A Taste of Algal Genomes from the Joint Genome Institute

Alan Kuo[1], Igor Grigoriev[1]

[1]Lawrence Berkeley National Laboratory

June 2012

DISCLAIMER

A Taste of Algal Genomes from the Joint Genome Institute.

Alan Kuo* (akuo@lbl.gov) and Igor Grigoriev, DOE Joint Genome Institute, Lawrence Berkeley National Laboratory

Algae and the JGI mission

Algae and the environment
50% of global photosynthesis
Base of food webs
Carbon fixation and cycling
Biomineralization, biogeochemistry
Blooms can be toxic, blooms can be HUGE
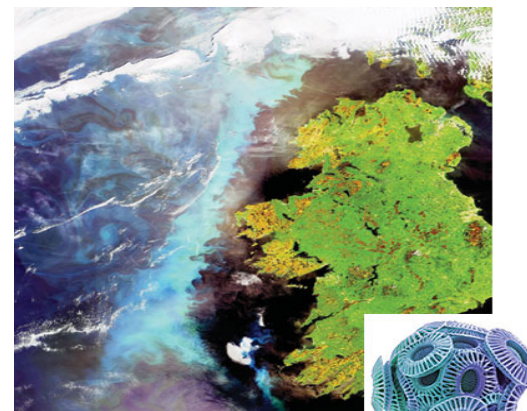
Algae and evolution
Very diverse, all over eukaryotic tree
Photosynthesis acquired many times by many clades
Endosymbiosis is a principal driver of evolution
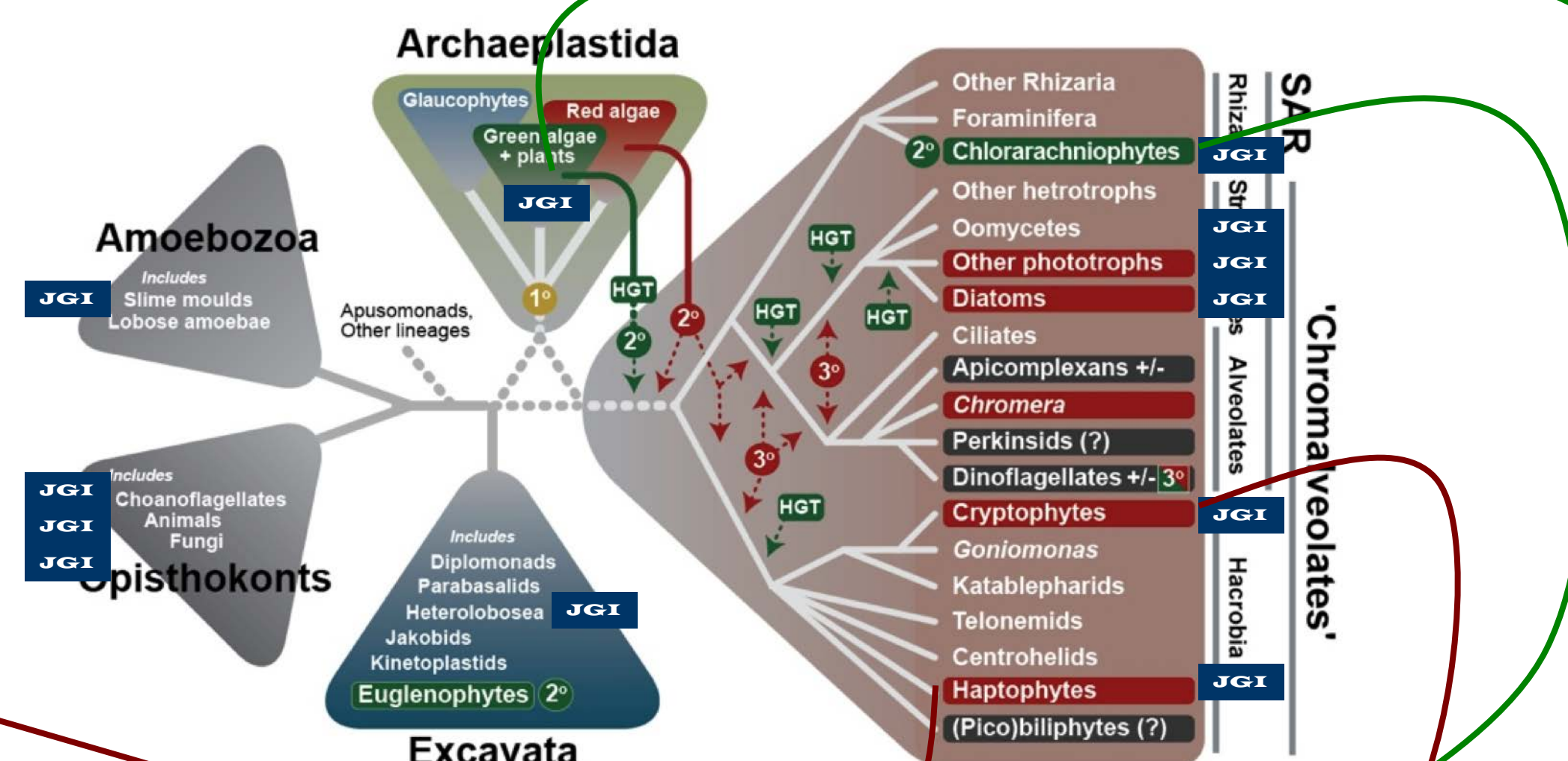Plastid loss: Oomycota, evil plant pathogens
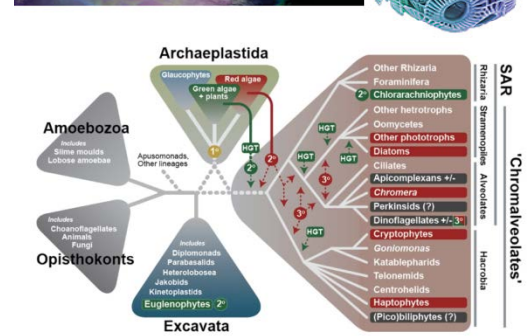
Algae and energy
Source of biofuels
Model systems for plants
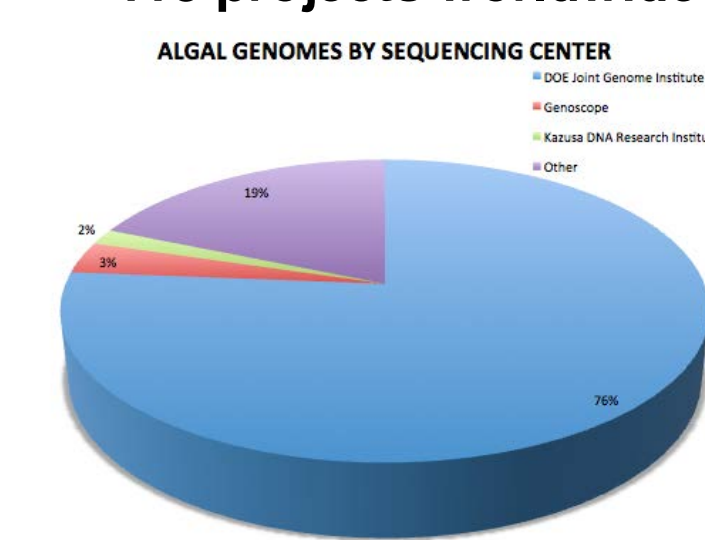Model systems for plant-fungal symbioses (lichens)

