

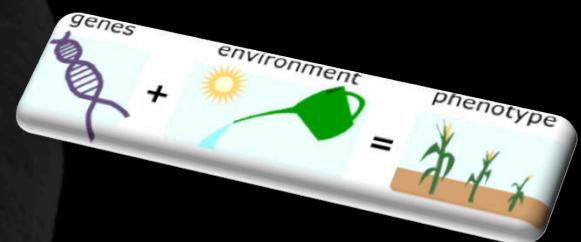
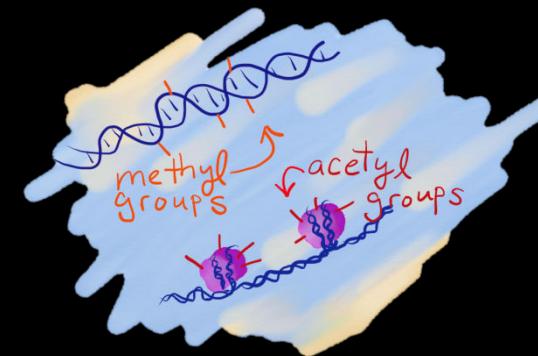


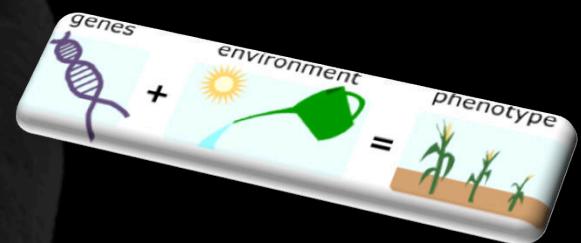
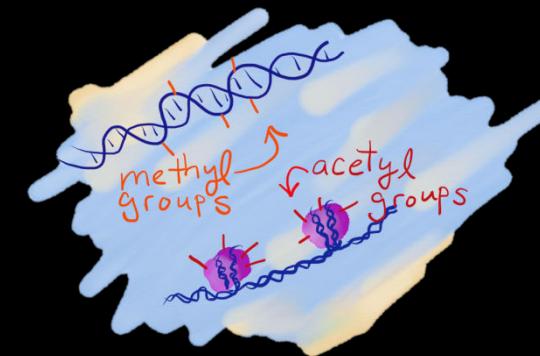
*David Roquis, Rémi Emans, Guillaume Mitta &
Christoph Grunau*

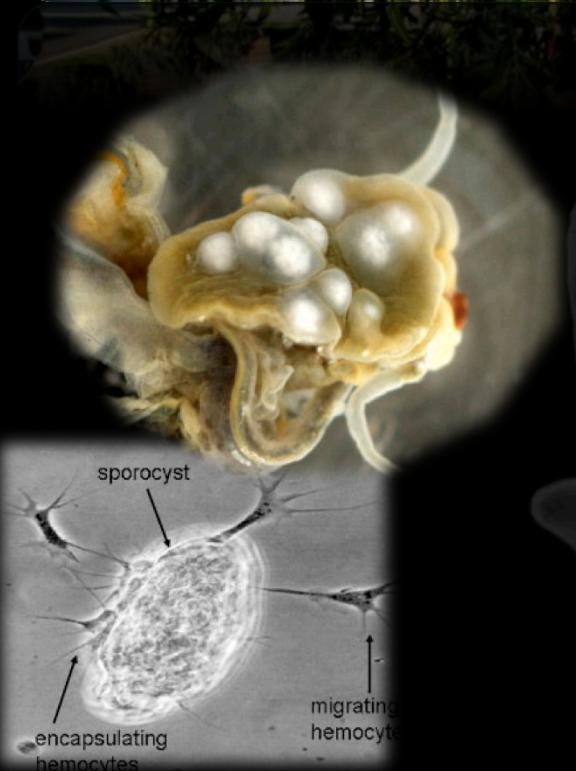
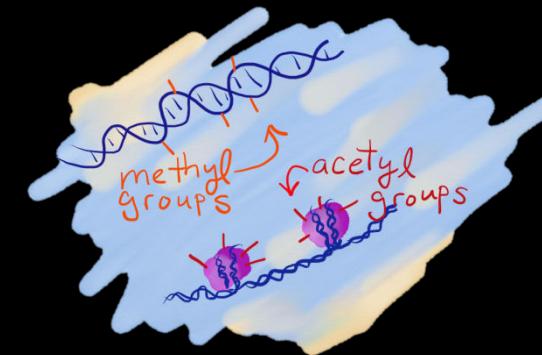
Galaxy for the rest of us

Ph.D. student, 2nd year









A human parasite



Schistosoma mansoni

A freshwater snail



Biomphalaria glabrata

A tropical coral



Pocillopora damicornis

An insect



Dinocampus coccinellae

- + Previously → Candidate gene approach.



- + Previously → Candidate gene approach.
- + Next Generation Sequencing more **available** and **affordable**

