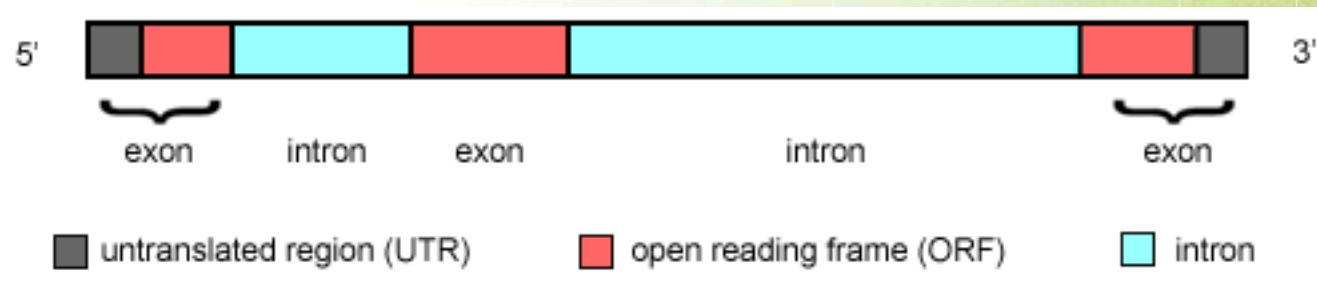
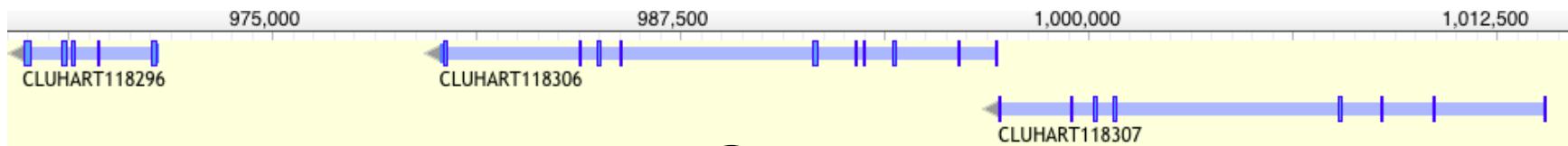


Functional annotation



Enabler for Life Sciences

Right, now we have our genes, but what do they do?



Insulin receptor?

Vesicle-trafficking protein?

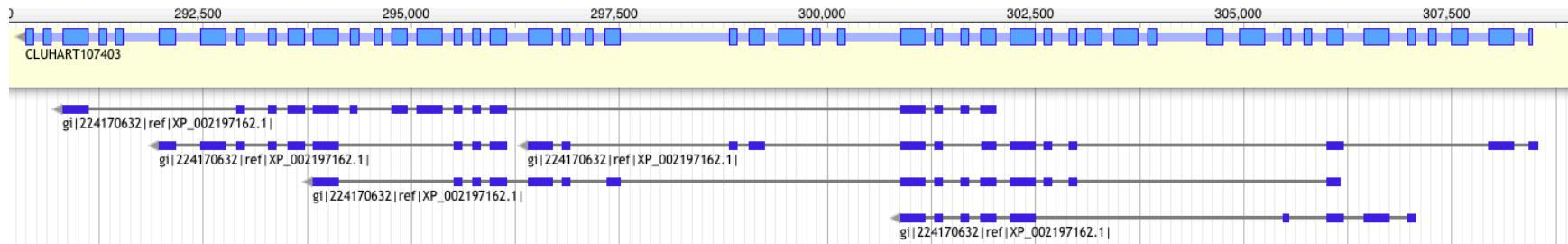
Alcohol dehydrogenase?

Aquaporin?

Transcription factor

MAP kinase kinase kinase?

But we have used proteins in our annotation!



It is actually kind of complex...