Lipid -> Diesel
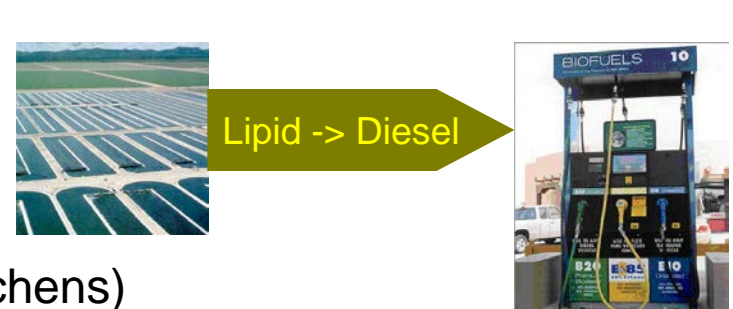
Abstract

Algae play profound roles in aquatic food chains and the carbon cycle, can impose health and economic costs through toxic blooms, provide models for the study of symbiosis, photosynthesis, and eukaryotic evolution, and are candidate sources for bio-fuels; all of these research areas are part of the mission of DOE's Joint Genome Institute (JGI). To date JGI has sequenced, assembled, annotated, and released to the public the genomes of 18 species and strains of algae, sampling almost all of the major clades of photosynthetic eukaryotes. With more algal genomes currently undergoing analysis, JGI continues its commitment to driving forward basic and applied algal science. Among these ongoing projects are the pan-genome of the dominant coccolithophore *Emiliania huxleyi*, the interrelationships between the 4 genomes in the nucleomorph-containing *Bigelowiella natans* and *Guillardia theta*, and the search for symbiosis genes of lichens.

Archaeplastida
Glaucophytes
Red algae
Green algae + pla ti ds
Other Rhizaria
Foraminifera
Chlorarachniophytes
Other hetrotrophs
Oomycetes
Other phototrophs
Diatoms
Ciliates
Apicomplexans +/-
Chromera
Perkinsids (?)
Dinoflagellates +/-
Cryptophytes
Goniomonas
Katablepharids
Telonemids
Centrohelids
Haptophytes
(Pico)biliphytes (?)

Amoebozoa
Includes
Slime moulds
Lobose amoebae

Opisthokonts
Includes
Choanoflagellates
Animals
Fungi

Excavata
Includes
Diplomonads
Parabasalids
Heterolobosea
Jakobids
Kinetoplastids
Euglenophytes

Algae sequenced by JGI

118 projects worldwide

11 genomes published by JGI (out of 13):
Thalassiosira (2004)
Ostreococcus, 2 sp. (2007)
Chlamydomonas (2007)
Phaeodactylum (2008)
Micromonas, 2 sp. (2009)
Volvox (2010)
Chlorella (2011)
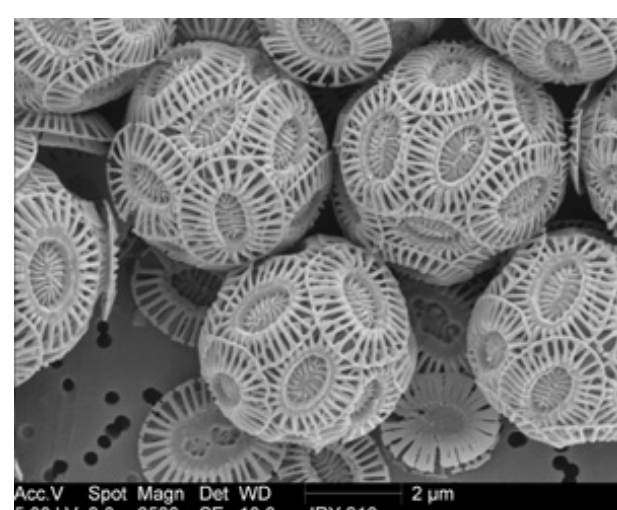*Aureococcus* (2011)
Coccomyxa (2012)
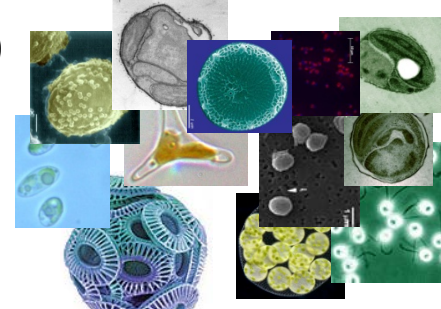
7 genomes released but not yet published:
Emiliania   *Asterochloris*
Ostreococcus (a 3rd sp)   *Guillardia*
Fragilariopsis   *Pseudo-nitzschia*
Bigelowiella

