

Galaxy Workflows on the EGI Federated Cloud

Miguel Caballer, Ignacio Blanquer (*)
Alain Franc, Jean-Marc Frigerio ()**



(*) Universitat Politècnica de València

(**) INRA – French National Institute for Agricultural Research



- Use case of Galaxy
 - Use case problem.
 - Galaxy architecture.
- Virtual elastic cluster approach
 - Components
 - Architecture for Galaxy
- Video demo (UPV)
 - www.youtube.com/watch?v=qJz5HRsApSI
- Medium-term plans

Applied Mathematics and Informatics for Biodiversity

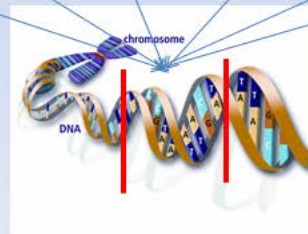
Few individuals



Many traits : genome wide cover



Many individuals



Few DNA regions of interest

Which tools do we have?

ADDED TOOLS

virtual BiodiversityL@b

MAFFT alignment (basic options)

BMGE block selection

Fasta to Phylip conversion

Fasta to Nexus conversion

PHYML wrapper (maximum likelihood phylogenies)

Draw ascii tree (newick)

BEAST wrapper

DECLIC

Disseq

Distance ape

Declic Graph

Declic MDS

DIAGNO-SYST (HIDDEN TOOLS)

Diagno-Syst char format
Formats charfile

Diagno-Syst

Disseq reads to taxo

Taxonomic inventory from

Alignment (multiple)

Phylogeny

Distance based methods

Smith-Waterman (exact)

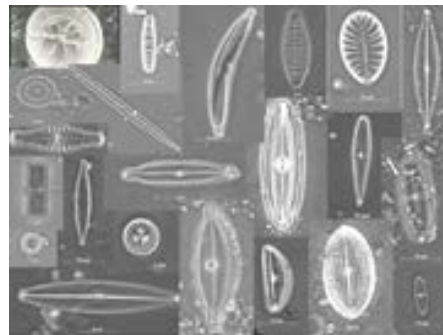
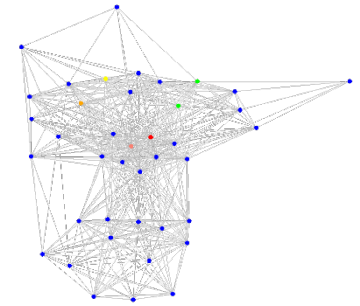
Evolution models

