Flexible use of virtualization for Galaxy

Marius van den Beek
Drosophila Genetics and Epigenetics
UMR7622 – Jussieu
mvandenb@snv.jussieu.fr
drosophile.com

Galaxy Day 04/12/13

Who we are

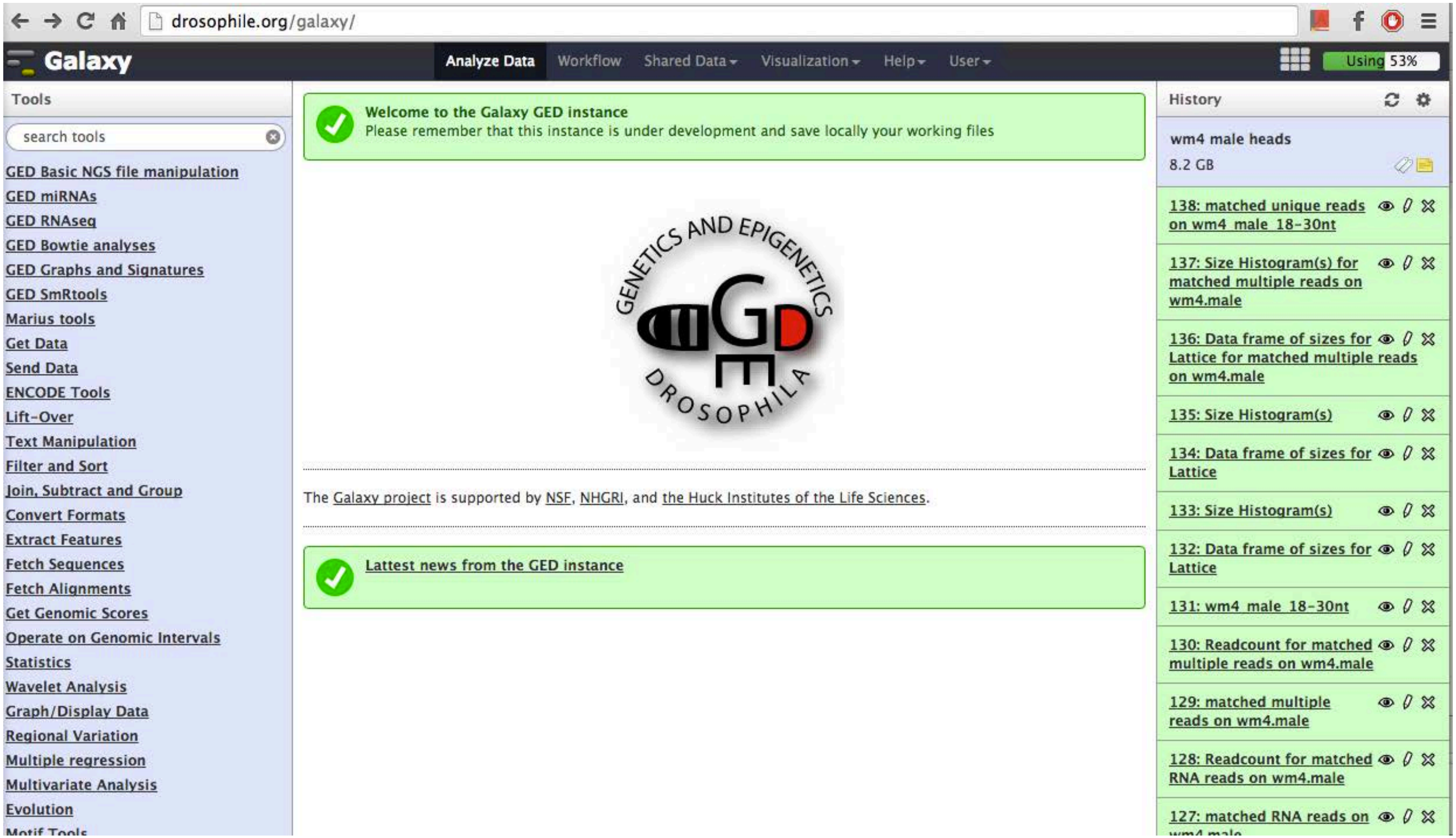
- Mostly genetics and mol. biology background
- Research focused on small RNA biology (miRNA, piRNA, siRNA)
- Over the years, a number of (command-line) workflows have been developed in the lab, for our own research or for collaborators

