A journey of a team of engineers in learning packaging technology

Valentin Marcon[1*], Laure Quintric[2], Durand Patrick[2], Olivier Inizan[1], Caroline Dussart[2], Valentin Loux[1], Maria Bernard[3], Géraldine Pascal[4]

1: Mathématique et Informatique Appliquées du Génome à l'Environnement (MaIAGE) – Institut National de la Recherche Agronomique (INRA) Jouy-en-Josas; 2: Service Ressources Informatiques et Communications (IMN/IDM/RIC) – Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER); 3: Système d'Information des GENomes des Animaux d'Elevage (SIGENAE) – Institut National de la Recherche Agronomique (INRA); 4: GENétique, PHYsiologie et Systèmes d'Elevage (GENPHYSE) – Institut National de la Recherche Agronomique (INRA)

PROBLEMATIC

Bioinformatics pipelines installation can take time for engineers and administrators because of **the amount of dependencies required**.
Facing this problematic, we, a team of four engineers responsible of tools installation where wondering **how to facilitate such complex installation for everyone for a pipeline on Galaxy**.

STUDY CASE, FROGS: 'Find Rapidly Otu with Galaxy Solution'

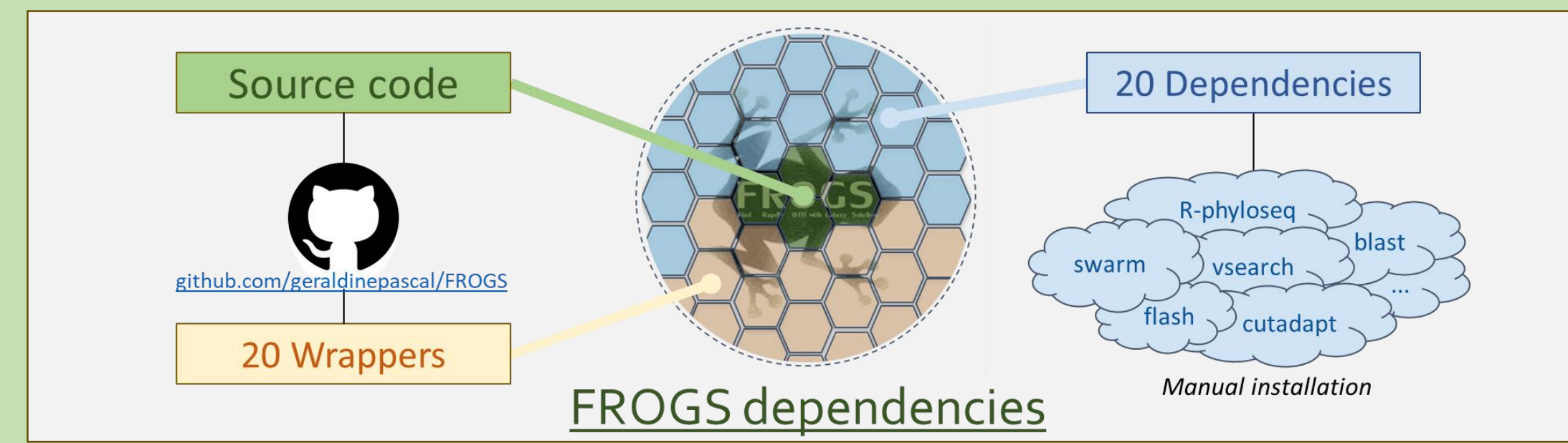### THE PIPELINE

This **metabarcoding** analysis pipeline is designed to give the **abundance table and the taxonomic affiliation of operational taxonomic units (OTUs)**. It's used as a **Galaxy workflow**, focused on user-friendliness, so it does not require bioinformatics or command lines skills.
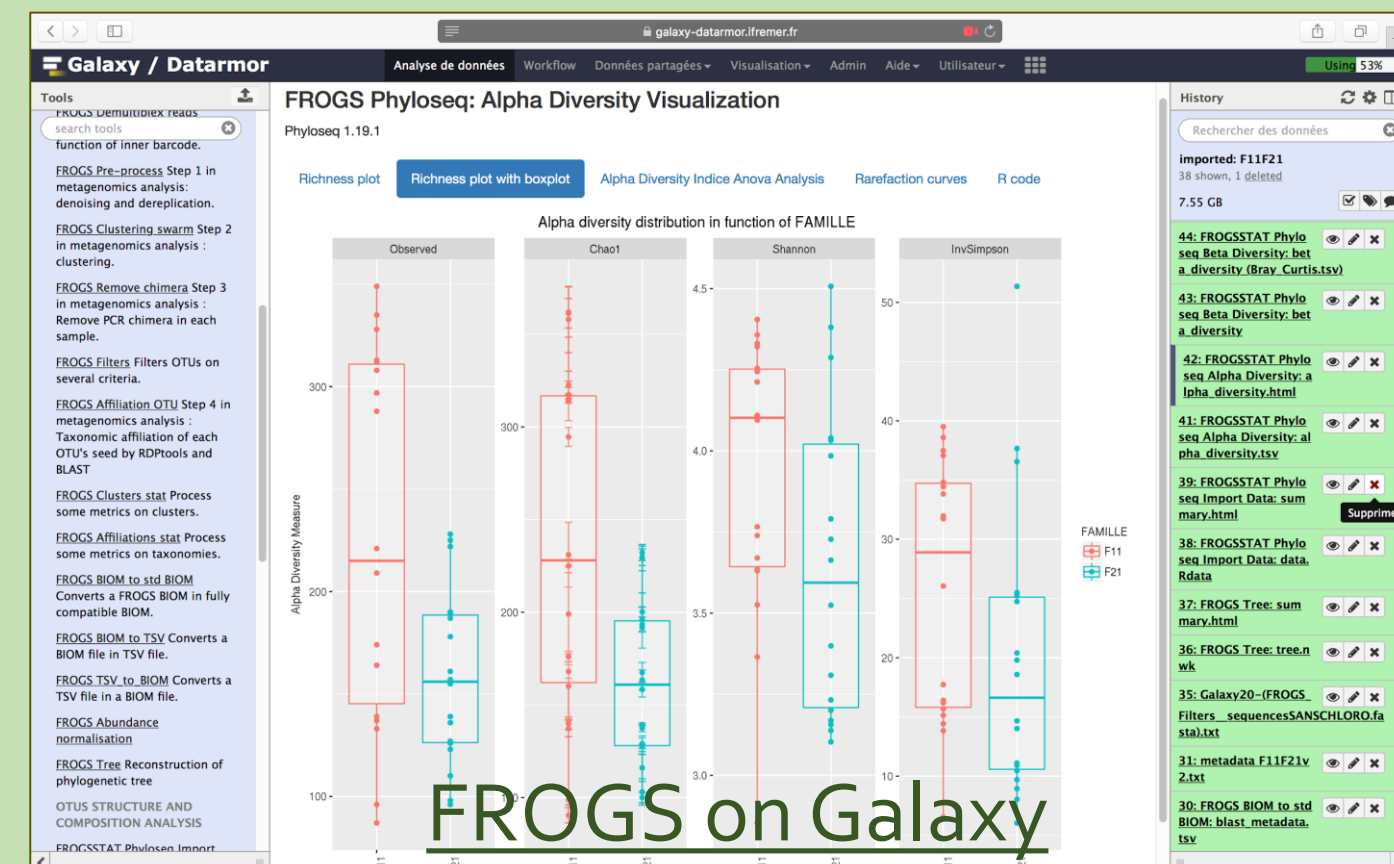
FROGS on Galaxy

WHY FROGS?

