

Overview of Bioinformatics Expertise



Hakim Tafer

University of Natural Resources and Life
Sciences
Extremophile Center
Vienna, Austria

January 24, 2019

Outline

Non-coding RNA

Annotation

Function

Extremophile Center

Genome Assembly

Gene Annotation

Functional Annotation

Comparative Genomics

Differential expression

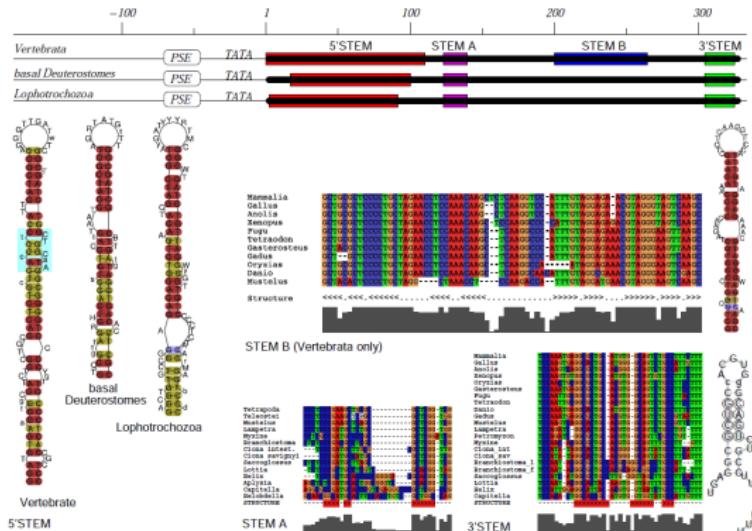
Chimeric RNA

Population Genomics: CSI Vienna

ncRNA Families

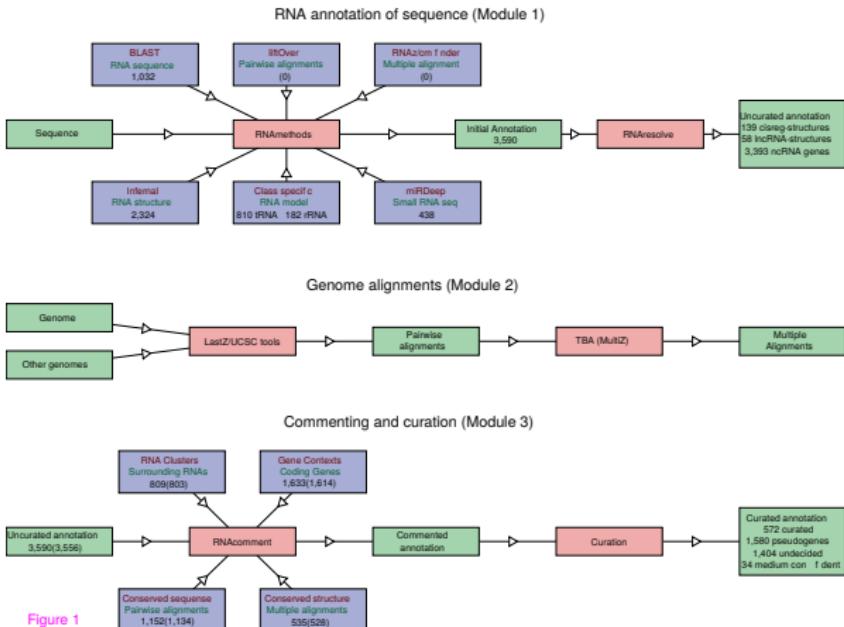
- ncRNAs play a key role in the cell metabolism
- miRNA, piRNA, 7SK: gene expression regulation
- snRNA: mRNA splicing
- Y-RNA: DNA replication
- snoRNA: rRNA modification, gene expression

Specific Annotation 7SK



- Sequence conservation
- Structure conservation
- Promoter sequence

Genome-wide annotation



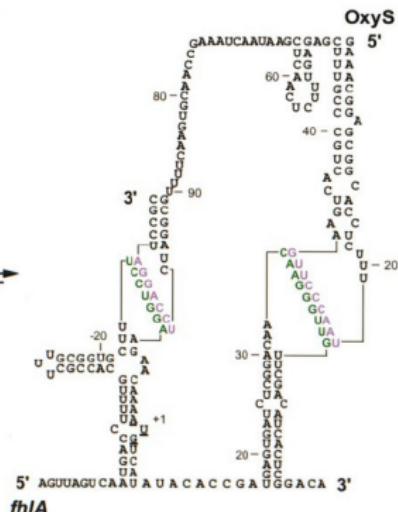
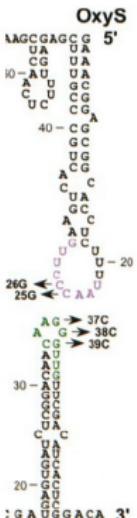
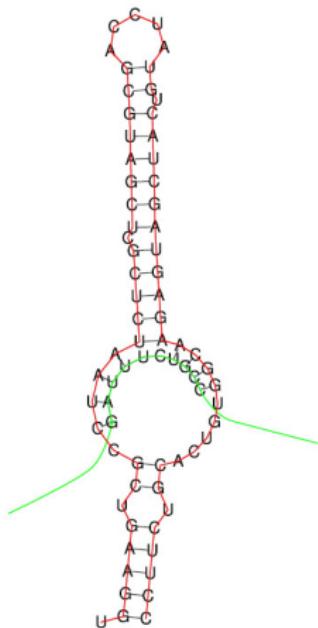
- Pig, Cow, Coelacanth, Duck
- Trichoplax
- Beet
- Fungi
- Bacteria

Figure 1

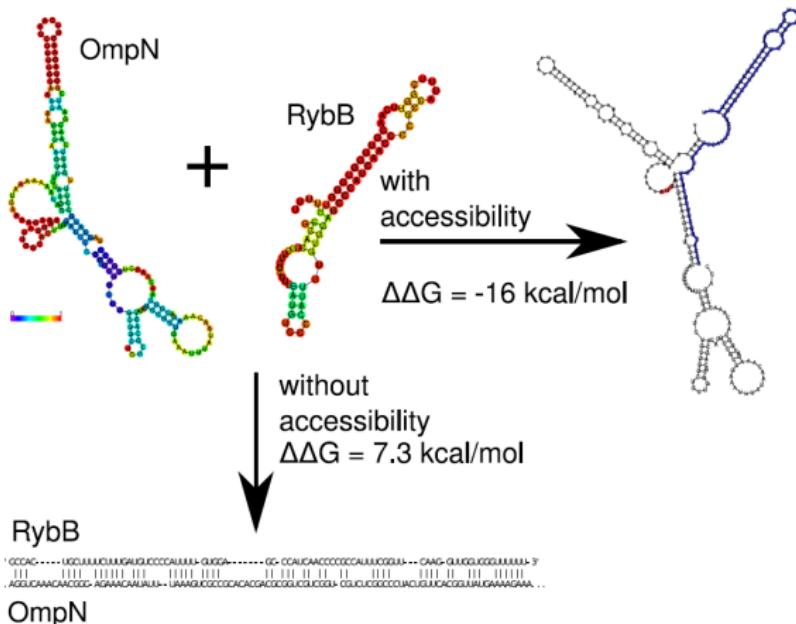
Function

- Function of ncRNAs is defined by their binding partners
- miRNAs modulate Gene expression by binding to mRNAs
- snoRNAs recognize their targets through basepairing
- snRNAs regulate splicing by binding intronic and exonic region
- sRNAs in bacteria modulate gene expression by modulation RBS accessibility

RNA-RNA interaction



RNA-RNA computation



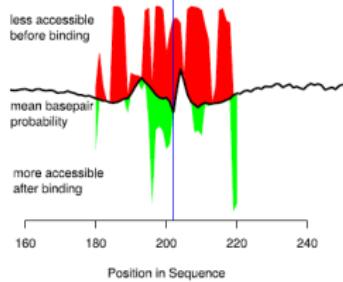
Applications: small RNAs

mRNA	sRNA	regulation	$\Delta\Delta G$	Position	Pos.lit.	cite
RyhB	sodB	-	-11.50	-18,+4	-4,+5	[78]
DsrA	hns	-	-14.60	-10,+11	+7,+19	[138]
MicA	ompA	-	-13.60	-21,-6	-21,-6	[201]
MicC	ompC	-	-15.80	-30,-15	-30,-15	[37]
MicF	ompF	-	-17.80	-11,+9	-11,+10	[37]
Spot42	galK	-	-17.00	-18,+30	-19,+21	[178]
SgrS	ptsG	-	-17.33	-28,-10	-28,+4	[115]
GcvB	dppA	-	-17.30	-30,-7	-31,-14	[227]
DsrA	rpoS	+	-14.52	-126,-97	-119,-97	[157]
RprA	rpoS	+	-15.90	-134,-94	-117,-94	[157]

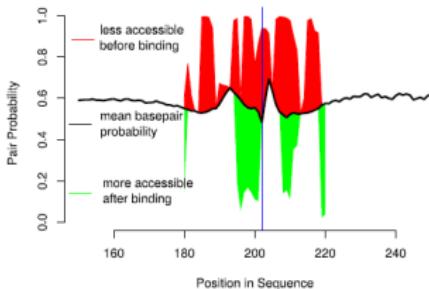
- Predicted and reported target site
- Developed RNAplex $O(n)$ instead of $O(n^3)$
- Whole genome search possible (RNAPredator)

Applications: small RNAs

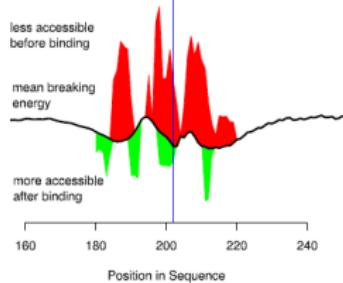
air Probability vs Sequence Position for RprA



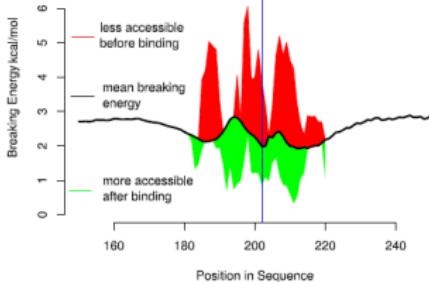
Pair Probability vs Sequence Position for DsrA



eaing Energy vs Sequence Position for RprA

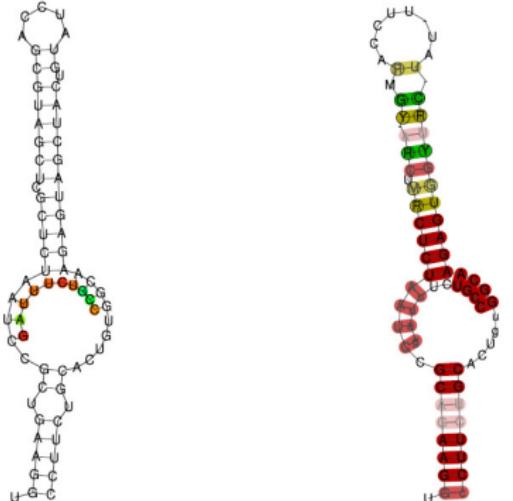


Breaking Energy vs Sequence Position for DsrA



- RBS is controlled remotely
- RBS accessible after binding of sRNA

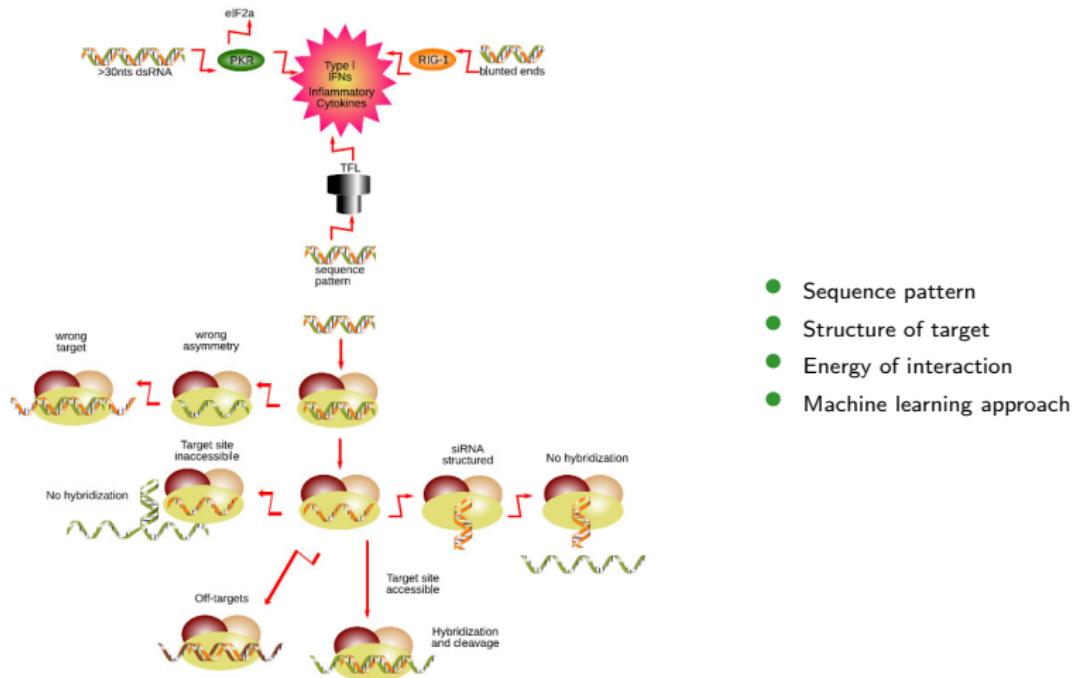
Applications: snoRNAs



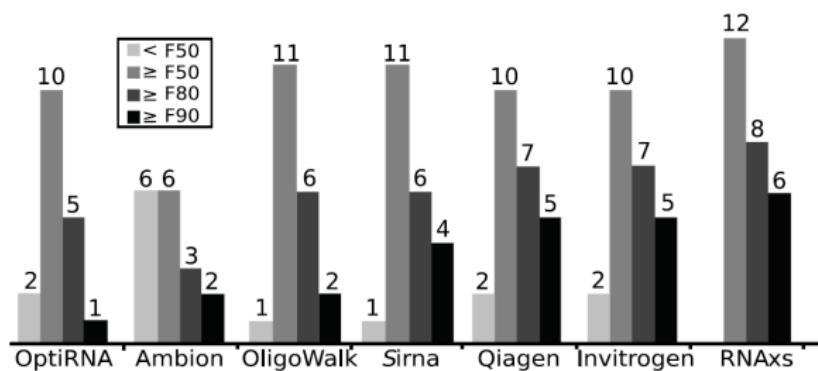
D. Melanogaster
D. Erecta
D. Persimilis
D. Sechellia
D. Simulans
D. Yakuba
D. Willistoni

- RNA-RNA interaction energy
 - Accessibility of the target
 - Stability of the stem
 - Stem structure
 - Conservation of the interaction
 - $O(n)$ instead of $O(n^4)$

