

# SUPERSMART - A Self-Updating platform for Estimating Rates of Speciation and Migration, Ages, and Relationships of Taxa

Hannes Hettling <sup>\*</sup>      Alexandre Antonelli <sup>†,‡</sup>      Rutger A. Vos <sup>\*</sup>

17<sup>th</sup> Bioinformatics Open Source Conference (BOSC) 2016, Orlando, FL, USA

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

**Repository:** <https://github.com/naturalis/supersmart>

**License:** MIT license (see <https://opensource.org/licenses/MIT>)

## Background

Massive volumes of biological data, including molecular sequences and fossils, are accumulating quickly and are frequently made available in public repositories. This wealth of data facilitates large-scale quantitative analyses revealing the evolutionary mechanisms that generate biodiversity. However, exploiting the data to generate time-calibrated phylogenies that represent the evolutionary relationships and history among species is generally a time-consuming and error-prone procedure involving a variety of different analysis tools.

## Results

Here we present the SUPERSMART *virtual research environment* which provides an integrative solution for automated mining, cleaning and assembly of molecular sequencing data and the inference of large, time-calibrated species phylogenies. Phylogenetic inference for large numbers of taxa and sequence data is accomplished by a recursive divide-and-conquer approach implemented in our platform, engineered for massive parallelization. Our software is therefore scalable to build phylogenies for tens of thousands of taxa. SUPERSMART comprises state-of-the-art analysis tools for taxonomic name resolution, sequence alignment (e.g. *muscle*, *mafft*), maximum likelihood and Bayesian tree inference (*RaXML*, *ExaML*, *ExaBayes*, *\*BEAST*) and molecular dating (*TreePL*, *\*BEAST*). Deployment of the platform including the above tools is accomplished using the virtualization frameworks *Docker* and *Vagrant*, rendering the installation of any dependencies unnecessary. A graphical user interface to SUPERSMART is provided via *Galaxy* web services. The core of SUPERSMART is implemented in Perl/BioPerl in a highly modular fashion. In the future, we aim to integrate single SUPERSMART modules, such as wrappers for the above analysis tools, into the OBF BioPerl library.

## Conclusions

SUPERSMART is an integrated platform for data mining and phylogenetic inference workflows. From simply a list of taxa of interest, users can generate dated species phylogenies in merely a few steps. Our software is easily deployable, platform independent and can run on personal computers and high-performance cluster infrastructures. SUPERSMART is originally a command-line application but can alternatively be accessed through *Galaxy*. Thereby, we give low-threshold access to phylogenetic inference workflows to researchers with little computer expertise.

---

<sup>\*</sup>Naturalis Biodiversity Center, Darwinweg 4, 2333 CR Leiden, the Netherlands. Email: [hannes.hettling@naturalis.nl](mailto:hannes.hettling@naturalis.nl)

<sup>†</sup>University of Gothenburg, Department of Biological and Environmental Sciences, Box 461, SE-405 30 Gothenburg, Sweden

<sup>‡</sup>Gothenburg Botanical Garden, Carl Skottsbergs gata 22A, SE-41319, Gothenburg, Sweden