Sub)	Trees	VLBs		n_{70} Values	
				Greens	'Chromists'	Gre	'Chr'
Plastid							
Polymerase	**1,911**	**4.79**				**365**	**1,802**
Photosynthesis	**5,717**	**1.63**				**101**	**95**
Ribosome	**2,074**	**3.56**				**208**	**172**
Mitochondrion	**3,106**	**3.51**				**176**	**n.c.**
Nucleus							
Proteasome	**3,102**	**2.34**				**501**	**n.c.**
Ribosome	**8,697**	**2.72**				**130**	**n.c.**
Varia	**3,980**	**1.99**				**305**	**n.c.**

Data sets split into functional classes indicate that rate heterogeneity does not impair our test.

Baurain et al. (2010) *Mol Biol Evol* (in press)

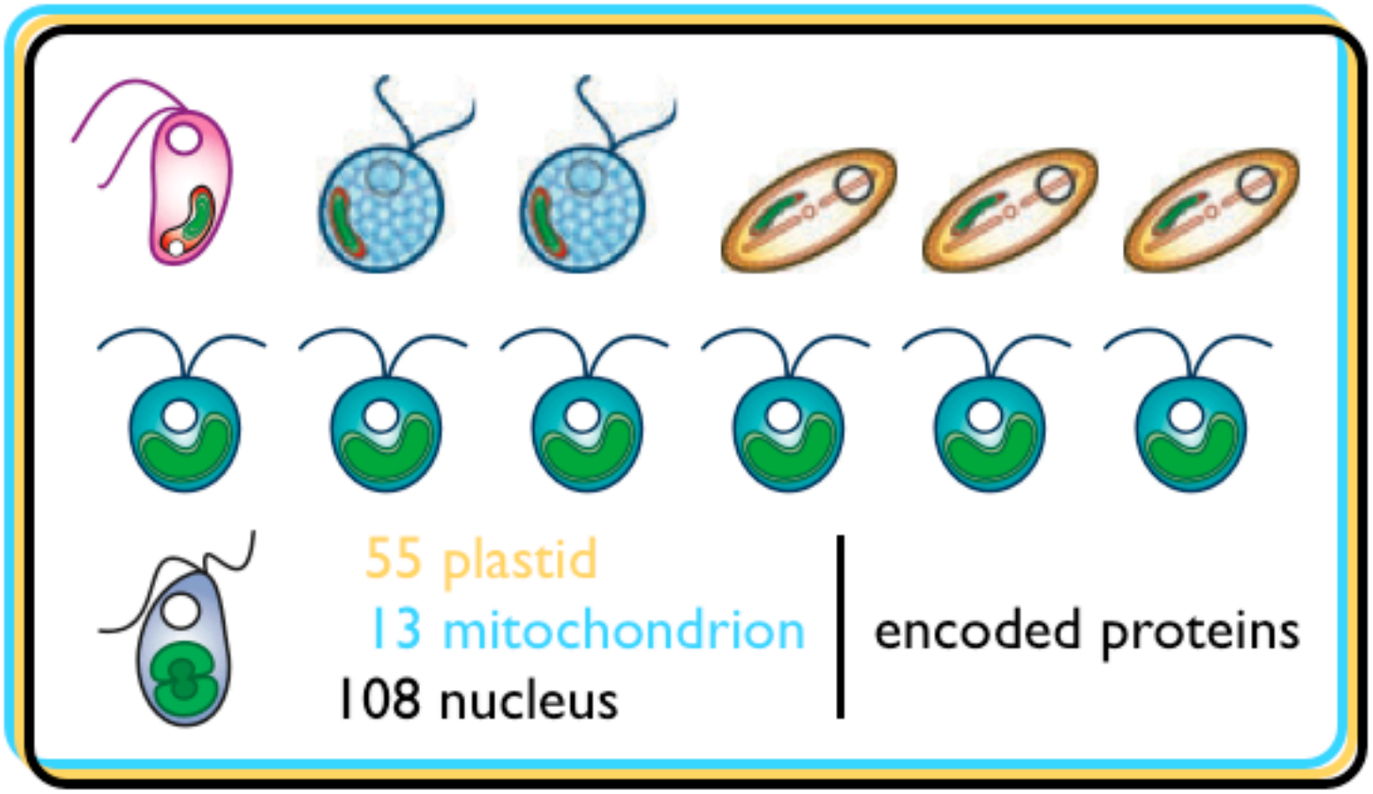
How to check the assumptions of our test?

1. *What if Plantae are actually paraphyletic?*
2. *How to deal with heterogeneous rates?*
3. *Do we have enough phylogenetic power?*
4. *Are we misled by phylogenetic artifacts?*
5. *Are we misled by undetected EGT?*

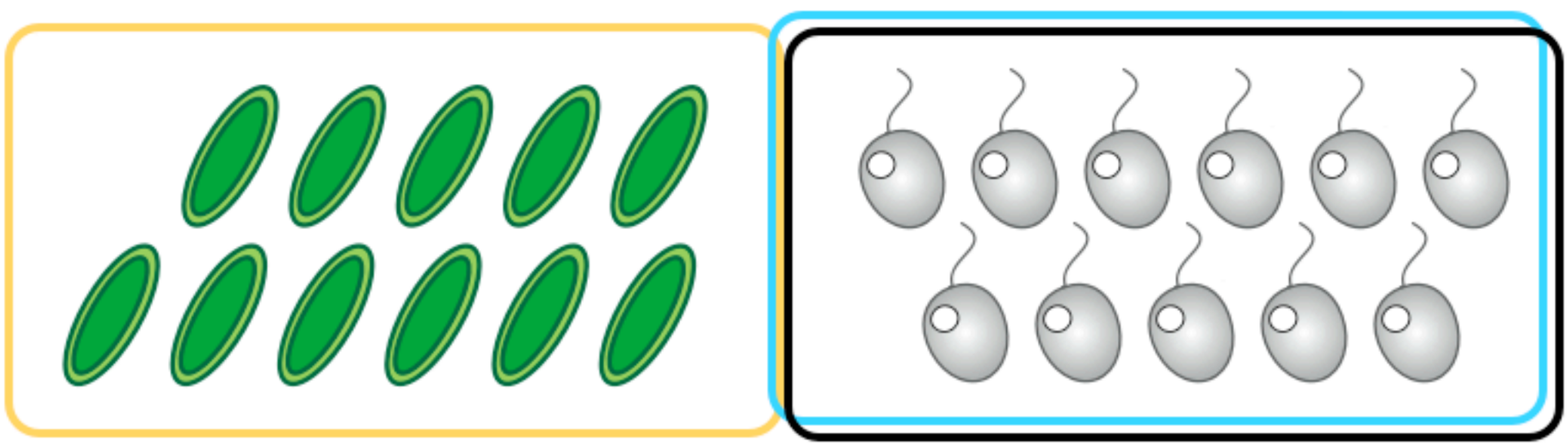
6 'Chromists'

6 green plants

1 glaucophyte



ingroup



11 cyanobacteria

11 opisthokonts

outgroup

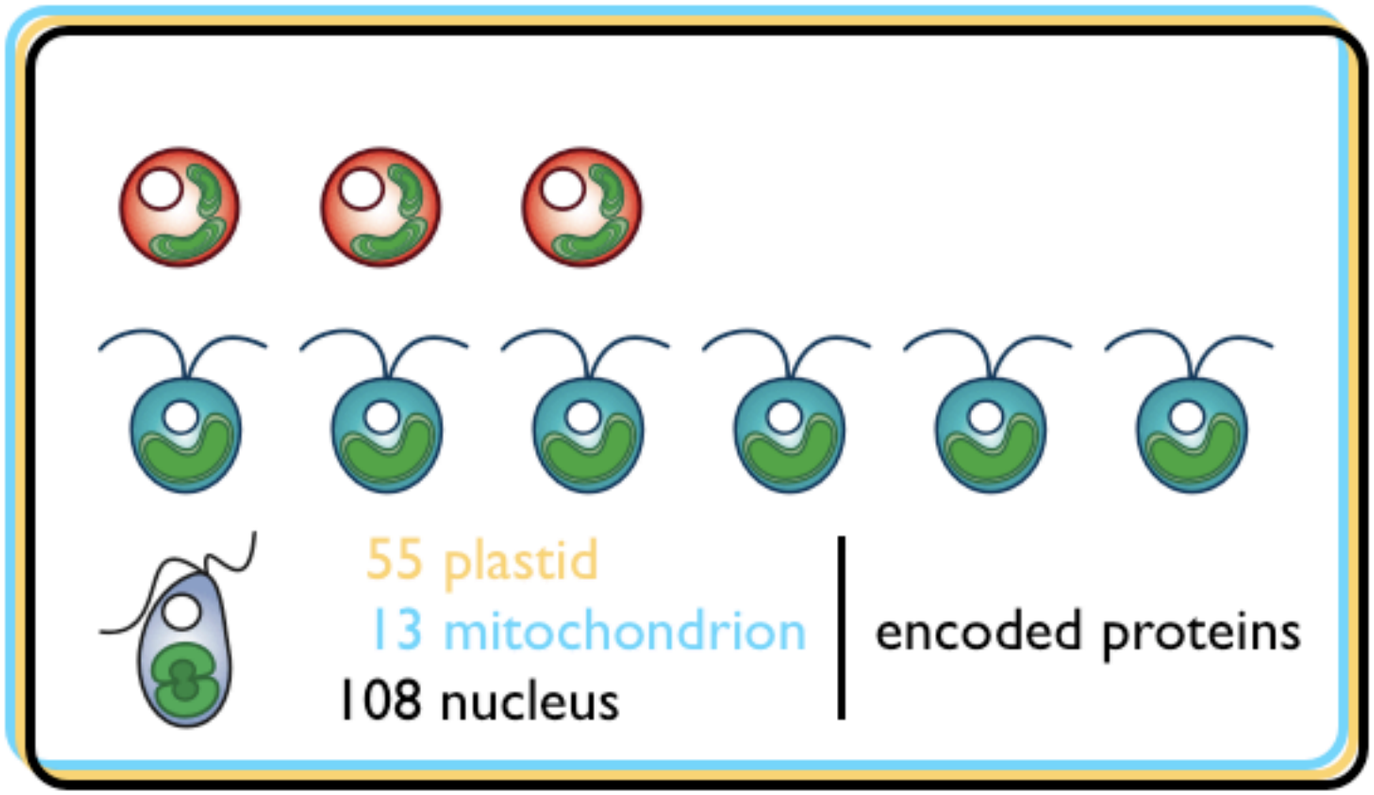
We replace 'Chromists' by red algae.