The need for a data analysis platform …

Lab publications

- AutomiG, a Biosensor to Detect Alterations in miRNA Biogenesis and in Small RNA Silencing Guided by Perfect Target Complementarity. Carré C., Jacquier C., Bougé A.-L, de Chaumont F., Besnard-Guerin C., Thomassin H., Pidoux J., Da Silva B., Chalatsi E., Zahra S., Olivo-Marin J.-C.,Munier-Lehmann H., Antoniewski C. **PLoS ONE. 2013. 8(9): e74296**
- Profiles of piRNA abundances at emerging or established piRNA loci are determined by local DNA sequences. *de Vanssay A, Bougé AL, Boivin A, Hermant C, Teysset L, Delmarre V, Ronsseray S, Antoniewski C.* **RNA Biol. 2013 Jul 15;10(8). [Epub ahead of print]**
- Lack of miRNA misregulation at early pathological stages in Drosophila neurodegenerative disease models. *Reinhardt A, Feuillette S, Cassar M, Callens C, Thomassin H, Birman S, Lecourtois M, Antoniewski C and Tricoire H.* **Front Genet. 2012; 3:226**
- Paramutation in Drosophila linked to emergence of a piRNA-producing locus. *de Vanssay, A., Bouge, A.L., Boivin, A., Hermant, C., Teysset, L., Delmarre, V., Antoniewski, C., and Ronsseray, S.* **Nature. 2012; 490:112-115**
- Convergent evolution of argonaute-2 slicer antagonism in two distinct insect RNA viruses. *van Mierlo, J.T., Bronkhorst, A.W., Overheul, G.J., Sadanandan, S.A., Ekstrom, J.O., Heestermans, M., Hultmark, D., Antoniewski, C., and van Rij, R.P.* **PLoS pathogens 2012. 8, e1002872**.
- Naive and primed murine pluripotent stem cells have distinct miRNA expression profiles. *Jouneau A, Ciaudo C, Sismeiro O, Brochard V, Jouneau L, Vandormael-Pournin, Coppe JY, Zhou Q, Heard E, Antoniewski C, Cohen-Tannoudji M.* **RNA. 2012.** Supplementary material.
- Visitor, an informatic pipeline for analysis of viral siRNA sequencing datasets. *Antoniewski C.* **Methods Mol Biol. 2011;721:123-42.**

