

Using Galaxy for the FMI NGS Pipeline

Hans-Rudolf Hotz (hrh@fmi.ch)

**Friedrich Miescher Institute for Biomedical Research
Basel, Switzerland**

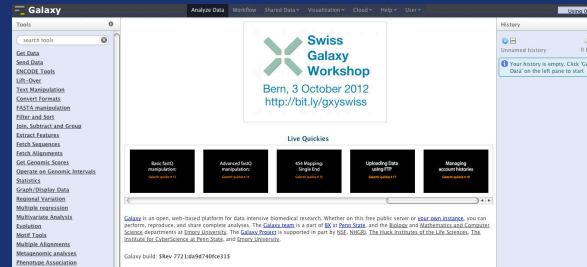
Galaxy at FMI is...

- in production for more than 5 years (3 years for NGS)
- used to reduce the workload of the core facilities
- used to bridge the gap

the “average”
lab scientist



the “modern”
lab scientist

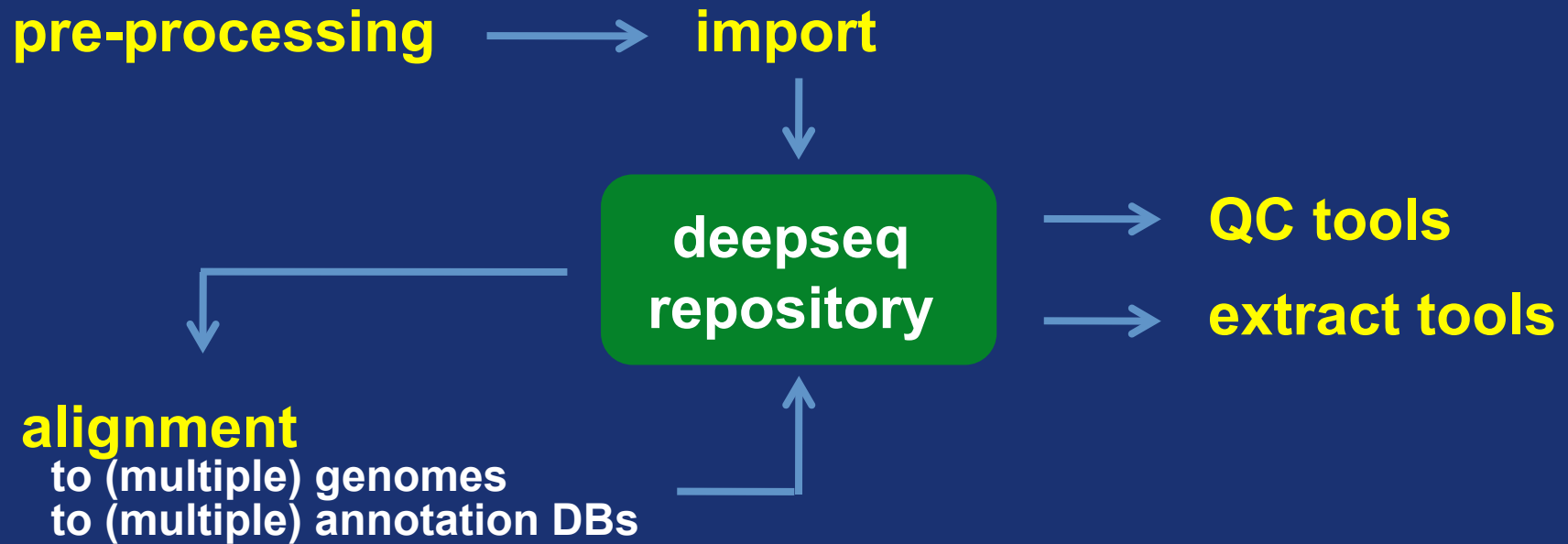


FMI

Friedrich Miescher Institute
for Biomedical Research

The (current) FMI Deep Sequencing Pipeline

(developed by Michael Stadler and Dimos Gaidatzis)



FMI

Friedrich Miescher Institute
for Biomedical Research

The FMI Deep Sequencing is


....just a bunch of Perl scripts *(currently)*

 which can be easily added to Galaxy

....just a simple file system *(currently)*

....which cannot be added to Galaxy.

(Galaxy uses its own data directory)

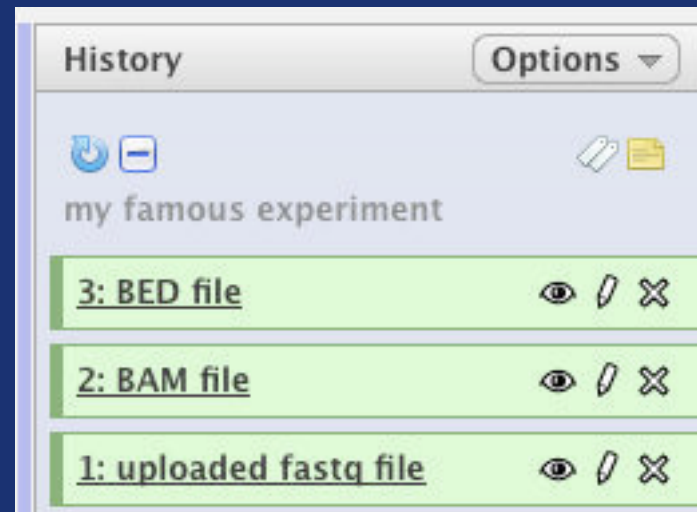
 we don't have to, we just have to give Galaxy access to the file system (without using "Data Libraries")