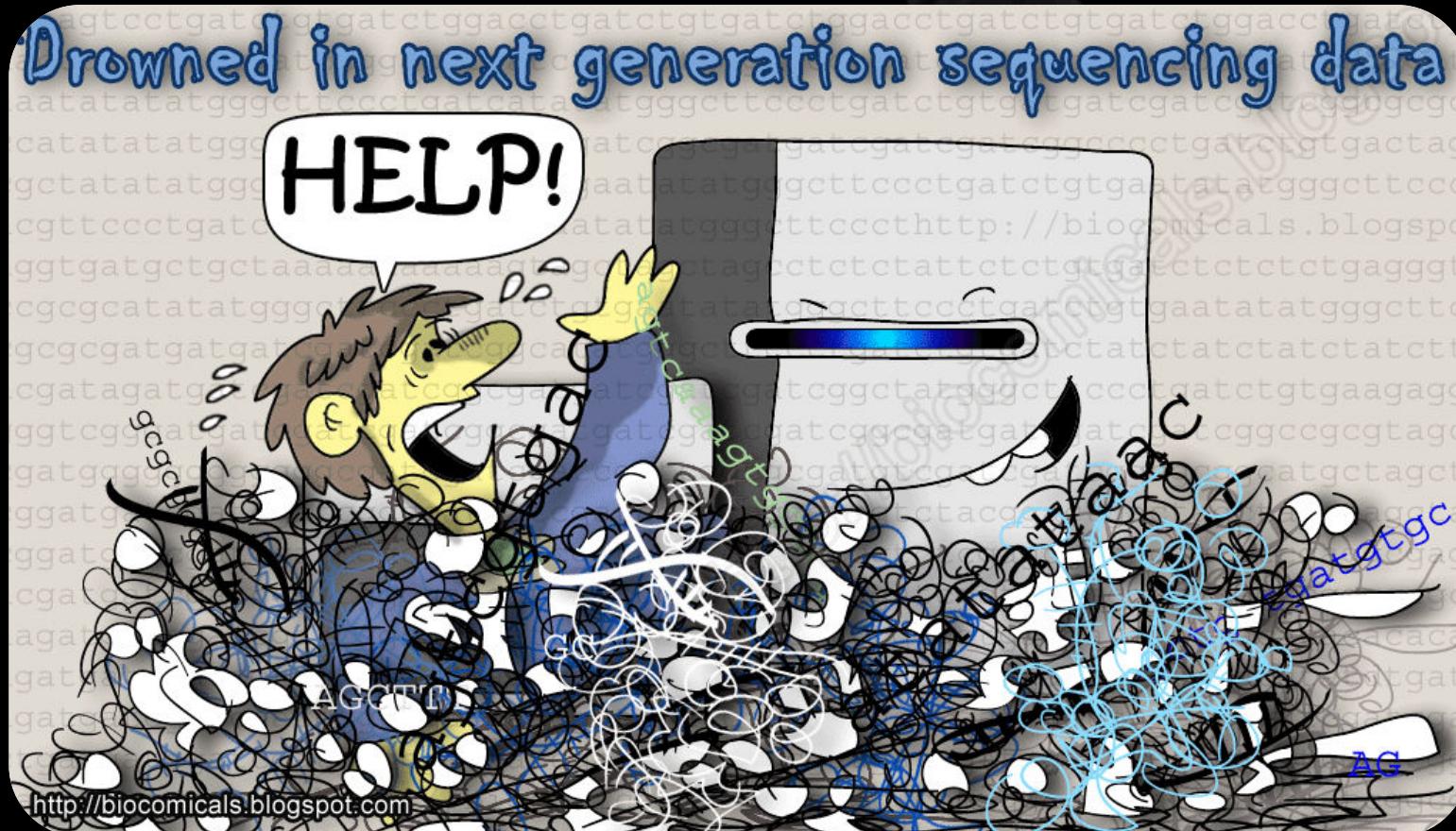


- + Previously → Candidate gene approach.
- + Next Generation Sequencing more **available** and **affordable**



- + Will to switch to genome-wide approaches to study **Genomes, Epigenomes, Transcriptomes**.

- + No bioinformatician.
- + No dedicated platform or hardware on the campus.

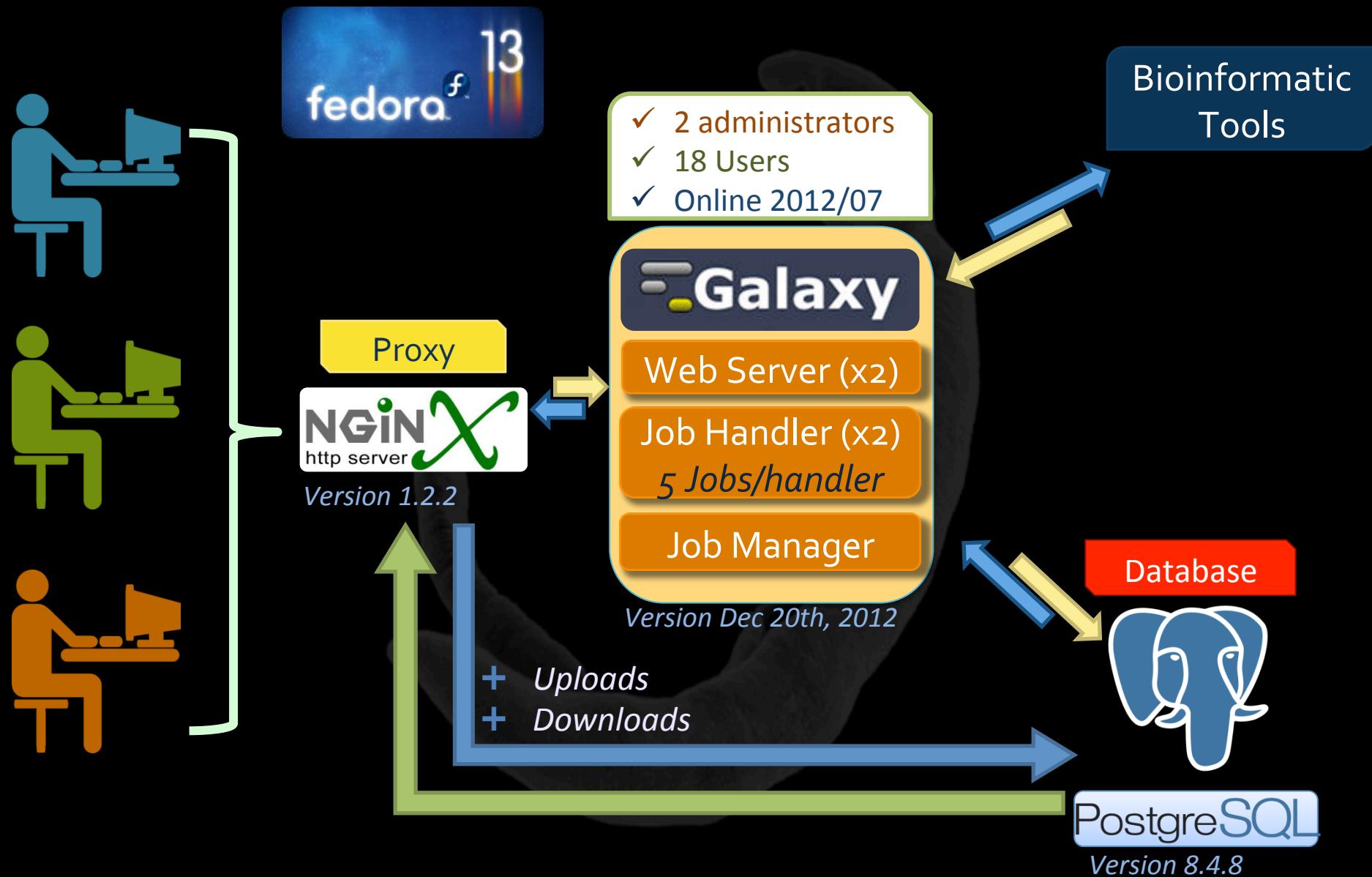


- + But, a few people with some bioinformatics knowledge...



But why a local instance?

- + Very different and specific needs.
- + Working on non-model organisms means that most of the tools need to be modified/adapted.
- + Flexibility.





