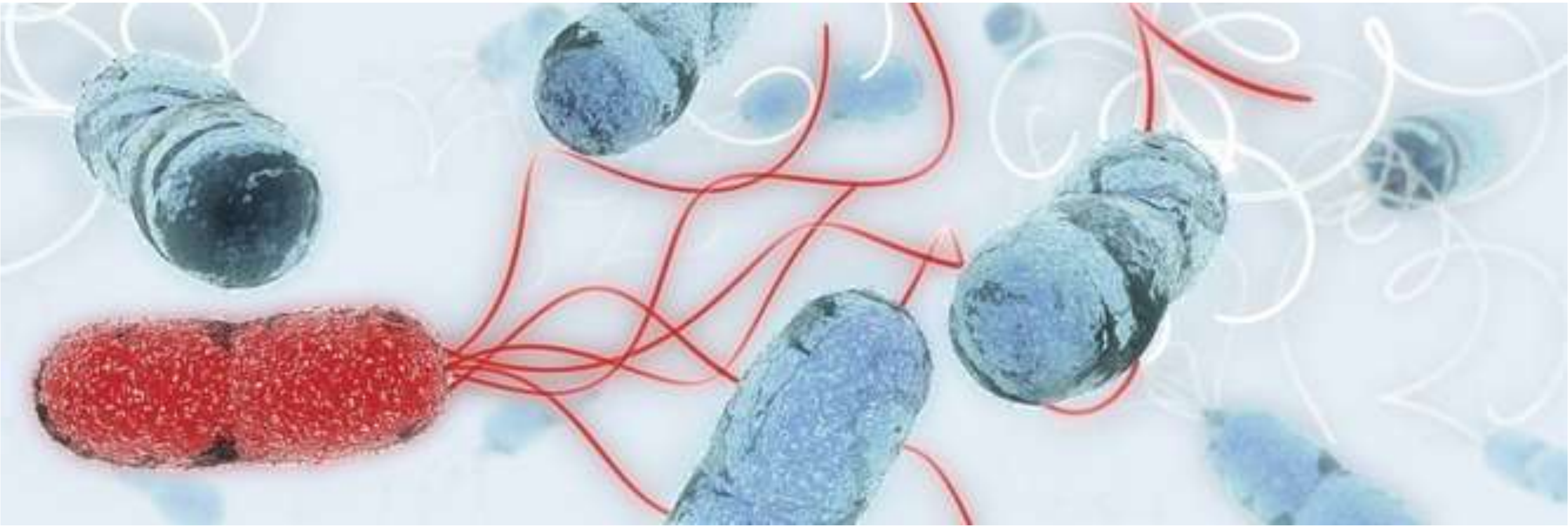
Aries

NGS solutions in Microbiology

(in 30 min)