90 projects at JGI

JGI Algal Portal: http://jgi.doe.gov/Algae/

JGI Community Sequencing Program:
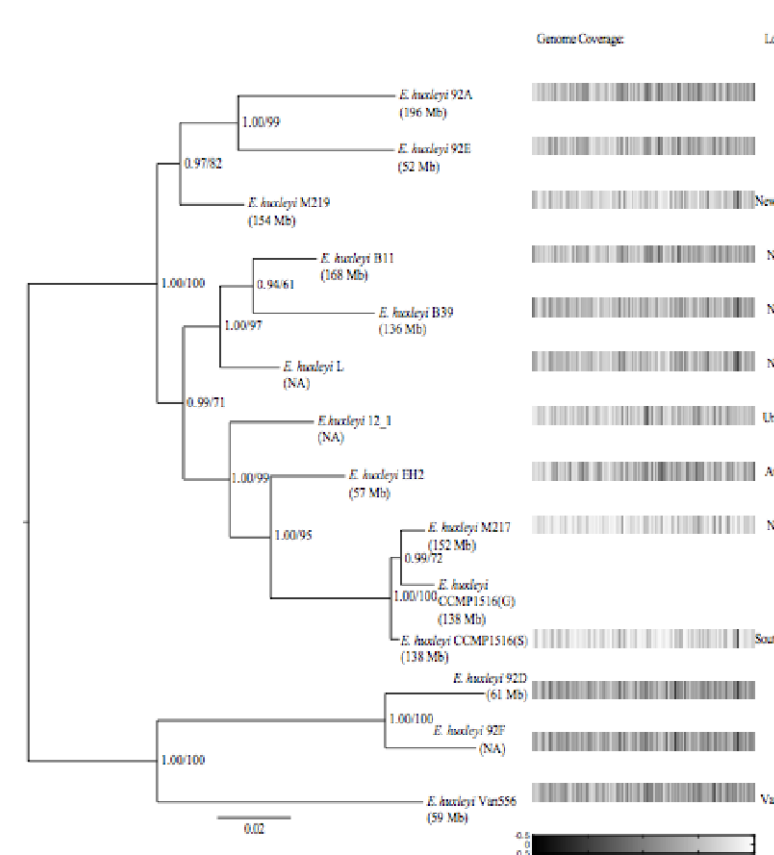http://www.jgi.doe.gov/CSP/

The dominant phytoplankter *Emiliania huxleyi*

Ehux is globally distributed, locally abundant, and well adapted to a wide range of environments.
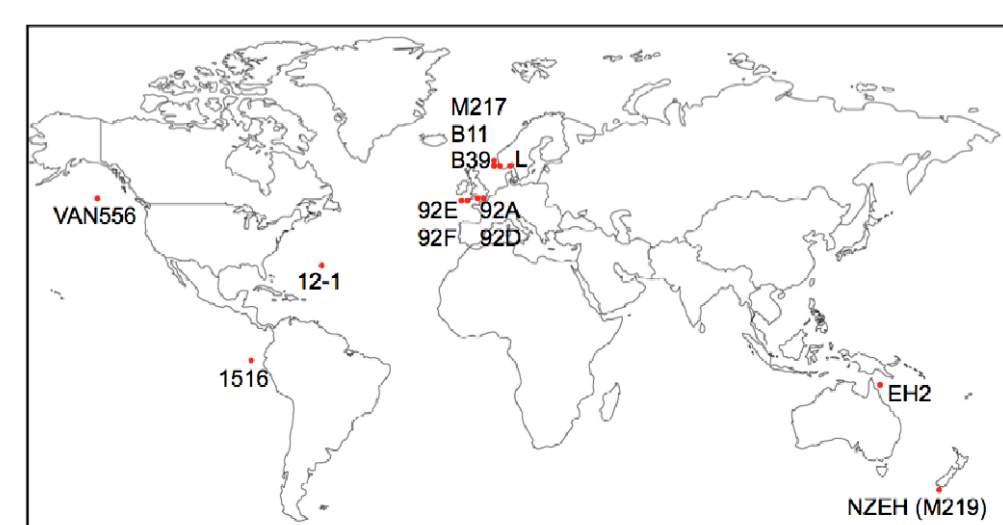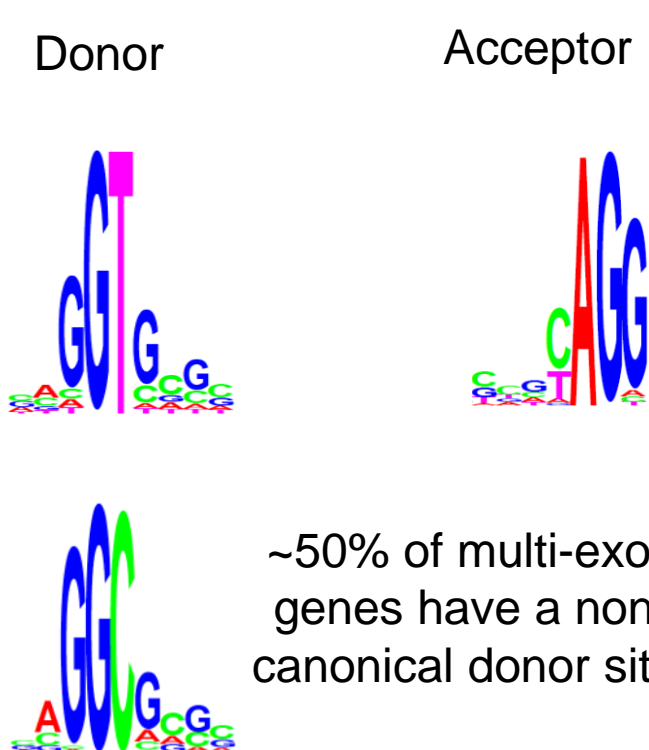
Geography and phylogeny of 14 strains of Ehux.

The reference genome of Ehux.
1st sequenced Haptophyta, Hacrobia

Genome size (Mbp)	168
# scaffolds	7809
# genes	39126
# genes (reduced)	33341

Donor   Acceptor

~50% of genome is repetitive.
~55% of genes are in multigene families.
~75% of genes have introns.

~50% of multi-exon genes have a non-canonical donor site.

