

User feedback of a Galaxy workflow conception

Claire Toffano-Nioche

eBio facility – Université Paris-Sud, Orsay

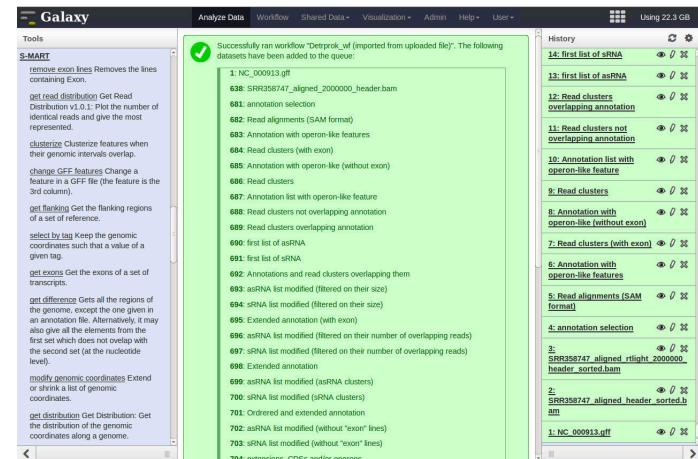
IGM - Institut de Génétique et Microbiologie

How have I built the Workflow?

step-by-step, biologist
Data = README file
s-smart tools, home-made scripts

```
### 1- select reads outside "gene" :
# select reads (-i VIBSP_chrl_smart.gff3) outside
# "annotation" (-j NC_011744_annot.gff3 -c -x), with
# keeping reads overlapping ATG (-e 10)
for i in "NC_011744" "NC_011753" ; do
    compareOverlapping.py -i ${i}_RNAseq_smart.gff3 -f gff3
    clusterize.py -i ${i}_cis_ReadsOutGene.gff3 -f gff3 -c -d 20 -o
    ${i}_cis_e10_clusters.gff3
done ;
### 2- clusterize reads before seed selection :
# 2 clusterization steps: strict + d < 20
for i in "NC_011744" "NC_011753" ; do
    clusterize.py -i ${i}_cis_ReadsOutGene.gff3 -f gff3 -c -o $
    {i}_cis_e10_clusters.gff3 ;
    clusterize.py -i ${i}_cis_e10_clusters.gff3 -f gff3 -c -d 20 -o
    ${i}_cis_e10_clusters20.gff3 ;
### 3- crossing cluster with seed :
for i in "NC_011744" "NC_011753" ; do
    seedGff.pl -i ${i}_annot.gff -p 25 -e 15 -o ${i}_cis_seed.gff ; #
    seed creation in -15 -25 bp from ATG
    compareOverlapping.py -i ${i}_cis_e10_clusters20.gff3 -f gff3 -j $
    {i}_cis_seed.gff -g gff3 -c -o ${i}_cis_e10_cluster20InSeed2515.gff3 ; #
    crossing cluster with seed
done ;
#
### 4- filters: nb_reads > 10 reads + length > 50 nt
for i in "NC_011744" "NC_011753" ; do
    selectByTag.py -i ${i}_cis_e10_cluster20InSeed2515.gff3 -f gff -o
    ${i}_cis_e10_cluster20InSeed2515_nbEUp10.gff3 -m "nbElements" -m
    10 -d 0 ; # nb_reads > 10 filter
done ;
...
```

Wrap tool, share (test/toolshed)
Import on my local Galaxy instance
Manual launch, follow the readme file