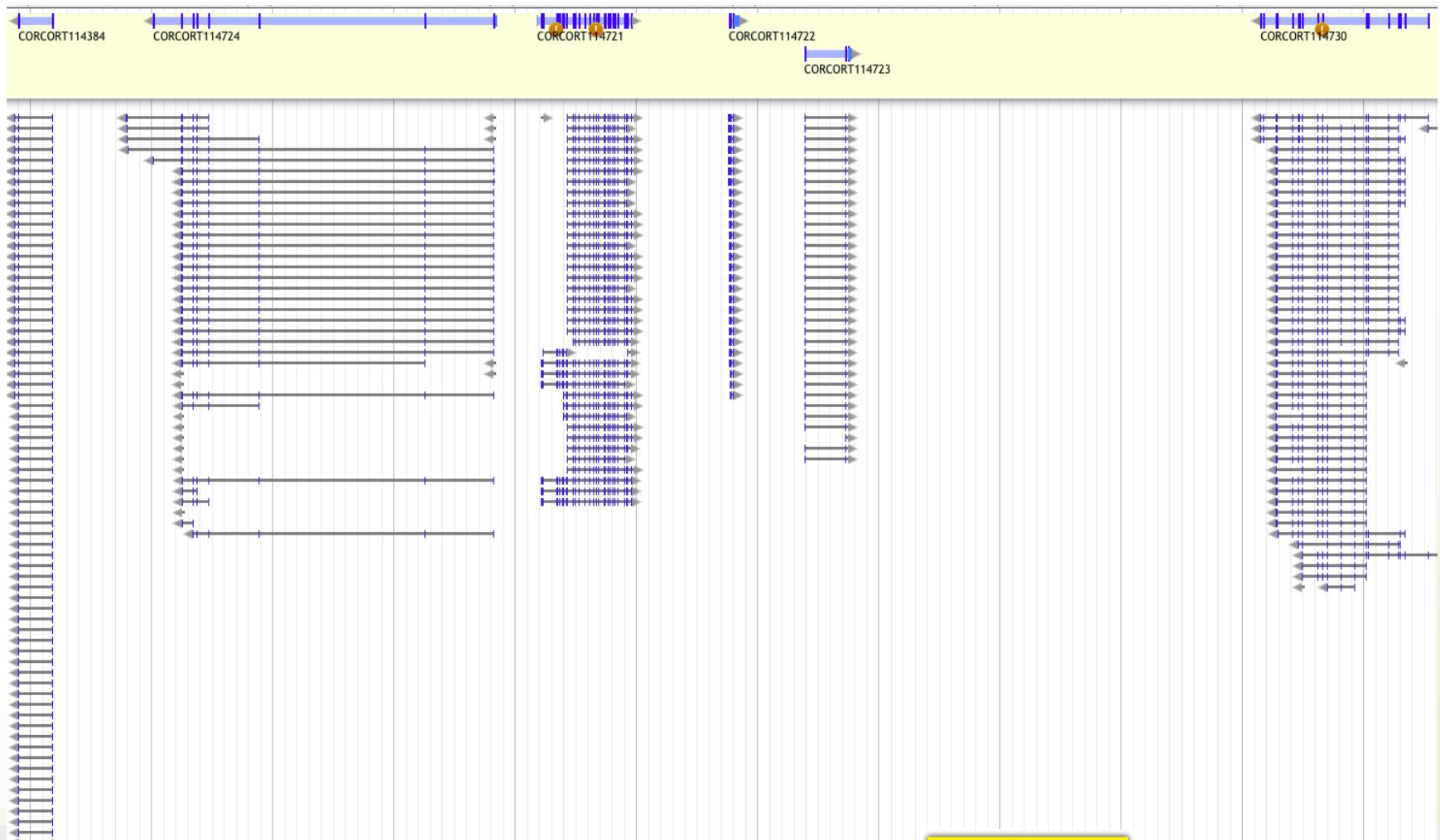
... and Maker does not do this for you.

Extract sequences -> functional annotation

- Extract sequences from Webapollo or use gffread (in Cufflinks package)
- Annotate the sequences functionally using Blast and InterProScan
- Or use the (mostly) commercial alternative



Blast-based annotation - proteins



Blast-workflow

- Blast the protein-sequences from your maker-run to a protein blast-database (e.g., uniprot) using blastp from the Blast+ package
- Use Annie to extract best hits from blast-hit list and the corresponding description from uniprot-headers
- Add this information to the annotation.gff using custom-script

Blast-based approach

- Fairly fast and easy
- Orthology not certain - best blast-hit does not equal orthologous!

