



LEIDEN UNIVERSITY MEDICAL CENTER

# Evaluation and wrap up

**Jeroen F. J. Laros**

**Leiden Genome Technology Center**

**Department of Human Genetics**

**Center for Human and Clinical Genetics**



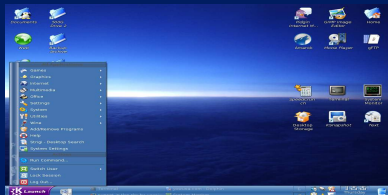
*Linux*

Figure 1: Window managers.

You have seen that the graphical user interface is not very hard to use.

- Applications are similar to that of other operating systems.
- Mostly point and click.

## *Why Linux?*

A popular operating system for scientific developers.

- Computer science.

## *Why Linux?*

A popular operating system for scientific developers.

- Computer science.
- Physics.
- Astronomy.
- ...

## *Why Linux?*

A popular operating system for scientific developers.

- Computer science.
- Physics.
- Astronomy.
- ...

It is free.

- The operating system is free.
- Almost all software is free.
  - Development software.

## *Why Linux?*

A popular operating system for scientific developers.

- Computer science.
- Physics.
- Astronomy.
- ...

It is free.

- The operating system is free.
- Almost all software is free.
  - Development software.

Not only free as in costless, but also *open source*.



## *Why the command line?*

```
1  bwa aln ref.fa s_1.fq > s_1.sai
2  bwa aln ref.fa s_2.fq > s_2.sai
3  bwa sampe ref.fa s_1.sai s_2.sai s_1.fq s2.fq > s.sam
4  samtools view -bt ref.fa -o s.bam s.sam
5  samtools sort s.bam s.sorted
6  samtools mpileup -f ref.fa s.sorted.bam > s.pileup
```

Listing 1: Bash script snippet.

If the command line works, it works in a script.

- Less chance of errors.
- Less work.
- Reproducible.



## *Programming*

We have done basic scripting, a form of programming.

For more sophisticated tasks, we use *higher programming languages*.

- Usually a lot faster.
- Easier to do general tasks.
- Less easy to connect programs, like we did in this course.

## *Programming*

We have done basic scripting, a form of programming.

For more sophisticated tasks, we use *higher programming languages*.

- Usually a lot faster.
- Easier to do general tasks.
- Less easy to connect programs, like we did in this course.

There are programming languages for specific purposes.

- Parallelised pipelines.

## *Clusters*

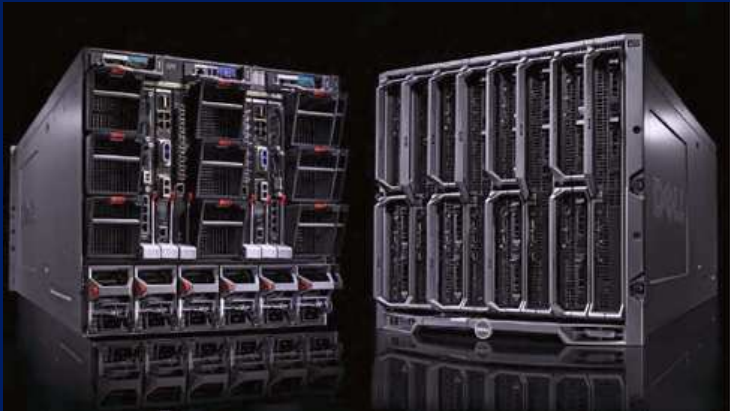


Figure 3: Dell M610 blade server.

## Clusters

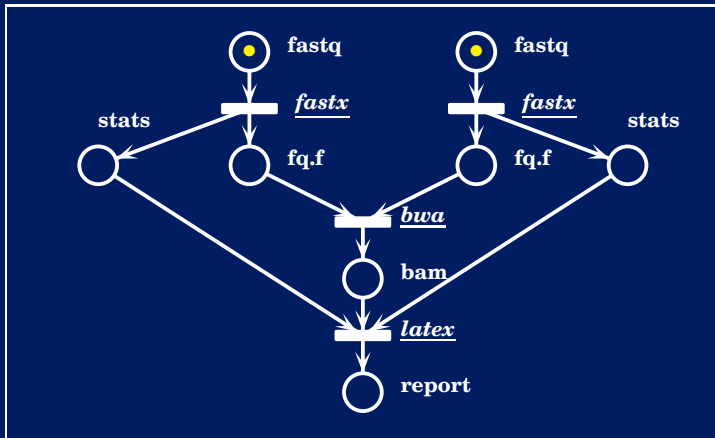


Figure 4: A parallel workflow.

## Clusters

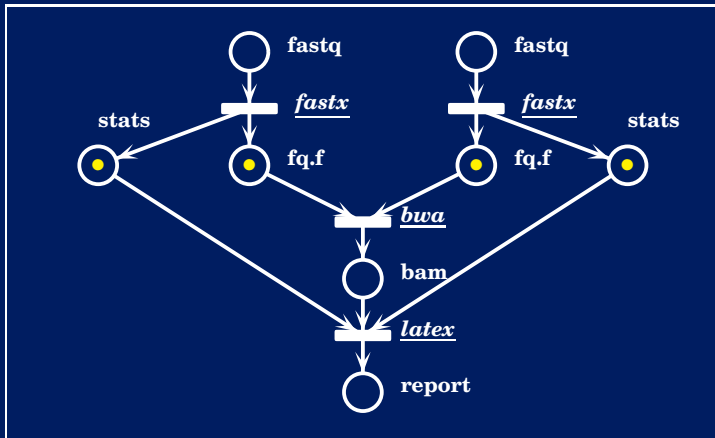


Figure 4: A parallel workflow.