The screenshot shows the Galaxy web interface. The main panel displays a workflow with 14 steps, including '1: NC_000913.gff', '681: annotation selection', '682: Read alignments (SAM format)', '683: Annotation with operon-like features', '684: Read clusters (with exon)', '685: Annotation with operon-like (without exon)', '686: Read clusters', '687: Annotation list with operon-like feature', '688: Read clusters not overlapping annotation', '689: Read clusters overlapping annotation', '690: first list of asRNA', '691: first list of sRNA', '692: Annotations and read clusters overlapping them', '693: asRNA list modified (filtered on their size)', '694: sRNA list modified (filtered on their size)', '695: Extended annotation (with exon)', '696: asRNA list modified (filtered on their number of overlapping reads)', '697: sRNA list modified (filtered on their number of overlapping reads)', '698: Extended annotation', '699: asRNA list modified (asRNA clusters)', '700: sRNA list modified (sRNA clusters)', '701: Ordered and extended annotation', '702: asRNA list modified (without "exon" lines)', '703: sRNA list modified (without "exon" lines)', and '704: extensions: CDSs and/or operons'. The right panel shows the 'History' section with a list of previous workflow runs, including '14: first list of sRNA', '13: first list of asRNA', '12: Read clusters overlapping annotation', '11: Read clusters not overlapping annotation', '10: Annotation list with operon-like feature', '9: Read clusters', '8: Annotation with operon-like (without exon)', '7: Read clusters (with exon)', '6: Annotation with operon-like features', '5: Read alignments (SAM format)', '4: annotation selection', '3: SRR35747_aligned_r1.fastq.header.sorted.bam', '2: SRR35747_aligned_header.sorted.bam', and '1: NC_000913.gff'.

Extract Workflow from History
Edit workflow: close or open parameter
(« set at run time »), add comments,
change name of the steps
Share Workflow (test/toolshed)

Technical issues

- Use « virtual_env »
- Dissociate « your » directories from the Galaxy distribution (configure the universe_wsgi.ini file)
 - galaxy-dist
 - galaxy_env
 - repository_dependencies
 - toolshed
 - upload_libraries
- « underlying » database :
 - SQLite => postgres
 - allow concurrent accesses, count DB in the resources needed
- Install a Galaxy instance on a stratuslab virtual machine during revision process of the publication



Feedbacks

User's point of view: local instance vs. Galaxy server ?

User : import workflows, not the tools (admin account).

Ok in a local instance, contact the admin in a server => not so easy !

Open questions (perhaps have you the answers?):

- How expliciting the need of specific tools? **dependencies?**
- « **Global** » **variables?** Launching a tool many times implies to re-enter the same values for open parameters. Environment variables?
- **Meta-workflow?** Begin = bam file (user chooses the mapper). How add the mapping step? another workflow and combine the two?



Methods

Volume 63, Issue 1, 1 September 2013, Pages 60–65

Diversity of the non-coding transcriptomes revealed by RNA-seq technologies



Detection of non-coding RNA in bacteria and archaea using the DETR'PROK Galaxy pipeline ☆

Claire Toffano-Nioche^a, Yufei Luo^b, Claire Kuchly^a, Claire Wallon^a, Delphine Steinbach^b, Matthias Zytynicki^b, Annick Jacq^a, Daniel Gautheret^a  

^a Université Paris-Sud, Institut de Génétique et Microbiologie, CNRS UMR 8621, Orsay F-91405, France

^b URGI, INRA, Versailles F-78026, France



[Open Access](#)

2279 valid tools on Oct 24, 2013

Search

- [Search for valid tools](#)
- [Search for workflows](#)

Valid Galaxy Utilities

- [Tools](#)
- [Custom datatypes](#)
- [Repository dependency definitions](#)
- [Tool dependency definitions](#)

All Repositories

- [Browse by category](#)

Available Actions

- [Login to create a repository](#)

Repositories with matching workflows

workflow name: **detrprok**
exact matches only: **False**

<input type="checkbox"/>	Repository name	Synopsis	Revision	Owner
<input type="checkbox"/>	detrprok_wf	ncRNA detection in prokaryote oriented RNAseq	f0ca0981eb86	clairetn
<input type="checkbox"/>	detrprok_wf	ncRNA detection in prokaryote oriented RNAseq	8f7eb127baf3	clairetn
<input type="checkbox"/>	detrprok_wf	ncRNA detection in prokaryote oriented RNAseq	bb71a378053a	clairetn

For 0 selected items:

