Meta-pipe analysis pipeline

Espen Mikal Robertsen, UiT – The Arctic University of Norway
Tromsø, 2018

Outline

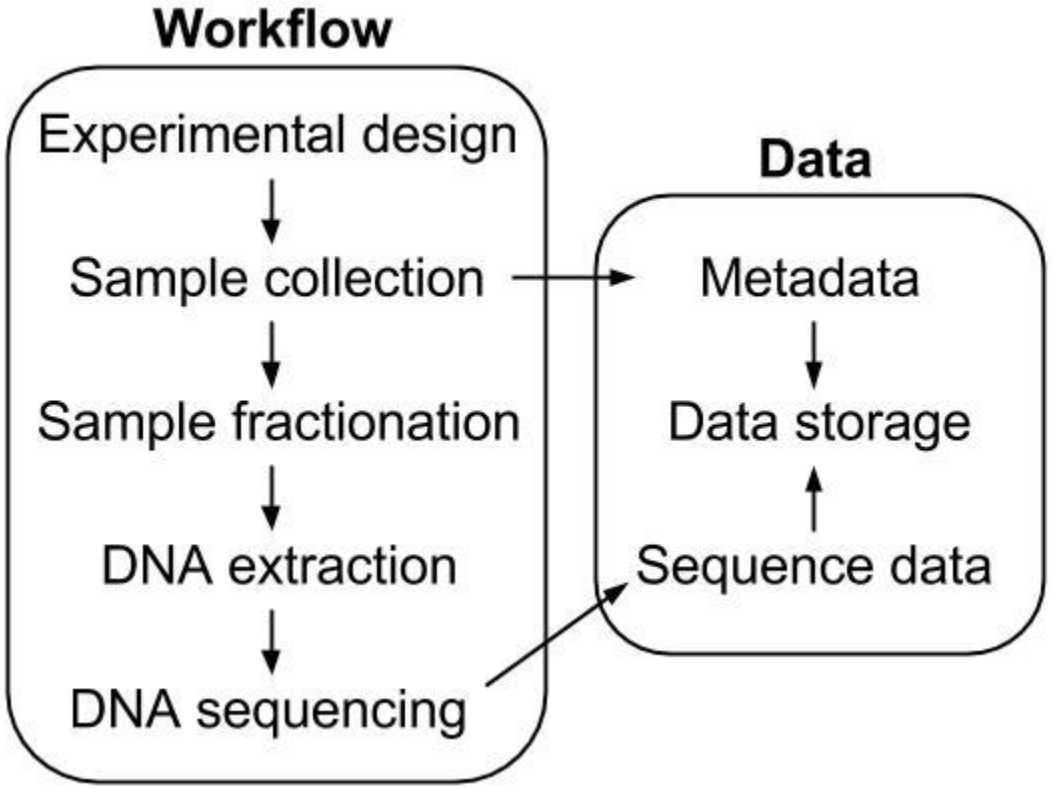
- Motivation / Why Meta-pipe
- Steps prior to bioinformatic analysis
- Overview of Meta-pipe
- Interfaces

Motivation behind META-pipe

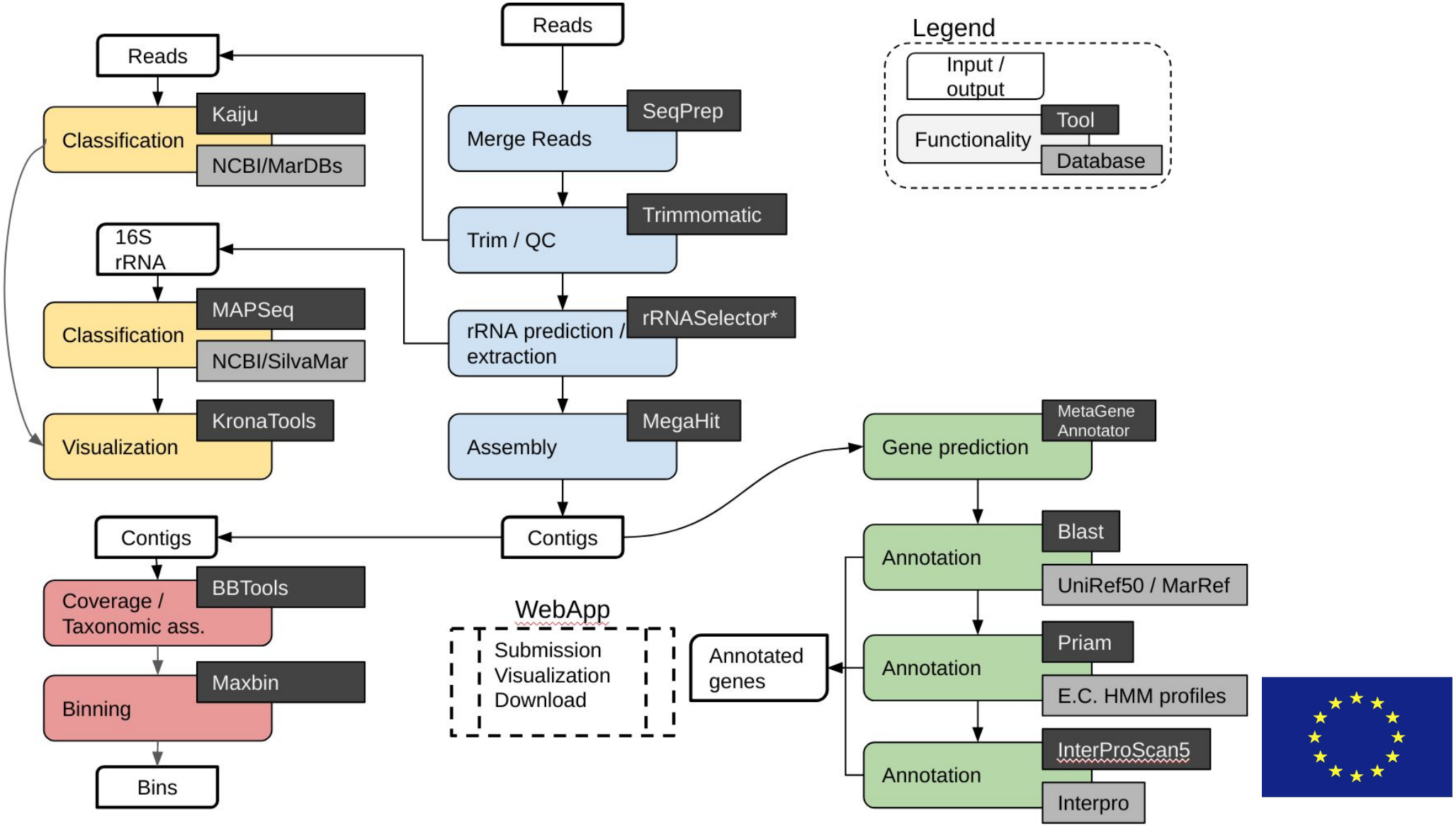
- Lack of metagenomics resources for the marine domain
- "Lack" of metagenomics resources for analysis of full-length genes
- User simplicity
- Resource availability

