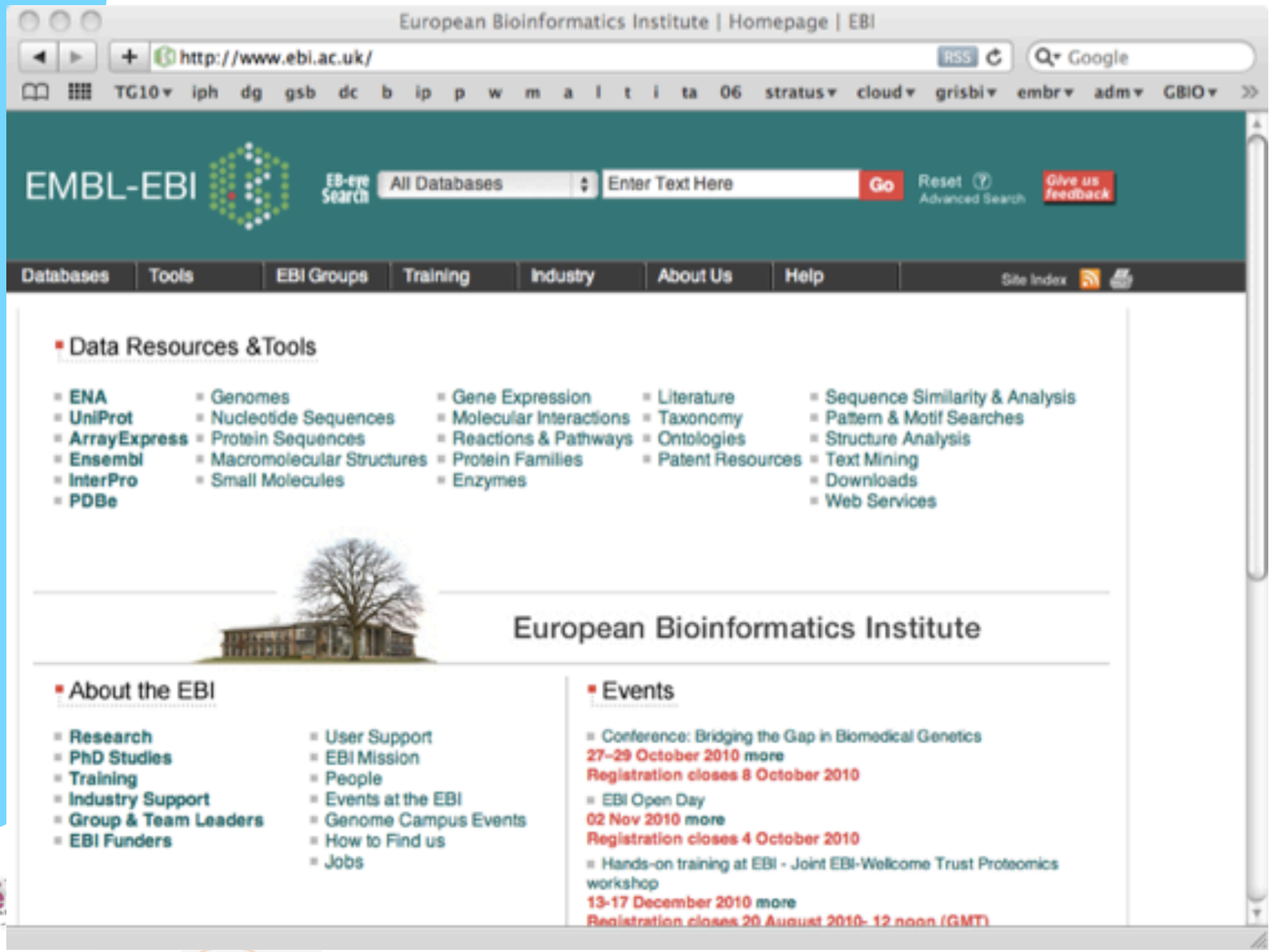
Bioinformatics applications

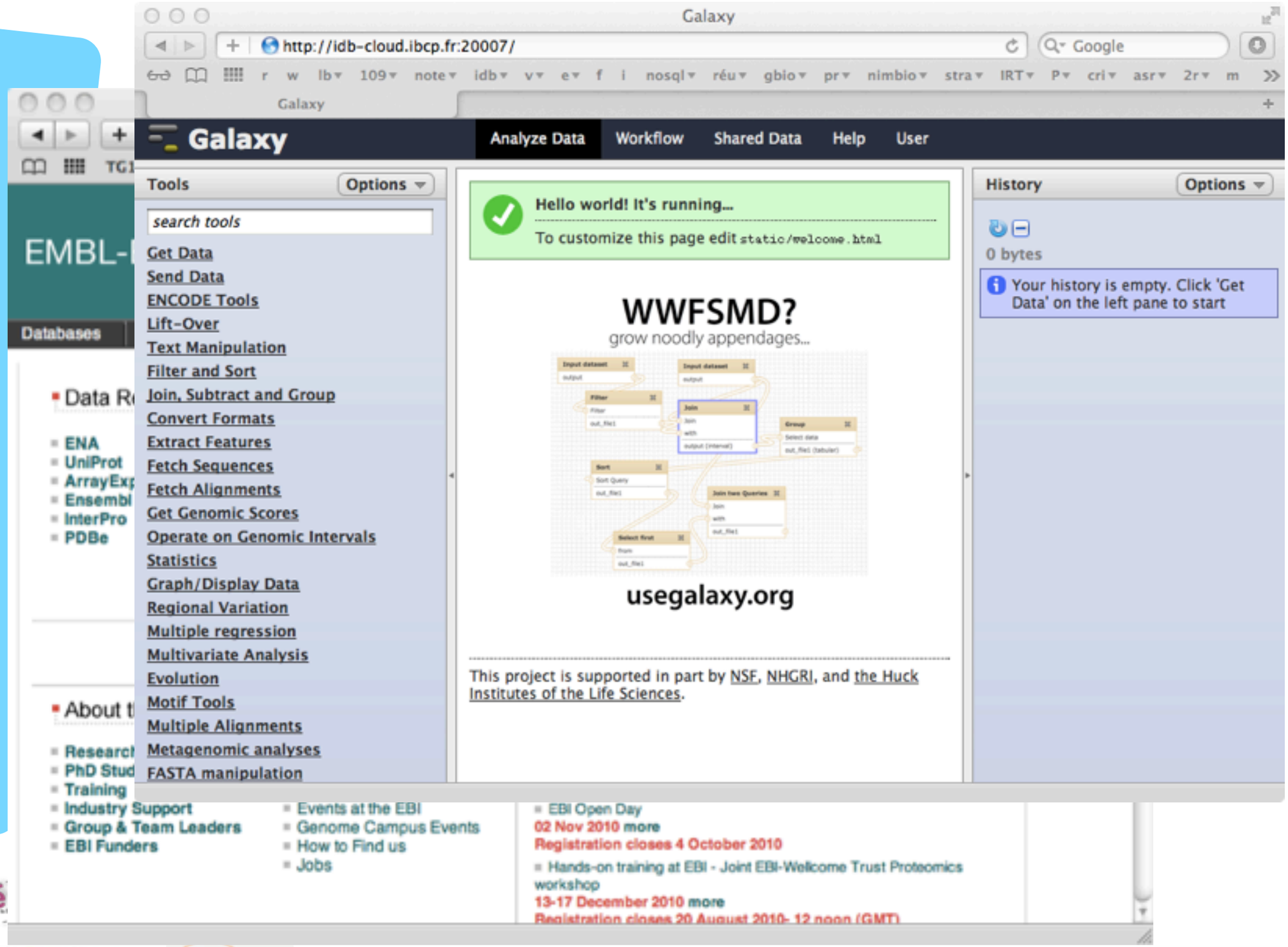
StratusLab Tutorial (Bordeaux, France)