Graph clustering

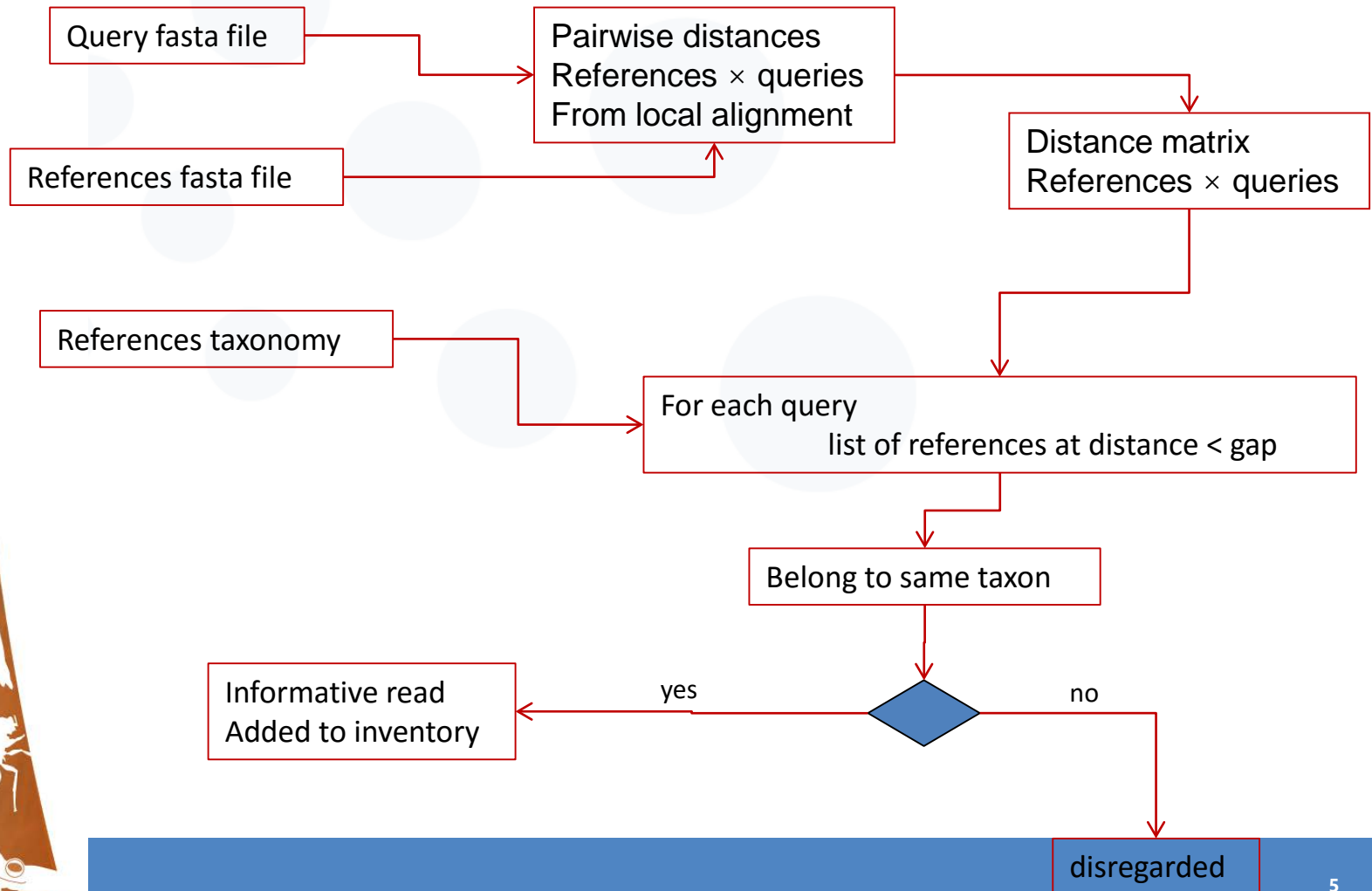
MDS

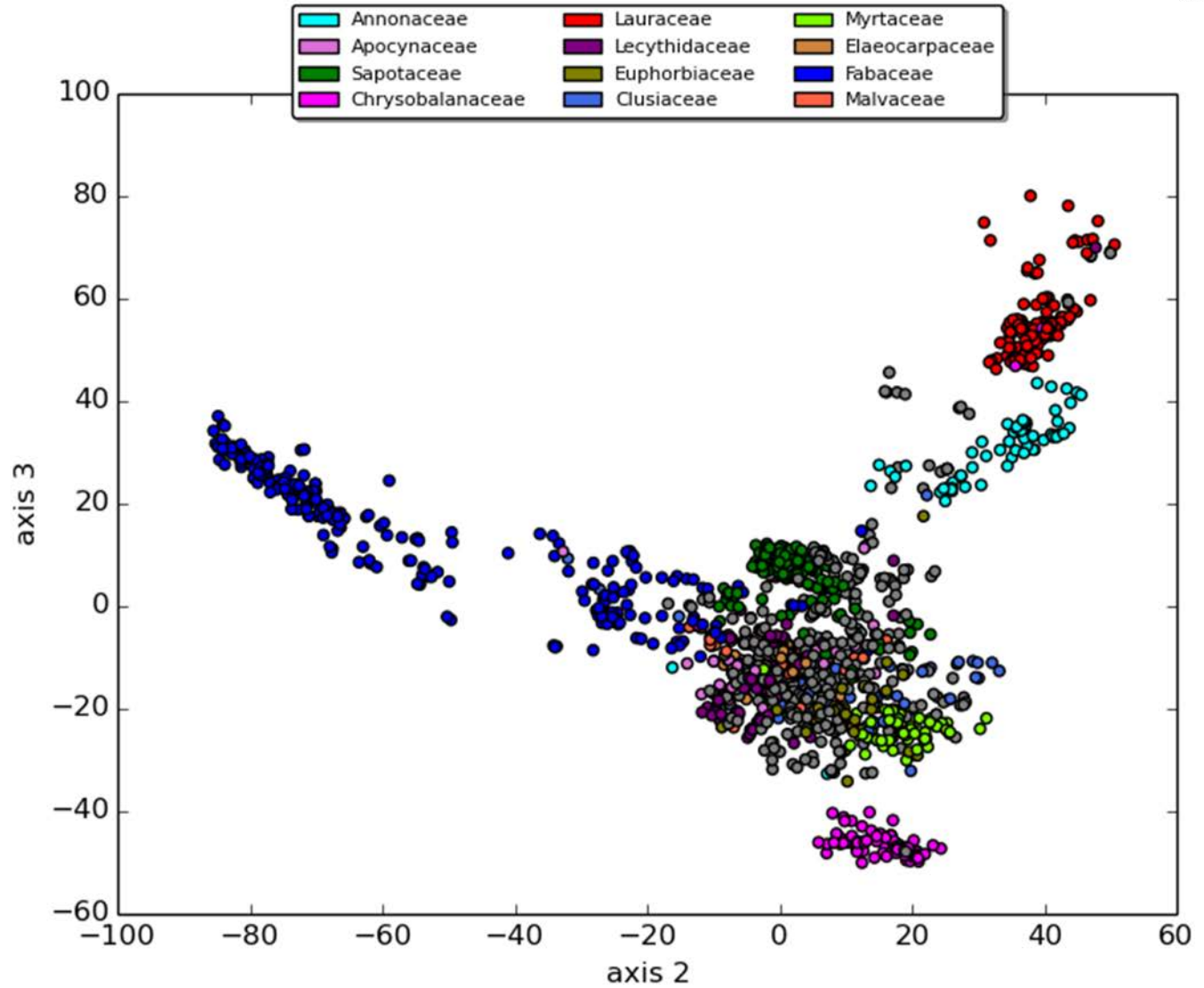