Mineta, Katsuhiko, Gojobori, Takashi 2016 Gene, ISSN: 1879-0038, Vol: 576, Issue: 2 Pt 1, Page: 724-8

Steps prior to bioinformatic analysis



Biological Overview of Meta-pipe

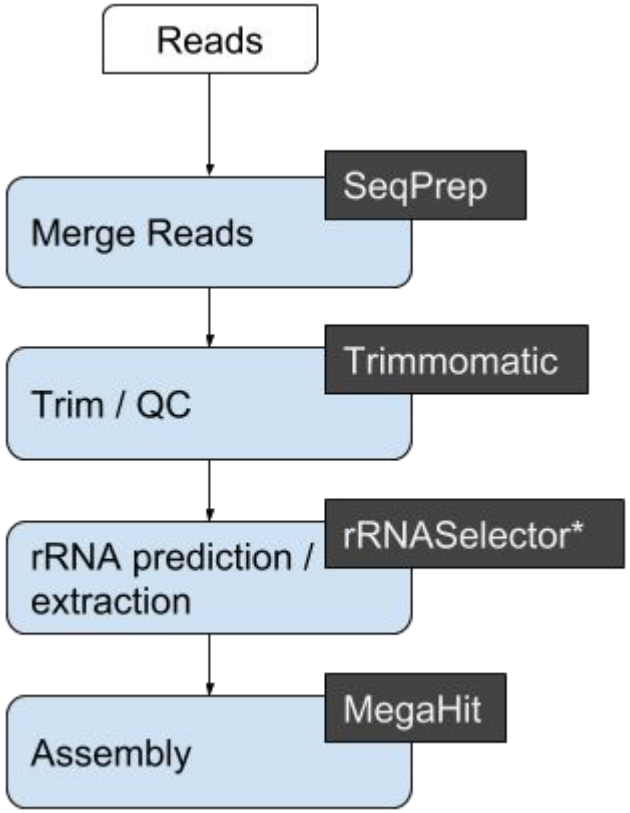


Tromsø, May,

Biological Overview

Preprocessing

- SeqPrep – Merging of Illumina reads
- Trimmomatic – QC of reads
- (DeconSeq) – Filter unwanted sequences with reference
- rRNASelector – Removal/Extraction of rRNA sequences
- MEGAHIT – Assembly

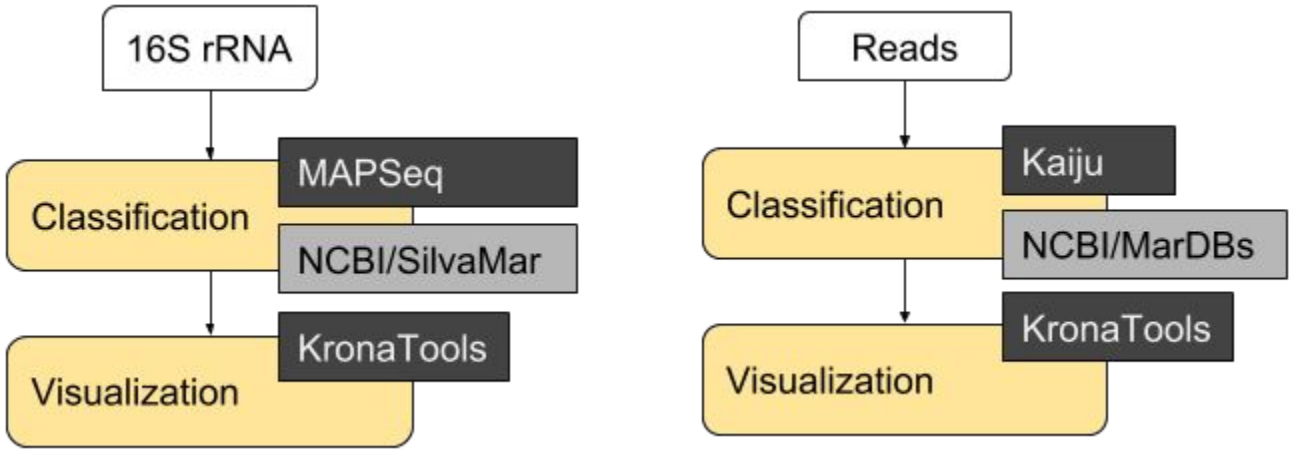


Biological Overview

Taxonomic Classification

- Kaiju - Classification of reads using protein reference databases
- Mapseq - High throughput rRNA sequence classification