Before Starting

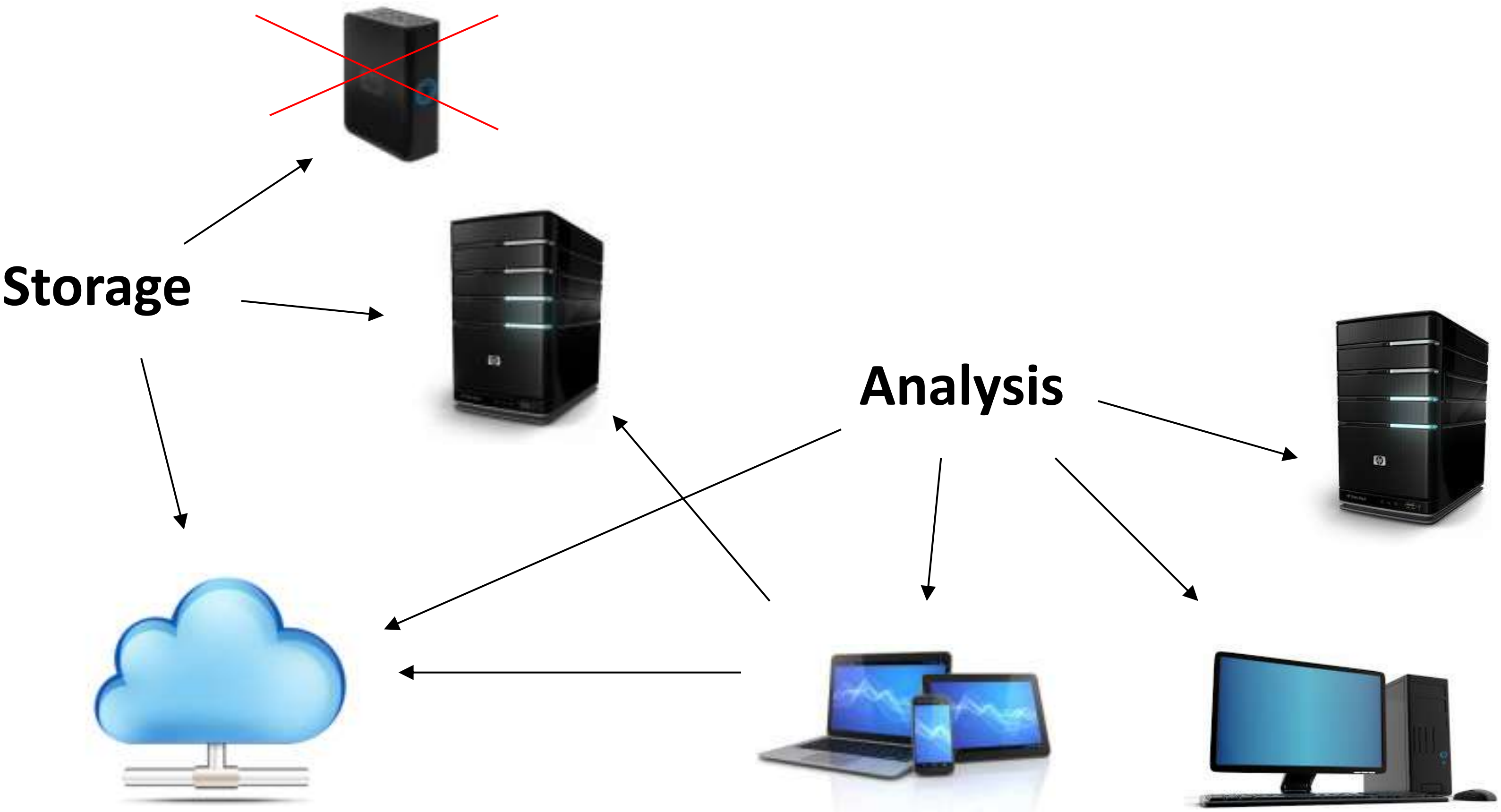
What I’ve to do with sequences?

What kind of results do I expect?

What are my applications?

What do my collaborators expect? Can I satisfy them?