25-26 October 2012

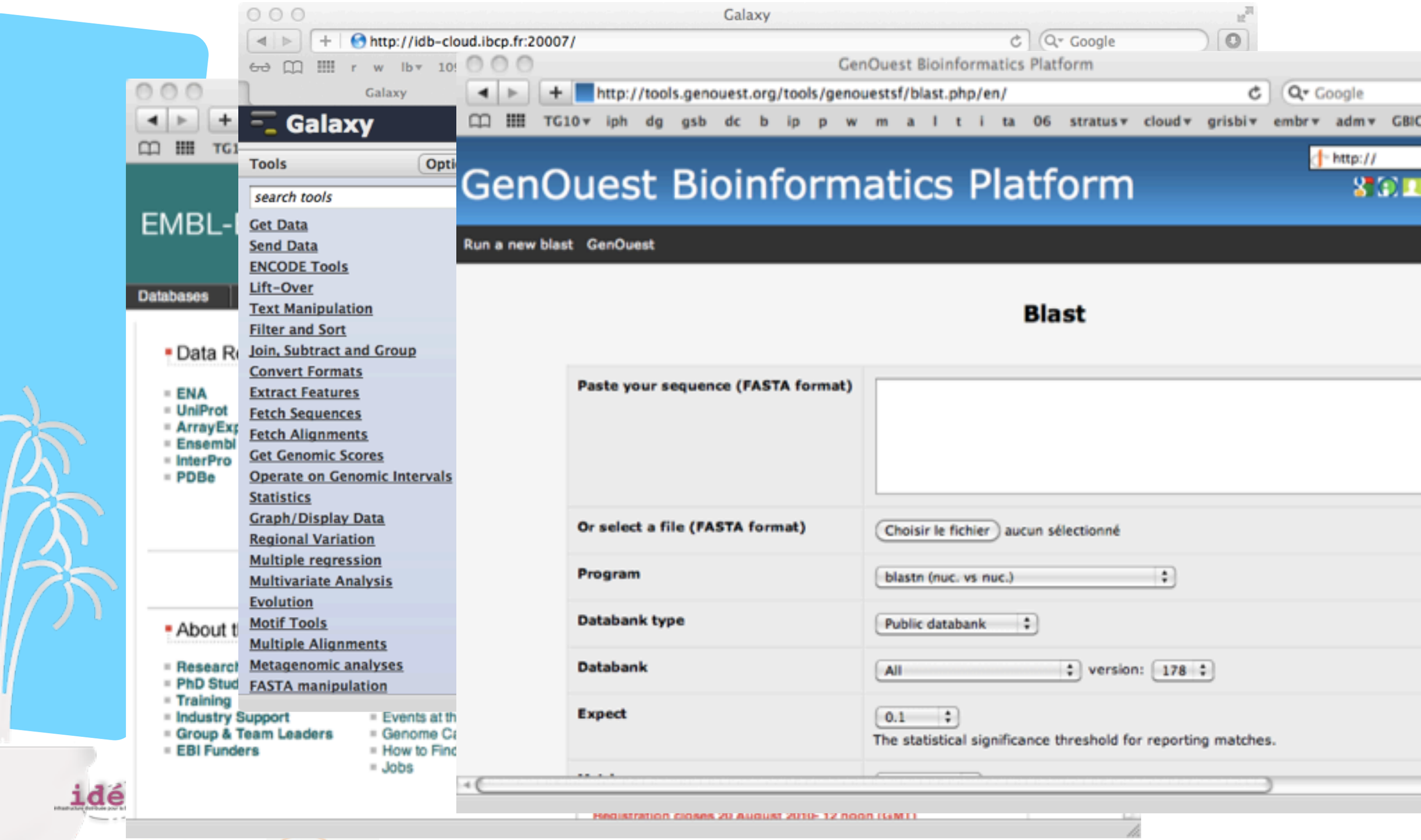
Infrastructure in Biology



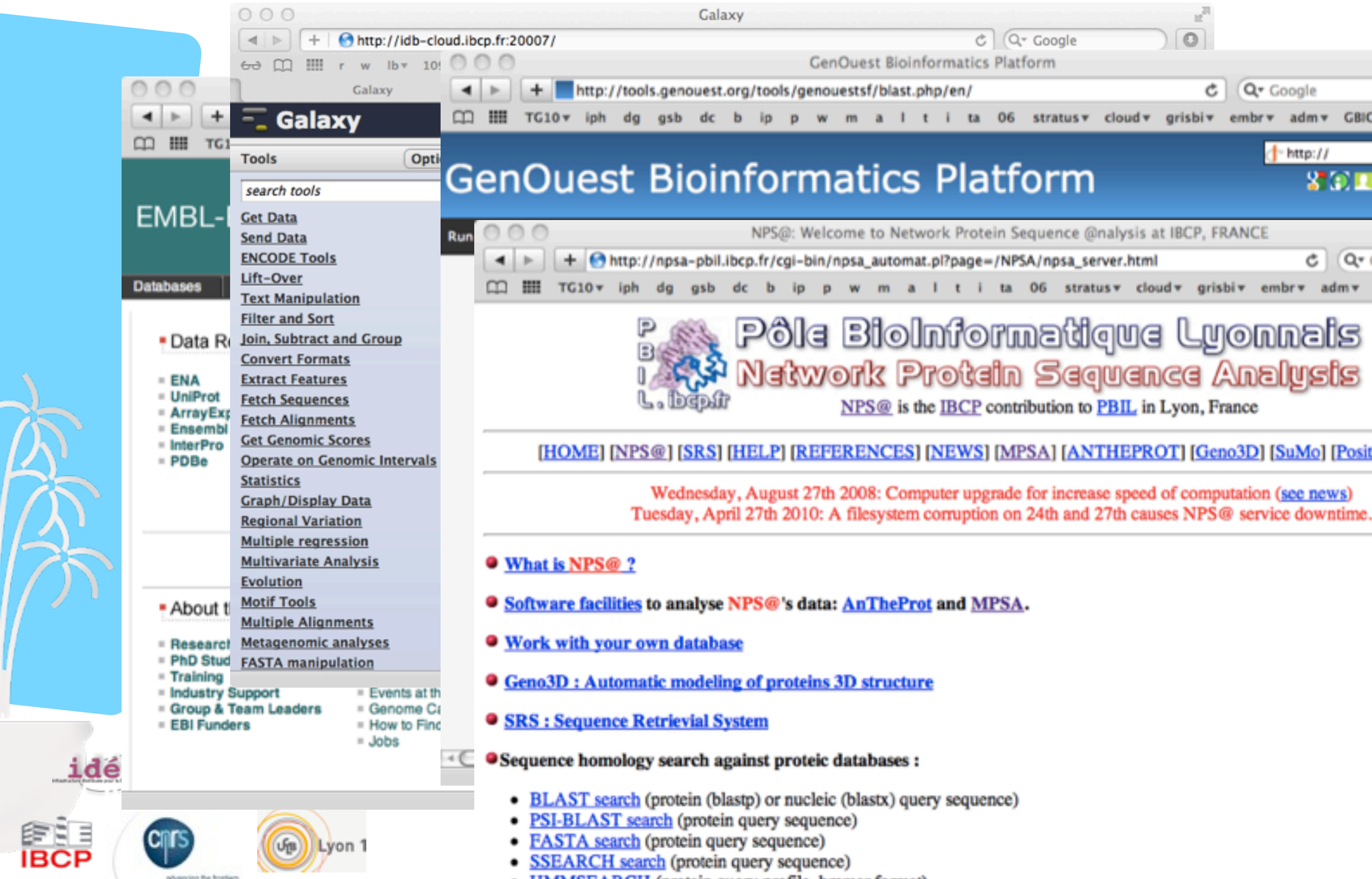
Infrastructure in Biology



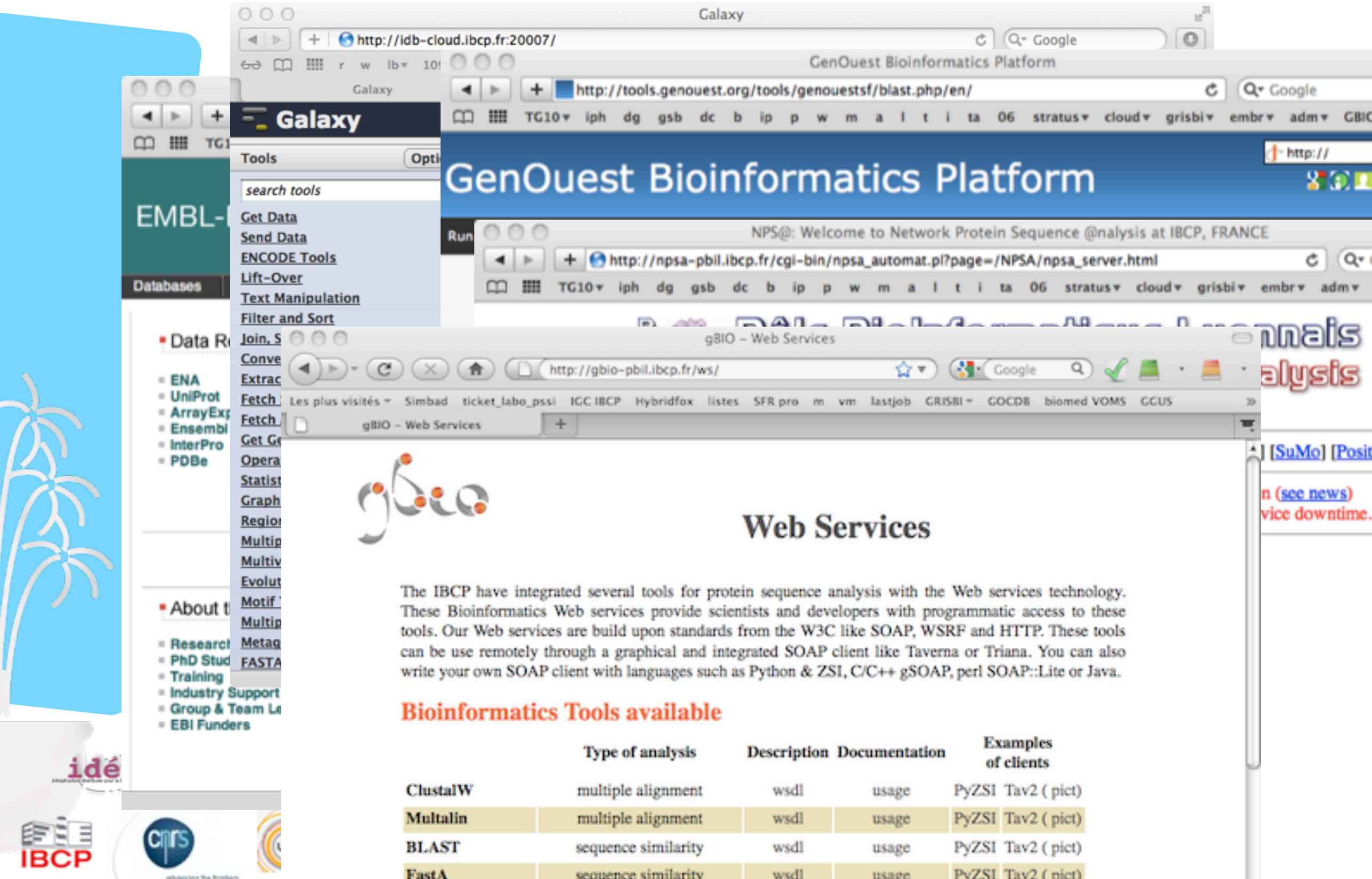
Infrastructure in Biology



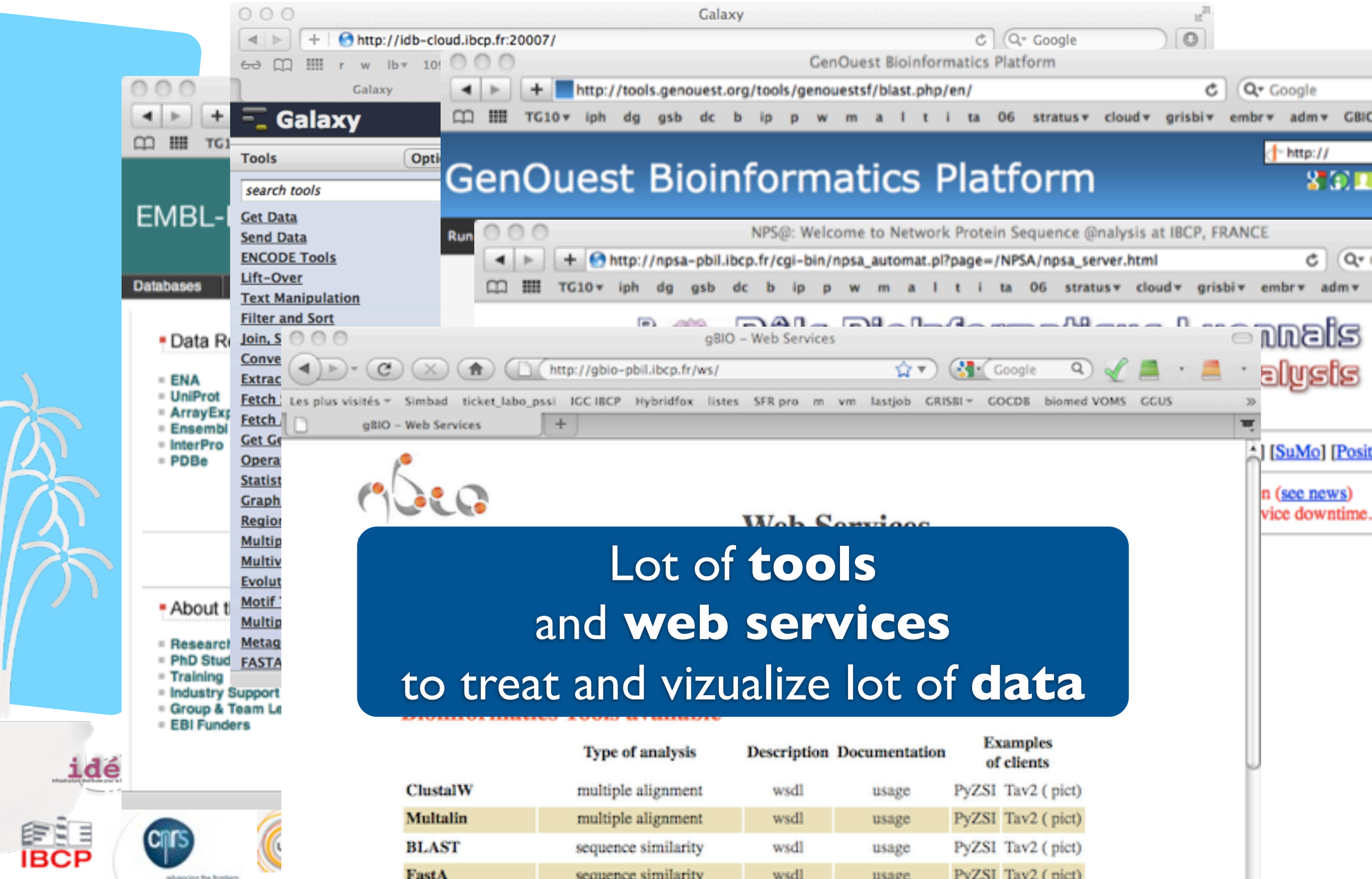
Infrastructure in Biology



Infrastructure in Biology



Infrastructure in Biology

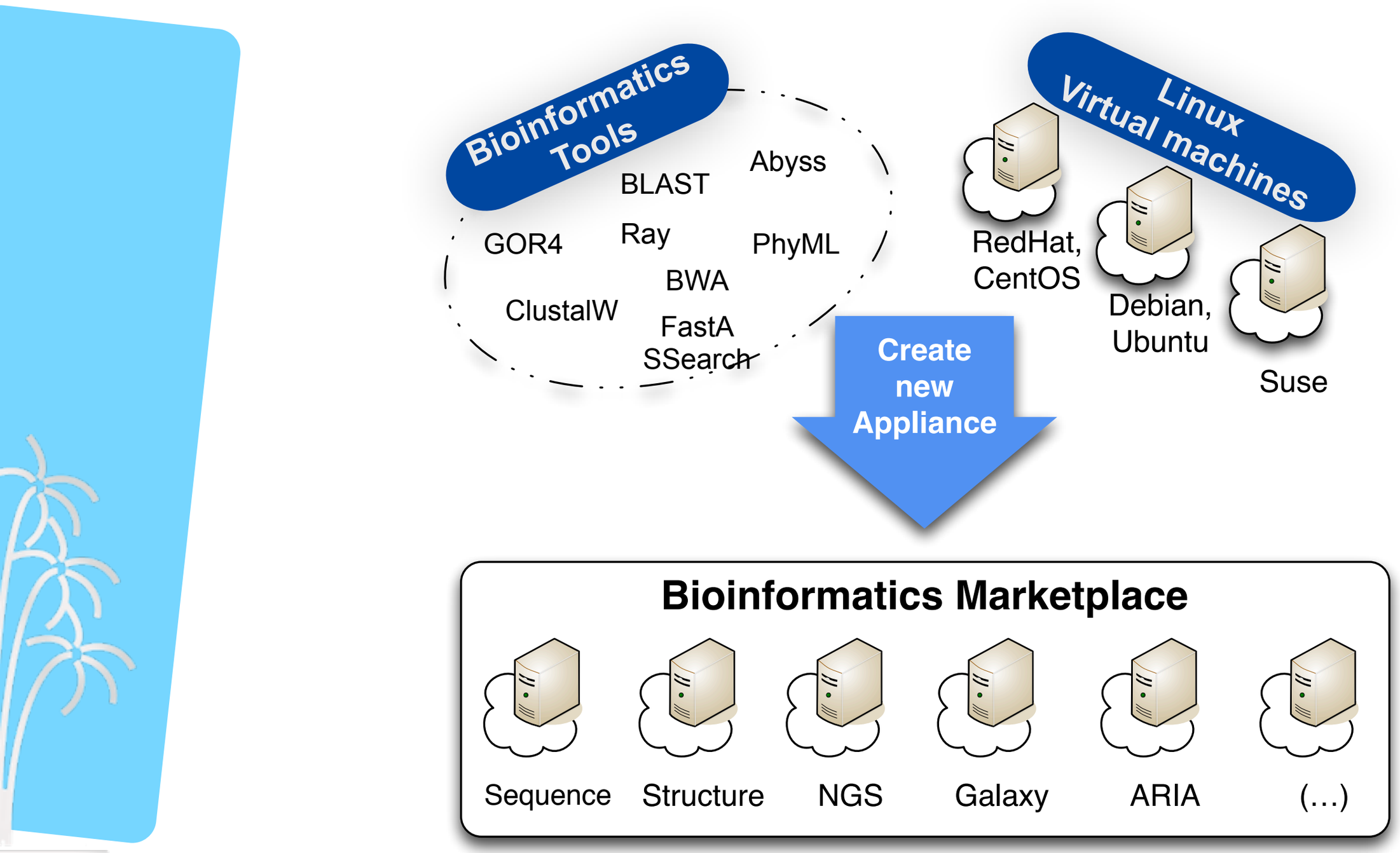


Constraints

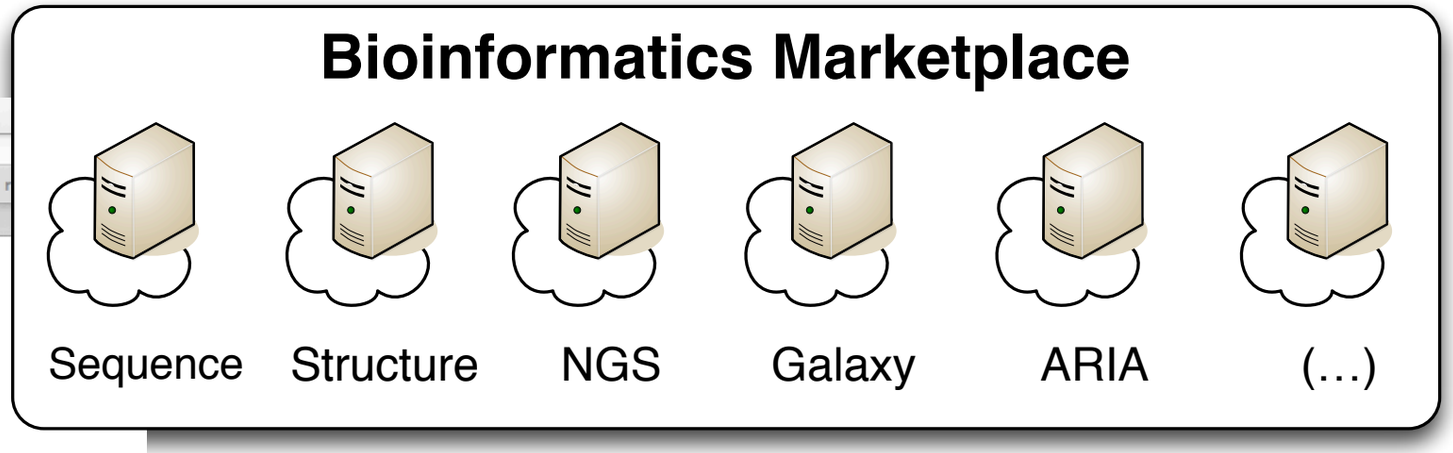