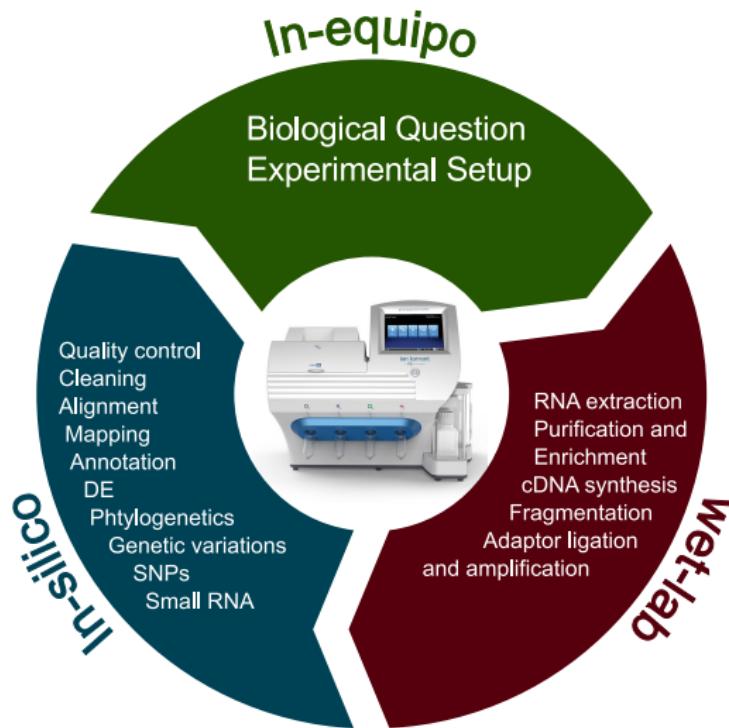
Applications: silencing RNAs



Applications: silencing RNAs



Extremophile Center



Genome Assembly



PAM

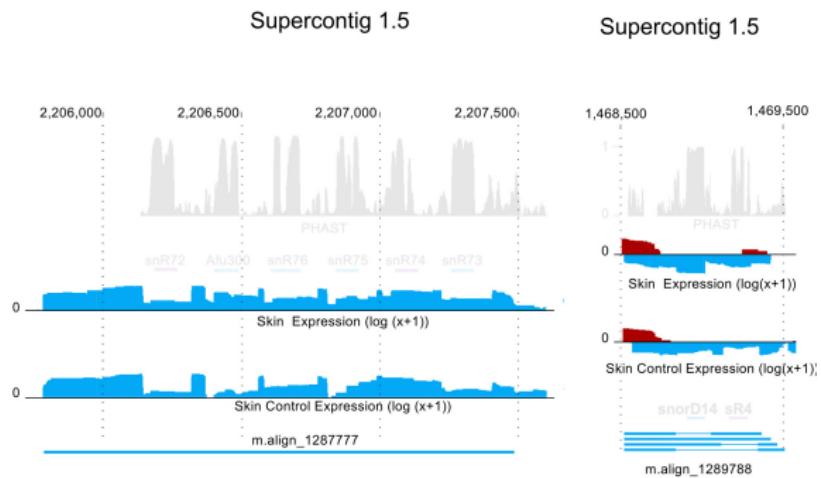
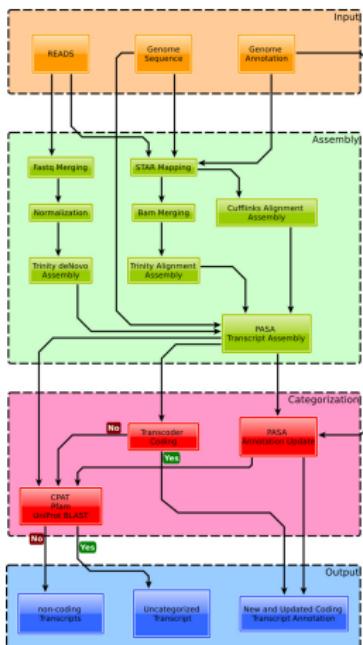
PKS1:N G K P K D V T H A K C V I
4742 - AATGgaaaGAAAGATGTCACTCATGCCAAGTGCCTGATCG - 4801

WT genome AATGgaaaGAAAGATGTCACTCATGCCAAGTGCCTGATCG

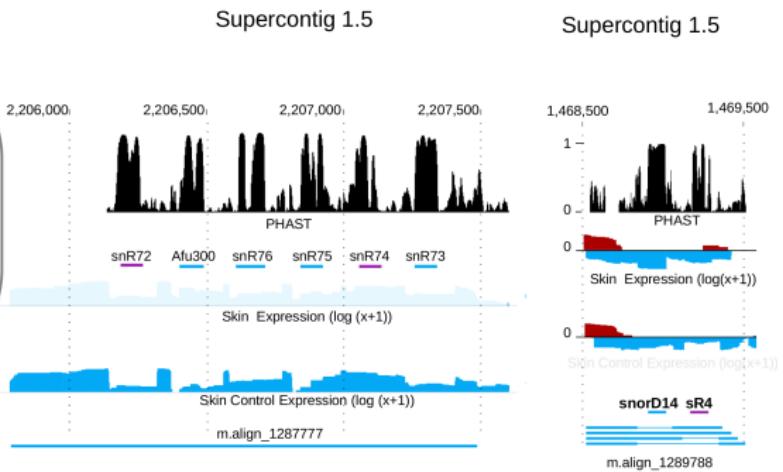
MT genome AATG----GAAAGATGTCACTCATGCCAAGTGCCTGATCG