Our galaxy server – alive since December '11…

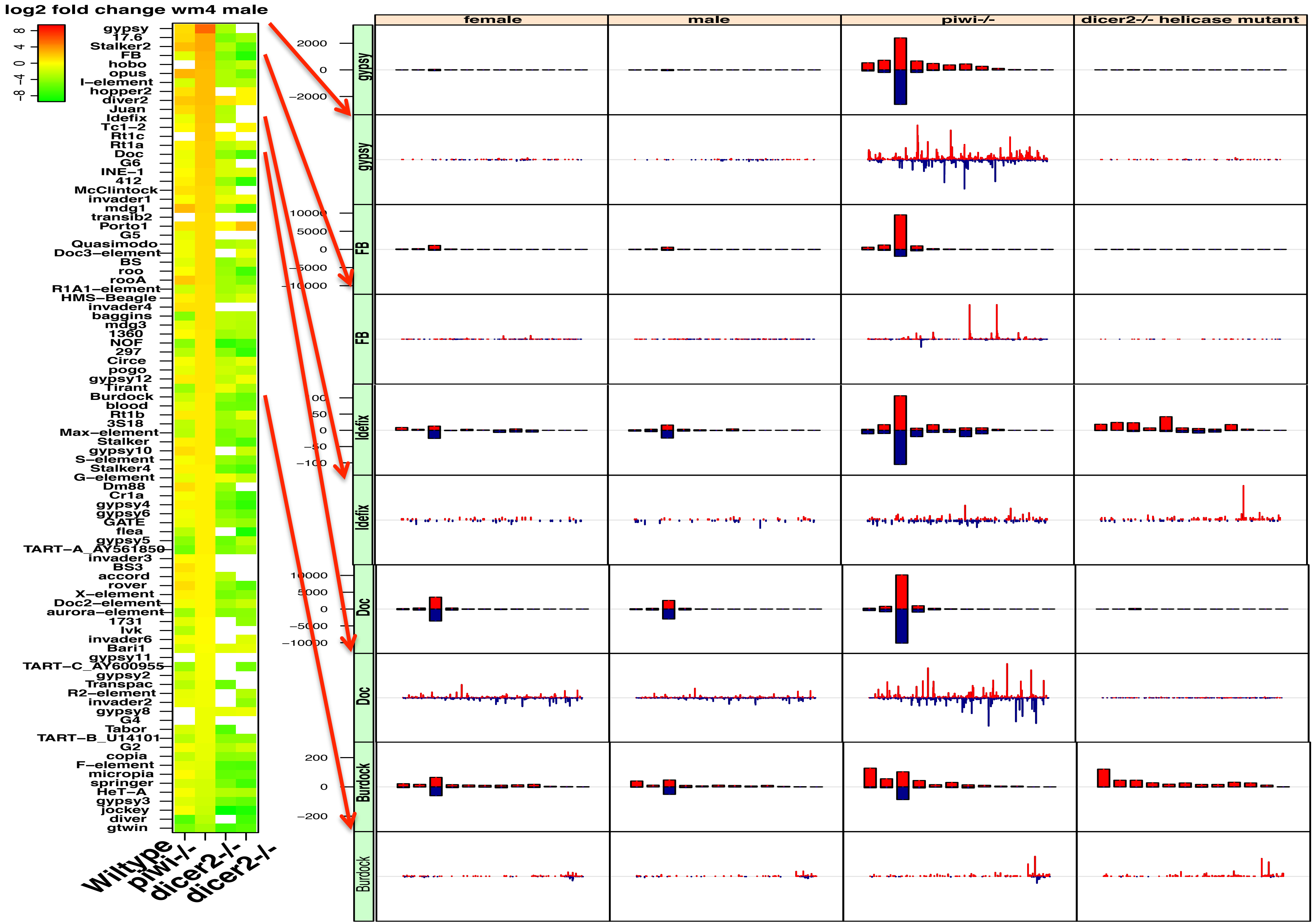


Our galaxy server – some stat's…

Dell PowerEdge T610
2x X5650 @ 2.67GHz, 12 cores total
32GB Ram
2x300 System / in RAID1, managed by LVM
6x4TB Storage in RAID6, 16TB net capacity, LVM
21 users, of which 6 regular users
~ 60 custom tools

mostly dedictated to **small RNAseq** and RNAseq analysis

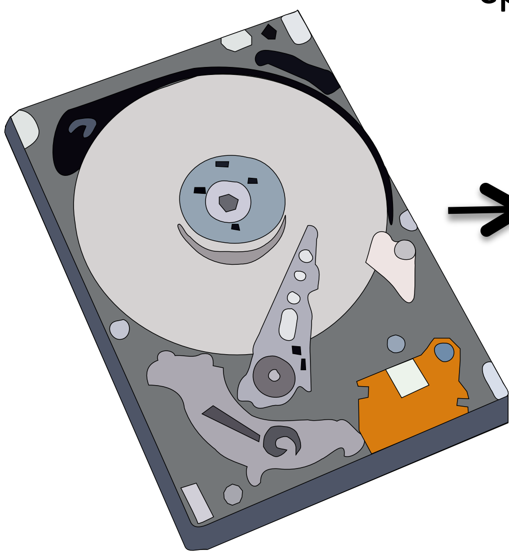
… and a second, more powerful one and a storage server are expected next week!



log2 fold change wm4 male
gypsy
17.6
Stalker2
FB
hobo
opus
I−element
hopper2
diver2
Juan
Idefix
Tc1−2
Rt1c
Rt1a
Doc
G6
INE−1
412
McClintock
invader1
mdg1
transib2
Porto1
G5
Quasimodo
Doc3−element
BS
roo
rooA
R1A1−element
HMS−Beagle
invader4
baggins
mdg3
1360
NOF
297
Circe
pogo
gypsy12
Tirant
Burdock
blood
Rt1b
3S18
Max−element
Stalker
gypsy10
S−element
Stalker4
G−element
Dm88
Cr1a
gypsy4
gypsy6
GATE
flea
gypsy5
TART−A_AY561850
invader3
BS3
accord
rover
X−element
Doc2−element
aurora−element
1731
Ivk
invader6
Bari1
gypsy11
TART−C_AY600955
gypsy2
Transpac
R2−element
invader2
gypsy8
G4
Tabor
TART−B_U14101
G2
copia
F−element
micropia
springer
HeT−A
gypsy3
jockey
diver
gtwin
Wiltype
piwi-/-
dicer2-/-
dicer2-/-
female
male
piwi-/-
dicer2-/- helicase mutant
gypsy
FB
Idefix
Doc
Burdock

