



## **Insyght : Analyse evolutionary conserved CDS, orthologs, syntenies, pan-genome, fusion, etc., for your bacteria of interest**

Thomas Lacroix, Valentin Loux, Jonathan Lorenzo, Bryan Brancotte,  
Christophe Blanchet, Jean-Francois Gibrat

### **► To cite this version:**

Thomas Lacroix, Valentin Loux, Jonathan Lorenzo, Bryan Brancotte, Christophe Blanchet, et al..  
Insyght : Analyse evolutionary conserved CDS, orthologs, syntenies, pan-genome, fusion, etc., for  
your bacteria of interest. Annual Research in Computational Molecular Biology (RECOMB), Apr  
2018, Paris, France. , 2018. hal-02736501

**HAL Id: hal-02736501**

**<https://hal.inrae.fr/hal-02736501>**

Submitted on 2 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# Analyse evolutionary conserved CDS, orthologs, syntenies, pan-genome, fusion, etc., for your bacteria of interest

Lacroix T.<sup>1,\*</sup>, Loux V.<sup>1</sup>, Lorenzo J.<sup>2</sup>, Brancotte B.<sup>2</sup>, Blanchet C.<sup>2</sup>, Gibrat J.F.<sup>1</sup>

<sup>1</sup> MalAGE, INRA, Université Paris-Saclay, 78350 Jouy-en-Josas, France ; <sup>2</sup> CNRS, UMS 3601 ; Institut Français de Bioinformatique, IFB-core, F-91400 Orsay, France ; \* Corresponding author: [thomas.lacroix@inra.fr](mailto:thomas.lacroix@inra.fr)

## Why ? How ?

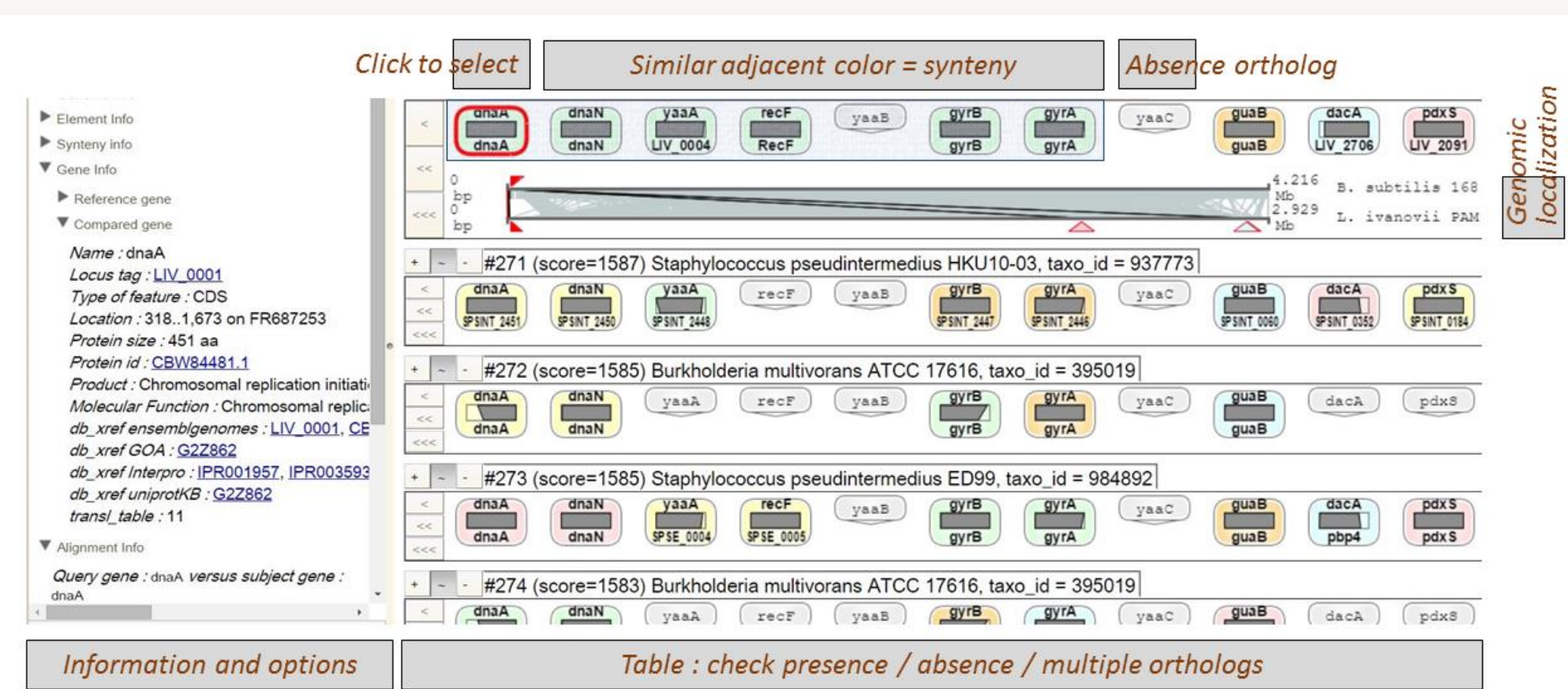
Do you wish to quickly grasp the extent of strongly evolutionary conserved CDS among your bacterial organisms ?