MT PKS1:N G K M S L M P S A *

Gene Annotation

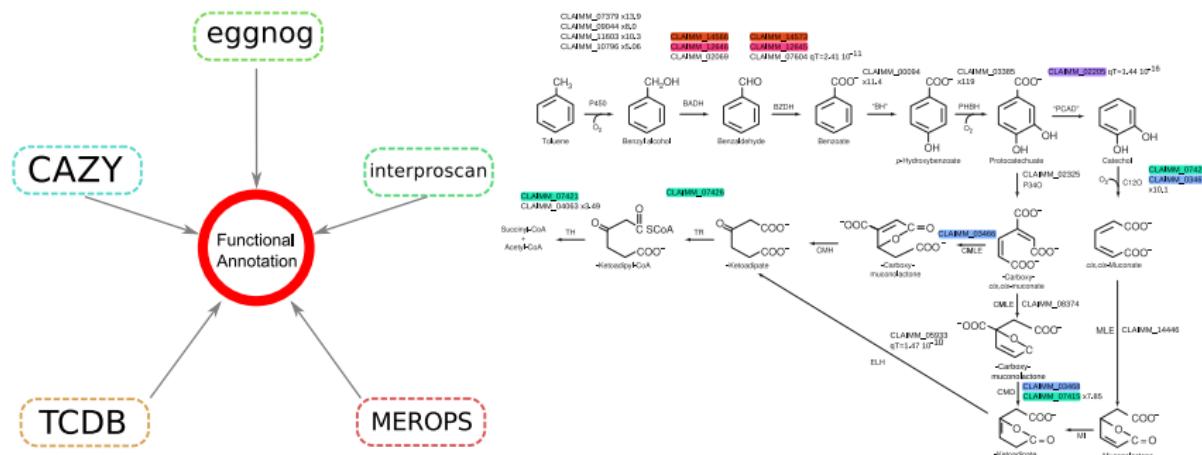


Gene Annotation

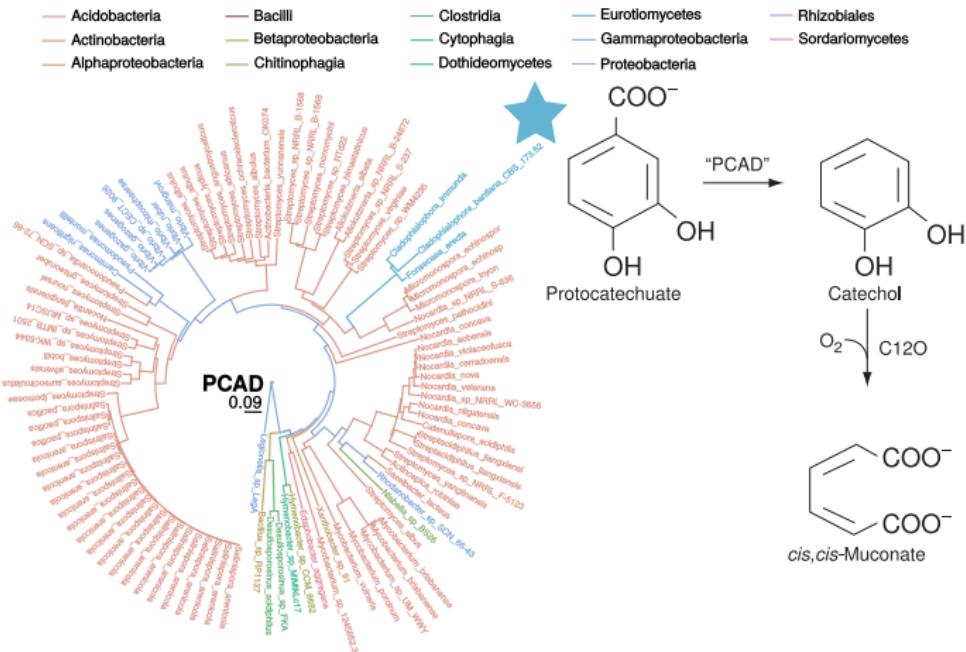


Functional Annotation

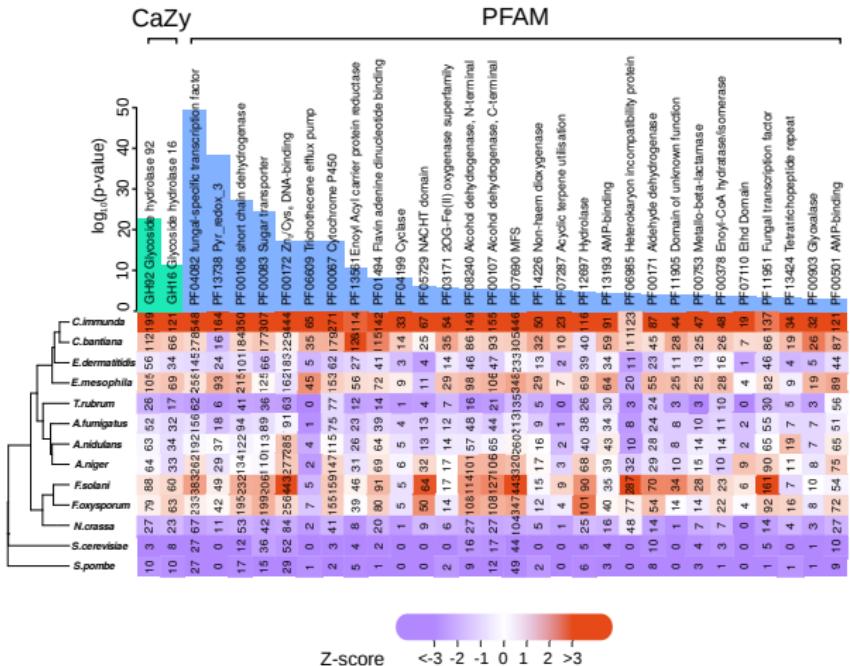
Cladophialophora Immunda



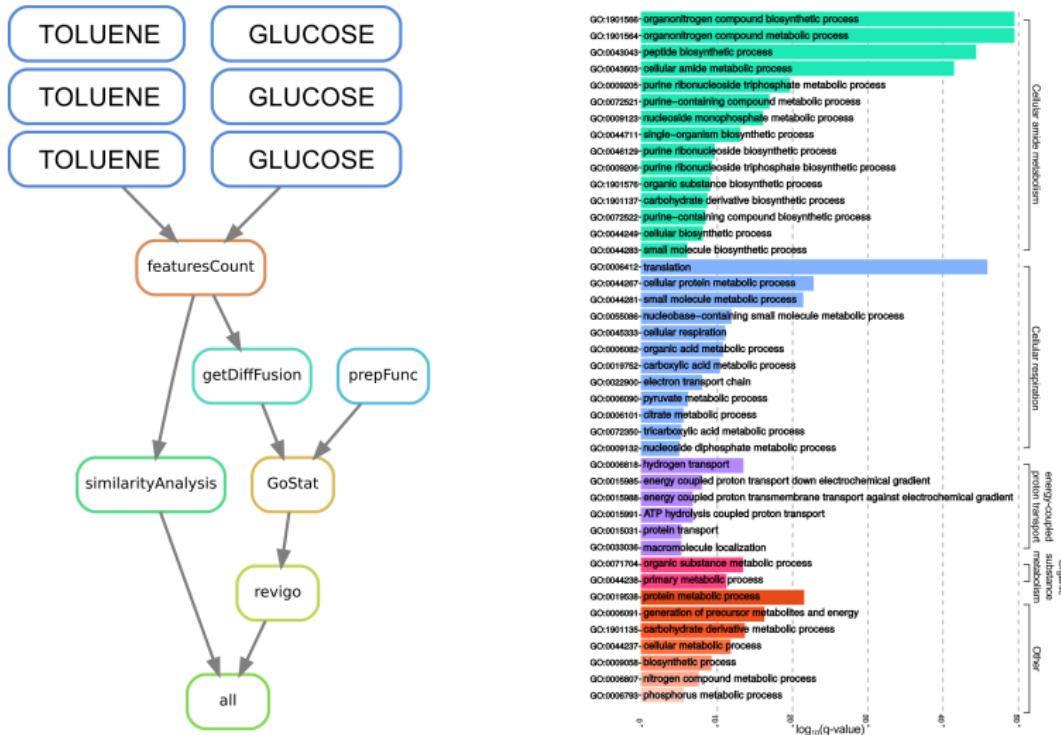
Comparative Genomics



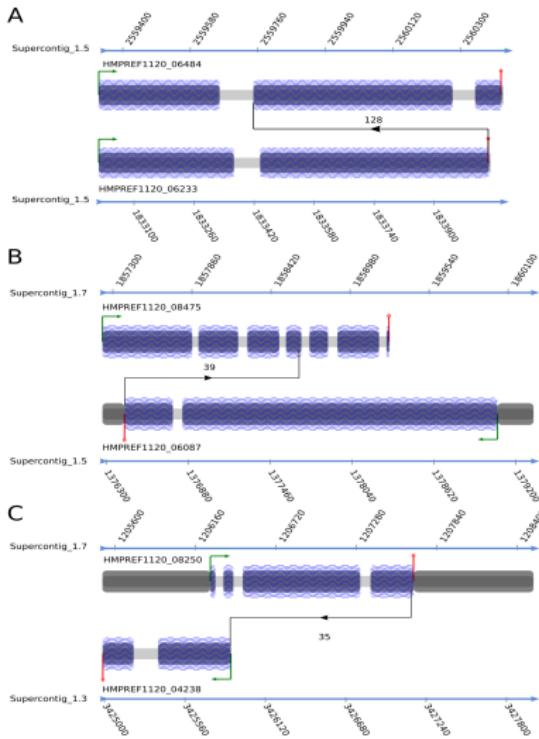
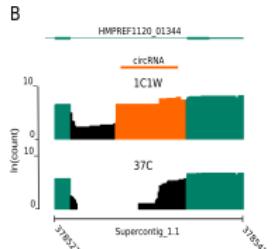
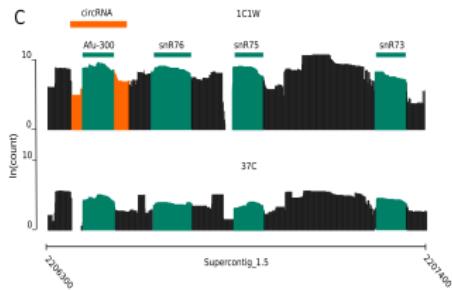
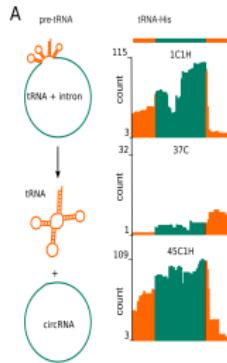
Comparative Genomics