- Core facility providers
  - How to deploy easily lot of (incompatible) tools ?
  - To make them connected to public databases ?
  - To limit transfer of huge data ?
  - To provide users with their own computing resources ?
  - With their own isolated storage ?

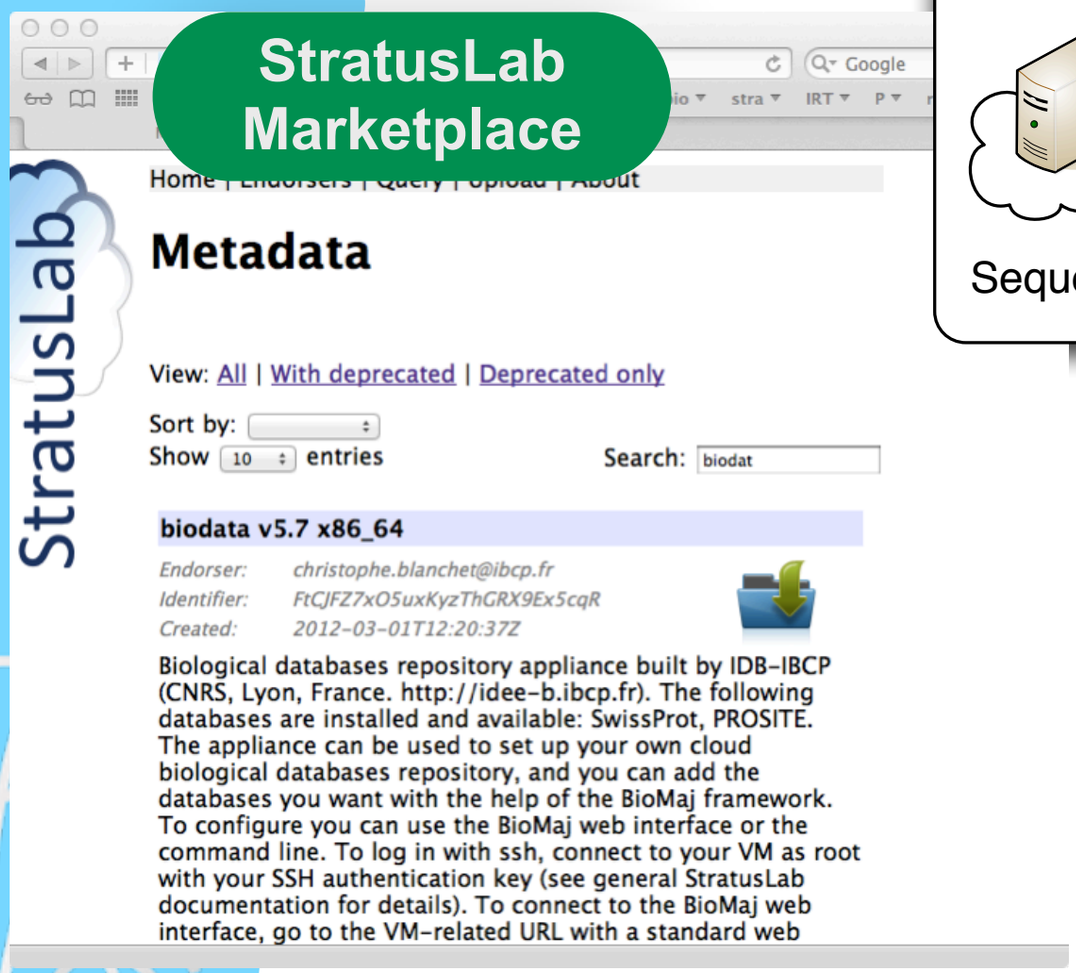
- Scientific users
  - How easy is it to access/use these tools ?
  - To adapt to your usage ?
  - To get your/other tools deployed on a datacenter ?
  - To combine them ?

Integrate Bioinformatics Tools in Cloud



Appliances for Bioinformatics

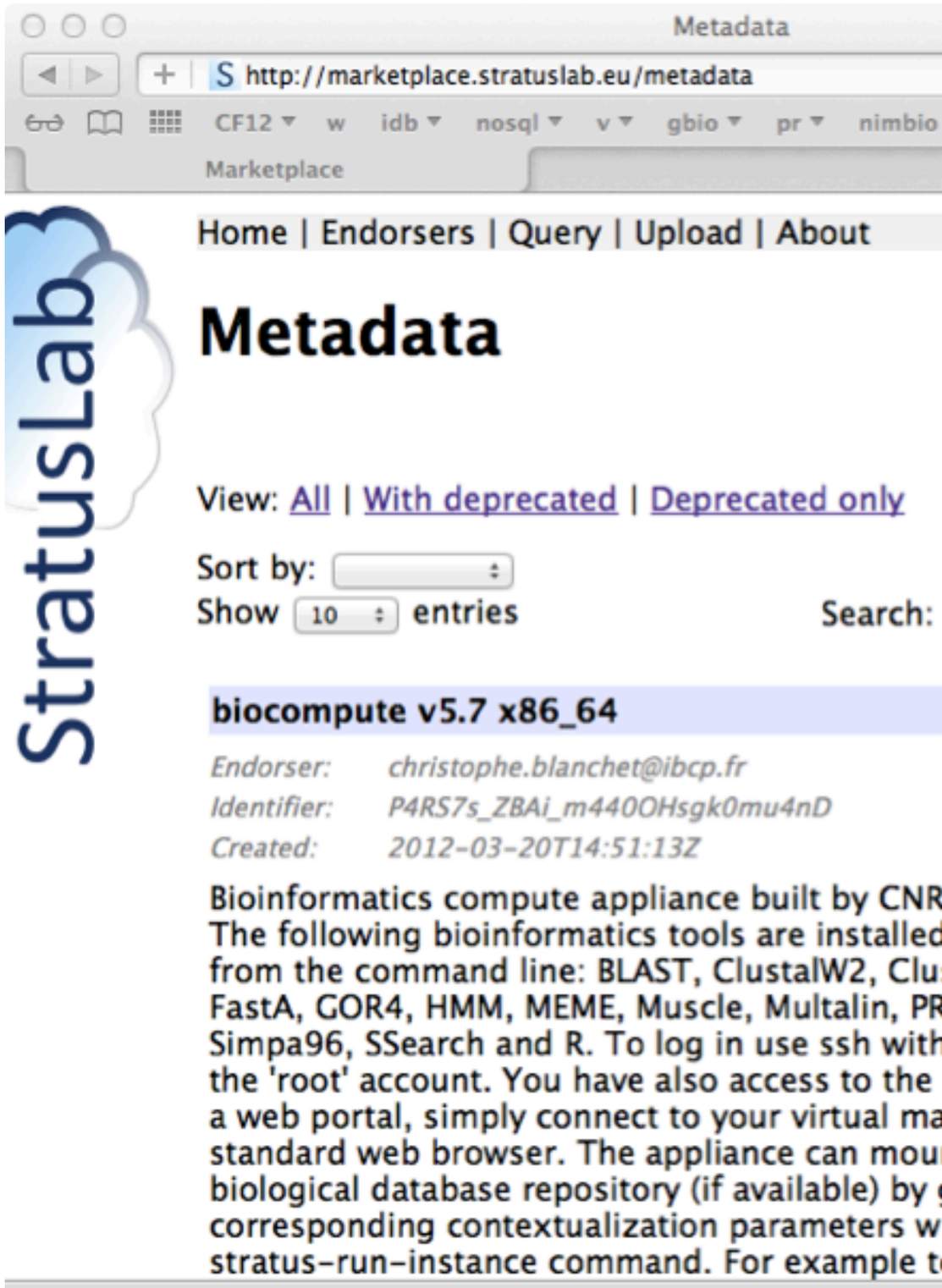




- Predefined virtual machines
  - small: few GB
  - easy to convert in most formats
- Installed and pre-configured bioinformatics tools
- Several ones already available
  - biodata, biocompute, galaxy, etc.
- with several tools
  - BLAST, Clustalw, ARIA, MEME, HMMer, Abyss, BWA, Ray, etc.

Example of an Appliance

- **‘biocompute’ appliance**
  - pre-installed with bioinformatics tools:
    - BLAST, FastA, SSearch,HMM,
    - ClustalW2, Clustal-Omega, Muscle, Multalin,
    - GOR4, PREDATOR, Simpa96,
    - MEME, R.
  - Deploy **your own** instances
    - **SSH** as root
    - **HTTP** with a standard web browser.
  - Connected to **local public** biological databases
    - Automaticaly parametrized
    - For example on the IBCP cloud
      - BIO_DB_SERVER=idb-databases.ibcp.fr



Bioinformatics Appliances

Protein identification

Endorser: *christophe.blanchet@ibcp.fr*
Identifier: *H6KPqxYlZRdlhPhs2ZKIENiiVyx*
Created: *2012-10-23T14:05:07Z*

Bioinformatics virtual appliance for protein identification from mass spectrometry data. Contains OMSSA, X!Tandem, PeptideShaker and SearchGUI tools. Constructed by IDB...

More...

Galaxy portal

Endorser: *christophe.blanchet@ibcp.fr*
Identifier: *OqucGN3bQD9FdIenGRlqZ4ZNNHW*
Created: *2012-10-11T15:11:59Z*

Bioinformatics gateway appliance configured with the GALAXY portal, built by CNRS IBCP-IDB. You have also access to the pre-installed bioionformatics tools through the web...

More...

Hadoop MapReduce

Endorser: *clement.gauthey@ibcp.fr*
Identifier: *PElfkAp5mOwULVh1KLsprFcji0s*
Created: *2012-10-11T14:42:36Z*

This appliance provides an easy way to deploy an Hadoop MapReduce cluster. You just need to run the bash script hadoop-create-cluster with a nodes list and an username in parameters and wait few...

More...

BIO compute node

Endorser: *christophe.blanchet@ibcp.fr*

Bioinformatics Appliances (...)

ARIA 2.3

Endorser: *christophe.blanchet@ibcp.fr*
Identifier: *N_zDsconV86gvkjZtt7D-ePv4M6*
Created: *2012-10-11T14:01:38Z*

This appliance is part of the StratusLab bioinformatics usecase TOSCANI (TOwards StruCtural AssignmeNt Improvement). The goal is to improve the determination of protein...

More...

CentOS 6

Endorser: *christophe.blanchet@ibcp.fr*
Identifier: *Bl8HibMjBB6uu231adQUkqyGtnI*
Created: *2012-09-27T12:04:15Z*

A minimal installation for CentOS 6.x. Only root account configured. Firewall enabled with SSH and HTTP port open. SELinux disabled. Enhanced StratusLab contextualization used...

More...

BIO data

Endorser: *christophe.blanchet@ibcp.fr*
Identifier: *FtCJFZ7xO5uxKyzThGRX9Ex5cqR*
Created: *2012-09-24T08:56:26Z*

Biological databases repository appliance built by IDB-IBCP (CNRS, Lyon, France. http://idee-b.ibcp.fr). The following databases are installed and available: SwissProt,...

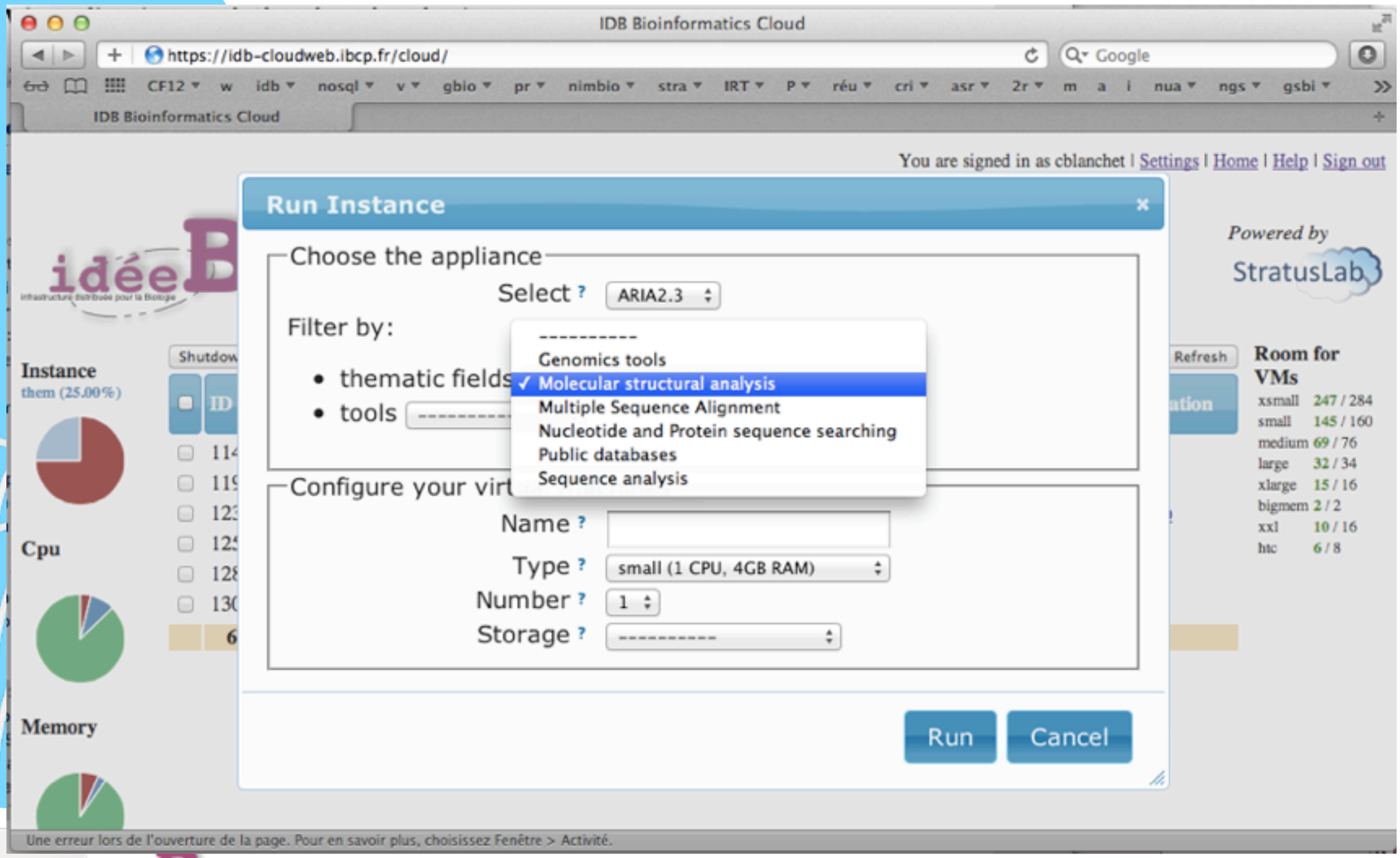
More...