Aggregative clustering

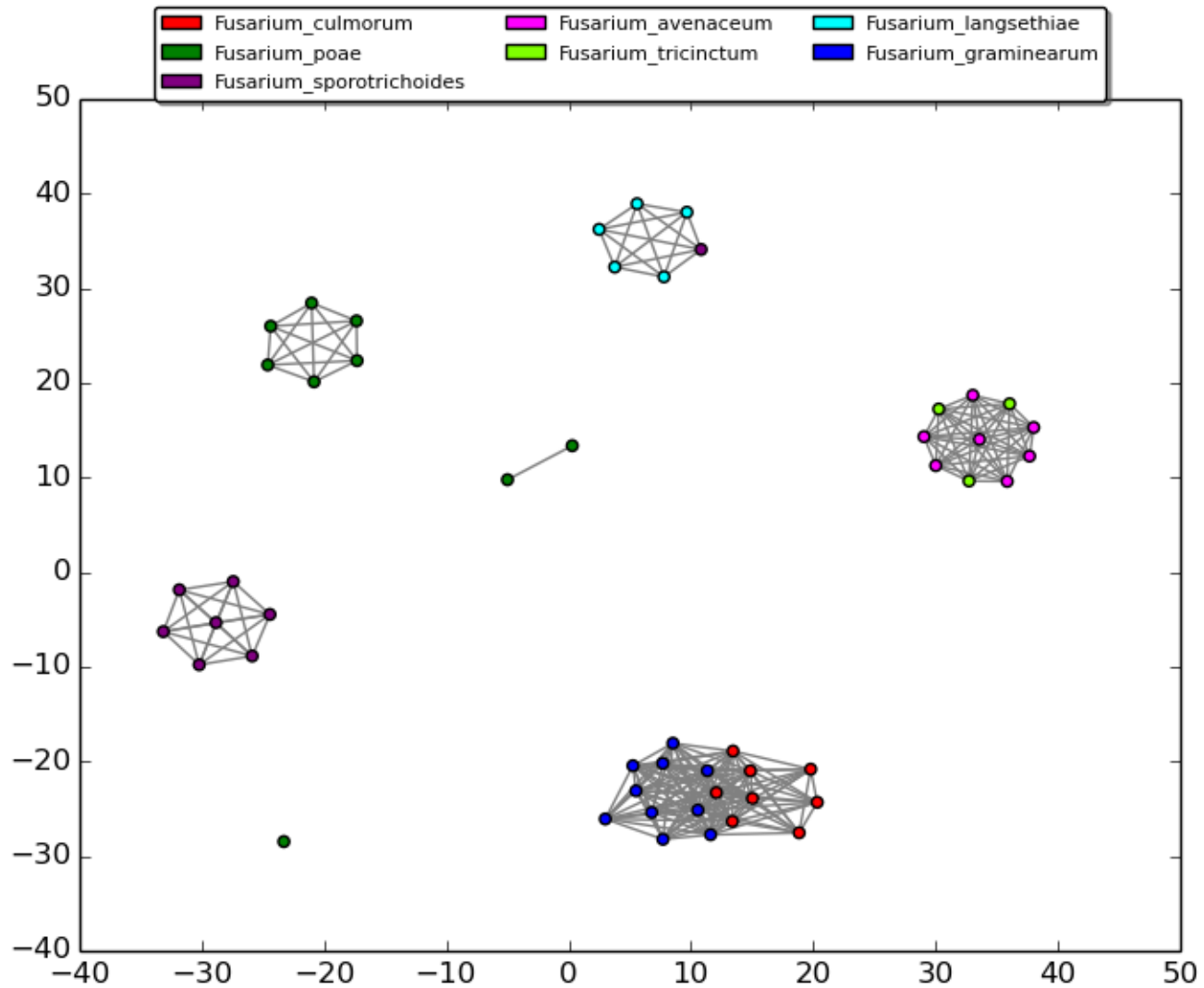
Diatoms_18S ; cc = 4 ; character = species ; gap = 3