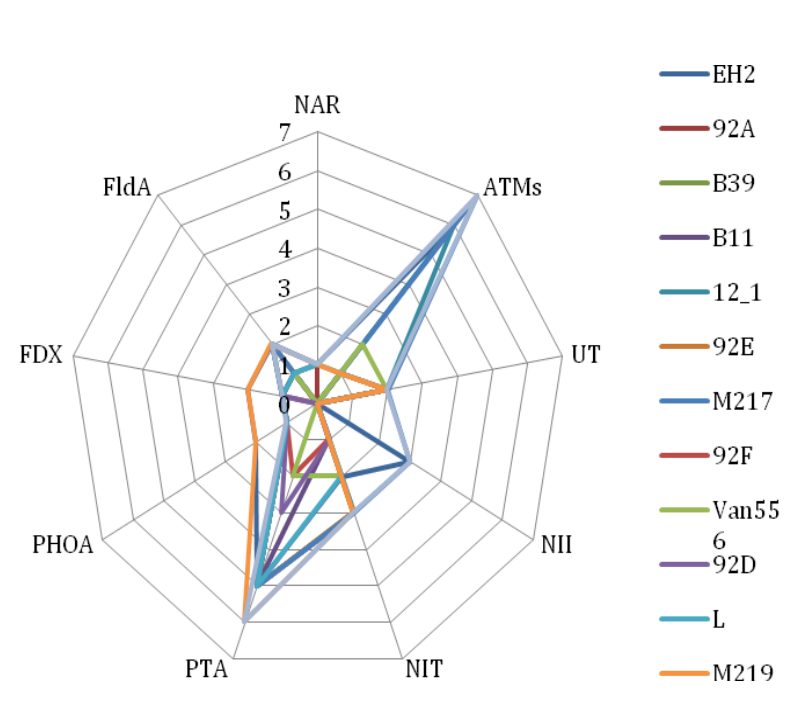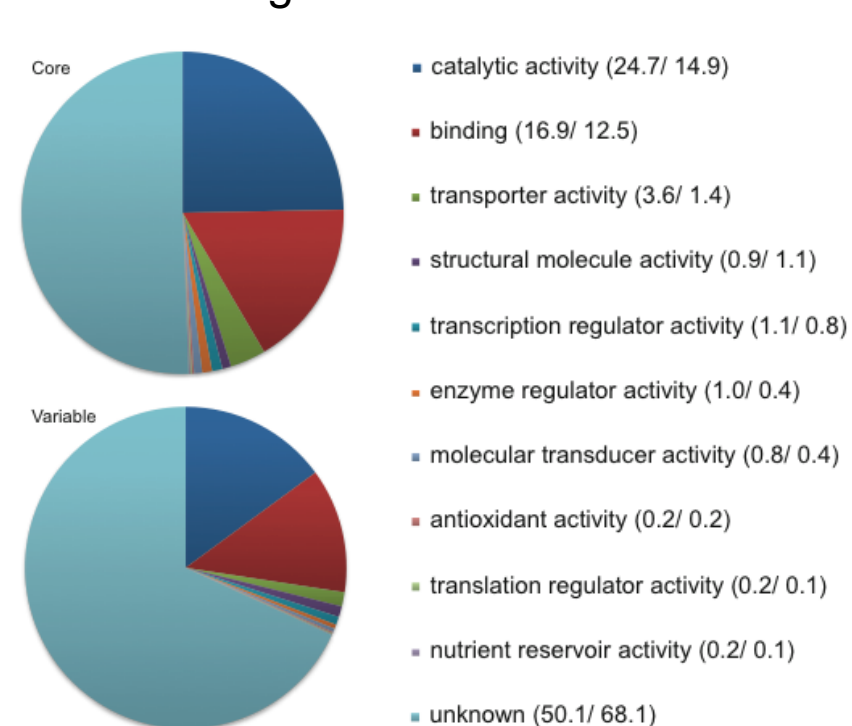**Resequencing of 13 other strains of Ehux from around the world.**

Collect 13 strains.
Sequence to 9-49x coverage.
Assemble into contigs.
Align to reference proteins.
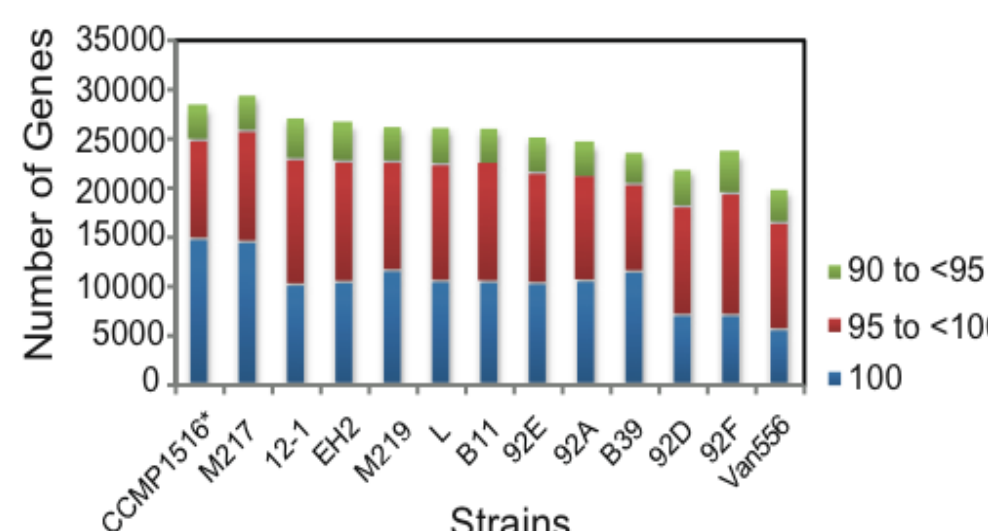
Many of the reference genes are absent from many of the 13 strains.

90 to <95
95 to <100
100

Ehux has core and variable genomes.

