



LiceBase

Model organism database and functional genomics tools for the sea lice research community

Michael Dondrup, Christian Andreetta, Inge Jonassen, Frank Nilsen

michael.dondrup@ii.uib.no

admin@licebase.org




Introduction



Motivation

- Building a community of sea lice experts
 - Provide first class tools and data
 - Work collaboratively on curation of genome and data


[wFleaBase](#) | [BLAST](#) | [BioMart](#) | [GBrowse Maps](#) | [Genomics](#) | [Help](#)

Search: [Gene pages](#) ↓ [\[?\]](#)

wFleaBase
Daphnia Genome project

Genome of *Daphnia pulex* and other species, including bulk data files, and all [gene pages](#)

Genomics Tools including microsatellites, cDNA, Cosmid and BAC libraries, GSS and ESTs, microarrays

Database Genome database development notes, tools and programs

Maps of *Daphnia* genome, and genome annotation tools

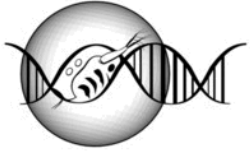
References to research literature

Help & Documents

News

[About this genome web](#)

[Send comments to](#)
 daphnia [AT] iubio.bio.indiana.edu



wFleaBase
Daphnia Water Flea
Genome Database

2010 April: *Daphnia Genes 2010* beta release.
2007 July: First public release of *Daphnia pulex* Genome is here
and see these [genome summaries](#).

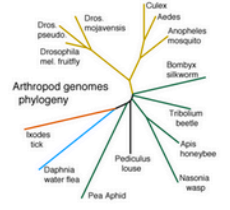
This web service provides gene and genomic information for species of the genus *Daphnia* - commonly known as the water flea.

The freshwater crustacean *Daphnia* is a model system for ecology, evolution and the environmental sciences. The rapidly growing genomic data for this organism is stimulating interdisciplinary research to understand the complex interplay between genome structure, gene expression, individual fitness, and population-level responses to chemical contaminants and environmental change.

wFleaBase includes data from all species of the genus, yet the primary species are *D. pulex* and *D. magna*, because of the broad set of genomic tools that have already been developed for these animals.

A complete sequence for *Daphnia pulex* is now available at this site. Please observe this [Data release policy](#). The data is a first characterization of the crustacean genome, which was made possible by the U.S. Department of Energy (DOE) [Joint Genome Institute](#) (JGI) in collaboration with the [Daphnia Genomics Consortium](#) (DGC) whose members were funded by the National Science Foundation.

Compare *Daphnia* to Arthropods



2010 poster
Daphnia genomes at the crossroads

The Salmon Louse Genome Project



- Genome size: ~ 600Mb
- Sequencing: 454, Illumina (SE, PE), Sanger
- Coverage: ~300X
- Assembly: LSalAtl2s - Newbler, Abyss, SSPACE, Linkage map,
- Scaffolds: ~33,000
- Gene prediction: MAKER pipeline, EBI/Ensembl
- Predicted gene count: ~13,081

Partners: The Institute for Marine Research (IMR), MPI Berlin and Cologne,
GenomeBC, SLRC, UiB, EBI

Salmon Louse Genomics Data Deluge



- Transcriptome of all developmental stages + tissues (23 and growing)
- RNA-interference experiments
- DNA re-sequencing from wild samples
- Microarray data + oligo sequences
- >100k ESTs
- Linkage map*
- Related organisms (*Caligus sp.*, *Calanus finmarchicus*)
- Salmon genome*
- Genetic markers
- SNP chip*
- Micro-biome
- Mapping spliced RNA-seq to genome (STAR, Bowtie, ...)
- normalized expr. values: CPM (edgeR)
- InterProScan 5 + GO
- GO-enrichment analysis for RNA-seq
- Aligned all Genbank ESTs
- Aligned all microarray oligos
- Ensembl Compara
- KAAS: Pathway reconstruction
- NR/Swissprot homologues, paralogues, *C.finmarchicus* homologues
- Meta-genomics pipeline

Computation

Sustainable - Established Technology

- Based on GMOD tools (gmod.org)



- Genome Browser:



- Genomics web-pages:

Drupal™



- Database backend:

CHADO

- Authentication: SAML SSO (FEIDE)



- Analysis pipelines (by end of 2014)

NeLS /



Search this site:

>> Logout

LiceBase

- ▷ In-situ images
- ▷ RNAi
- Search Blast Results
- Search Publications
- Forum

Michael Dondrup

- ▽ Search Biological Data
 - Features
 - Organisms
 - Analyses
- Analysis
- Analysis: Blast
- Analysis: Interpro
- Analysis: KEGG
- Analysis: Unigene
- ▷ Blogs
- Browse in-situ images