Network

 $2 \times 100 \text{ mb/sec}$ 

Dell PowerEdge R820
(~5 000 €)

- + 16 cores, Xeon 2,4 GHz
- + 96 Gb RAM
- + 1.67 Tb hard drive



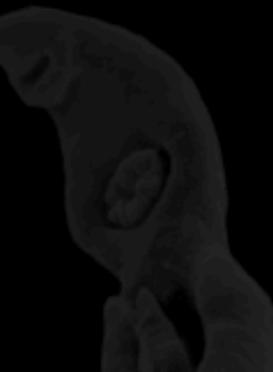
Network

 $2 \times 100 \text{ mb/sec}$ 

Dell PowerEdge R820
(~5 000 €)

- + 16 cores, Xeon 2,4 GHz
- + 96 Gb RAM
- + 1.67 Tb hard drive

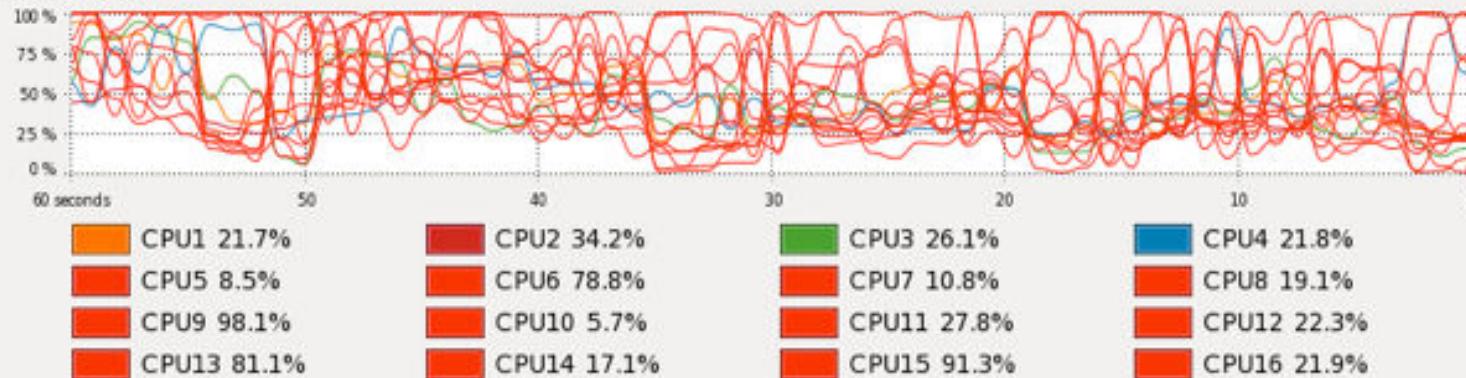
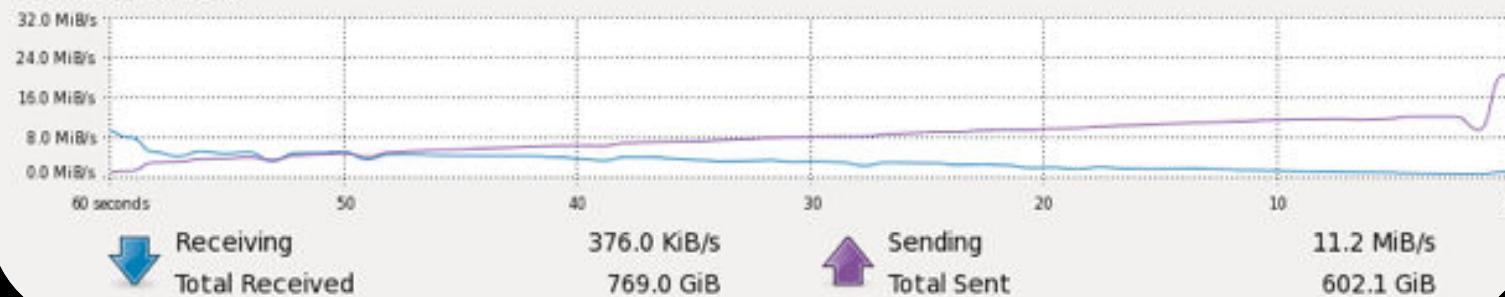
Network

 $2 \times 100 \text{ mb/sec}$ 

Buffalo TerraStation 5400 WSS
(~2 000 € x 2)

- + 20 Tb hard drive
- + Main and backup



CPU History**Memory and Swap History****Network History**





Local storage by
project manager





Local storage by project manager



Data Libraries

 [Advanced Search](#)

Data library name ↓

[B.glabrata annotation - genome](#)

Data library description

FASTA and gff files from VectorBase

[B.glabrata annotation - results](#)

B.glabrata analysis results

[B.glabrata annotation - RNA-Seq I](#)

RNA-Seq data provided by WashU (fastq f/r)

[B.glabrata annotation - RNA-Seq II](#)

RNA-Seq data provided by WashU (aligned BAM and assembled fasta)

[B.glabrata annotation - RNA-Seq III](#)

sorted BAM files of RNA-Seq suitable for upload to IGV (from G.Oliveira)

[B.glabrata - assemblies](#)

assemblies for read dependency

[B.glabrata RNA-Seq \(clean\)](#)

RNA-Seq without adaptors, quality clipped

[B.glabrata RNA-Seq \(paired\)](#)

fastq custom upload N.Dheilly 16/01/13

Data libraries by Galaxy Admin





Local storage by
project manager



Data Libraries

 [Advanced Search](#)

Data library name ↓

[B.glabrata annotation - genome](#)

Data library description

FASTA and gff files from VectorBase

[B.glabrata annotation - results](#)

B.glabrata analysis results

[B.glabrata annotation - RNA-Seq I](#)

RNA-Seq data provided by WashU (fastq f/r)

[B.glabrata annotation - RNA-Seq II](#)

RNA-Seq data provided by WashU (aligned BAM and assembled fasta)

[B.glabrata annotation - RNA-Seq III](#)

sorted BAM files of RNA-Seq suitable for upload to IGV (from G.Oliveira)

[B.glabrata - assemblies](#)

assemblies for read dependency

[B.glabrata RNA-Seq \(clean\)](#)

