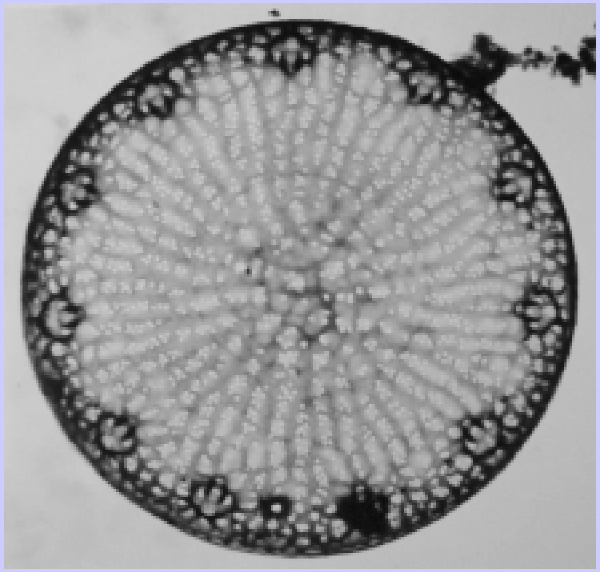
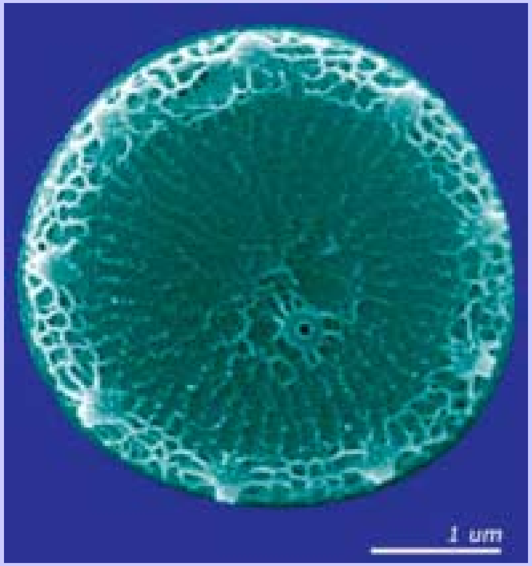
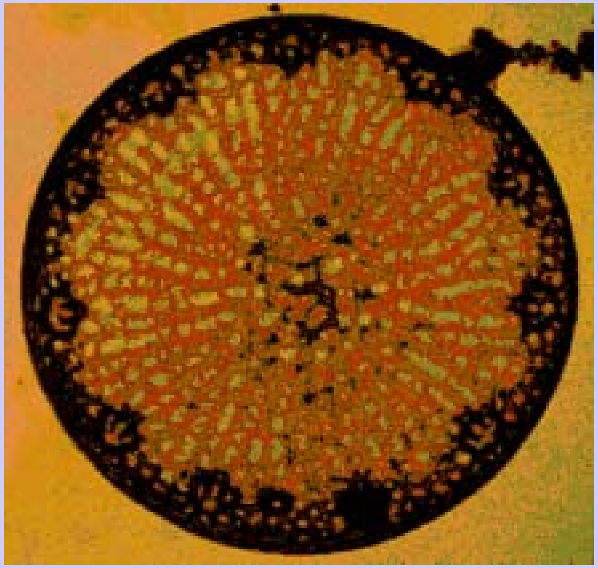
The *Thalassiosira pseudonana* („Thaps“) genome project

Coordinated by Ginger Armbrust, UW, Seattle
Sequenced by the Joint Genome Institute (JGI), Walnut Creek, CA
Financed by Dept. of Energy, i.e. the US Government!







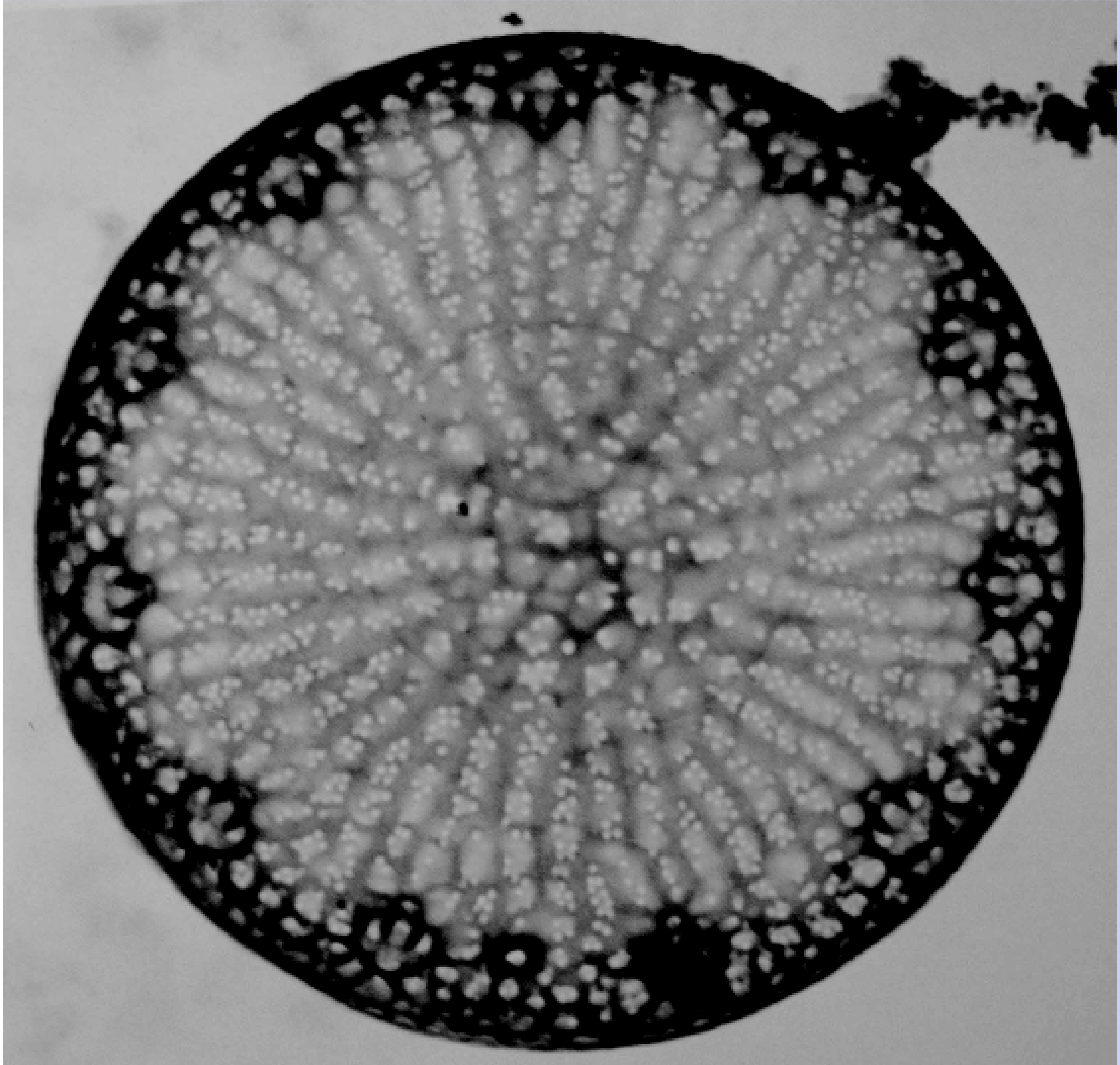


Figure from

Grethe Hasle, Oslo

The JGI:

publically funded, capacity 1 billion bases/month raw data

(i.e. 3 months for the human genome)