The premier resource for sea louse genomics

Welcome to LiceBase

Track

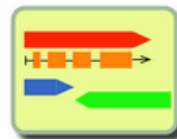
Please see our featured blog post: **Feature: Help, my favorite gene is missing in the genome!**



BLAST



RNA interference



GBrowse



Biological data

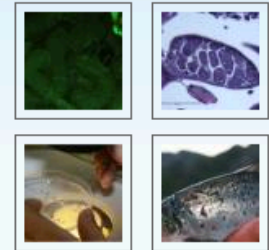


Publications

Sea lice like the salmon louse are a major factor for economical loss in the salmon farming industry world wide. LiceBase is developed at the Sea Lice Research Centre (SLRC), a new centre for research based innovation (CRI) established through the Norwegian Research Council and hosted by University of Bergen. The CRI has a long term funding (5+3 years) from both the research council and the consortium partners (4 academic partners: University of Bergen, Institute of Marine Research, Norwegian School of Veterinary Science, UNI and 5 industrial partners: Novartis Animal Health, EWOS Innovation, Patogen Analyse AS, Marine Harvest ASA, Lerøy Seafood Group ASA). The research is organized in 6 different work packages.

LiceBase is a system for storing and mining the data from within this and related projects and is aimed to be the preferred site for the global research community for access to genome data on sea lice. The work will involve both development and application of bioinformatics tools to facilitate data representation and analysis (e.g., genome annotation, integration of data from genome-wide

Recent Images



Recent blog posts

- Feature: Help, my favorite gene is missing in the genome!
- News: BLAST and other major updates
- News: Reporting Tool for Annotation Errata
- News: Added Raw-data downloads, publications and other goodies
- [IMPORTANT] Mandatory Primary Target ID in RNAi
- Enabled multi-target RNAi experiments.
- Migration to new db host
- [NEWS] New tracks with all

BLAST



Search this site:

>> Logout

LiceBase

- ▷ In-situ images
- ▷ RNAi
- Search Blast Results
- Search Publications
- Forum

Michael Dondrup

- ▽ Search Biological Data
 - Features
 - Organisms
 - Analyses
- Analysis
- Analysis: Blast
- Analysis: Interpro
- Analysis: KEGG
- Analysis: Unigene
- ▷ Blogs
- Browse in-situ images
- Feature
- Galleries
- Gallery
- Image
- In-situ image
- My account

Run BLAST

Query Comment (optional, will be added to output for your use):

NOTE: If the input sequence is less than 30 letters you should change the default Cutoff Score value to something less than 100 or you can miss matches.

Upload Local TEXT File: FASTA, GCG, and RAW sequence formats are okay

WORD Documents do not work unless saved as TEXT. No file chosen

Type or Paste a Query Sequence : (No Comments, Numbers Okay)

- L. salmonis scaffolds: LSalAtI2s (low-complexity masked)
Everything L. salmonis – all DNA sequences(assemblies, ESTs, mitoch., rDNA, ...)
- L. salmonis scaffolds: LSalAtI2s (unmasked)
- L. salmonis extra sequences: Mitochondrial + rDNA
- ESTs: Consensus Clusters(CL) and singletons
- L. salmonis masked ESTs only
- L. salmonis all Genbank ESTs
- L. salmonis Ensembl predicted protein sequences
- L. salmonis mitochondrium, complete genome sequence
- Calanus finmarchicus transcriptome shotgun assembly (mRNA)
- clcAll50 – Lepeophtheirus salmonis intermediate assembly scaffolds (depricated)
- L. salmonis masked ESTs only
- Swissprot
- NR (non-redundant protein database)
- NT (dna database)




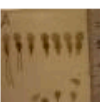
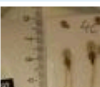
% Identity :

Expect threshold (E threshold) :

Others options :