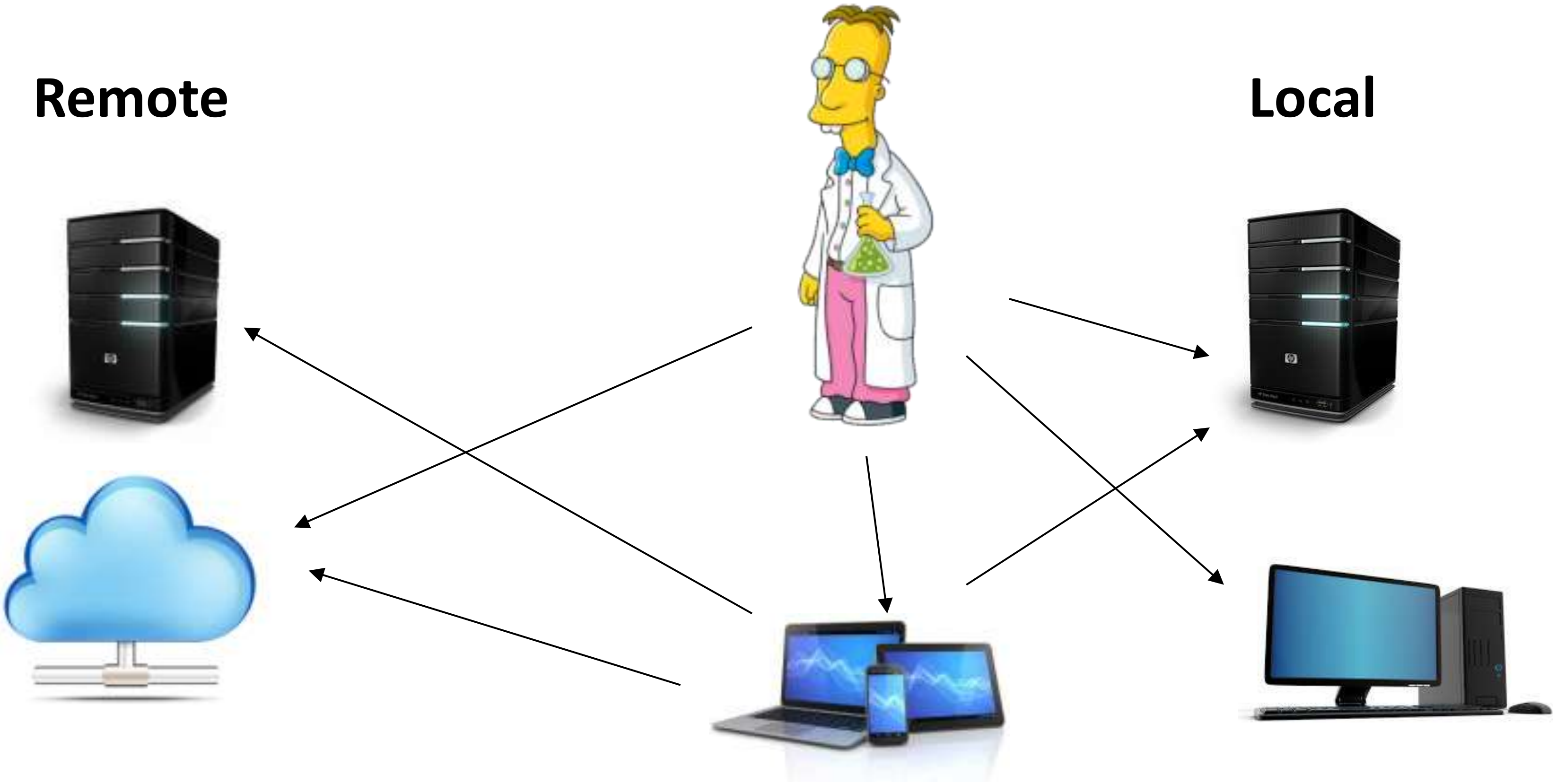
My brand-new sequencer has just produced 200GB of data… What can I do with them?



