



UNIVERSITY OF
GOTHENBURG

Pipeline for large scale data mining and mega-chronogram inference

Ruud Scharn, Alexandre Antonelli, Rutger Vos, Karin Nilsson, Mats Töpel

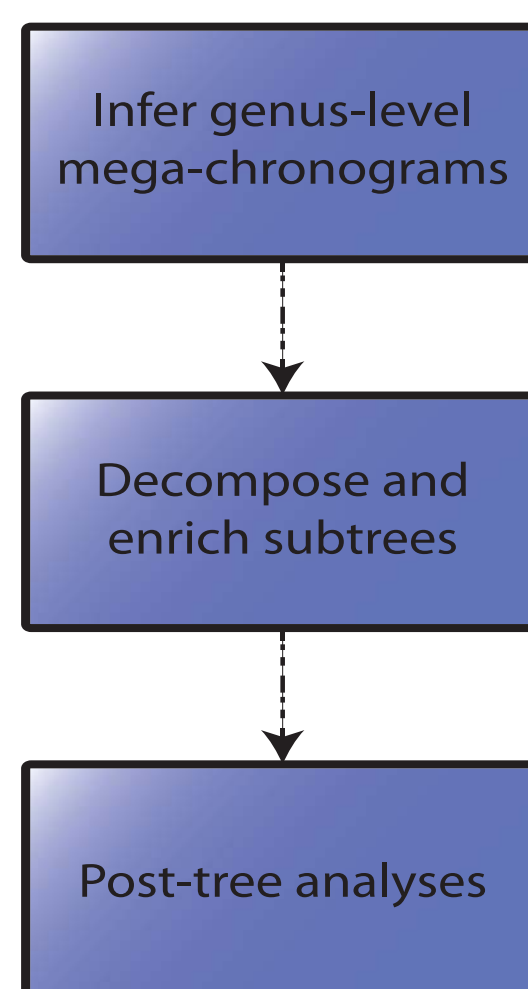


Introduction

Large scale molecular analyses are relying on data from internet databases such as GenBank and Gbif in increasingly large amounts. It is thus getting more important to optimize the use of the databases at hand. The pipeline called SUPERSMART* described here will use extensive data mining to produce up to date mega-chronograms of plant, animal and fungi kingdoms that can be easily pruned on area, clade of interest, biome, etc.

*Self Updating Platform for Estimating Rates of Speciation and Migration, Ages and Relationships of Taxa