-options=value

Unique RNA-interference LIMS

Title▲	Target	Dev Stage	Phenotype	# individuals (start/end)	Image	Organism	Posted	Efficacy of Knock Down
 RNAi E, F34 (LsRaptor)	Raptor	PreAdult2 Adult	egg string deformation, female reproduction, no eggstrings	39 / 24	2c_F34	Lepeophtheirus salmonis	12.03.2013 - 10:05 Christiane.Troesse	
 RNAi E, F42 (g13677_Rheb) del_me		PreAdult2 Adult	feeding, female reproduction, no blood uptake, no eggstrings, reduced nutrient uptake Low reproduction and most are missing blood in the gut	39 / 25		Lepeophtheirus salmonis	12.03.2013 - 11:38 Heidi	
 RNAi E, F50 (Ferr_sub)	ferr_sub	PreAdult2 Adult	female reproduction, no eggstrings Most lice didn't have egg strings at termination (but had first and second which had fallen off and were found in the filter). No reproduction.	38 / 18	7a	Lepeophtheirus salmonis	21.02.2013 - 12:54 Christiane.Troesse	F>85.00%
 RNAi E, F51 (Ferr_LH)	ferr_LH	PreAdult2 Adult	female reproduction, no eggstrings Most lice didn't have egg strings at termination (but had first and second which had fallen off and were found in the filter). No reproduction.	39 / 23	8a	Lepeophtheirus salmonis	28.06.2013 - 11:43 Christiane.Troesse	F>78.00%
RNAiE-IR25a/IR8a	20261	PreAdult2 Adult	no visible phenotype no visible phenotype	39 / 17		Lepeophtheirus salmonis	17.03.2013 - 11:52 Anna	F>39.00%
RNAiE-IR25a/IR8a	IR25a	PreAdult2 Adult	no visible phenotype no visible phenotype	39 / 17		Lepeophtheirus salmonis	17.03.2013 - 11:52 Anna	F>39.00%
 RNAi H_FK506		PreAdult2 Adult female	female reproduction, no eggstrings no reproduction	30	f60	Lepeophtheirus salmonis	09.04.2013 - 13:26 Sussie	
 RNAiH_g9531		PreAdult2 Adult	no visible phenotype		4c	Lepeophtheirus salmonis	09.04.2013 - 13:37 Sussie	

Pathway reconstruction

KEGG Analysis Reports

Select a KEGG report to view:

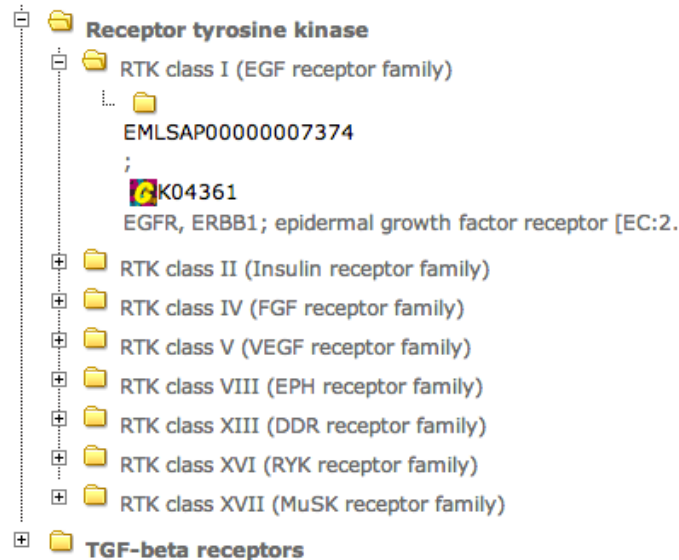
KAAS/KEGG

Any analysis with KEGG results related to this organism are available for viewing. For further information, see the analysis information page.

KEGG BRITE Hierarchy

- Cytoskeleton proteins
- Cell Adhesion Molecules and Their Ligands
- Exosome
- SNAREs
- Ubiquitin System
- CD Molecules
- Cytokine Receptors
- Ion Channels
- GTP-Binding Proteins
- G Protein-Coupled Receptors
- DNA Repair and Recombination Proteins
- Nuclear Receptors
- Chaperones and folding catalysts
- Proteasome
- Spliceosome
- Chromosome and associated proteins
- DNA Replication Proteins
- Mitochondrial biogenesis
- Transcription Machinery
- Transfer RNA Biogenesis
- Translation Factors
- Ribosome
- Ribosome Biogenesis