algal icons from Reyes-Prieto et al. (2007) *Annu Rev Genet* 41:147-168

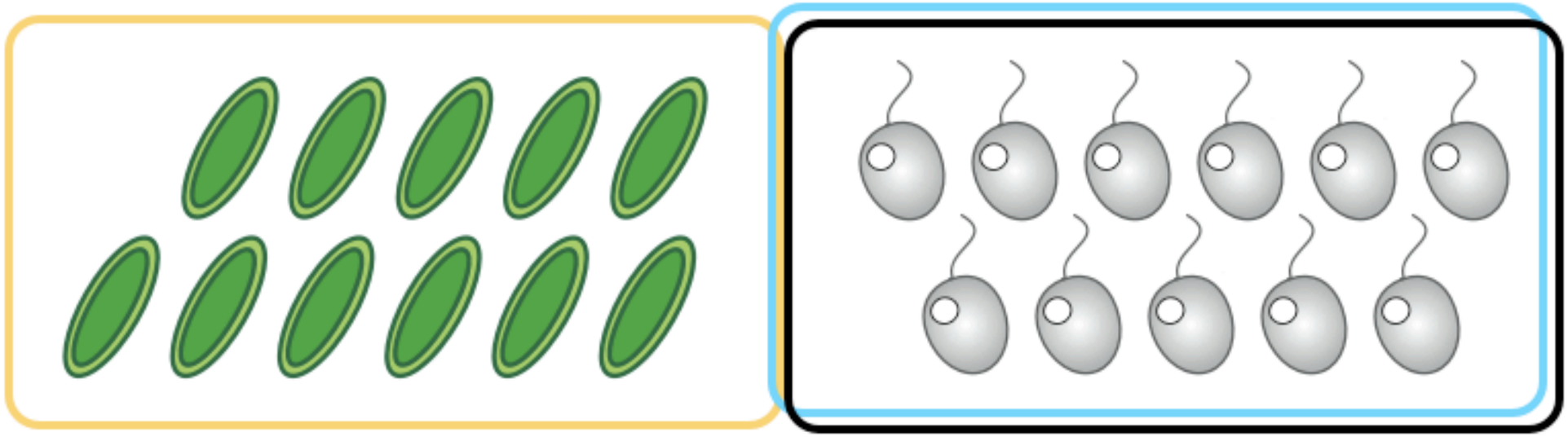
3 red algae

6 green plants

1 glaucophyte



ingroup



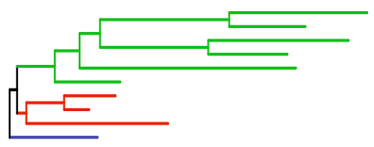
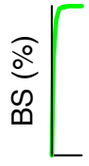
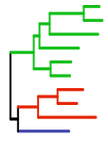
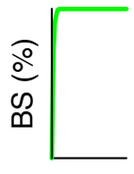
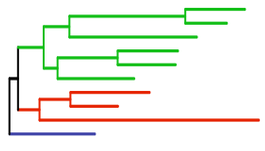
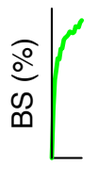
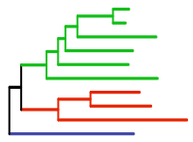
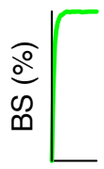
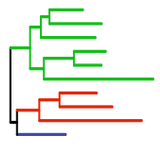
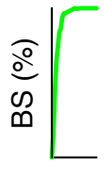
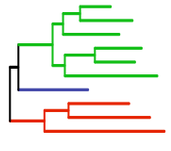
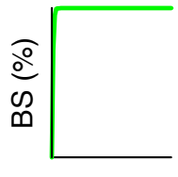
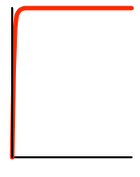
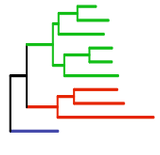
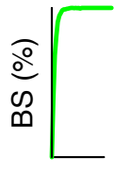
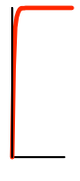
11 cyanobacteria

11 opisthokonts

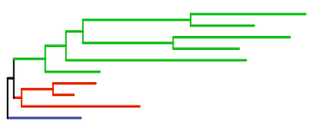
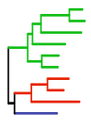
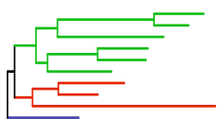
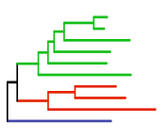
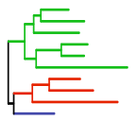
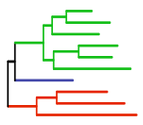
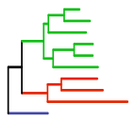
outgroup

We replace ‘Chromists’ by red algae.

algal icons from Reyes-Prieto et al. (2007) *Annu Rev Genet* 41:147-168

	# Pos	f(Sub)	Trees	VLBs		n_{70} Values	
				Greens	Reds	Gre	Reds
Plastid							
Polymerase	**1,911**	**3.55**				**137**	**1,226**
Photosynthesis	**5,717**	**1.17**				**102**	**120**
Ribosome	**2,074**	**2.92**				**577**	**853**
Mitochondrion	**3,106**	**2.29**				**174**	**109**
Nucleus							
Proteasome	**3,102**	**1.77**				**371**	**237**
Ribosome	**8,697**	**2.05**				**91**	**130**
Varia	**3,980**	**1.46**				**221**	**128**

Baurain et al. (2010) *Mol Biol Evol* (in press)

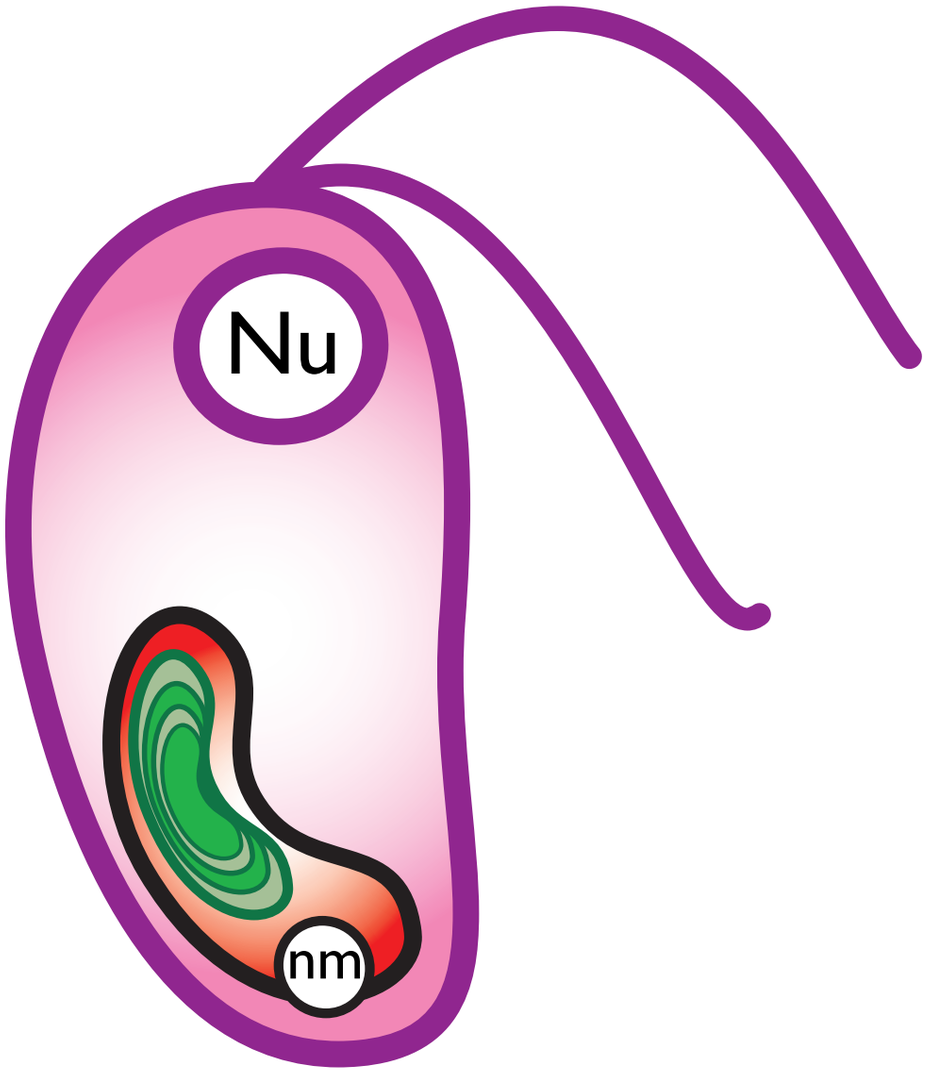
	# Pos	f(Sub)	Trees	VLBs		n_{70} Values	
				Greens	Reds	Gre	Reds
Plastid							
Polymerase	1,911	3.55				137	1,226
Photosynthesis	5,717	1.17				102	120
Ribosome	2,074	2.92				577	853
Mitochondrion	3,106	2.29				174	109
Nucleus							
Proteasome	3,102	1.77				371	237
Ribosome	8,697	2.05				91	130
Varia	3,980	1.46				221	128