~18k core genes

Core
catalytic activity (24.7/ 14.9)
binding (16.9/ 12.5)
transporter activity (3.6/ 1.4)
structural molecule activity (0.9/ 1.1)
transcription regulator activity (1.1/ 0.8)
enzyme regulator activity (1.0/ 0.4)
molecular transducer activity (0.8/ 0.4)
antioxidant activity (0.8/ 0.4)
translation regulator activity (0.2/ 0.1)
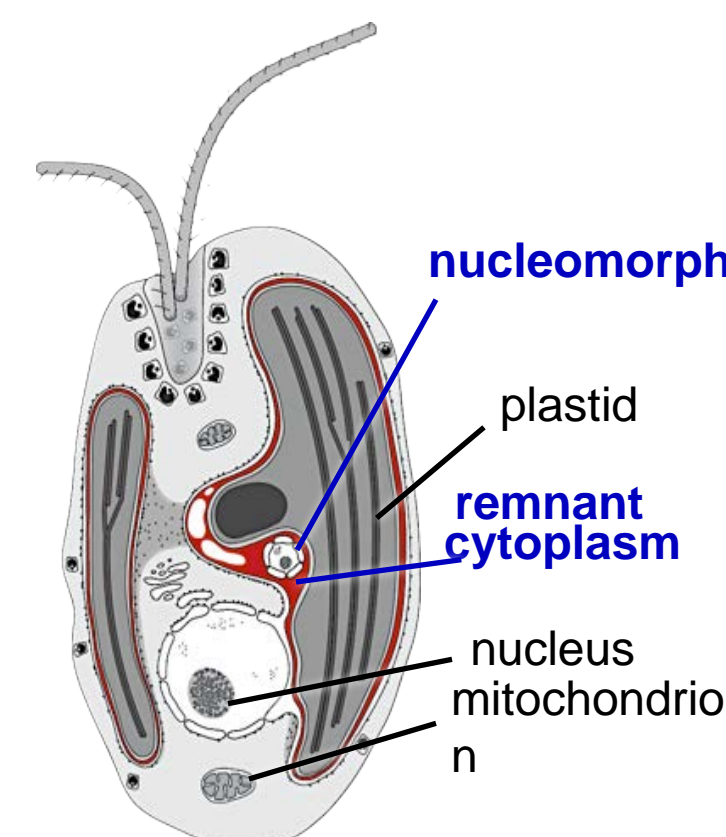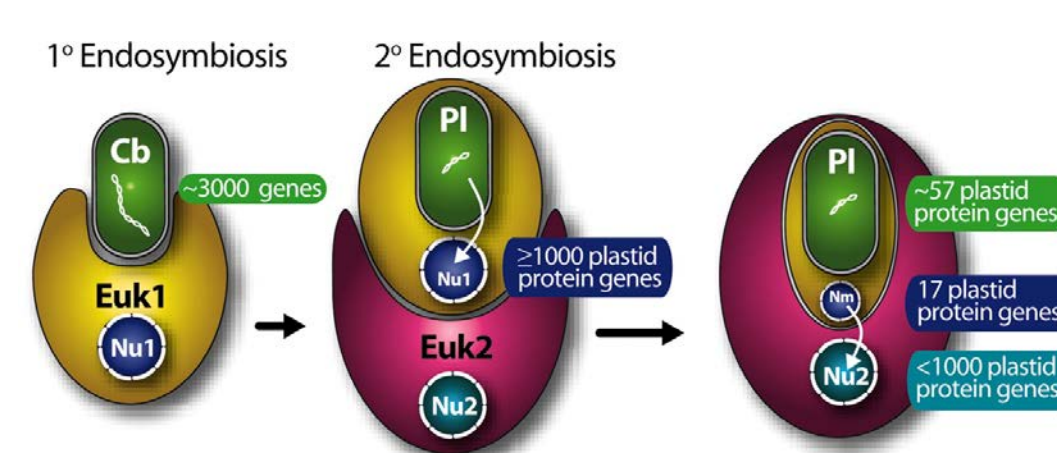nutrient reservoir activity (0.2/ 0.1)
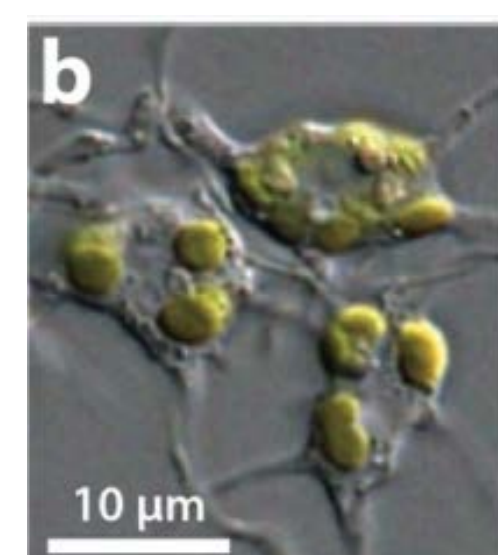unknown (50.1/ 68.1)