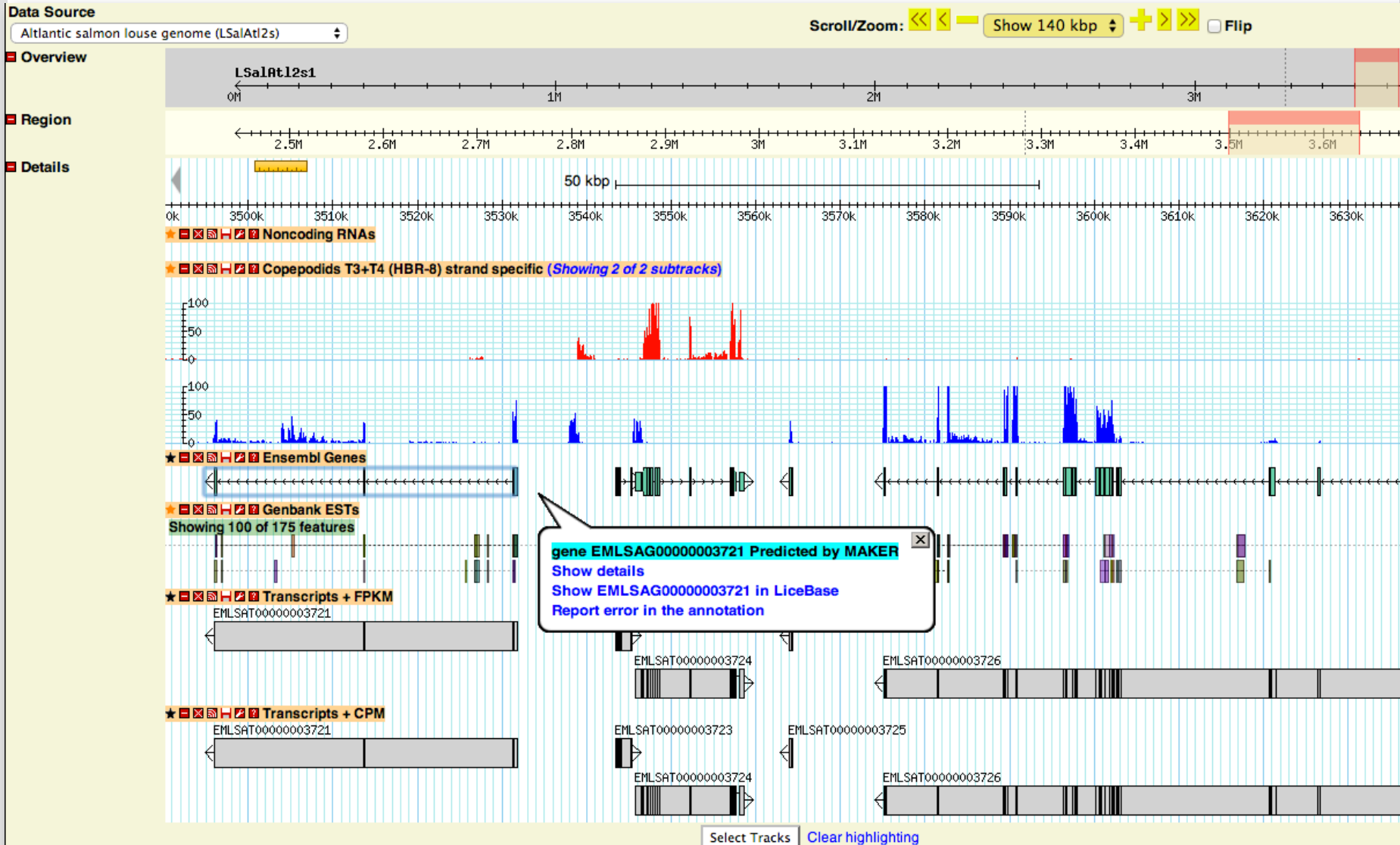
Hierarchy: Cytokine Receptors



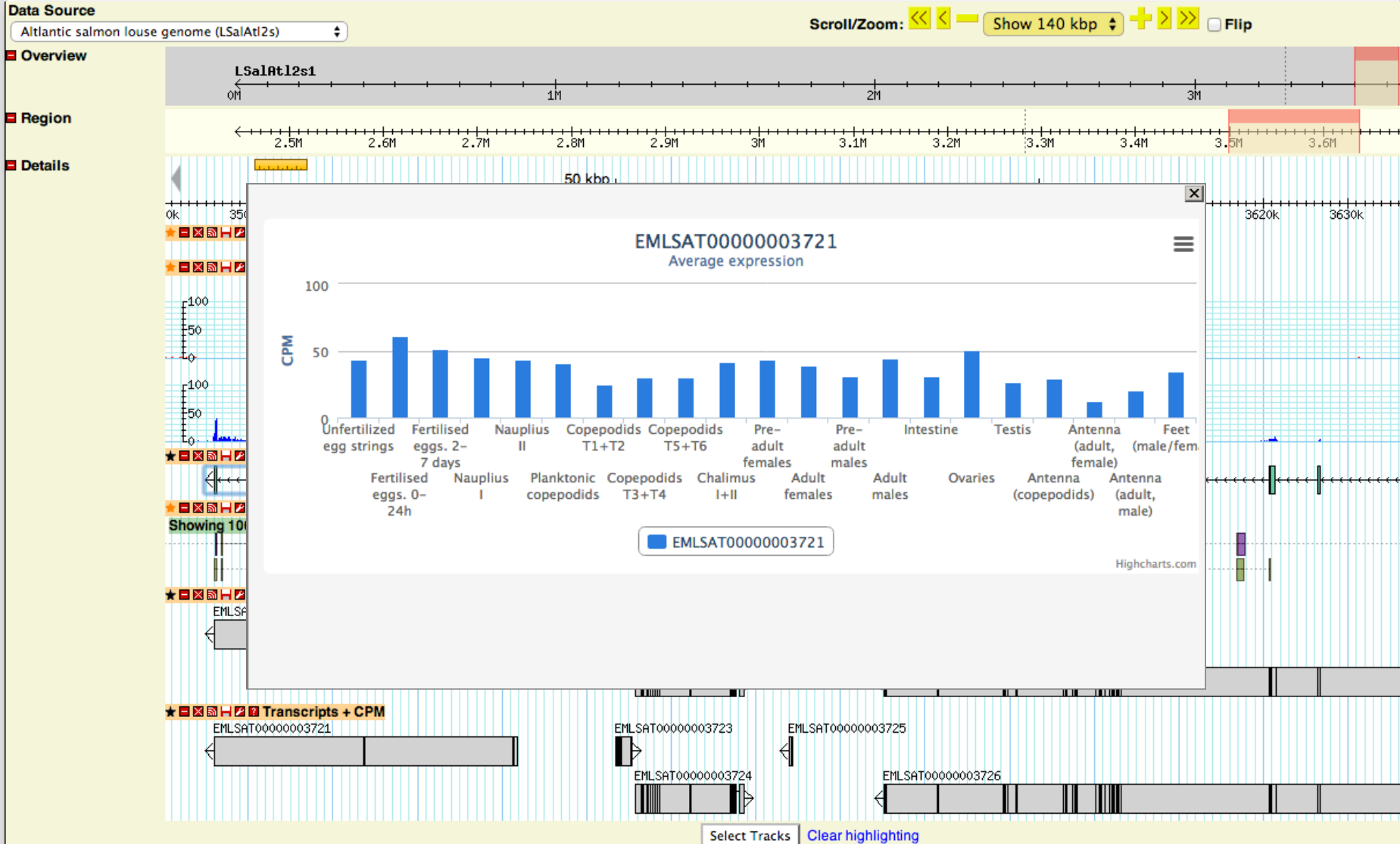
Resources

- Organism Details
- KEGG Analysis Reports
- Data Type Summary
- Feature Browser
- GO Analysis Reports

GBrowse



GBrowse (with RNA-seq)



GO Browser



LiceBase

Genome tools RNAI

Search this site:

>>Logout

LiceBase

- ▷ In-situ images
- ▷ RNAi
- Search Blast Results
- Forum

lb_admin

- ▽ Search Biological Data