Interproscan

www.ebi.ac.uk/interpro/interproscan.html

Reader

An Introduction to the Gene Ontology

About InterProScan 5 < InterPro < EMBL-EBI

Services | Research | Training | About us

EMBL-EBI

InterPro
Protein sequence analysis & classification

Search InterPro... Examples: IPR020405, kinase, P51587, PF02932, GO:0007165

Home | Release notes | Training & tutorials | FAQs | Download | About InterPro | Contact

About InterProScan

What is InterProScan?

InterProScan is the software package that allows sequences (protein and nucleic) to be scanned against InterPro's signatures. Signatures are predictive models, provided by several different databases (referred to as member databases), that make up the InterPro consortium.

The software is available:

- As a web-based tool, using the sequence search box on the [InterPro homepage](#), for the analysis of single protein sequences (also available in the [EBI tool section](#))
- Programmatically via Web services that allow up to 25 sequences to be analysed per request (both [SOAP](#) and [REST](#)-based services are available)
- As a downloadable package for local installation from the EBI's FTP server, for instructions see the [detailed documentation pages](#).

InterProScan is run regularly against UniProtKB and the results are made available via the InterPro website.

More information

For more information, and for instructions on how to obtain, install and run InterProScan, please see the [detailed documentation pages](#).

Publications

InterProScan 5: genome-scale protein function classification
Philip Jones, David Binns, Hsin-Yu Chang, Matthew Fraser, Weizhong Li, Craig McAnulla, Hamish McWilliam, John Maslen, Alex Mitchell, Gift Nuka, Sébastien Pesquet, Antony F. Quinn, Amaia Sangrador-Vegas, Maxim Scheremetjew, Siew-Yit Yong, Rodrigo Lopez, and Sarah Hunter
Bioinformatics, Jan 2014
(doi:10.1093/bioinformatics/btu031)
[HTML](#) - [PDF \(324Kb\)](#)