Differential expression



Chimeric RNA



FunBench

localhost:8080/?tool_id=toolshed.g2.bx.psu.edu%2frepos%2ferasmus-medical-center%2fdr_disco%2fdr_disco_fix%2f0.14. 110% [...
Most Visited](#) [Getting Started](#) [Log in to Impulsedyn...](#) [Wetrepublique Categ...](#) [AllCompared Global...](#) [Save to Mendeley](#) [Installing MySQL wit...](#) [Save to Mendeley](#) [bead beating - Googl...](#)

Analyze Data Workflow Shared Data Visualization Help Login or Register

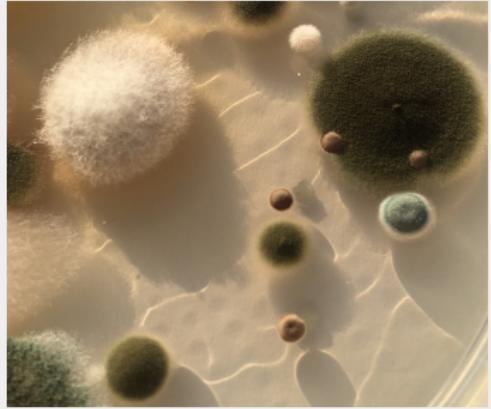
Galaxy /

Tools

- [break-points in RNA-seq](#)
- [NASTseq Identify cis-NASTs using ssRNA-seq](#)
- [PPmIR PIPELINE a method to identify novel plant miRNA](#)
- [ElaMapper detects small ncRNA derived fragments in small RNA-Seq data](#)
- [Filter with SortMeRNA of ribosomal RNAs in metatranscriptomic data](#)
- [STAR-fusion detect fusion genes in RNA-Seq data](#)
- [RNA STAR Gapped-read mapper for RNA-seq data](#)
- [featureCounts Measure gene expression in RNA-Seq experiments from SAM or BAM files.](#)
- [Salmon Transcript Quantification from RNA-seq data](#)
- [Sealfish transcript quantification from RNA-seq data](#)
- [Trinity de novo assembly of RNA-Seq data](#)
- [Hisat2 A fast and sensitive alignment program](#)
- [htseq-count - Count aligned reads in a BAM file that overlap features in a GFF file](#)

Your FunBench Galaxy instance is ready!

Configuring Galaxy » [Installing Tools »](#) [Guided Tour »](#)



History

search data
Unnamed his
(empty)

This hist
load your
from an r

23 of 31 [Recent Determinants differentially expressed](#)

Case-Study

- Custom officers found three possibly smuggled marble statues of an unknown origin, two of them representing human torsos, and one representing a small young girl head, at a flea market in Vienna. The seller said he had found them under the rubble of an old building.
- The statues were taken to the Museum of Art History in Vienna and analyzed by experts of the museum
- Sampling and molecular analyses were performed by our team (BOKU) in order to answer some intriguing questions
- Is it possible to reconstruct the history of the storage of each single object?
- Were the statues stored individually or together?
- Is it possible to elucidate the geographical shaft of some of the objects?