 - Features
 - Organisms
 - Analyses
- Analysis
- Analysis: Blast
- Analysis: Interpro
- Analysis: KEGG
- Analysis: Unigene
- ▷ Blogs
- Browse in-situ images
- Feature

Lepeophtheirus salmonis

GO Analysis Reports

Select a GO report to view:

InterproScan 5

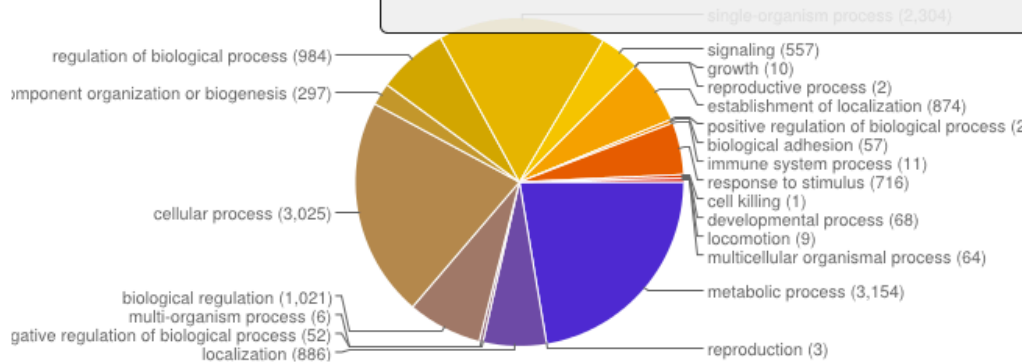
Any analysis with GO results related to this organism are available for viewing. For more information see the analysis information page.