Mobyle

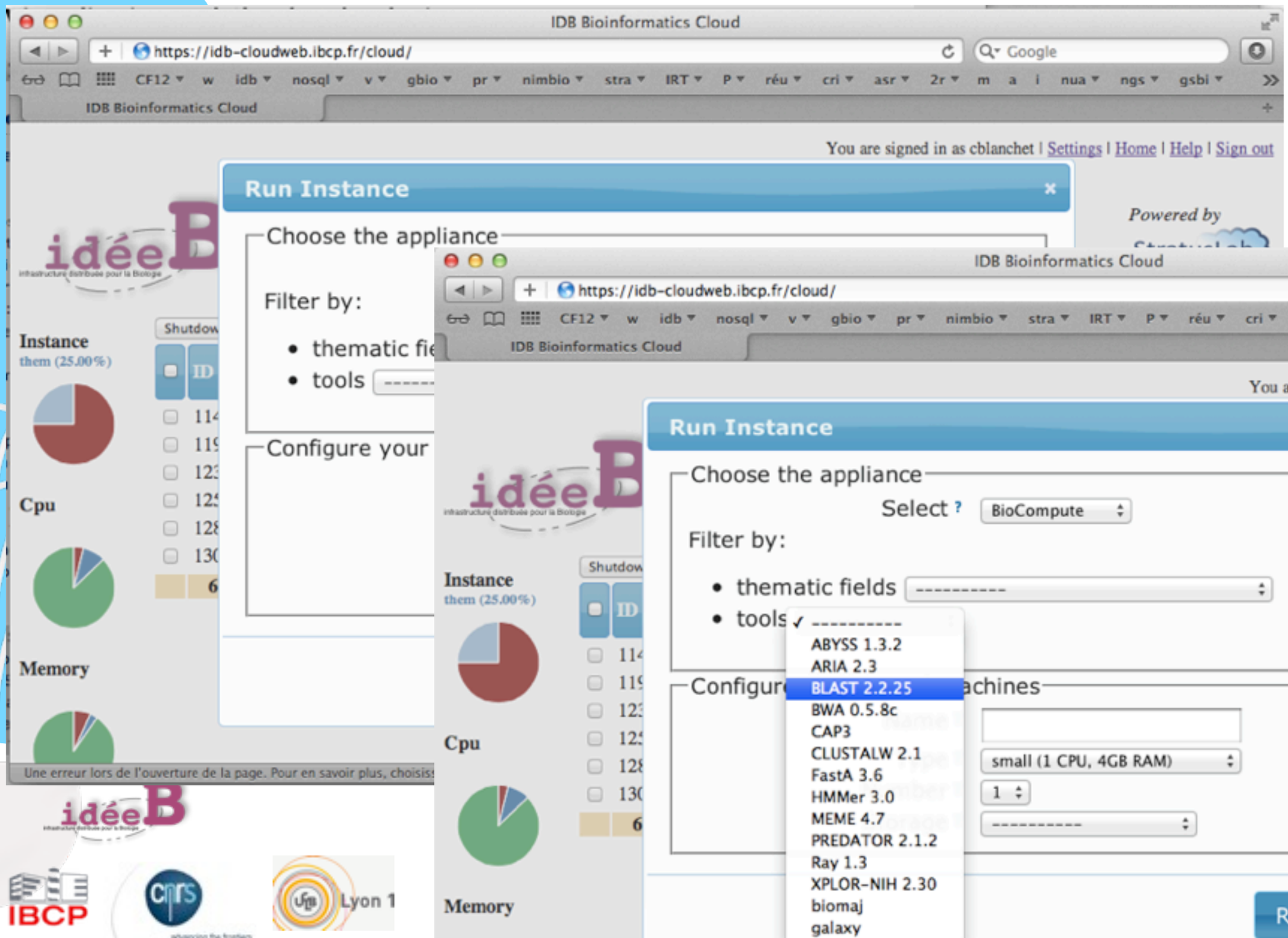
Endorser: *christophe.blanchet@ibcp.fr*
Identifier: *NaCGZfy9NxClc3ISU158RHrG0ik*
Created: *2012-09-07T14:20:38Z*

This appliance provides cloud users with a fully functional Mobyle portal. Mobyle is a framework and web portal

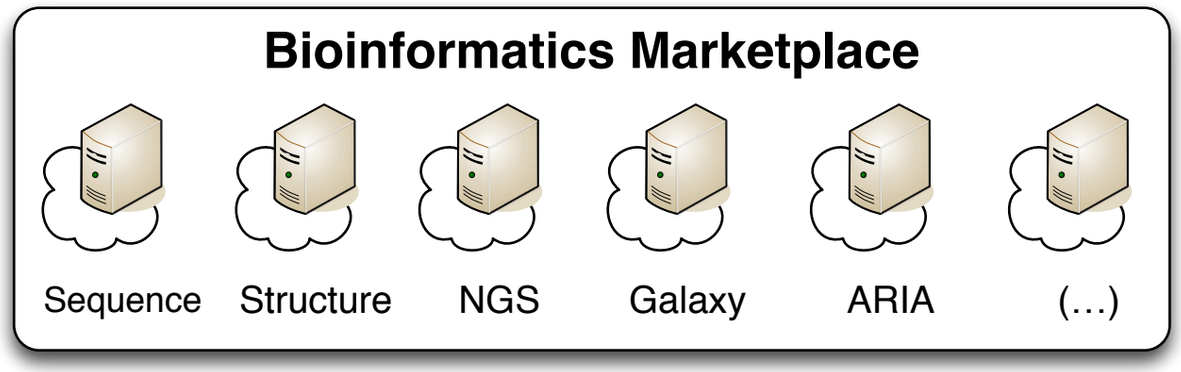
A bioinformatics web interface

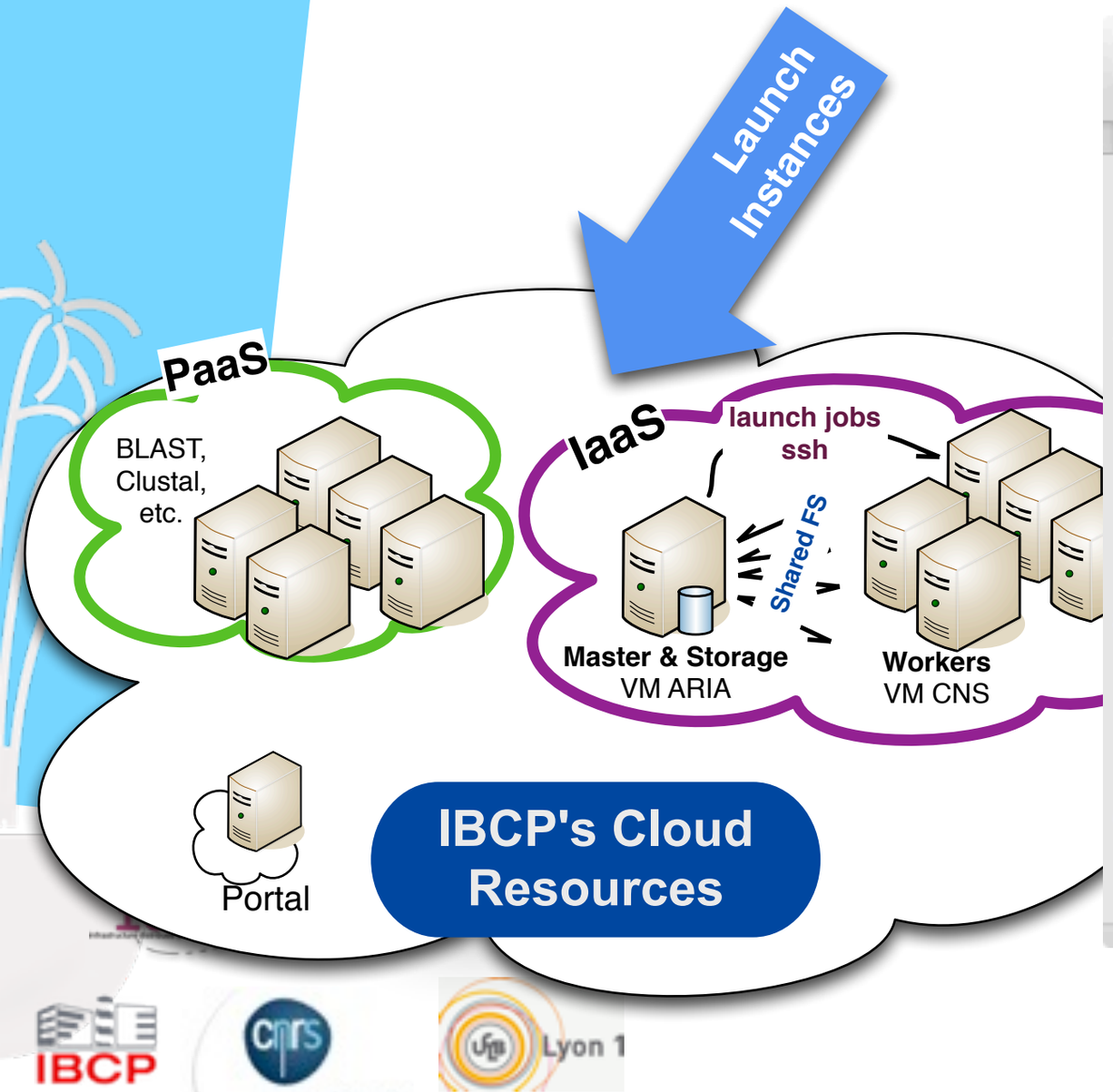


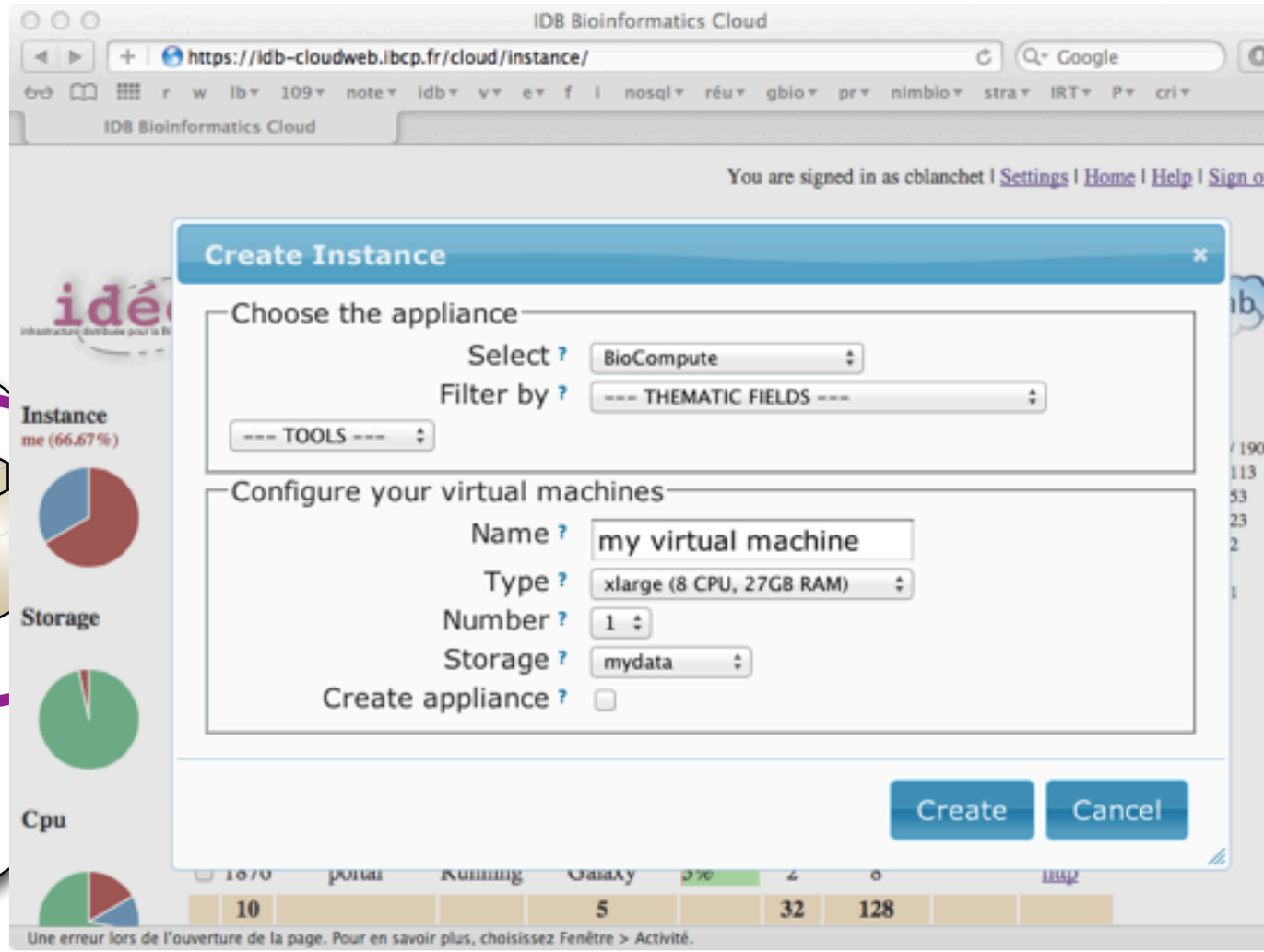
A bioinformatics web interface



Run Bioinformatics Cloud Instances







Manage your Cloud Instances

IDB Bioinformatics Cloud

https://idb-cloudweb.ibcp.fr/cloud/instance/

Google

r w lb 109 note idb v e f i nosql réu gbio pr nimbio stra IRT P cri

IDB Bioinformatics Cloud

You are signed in as cblanchet | Settings | Home | Help | Sign out

idéeB infrastructure distribuée pour la Biologie

Bioinformatics cloud

Powered by StratusLab

Instance
me (66.67%)