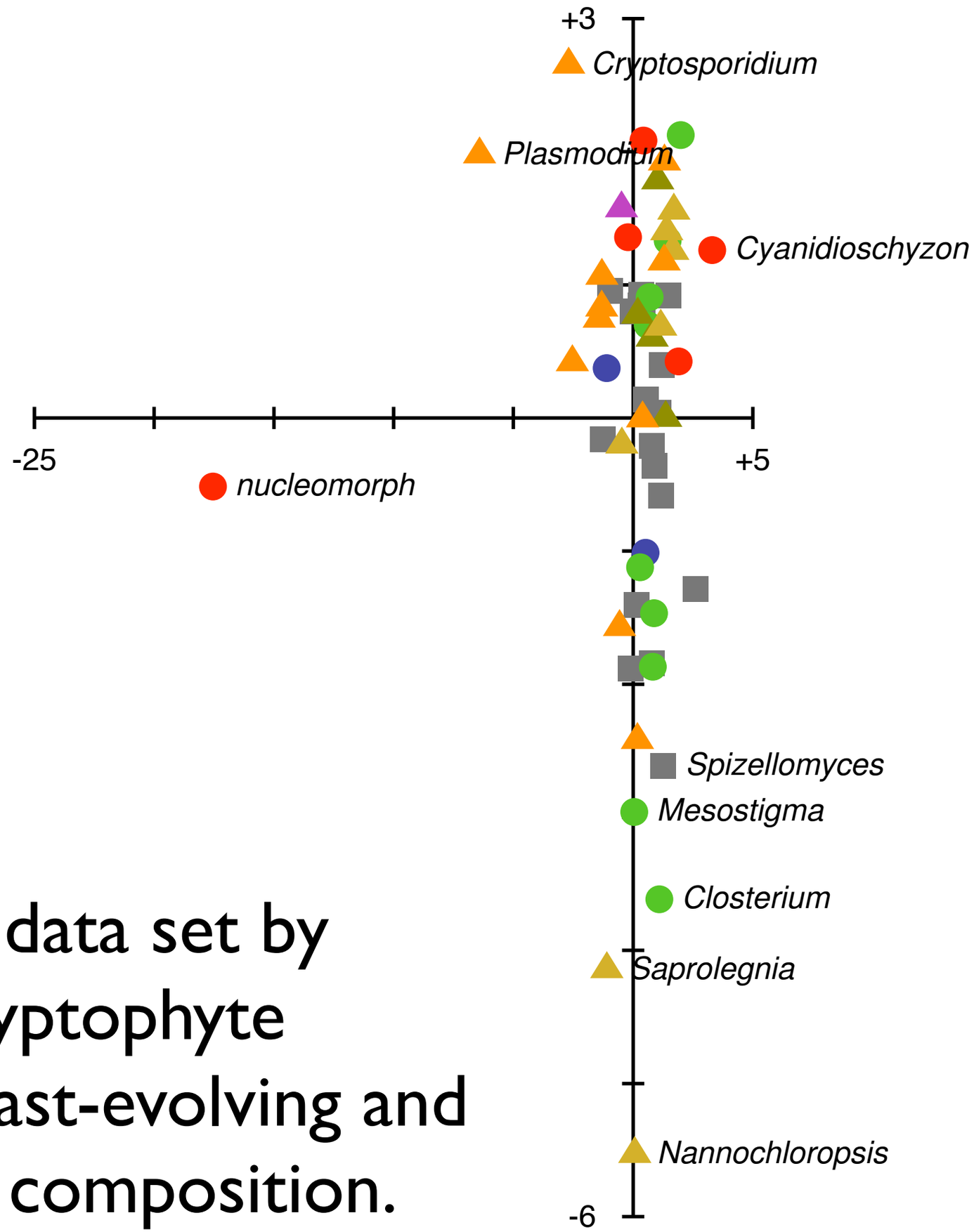
Easy recovery of the monophyly of red algae shows that our data sets have ample phylogenetic power.

Baurain et al. (2010) *Mol Biol Evol* (in press)

How to check the assumptions of our test?

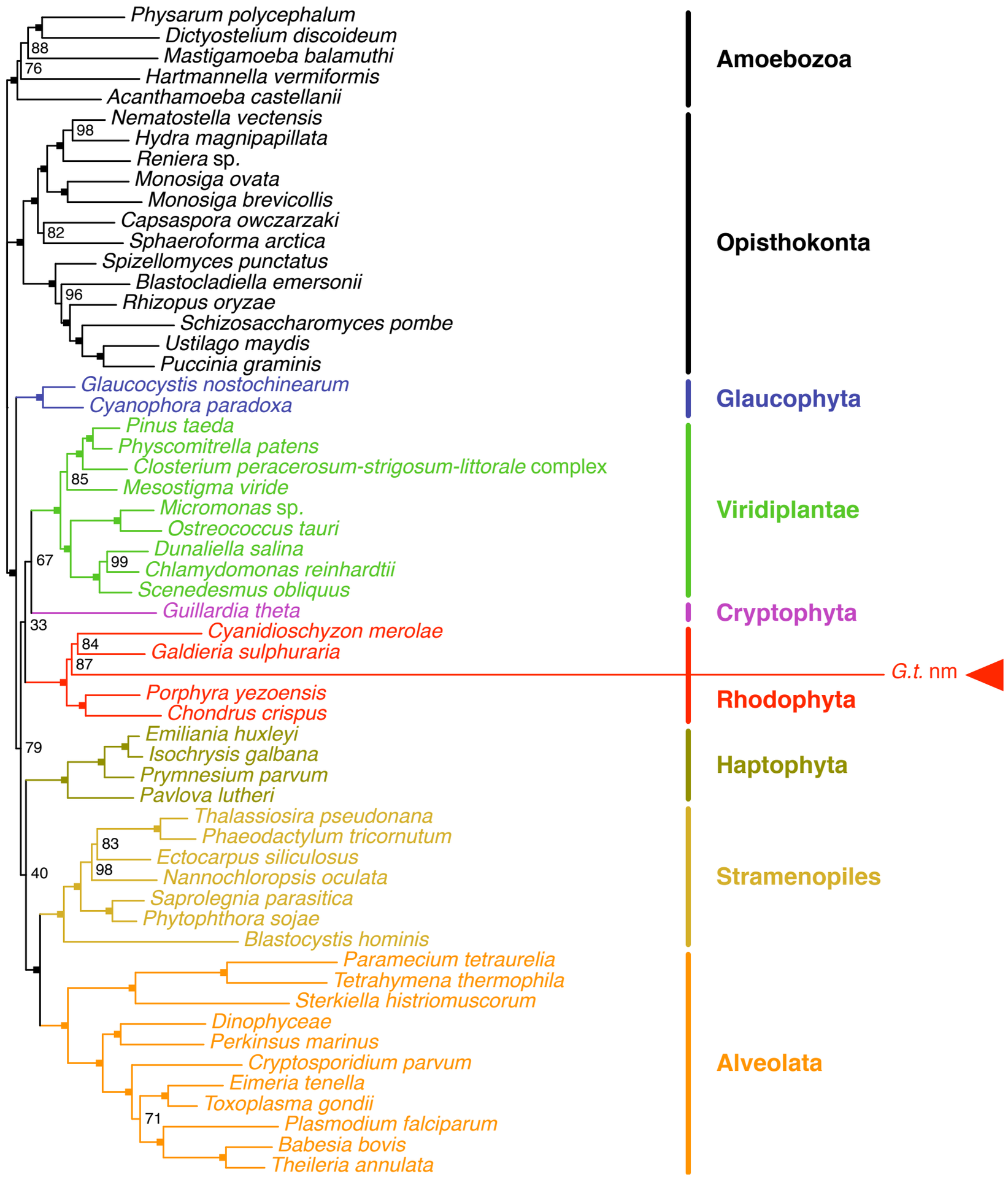
1. *What if Plantae are actually paraphyletic?*
2. *How to deal with heterogeneous rates?*
3. *Do we have enough phylogenetic power?*
4. *Are we misled by phylogenetic artifacts?*
5. *Are we misled by undetected EGT?*



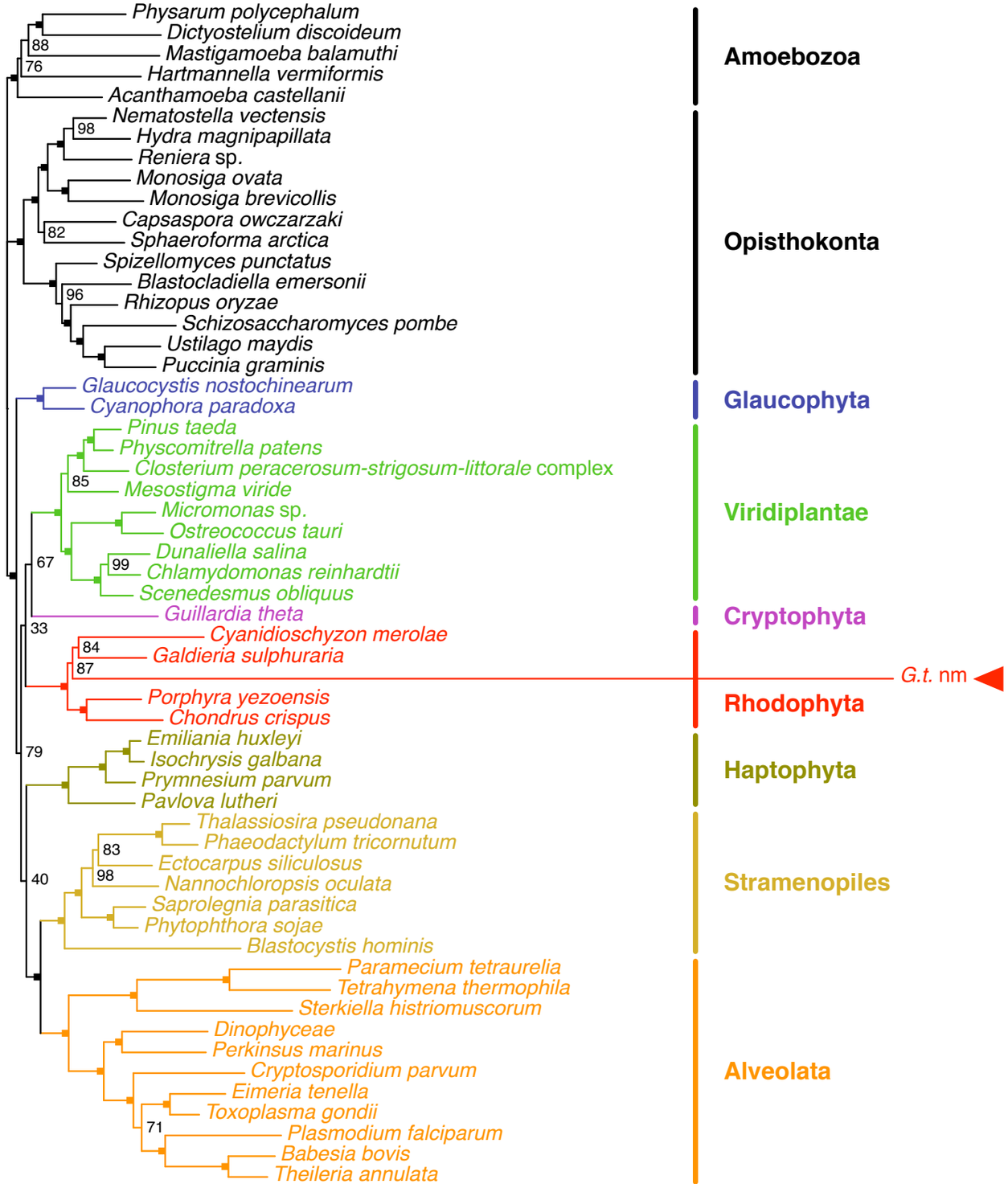


We extend our nuclear data set by notably including the cryptophyte nucleomorph, which is fast-evolving and shows a strongly biased composition.