RNA-Seq without adaptors, quality clipped

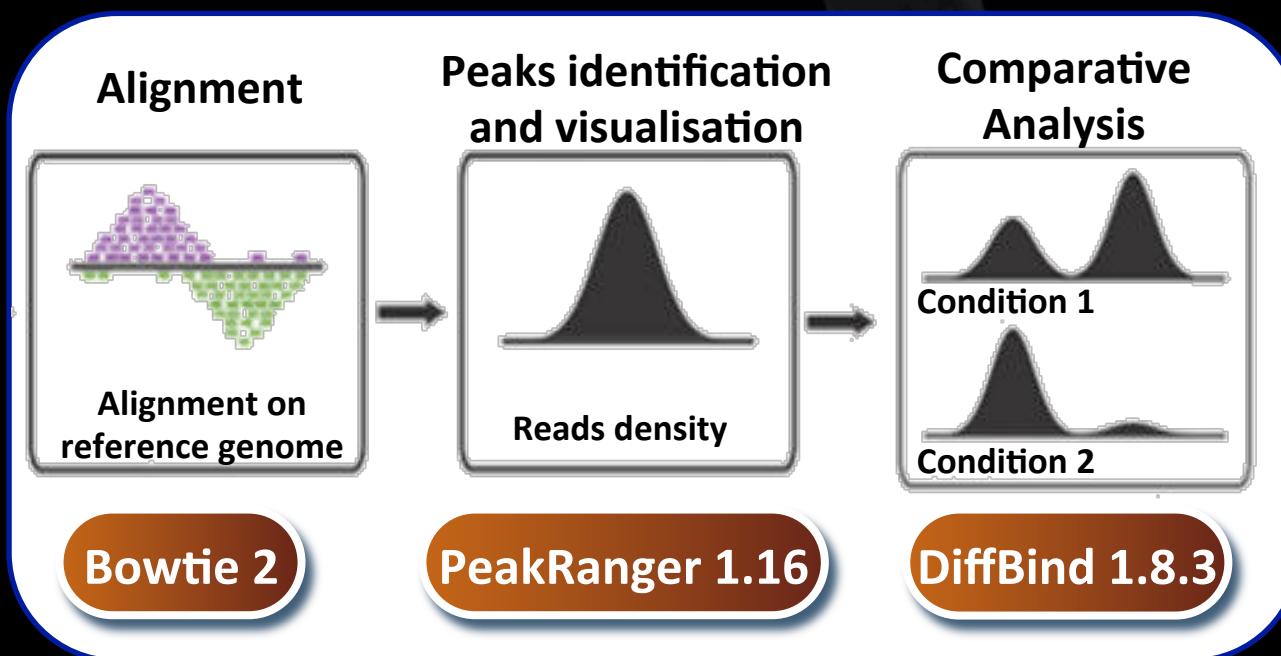
[B.glabrata RNA-Seq \(paired\)](#)

fastq custom upload N.Dheilly 16/01/13

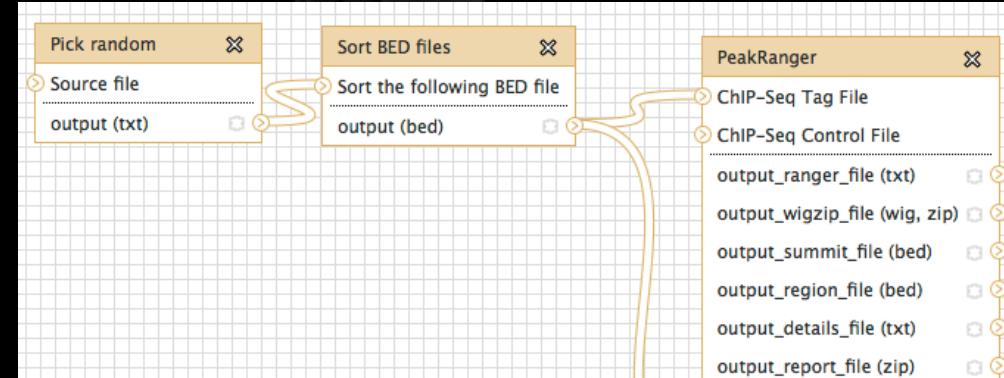
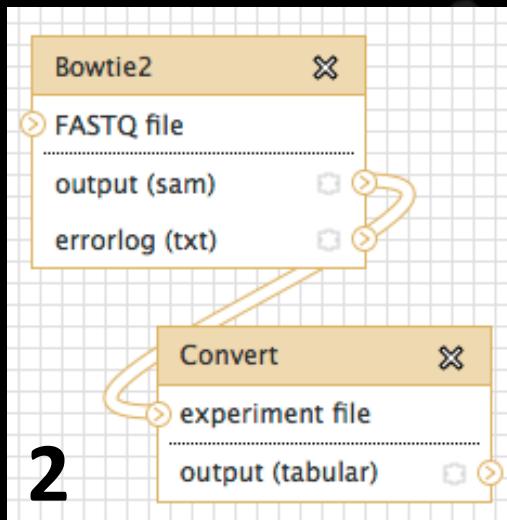
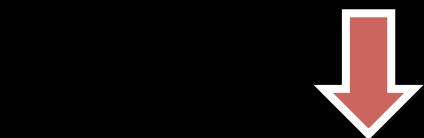
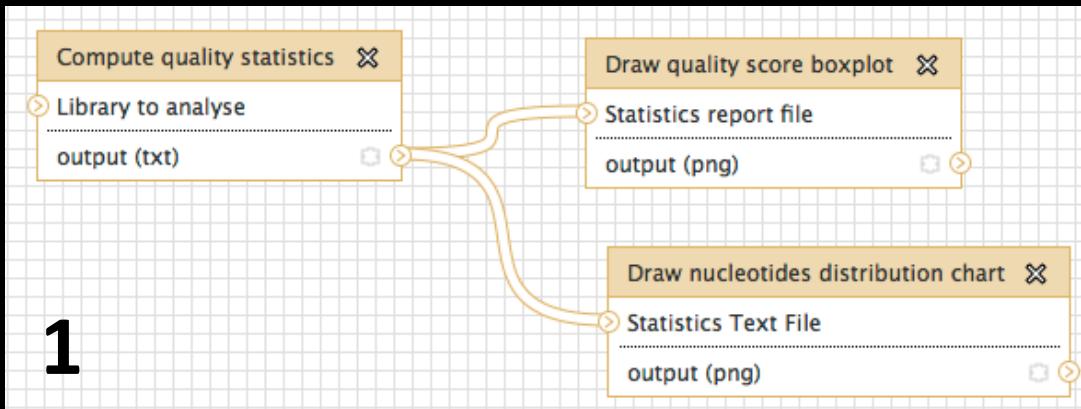
Data libraries by
Galaxy Admin



SFTP?



