

BioHackathon 2018 - Paris http://bh2018paris.info/



Bioconda packaging of the Regulatory Sequence Analysis Tools (RSAT)



Proponent: Jacques van Helden

- Institut Français de Bioinformatique (IFB), France
- Aix-Marseille Université (AMU), France





Project links

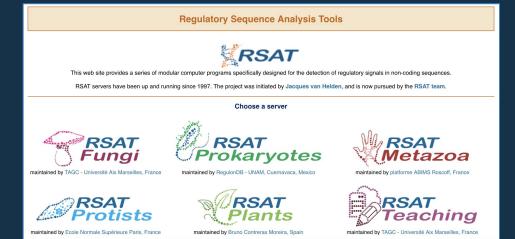
- RSAT: http://rsat.eu/
- Biohackathon project: https://goo.gl/5VhtEK



Background information

Regulatory Sequence Analysis Tools (RSAT)

- A software suite for the analysis of regulatory motifs in nucleic sequences.
- A 20-year project, supported by an international team
- Access: http://rsat.eu/
- Access and usage
 - 6 Web servers: Metazoa, Plants, Fungi,
 Prokaryotes, Protists + Teaching (all taxa)
 - 51 Web-interfaced tools
 - ~10,000 queries / month (2017)
 - Web services (SOAP/WSDL)
 - ~8,000 queries / month (2017)
 - Downloadable stand-alone version
 - 234 downloads in 2017
 - Virtual machine on the cloud of the Institut Français de Bioinformatique (IFB)
- Downloadable from http://teaching.rsat.eu/
 - Tar archive with the tools
 - Virtual machine (ova, 10Gb)



Nguyen NTT, Contreras-Moreira B, Castro-Mondragon JA, Santana-Garcia W, Ossio R, Robles-Espinoza CD, Bahin M, Collombet S, Vincens P, Thieffry D, van Helden J, Medina-Rivera A, Thomas-Chollier M. *RSAT 2018: regulatory sequence analysis tools 20th anniversary*. Nucleic Acids Res. 2018 Jul 2;46(W1):W209-W214. doi:10.1093/nar/gky317. PMID: 29722874

RSAT logos designed by Mauricio Guzman (http://www.altamirastudio.com/)



Difficulties for RSAT installation

- Multiple languages
 - ~250 perl scripts + ~40 python scripts + 7 C/C++ compile tools
- Multiple library dependencies
 - o Perl, python, R libraries
 - Installing these libraries requires some admin intervention (must be available to Apache user)
- Package segmentation
 - The current tarball includes tools + data (motif databases + demo datasets) -> a bit heavy
- Automatisation of the installation
 - Installation is managed via bash, make and some manual steps
- Portability
 - Some dependencies are hard to install on some platforms (e.g. Mac OS X)
 - The names of the packages required for dependencies vary between package managers (apt-get, yum, brew, ...).
 - Install scripts are based on apt-get, and package names are different in other package managers



Goals of the hacking project

https://goo.gl/gkYYNc

Goal and expected outcome

- General goal of the hacking project
 - Facilitate RSAT diffusion by enabling its installation via conda
- Expected results at the end of the hackathon
 - A git repository with the conda packaging scripts and metadata.
 - Be able to install the components of RSAT with simple conda commands, in a first time in a custom channel
 - conda install -c rsat rsat-core ## Install the core tools for command-line use
 - conda install -c rsat rsat-data ## Install RSAT data (motif databases)
 - conda install -c rsat rsat-web ## Install a local instance of the RSAT Web server
 - After testing, the packages will be submitted to bioconda (during or after biohackathon).

0

Post-biohackathon perspectives

- RSAT deployment on IFB National Network of Computing Resources (NNCR)
 - New appliance on the federation of clouds, via Biosphere (https://biosphere.france-bioinformatique.fr/)
 - Federation of clusters, for command-line access
 - o Transfer some persistent Web servers (RSAT Fungi, Teaching) on IFB-core-cluster
- Create Docker/singularity containers
 - Will be greatly facilitated by conda packaging



Hack organisation

Organisation of the hacking project

- Duration: 2 hacking days (Tue 13 + Wed 14 Nov)
- Project representation at the biohackathon
 - <u>Jacques van Helden</u> head and main developer of the RSAT project
- Invited resource people and associated expertise
 - Loic Gouarin conda packaging
 - Aurélien Naldi conda packaging + bioinformatics analysis of genetic regulation

- Any additional expertise and contribution from biohackers is welcome!
 - Register via the biohackathon gitter



Steps and tasks

- Presentation of RSAT structure to conda expert.
- Identification of existing conda resources for the dependencies (Perl, python, R libraries).
 - Packaging of the other dependencies
- Evaluating if some reorganisation of the code is required for conda porting.
- Conda packaging in custom channel (conda -c rsat)
 - the core tools (stand-alone Unix commands).
 - RSAT data (motif databases, demo datasets, ...).
 - Installation of RSAT Web server and Web services (requires prior installation of an Apache server).
- Testing of the conda packages on different OS
 - Linux distributions via VMs on the IFB cloud
 - Mac OS X
- Submission of the conda packages to bioconda



Contact and links

- Contact
 - o <u>Jacques.van-Helden@univ-amu.fr</u>
- Links related to the project
 - o RSAT Portal: http://rsat.eu/
 - Biohackathon project: https://goo.gl/5VhtEK