Biological Process

Expand the tree to browse term counts

biological_process

Number of Biological



Term Information

Close [X]

View Edit Outline Track

Term biological_process

Accession 0008150

Ontology biological_process

Definition Any process specifically pertinent to the functioning of integrated living units: cells, tissues, organs, and organisms. A process is a collection of molecular events with a defined beginning and end. [GOC:go_curators, GOC:isa_complete]

narrow: biological process unknown

Synonyms exact: physiological process

exact: biological process

Internal ID 5651

Download sequences GO_0008150.fasta

Terms From InterproScan 5 Analysis

single-organism process (2,304)

Resources

- KEGG Analysis Reports
- Data Type Summary
- Feature Browser
- GO Analysis Reports

GO term overrepresentation analysis in sea lice

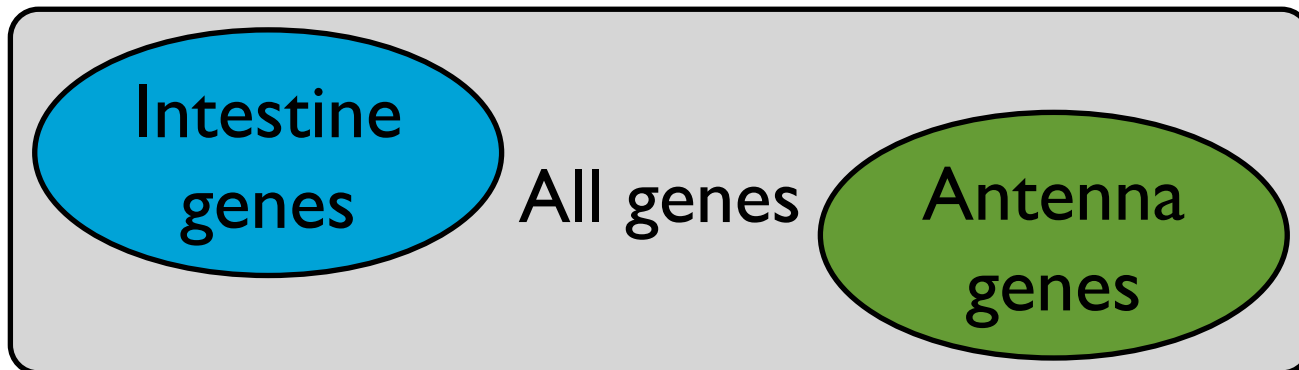
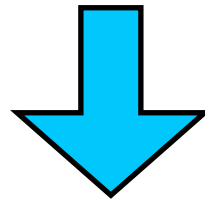
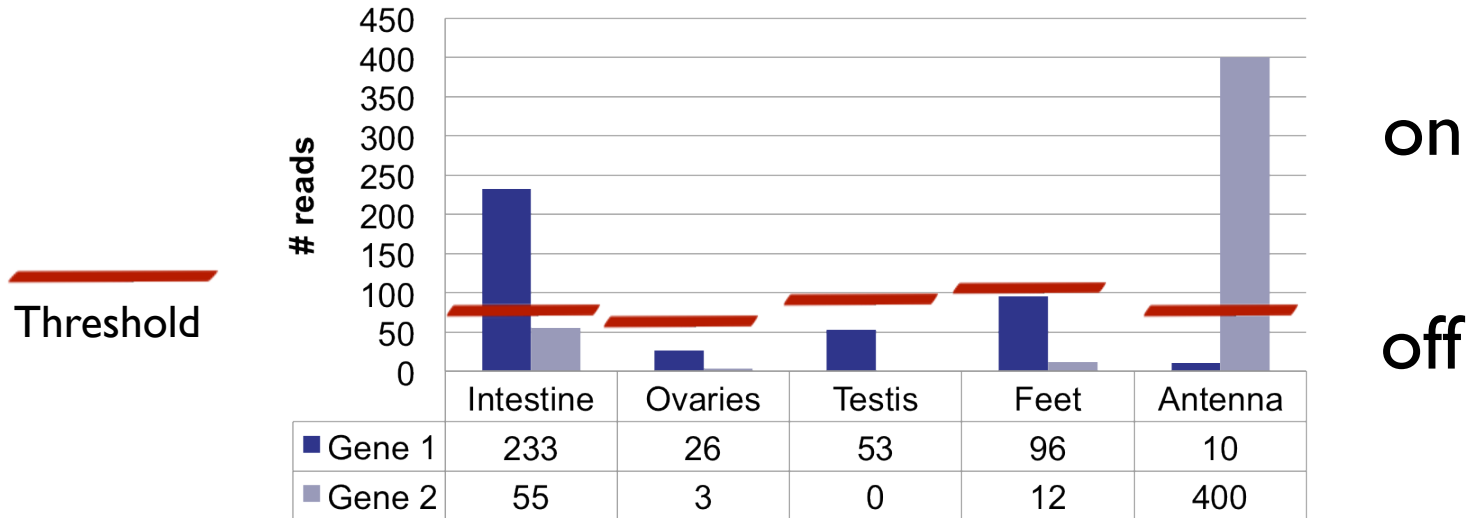


- ▶ GO: Ontologies that describe gene products in terms of their associated biological processes, cellular components and molecular functions in a species-independent manner.
- ▶ For all *L. salmonis* genes, GO terms are assigned to Ensembl **predicted** proteins (EMLSAP) **automatically** by InterProScan 5 pipeline.
- ▶ Gene-sets are made from RNA-seq data: Gene either **on** or **off** in a certain condition based on sample specific threshold.