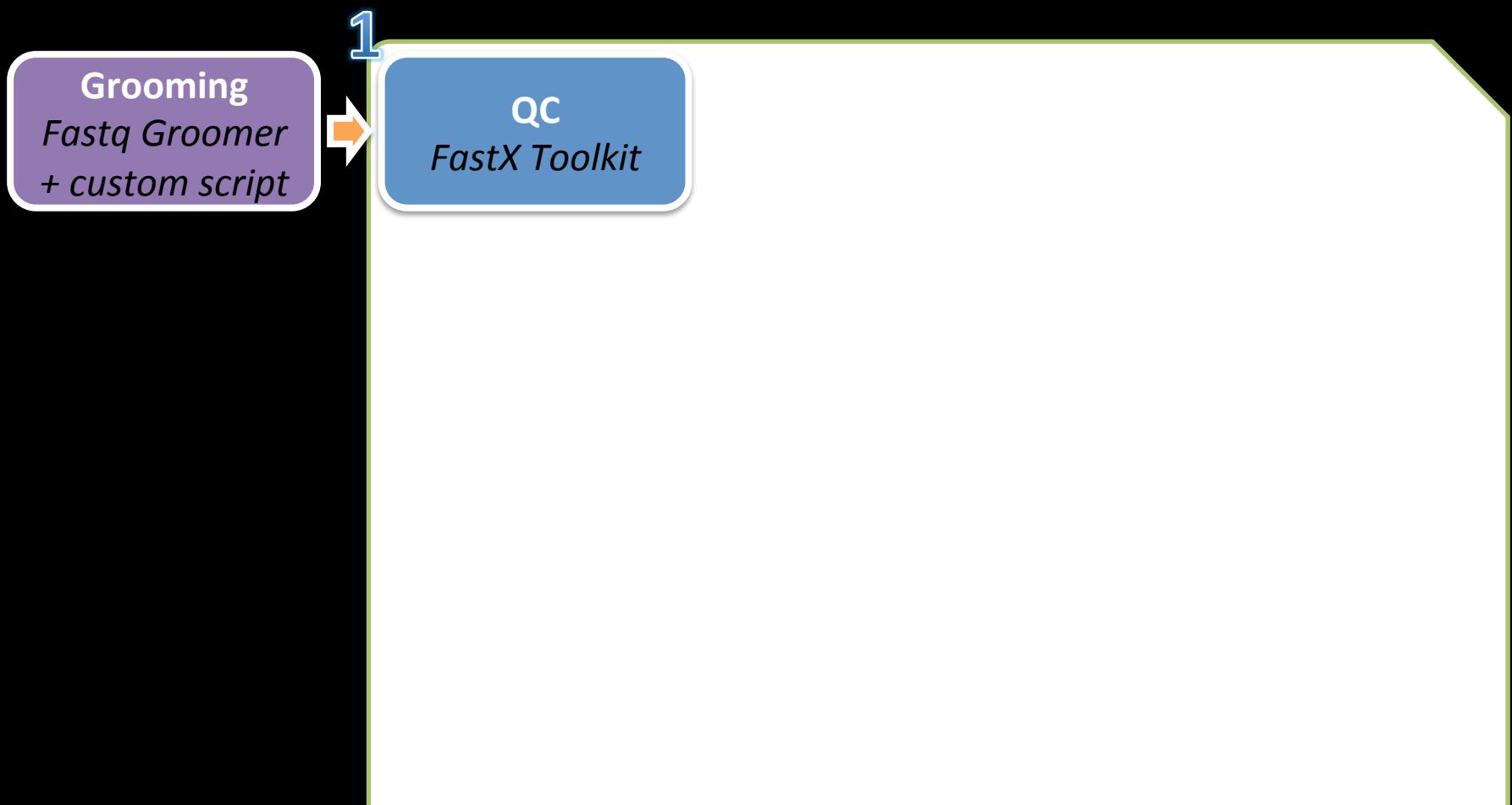
Adapted from Bardet et al. (2012)