EMBL-EBI

News
Brochures
Contact us
Intranet

Services

By topic
By name (A-Z)
Help & Support

Research

Overview
Publications
Research groups
Postdocs & PhDs

Training

Overview
Train at EBI
Train outside EBI
Train online
Contact organisers

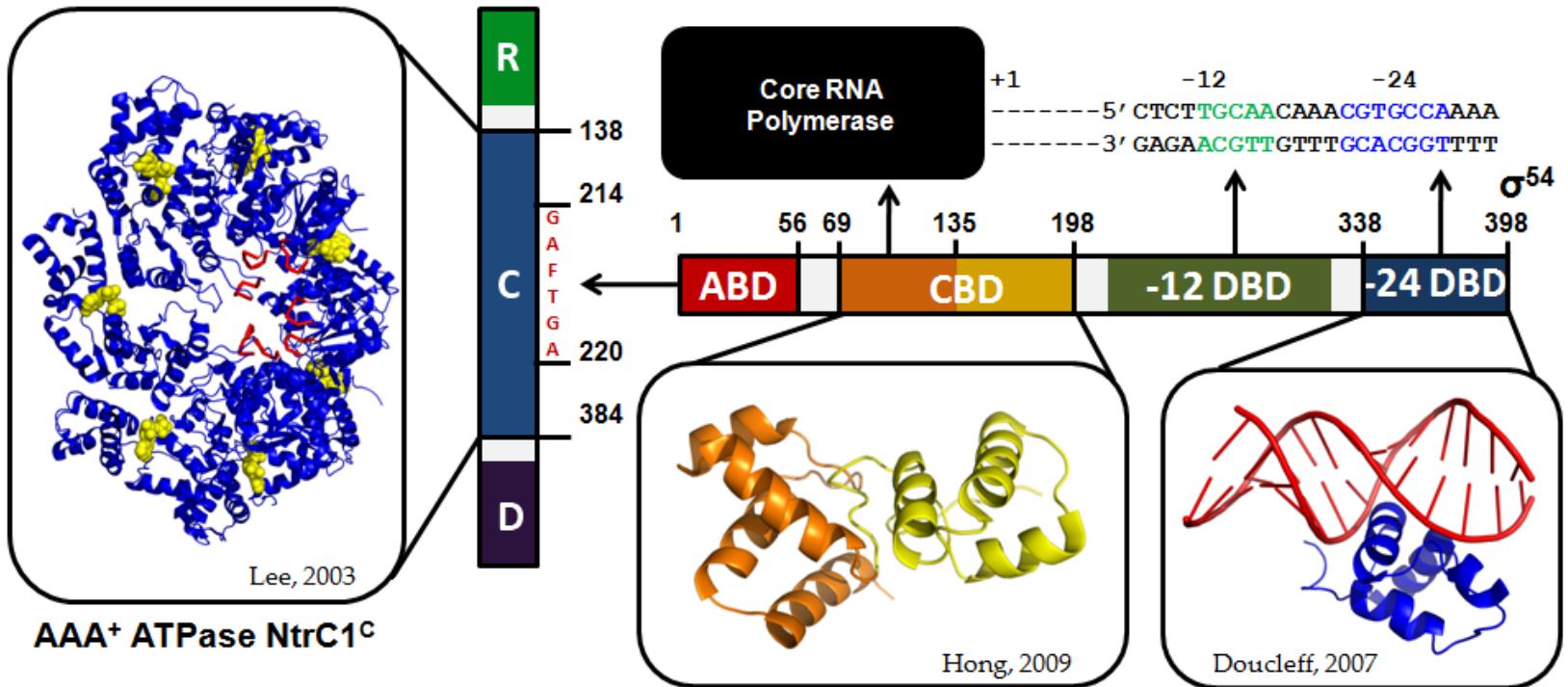
Industry

Overview
Members Area
Workshops
SME Forum
Contact Industry programme

About us

Overview
Leadership
Funding
Background
Collaboration
Jobs
People & groups
News

Sequence domains

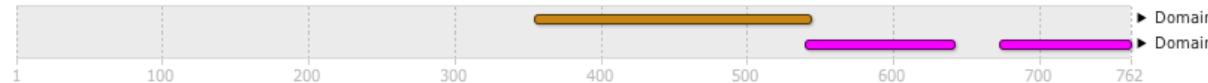


Interproscan results

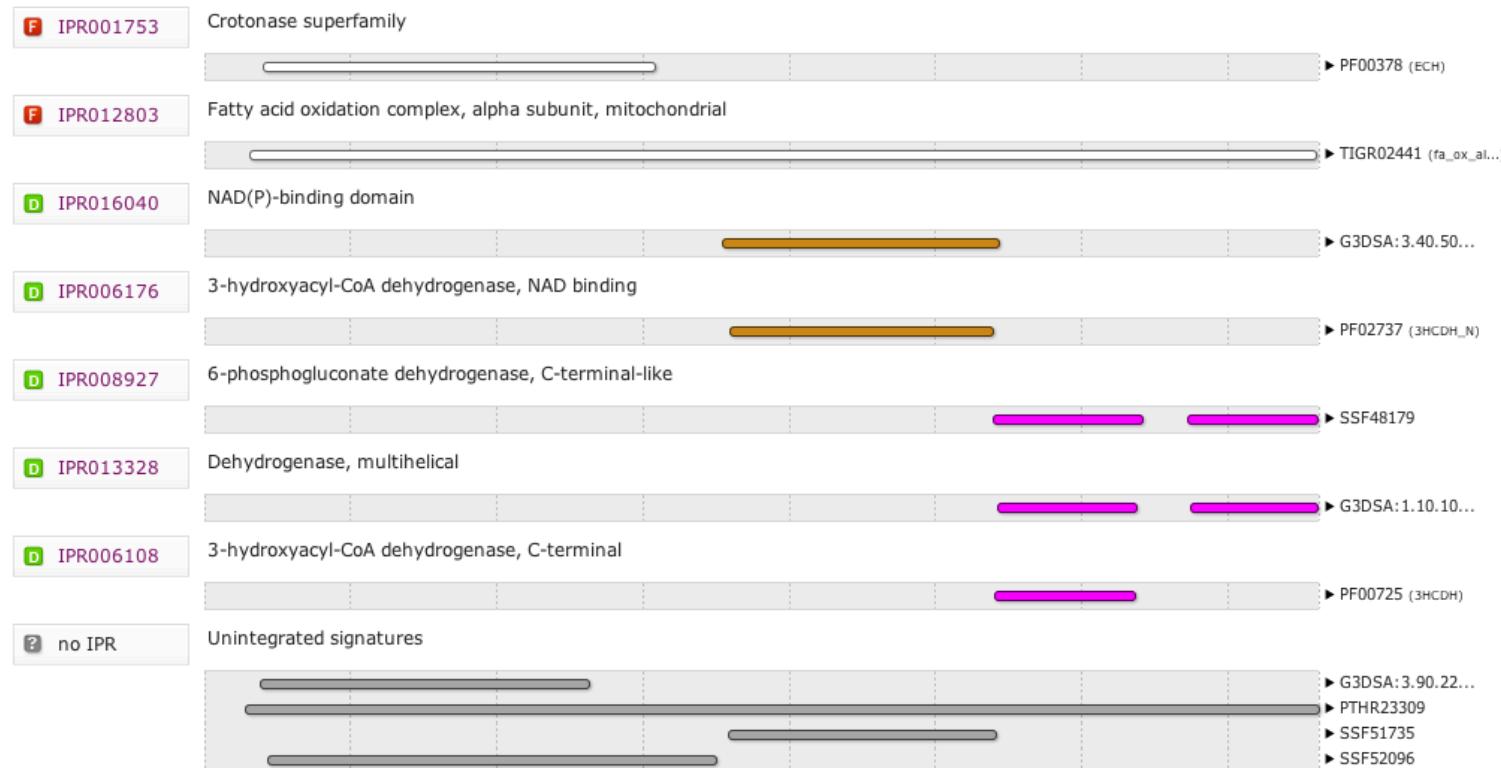
Protein family membership

- ↳ F Crotonase superfamily (IPR001753)
 - ↳ F Fatty acid oxidation complex, alpha subunit, mitochondrial (IPR012803)

Domains and repeats



Detailed signature matches



Interproscan results - GO terms

GO term prediction

Biological Process

-  [GO:0006631](#) fatty acid metabolic process
-  [GO:0006635](#) fatty acid beta-oxidation
-  [GO:0008152](#) metabolic process
-  [GO:0055114](#) oxidation-reduction process

Molecular Function

-  [GO:0003824](#) catalytic activity
-  [GO:0003857](#) 3-hydroxyacyl-CoA dehydrogenase activity
-  [GO:0004300](#) enoyl-CoA hydratase activity
-  [GO:0016491](#) oxidoreductase activity
-  [GO:0016616](#) oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor
-  [GO:0050662](#) coenzyme binding

Cellular Component

-  [GO:0005739](#) mitochondrion
-  [GO:0016507](#) mitochondrial fatty acid beta-oxidation multienzyme complex

Gene Ontology

The Gene Ontology

Genome assembly Wiki SILL BILS projects Wikipedia Google Maps

the Gene Ontology

Search go!

Downloads Tools Documentation Projects About Contact

Welcome to the Gene Ontology website!

The Gene Ontology project is a major bioinformatics initiative with the aim of standardizing the representation of gene and gene product attributes across species and databases. The project provides a controlled vocabulary of terms for describing gene product characteristics and gene product annotation data from GO Consortium members, as well as tools to access and process this data. [Read more about the Gene Ontology...](#)

Search the Gene Ontology Database

Search for genes, proteins or GO terms using AmiGO:

GO!

AmiGO is the official GO browser and search engine.

The Gene Ontology project very much encourages input from the community into both the content of the GO and annotation using GO. We are very happy to work with others to ensure that the GO is both complete and accurate, and we also very much encourage communities to submit GO annotations for inclusion in the GO database. [Please contact us.](#)

The Gene Ontology Consortium is supported by a U41 grant from the National Human Genome Research Institute (NHGRI) [grant HG002273]. See the full list of funding sources. The Gene Ontology Consortium would like to acknowledge the assistance of many more people than can be listed here. Please visit the acknowledgements page for the full list.