Variable

~ 10k variable genes

The nucleomorph-retaining algae *Bigelowiella natans* and *Guillardia theta*

Nucleomorph is the residual nucleus of an algal 2° endosymbiont

1° Endosymbiosis   2° Endosymbiosis

nucleomorph
plastid
remnant cytoplasm
nucleus
mitochondrion

**Bnatans and Gtheta each has 4 genomes.**

Bnatans
1st sequenced Chlorarachniophyta, Rhizaria.
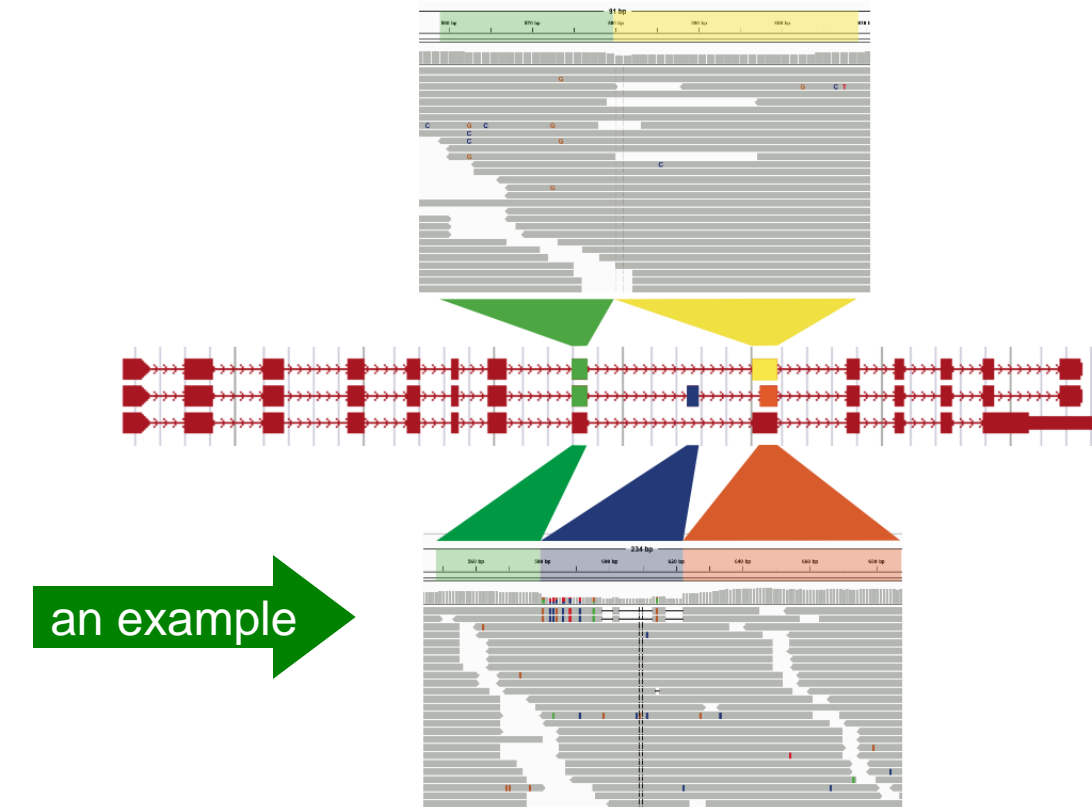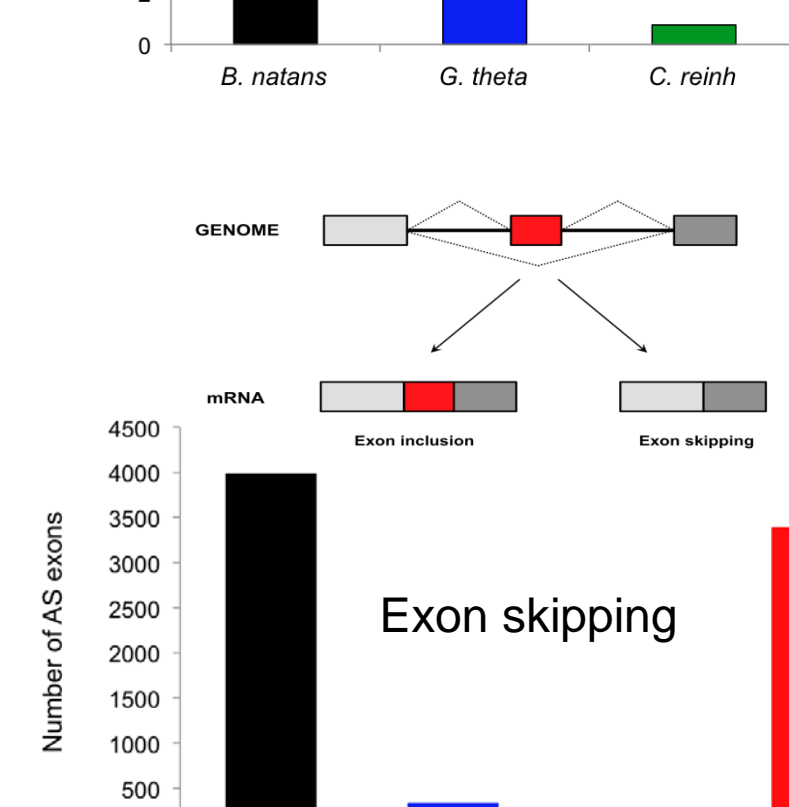
Gtheta
1st sequenced Cryptophyta

genome	size	# chromo.	# genes	size	# chromo.	# genes
nucleomorph	373 kbp	3	283	551 kbp	3	485
chloroplast	69 kbp	1	61	122 kbp	1	147
mitochondrion	37 kbp	1	22	? kbp	1	?
nucleus	95 Mbp	< 302	21380	87 Mbp	< 670	24525

Bnatans has unexpectedly large amounts of alternative splicing (AS)
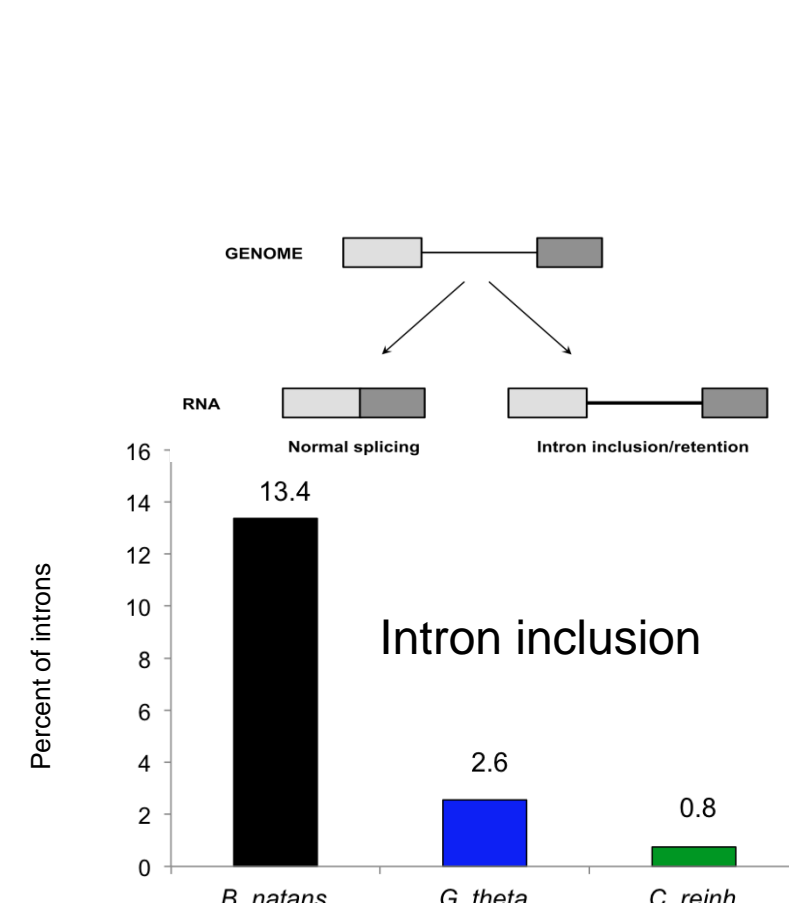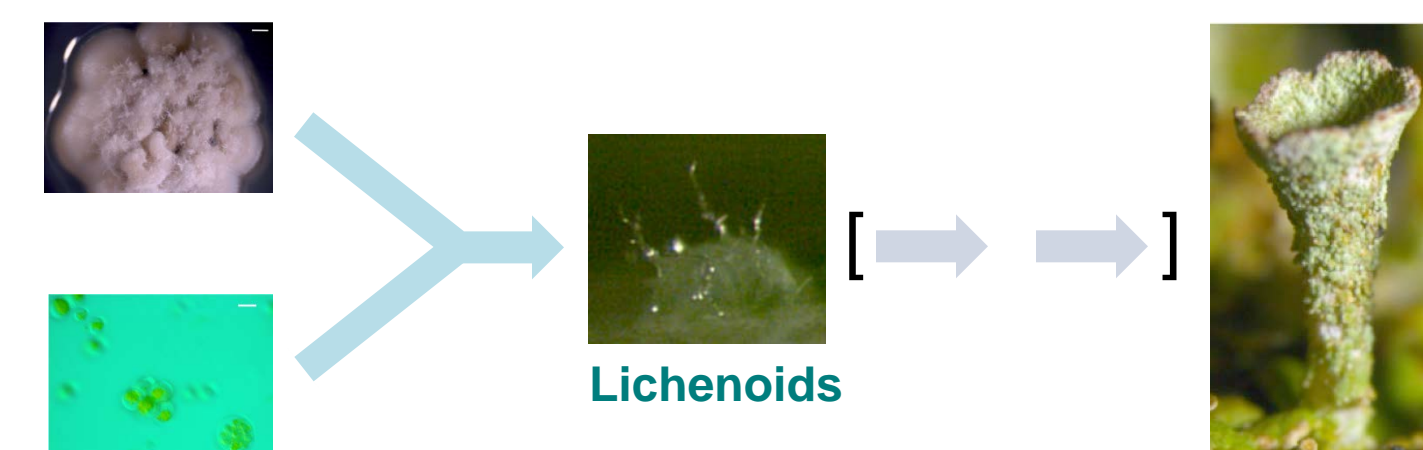
	Bnatans	Gtheta
# genes	21380	24525
# transcripts	27065	25438
Avg. # exons / transcript	9.4	7.1