Scientists can apply for a certain species

(Future projects: *E.hux.*, *Phaeocystis*, *F. cylindrus*???)

Why a diatom?

Ecologically most important

Interesting physiology (PS, Silica, Iron ...)

Evolutionary interesting (secondary ES)

Why Thaps?

Ecologically relevant

Axenic cultures available

Significant community (incl. Tw.)

Small genome (32 mb)

Approach

(quick, dirty, cheap, and non-elegant)

Subclone genome into (overlapping) BACs/YACs (10^6 – 10^5 bp)
Cut BACs/YACs into small pieces (1- 5 kb), e.g. by shearing
Blunt-end and clone into plasmids
Shotgun sequencing of plasmids by robots
Gene prediction by grail or genewise
Blast databanks against results (Swissprot, ESTs, genomes ...)

Results

~ 1000 contigs sequenced (1 million – a few hundred base pairs), 2 mio runs, 1400 gaps with known sizes, 500 unknown
~ 8.8-fold redundancy, 95% of genome covered
~ 6000 genes with similarities to known genes + 8000 unknown small and few introns

Bearbeiten Ansicht Gehe Communicator Hilfe

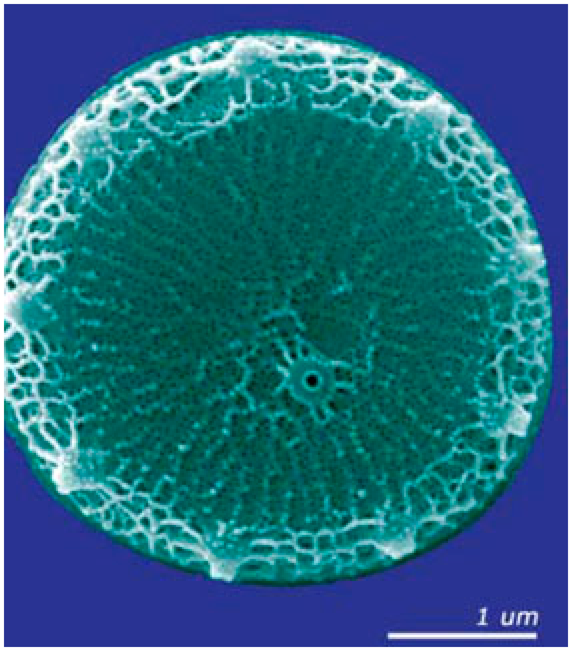
Lesezeichen Adresse: http://shake.jgi-psf.org/thaps1/thaps1.home.html Verwandte Objek