Storage

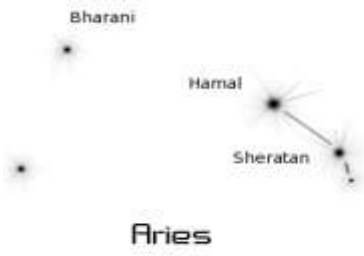
Analysis

Let’s focus on Analysis



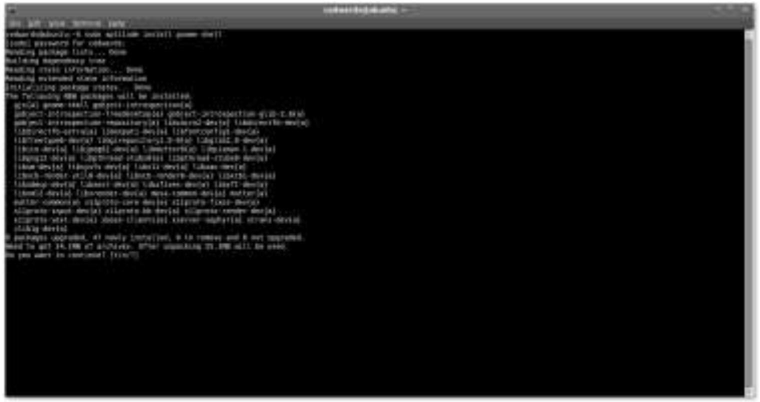
Remote

Local

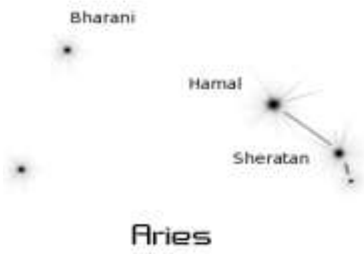


Remote Servers

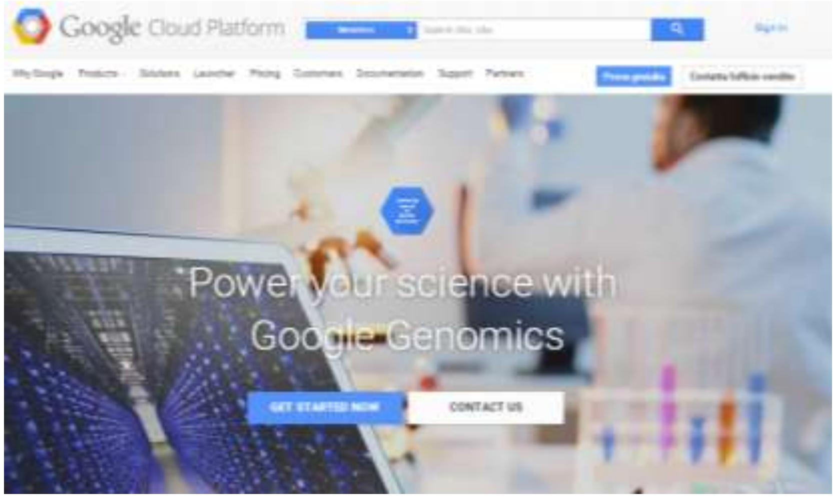
amazon
webservices™



Are you familiar with ssh?
Command line?
Modules environment?
Sftp?
Pay what you use…



Cloud Solutions



Are you familiar with API?
Low flexibility
High powerful resources
Do you trust in data security?
Reachable anywhere

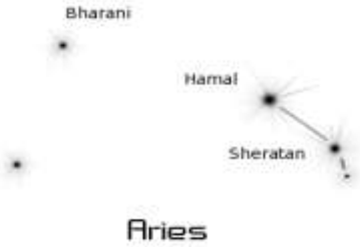
Local Solutions

Workstation

Small scale projects
Affordable costs
Low management efforts

Local Server

Higher costs, scalabilty
Storage
medium scale requests
People to manage it !!

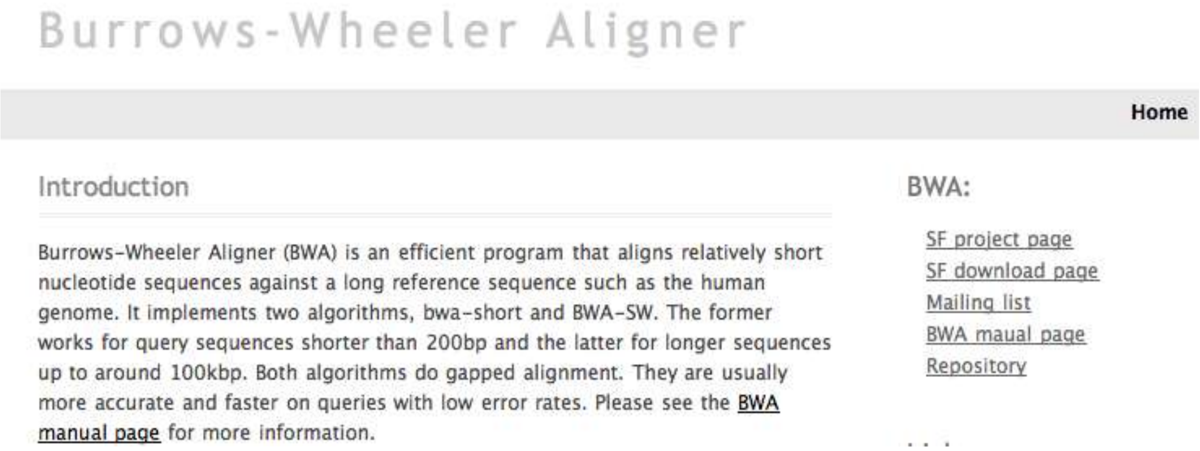


And Software???

CLC GENOMICS WORKBENCH

A comprehensive and user-friendly analysis package for analyzing, comparing, and visualizing next generation sequencing data.

Open or Commercial?



Burrows-Wheeler Aligner

Introduction

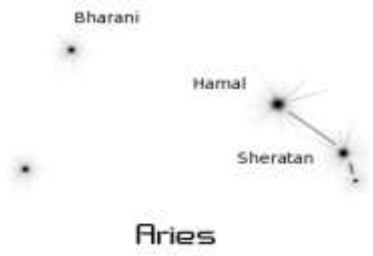
Burrows-Wheeler Aligner (BWA) is an efficient program that aligns relatively short nucleotide sequences against a long reference sequence such as the human genome. It implements two algorithms, bwa-short and BWA-SW. The former works for query sequences shorter than 200bp and the latter for longer sequences up to around 100kbp. Both algorithms do gapped alignment. They are usually more accurate and faster on queries with low error rates. Please see the BWA manual page for more information.

BWA:

- SF project page
- SF download page
- Mailing list
- BWA maual page
- Repository

. . .

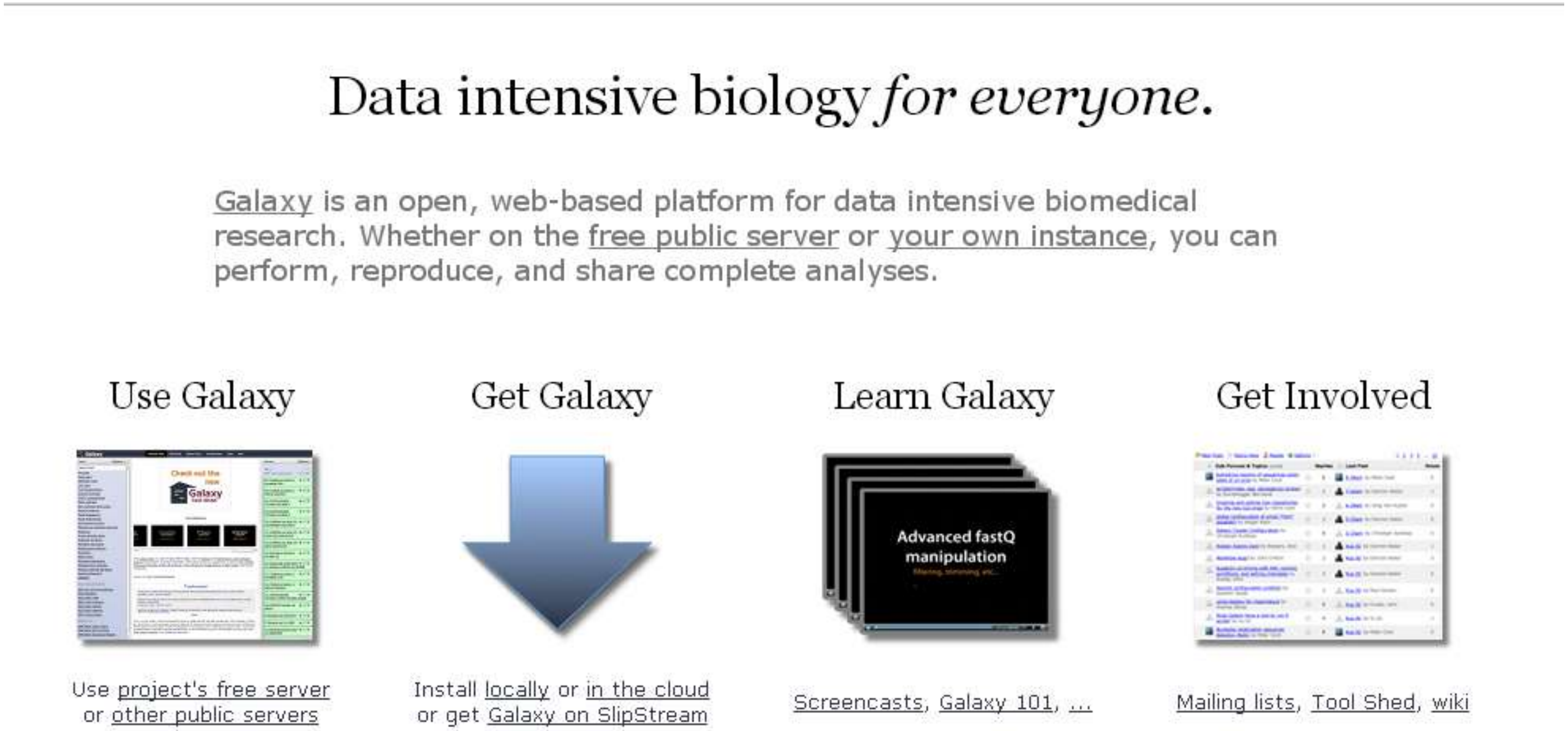
And People???



Did you think about these aspects?

Would you tell us your experience…?

Where is Galaxy framework in this scenario?



Data intensive biology *for everyone.*

Galaxy is an open, web-based platform for data intensive biomedical research. Whether on the free public server or your own instance, you can perform, reproduce, and share complete analyses.

