

New Generation Phylogeny.fr

Refactoring Phylogeny.fr for innovative phylogenetic services



Damien Correia¹, Fabien Mareuil², Olivia Doppelt-Azeroual², Vincent Lefort³, Sarah Cohen-Boulakia⁴, Olivier Gascuel^{2,3}
¹Institut Français de Bioinformatique (IFB); ²Centre de Bioinformatique, Biostatistique et Biologie Intégrative (C3BI), USR 3756, Institut Pasteur;
³Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM);
⁴Laboratoire de Recherche en Informatique (LRI).

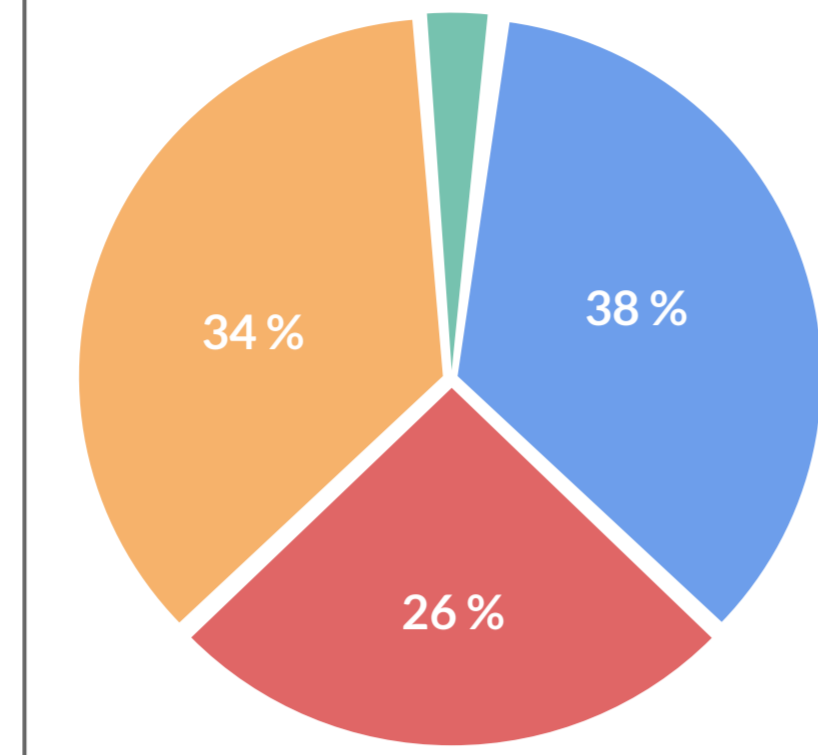


Introduction

Phylogeny.fr[1] is a French website made to help biologists with no experience in phylogeny analyze their data in a robust way. It provides a set of tools to perform online phylogenetics analyses from alignment to tree visualisation.

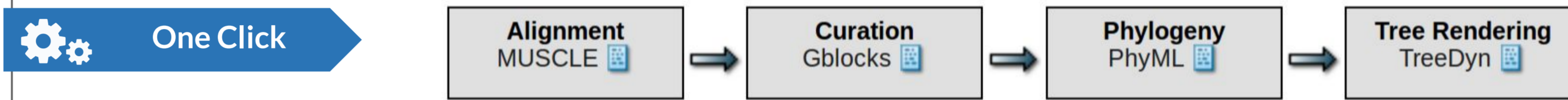
Based on users expectations and inspired from the previous version of Phylogeny.fr, the new generation will propose a redesigned interface, based on a robust backend workflow management system.

How Phylogeny.fr is used ?