FMI

Friedrich Miescher Institute
for Biomedical Research

a simple NGS workflow

- your famous aligner
- your famous extract tool

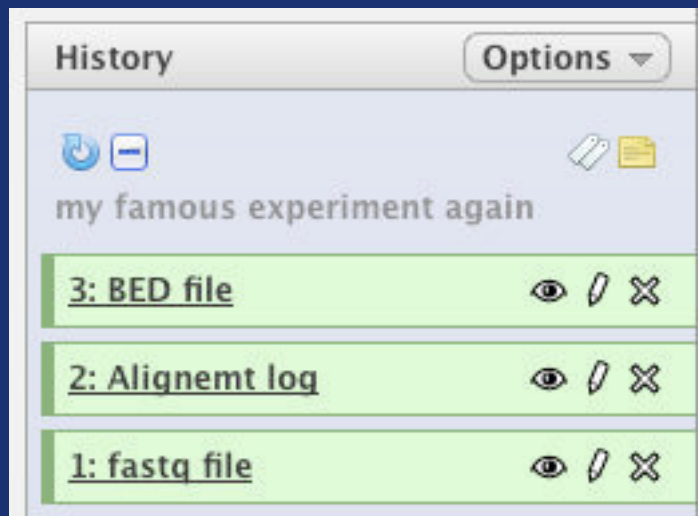


FMI

Friedrich Miescher Institute
for Biomedical Research

a simple NGS workflow

- do you need the first result (ie the alignment) as a new history item?
- does your tool require a Galaxy history item as input?