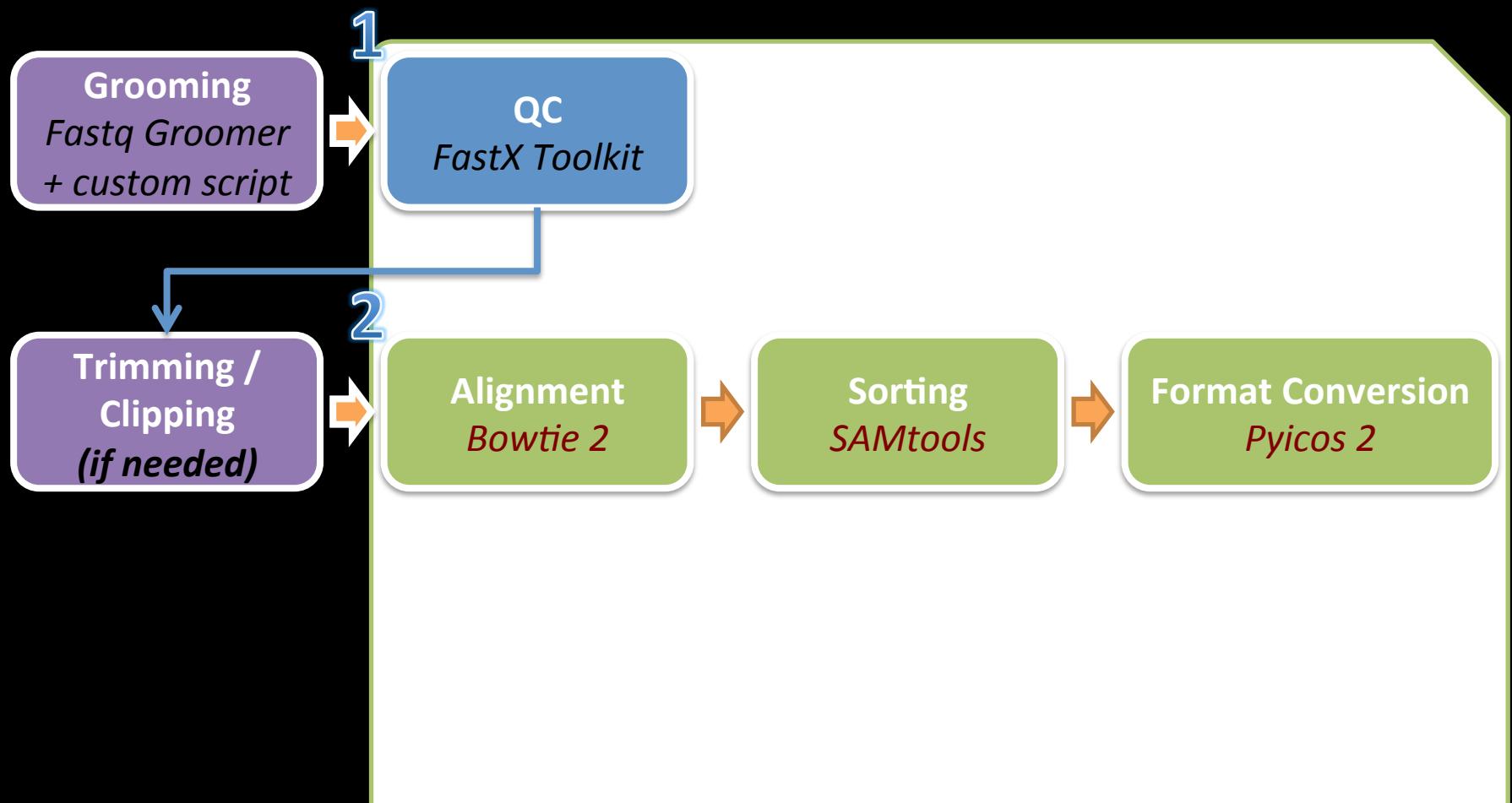


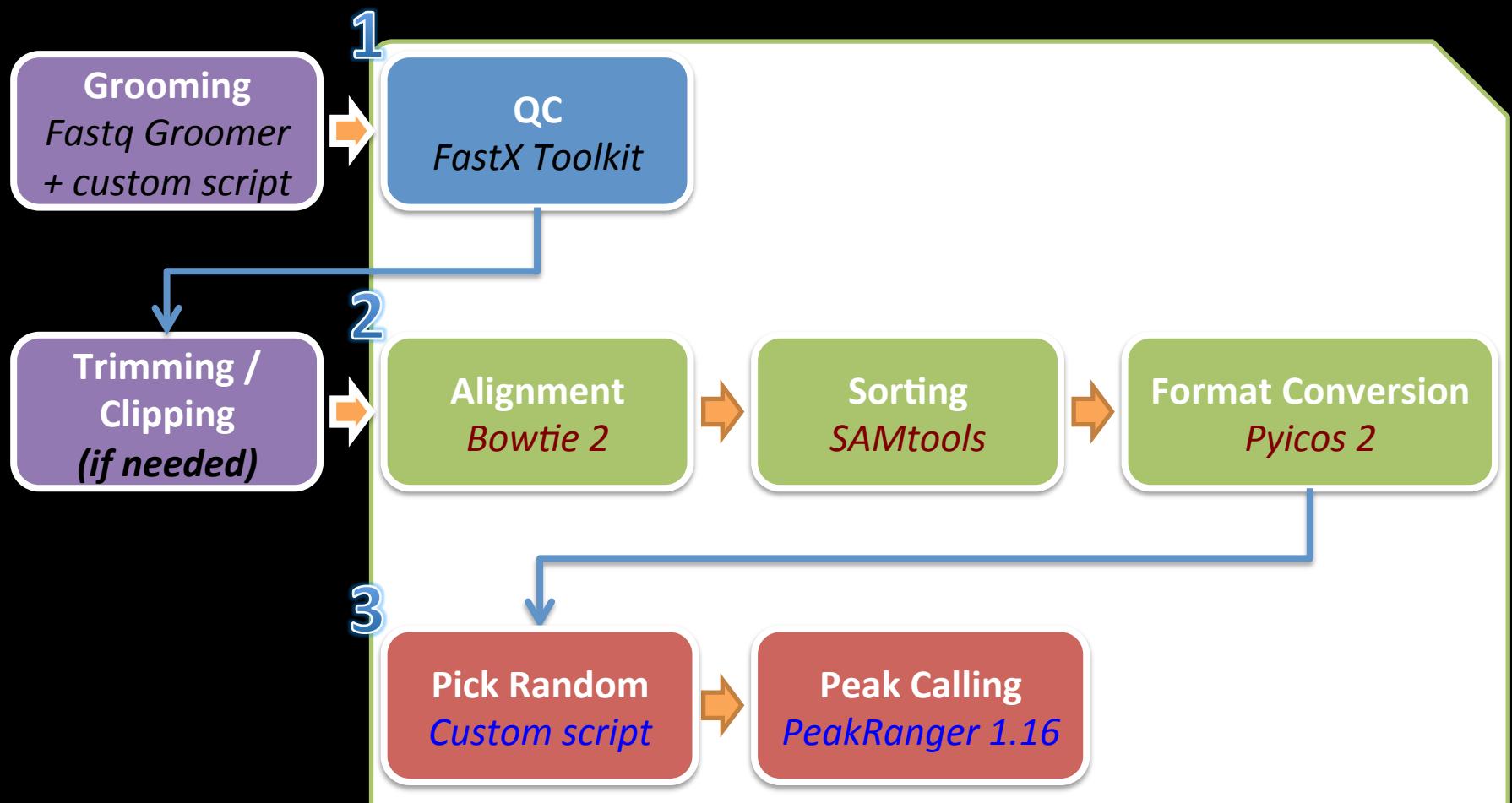
3

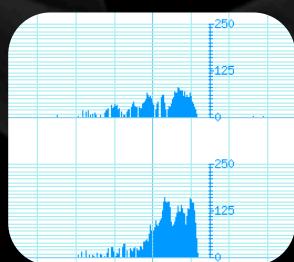
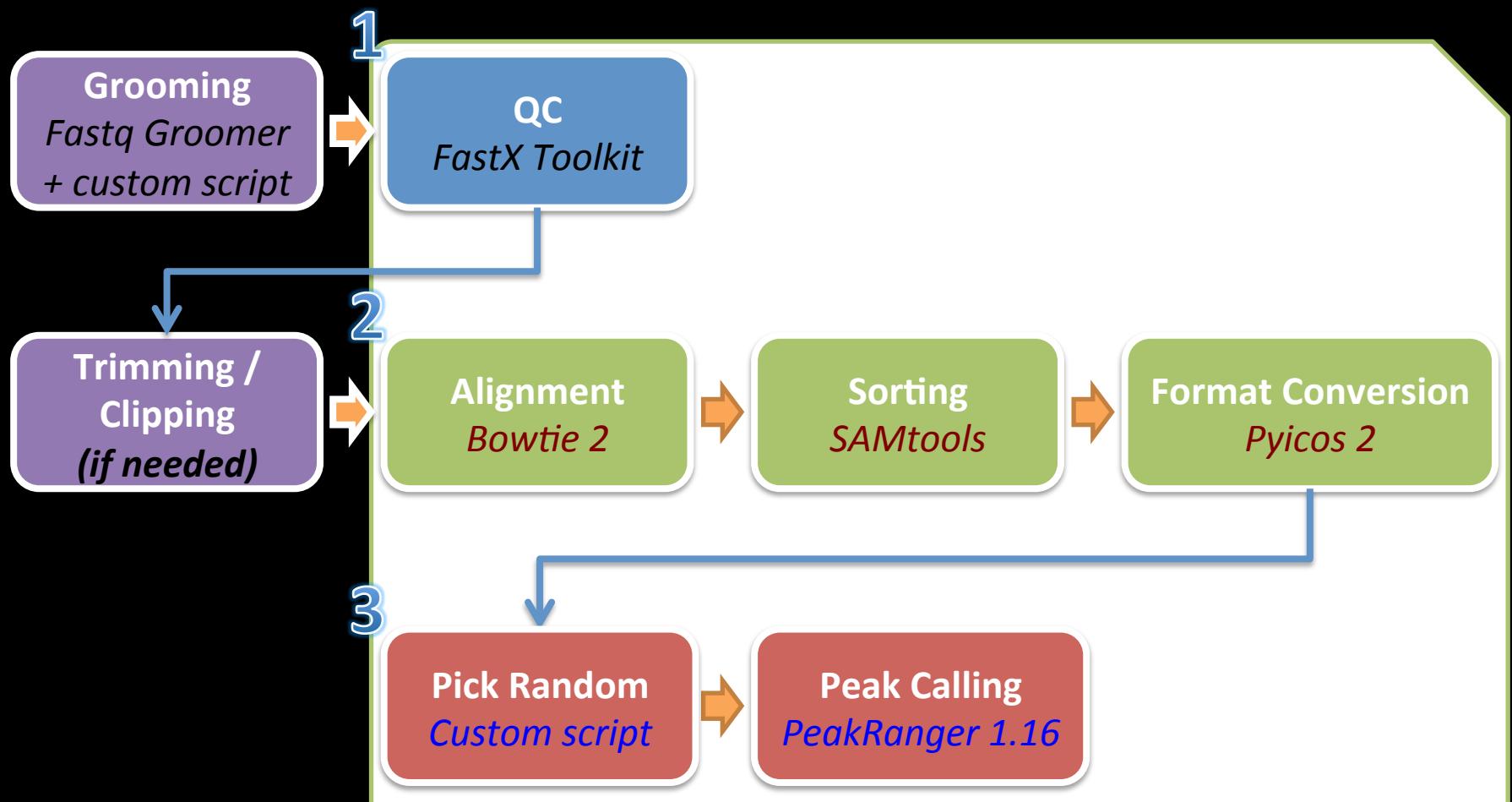
Grooming
Fastq Groomer
+ custom script