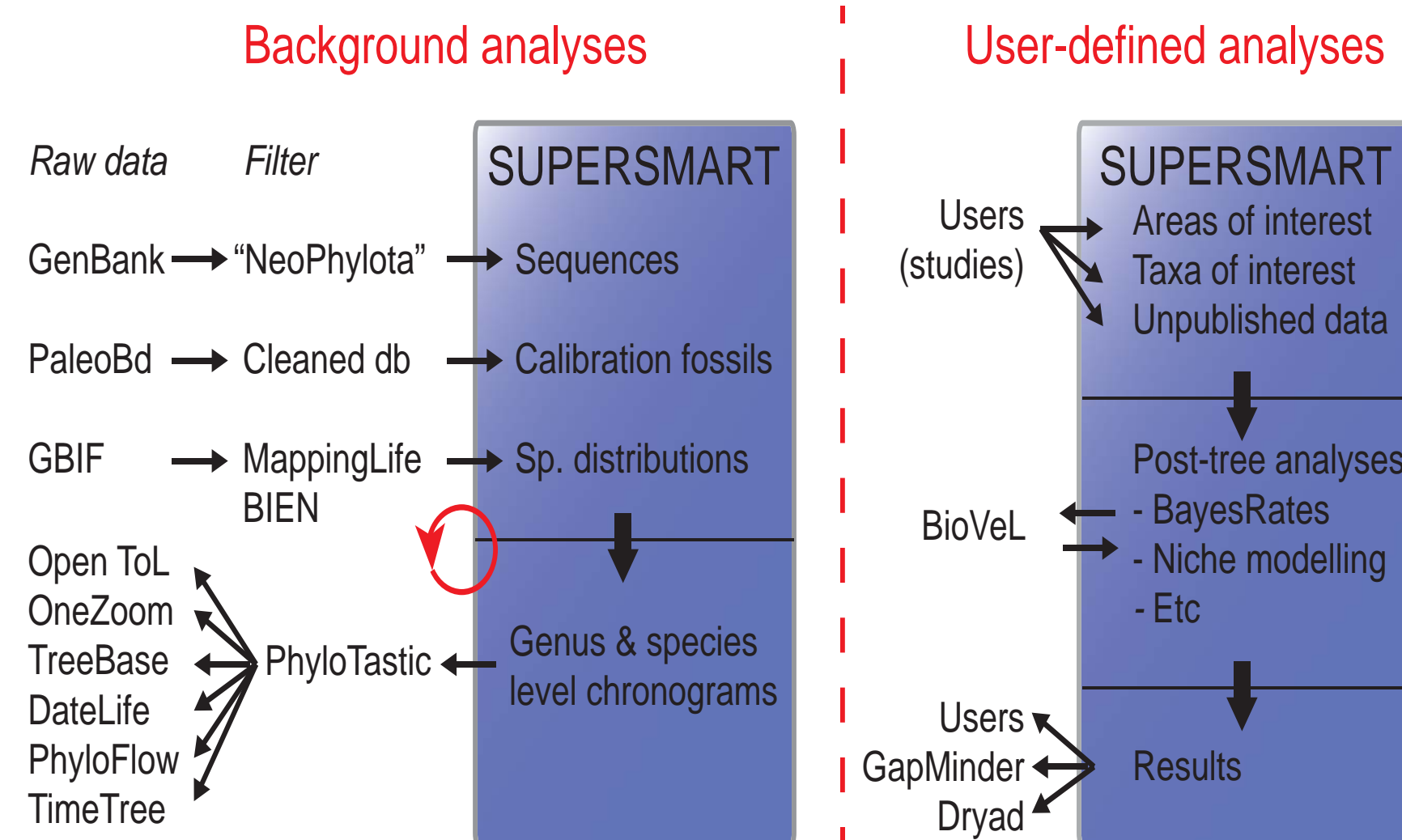
Aims



Create a Pipeline that is:

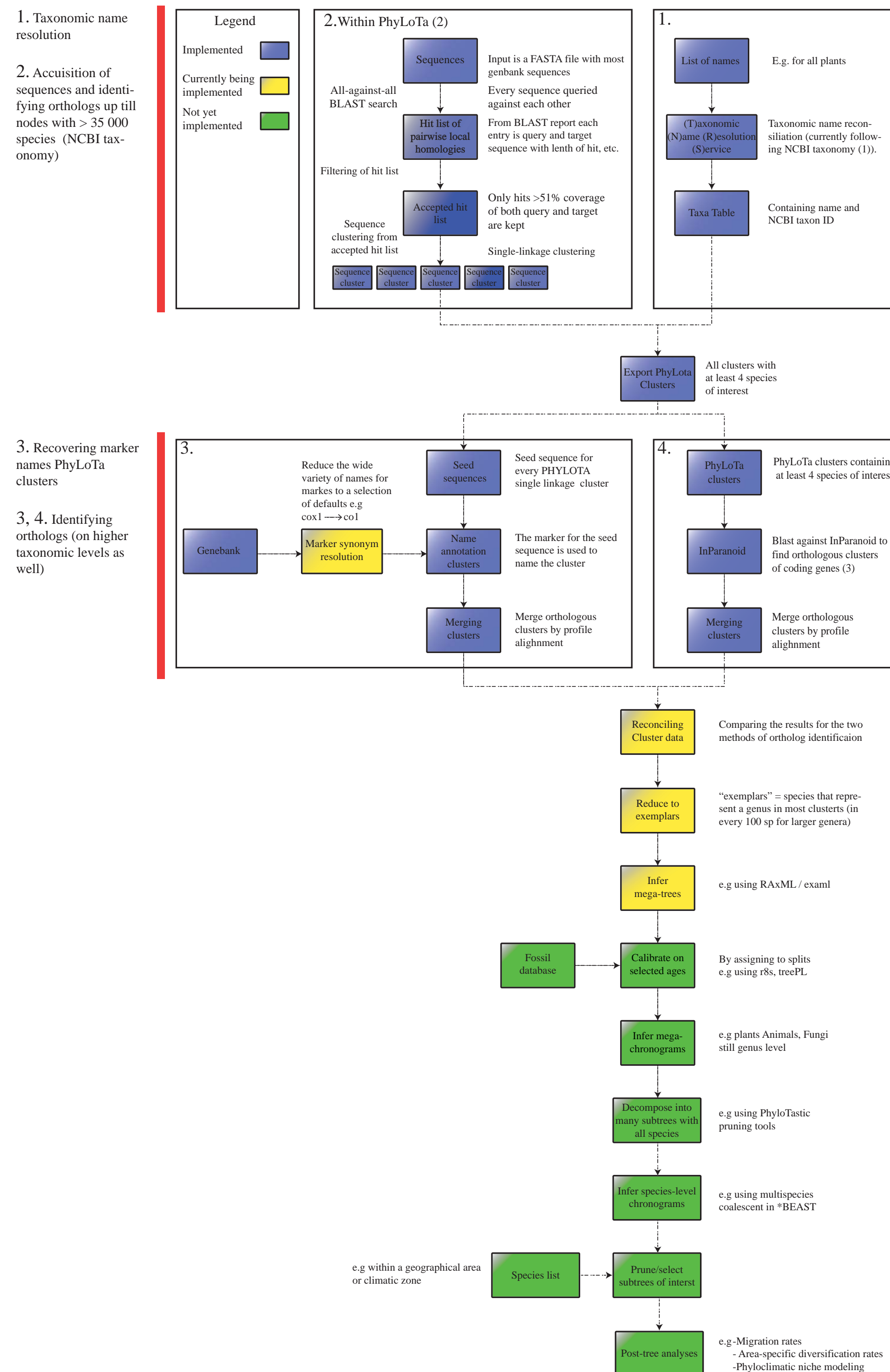
- Self Updating
- Highly Modular and self adapting
- Allows for many different types of post-tree analyses

