Galaxy and South Green
Galaxy for tropical and mediterranean plants
What’s new since Chicago?

Alexis Dereeper (IRD, UMR IPME)
Jean-François Dufayard (CIRAD, UMR AGAP)
Galaxy instance for plant genomics/genetics

25 bioinformaticians
Galaxy instance for plant genomics/genetics

400+ tropical and mediterranean (from the south) plant experts

25 bioinformaticians
425+ scientists proficient in bioinformatic, thanks to Galaxy
Galaxy role in South Green platform

Public main instance, in production
http://galaxy.southgreen.fr/galaxy/

Private secondary instance, in production

Dev instance (intranet, private)

New functionalities

Agilathon/Galaxython: developer working sessions
Galaxy, some statistics

400+ registered users

100+ original tools

30 shared workflows

~36,000 jobs / year

~53,000 CPU hour / year
Galaxy as a training platform

Data sharing facilities between trainees

<table>
<thead>
<tr>
<th>Data library name</th>
<th>Data library description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formation</td>
<td>Jeux de données utilisés dans les formations Southgreen</td>
</tr>
<tr>
<td>Formation Galaxy</td>
<td>Jeux de données simples pour la formation à Galaxy</td>
</tr>
<tr>
<td>Formation ISS</td>
<td>Fichiers utilisés à la formation phylogénie sous Galaxy</td>
</tr>
<tr>
<td>Formation phylogénie</td>
<td>Fichiers utilisés à la formation phylogénie sous Galaxy</td>
</tr>
<tr>
<td>Formation Structure 3D</td>
<td>Documents correspondant au TDS Structure 3D des Problèmes</td>
</tr>
<tr>
<td>Formation brepli</td>
<td>Données pour la formation Galaxy de février 2015</td>
</tr>
</tbody>
</table>

Workflows we hope “intuitive”


Montpellier (2011 to 2017)

To the north

To the south
- Dedicated to Galaxy, federated by French Bioinformatic Institute
- Federation of the French community around Galaxy

Many platforms involved

Many missions:
- Sharing
- Training
- Project initiation
- ...

IFB Galaxy Working Group
8 highlighted workflows

Effort to spotlight our main workflows, relevant for our thematics:
=> 8 preconfigured and validated workflows:
(comprehensive carousel images, facilitated access, documentation…)

Welcome to GALAXY

Our pre-configured and validated workflows

How to load big datasets?

1. My file
2. HPC Cluster
3. My Galaxy History

These workflows as part of the services provided by South Green

Chromosome reconstruction

Scaffhunter tools assemble scaffolds into pseudomolecules using markers genotyped in a population (Martín et al., 2016)

Input: Fastq + FASTA

Access workflow

In order to figure out which tools were made available by our team, please activate the "tool search" functionality from the Options drop-down and type "south green" in the lookup filter.
Highlight 1: Phylogeny and Gene families

Build gene families:
1) by full proteome clustering
2) from sequences of interest, searching for homologs

Compute phylogenetic trees

Integrate data around gene families (syntenies, proteic domains, gene annotations, gene expression…)

Transfer annotation by comparative analyses

Visualize results
Highlight 2: Structural variations

**Structural variations**

Scaffremodeler can be used to detect large structural variations between a reference sequence and a resequenced genome (Martin et al, 2016)

**Input:** Fastq + FASTA

Access workflow

---

*No overlap of discordant regions*
- Consistent with SV hypothesis

*No concordant reads linking SV boundaries*
- Consistent with homozygous SV

Discordant reads
- (reverse-reverse orientation)

Concordant reads
- (forward-forward orientation)
Highlight 3: SNP analysis and GWAS workflow

**SNP analysis**

- **SNiPlay3** complete workflow: a package for exploration and large scale analyses of SNP polymorphisms (filtering, SNP density, diversity, linkage disequilibrium) (Dereeper et al., 2015)
  - Input: VCF

**GWAS**

- **SNiPlay3** GWAS workflow (GLM model) including population structure and correction for structure (Dereeper et al., 2015)
  - Input: VCF + Phenotypic tabulated file

---

**Workflow SNiPlay Galaxy**

1. VCF → Ref → GFF
   - snpEff
   - snpSift
   - VCFTools Filter
   - Annotation FromVCF
2. VCFTools Stat → SlidingWindows → Beagle
   - GetHaplotypes FromVCF → PedToBed → SNMF → MDSPlot
   - Network
3. PedToFasta → ReadSeq → FastME → SNPdensity
   - Diversity ByGene
4. VCF2HapMap

---

Winter Asia
Winter Europe
Winter America

G. Andres
A. Dereeper
8 preconfigured and validated workflows

NGS analyses
We propose several workflows for NGS analyses in different scenarios (transcriptome vs transcriptome, transcriptome vs genome...). It includes cleaning and mapping steps using commonly used softwares.

- **Input**: Fastq files + FASTA for reference
- **Output**: BAI alignment files
- **Access workflow**

SNP calling
The SNP Calling is based on the GATK toolkit, using either UnifiedGenotyper or HaplotypeCaller module.

- **Input**: BAI alignment files + FASTA for reference
- **Output**: VCF (Variant call Format) file
- **Access workflow**

SNP analysis
SNIPay3 complete workflow: a package for exploration and large scale analyses of SNP polymorphisms (filtering, SNP density, diversity, linkededisequilibrium) (Dereeper et al., 2015)

- **Input**: VCF
- **Access workflow**

GWAS
SNIPay3 GWAS workflow: Tassel-based GWAS workflow (GLM model) including population structure and correction for structure (Dereeper et al., 2015)

- **Input**: VCF + Phenotypic tabulated file
- **Access workflow**
8 preconfigured and validated workflows

**Chromosome reconstruction**

*Scaffrehunter* tools assemble scaffolds into pseudomolecules using markers genotyped in a population (Martin et al, 2010)

*Input:* Fastq + FASTA

Access workflow

G. Martin

G. Droc

**Metagenomics**

*FROGS:* Find Rapidly OTU with Galaxy Solution (Pascal et al, 2015)

*Input:* Fastq files

Access workflow

GenPhySE

*Génétique Physiologie et Systèmes d’Élevage*
Documentation on our workflows

Galaxy pages including:
- workflow(s)
- datasets
- history of analysis
- documentation
Efforts for development of visualization

As Galaxy visualization plugin (Highcharts)
Efforts for development of visualization

As Galaxy visualization plugin (Highcharts)

As embedded HTML/PHP viewer (CircosJS)
Efforts for development of visualization

As Galaxy visualization plugin (Highcharts)

As embedded HTML/PHP viewer (CircosJS)

By sending Galaxy datasets to external viewer (Jbrowse, CytoscapeJS)
Efforts for development of visualization

As Galaxy visualization plugin (Highcharts)

As embedded HTML/PHP viewer (CircosJS)

By sending Galaxy datasets to external viewer (Jbrowse, CytoscapeJS)
Sharing our workflows

- Some of the workflows Galaxy are available through the main Tool Shed
- Tests with Planemo
- Our wrappers are available through GitHub South Green

Perspectives:
- Management of dependencies using Conda
- Towards Docker for containerizing our Galaxy instance
Thank you for your attention!!!! Questions ???