Use Galaxy

Use project's free server
or other public servers

Get Galaxy

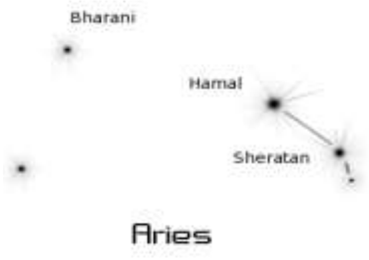
Install locally or in the cloud
or get Galaxy on SlipStream

Learn Galaxy

Screencasts, Galaxy 101, ...

Get Involved

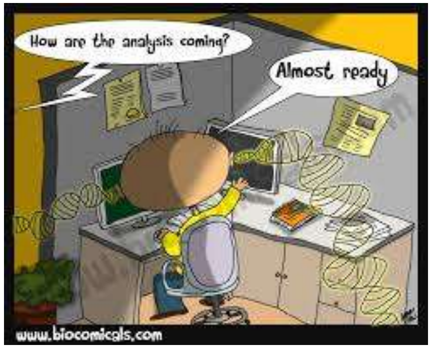
Mailing lists, Tool Shed, wiki

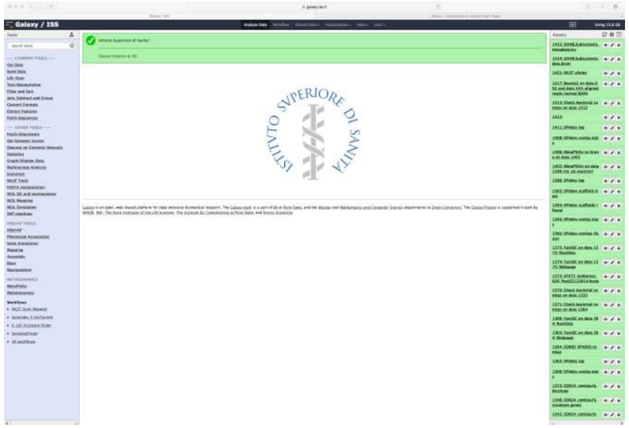


What is Galaxy?

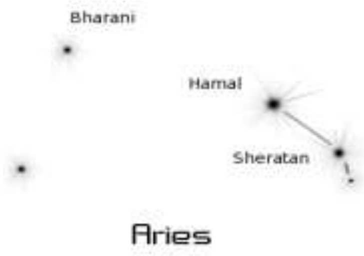
an open **web based** platform

- **Accessible**, reproducible, and transparent community standard
- **Personalized** data access and specialized tools for data analysis
- Use “Big Data” technologies to help **scale** computation





Layer between you and hardware resources



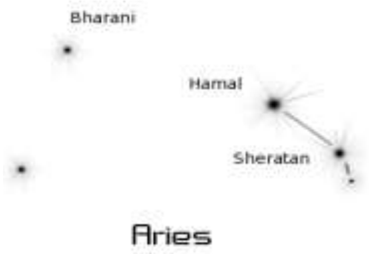
Galaxy Main Features

- Galaxy integrates a multitude of different tools by giving them the same “**look and feel**”

- **Histories** Every step of your analyses is recorded in Galaxy's history system…

- **Share your analyses** share histories between users

- **Results are reproducible** replace “the data were analyzed using a collection of in-house scripts” with an url pointing to Galaxy's history



Galaxy Specialized Istances

Public Galaxy Servers
and *still* counting

http://gcc2015.tsl.ac.edu
4-8 July
Register Now

Use Galaxy

Servers • Learn
Main • Choices
Share • Search

Communicate

Support • Biostar
Events • Mailing Lists

Publicly Accessible Galaxy Servers

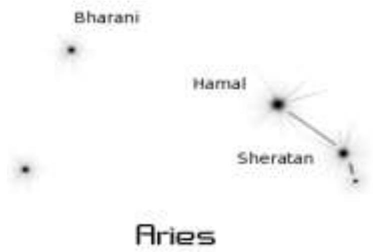
The Galaxy Project's public server (UseGalaxy.org, *Main*) can meet many needs, but it is not suitable for everything (see Choices for why) and cannot possibly scale to meet the entire world's needs.