Copyright © 1999-2013 the Gene Ontology.
[Helpdesk](#) • [Cite](#) • [Terms of use](#) • [News](#) • [RSS](#)
Member of the [Open Biological and Biomedical Ontologies](#)

Quick Links

- Tools
- AmiGO browser A
- Submit GO Annotations
- OBO-Edit ontology editor
- Ontology downloads
- Annotation downloads
- Database downloads
- Documentation
- GO FAQ
- GO on SourceForge
- Contact GO

News

- GO on Twitter
- Finding updates...
- GO newsdesk
- GO news RSS feed
- GO on Facebook

Open "http://www.geneontology.org/GO.downloads.annotations.shtml" in a new tab

Gene Ontology

- C: Cellular Component, e.g., endoplasmatic reticulum, integral to plasma membrane
- P: Biological Process, e.g., pyrimidine metabolic process
- F: Molecular Function, e.g., catalytic activity, transporter activity

Interproscan workflow

- Run Interproscan on the protein fasta file created by maker
- Use Maker-supplied scripts to merge the interproscan-results to the Maker annotations.gff file

Interproscan approach

- Looks for conserved domains, so might be more reliable than blast?
- How to go from conserved domains to assigning a function for your protein?

- Combines a blast-based search with a search for functional domains
- Blast at NCBI -> picks out GO terms based on blast hits and uniprot -> statistical significance test -> done!
- Interproscan

Blast2GO

/Users/hobbe/Documents/Artemis_files_current/blast2go_20101001_0816.dat - Blast2GO V.2.4.4

File Blast Mapping Annotation Analysis Statistics Select Tools View Info

GO:0007067;GO:0016021 transport;binding;apoptos SPO_2518.DDX18_HUMAN

nr	sequence name	seq description	length	#...	min. eValue	sim mean	#G...	GO IDs	Enzyme	InterPro
3884	gene_3884 GeneMark...	c6 transcription	977	20	1.0E-171	59.85%	7	F transcription factor activity; F zinc ion binding; P regulation of transcription, DNA-dependent; C transcription factor complex; F transporter activity; C membrane; P transmembrane transport		IPR005829; IPR007219
3885	gene_3885 GeneMark...	hypothetical protein NFIA_039100 [Neosartorya fischeri NTRL 181]	312	20	1.0E-39	63.15%	1	C viral capsid		no IPS match
3886	gene_3886 GeneMark...	sin3 complex subunit	870	20	0.0	73.2%	0			-
3887	gene_3887 GeneMark...	mitochondrial intermembrane space translocase subunit	87	20	1.0E-40	88.55%	5	F metal ion binding; P rotein import into mitochondrial inner membrane; C mitochondrial inner membrane; C mitochondrial intermembrane space protein transporter complex; P transmembrane transport		IPR004217; PTHR11038 (PANTHER), PTHR11038:SF8 (PANTHER)
3888	gene_3888 GeneMark...	lysyl-tRNA synthetase	592	20	0.0	73.55%	7	C ytosol; P auxin biosynthetic process; F nucleic acid binding; F lysine-tRNA ligase activity; P lysyl-tRNA aminoacylation; F ATP binding; P lysine biosynthetic process	EC:6.1.1.6	IPR004364; IPR004365; IPR006195; IPR012340; IPR016027; IPR018149; IPR018150; G3DSA:3.30.930.10 (GENE3D), SSF5568 (SUPERFAMILY)
3889	gene_3889 GeneMark...	transcription factor conserved hypothetical protein [Aspergillus clavatus NTRL 1]	1569	20	0.0	70.9%	0			-
3890	gene_3890 GeneMark...	udp-gluc gal endoplasmic reticulum nucleotide	240	20	1.0E-51	56.25%	0	C integral to membrane; C endoplasmic reticulum membrane; P transmembrane transport; P carbohydrate transport		IPR013657; PTHR10778 (PANTHER)

