

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)

Regulatory Sequence Analysis Tools




This web site provides a series of modular computer programs specifically designed for the detection of regulatory signals in non-coding sequences.
RSAT servers have been up and running since 1997. The project was initiated by Jacques van Helden, and is now pursued by the RSAT team.


Choose a server




maintained by TAGC - Université Aix Marseilles, France




maintained by RegulonDB - UNAM, Cuernavaca, Mexico




maintained by plateforme ABIMS Roscoff, France



maintained by Ecole Normale Supérieure Paris, France



maintained by Bruno Contreras Moreira, Spain



maintained by TAGC - Université Aix Marseilles, France

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

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

Difficulties for RSAT installation

- Multiple languages
 - ~250 `perl` scripts + ~40 `python` scripts + 7 `C/C++` compile tools
- Multiple library dependencies
 - `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 (125Mb)
- 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: be able to install the components of RSAT with simple conda commands
 - `conda install -c bioconda rsat-core` ## Install the core tools for command-line use
 - `conda install -c bioconda rsat-data` ## Install RSAT data (motif databases)
 - `conda install -c bioconda rsat-web` ## Install a local instance of the RSAT Web server

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
 - 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
 - [Jacques van Helden](#) 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
- **Call for additional expertise from biohackathon attendees**
 - **Loic & Aurélien, do you think we should call for complementary expertise?**

Steps and tasks

- Presentation of the structure to conda expert
- Identification of existing conda resources for the dependencies (Perl, python, R libraries)
- Evaluation if some reorganisation of the code is required for conda porting
- Porting the core tools (stand-alone Unix commands)
- Porting of the accompanying data (motif databases, demo datasets, ...)
- Porting of the Web server

Links and contact

Contact: Jacques.van-Helden@univ-amu.fr

Links

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