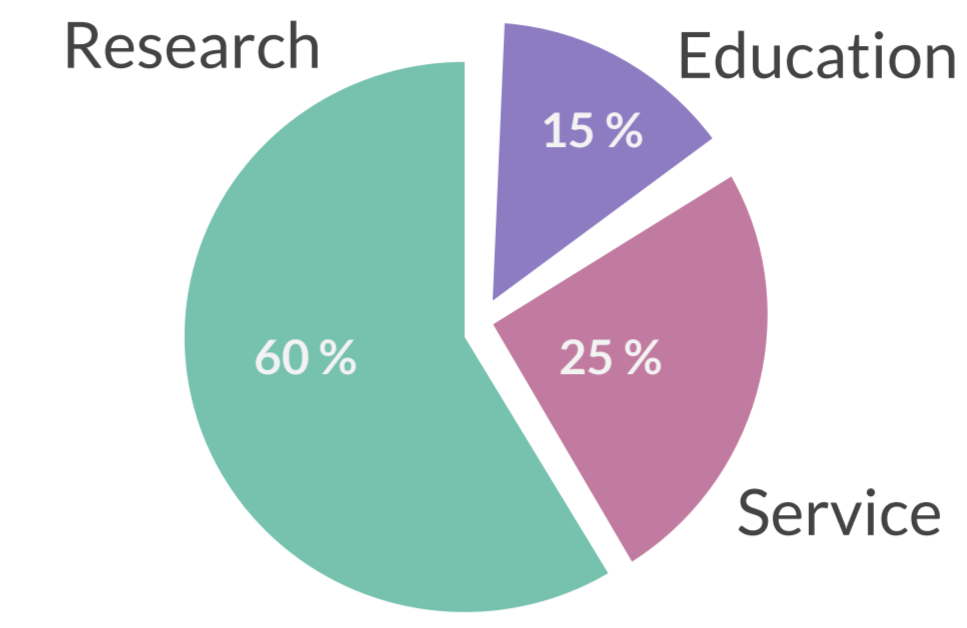
- **One Click:** Predefined workflows with auto compute options.
- **Advanced:** Predefined workflows with the possibility to set up each option of the tools.
- **A la carte:** Users make their own workflows by choosing tools and setup options.
- **Other:** Standalone tools