GO Graphs Application Messages Blast/IPS Results Statistics Kegg Maps

```

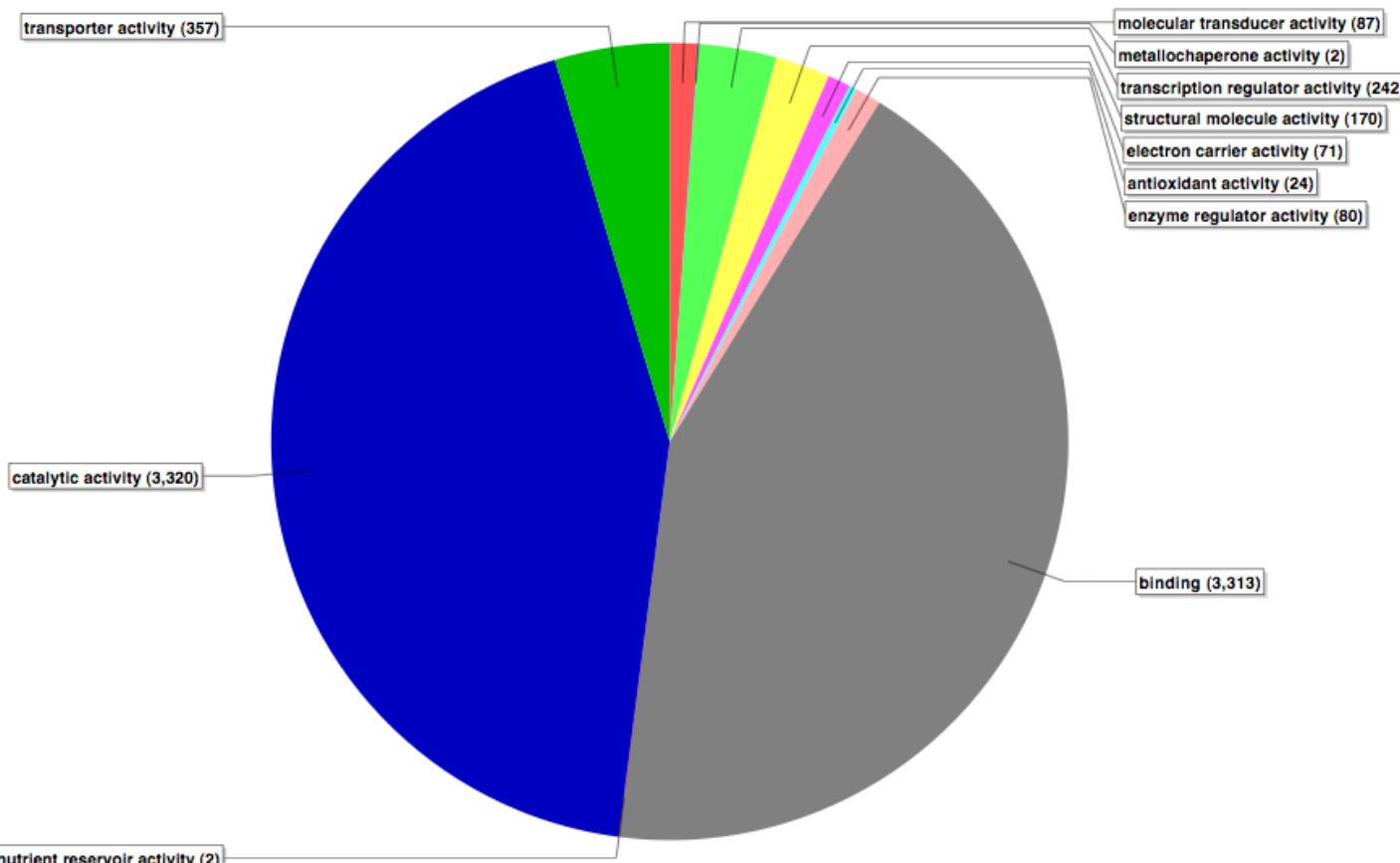
17:59 InterProScan for gene_8871|GeneMark.hmm|286_aa done.
17:59 -----
17:59 InterProScan Result:
17:59 InterProId: IPR001715
17:59 InterProName: Calponin-like actin-binding
17:59 InterProType: Domain
17:59 DB-Name: GENE3D - G3DSA:1.10.418.10
17:59 InterProId: IPR016146
17:59 InterProName: Calponin-homology
17:59 InterProType: Domain
17:59 DB-Name: SUPERFAMILY - SSF47576
17:59 InterProId: noIPR
17:59 InterProName: unintegrated
17:59 InterProType: unintegrated
17:59 DB-Name: PANTHER - PTHR19961
17:59 DB-Name: PANTHER - PTHR19961:SF9