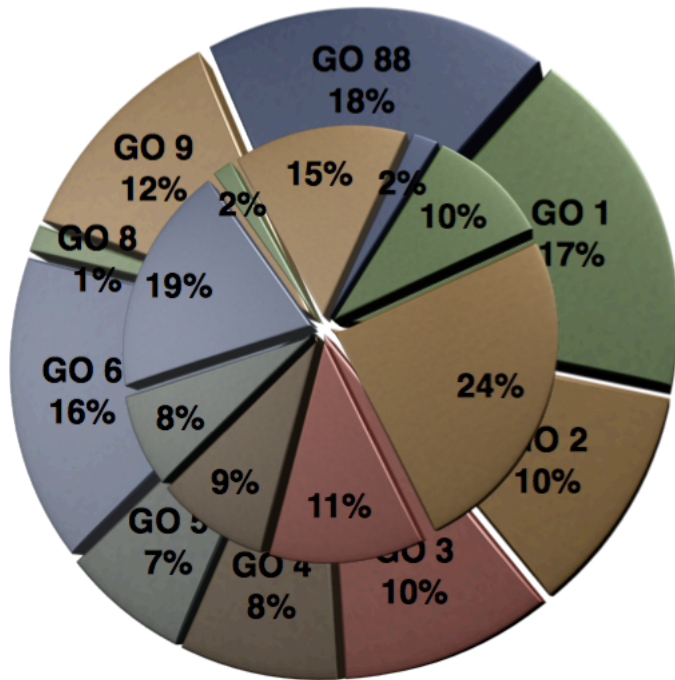
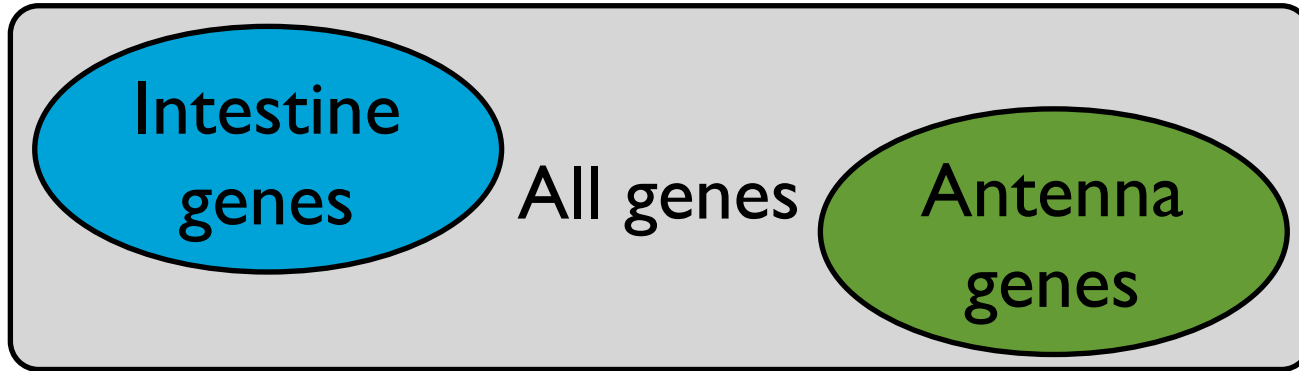
How good is the automatic annotation?