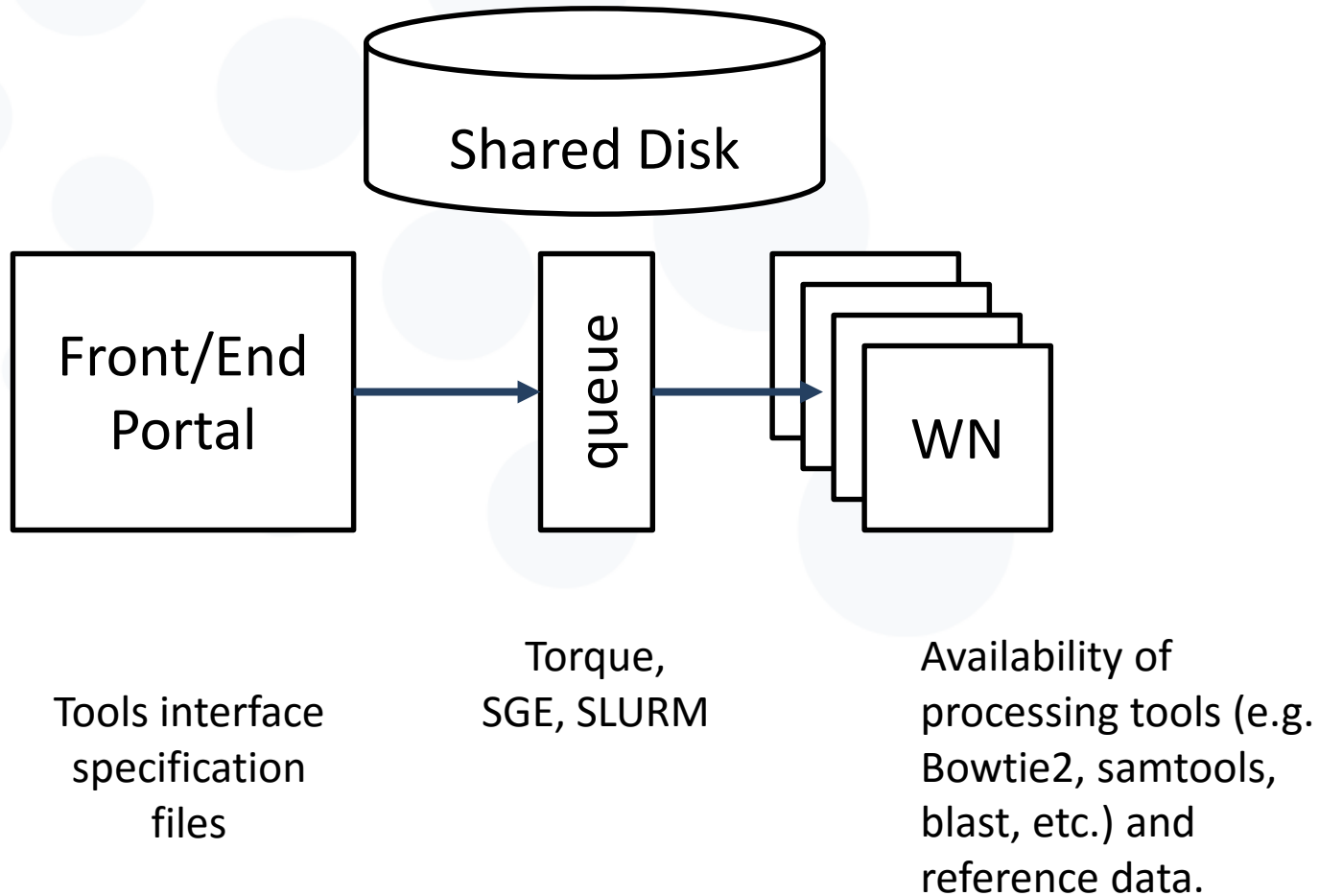
An example for taxonomic Annotation from NGS