Baurain et al. (2010) *Mol Biol Evol* (in press)



Baurain et al. (2010) *Mol Biol Evol* (in press)



108 nucleus-encoded proteins
57 ‘species’ x 15,392 AA

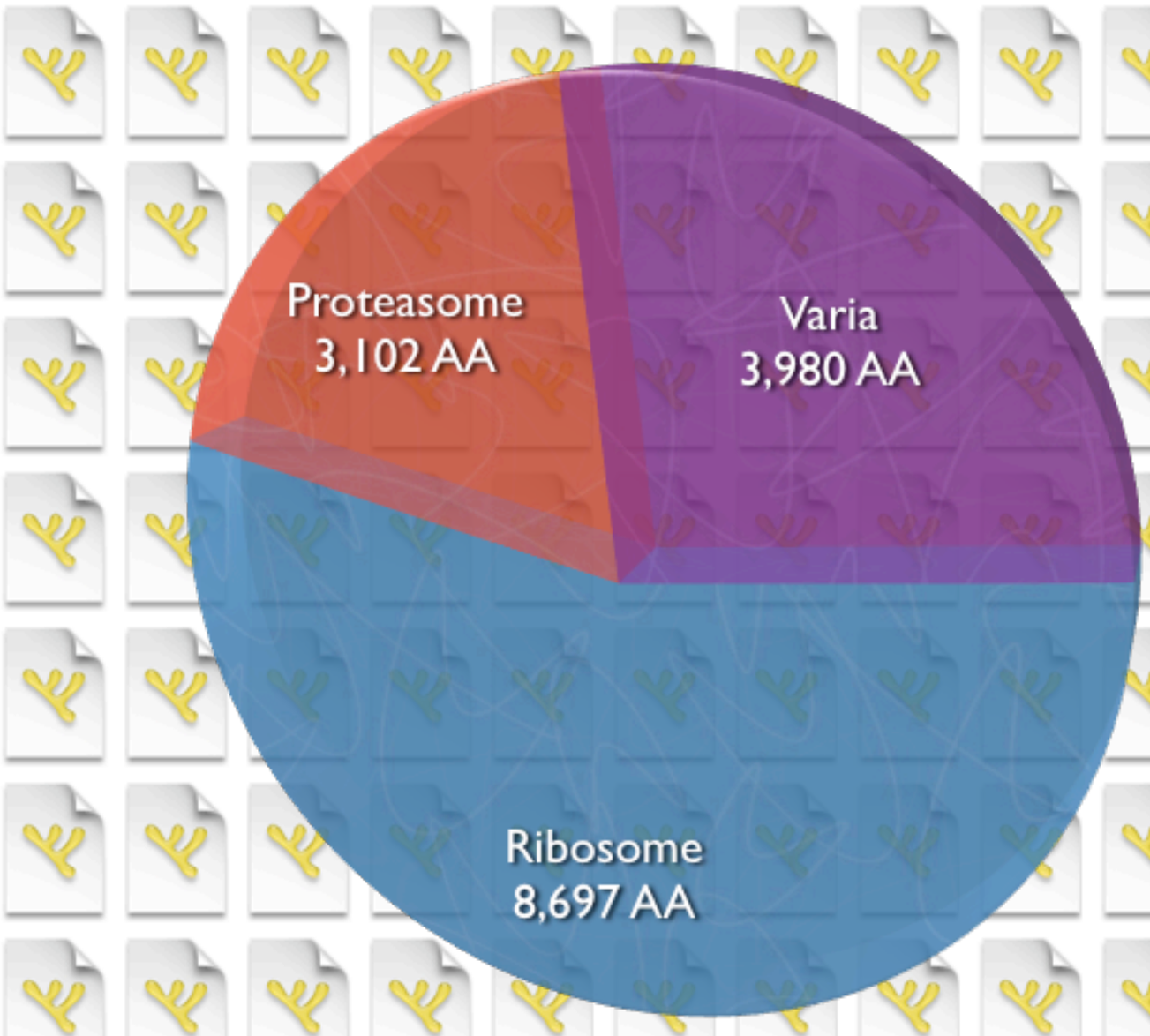
CAT+Γ_4 model
100 bootstrap replicates

Robustly locating the nucleomorph rules out artifacts when resolving less aberrant ‘Chromalveolates’.

Baurain et al. (2010) *Mol Biol Evol* (in press)

How to check the assumptions of our test?

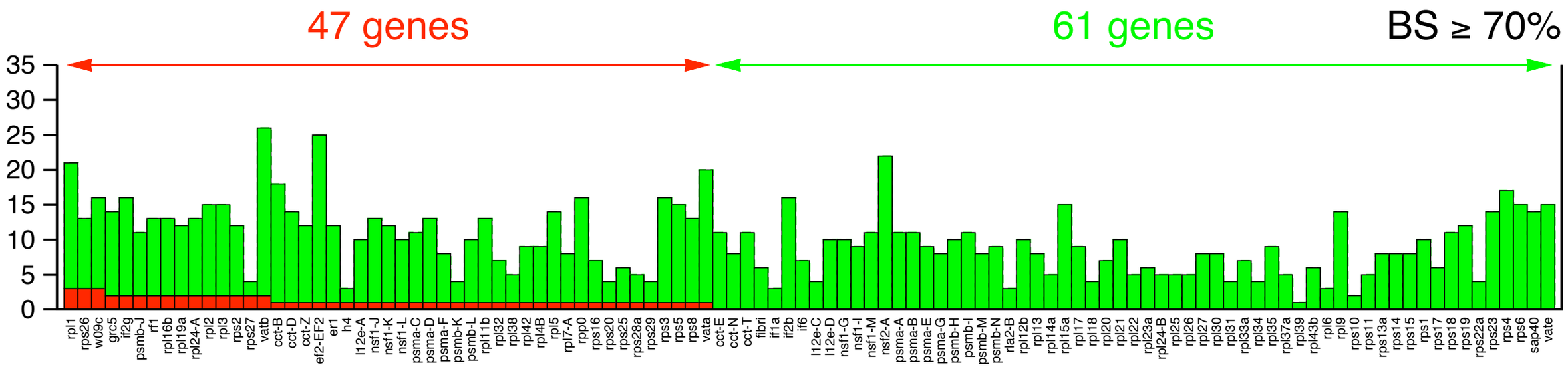
1. *What if Plantae are actually paraphyletic?*
2. *How to deal with heterogeneous rates?*
3. *Do we have enough phylogenetic power?*
4. *Are we misled by phylogenetic artifacts?*
5. *Are we misled by undetected EGT?*



Concat. Tree

We analyze separately each nuclear gene and look for statistically supported nodes that are incongruent with the concatenation.

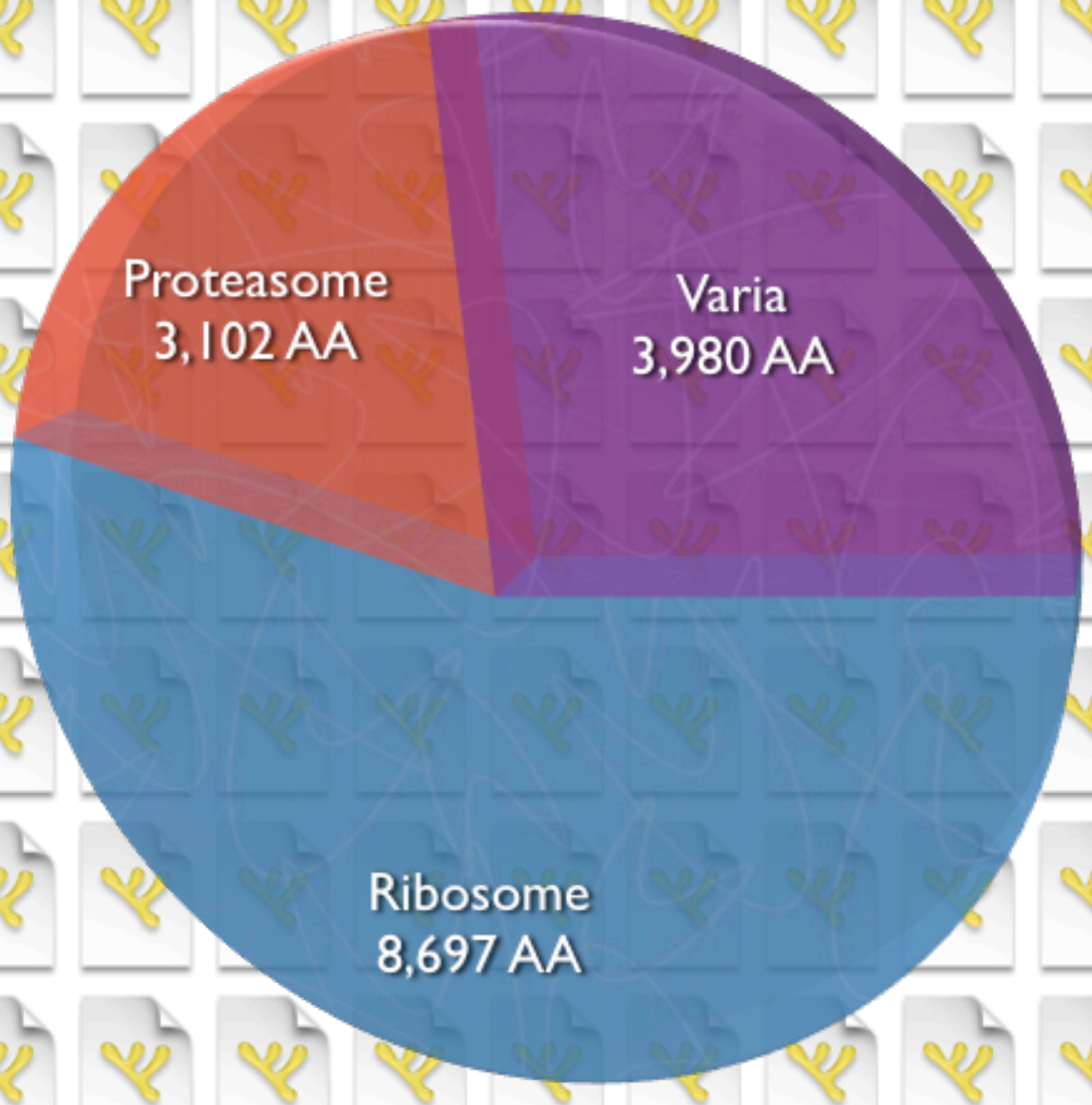
trees icons from TreeView X (Rod Page, Glasgow); magnifying glass from http://www.multimedialab.be/



Causes of Incongruence					
NNI	local	LBA	EGT	unexp.	total
30	22	9	0	4	65

Congruence analysis of individual genes does not yield any evidence for EGT in our nuclear data set.