Insyght (<http://genome.jouy.inra.fr/Insyght/>) helps you to quickly generate the relevant data and navigate among abundant orthologs, syntenies and gene fusions. The navigation among the various pieces of information (annotations, localization) and comparisons (presence / absence, gene set, compared organisms, etc...) is practical and extensive, even within large datasets.

It detects strongly evolutionary conserved CDS by comparing proteomes (bi-directional best hit) and by leveraging information on syntenies (using dynamic programming) and gene fusions. The presence of strongly evolutionary conserved CDS is a compelling evidence in analysis such as inference of functions, core genome, or phylogenomic relatedness. A functionality highlights overrepresented orthologs among group of genomes, which is useful to investigate niche-specific CDS for a particular phenotype of interest.

## A graphical way to analyze data

### Orthologs table view



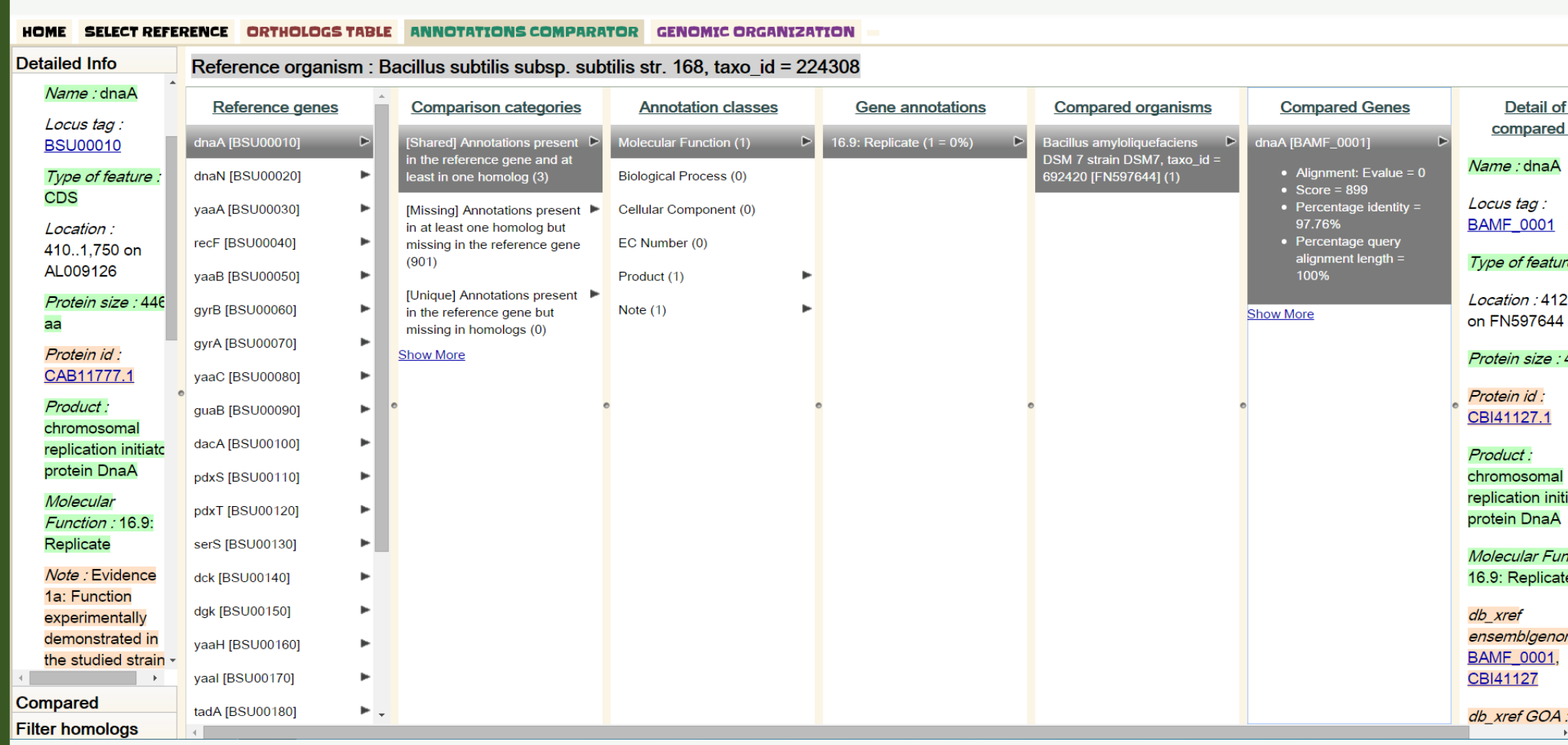
A spreadsheet adapted to browse orthologs :