Zurück Vor Neu laden Anfang Suchen Guide Drucken Sicherheit Shop Stop

GI

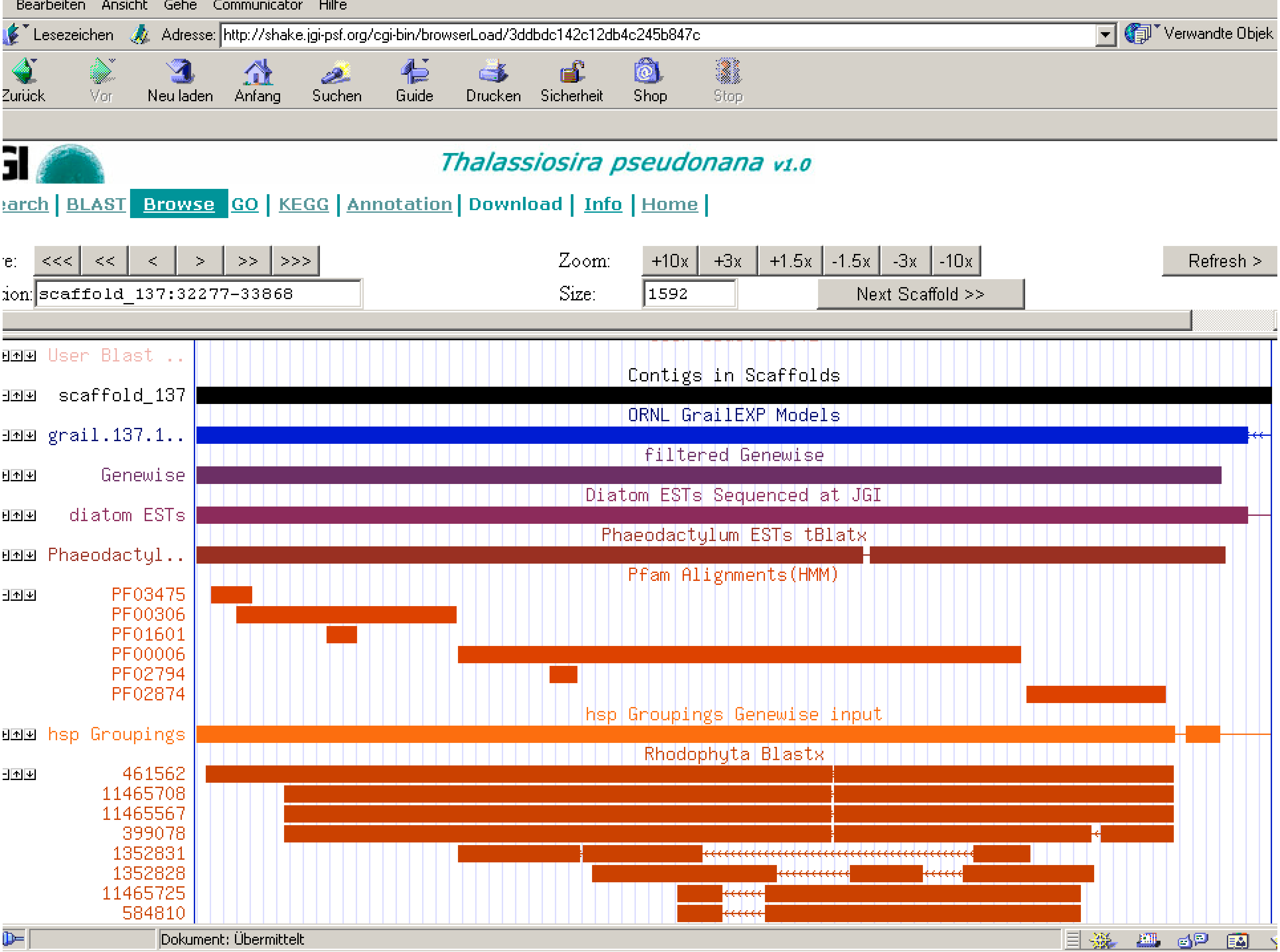
Thalassiosira pseudonana v1.0

Search | BLAST | Browse | GO | KEGG | Annotation | Download | Info | Home



nore information about this organism, contact: Diego Martinez

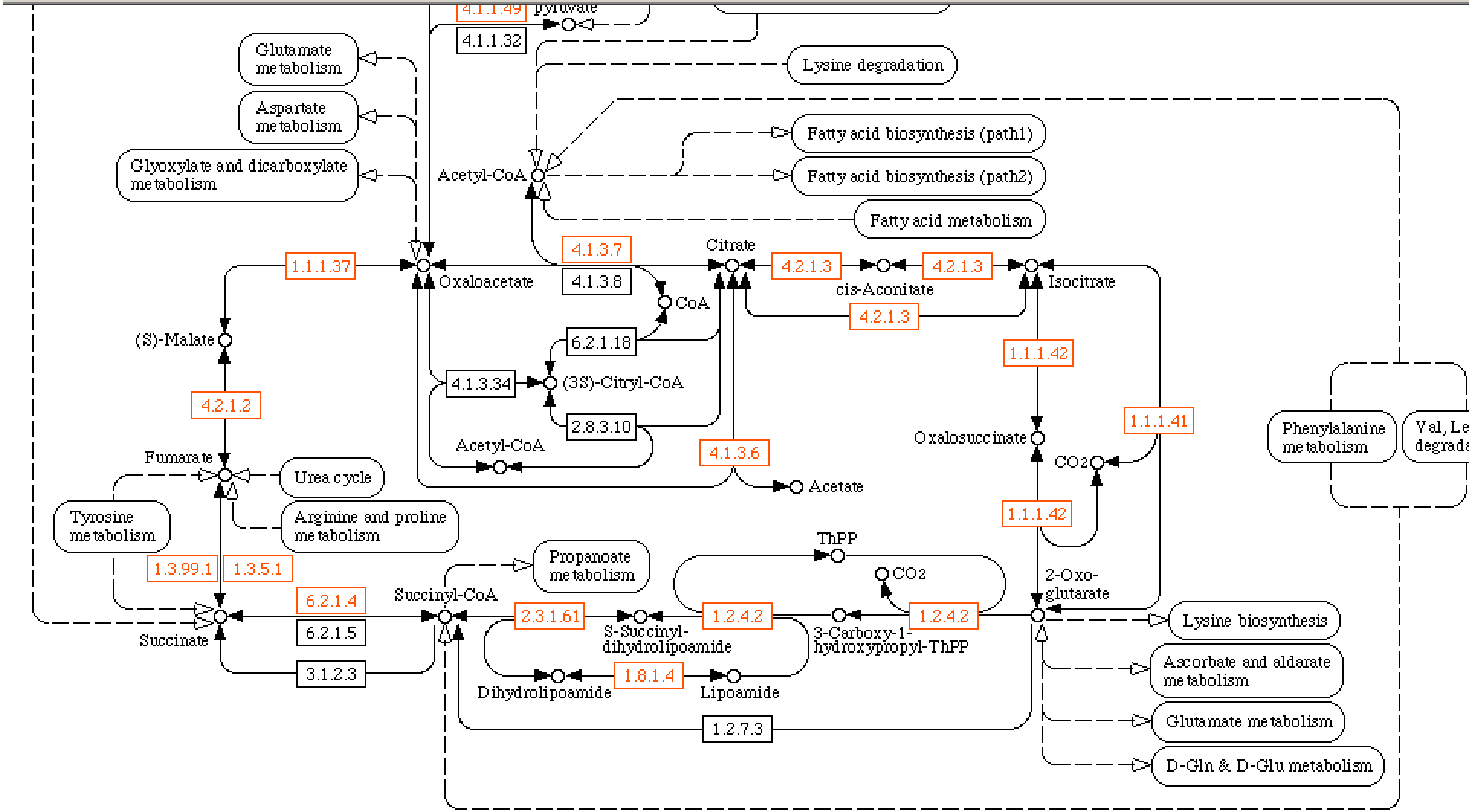
Dokument: Übermittelt


Bearbeiten Ansicht Gehe Communicator Hilfe
Lesezeichen
Adresse: http://shake.jgi-psf.org/cgi-bin/browserLoad/3ddbdc142c12db4c245b847c
Verwandte Objek
Zurück
Vor
Neu laden
Anfang
Suchen
Guide
Drucken
Sicherheit
Shop
Stop
Thalassiosira pseudonana v1.0
earch | BLAST | Browse | GO | KEGG | Annotation | Download | Info | Home |
<<< << < > >> >>>
Zoom: +10x +3x +1.5x -1.5x -3x -10x
Refresh >
scaffold_137:32277-33868
Size: 1592
Next Scaffold >>
User Blast ..
Contigs in Scaffolds
scaffold_137
ORNL GrailEXP Models
grail.137.1..
filtered Genewise
Genewise
Diatom ESTs Sequenced at JGI
diatom ESTs
Phaeodactylum ESTs tBlatx
Phaeodactyl..
Pfam Alignments(HMM)
PF03475
PF00306
PF01601
PF00006
PF02794
PF02874
hsp Groupings Genewise input
hsp Groupings
Rhodophyta Blastx
461562
11465708
11465567
399078
1352831
1352828
11465725
584810
Dokument: Übermittelt

Bearbeiten Ansicht Gehe Communicator Hilfe

Lesezeichen Adresse: .1.1.37+ec:6.2.1.4+ec:6.4.1.1+ec:1.3.5.1+ec:1.1.1.41+ec:1.1.1.42+ec:2.3.1.61+ec:4.2.1.2+ec:4.2.1.3+ec:4.1.1.49+ec:4.1.3.6+ec:4.1.3.7 Verwandte Objek