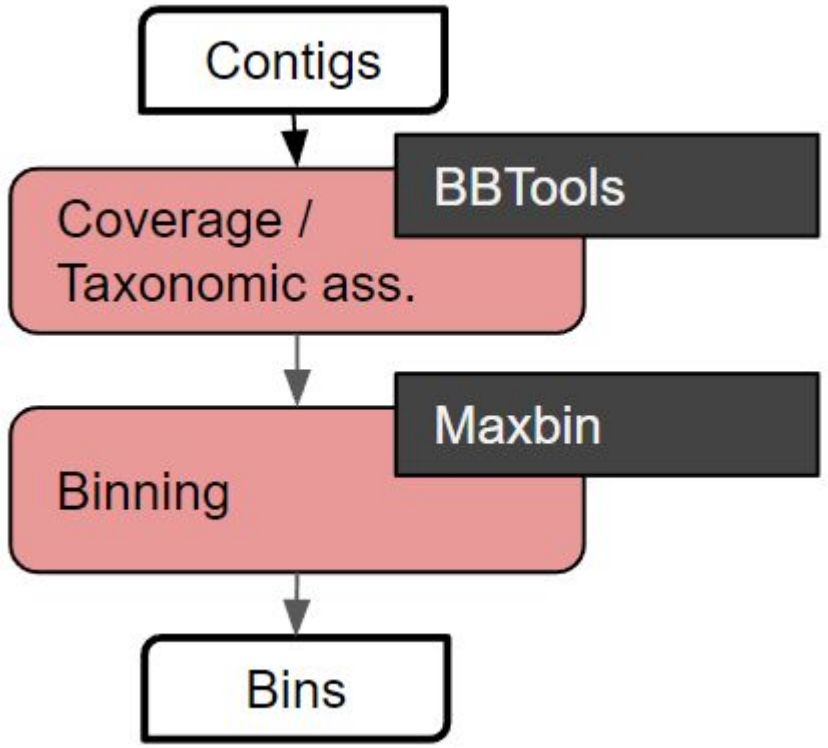
- KronaTools – Visualization of output



Biological Overview

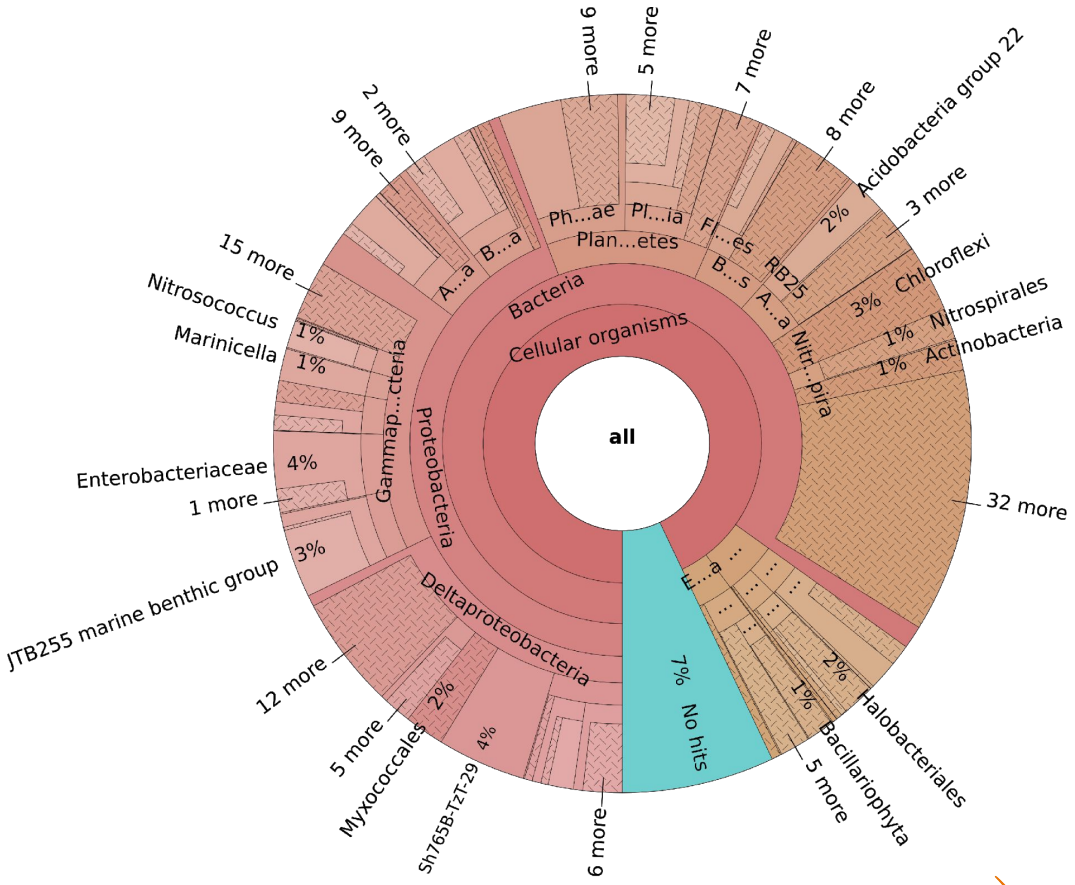
Binning

- BBMap - Coverage calculation
- MaxBin - Binning of contigs
- BBSketch - Taxonomic assignment of bins



