

## Marine metagenomics – towards domain specific service

The study of genetic material sampled directly from marine environmental sources is still in its infancy, but is rapidly expanding. To prevent that processing and analysis of these samples becomes a bottleneck, where data production is faster than the speed users are able to make use of it, there is an urgent need to establish dedicated data management e-infrastructure and bioinformatics pipelines specialized for marine research. While EBI has developed EBI-metagenomics, a generic pipeline, which aims to provide insights into the phylogenetic diversity as well as

the functional and metabolic potential of the samples, the Norwegian node has developed Meta-pipe in the direction coupled with marine bioprospecting. In this pilot project the two pipelines will be harmonized in terms of interoperability in order to establish long-term sustainable service platforms and build a user community for marine metagenomics analysis in ELIXIR.

### Deliverables

[1] Harmonizing existing metagenomics pipelines (EBI-metagenomics and Meta-Pipe) to ensure interoperability.

[2] Assess specialized databases such as Merops, Panther and SFLD (Structure Function Linkage-Database) to evaluate whether these can enrich the output further.

[3] Explore and prototype with EMBL-EBI cloud based technologies.

[4] Report on gap analysis related to establishment of reference genomes for the marine environment. Also includes a workshop with ESFRI initiatives, EU FP7 projects and other appropriate groups/networks to discuss current status and needs of marine reference genomes.

[5] Report about the needs for specific investments in connection to services for the marine sector.

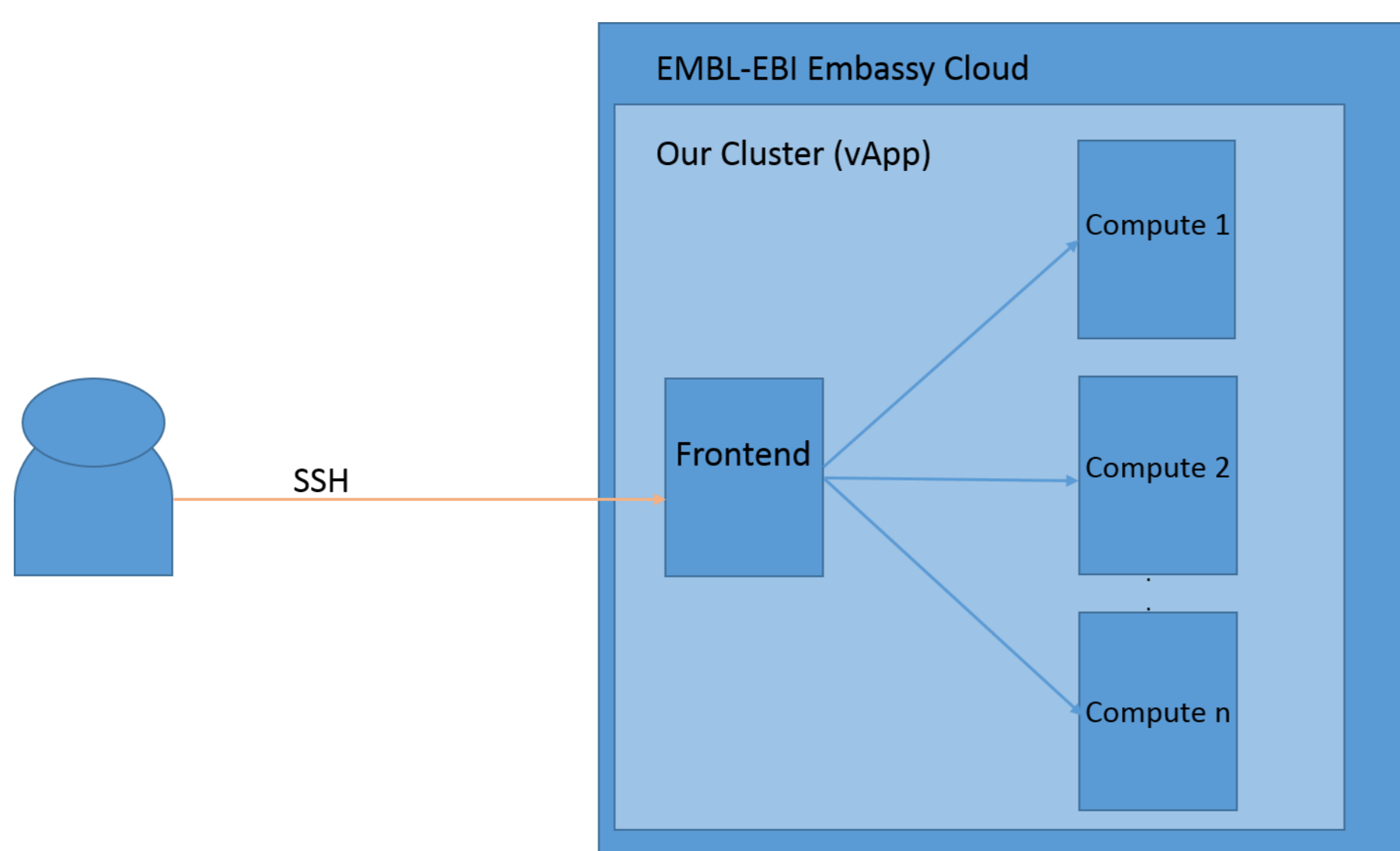


Figure 2: A schematic overview of the prototyping process with EMBL-EBI Embassy Cloud. Access to the virtual cluster via SSH, with one front end and 14 compute nodes.