Zurück Vor Neu laden Anfang Suchen Guide Drucken Sicherheit Shop Stop



0020 9/27/01

map00071.html

Prot name	Prot Id	E-value	Top KEGG hit	EC Num	Enzyme	Annotation
grail.14.167.1	100042	0	KEGG_ath:At2g20420 F11A3.3; succinyl-CoA ligase beta subunit [EC:6.2.1.4] taxId: 3702 taxName: [Arabidopsis thaliana]	6.2.1.4	Succinate--CoA ligase (GDP-forming).	1
grail.2.215.1	101918	4.6e-08	KEGG_dme:CG12231 putative isocitrate dehydrogenase (NAD+) [EC:1.1.1.41] taxId: 7227 taxName: [Drosophila melanogaster]	1.1.1.41	Isocitrate dehydrogenase (NAD+).	1
grail.2.258.1	101961	0	KEGG_vch:VC1141 isocitrate dehydrogenase [EC:1.1.1.42] taxId: 686 taxName: [Vibrio cholerae O1 biovar eltor]	1.1.1.42	Isocitrate dehydrogenase (NADP+).	1
grail.25.38.1	103317	0	KEGG_dme:CG3283 SdhB; succinate dehydrogenase (ubiquinone) iron-sulfur protein precursor [EC:1.3.5.1] [SP:DHSB_DROME] taxId: 7227 taxName: [Drosophila melanogaster]	1.3.5.1	Succinate dehydrogenase (ubiquinone).	1
grail.351.1.1	105476	0	KEGG_sce:YGL062W PYC1, PYV; pyruvate carboxylase [EC:6.4.1.1] [SP:PYC1_YEAST] taxId: 4932 taxName: [Saccharomyces cerevisiae]	6.4.1.1	Pyruvate carboxylase.	1
grail.38.71.1	105822	0	KEGG_cje:Cj0932c pckA; ATP; phosphoenolpyruvate carboxykinase [EC:4.1.1.49] [SP:PPCK_CAMJE] taxId: 197 taxName: [Campylobacter jejuni]	4.1.1.49	Phosphoenolpyruvate carboxykinase (ATP).	1
grail.398.1.1	105975	0	KEGG_mmu:18563 Pcx; pyruvate decarboxylase [EC:6.4.1.1] [SP:PYC_MOUSE] taxId: 10090 taxName: [Mus musculus]	6.4.1.1	Pyruvate carboxylase.	1
grail.398.2.1	105976	0	KEGG_spo:SPBC17G9.11c pyr1; pyruvate carboxylase [EC:6.4.1.1] taxId: 4896 taxName: [Schizosaccharomyces pombe]	6.4.1.1	Pyruvate carboxylase.	1
grail.4.74.1	106052	0	KEGG_sty:STY0181 acnB; aconitate hydratase 2 (citrate hydro-lyase 2) [EC:4.2.1.3] taxId: 601 taxName: [Salmonella typhi]	4.2.1.3	Aconitate hydratase.	1
grail.40.55.1	106348	0	KEGG_ath:At5g66760 MSN2.16; succinate dehydrogenase flavoprotein alpha subunit [EC:1.3.5.1] [SP:DHSA_ARATH] taxId: 3702 taxName: [Arabidopsis thaliana]	1.3.5.1	Succinate dehydrogenase (ubiquinone).	1
grail.42.45.1	106604	0	KEGG_ath:At1g48030 T2J15.6, F21D18.28; lipoamide dehydrogenase precursor [EC:1.8.1.4] taxId: 3702 taxName: [Arabidopsis thaliana]	1.8.1.4	Dihydrolipoamide dehydrogenase.	1
grail.44.10.1	106769	1.69978e-42	KEGG_bme:BMEI1016 fumarate hydratase class I, aerobic [EC:4.2.1.2] taxId: 29459 taxName: [Brucella melitensis]	4.2.1.2	Fumarate hydratase.	1
grail.44.11.1	106770	0	KEGG_bme:BMEI1016 fumarate hydratase class I, aerobic [EC:4.2.1.2] taxId: 29459 taxName: [Brucella melitensis]	4.2.1.2	Fumarate hydratase.	1
grail.50.12.1	107624	6.7e-33	KEGG_mtu:Rv3339c icd1; Isocitrate dehydrogenase [EC:1.1.1.42] [SP:IDH_MYCTU] taxId: 83332 taxName: [Mycobacterium tuberculosis H37Rv]	1.1.1.42	Isocitrate dehydrogenase (NADP+).	1
grail.58.60.1	108311	1e-20	KEGG_dme:CG1516 putative pyruvate carboxylase [EC:6.4.1.1] taxId: 7227 taxName: [Drosophila melanogaster]	6.4.1.1	Pyruvate carboxylase.	1

Bearbeiten Ansicht Gehe Communicator Hilfe

Lesezeichen Adresse: http://shake.jgi-psf.org/cgi-bin/dispGeneModel.v4?db=thaps1&id=100042 Verwandte Objek

Zurück Vor Neu laden Anfang Suchen Guide Drucken Sicherheit Shop Stop

:MAREST	SCS1

iotate this protein

note this curated Model

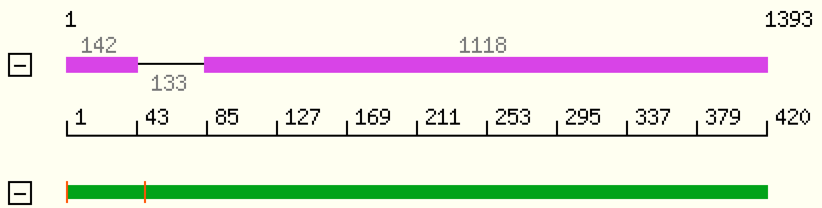
w nucleotide and 3-frame translation   To Genome Browser

ıl hits: 203

: shown: 10

ısmodium: 1 Synechocystis: 1 Anabaena: 0 Thermosynechococcus: 0 Arabidopsis: 5

grail.14.167.1   To Genome Browser



.p Start	End	Len	Pct	Score	Description [taxName]
24	416	422	93%	1101	At2g20420-68297.m02019 F11A3.3 succinyl-CoA ligase beta subunit ;supported by full-length [Arabidopsis thaliana]
24	416	421	93%	1101	SUCB_ARATH Succinyl-CoA ligase [GDP-forming] beta-chain, mitochondrial precursor (Succinyl-CoA [Arabidopsis thaliana]
24	416	421	93%	1101	(NM_127601) succinyl-CoA ligase beta subunit; protein id: At2g20420.1, supported by cDNA: [Arabidopsis thaliana]
24	416	421	93%	1101	ath:At2g20420 F11A3.3; succinyl-CoA ligase beta subunit [EC:6.2.1.4] custom_KEGG [Arabidopsis thaliana]
24	416	421	93%	1095	(AY087596) succinyl-CoA ligase beta subunit [Arabidopsis thaliana]
51	459	466	88%	1063	(AAAB01008968) agCP2799 [Anopheles gambiae str. PEST]
35	445	502	82%	1062	(AE003681) CG11963-PA [Drosophila melanogaster]
35	445	465	88%	1062	dme:CG11963 putative succinate--CoA ligase (ADP-forming)