## Clusters

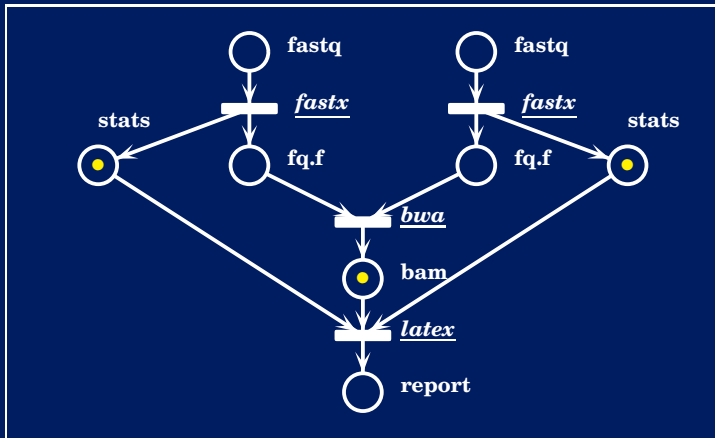


Figure 4: A parallel workflow.

## Clusters

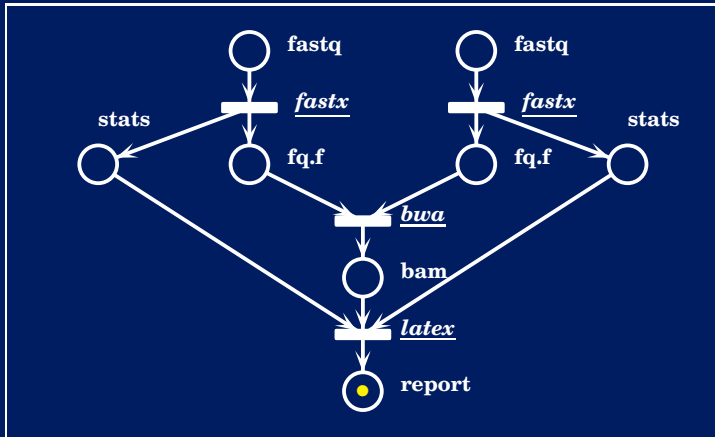


Figure 4: A parallel workflow.

### *Overview*

We offer the following courses:

- Introduction to clusters.
- Introduction to Git.
- Next generation sequencing (NGS) data analysis.
- Python programming.



*Introduction course clusters*

Half day course.

- Connecting to a cluster.
- Using the Sun Grid Engine.
- Do's and don'ts.
- Makefiles.

*Git*

Everyone in the Bioinformatics field:

- Software development.
- Project management.
- Collaboration.

**M. Vermaat**

<https://humgenprojects.lumc.nl/trac/humgenprojects/wiki/git>

## *Git*

Everyone in the Bioinformatics field:

- Software development.
- Project management.
- Collaboration.

Topics:

- Git Basics
- Branching
- Remotes
- Project skeleton / git annex

**M. Vermaat**

<https://humgenprojects.lumc.nl/trac/humgenprojects/wiki/git>

*Next Generation Sequencing data analysis*

Three day course:

- Discussion of different platforms and produced data.
  - Illumina, Roche, ABI, Ion Torrent, etc.

**J.M. Boer, J.T. den Dunnen, W. van IJcken**

## *Next Generation Sequencing data analysis*

Three day course:

- Discussion of different platforms and produced data.
  - Illumina, Roche, ABI, Ion Torrent, etc.
- General techniques.
  - Resequencing, structural variation, de novo assembly, visualisation, pipelines, etc.

**J.M. Boer, J.T. den Dunnen, W. van IJcken**

## *Next Generation Sequencing data analysis*

Three day course:

- Discussion of different platforms and produced data.
  - Illumina, Roche, ABI, Ion Torrent, etc.
- General techniques.
  - Resequencing, structural variation, de novo assembly, visualisation, pipelines, etc.
- Specific applications.
  - QC, statistics, expression, ChIP-seq, metagenomics, etc.

**J.M. Boer, J.T. den Dunnen, W. van IJcken**

## *Next Generation Sequencing data analysis*

Three day course:

- Discussion of different platforms and produced data.
  - Illumina, Roche, ABI, Ion Torrent, etc.
- General techniques.
  - Resequencing, structural variation, de novo assembly, visualisation, pipelines, etc.
- Specific applications.
  - QC, statistics, expression, ChIP-seq, metagenomics, etc.
- Practical sessions.
  - Galaxy, NextGENe, CLCbio.

Room for 60 people, always full.

**J.M. Boer, J.T. den Dunnen, W. van IJcken**

## *Programming in Python*

Four day course.

- Python basics.
- Standard data structures.
- Working with NumPy arrays.
- Plotting with matplotlib.
- Object-oriented programming.
- The Biopython library.

M. Vermaat, W. Arindrato, Z. Tatum

<https://humgenprojects.lumc.nl/trac/programming-course>





Tugce Güler  
 Zuotian Tatum  
 Maarten van Itersson  
 Martijn Vermaat  
 Rob Marissen  
 Magnus Palmblad  
 Jeroen Laros