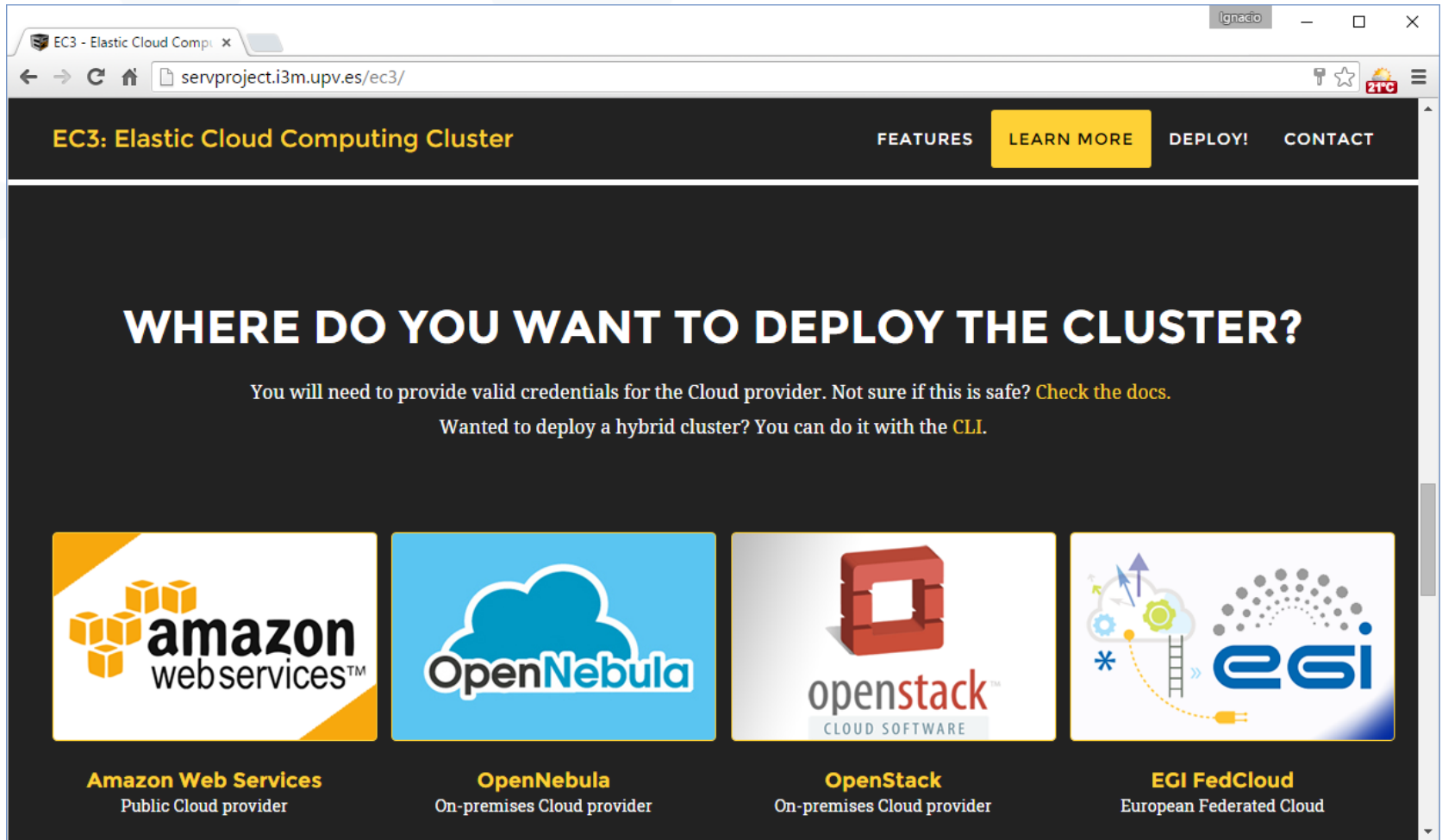
- Galaxy can be installed locally on a machine or as a front/end to a batch queue.
- Galaxy exposes a web interface and executes all the interactions (including data uploading) as jobs in a batch queue.
- Requires a shared directory among the working nodes and the front/end.
- It supports a separate storage area for different users, managing them through the portal.