Biological Overview

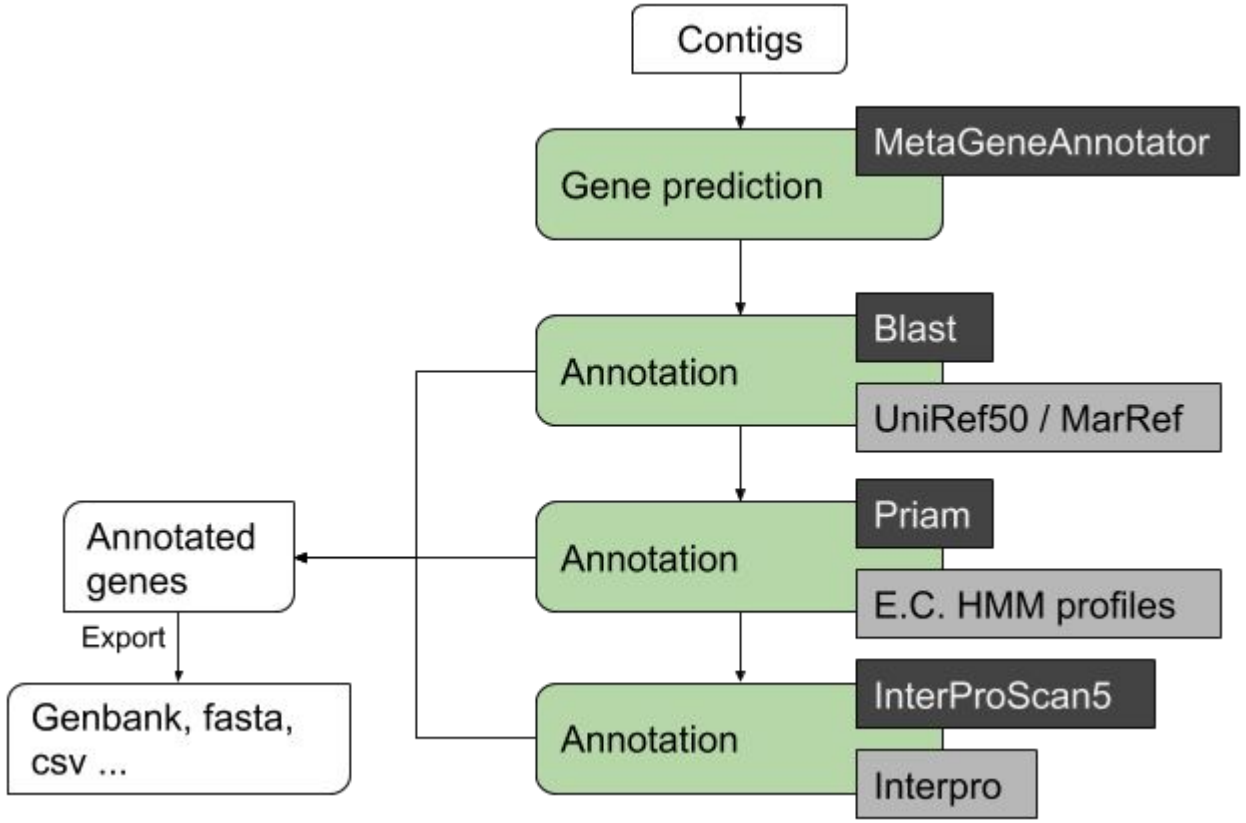
Taxonomic Classification



Biological Overview

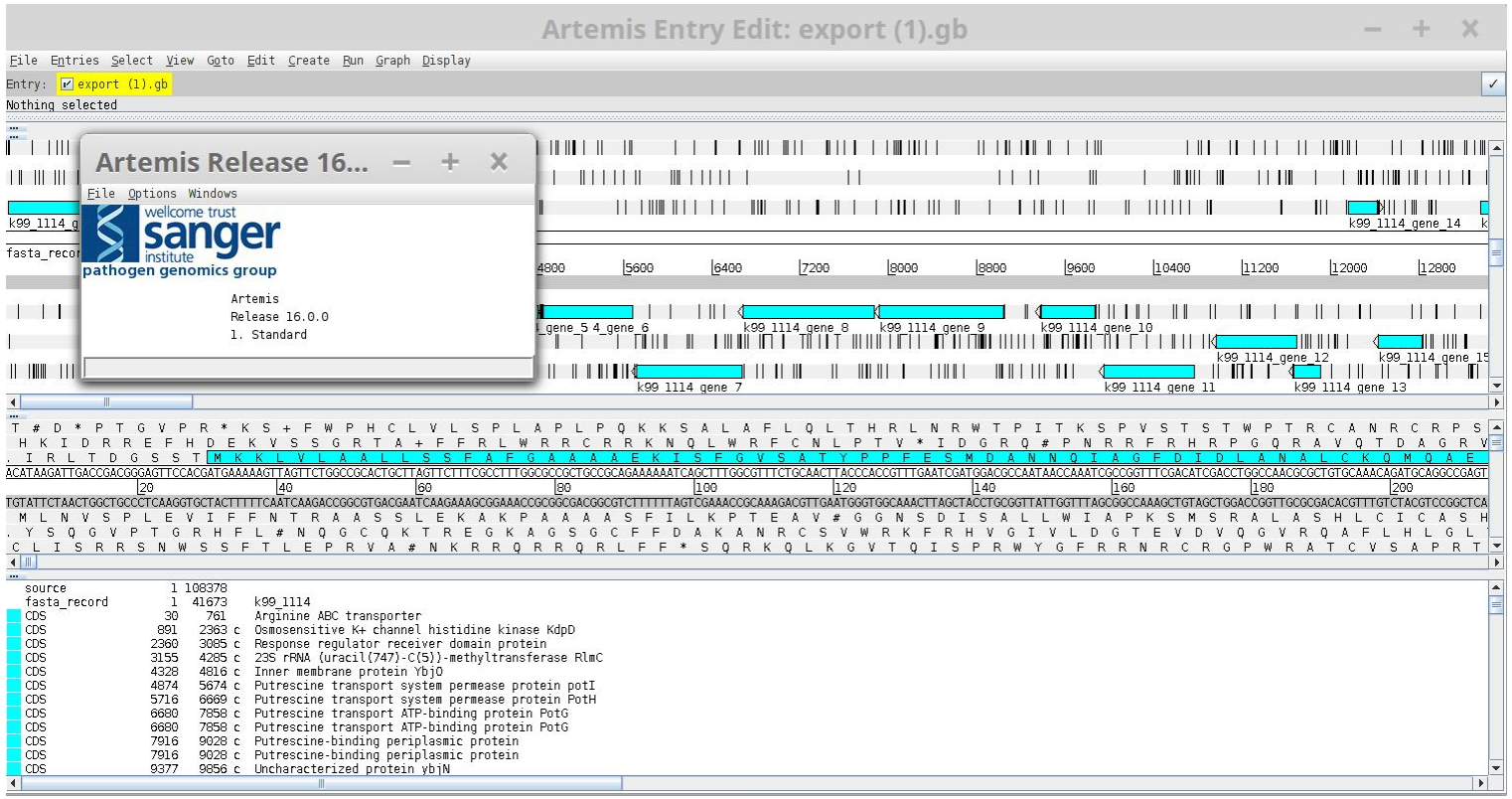
Functional Analysis

- PRIAM – Annotation of EC-numbers with RPS-blast
- Uniref50 – Clustered UniProt using Blast
- InterPro – Collection of 14 databases queried with InterProScan



Biological Overview

Functional Analysis



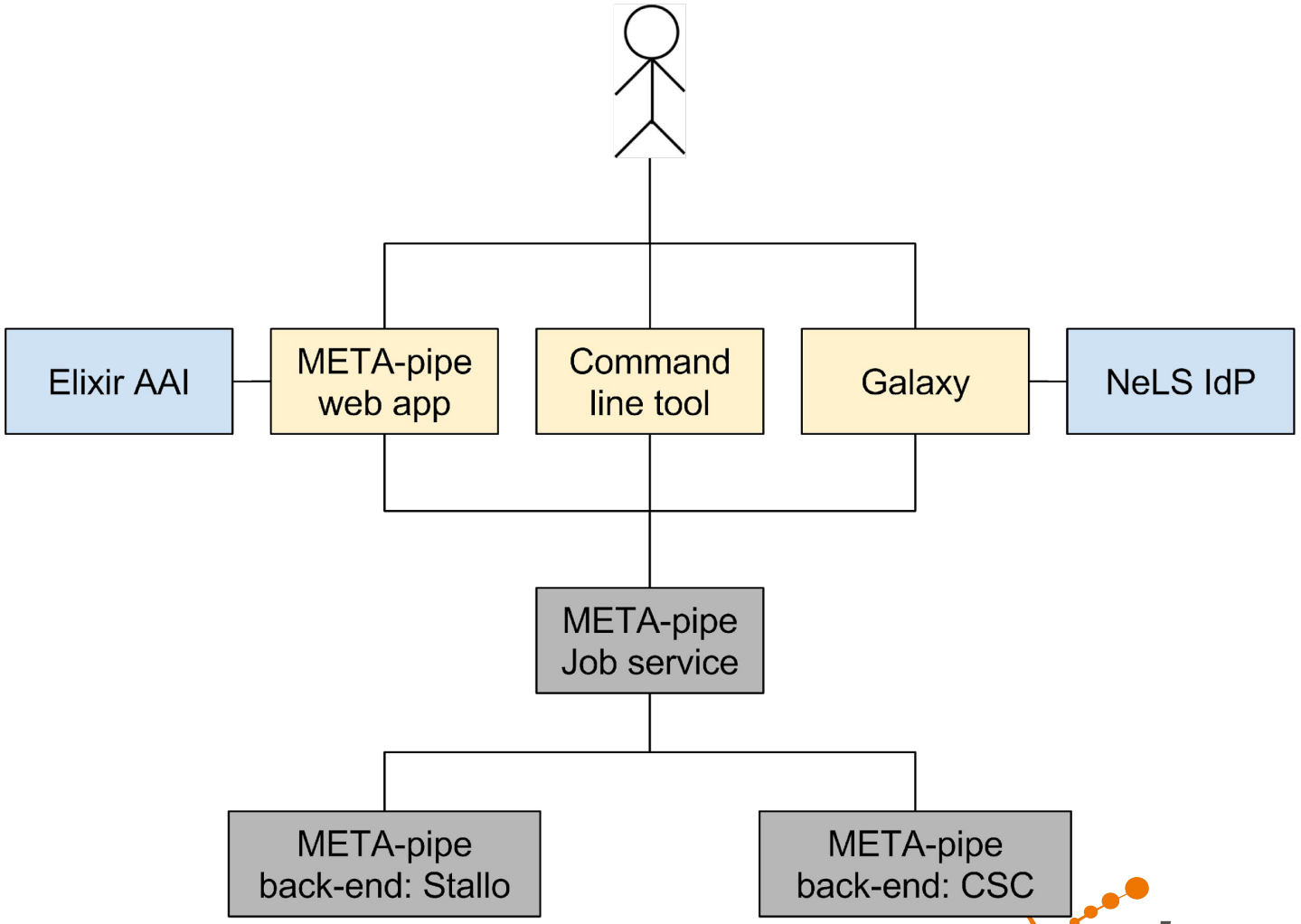
Inclusion and processing of MarDBs

- MarRef integrated with Meta-pipe
- MarDB scheduled for integration with Meta-pipe
- Meta-pipe currently processing samples for MarCat