20 training
200 people trained
17 citations
~400 visit/month
frogs.toulouse.inra.fr

Source code — github.com/geraldinepascal/FROGS — 20 Wrappers
20 Dependencies — R-phyloseq, swarm, blast, vsearch, flash, cutadapt — Manual installation
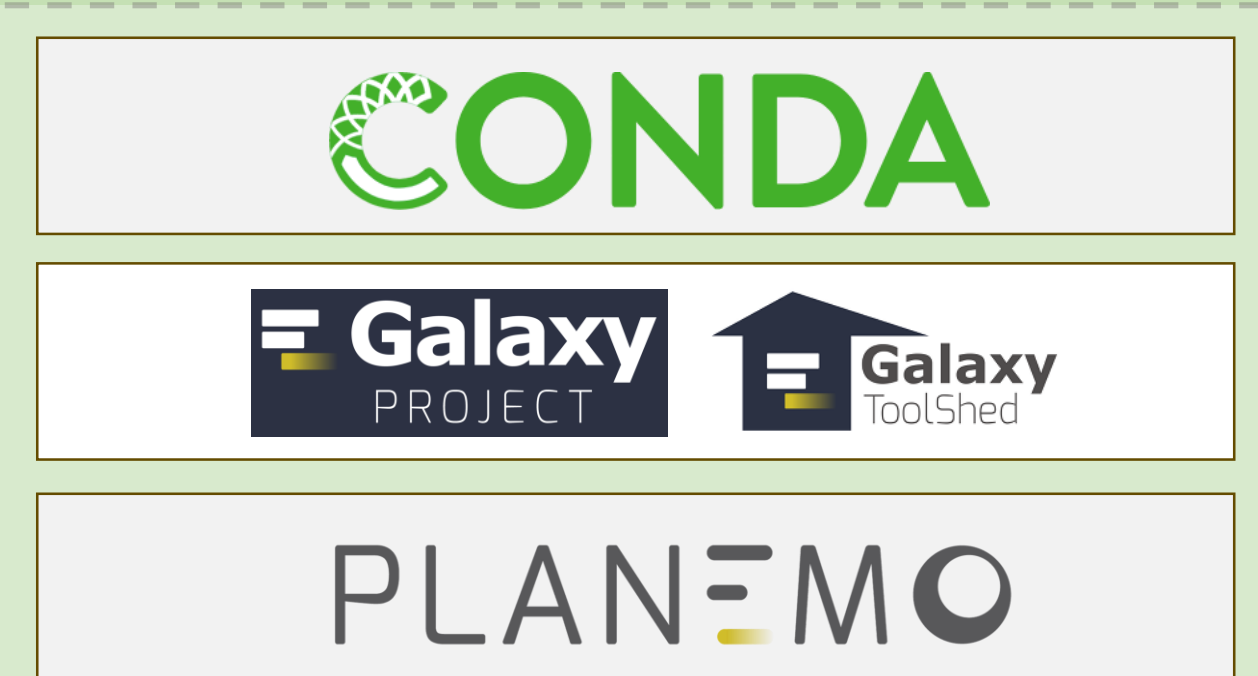FROGS dependencies

DEPENDENCIES

FROGS pipeline is **a good case of study** because it contains about **20 wrappers** (tools in Galaxy) and needs about **20 dependencies** to work. **Wrappers** and source code are host on the Github page of developers and to be **installed manually**. Dependencies needs to be found in various websites and to be **installed manually too**.

GOAL

FROGS is **accessible for every users**, the ones familiar with the command line and the others that are not, thanks to the Galaxy interface. It is way **much difficult for the administrators** because they have to install all wrappers, sources and dependencies manually.
So, our objective was to **improve the overall installation process of the FROGS pipeline**.

Cmd Line >_ | GUI Galaxy PROJECT | Dependencies & wrappers — Manual installation
Accessibility for Users ✓ | FROGS accessibility | Accessibility for Admins to improve (people in charge of tools installation)

TECHNOLOGIES

The key issue was that **we did not have a good knowledge in packaging technologies**. But we don't made a huge work of research, we just **followed the advices of the Galaxy Community**.
The packaging technology provided by Galaxy is **Conda**, which allow you to fully make a tool ready to use with only one command (as "apt-get install") and to manage virtual environments. Then, we wanted to publish the **wrappers (Galaxy integrated tools) on the Toolshed** (the galaxy tools catalogue). For that purpose we used **Planemo**v(a tool that help people to create, validate and test their own wrapper) to ensure that the wrappers made by the FROGS developers team where following the **best practices** of designing wrappers.
Finally, we wanted to figure out whether or not this technology will be **easy to learn** and if we will be able to package FROGS in **a reasonable amount of time**.