A flexible backup strategy

Root backups are performed with make_snapshot.sh in an incremental fashion:



mv folder backup.0 to backup.1,
cp –al backup.1 to backup.0

↓

LVM snapshot,
mounted read only

↓

mount backup filesystem,
rsync into folder backup.0

↓

unmount & destroy snapshot,
remount backup file-system read only

```
Nov 30 16:00 daily.0
Nov 29 16:00 daily.1
Nov 28 16:01 daily.2
Nov 27 16:00 daily.3
Nov 26 16:00 daily.4
Nov 25 16:00 daily.5
Nov 24 16:00 daily.6
Dec  2 00:00 hourly.0
Dec  1 20:00 hourly.1
Dec  1 16:00 hourly.2
Dec  1 12:00 hourly.3
Nov  3 16:00 monthly.0
Sep 29 16:00 monthly.1
Sep  1 16:00 monthly.2
Nov 24 16:00 weekly.0
Nov 17 16:00 weekly.1
Nov 10 16:00 weekly.2
Nov  3 16:00 weekly.3
```

Virtualizing a backup …

We can now use any of the backups to create a bootable virtual hard disk! p2v.sh does all of the following:

Create a virtual disk, format with a filesystem, mount it, copy a root file system, set new hostname, deletes old udev rules, changes /etc/fstab and install grub.

The resulting image is bootable on OSX, Windows and Linux, as long as it is the 64 bit version!

Perfect start for beginners, students, collaborators, to test the latest upgrades and especially **to test new tools!**

Using multiple virtual machines to test cluster setup

To setup a cluster, all we need is a shared nfs folder for user files and a job_workflow directory.

Then install the opengrid batch-queing system, inform the galaxy job runner and we're good to go:

```
galaxy@vboxgalaxy:/var/storage$ qstat -f
queuename                      qtype resv/used/tot. load_avg arch          states
---------------------------------------------------------------------------------
main.q@galaxy-exec2            BIP   0/1/1          0.23     lx26-amd64
    252 0.25000 g34786_GED galaxy       r     12/02/2013 00:57:37     1
---------------------------------------------------------------------------------
main.q@galaxy-exec3            BIP   0/1/1          -NA-     lx26-amd64    au
    166 0.75000 g34700_GED galaxy       dr    11/30/2013 16:22:48     1
---------------------------------------------------------------------------------
main.q@galaxy-exec4            BIP   0/1/1          0.01     lx26-amd64
    251 0.25001 g34785_GED galaxy       r     12/02/2013 00:57:37     1
---------------------------------------------------------------------------------
main.q@vboxgalaxy              BIP   0/1/1          0.45     lx26-amd64
    253 0.25000 g34787_GED galaxy       r     12/02/2013 00:57:37     1
---------------------------------------------------------------------------------
main.q@virtualged              BIP   0/1/1          0.43     lx26-amd64
    250 0.25001 g34784_GED galaxy       r     12/02/2013 00:57:37     1
galaxy@vboxgalaxy:/var/storage$
```

Running a bowtie Workflow on the server vs. 4 iMacs …

gedserver	galaxy-exec2	VirtualGED	vboxgalaxy	galaxy-exec4
20:11:00,335	19:45:07	19:45:07	19:45:07	19:45:07
20:18:02,161	19:57:00	19:56:23	19:57:35	19:56:12
~7 minutes	~11 minutes	~11 minutes	~11 minutes	~11 minutes

DROSOPHILA
GENETICS & EPIGENETICS

GED lab

Christophe ANTONIEWSKI
Clement CARRE
Bruno DA SILVA
Hélène THOMASSIN-BOURREL
Margarita Angelova

Bioinformatics Post-doc available!

Thanks!

mvandenb@snv.jussieu.fr