MARINE
REFERENCE
DATABASES

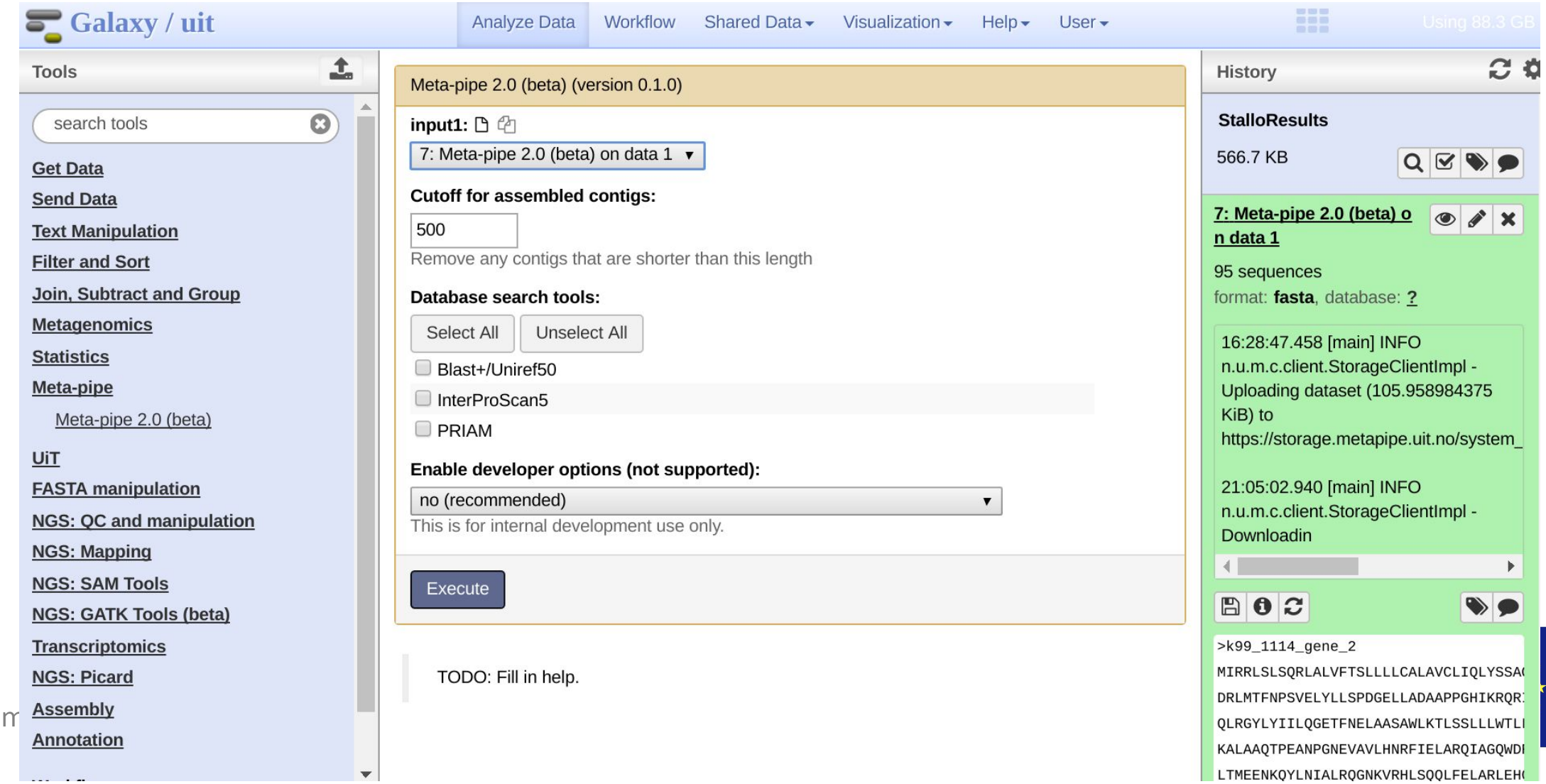
Tromsø, May, 2018

User Access


Elixir AAI
META-pipe web app
Command line tool
Galaxy
NeLS IdP
META-pipe Job service
META-pipe back-end: Stallo
META-pipe back-end: CSC