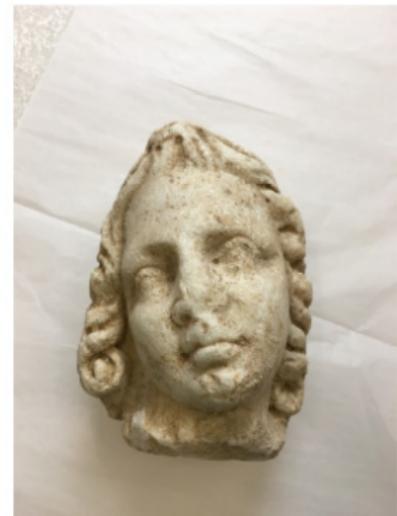
Statues



S1

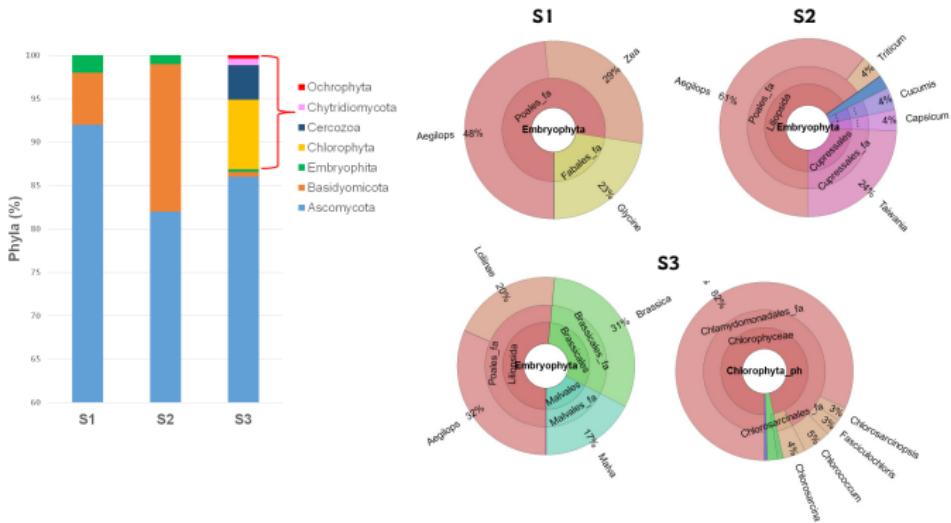


S2



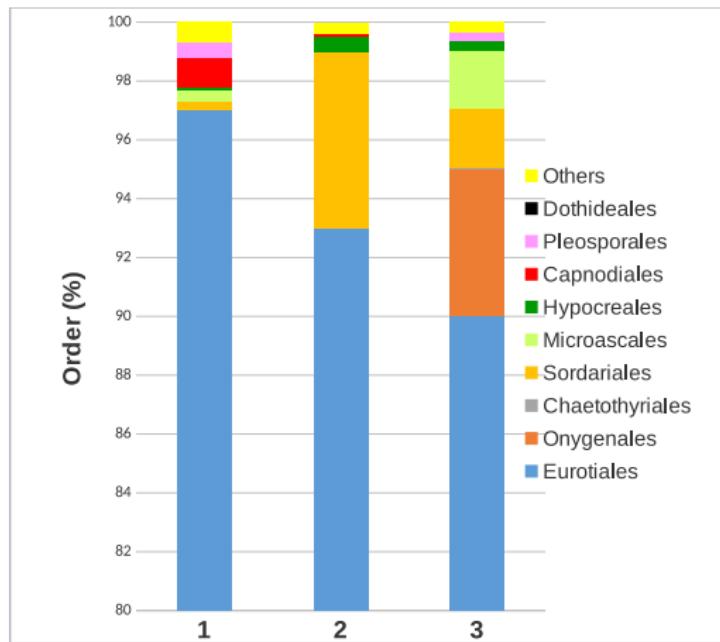
S3

Eukaryote Community ITS



- S1 and S2: Biological markers for land plants, stored (or buried) probably in agricultural soils!, S2:Taiwania is geolocated in south-eastern asia
- S3 Green algae of the phylum Chlorophyta

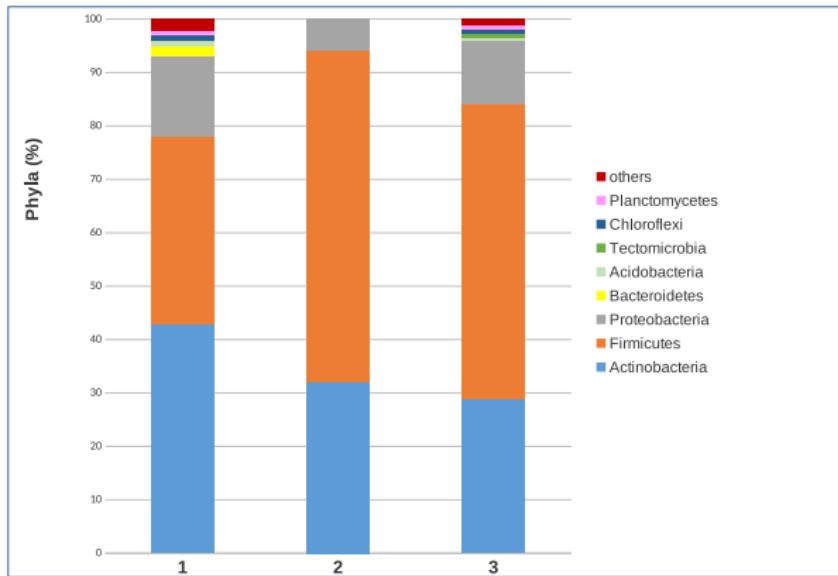
Fungal Community ITS



● Hypersaline environment: S1, S3

● Plant-associated/Phytopathogen
S1,S2,S3

Bacterial Community 16S



- Saline/Aquatic environment:
S1, S3

- Agricultural soil S1,S2,S3

- Soil from Animal's farm
S1,S2

Conclusion

- No strong indication that all three statues were together under house dust
- S1,S2 were probably stored together in the recent past
- S3 in salty water for a large amount of time
- S2 was in contact with *Taiwania* trees

ERC Grant

- Goal: Understand Evolution of Energy harnessing from available environmental sources
- Method: Comparative Phylogenetic Analysis of genes involved in Physiology, geochemical records of available environmental energy sources
- Check for monophyly of genes involved in pathway of interest (Sousa, Weiss 2016)
- Compare phylogenetic tree with Geochemical records
- ProteinOrtho/iqtree instead of OrthoMCL/RAxML
- Anchor time points can be derived from Marin et al. 2016 (SSU-derived tree). Could be putatively improved by taking rRNA structure into account (RNAsalsa Stocsits et al. 2009)

Acknowledgments

- Prof. Katja Sterflinger
- Caroline Poyntner
- Barbara Blasi
- Guadalupe Pinar
- Ksenija Lopandic
- Prof. Renee Schroeder
- Prof. Peter Stadler
- Prof. Ivo Hofacker
- Prof. Jan Gorodkin