Storage

Cpu

Shutdown | Go | New Instance | New Storage | Show Instances | Show Storages

	ID	Name	State	Appliance	CPU %	CPU	Mem. (GB)	Storage	Access
☐	1776	cluster	Running	Hadoop GRNet	0%	2	8		ssh
☐	1787	my clustalw node	Running	BioCompute	8%	4	16		ssh http
☐	1789	aria IaaS	Running	ARIA2.3	0%	4	16		ssh
☐	1790	aria IaaS	Running	ARIA2.3	0%	4	16		ssh
☐	1791	aria IaaS	Running	ARIA2.3	0%	4	16		ssh
☐	1792	aria IaaS	Running	ARIA2.3	0%	4	16		ssh
☐	1793	aria IaaS	Running	ARIA2.3	0%	4	16		ssh
☐	1874	Dble lien	Running	Apache AuthN	0%	2	8		ssh http
☐	1875	blast	Running	BioCompute	2%	2	8		ssh http
☐	1876	portal	Running	Galaxy	3%	2	8		http
	10			**5**		**32**	**128**		

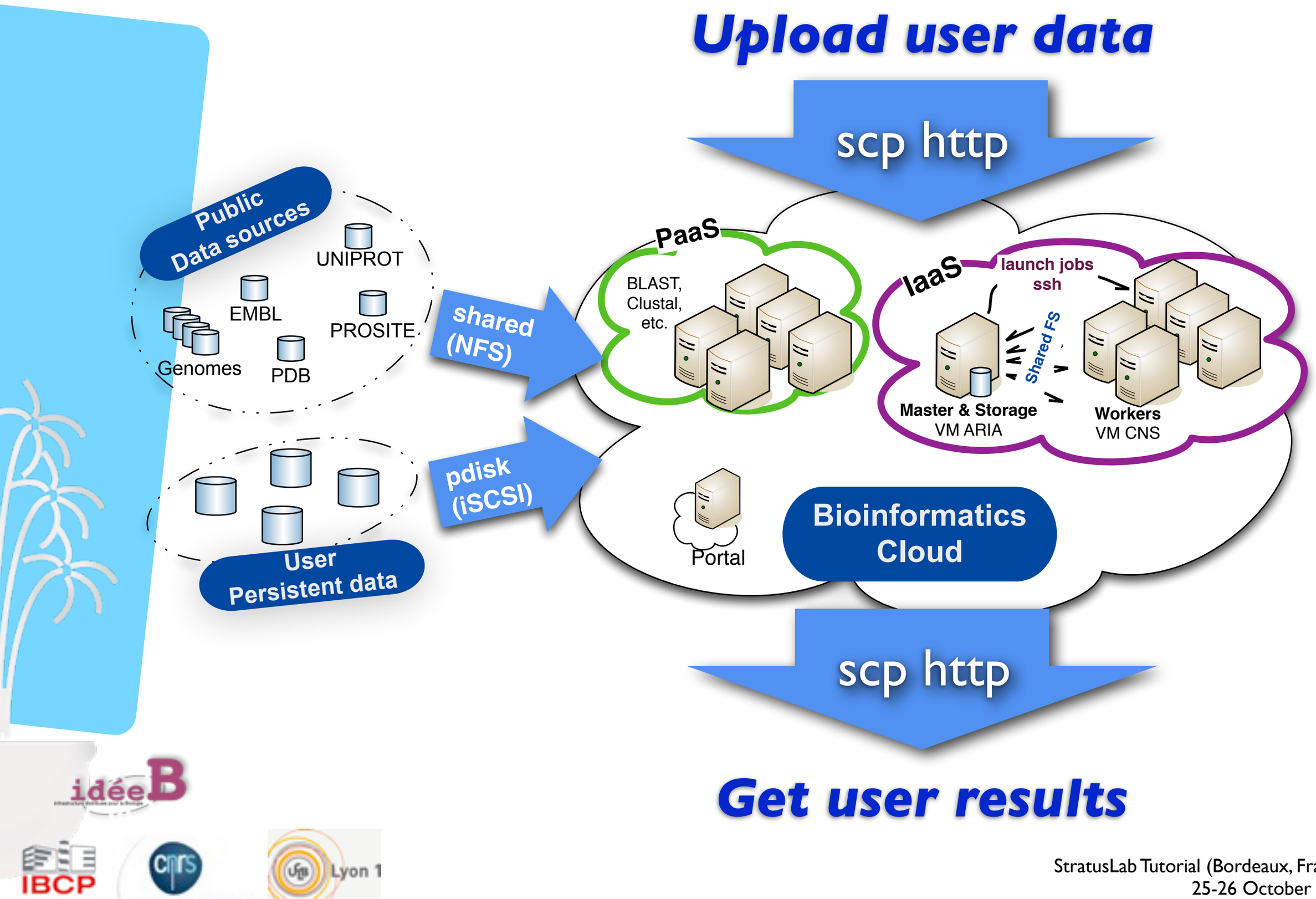
Room for VMs

xsmall	132 / 190
small	70 / 113
medium	33 / 53
large	14 / 23
xlarge	7 / 12
bigmem	1 / 1
xxl	6 / 11
htc	4 / 4
htc+	1 / 1

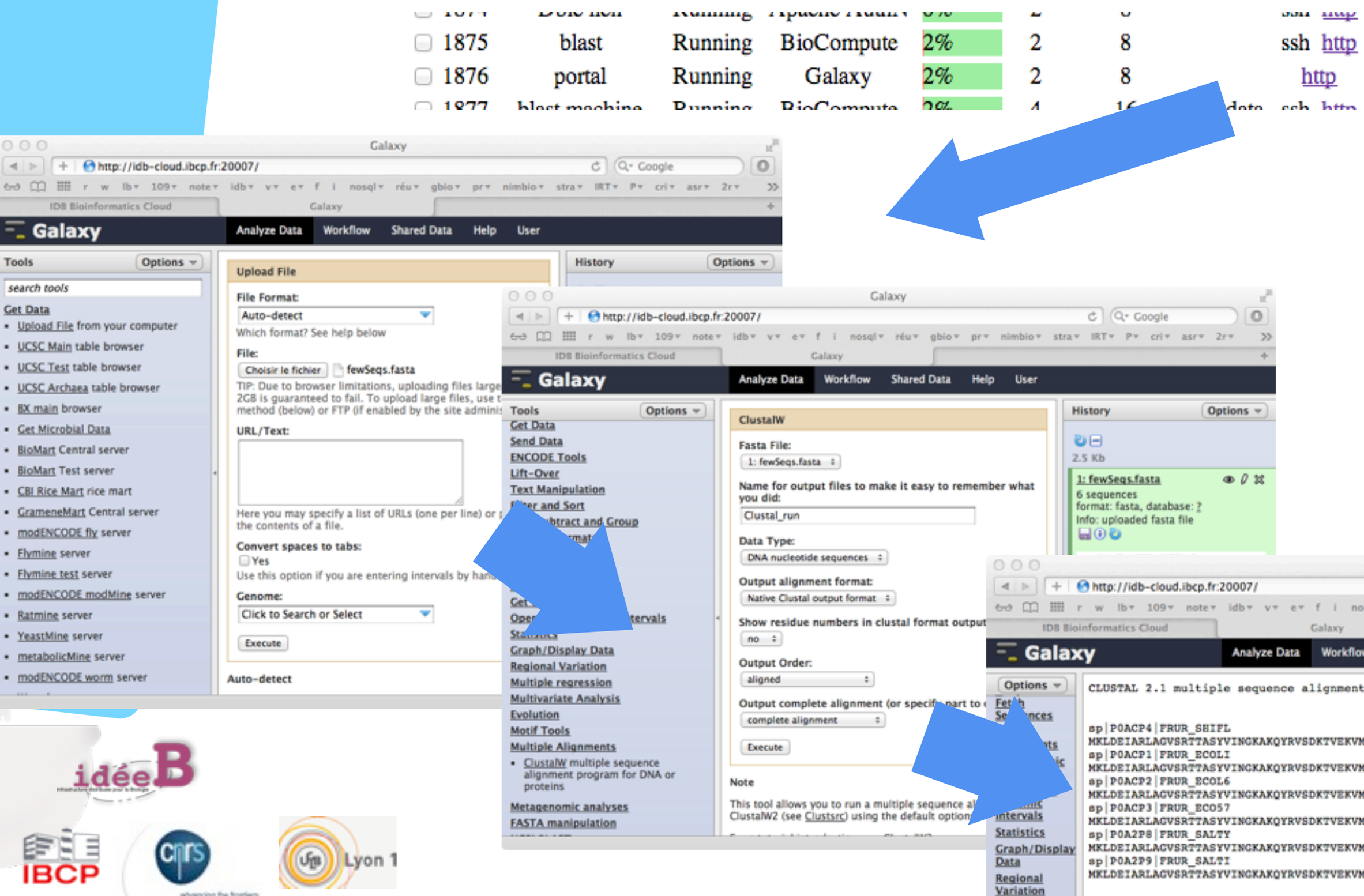
Une erreur lors de l'ouverture de la page. Pour en savoir plus, choisissez Fenêtre > Activité.