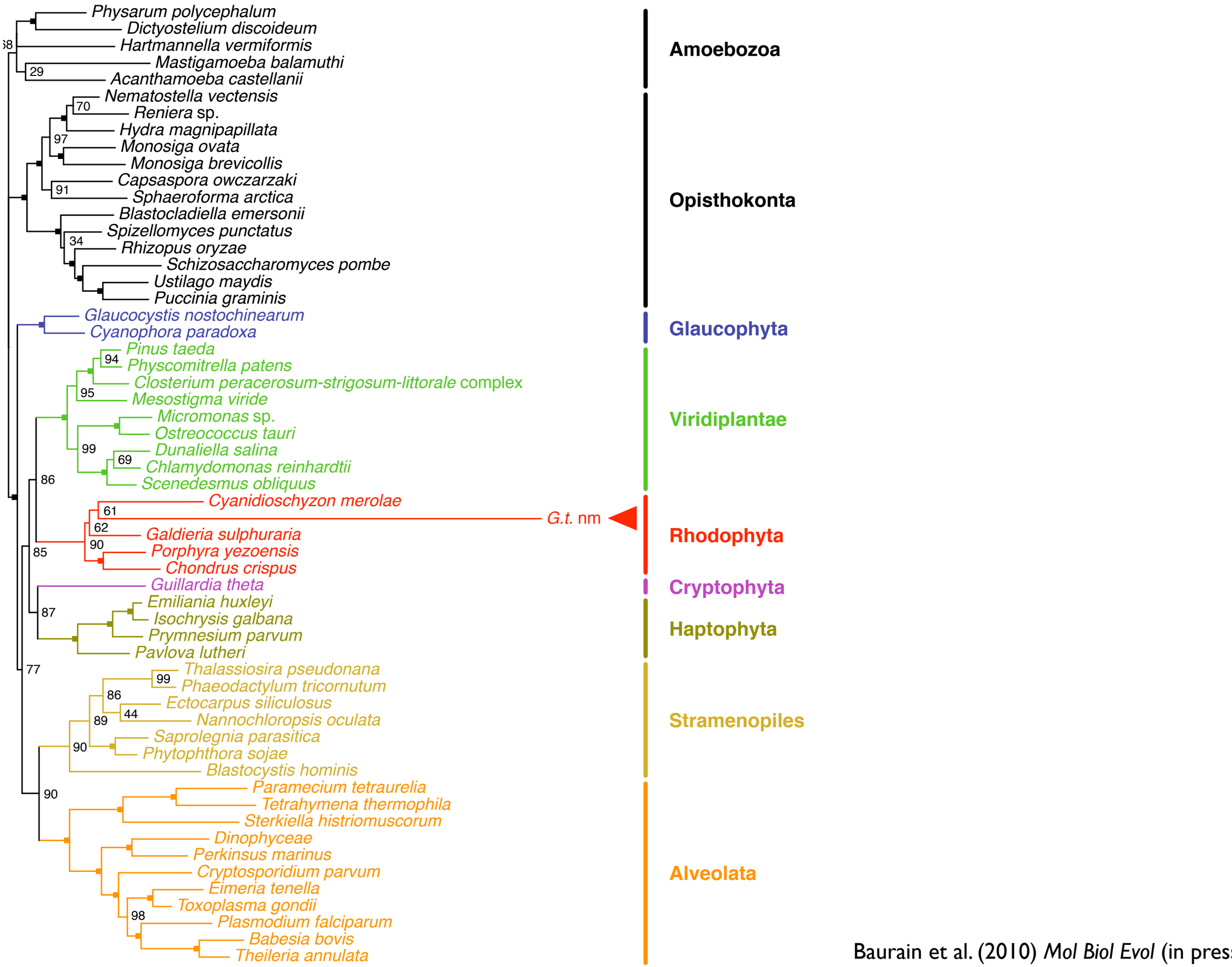
Baurain et al. (2010) *Mol Biol Evol* (in press)



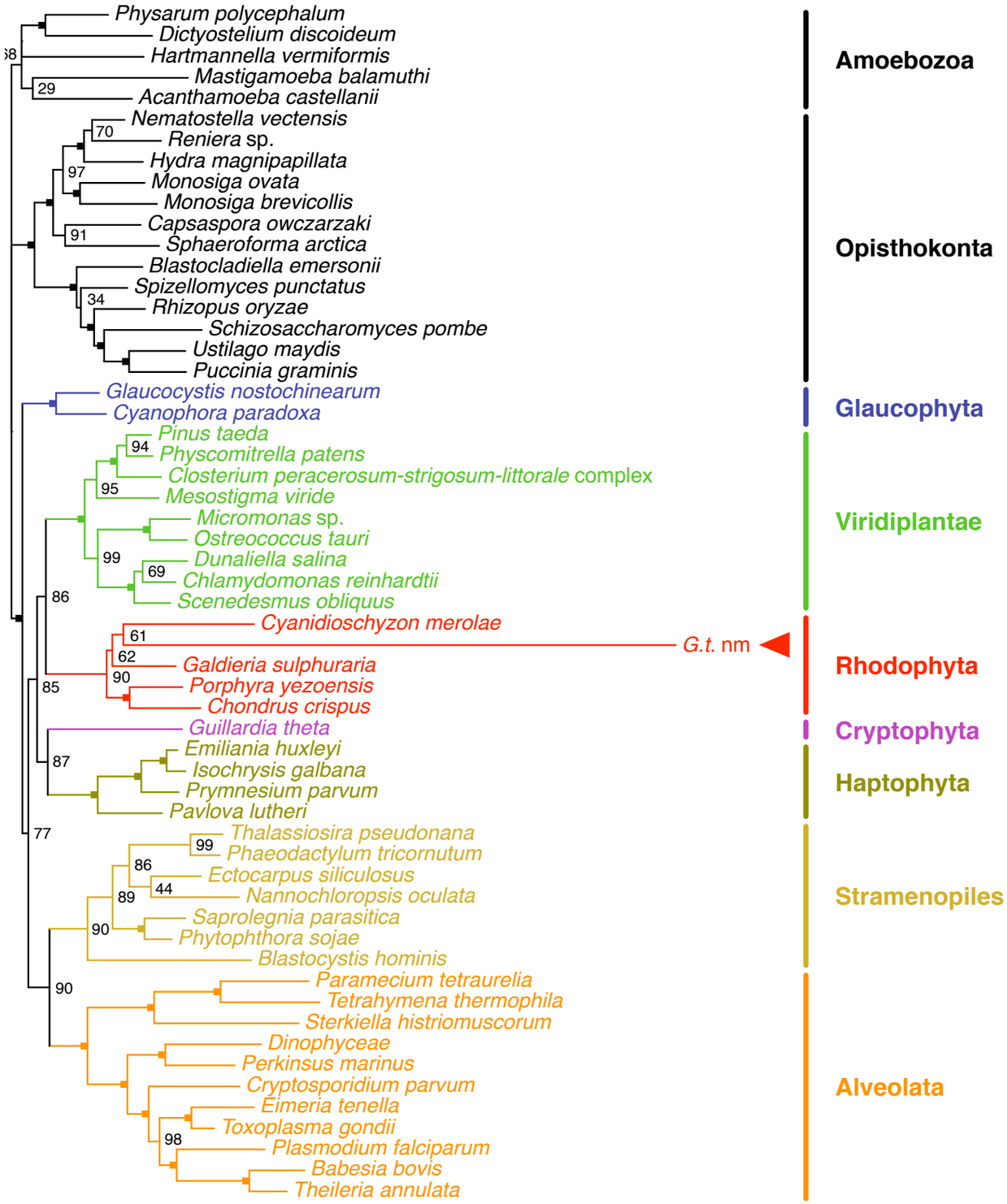
Reds

We analyze separately each nuclear gene and look for those that statistically support the monophyly of red algae.

trees icons from TreeView X (Rod Page, Glasgow); magnifying glass from http://www.multimedialab.be/



Baurain et al. (2010) *Mol Biol Evol* (in press)



21 nucleus-encoded proteins supporting red algae (BS ≥ 70%)

57 ‘species’ x 5,838 AA

CAT+Γ_4 model
100 bootstrap replicates

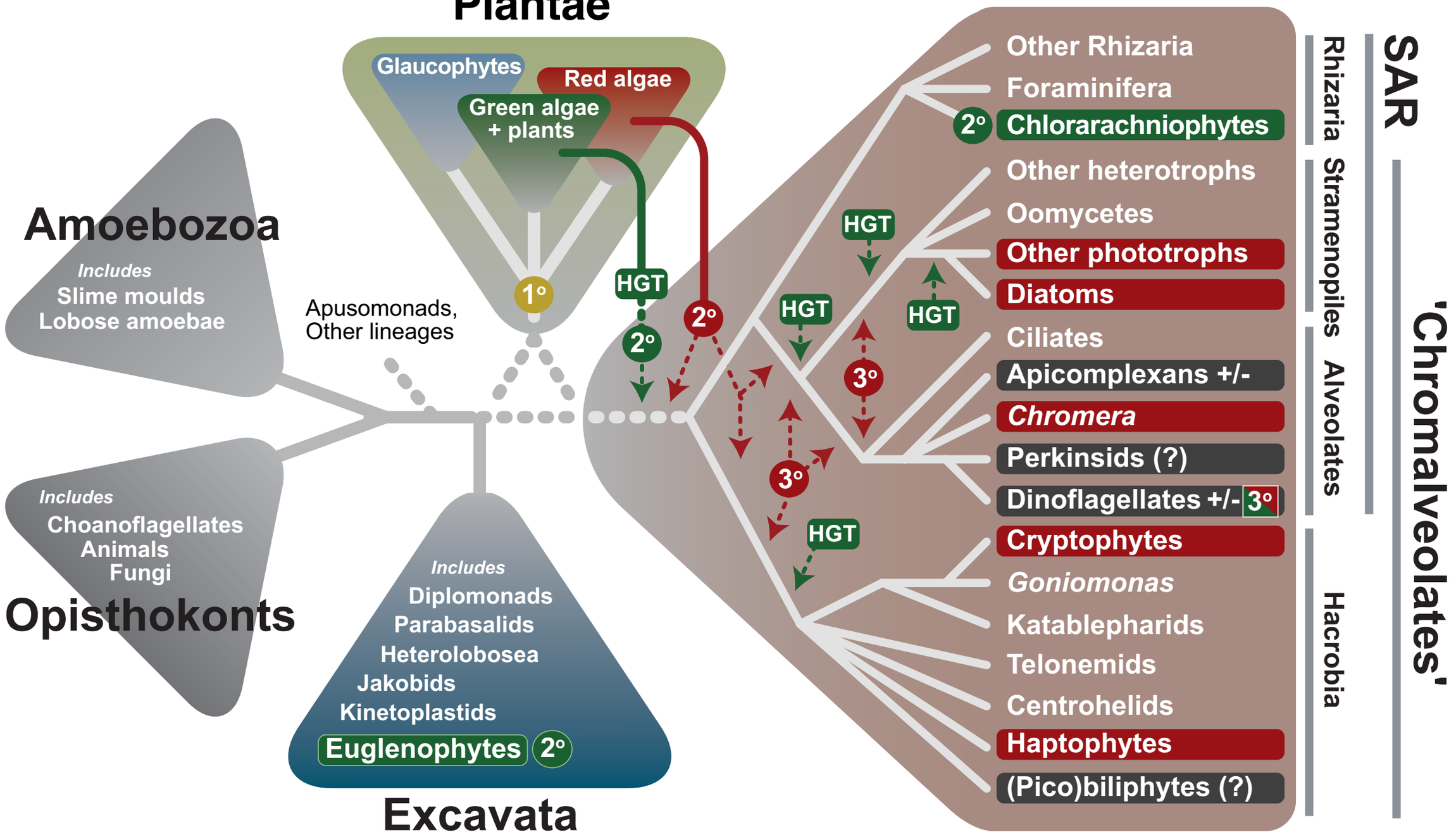
Clearly non-transferred genes do not recover the monophyly of ‘Chromalveolates’.