Fortunately the Galaxy Community is helping out by installing Galaxy at

Indice

1. General Purpose Servers
   1. Andromeda
   2. Biomina
   3. CBiB Galaxy

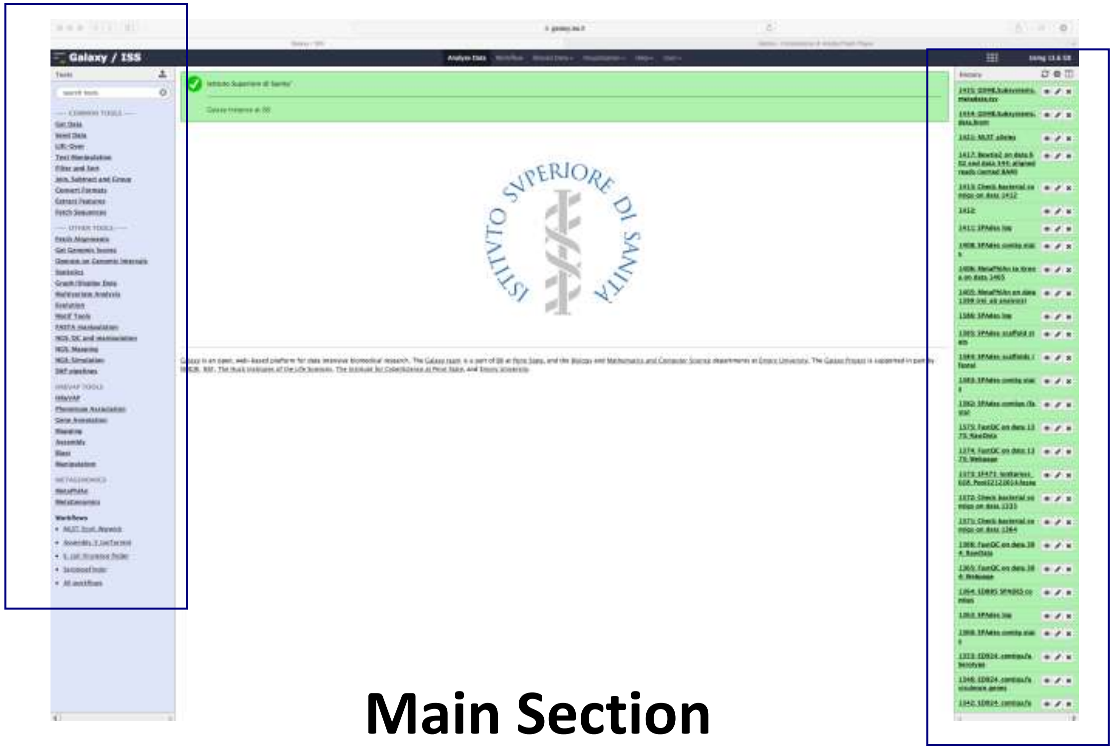
https://wiki.galaxyproject.org/PublicGalaxyServers



How does ARIES look like?

https://aries.iss.it

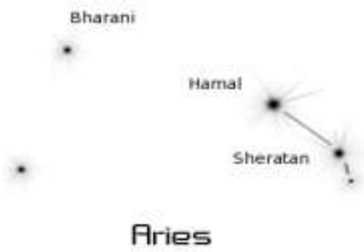
Aries’ three main sections



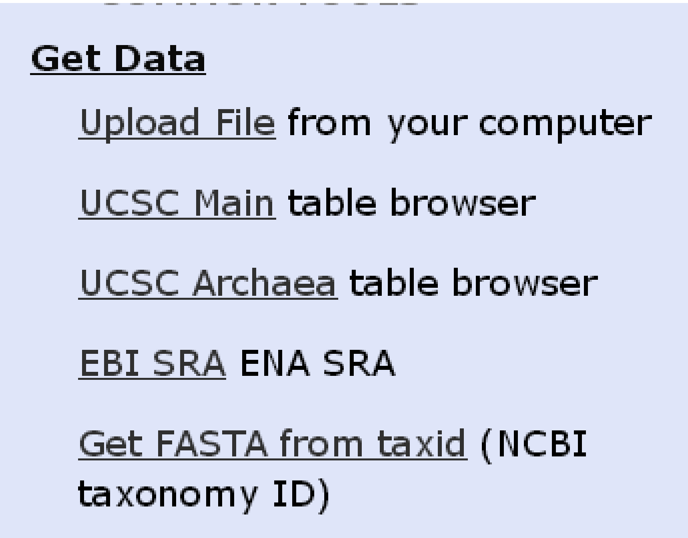
tools

History

Main Section
Input/Results



Are you ready?



Get Data

- Upload File from your computer
- UCSC Main table browser
- UCSC Archaea table browser
- EBI SRA ENA SRA
- Get FASTA from taxid (NCBI taxonomy ID)

GO!
…load data in your histories…