*Charts extract from users survey



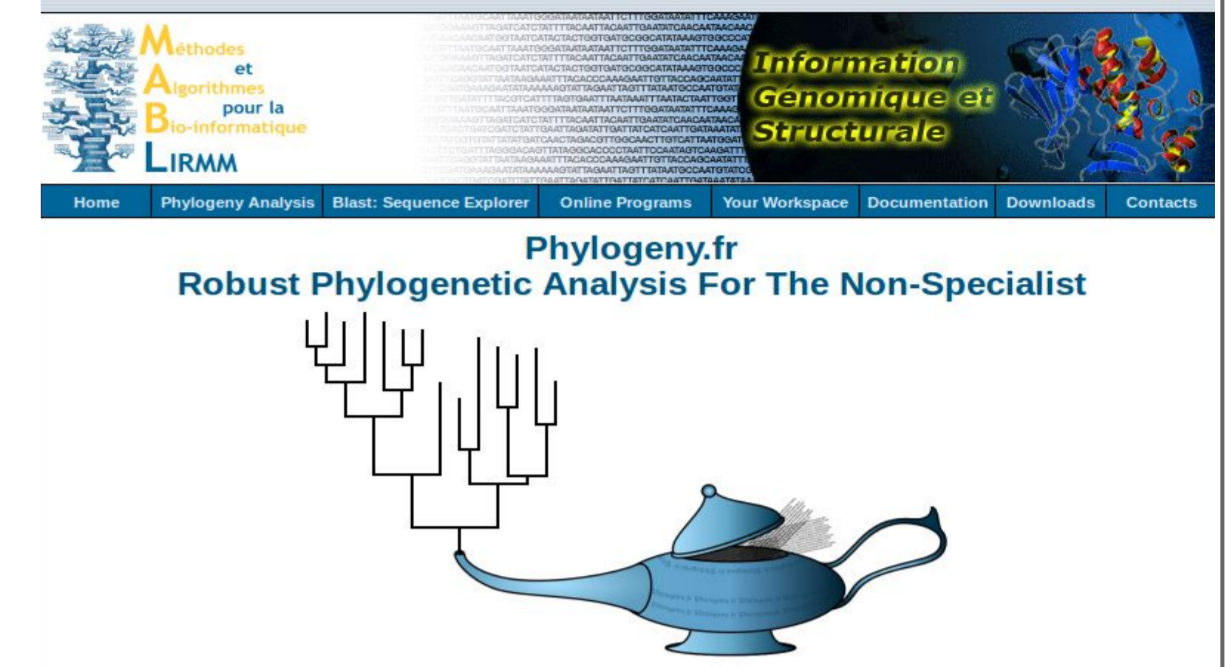
1 Phylogeny.fr

In which context ?



Today

- +50,000 analyses / month
- +2,000 citations
- +100 Trees inference / Day



Phylogeny.fr is a free, single to use web service dedicated to reconstructing and analyzing phylogenetic relationships between molecular sequences.

2 NGPhylogeny.fr

The new version will provide a large set of updated phylogenetic analysis programs and a set of optimized, robust and pre-designed workflows, with the ability to extend the basic phylogenetic workflow in various and rich contexts. This project aims to build an easy-to-install environment dedicated to phylogenetic analyses, coupled with a totally new visualization layer, on top of the workflow management system.

What's new ?

- ★ Ease to update tools
- ★ Fastest and replayed analyses
- ★ More available methods and tools
- ★ Downloadable and extensible workflows
- ★ Available on Virtual Machine

3 Workflow drive by Galaxy

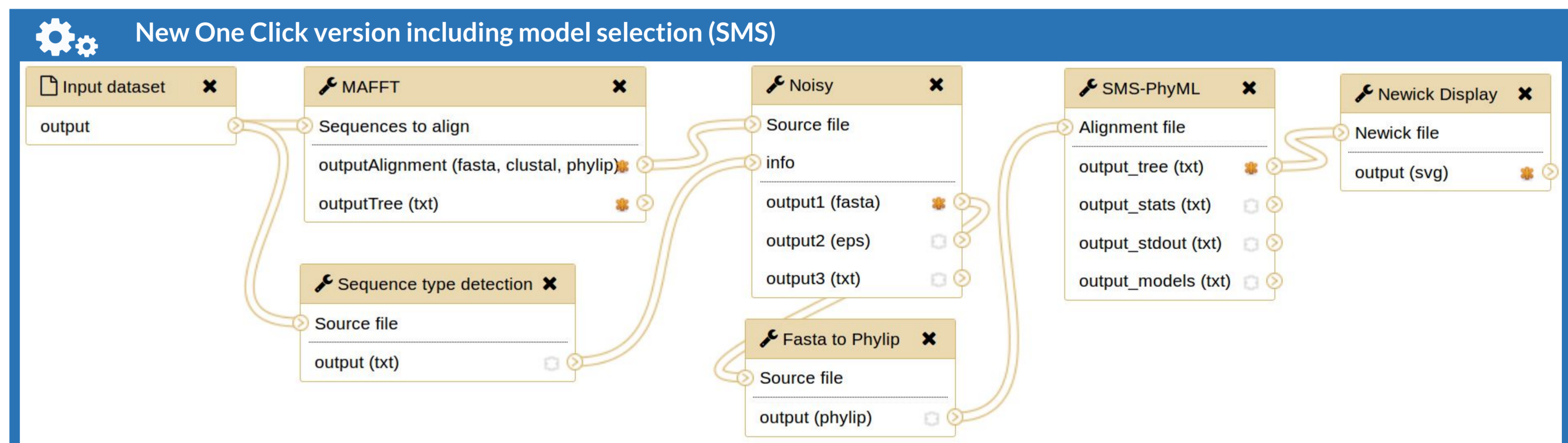
Galaxy[2], is a scientific workflow management system, which provides means to build multi-step computational analyses akin to a recipe. All programs are encapsulated into Galaxy thus making the system able to deal with large groups of users and/or large sets of data. One last very interesting functionality is reproducibility essential for the kind of analysis realized in phylogeny project.