- ☞ Familiar layout: gene = col, organism = row
- ☞ Info on annotations, alignments, location, etc. at your fingertip
- ☞ Genes in adjacent columns with similar background color = synteny
- ☞ Multiple "off shoots" homologs stacked in 1 cell
- ☞ Build your own gene set
- ☞ More: sort the table, quickly navigate CDS, etc.

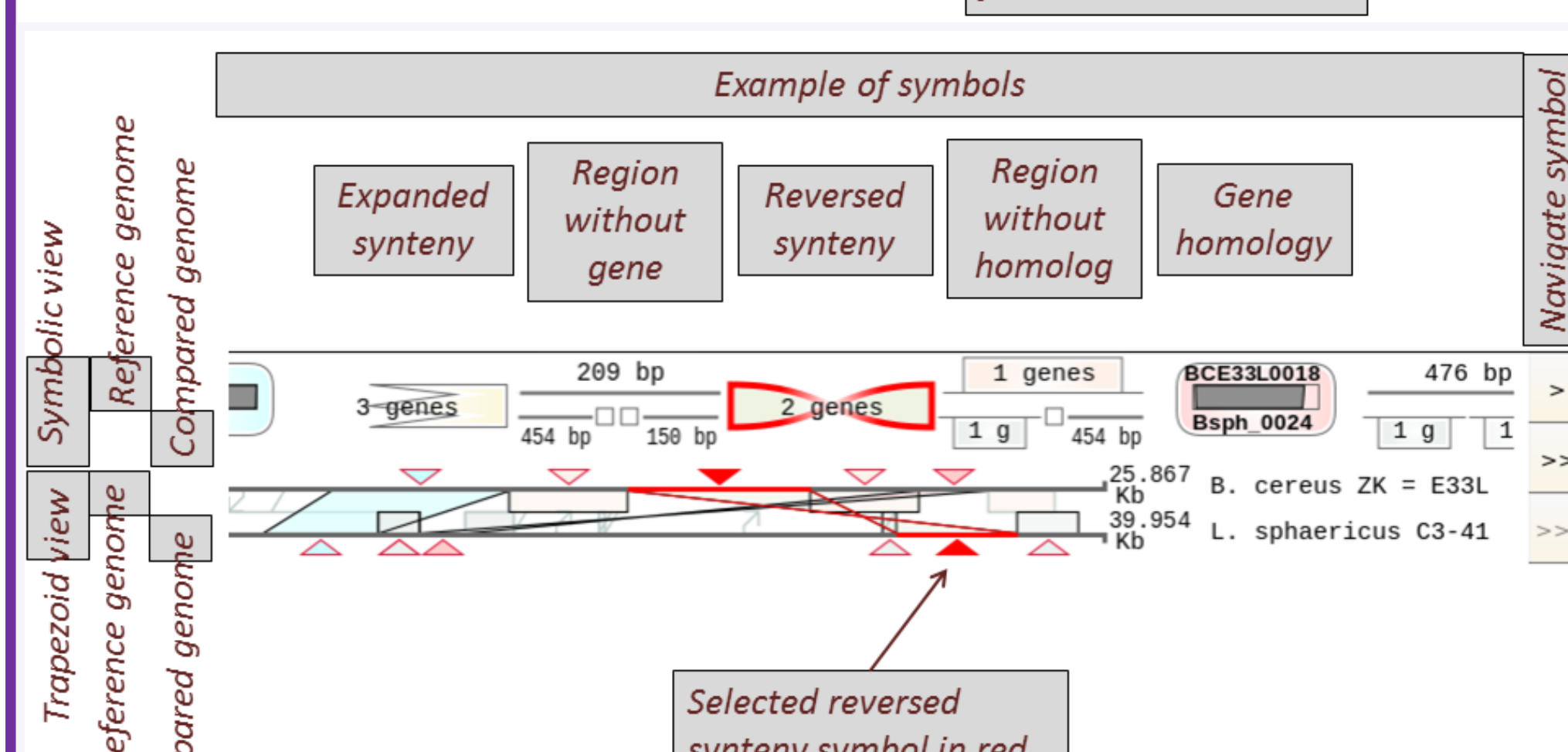
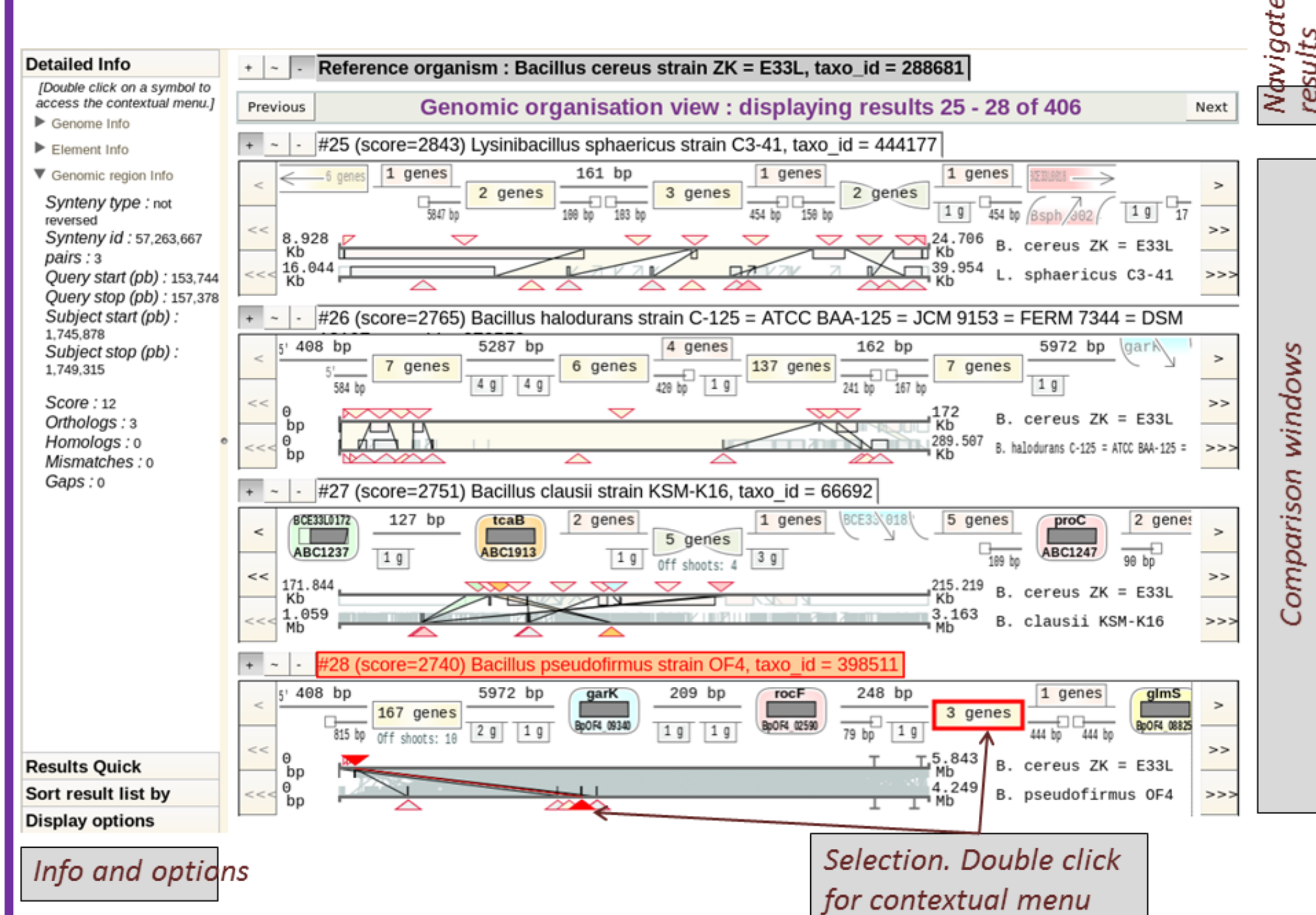
### Annotations comparator

For a given CDS and its orthologs, what are the functional annotations that are shared or not ? The annotations are classified into 3 categories (Shared, Missing, and Unique) :

- ☞ Browse those 3 categories and subcategories: functional annotation, homologous genes, sequence alignment, etc.
- ☞ Restrict the set of organisms considered, filter homologs, etc.



