

Role of Galaxy in a bioinformatic plant breeding platform



Vincent Maillol¹, Roberto Bacilieri¹, Stéphanie Sidibe Bocs³, Jean-Michel Boursiquot^{1,2}, Grégory Carrier², Alexis Dereeper⁴, Gaétan Droc³, Cécile Fleury³, Pierre Larmande⁴, Loïc Lecunff², Jean-Pierre Péros¹ Bertrand Pitollat³, Manuel Ruiz³, Gautier Sarah¹, Guilhem Sempéré³, Marilyne Summo³, Patrice This¹, and **Jean-Francois Dufayard**³

- ¹ INRA – Montpellier SupAgro, UMR 1334 AGAP, DAVEM team, 2 Place P. Viala, 34060 Montpellier, France.
- ² Institut Français de la Vigne et du Vin - Unité Mixte Technologique Géno-Vigne, 2 Place P. Viala, 34060 Montpellier, France.
- ³ CIRAD - UMR 1334 AGAP, ID team, avenue Agropolis, 34398 Montpellier Cedex 5, France.
- ⁴ IRD - 911 avenue Agropolis, 34394 Montpellier, France.

[Slides](#), [Video](#)

With NGS development, bioinformatics has become central in plant breeding laboratories, and researchers are in need of some autonomy in its use. The Southgreen platform (CIRAD, IRD, INRA) performs bioinformatics analyses for many plant breeding research teams in Montpellier (France), and offers many systems to users: for example GNPAnnot (automatic genomic sequence annotation), Greenphyl (phylogenetic orthology prediction), ESTtik (EST annotation) or the Bacchus analysis pipeline. Most of these systems have been translated into Galaxy workflows.

As for the Bacchus pipeline, it has been created at INRA Montpellier (France) to investigate clonal diversity in grapevine genomes. For this task, many softwares have been wrapped in the Galaxy framework. Bacchus can be decomposed in three steps: i) Genome reconstruction, ii) test of reconstruction results, and iii) diversity analysis. This last step is done using SNP's and structural variations. To detect SNP's, the latest Freebayes version was used, while the IDfixe software was developed for structural variation detection. Some of the softwares developed for this pipeline are now used in the international project Grapereseq.

Today, the Galaxy framework is widely used by Southgreen platform users as an alternative to the command line system. In this context, dozens of users have already been trained in Galaxy-using bioinformatics. During weekly collective pair-programming sessions, platform engineers and interested scientists integrate new tools and functionalities. Thus, Galaxy is now a core component of the plant breeding community around the Southgreen platform, and the main access portal for non- bioinformatics specialists to our computing clusters.