### Status

[1] Specific intermediate pipeline steps has been identified where components needs to be improved or developed, such as pre-processing and removing 16S sequences before assembly. Additionally, assessing new databases as mentioned in [2] and implementing Interproscan in Meta-pipe is in progress.

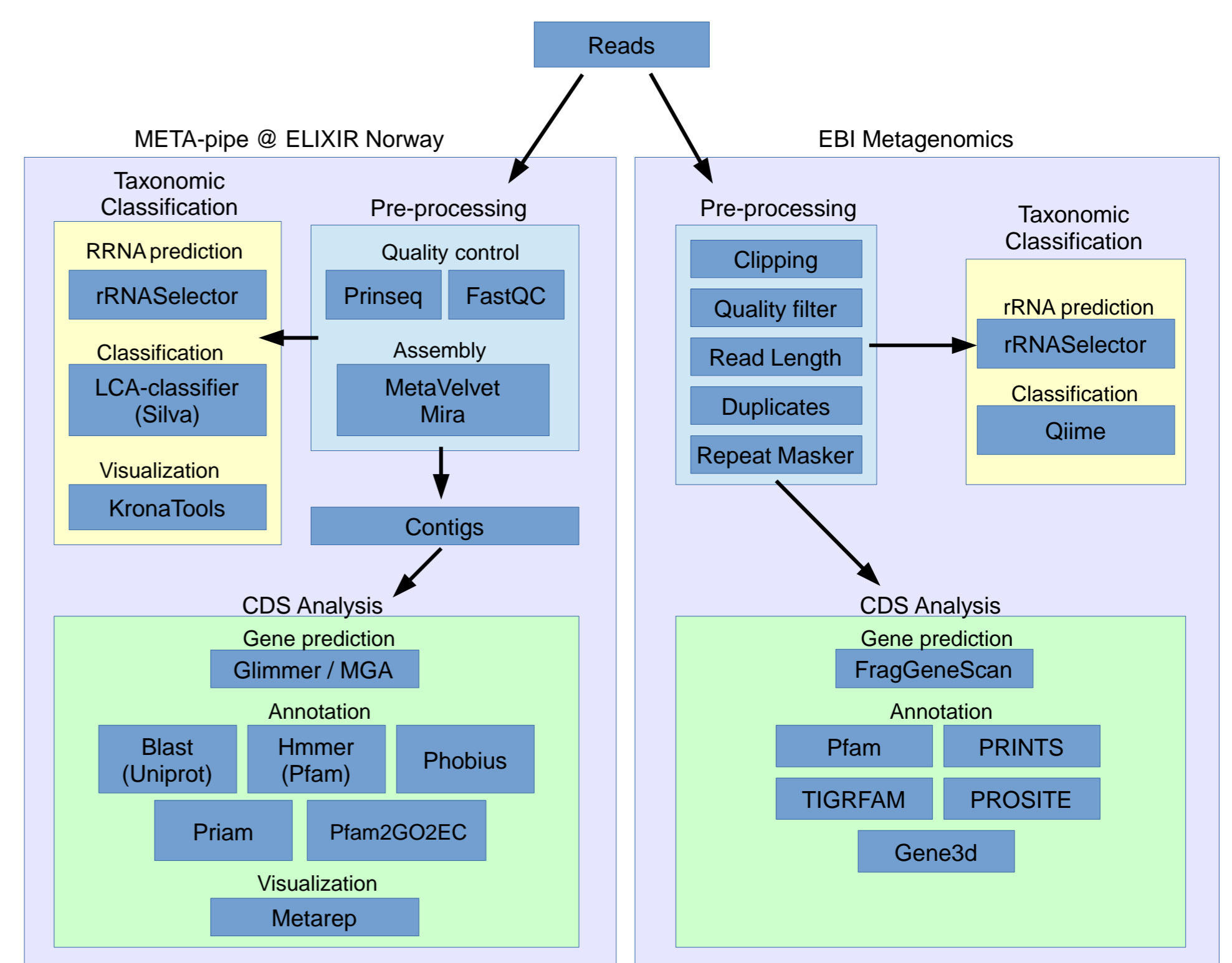


Figure 1: A comparison between the two pipelines Meta-pipe (left) and Ebi Metagenomics (right). Both pipelines provide similar functionality, but with different modules and approaches. Most noteworthy, Meta-pipe utilizes assembled contigs as input to produce mainly full length sequences, while EBI metagenomics accepts reads as input to produce a higher quantity, but mostly fragmented set of sequences

[3] EMBL-EBI Embassy Cloud services has been used to set up a virtual cluster with 15 nodes total and a shared NFS file system. To experiment with the possibilities this cloud service provides, a Blast job against UniprotKB using Slurm as a scheduler has been computed successfully.

[4] ELIXIR Collaboration Workshop on Marine Informatics arranged at ELIXIR Hub, Hinxton, 16-17 March 2015.

### Contact

META-pipe@ELIXIR Norway

EBI Metagenomics Portal

info@elixir-norway.org  
www.elixir-norway.org

www.ebi.ac.uk/metagenomics/  
www.ebi.ac.uk/support/metagenomics