Methods Available soon

- ★ Alignment
- ★ Curation/cleaning
- ★ Model selection
- ★ Tree rendering
- ★ Phylogeny inference (ML/Bayes)

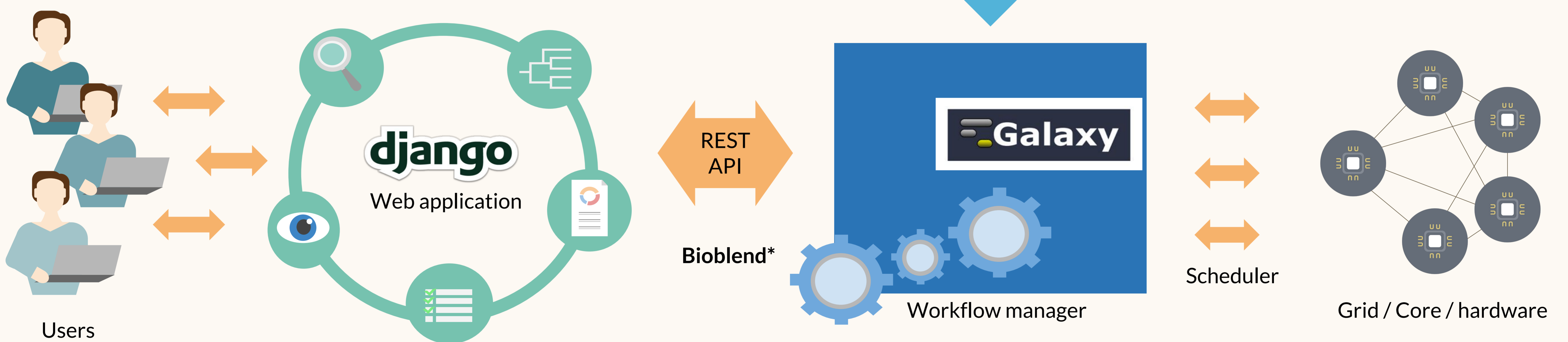
Galaxy community

The Galaxy community provides a tool sharing system called 'toolshed' which aims to ease the tools installation in Galaxy Instances. People work internationally to add bioinformatic tools and most of them are already integrated.