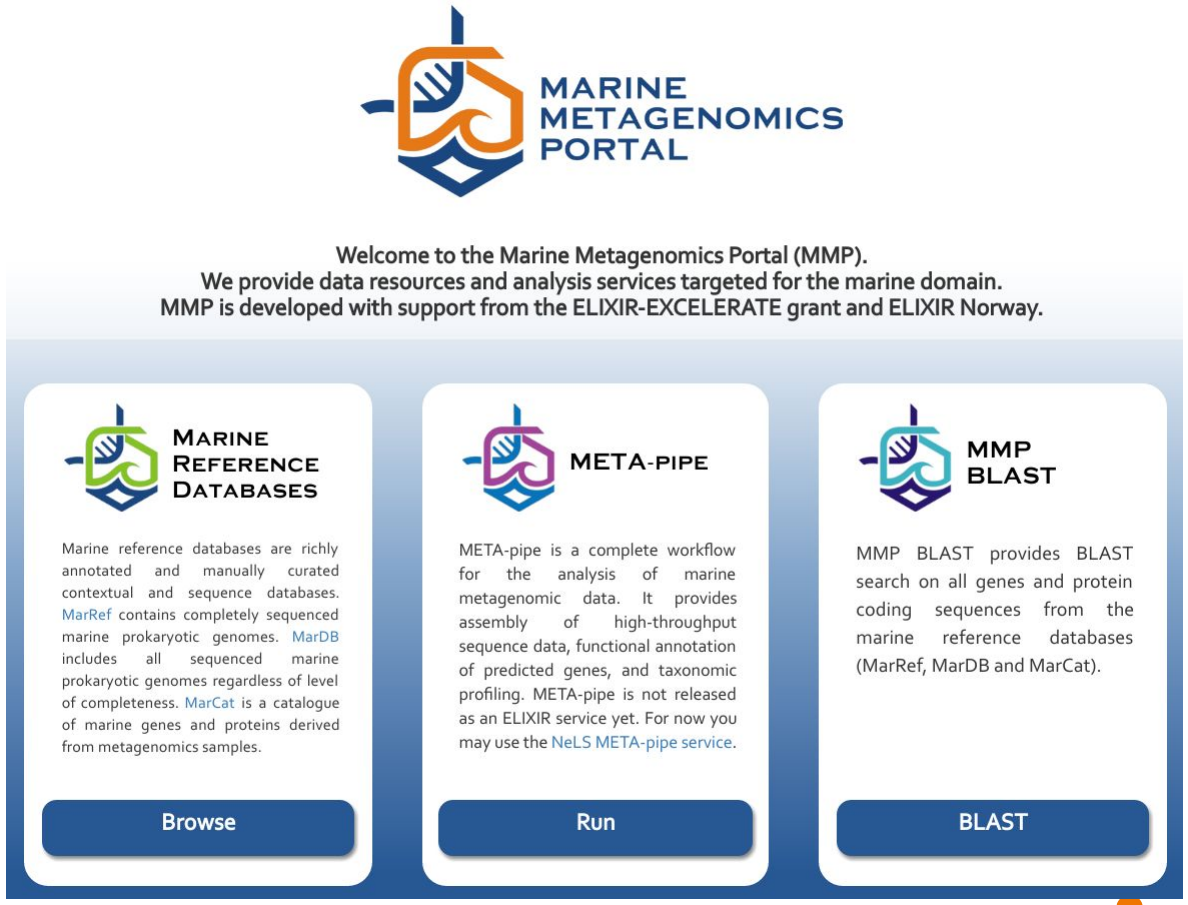
User Access

Galaxy with NeLS-idp



User Access

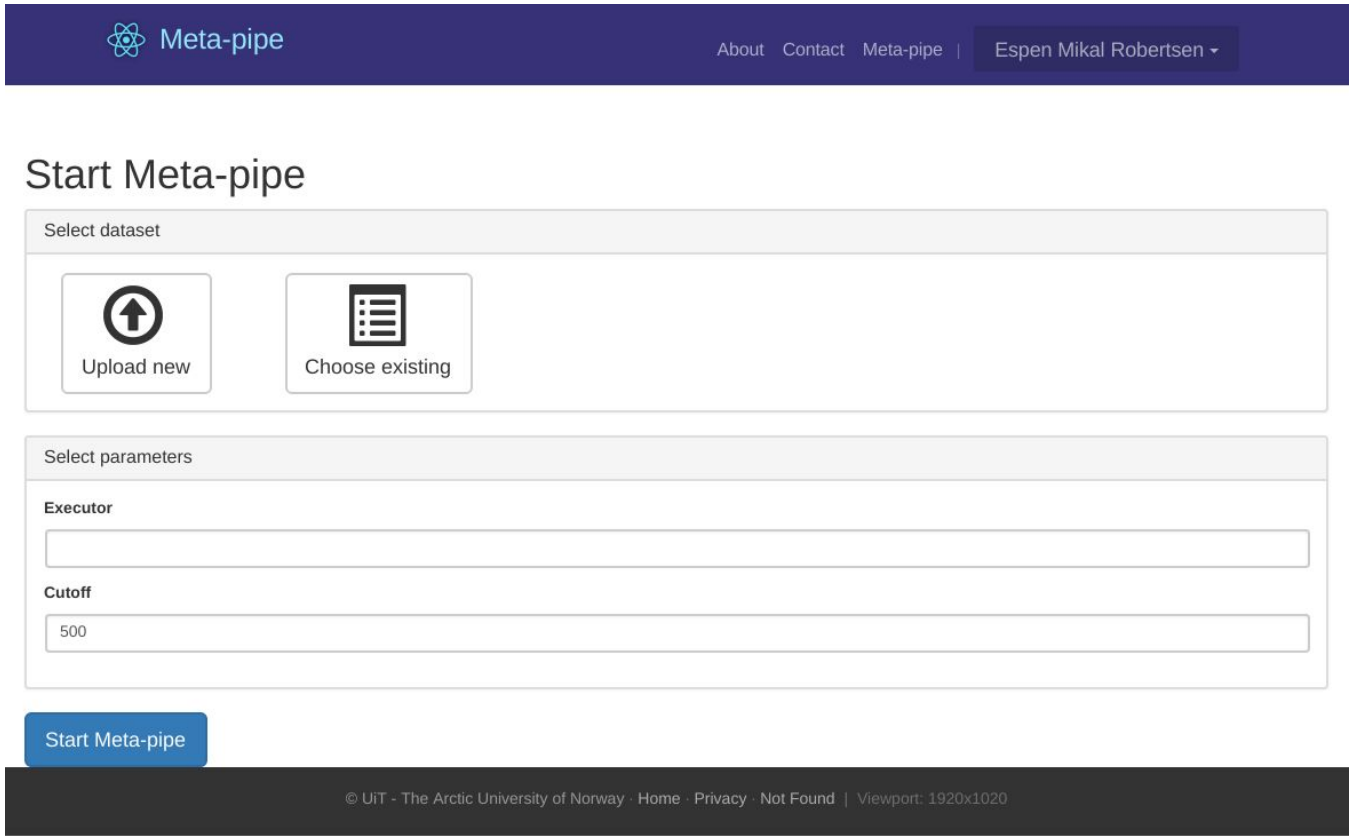
Webapp with ELIXIR-AAI



Tromsø, May, 2018

User Access

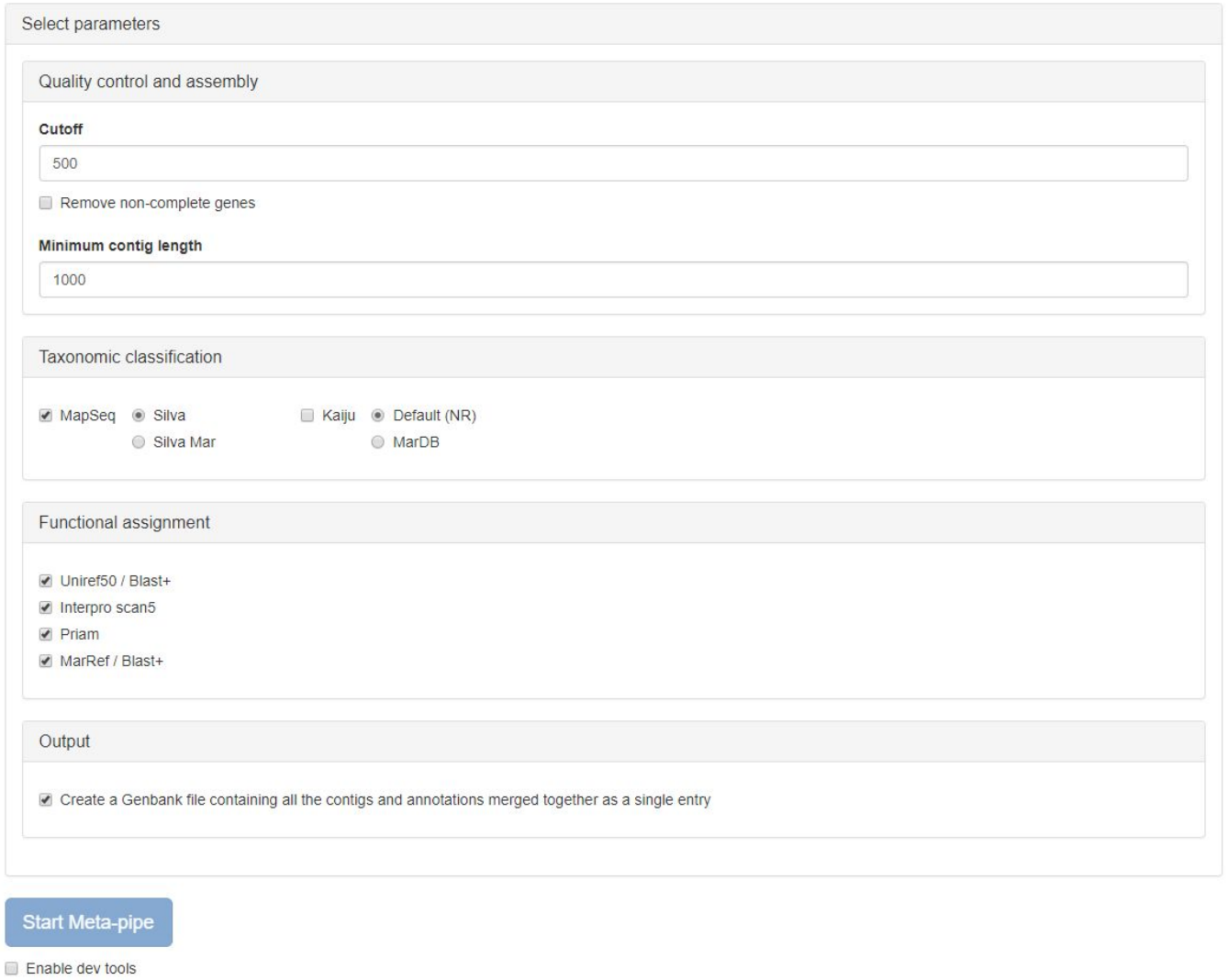
Webapp with ELIXIR-AAI



Tromsø, May, 2018

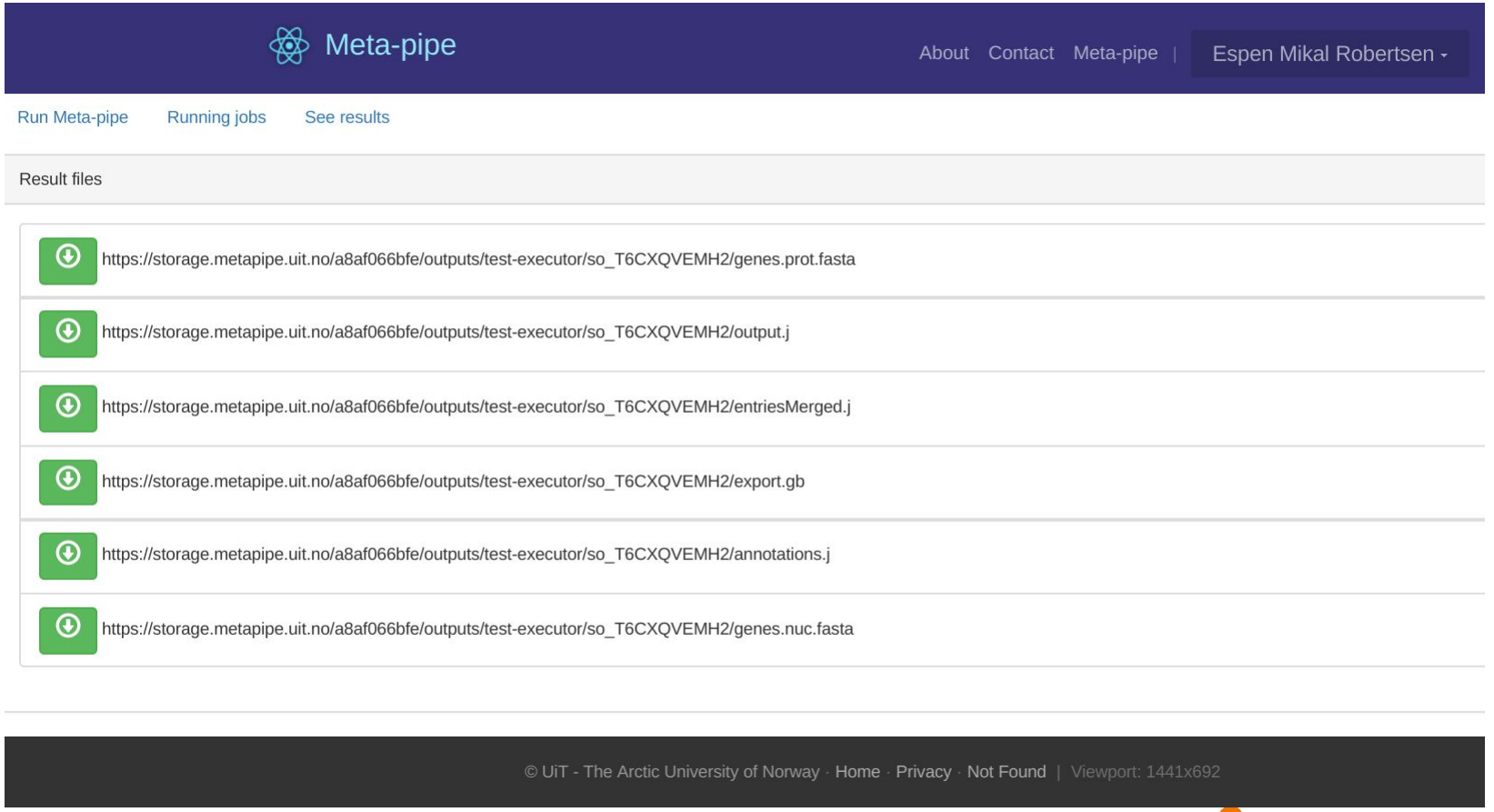
User Access

Webapp with ELIXIR-AAI



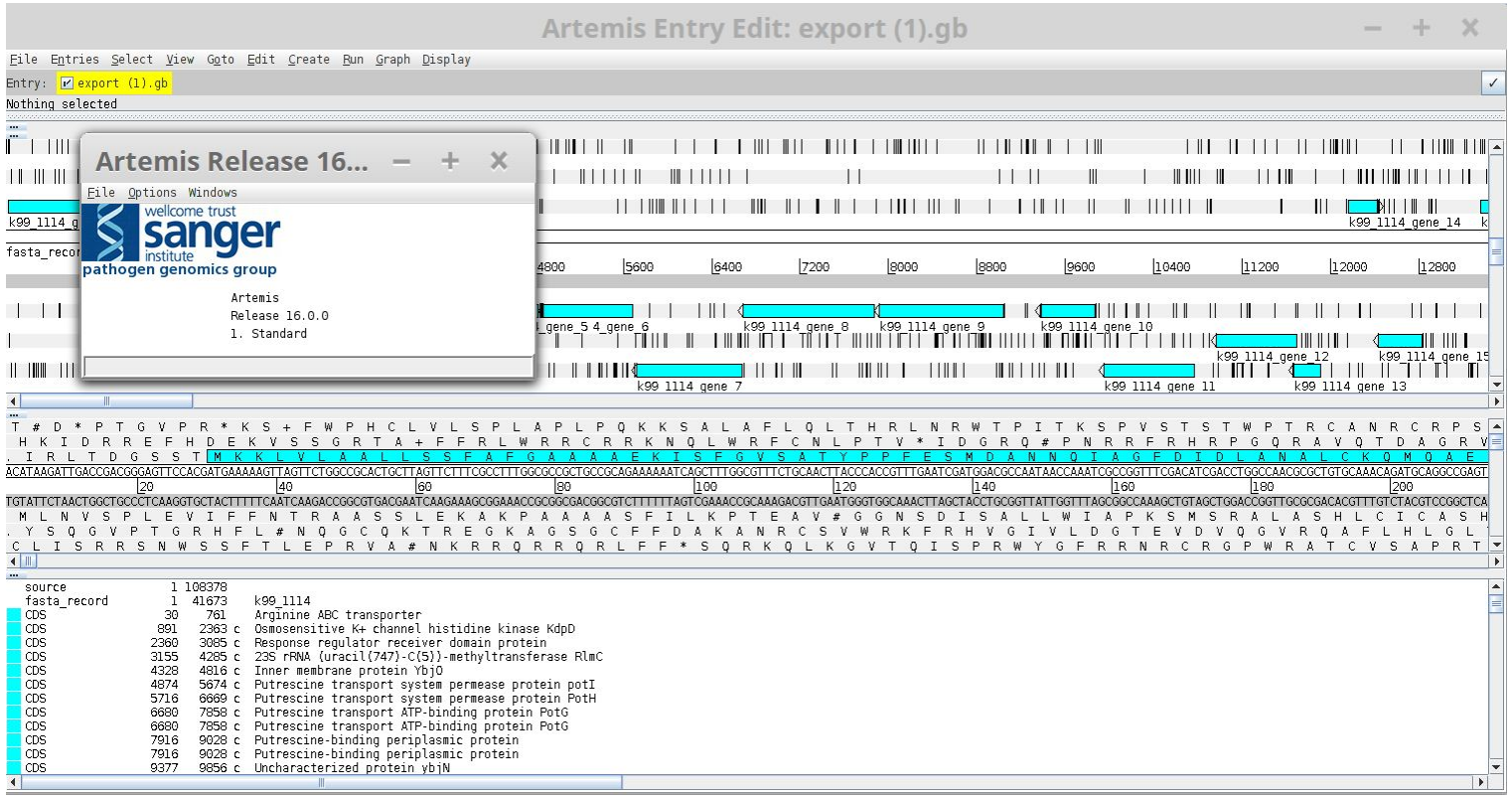
Tromsø, May, 2018

Output


Meta-pipe
About
Contact
Meta-pipe
Espen Mikal Robertsen
Run Meta-pipe
Running jobs
See results
Result files
https://storage.metapipe.uit.no/a8af066bfe/outputs/test-executor/so_T6CXQVEMH2/genes.prot.fasta
https://storage.metapipe.uit.no/a8af066bfe/outputs/test-executor/so_T6CXQVEMH2/output.j
https://storage.metapipe.uit.no/a8af066bfe/outputs/test-executor/so_T6CXQVEMH2/entriesMerged.j