```

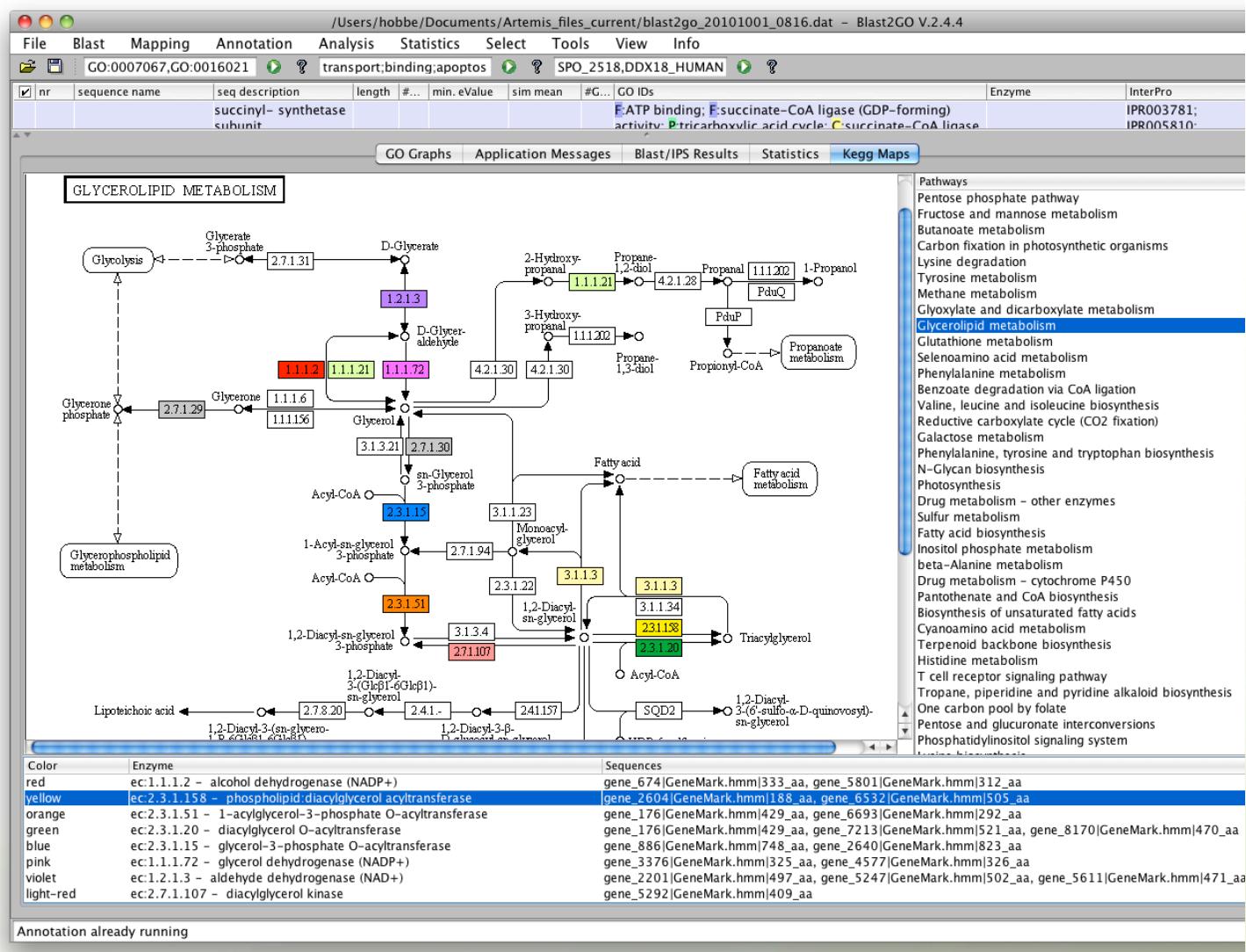
Annotation already running



molecular_function Level 2



KEGG-mapping



Liftovers are very useful for orthology determination

- Kraken
- Align the two genomes (Satsuma) and then transfer annotations between aligned regions