Baurain et al. (2010) *Mol Biol Evol* (in press)

How to check the assumptions of our test?

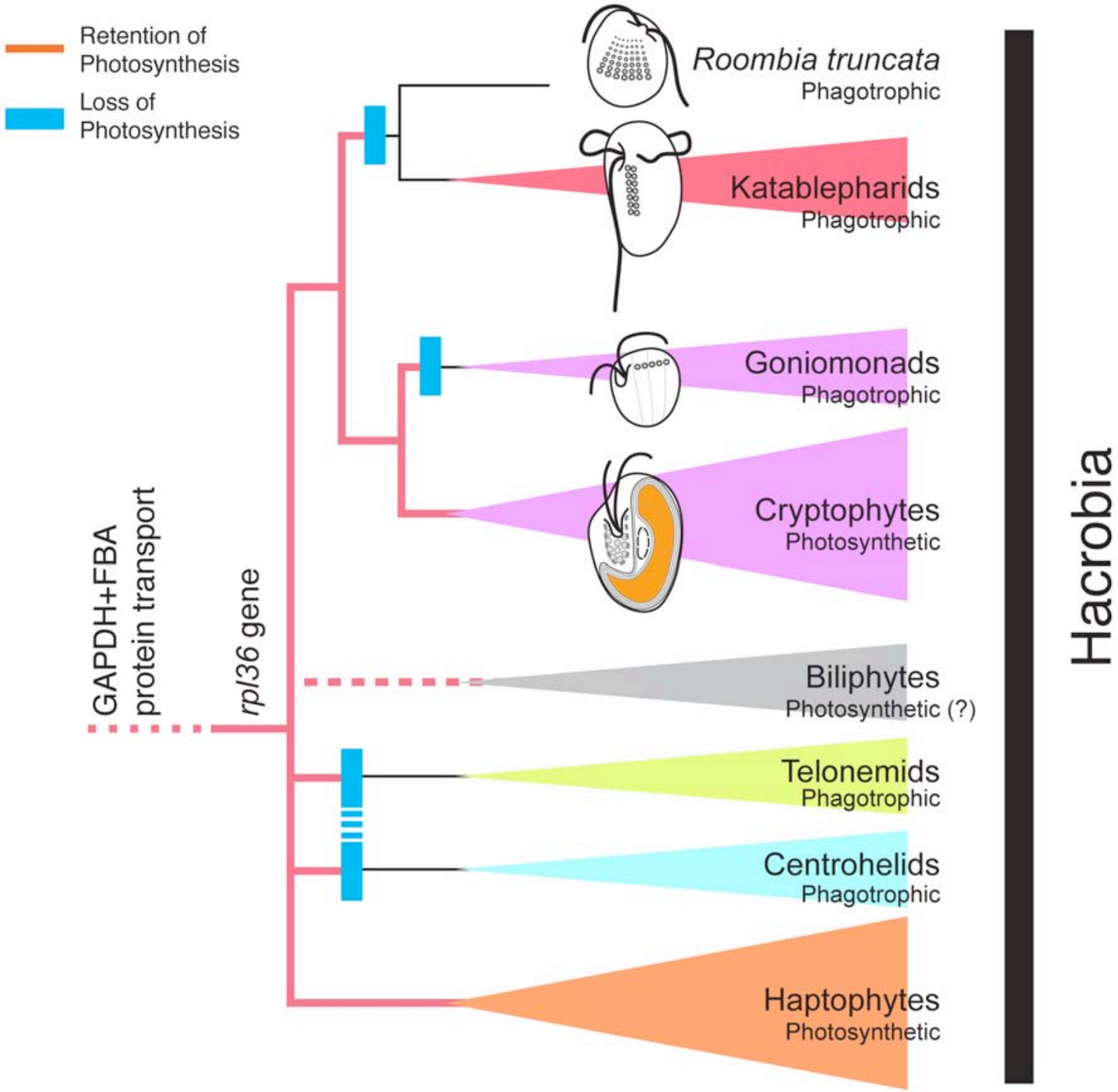
Control experiments show that the failure of the nuclear compartment to recover the monophyly of 'Chromalveolates' cannot be explained by wrong phylogenetic assumptions, heterogeneous evolutionary rates, a lack of phylogenetic power, combined phylogenetic artifacts, or undetected endosymbiotic gene transfer.

Can we specify an alternative hypothesis?



Hacrobia and the SAR clade each resemble 'Chromalveolates' in miniature.

adapted from Elias and Archibald (2009) *BioEssays* 31:1273-1279



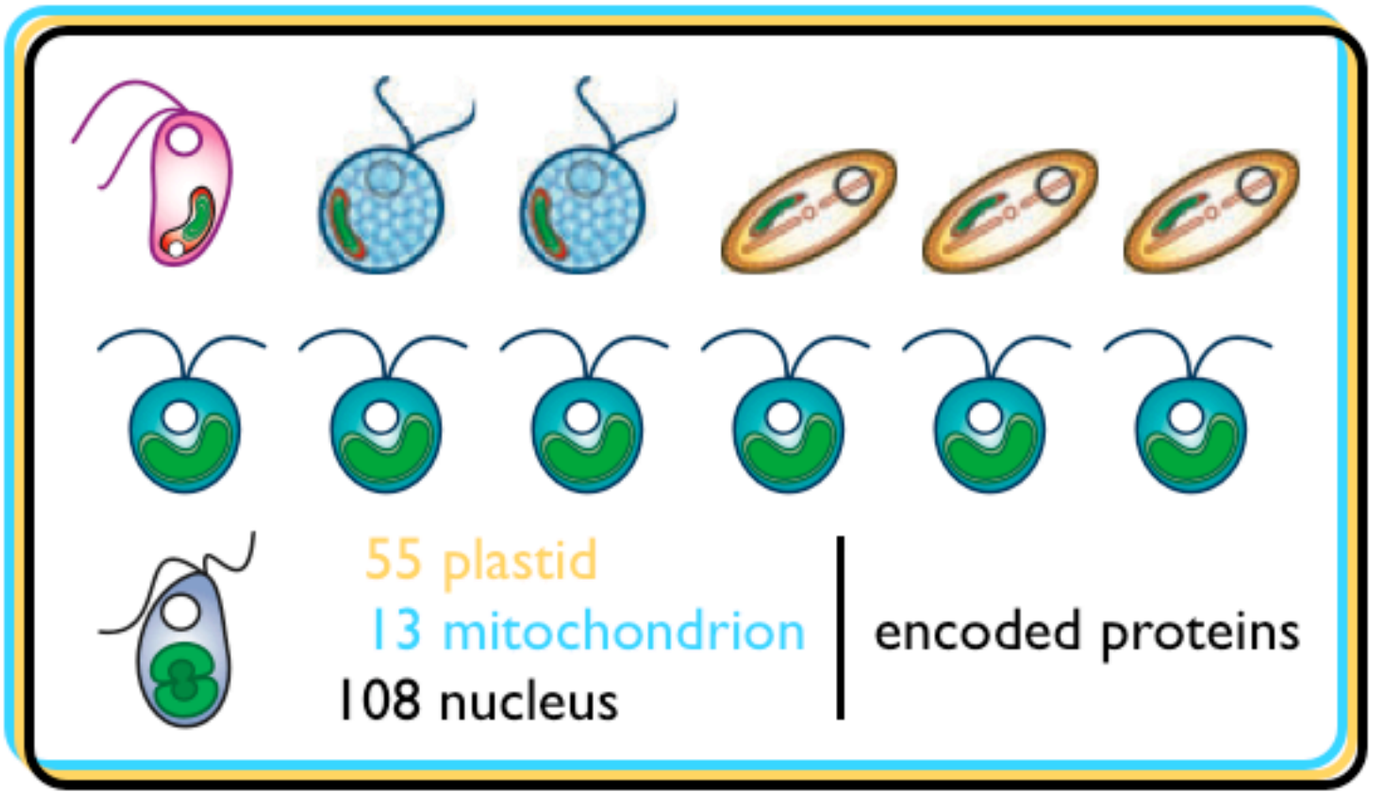
Hacrobia are supposed to be a monophyletic group.