IBCP cnrs Lyon 1

Integrated Bioinformatics Cloud Instances



Examples of use



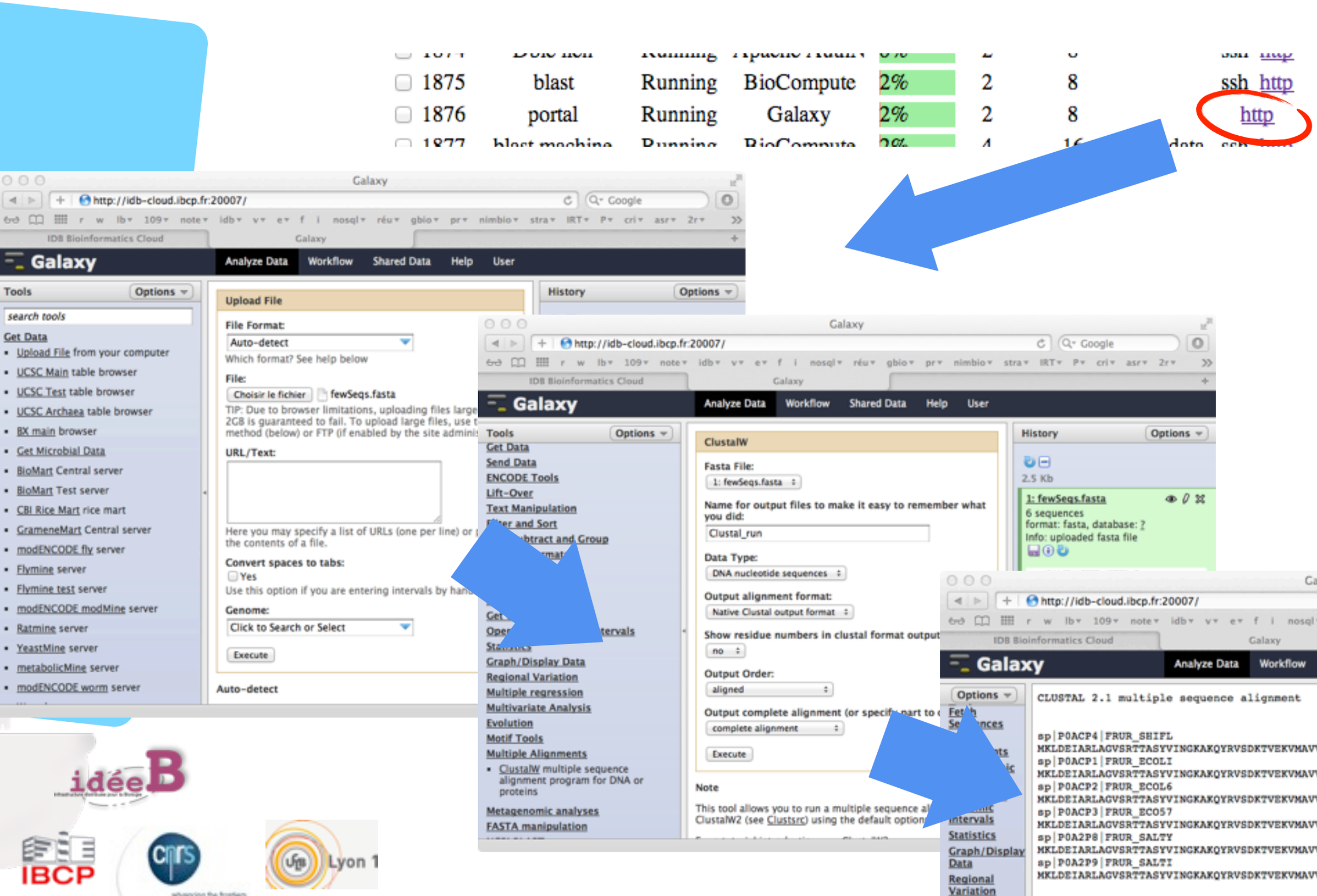
idéeB

IBCP

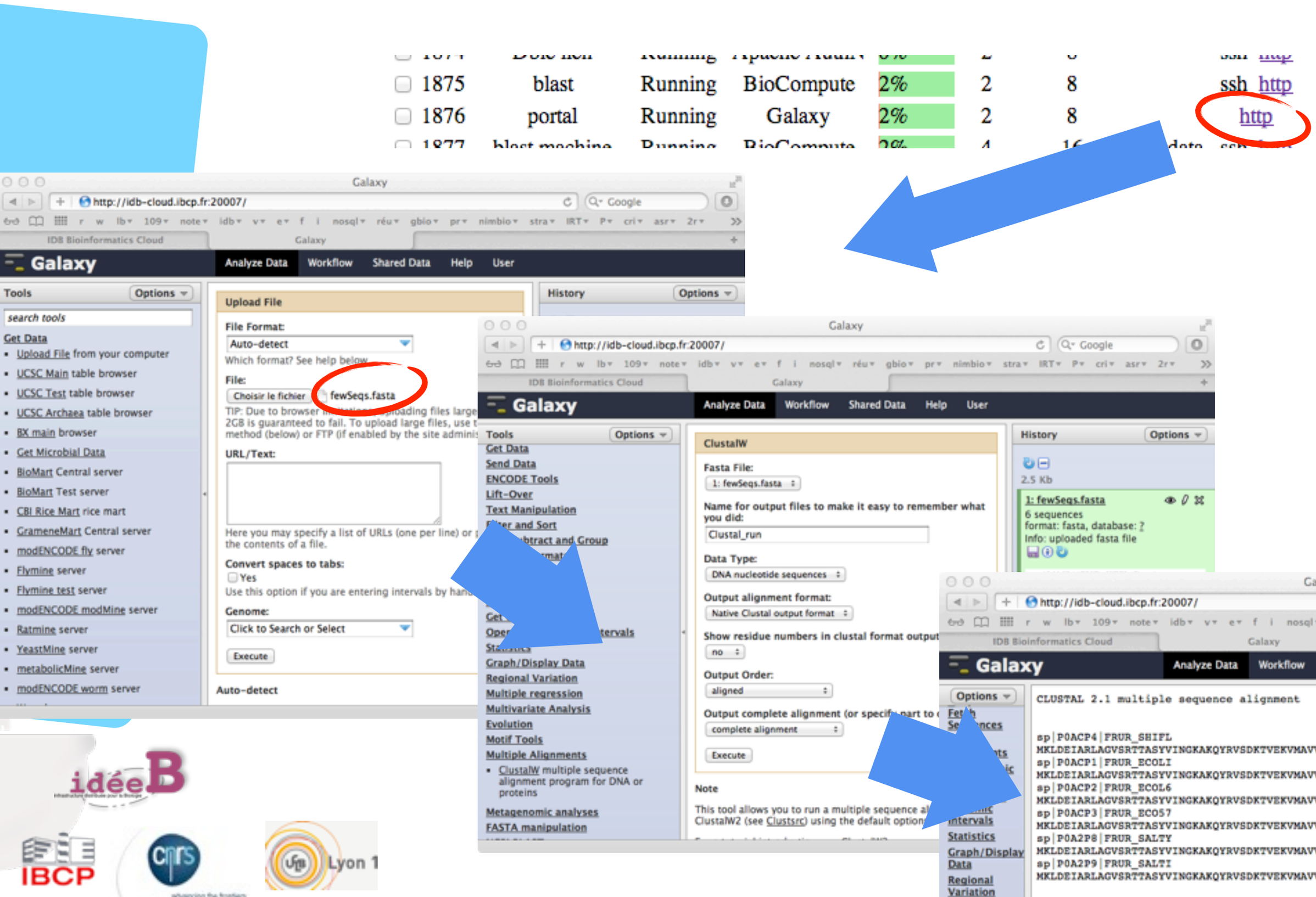
cnrs

Lyon 1

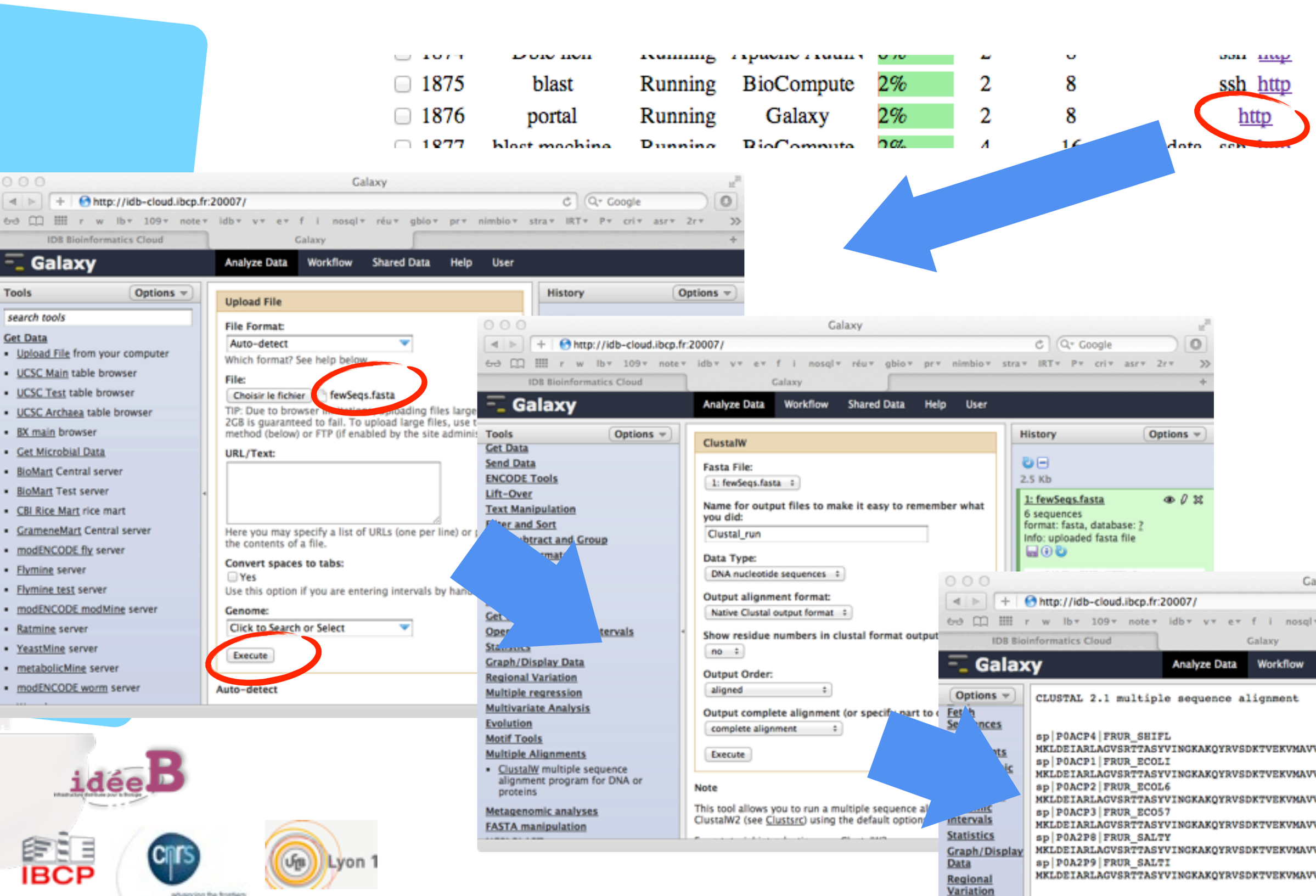
Examples of use



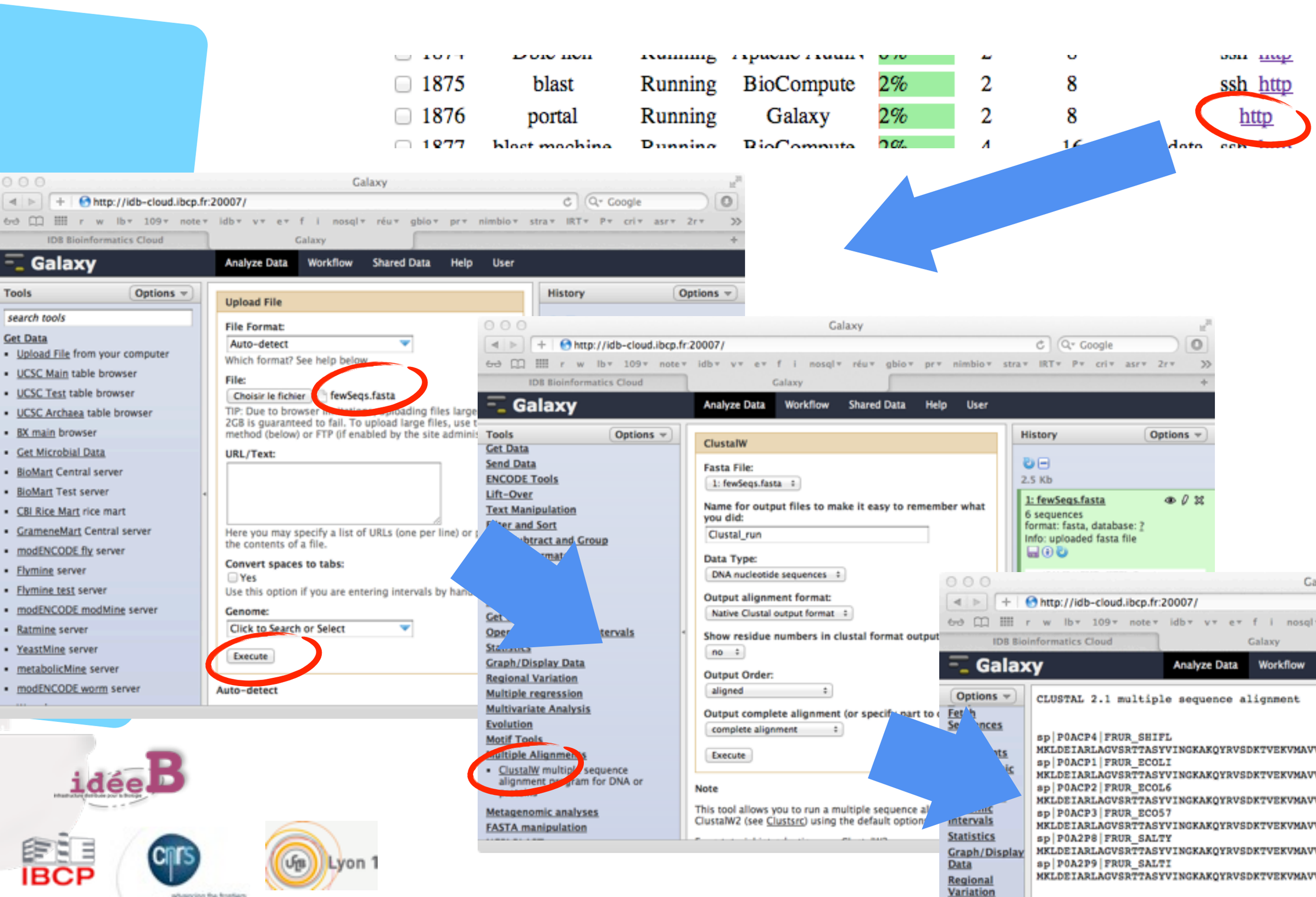
Examples of use



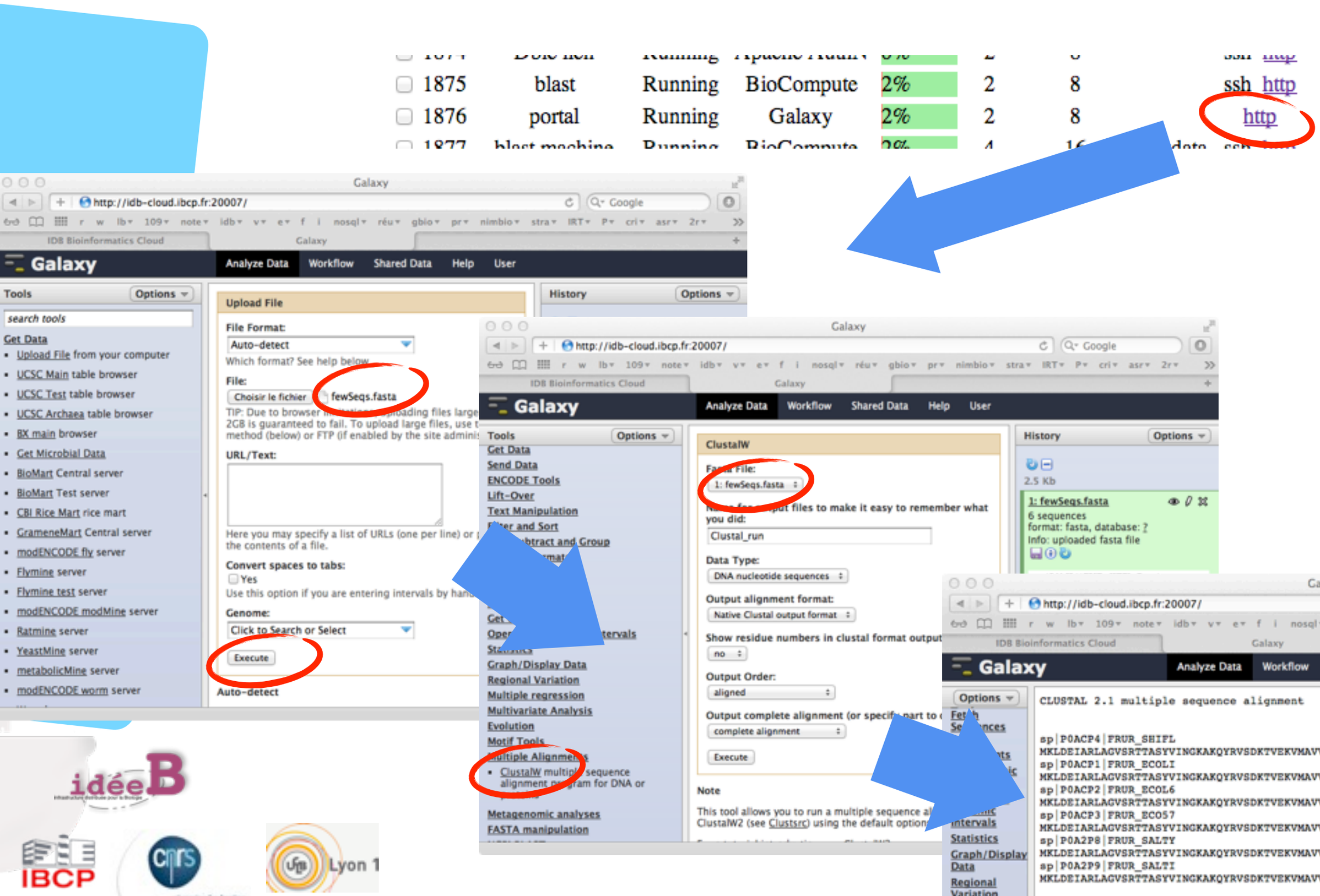
Examples of use



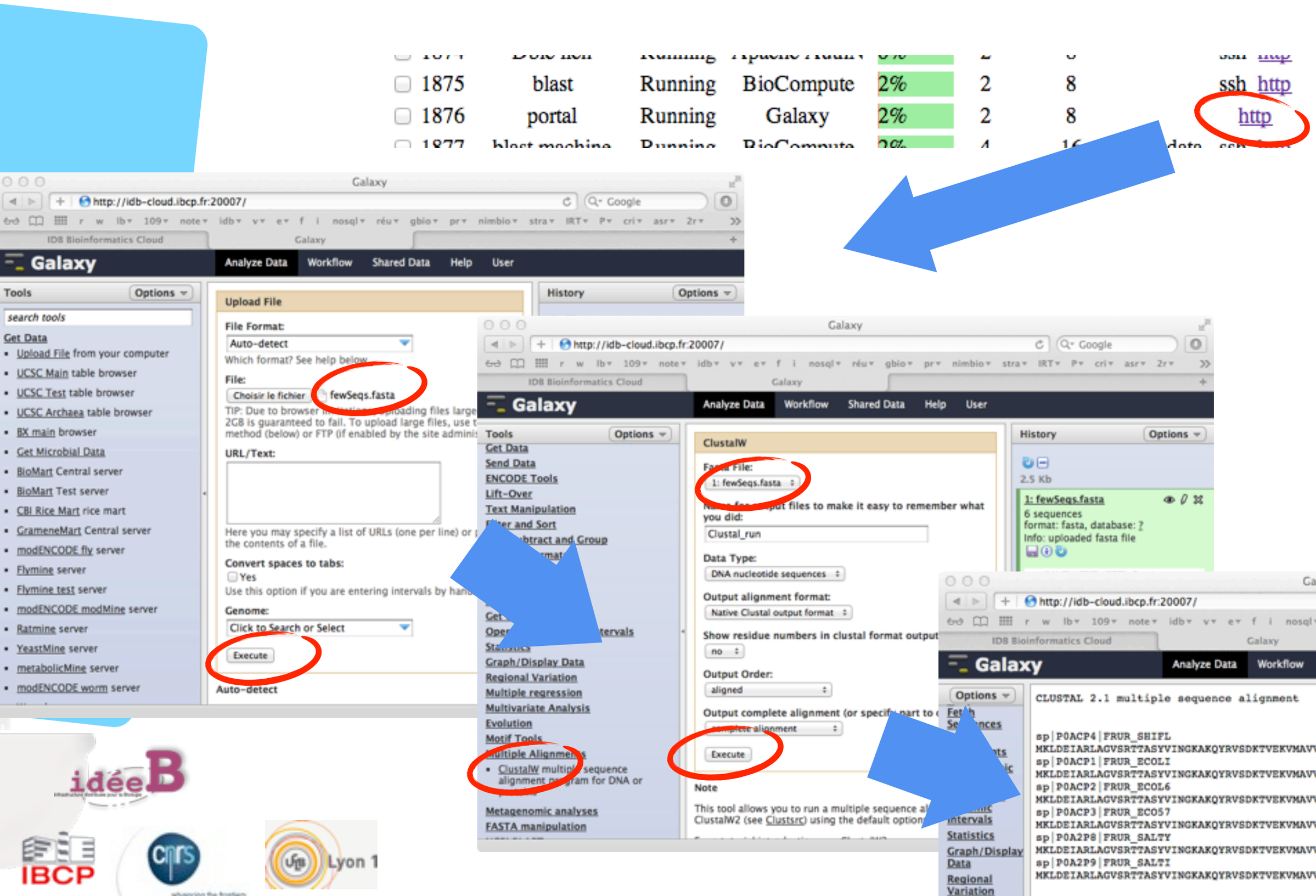
Examples of use



Examples of use



Examples of use



- Questions and Discussion

Exercise

- Sign in
  - connect to the bioinformatics-devoted cloud
    - http://idb-cloudweb.ibcp.fr
    - user ‘tbdxnn’
  - Configure you account : ‘setttings’
    (don’t forget to input your SSH pubkey)
- Run instance
  - select an appliance
  - create an instance
  - connect and run tools (ssh or http)
- Store data
  - create a storage
  - use it with an instance

Platform ‘ Infrastructure Distributed for Biology - IDB ’ acknowledges co-funding by the European Community's Seventh Framework Programme (INFSO-RI-261552) and the French National Research Agency's Arpege Programme (ANR-10-SEGI-001)

http://idee-b.ibcp.fr