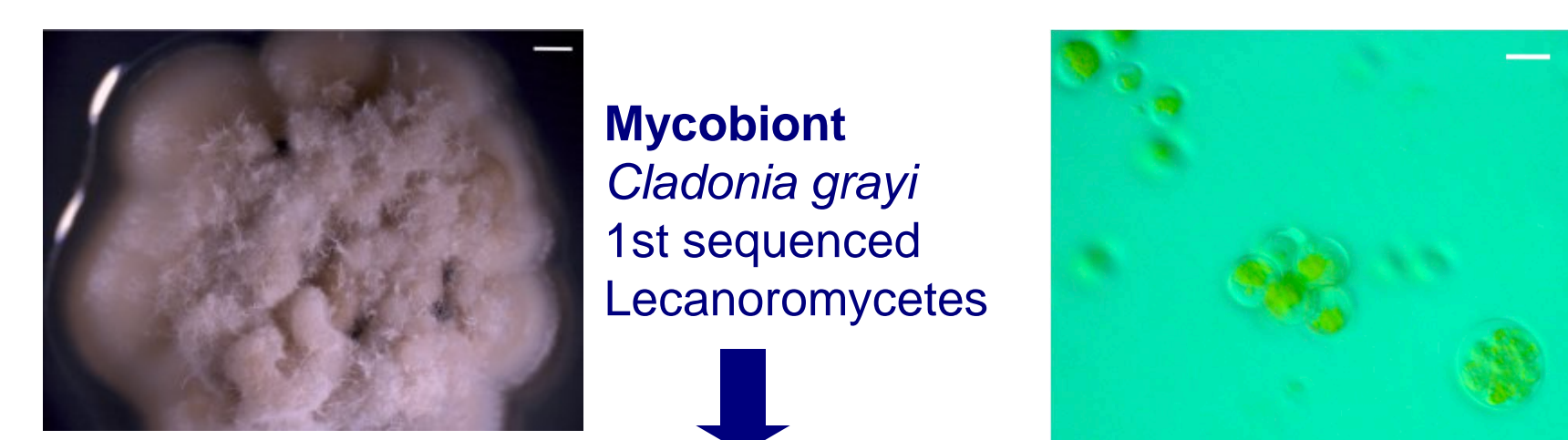Lots of introns.
15% of genes are AS.
33% of transcripts are AS.

Intron inclusion
13.4
2.6
0.8
B. natans   *G. theta*   *C. reinh*

Exon skipping
an example
B. natans   *G. theta*   *C. reinh*   *Mouse*

The lichen photobiont *Asterochloris* sp.

The lichen symbiosis, and its *in vitro* reconstitution.

Lichenoids   [   ]