Okamoto et al. (2009) *PLoS ONE* 4:e7080

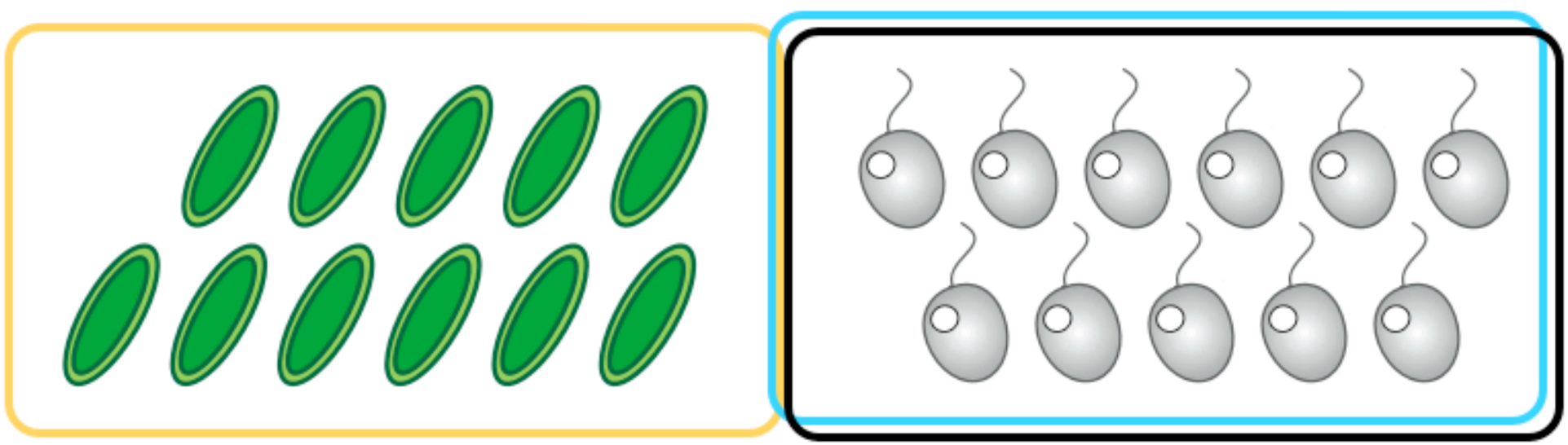
6 'Chromists'

6 green plants

1 glaucophyte



ingroup



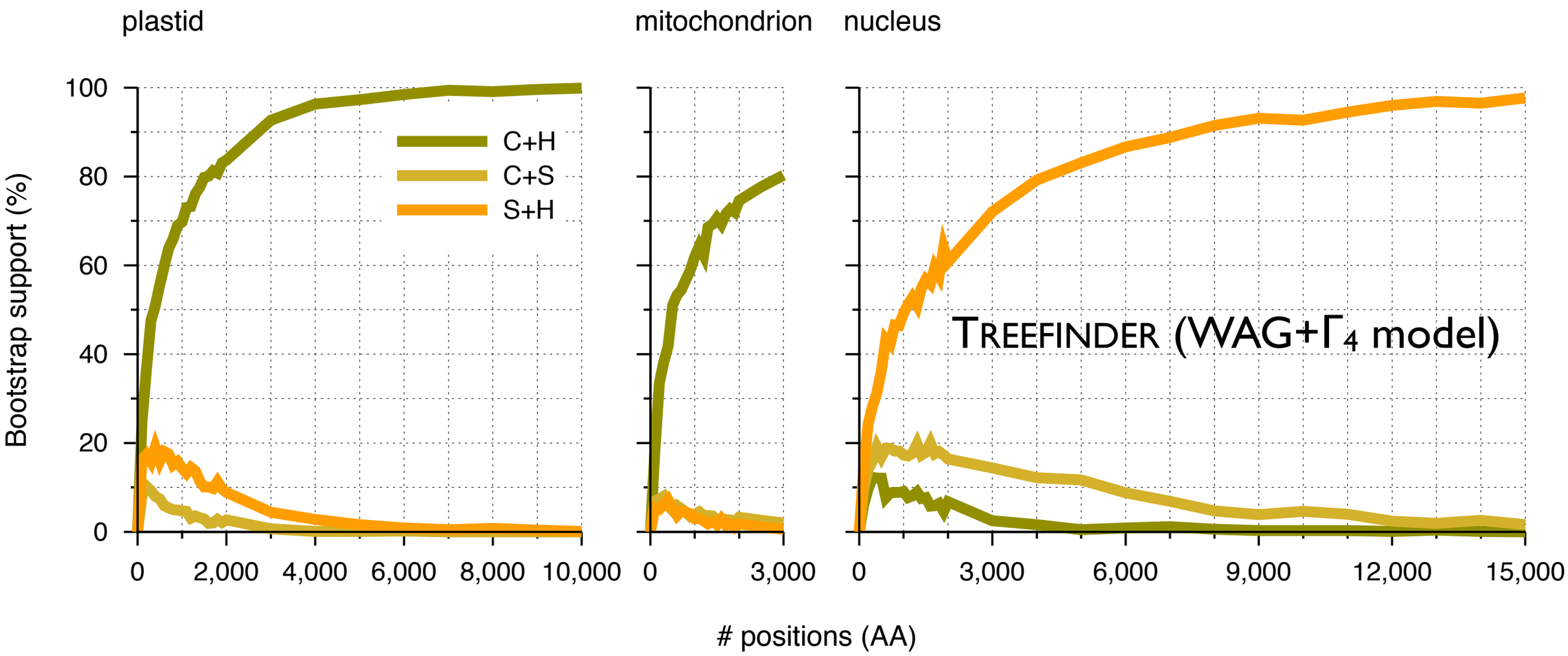
11 cyanobacteria

11 opisthokonts

outgroup

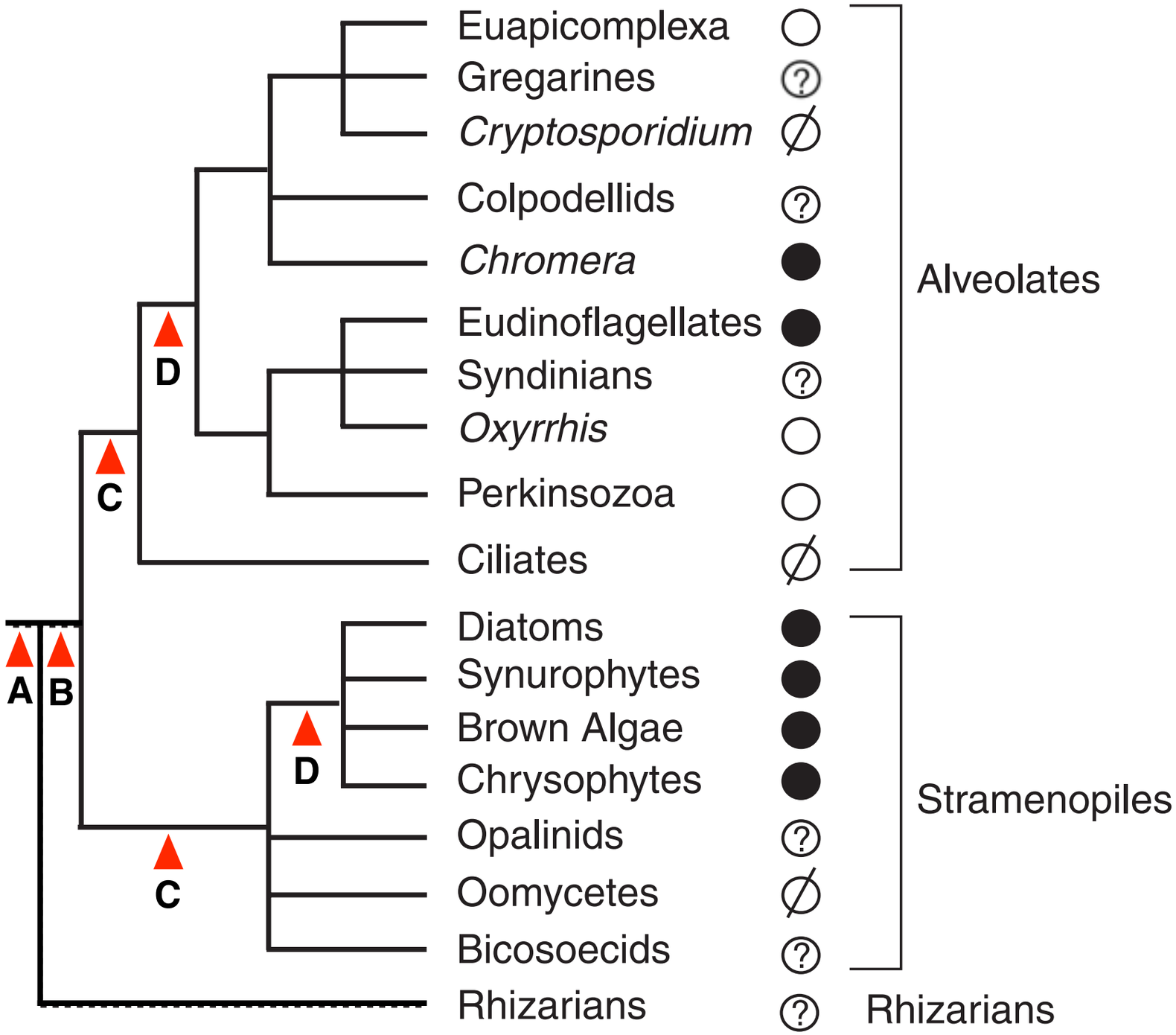
Our data sets can be used to estimate the strength of the signal for the monophyly of Hacrobia.

algal icons from Reyes-Prieto et al. (2007) *Annu Rev Genet* 41:147-168



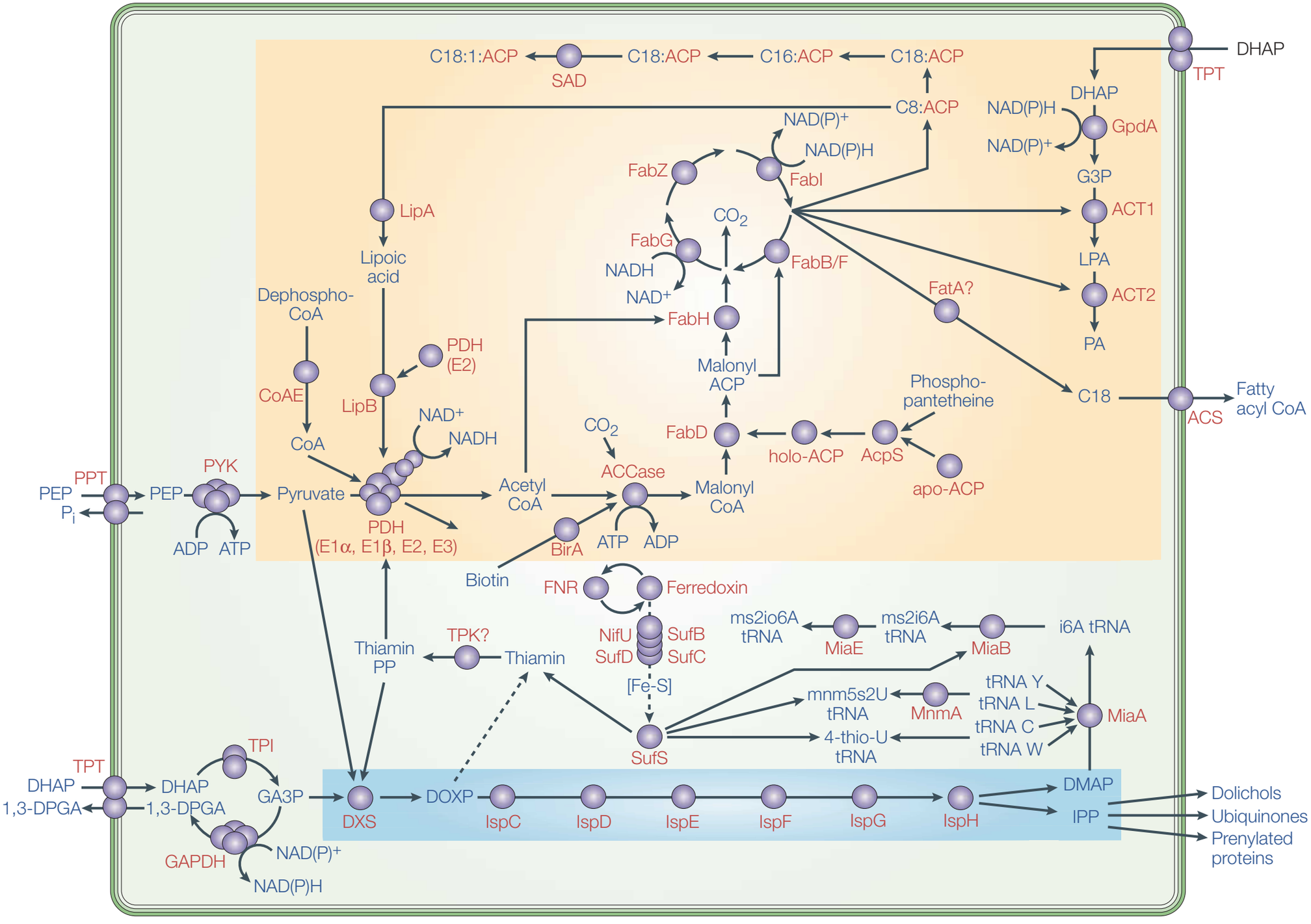
Support for Hacrobia is ambiguous, depending on compartment, gene sampling and evolutionary model.