Dokument: Übermittelt

Bearbeiten Ansicht Gehe Communicator Hilfe

Lesezeichen Adresse: http://shake.jgi-psf.org/cgi-bin/noNameYet?db=userBlast&scaffoldDb=thaps1&name=query&table=blast_1037884797_28098 Verwandte Objek

Zurück Vor Neu laden Anfang Suchen Guide Drucken Sicherheit Shop Stop

GI *Thalassiosira pseudonana* v1.0

earch | **BLAST** | Browse | GO | KEGG | Annotation | Download | Info | Home |

ize:

)0

ize:

ıScore:

ı%id:

ı%pos:

ıLength:

ırt:

d:

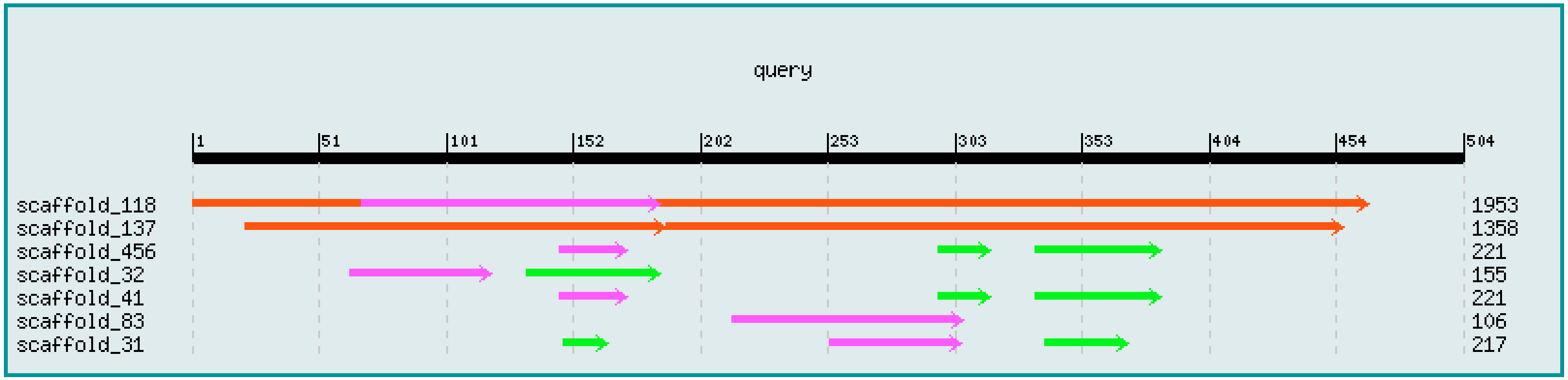
)4

et stuff

Name	query
DefLine	query
Length	504
Hits	7

mouse over to get info

7 hits shown



http://shake.jgi-psf.org/cgi-bin/go3.cgi?species_db=thaps1&dbname=go_eu

My personal interest and approach

1 – Look for interesting pathways/genes, especially with respect to their evolutionary history (sec genes, GOGAT, clp genes ...)

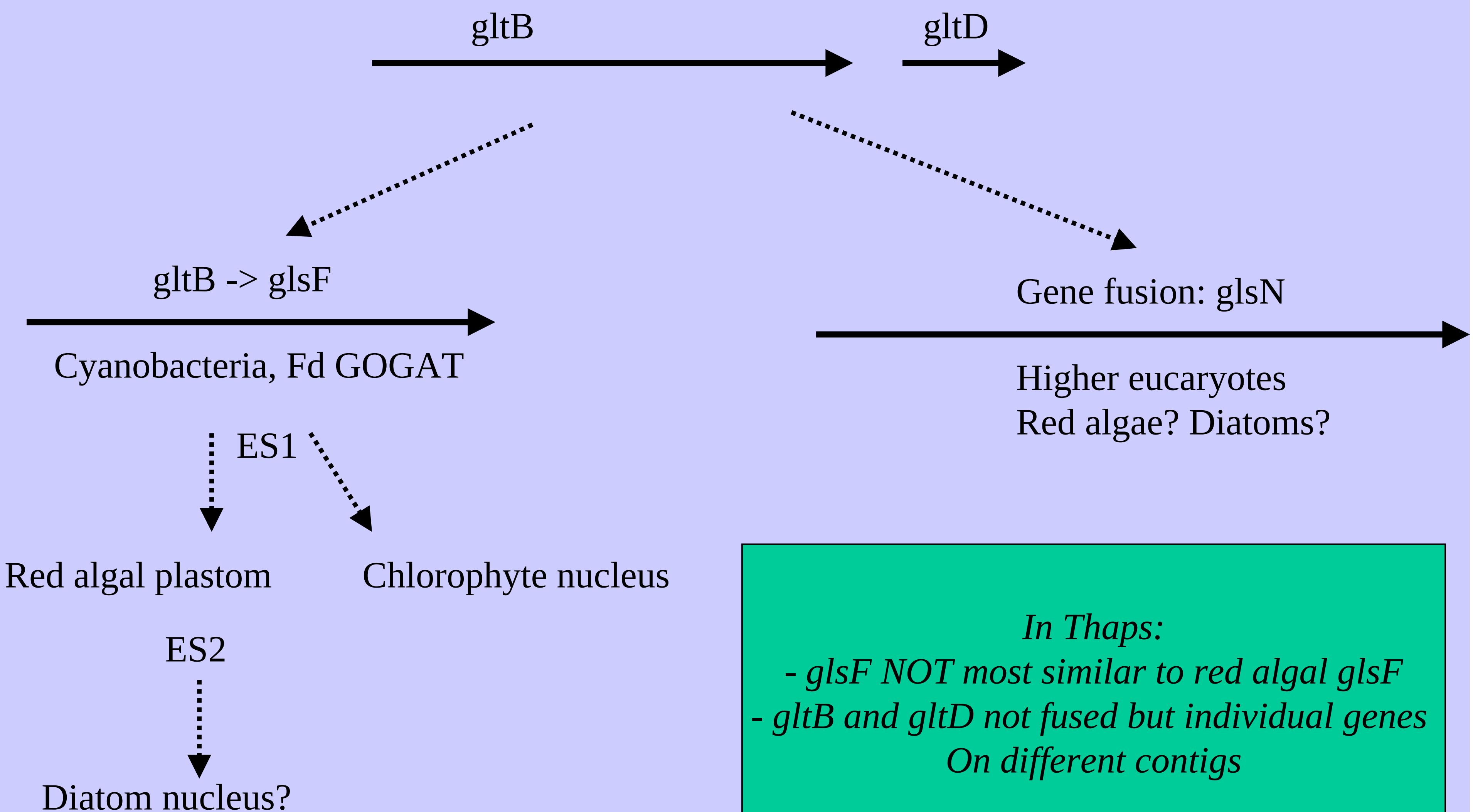
2 – Look for gene transfer events:

a. From the red algal plastome to the diatom nucleus

b. From the red algal nucleus to the diatom nucleus

Interesting genes 1, GOGAT (Glutamate synthase)