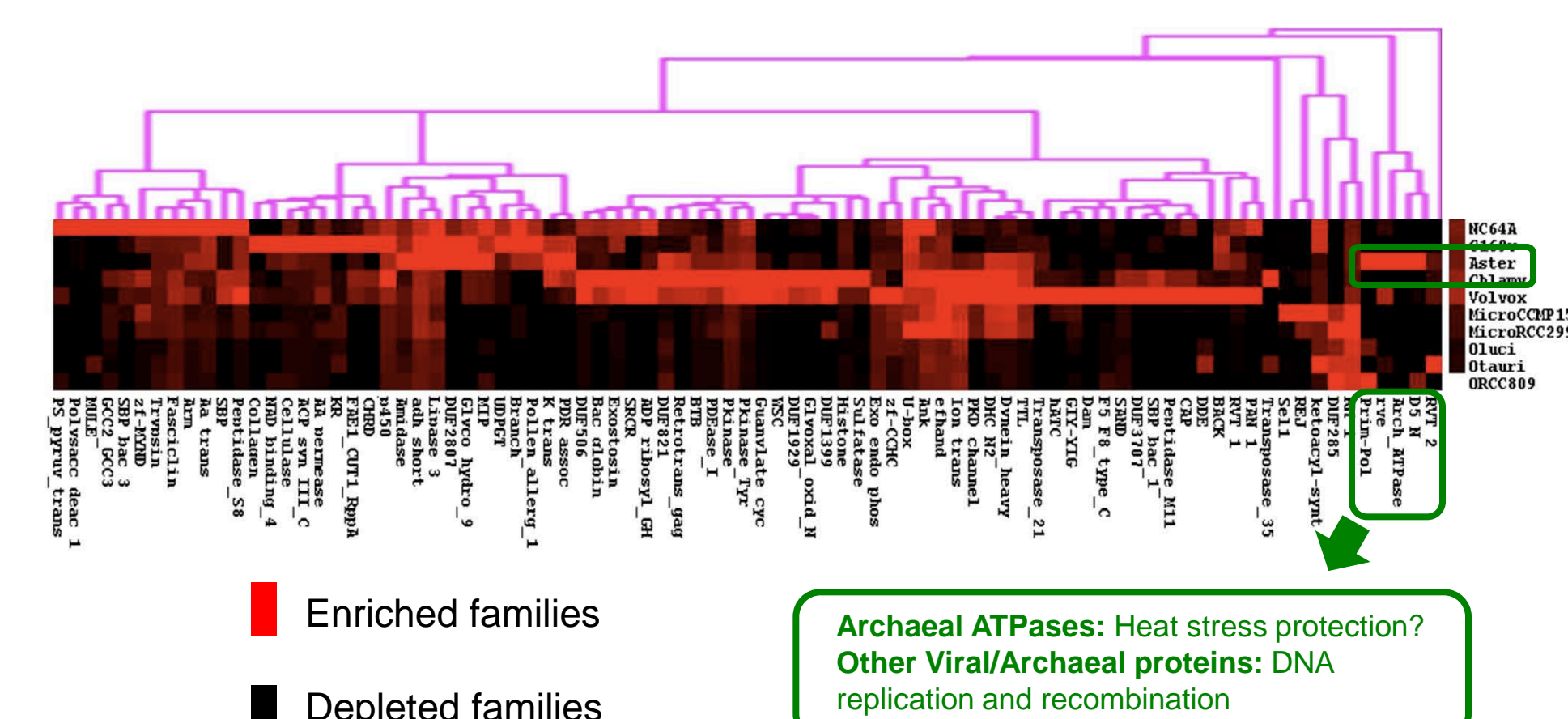**Genomes of the 2 lichen symbionts.**

Mycobiont
Cladonia grayi
1st sequenced
Lecanoromycetes

Photobiont
Asterochloris
3rd sequenced
Trebouxiophyceae,
after *Chlorella* and *Coccomyxa*

Genome size (Mbp)	35	56
# scaffolds	414	151
# genes	11274	9636
# transcripts	11389	10025
Avg. # exons / transcript	3.0	9.0

Gene expression changes in the lichenoids.

Changes in fungal gene expression on day 21 of reconstitution
Mycobiont

Changes in algal gene expression on day 21 of reconstitution
Photobiont
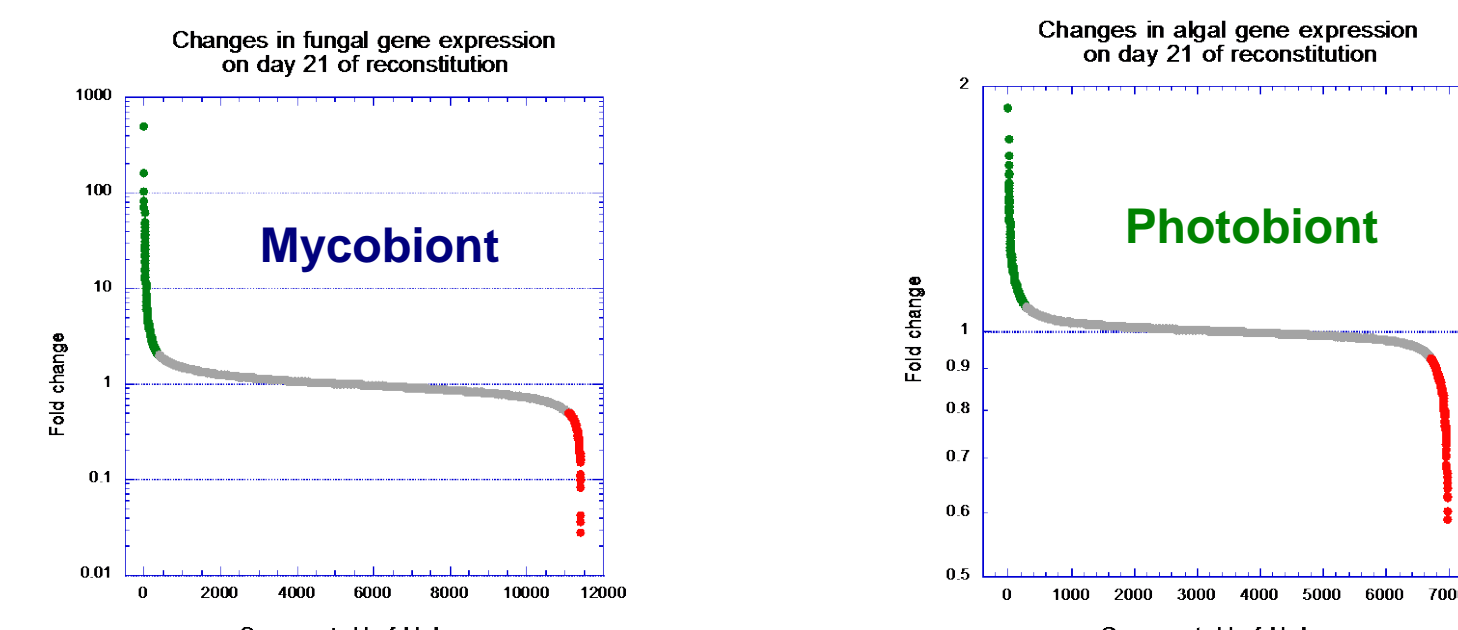
Enrichment or depletion of Pfams in families clustered from *Asterochloris* and 10 other species of Chlorophyta.

Enriched families
Depleted families

Archaeal ATPases: Heat stress protection?
Other Viral/Archaeal proteins: DNA replication and recombination

Credits
These collaborators provided much data and many figures:
Ehux: Betsy Read, Xiaoyu Zhang
Bnatans+Gtheta: John Archibald, Bruce Curtis, Eunsoo Kim, Manuel Irimia
Lichen: Danielle Armaleo, Olaf Mueller