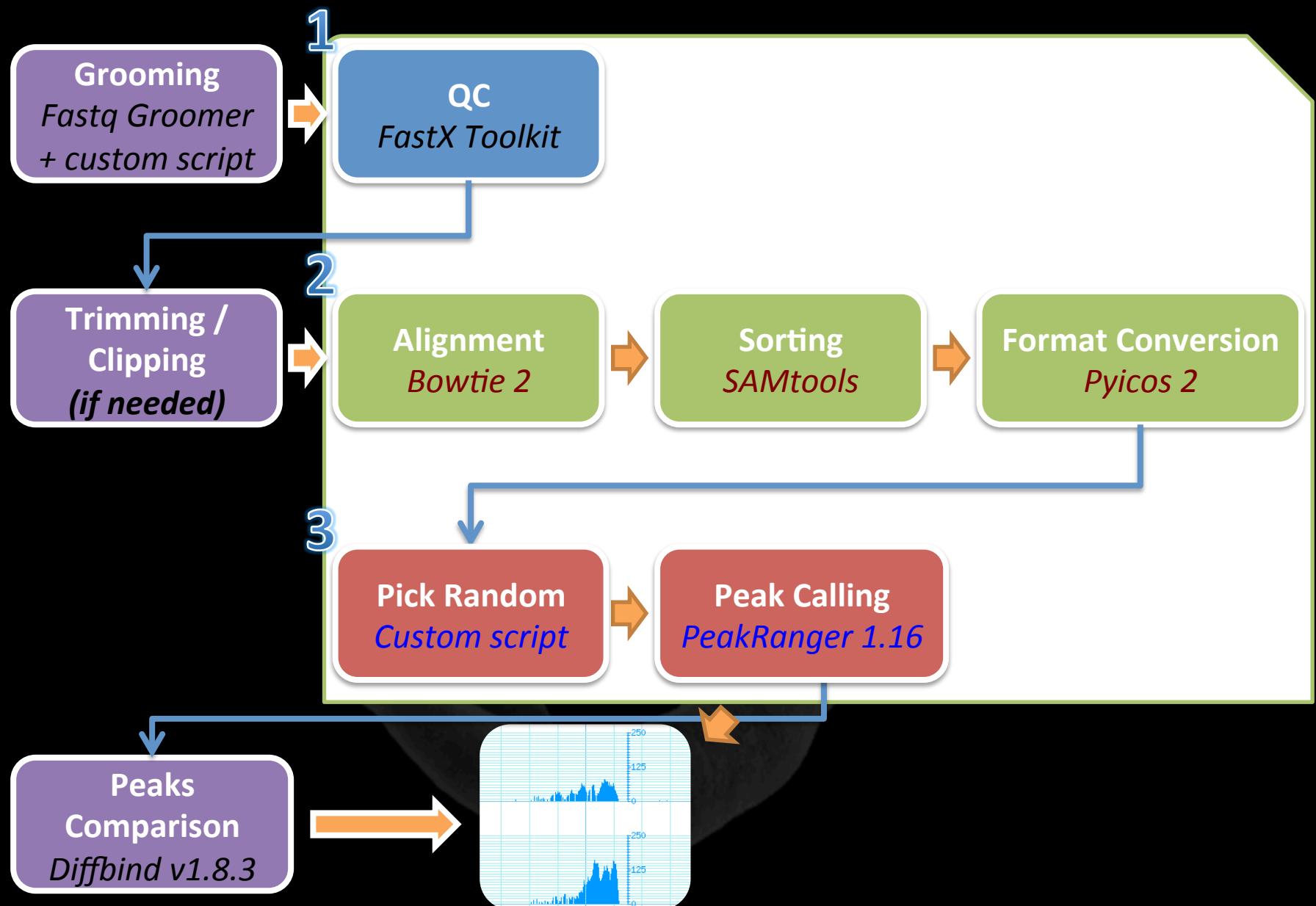










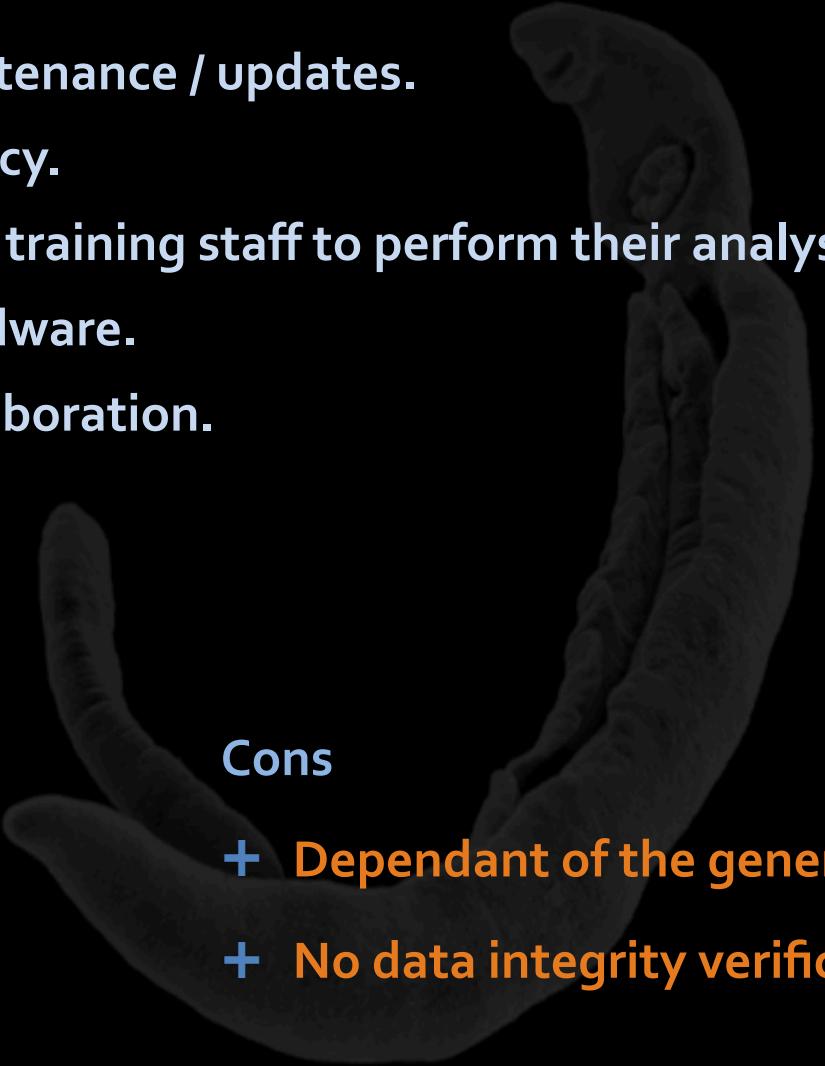


Pros

- + Very flexible maintenance / updates.
- + No data redundancy.
- + Less time spent in training staff to perform their analyses.
- + No expensive hardware.
- + International collaboration.

Pros

- + Very flexible maintenance / updates.
- + No data redundancy.
- + Less time spent in training staff to perform their analyses.
- + No expensive hardware.
- + International collaboration.



Cons

- + Dependant of the general network speed.
- + No data integrity verification.



System Administrator

Rémi Emans



Supervisor

Céline Cosseau



Supervisor

Christoph Grunau



Lab Director

Guillaume Mitta

AGENCE NATIONALE DE LA RECHERCHE

ANR

EPIGEVOL

ÉPIGÉnétique
et EVOution

2 énergie
environnement

Ecole Doctorale UPVD

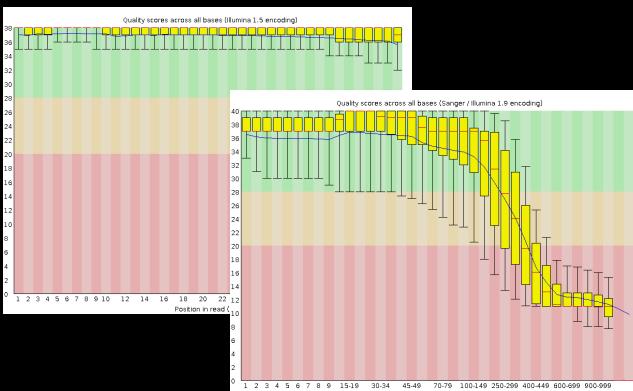
Université de Perpignan Via Domitia

Ecole Doctorale UPVD

Université de Perpignan Via Domitia

Example of application

```
@HWI-ST609:177:D16HYACXX:3:1308:14623:59114 1:N:0:AGTTCC
TGAGACTAAGGCATATTATAATTGCCGCGAACTGTATAGTAAAAGCAC
+
@CCFFFFFGHHGHIIJJJJJJJJJIJIGGGJIIIGIIIIGIJJIIBHI