GO term overrepresentation

RNA-seq tissue samples



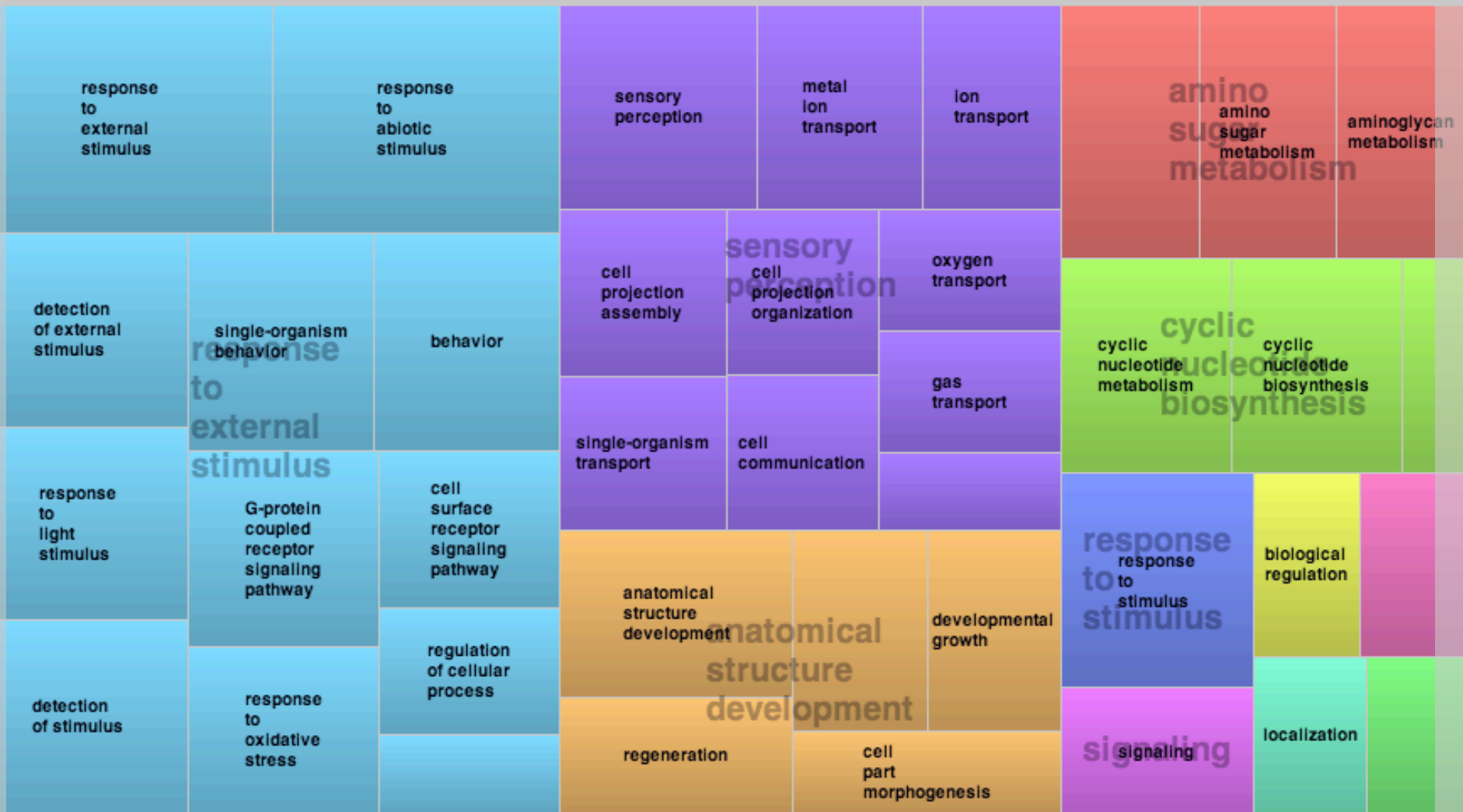
GO term overrepresentation



GO Term	p-value
GO 2	0.01
GO 6	0.04
GO 9	0.1
...	...
...	...

Results in REVIGO

Antenna (copepod)



Intestine



Outlook



- Integrate pipelines for user data analysis
- Upload and annotate more, related (copepod) genomes for comparative genomics
- Integrate LiceBase + SalmonBase*
- Integrate WebApollo, a tool for community annotation
- Integrate BioMart for data-mining
- Open LiceBase for the general public at the same time with publication of the genome

Conclusions



- The key for effective sea lice control is possibly in its genome
- LiceBase - the premier resource for sea lice genomics
- Bioinformatics analyses and model organism databases are aiding our understanding of host-parasite interactions of sea lice
- Please visit <https://licebase.org>
- Please contact admin@licebase.org

Acknowledgements



CBU Service Group:

 David Fredman

 Kjell Petersen

 Kidane M. Tekle

IT admin:

 Stanislav Oltu

Institute of Marine Research:

 Kjetil Malde

 Tomasz Furmanek

 Rasmus Skern-Mauritzen



- Damien Geneste, Univ. Poitiers, France
- Timothy Parnell, @HCI, Univ. Utah
- InterProScan help desk @EBI
- Alexander Dobin, STAR aligner
- Stephen Ficklin, developer of Tripal
- Scott Cain, GBrowse
- Gaëtan Droc, @Banana Genome hub
- Michael Nuhn @EBI, Ensembl annotation



Norges miljø- og
biovitenskapelige
universitet



Department of Informatics, UiB

New insights: pancrustacean evolution