https://storage.metapipe.uit.no/a8af066bfe/outputs/test-executor/so_T6CXQVEMH2/export.gb
https://storage.metapipe.uit.no/a8af066bfe/outputs/test-executor/so_T6CXQVEMH2/annotations.j
https://storage.metapipe.uit.no/a8af066bfe/outputs/test-executor/so_T6CXQVEMH2/genes.nuc.fasta
© UiT - The Arctic University of Norway · Home · Privacy · Not Found | Viewport: 1441x692

Output

Visualization of contigs in detail



Output

Visualization prototype

Overview | Sequence Distribution | Taxonomic Classification | Functional Assignment | Geographic Context To MarCat | Downloads

Histogram | Kernel density estimation

View histogram for:

Reads | Contigs | Genes, complete | Genes, truncated

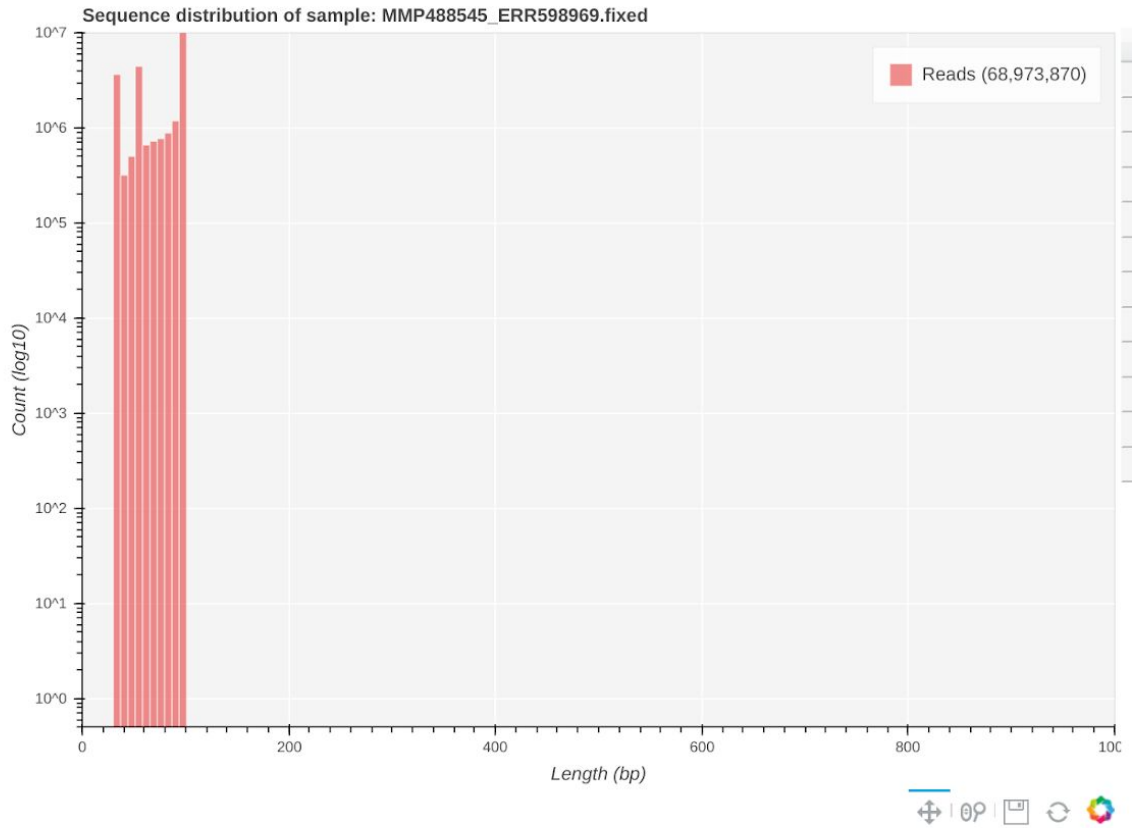
Y axis type

log

X scale: **0**

View KD-estimation for:

Reads | Contigs | Genes, complete | Genes, truncated



#	Metric	Reads	Contigs	Genes
0	Count (#)	68,973,870	162,298	62,591
1	Length (bp)	6,374,838,210	150,772,424	36,358,786
2	Over 100 bp	52,685,311	162,298	62,591
3	Over 500 bp	0	161,652	27,753
4	Over 1000 bp	0	39,001	8,392
5	Over 5000 bp	0	843	30
6	Over 10000 bp	0	177	0
7	Largest (bp)	101	166,928	8,936
8	Smallest (bp)	30	500	119
9	Average length (bp)	92	929	581
10	Median (bp)	101	708	452
11	N50	101	923	764

Thank you!

The Center for Bioinformatics team (SfB)

Alexandr Agafonov
Juan Fu
Espen Robertsen
Espen Åberg
Erik Hjerde
Inge Alexander Raknes
Lars Ailo Bongo
Nils Peder Willassen
Terje Klemetsen

SfB Center for Bioinformatics

UNIVERSITY OF TROMSØ

UiT / THE ARCTIC UNIVERSITY OF NORWAY

Tromsø, May, 2018

excelerate