```



Data Libraries	
search dataset name, info, message, dbkey	<input type="button" value="🔍"/>
Advanced Search	
Data library name ↓	Data library description
B.glabrata annotation - genome	FASTA and gff files from VectorBase
B.glabrata annotation - results	B.glabrata analysis results
B.glabrata annotation - RNA-Seq I	RNA-Seq data provided by WashU (fastq f/r)
B.glabrata annotation - RNA-Seq II	RNA-Seq data provided by WashU (aligned BAM and assembled fasta)
B.glabrata annotation - RNA-Seq III	sorted BAM files of RNA-Seq suitable for upload to IGV (from G.Oliveira)
B.glabrata - assemblies	assemblies for read dependency
B.glabrata RNA-Seq (clean)	RNA-Seq without adaptors, quality clipped
B.glabrata RNA-Seq (paired)	fastq custom upload N.Dheilly 16/01/13

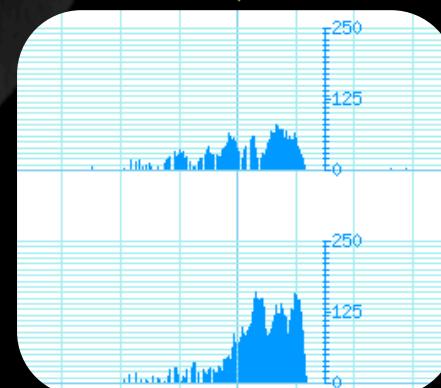


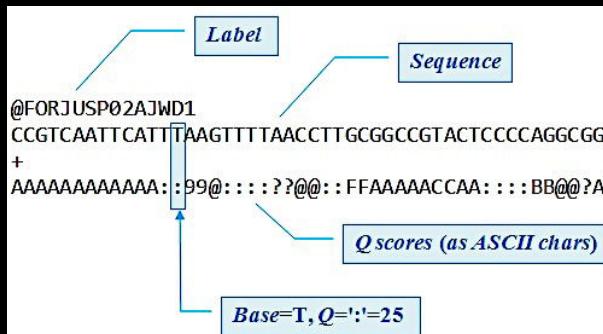
+ Align Reads

- + Format Conversion
- + Sorting

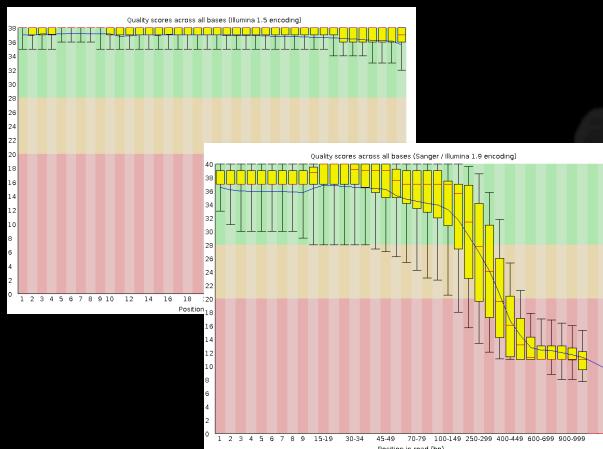
PeakRanger

+ Find Enriched Regions





↓ + FastQC



+ Format Conversion

+ Trimming depending on QC



+ Align Reads

+ Format Conversion

+ Sorting

PeakRanger

+ Find Enriched Regions