4 How does it work ?

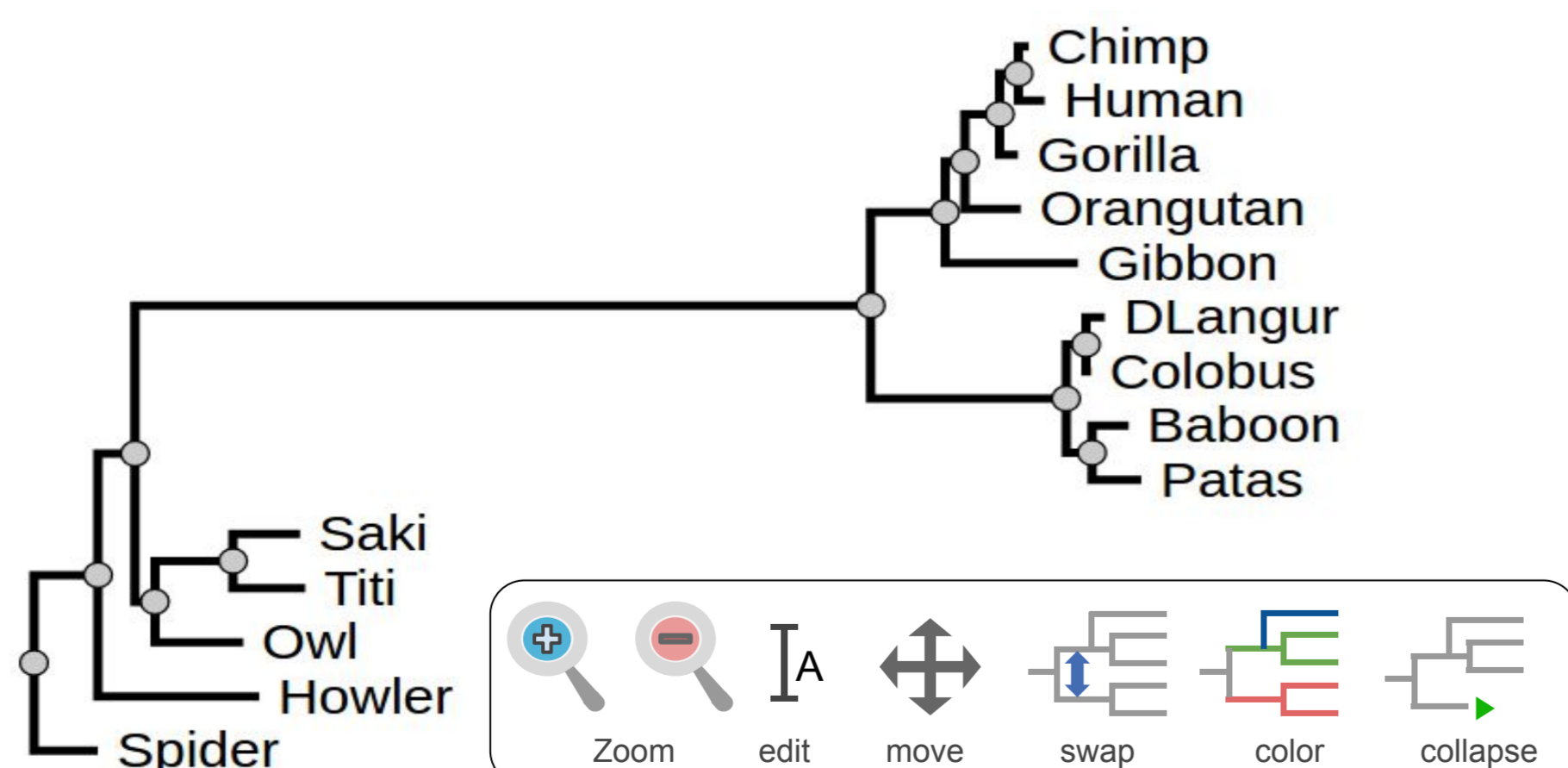
NGPhylogeny.fr



*Bioblend[3] is a python library built to remotely interact with Galaxy and use an API key authentication to launch tools.

5 Visualization

A new visualization tool based on Javascript will be developed to dynamically interact with phylogenetic trees.



6 Distribution

NGPhylogeny.fr is designed to be Open source software distributed under the GPL licence. It can be easily deployed on any Linux computer, provided that a Galaxy instance is already installed. Moreover we will provide several ways to use and install the application, such as virtual machines and Docker repositories.



Conclusion

With its new architecture and updated tools, NGPhylogeny.fr will offer a robust and accessible analysis platform for the scientific community. It will be available through both the website NGPhylogeny.fr and a package offering an easy-to-install distribution on any sufficiently powerful machine.

References: [1] Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O. Nucleic Acids Res. 2008, 36(Web Server issue):W465-9
 [2] Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Goekse J, Nekrutenko A, Taylor J; Galaxy Team. Genome Biol. 2010;11(8):R86.
 [3] BioBlend: automating pipeline analyses within Galaxy and CloudMan. Sloggett C1, Goonasekera N, Afgan E. Bioinformatics. 2013 Jul 1;29(13):1685-6.

Keywords: workflow, phylogeny, galaxy, educational, end-users