(Gln + OG -> 2 Glu, Fd or NADPH dependent)



In Thaps:

- glsF NOT most similar to red algal glsF

- gltB and gltD not fused but individual genes

On different contigs

Interesting genes 2, clp protease

protease, chaperones, heat shock proteins

Catalytical SU (clpP), regulatice SU (clpABC)

clpABC: Plastid copy in reds and browns, nuclear copy of plastid homologue in A. thal.
clpC: Plastid copy in green, nuclear copy in reds/browns?

In Thaps:

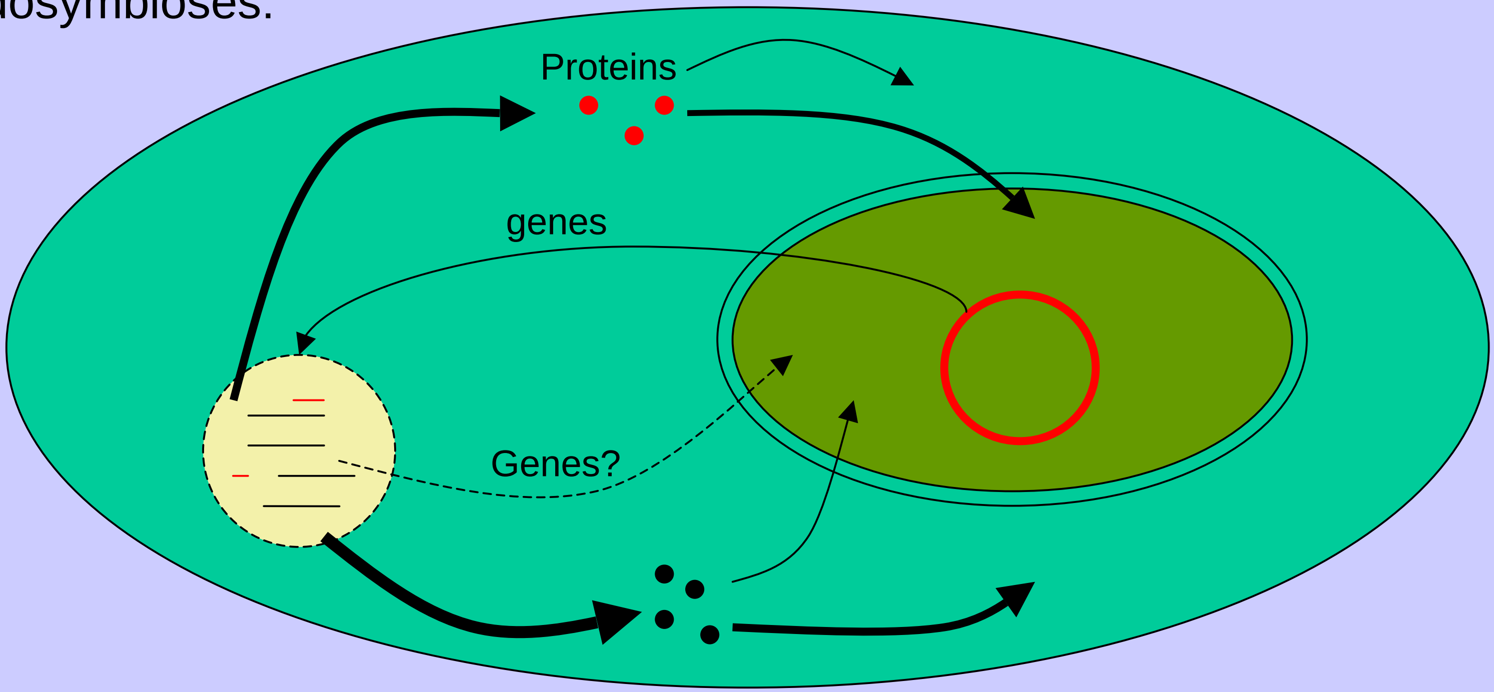
1 plastid clpC copy + 5-6 clpABC copies in the nucleus
6-7 clpC copies in the nucleus

Alpha purple-like sequences for both SU: mitochondrial?

1-2 clpC homologues are clearly nucleomorph-like: secondary gene transfer

Gene transfer during Endsymbiosis establishment

Primary endosymbioses:



Process

Gene transfer pt/nu
Gene recruitment pt
Gene recruitment cyt
Gene transfer nu-pt

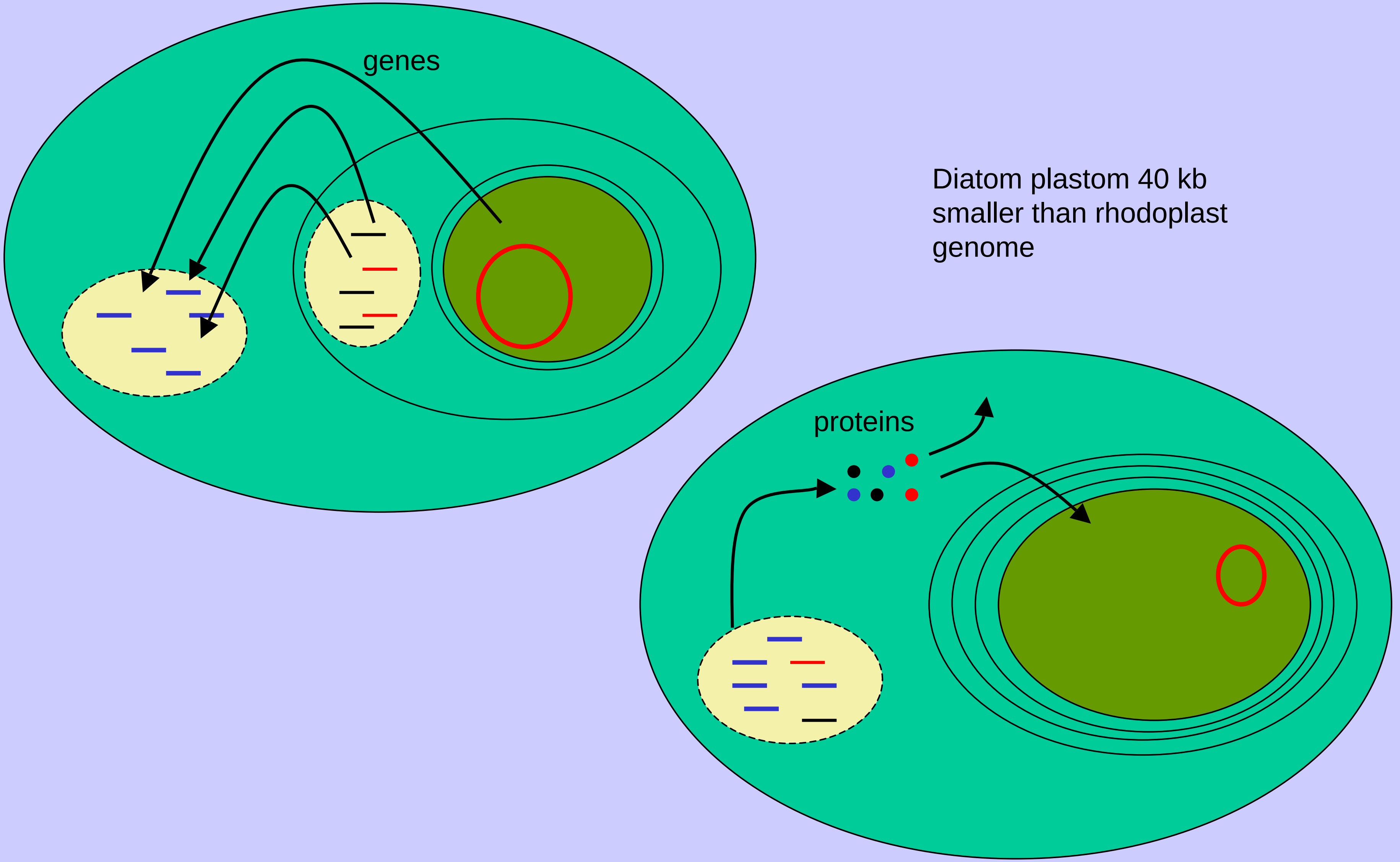
Examples

rbcS and > 1000 others
N-Gogat, GS, HSPs
glycolysis
???

Mechanisms

Whole genome transfer
mRNA transfer, RT
Random generation of transit peptides or by exon shuffling
Develop import machinery from cyanobacterial proteins?

Secondary Endosymbiosis



Diatom plastom 40 kb
smaller than rhodoplast
genome

Questions - approaches

- Are there any former red algal plastid genes in the Thaps genome?

- Are there any former red algal nuclear encoded plastid genes in the Thaps genome?

- Are there any fomer red algal nuclear genes in the Thaps genome

-… what about the red mitochondria?

- Blast red algal plastid genes against Thaps, namely those missing in Diatoms.

- Blast red algal nuclear genes against Thaps, namely those encoding plastid proteins

- Blast *G. theta* nucleomorph genes against Thaps.

- Blast red algal mitochondrial genes against Thaps.

Some results

The Mitchondrial genome:

- 53 genes from *Chondrus*, *Cyanidioschyzon*, and *Porphyra* tested
- 29 (!!!) genes did not produce a hit, including all ORFs, most *rp*SL genes, RT, most *sdh* genes, *ccm*C, *secY*.
- 5 genes only matched pt homologues.
- All of the hits, exept one, were to 4 contigs: 235, 325, 347, and 377. These seem to be mitochondrial.
- Only one hit was to a nuclear contig, i.e. contig 25. This was a subunit of succinacte dehydrogenase (*sdh*2). The Thaps homologue was most similar to animal homologues, not to red algal ones
- Two Thaps genes (*nad*1, *cox*1) matched best to chlorophytic homologues from *Nephroselmis*.

Conclusions

- Many mt genes are missing in Thaps – probably the mt genome is not complete
- There is no evidence for gene transfer from the red algal mt genome to the Diatom nucleus.

Some results

The rhodoplast genome:

- ca. 70 genes analysed
- > 40 genes did not produce a hit
- 16 genes matched best to homologues not from red algae or higher plants
- 4 genes (cfxX, ycf53, cobA, groEL) were of red algal origin
- 8 genes (hisH, 2xycf19, trX, frxB, 3xftsH) showed a chlorophytic signature

- psbW (pt-encoded in rhodophytes and in diatoms) exists in two copies, one on the plastome, the second in the nucleus, indicating on-going gene transfer.
- trpAB, consisting an operon in red algal plastoms, are fused on the nuclear genome in Thaps as in yeast. They are not fused in green plants

Conclusions

- In comparison to rhodoplasts, diatom plastoms miss some 40-80 genes. Most of them were NOT transferred from the rhodoplast to the diatom nucleus.

- Only 4 genes can still be traced back to the red algal ancestor

- The majority of the missing genes was likely replaced with existing nuclear copies, some of which are of chlorophytic origin

Some results

The red algal nuclear genome:

- 39 genes analysed, some of them producing many hits.
- 16 genes did not produce a hit.
- Of the remaining 23 genes 9 produced 41 matches that did not hit red algae or higher plants.
- 24 hits were of red algal origin, representing 6 genes (lhcA, sat2, ftsZ, ubi, atpC, psbU).
- 17 hits (representing 8 genes, lhcA, gabA, sat2, frx, fabH, glsF, atpC, pmsR) showed a chlorophytic signature.

- Interestingly for several genes (lhcA, sat, atpC) there were green *and* red maches.

Conclusions

- Many former red algal nuclear genes can be traced to the Thaps genome, i.e. a higher percentage than for the plastid genome. Most of these encode plastid proteins

- A significant number of genes show a chlorophytic signature, again mostly ancoding plastid protein.

Some results
The Gtheta genome

- 51 gene and ORFs of chromosome 1 analysed
- 25 produced (sometimes many) hits
- 26 (mainly ORFs) produced no hit
- 7 best matches against Gtheta, mostly plastid proteins
- 14 best matches with green plants, mostly cytoplasmatic proteins

Interesting genes 3, psbU

Nuclear gene in red algae
Known also in cyanobacteria
Not found in green plants, e.g. Arabidopsis

Nuclear psbU copy in diatoms?