Baurain et al. (unpublished)



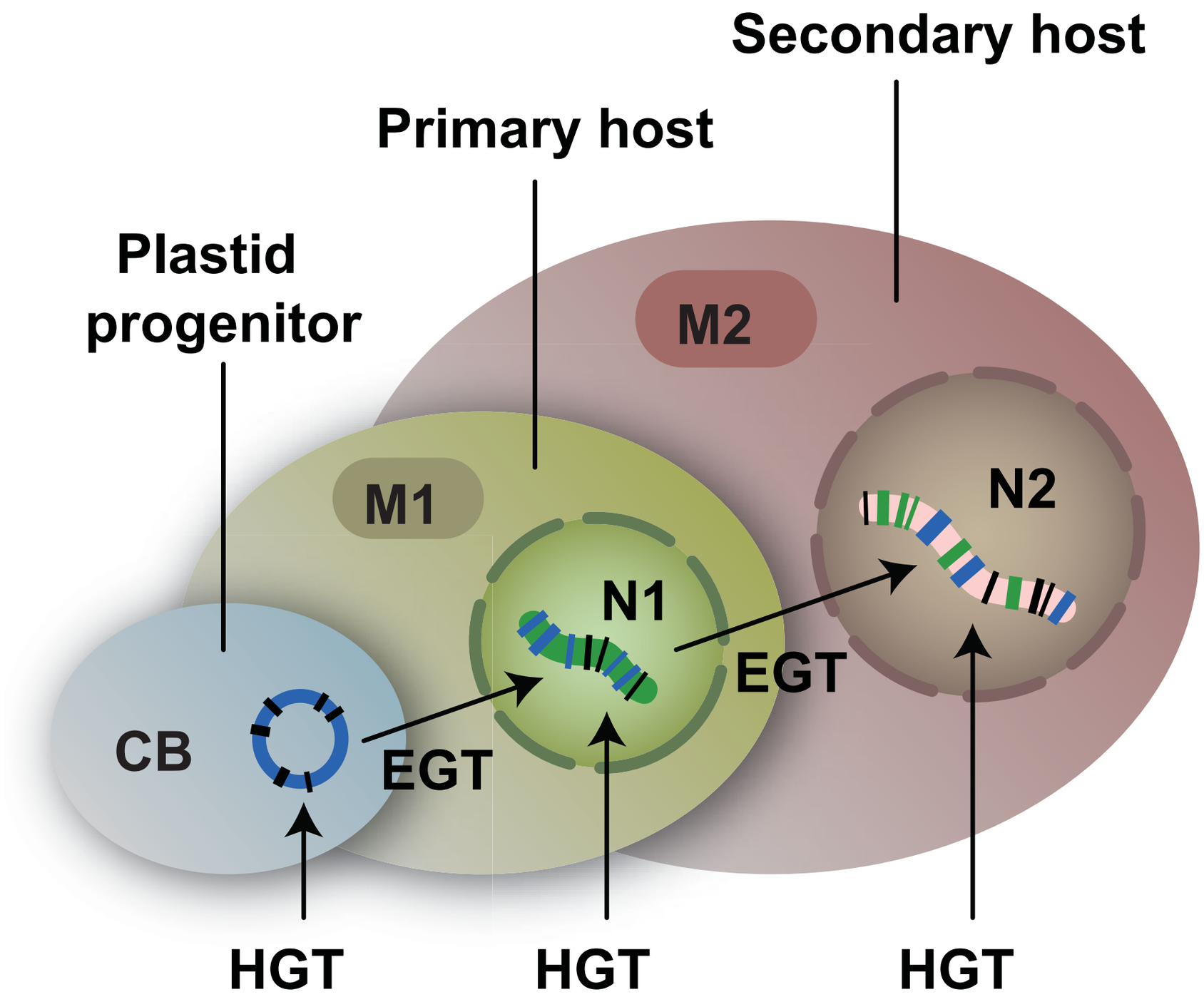
Cryptic plastids (or plastid remains) have been found in a number of heterotrophic SA[R] lineages.

adapted from Keeling (2009) *J Eukaryot Microbiol* 56:1-8



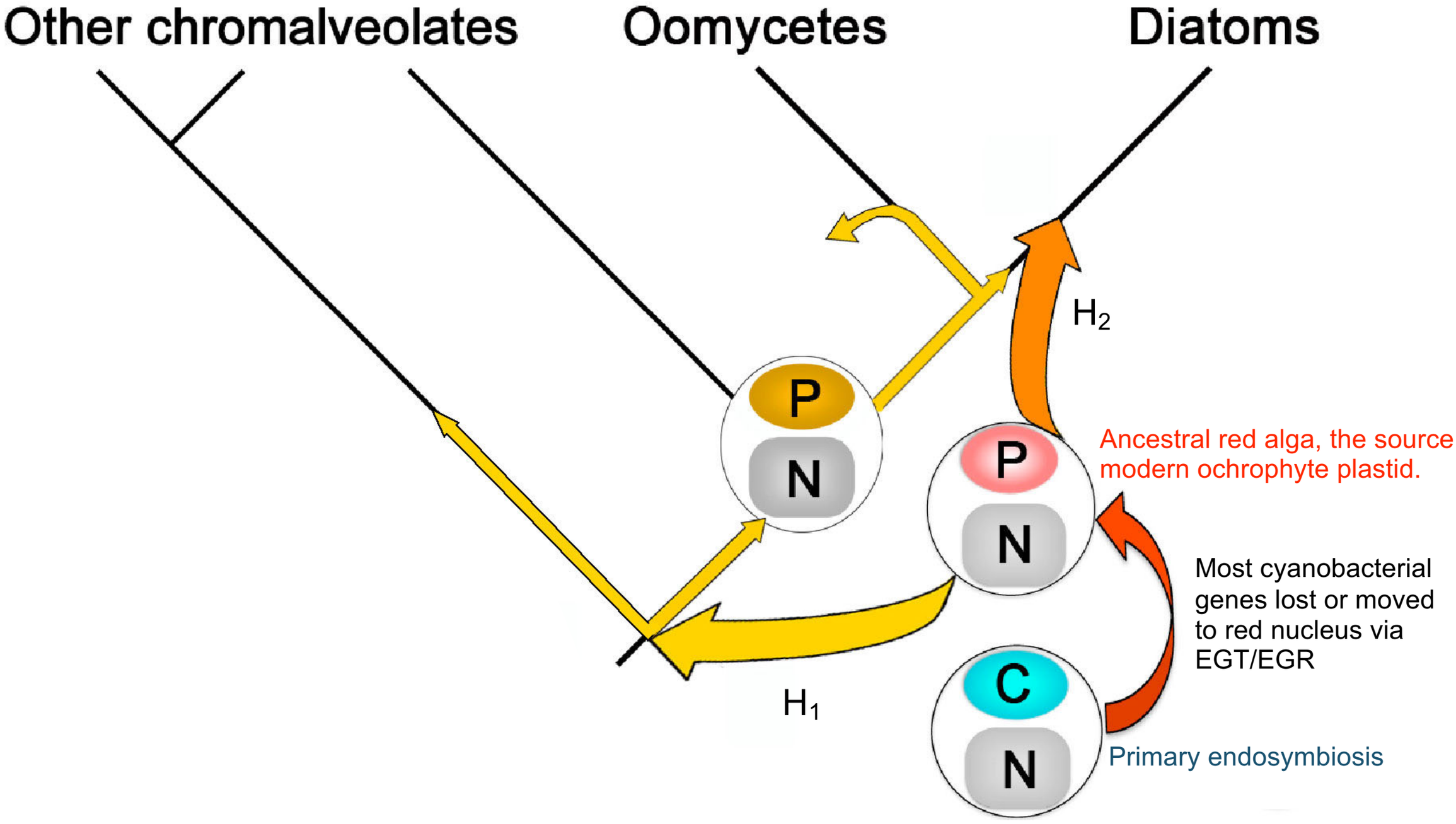
The apicoplast of apicomplexan parasites is an example of a non-photosynthetic plastid that cannot be lost.

Ralph et al. (2004) *Nat Rev Microbiol* 2:203-216



In case of complete plastid loss,
some ‘algal’ genes should still be present.

Elias and Archibald (2009) *BioEssays* 31:1273-1279



A priori testing shows that 'algal' genes found in Oomycetes are not proof of a photosynthetic past.

Stiller et al. (2009) *BMC Genomics* 10:484

Can we specify an alternative hypothesis?

The evolution of chlorophyll *c*-containing algae is better explained by a single secondary endosymbiosis with a red alga followed by at least one tertiary (quaternary...) endosymbiosis.

Issues surrounding Hacrobia and the SAR clade are reminiscent of those of 'Chromalveolates'.

As plastids are difficult to loose, 'algal' genes in heterotrophic lineages should be reassessed.

Do 'Chromalveolates' really exist?

Do 'Chromalveolates' really exist?

Whatever the meaning, most likely no.

Even if falsified, the 'Chromalveolate hypothesis' has provided a great impetus to the study of eukaryotic evolution (just like Archezoa).

Genome sequencing and cell biology are both needed to propose a testable alternative model.