CONDA
Galaxy PROJECT / Galaxy ToolShed
PLANEMO

THE JOURNEY

01 Follow the Galaxy training network for conda recipe	02 Check availability of FROGS dependencies as Conda packages
03 Create a conda env. with FROGS dependencies	04 Install FROGS manually and run unit tests
05 Write FROGS conda recipe	06 Run FROGS unit tests
07 Upload FROGS Conda package on a personal Anaconda Cloud account	08 Push the package on the channel « Bioconda »
09 Improve wrappers with Good Practices and Planemo	10 Publish FROGS wrappers on the test Toolshed
11 Deploy in a Galaxy instance for a 3 days training	… To do: Publish on the toolshed

We first focused on creating the **Conda package** for the FROGS pipeline: **(1)** Obviously, we followed the **documentation** of the Galaxy training network (=read the manual). **(2)** We searched the availability for all **dependencies as a Conda package**. **(3)** We created a **virtual environment** with Conda, installing all found dependencies. **(4)** We installed the source code of the FROGS pipeline to be sure that all the **dependencies worked well together**. We add some problem to manage version conflicts but finally whe decides do fix the major and minor versions of each. **(5)** When all the dependencies conflict where resolved we designed the FROGS Conda package **writing the installation recipe. (6)** We have run FROGS **unit test** to ensure the good functionnemnt of the package. **(7)** We **upload the package** on a personal Conda chanel to test its installation in private. **(8)** When the all was ok we **push the recipe on the "Bioconda" channel** (github). After that, we focused on the Galaxy wrappers: **(9)** We **improved** the Galaxy wrappers made by the developers team of FROGS in order to follow the **Best Practices** provided by the Galaxy community.
(10) We uploaded the galaxy package on **the toolshed test** to make sure it will install FROGS wrappers correctly. **(11)** We deployed the wrappers on a **new Galaxy instance**, from the testtoolshed, and we achieved a **3 day training** on this one. **(…)** The final step will be to **publish** the FROGS wrappers on **the main Toolshed** after fixing the remaining issues.

CONCLUSION

Today FROGS pipeline is available as a **Conda package** through **the Bioconda channel**, ready to use on the command line (https://anaconda.org/bioconda/frogs). The **Galaxy wrappers** are accessible from the **TestToolshed** in order to install it quickly from your Galaxy administrator interface (https://testtoolshed.g2.bx.psu.edu/view/oinizan/frogs_2_0_0). Some issues need to be fixed, and we will publish it soon on the Toolshed. Finally we want to insist on particular advices according to our personal experience. First for **the tools packaging** : Provide software tools as Conda packages;Do not let Conda install default dependencies;Be careful with Conda channels order;Provide separate Conda packages fr complex softwares. Then for the **tools development**: Provide softwares with tests;Use a logging framework instead of "/dev/null" redirection.

REFERENCES

1. FROGS: Find, Rapidly, OTUs with Galaxy Solution. [Paper in "Bioinformatics"; doi: 10.1093/bioinformatics/btx791] *Escudié F, Auer L, Bernard M, Mariadassou M, Cauquil L, Vidal K, Maman S, Hernandez-Raquet G, Combes S, Pascal G.*
2. The journey of a team of engineers in learning packaging technology. [Talk on the "Galaxy Community Conference and Bioinformatics Open Source Conference"; web: http://sched.co/EYCM] *Quintric L, Durand P, Inizan O, Marcon V, Dussart C, Loux V, Bernard M, Pascal G.*

CONTACTS

pgdurand@ifremer.fr
laure.quintric@ifremer.fr
olivier.inizan@inra.fr
valentin.marcon@inra.fr