In Thaps:

1-2 nuclear psbU copies
More similar to red algal homologues than to cyanobacterial ones
Indicates early secondary gene transfer from red algal nucleus
To diatom (host) nucleus

Interesting genes 4, psbW

Plastid gene in red algae and another diatom (Odontella), nuclear-encoded in green plants

In Thaps:

Two psbW copies, one on the plastome 80% similar to Odontella, the other in the nucleus, 70% similar.

Indicates ongoing gene transfer from the complex plastid to the diatom nucleus

What else is possible?

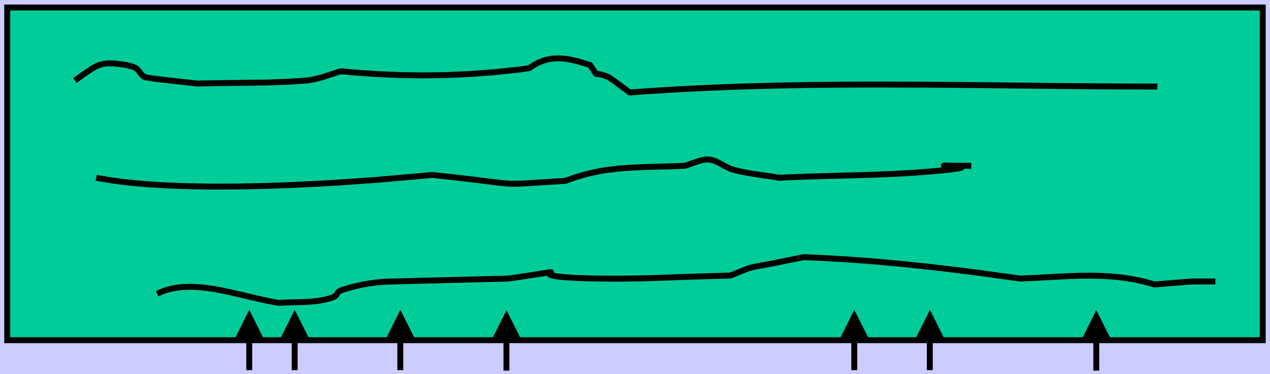
- Look for the presence of biochemical pathways (silica, PKS, etc.)
- Use diatom sequences for primer design of desired genes
- Reconstruct gene pyhlogenies using multi-gene alignments
- DNA-chips and proteomics

How can you access the data?

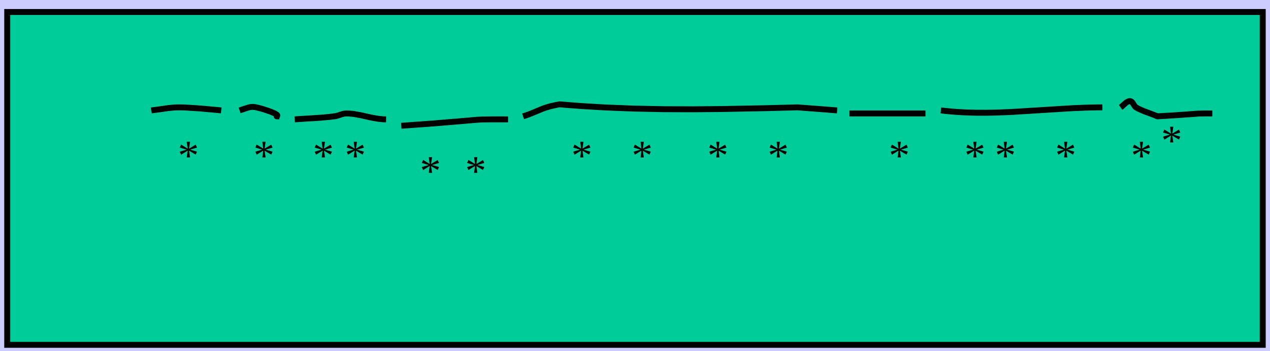
- Data will become publically available in may
- Right now I can search for genes and oligos for you

Nice new method – „optical mapping“ (OM) of genomes

Linearize chromosomes on slide

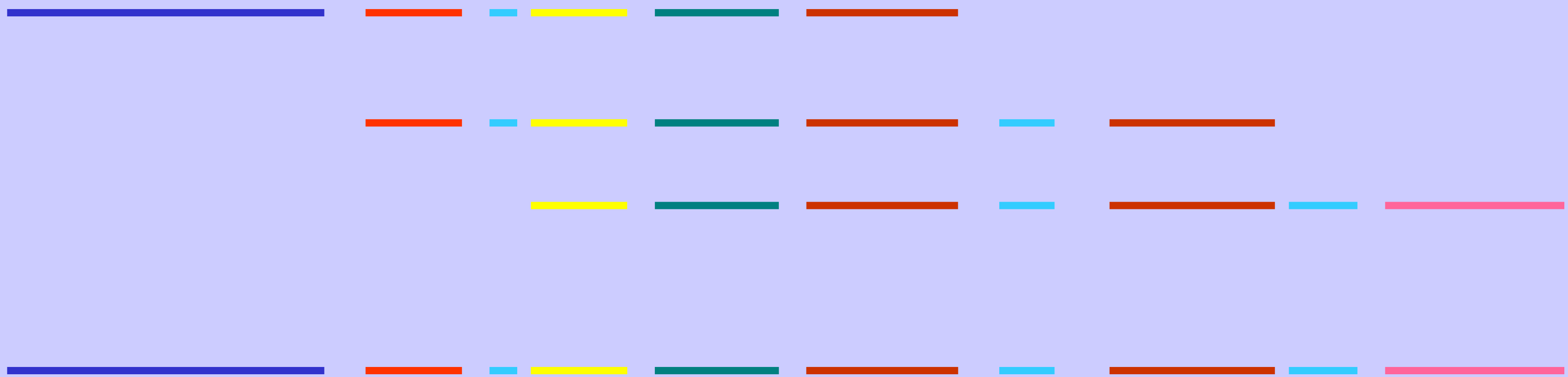


Cut with restriction enzymes



Label and scan with a fluorescence microscope
(about 60000 individual molecules/30 hr)

Align patterns, reconstruct contigs or entire genomes



→ String of masses: „bar code of the genome“

Applications:

- Create in *silico maps* and compare to OM
- Compare genome variability between strains, i.e. toxic/non-toxic, high light/low light, colony-forming/non colony-forming, …
- Determine chromosome numbers, genome sizes, ploidity, complexity (rep. seq.),

OM results in Thaps

- Compare real OM to *in silico* OM
- Determine genome size
- Determine number of chromosomes, identifiy differences between sister chromosomes
- Determine structure of the plastid genome, i.e. the plastome exists in circular trimers and hexamers of a 133 kb monomer