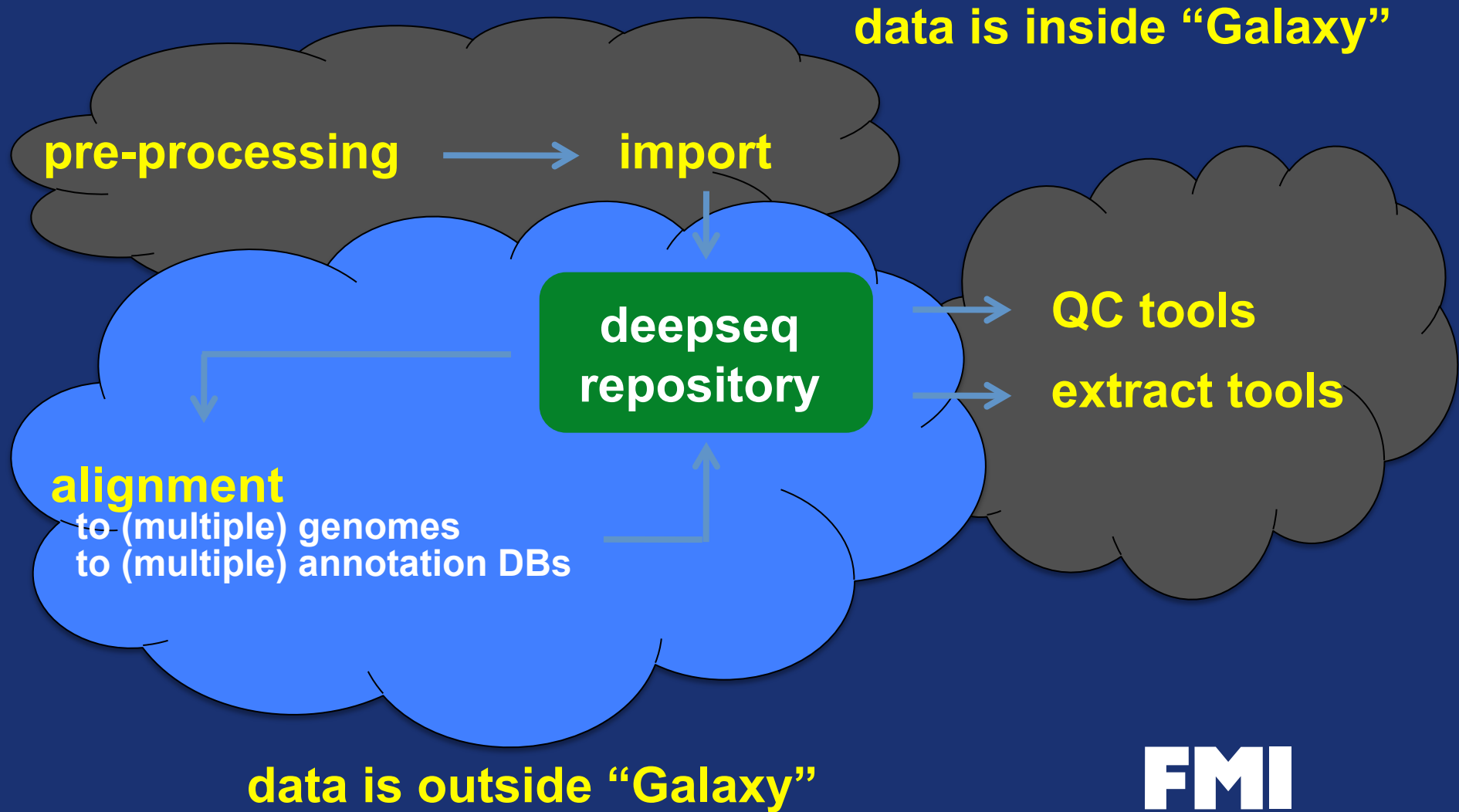


- the 'famous aligner' has a wrapper storing the BAM file in the central NGS repository and creating just a log file for Galaxy
- your 'famous extract tool' knows the location of the NGS repository

FMI

Friedrich Miescher Institute
for Biomedical Research

The FMI Deep Sequencing Pipeline



FMI

Friedrich Miescher Institute
for Biomedical Research

storing data outside of Galaxy

makes it easier to share with non-Galaxy users



The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy / FMI Xenon1', 'Workflow', 'Shared Data', 'Lab', 'Visualization', 'Admin', 'Help', and 'User'. The main content area is split into two panels. The left panel is a terminal window displaying the text: 'successfully finished annotation of sample_20110518 to dm3-dmV01-aln2'. The right panel is the 'History' view, which shows a list of workflow steps. The top step is 'my famous experiment again'. Below it are three steps, each with a green background and icons for viewing, editing, and deleting:

- 3: BED file
- 2: Alignment log
- 1: fastq file

FMI

Friedrich Miescher Institute
for Biomedical Research

makes it easier to share with non-Galaxy users

successfully finished annotation of
sampleId_20110518 to dm3-dmV01-aln2

and now the command line geek can do

```
[geek@xenon1 ~]$ extractData.pl -f -s p -m  
100 -i mySampleId_20110518 dm3-dmV01-aln2  
genome |frag2bed.pl -t -q -U - | head -5  
track name='mySampleId_20110518'  
chr2L    10493    10528    sq39319  1        +  
chr2L    10736    10764    sq74484  1        +  
chr2L    11442    11477    sq1340   1        +  
chr2L    13799    13834    sq84955  1        +  
[geek@xenon1 ~]$
```

makes it easier to share with non-Galaxy users

command line

```
extractData.pl -f -s p -m 100 -i  
mySampleId_20110518 dm3-dmV01-aln2 genome |  
frag2bed.pl -t -q -U -
```

Galaxy tool definition file

```
#elif ($summary.mode=="bed") #extractData.pl  
-f $strand $maxhits $signCnts  
$sampleSelect.sampleId $genome-$annot-aln2  
genome | frag2bed.pl -t -q $summary.ucsc -  
> $output
```

FMI

Friedrich Miescher Institute
for Biomedical Research

and doing the same in Galaxy

Extract data (step 1 of 2)

Sample selection:

Extract data (step 2 of 2)

Strand selection:

Ignore strand (use all)

```
track name='mySampleId_20110518'  
chr2L 10493 10528 sq39319 1 +  
chr2L 10736 10764 sq74484 1 +  
chr2L 11442 11477 sq1340 1 +  
chr2L 13799 13834 sq84955 1 +  
chr2L 13940 13974 sq9998 1 +  
chr2L 13948 13979 sq2852 1 +  
chr2L 14266 14301 sq29828 1 +  
chr2L 14381 14414 sq62373 1 +  
chr2L 14612 14645 sq50170 1 +  
chr2L 15215 15250 sq7575 1 +  
chr2L 18459 18490 sq20174 1 +  
chr2L 21264 21295 sq20174 1 +  
chr2L 67455 67489 sq31577 1 +  
chr2L 72882 72916 sq470 1 +  
chr2L 75216 75251 sq16959 1 +  
chr2L 75381 75416 sq21962 1 +  
chr2L 75416 75451 sq58948 1 +  
chr2L 76053 76088 sq54784 1 +  
chr2L 85320 85355 sq58664 1 +  
chr2L 101308 101343 sq2012 1 +  
chr2L 102620 102655 sq9815 1 +  
chr2L 103097 103132 sq63047 1 +  
chr2L 103605 103640 sq50914 1 +  
chr2L 103769 103802 sq69218 1 +  
chr2L 103855 103890 sq58865 1 +
```

History

Options



my famous experiment again



17.9 Mb

3: BED file



2: Alignment log



1: fastq file



Acknowledgment

Michael Stadler Lukas Burger

Dimos Gaidatzis

Tim Roloff Stefan Grzybek

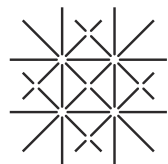
Anita Lerch

Thomas Übermeier Jan Welker

....and all the people from the “Galaxy”



Swiss Institute of
Bioinformatics



UNI
BASEL



FMI

Friedrich Miescher Institute
for Biomedical Research