- Galaxy on the cloud (<https://wiki.galaxyproject.org/Cloud>) is a branch using CloudMan (<https://wiki.galaxyproject.org/CloudMan>) for the automation of the deployment.
- It can be easily deployed on Amazon AWS (there are already pre-configured VMIs)
 - It can work with Ostack and ONE but you need to create the VMI.
 - However it still requires the manual configuration of the individual VMs and elasticity is not fully automatic.
 - No support for OCCl.

Virtual Elastic Cluster: Components

- EC3 (Elastic Compute Cluster – www.grycap.upv.es/ec3) & IM (Infrastructure Manager – www.grycap.upv.es/im) are tools compatible with EGI Federated Cloud that facilitate the configuration and management of virtual clusters
 - Fully automation of the configuration process.
 - Automatic Elasticity based on the LRMS queue size.
 - Compatible with multiple providers (including OCCI interface).
 - It can use plain “vanilla” Linux images and configure on the fly
 - Therefore the same configuration recipe works for different cloud providers.



The screenshot shows a web browser window with the URL `servproject.i3m.upv.es/ec3/`. The page title is "EC3: Elastic Cloud Computing Cluster". The navigation menu includes "FEATURES", "LEARN MORE" (highlighted in yellow), "DEPLOY!", and "CONTACT". The main heading is "WHERE DO YOU WANT TO DEPLOY THE CLUSTER?". Below the heading, there is a note: "You will need to provide valid credentials for the Cloud provider. Not sure if this is safe? [Check the docs.](#)" and another note: "Wanted to deploy a hybrid cluster? You can do it with the [CLI.](#)". The page features four deployment options in a grid:

Provider	Type
 Amazon Web Services Public Cloud provider	Public Cloud provider
 OpenNebula On-premises Cloud provider	On-premises Cloud provider
 OpenStack On-premises Cloud provider	On-premises Cloud provider
 EGI FedCloud European Federated Cloud	European Federated Cloud

- It uses a single VMI image (plain Ubuntu 14.04 LTS).
- It requires to have the ec3 client and the galaxy configuration recipes
 - They can be downloaded from <https://github.com/grycap/ec3>.
 - Currently we support main galaxy framework + bowtie2 + samtools + reference data (Drosophila Melanogaster) for demonstration
 - More tools can be added to the automation system (preferred) or manually after the deployment.
 - The version for INRA also includes 3 processing tools from the previous workflow.
 - Including the back-end processing tools and the XML interface files for the Front / End.
- Credentials must be provided in the proper file.

- www.youtube.com/watch?v=qJz5HRsApSI



EC3 Demo: Galaxy Portal with Torque LRMS on EGI FedCloud site

Grid y Computación de Altas Prestaciones
GRyCAP

 **Elastic Cloud Computing Cluster**

Demo

Galaxy Portal + Torque Cluster
On EGI FedCloud infrastructure

<http://www.grycap.upv.es/ec3>

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

i3M

0:03 / 7:28

- Extend the catalogue of services supported.
- Support for a wider catalogue of tools and workflows.
- Instructions to facilitate the customisation of the system by new users.
- Support to MPI.
- Intensive testing on multiple users.

Thank you for your attention.

Questions?



www.egi.eu

This work by Parties of the EGI-Engage Consortium is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