Interactions with other initiatives

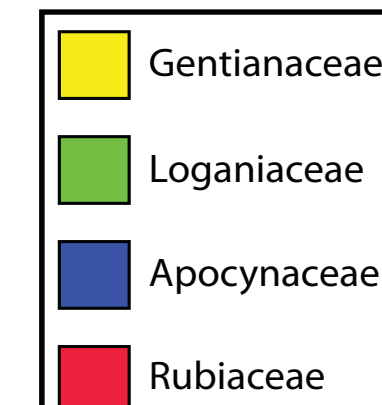


Overview pipeline

Flowchart of the pipeline detailing its current implementation and its progress



Results



Species level tree of the order of the Gentianales generated in RAxML using the data-matrix generated by step 1 through 4 of the pipeline.

Sources

- (1) Boyle, B. et.al. The taxonomic name resolution service: an online tool for automated standardization of plant names. BMC Bioinformatics. 2013, 14:16.(?)
- (2) Sanderson, M. J., D. Boss, D. Chen, K. A. Cranston, and A. Wehe. 2008. The PhyLoTA Browser: processing GenBank for molecular phylogenetics research. Syst. Biol. 57:335-346.
- (3) O'Brien Kevin P, Remm Mado and Sonhammer Erik L.L. "Inparanoid: A Comprehensive Database of Eukaryotic Orthologs" NAR 33:D476-D480 (2005)
- (4) Alexandros Stamatakis : "RAxML-VI-HPC: Maximum Likelihood-based Phylogenetic Analyses with Thousands of Taxa and Mixed Models", Bioinformatics

Contact information



Ruud Scharn,
Master student at Gothenburg University
Currently presenting

Email: ruud.scharn@gmail.com
Address: St Pauligatan 416 60 SWEDEN



Dr. Alexandre Antonelli
Assistant professor Gothenburg University
Project leader on the SUPERSMART project.
Email: alexandre.antonelli@bioenv.gu.se
Address: Department of Biological and Environmental Sciences, Carl Skottsbergs gata 22B, SE-413 19 SWEDEN



Dr. Rutger Vos
Bioinformaticist Naturalis Biodiversity Center, Leiden
Development lead on the SUPERSMART project
Email: rutger.vos@naturalis.nl
Address: Postbus 9517, 2300 RA, Leiden, THE NETHERLANDS

More information

<http://www.supersmart-project.org/>