RapGreen, a Java package to analyse, display and explore interactively phylogenetic trees
Jean-François Dufayard, Delphine Larivière, Manuel Ruiz, Frederic de Lamotte Guéry

To cite this version:
Jean-François Dufayard, Delphine Larivière, Manuel Ruiz, Frederic de Lamotte Guéry. RapGreen, a Java package to analyse, display and explore interactively phylogenetic trees. Bioinformatics, Oxford University Press (OUP), 2018, In Press. hal-01885400

HAL Id: hal-01885400
https://hal.archives-ouvertes.fr/hal-01885400
Submitted on 1 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NoDerivatives 4.0 International License
RapGreen, a Java package to analyse, display and explore interactively phylogenetic trees

Jean-François Dufayard¹, Delphine Larivière¹, Manuel Ruiz¹ and Frédéric de Lamotte²

¹ AGAP, Univ Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France
² CIRAD, UMR AGAP, F-34398 Montpellier, France - Univ Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France

Short structured abstract

The development of powerful but easy to use software is a permanent challenge, particularly in the field of bioinformatics as biologist users have various background, sometimes pretty far away from computer science. The RAP software was developed (Dufayard et al., 2005) in order to i) automatically reconcile phylogenetic trees with species tree, ii) display phylogenetic trees and iii) explore phylogenetic tree collections using the FamFetch interface. The main weaknesses of this implementation was its lack of portability because of its standalone Java software nature. It was difficult to install the service on a user infrastructure. It was also uneasy for users learn how to use it before being able to obtain satisfactory results. Simple tasks like adding new trees required more than basic skills.

In order to offer a more powerful and more friendly service, a new package called RapGreen has been developed. It is composed of 3 modules: i) a Java package to compute tree manipulations like reconciliation, rooting or several statistics, ii) a web interface in order to explore phylogenetic tree collections using a tree pattern, and iii) a web visualisator able to aggregate heterogeneous data around tree structures. This new implementation is easy to install and it offers two interfaces for directly available on websites.

Phylogenetic tree management Java package

Any careful phylogenetic analysis requires to find a convenient root, and infer events like gene duplications or losses. The RapGreen Java package addresses such requirements as it allows many manipulations and comparisons on phylogenetic trees.

The main features are i) the RapGreen tree reconciliation, which can infer gene duplications and allows to root a tree comparing the gene tree with the corresponding species tree ; ii) a daemon which is able to communicate with the tree pattern matching interface in order to manage the computational side of this service on a distant computing cluster. The tree data structure is recursive, and algorithms are implemented as described in the RAP original publication (Dufayard et al. 2005). Thanks to a full code refactoring, this version is now compatible with most recent JDK versions, and benefits of recent improvements of the Java virtual machine.

InTreeGreat, an integrative tree visualisator

In the post genomics era, data sets have become bigger and bigger. Their analysis can be cumbersome and requires a very efficient interface; friendly enough to help the end-user to be the main actor of the phylogenetic analysis of his data. InTreeGreat is a Javascript/PHP interface, compatible with every standard web navigator without plugin nor add-on requirement. It is able to display and explore any tree in Newick or Newick extended format. It is able to deal with branch and leaf coloring, branch lengths, branch support (or any other branch labels), and can aggregate heterogeneous data (annotations, expression profiles, and so on).

The installation is very simple as it mainly consists copying one file on an apache web server. Configuration of options can be done incrementally, but could include some Javascript/PHP modifications. InTreeGreat can display trees of more than 20,000 leaves while remaining very responsive.
Tree pattern matching

Tree pattern matching consists in the definition of an evolutionary scenario (called tree pattern) that is then searched for in a phylogenetic tree collection. It can be used, for example, to retrieve orthologous candidates in large comparative genomic datasets. It allows sophisticated queries like: recently duplicated, lost at a defined point of the species history, etc.

The tree pattern matching algorithm initially available in the FamFetch software has been implemented as i) a Javascript/PHP user interface to edit patterns and explore results, and ii) a Java daemon, that can be installed in any Linux infrastructure, in order to manage the computational part using a client/server protocol between the computing cluster and the webservers.

Results can be displayed in the InTreeGreat interface, but also exported on different visualisation clients (like with ancestrum and genomicus). This tool has been tested on huge data collections (HOGENOM, Penel et al. 2009, several billions of sequences and more than 300,000 trees).

Availability:
Every files are available freely on GitHub without mandatory identification:
http://southgreenplatform.github.io/rap-green/
https://github.com/SouthGreenPlatform/rap-green

Graphical interfaces can already be tested here:
http://phylogeny.southgreen.fr/treepattern/

The whole Java documentation is available here:
Fig. 1. On the top, the tree pattern edition interface, which allows to define an evolutionary scenario and search it on a phylogenetic tree collection. On the bottom, the InTreeGreat interface, that can display phylogenetic trees, founded tree pattern (in dotted lines) and heterogeneous related data.