### Genomic context view



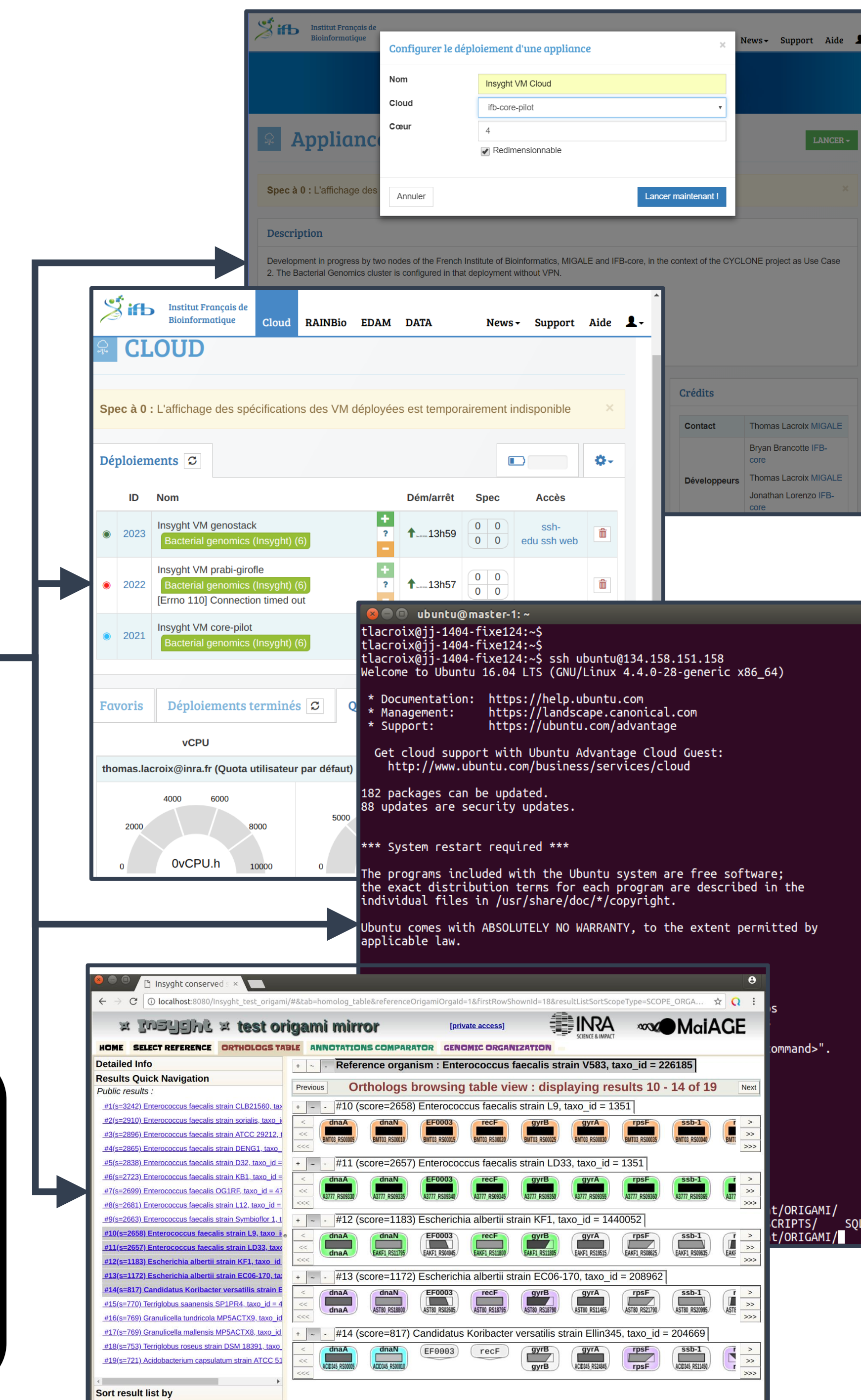
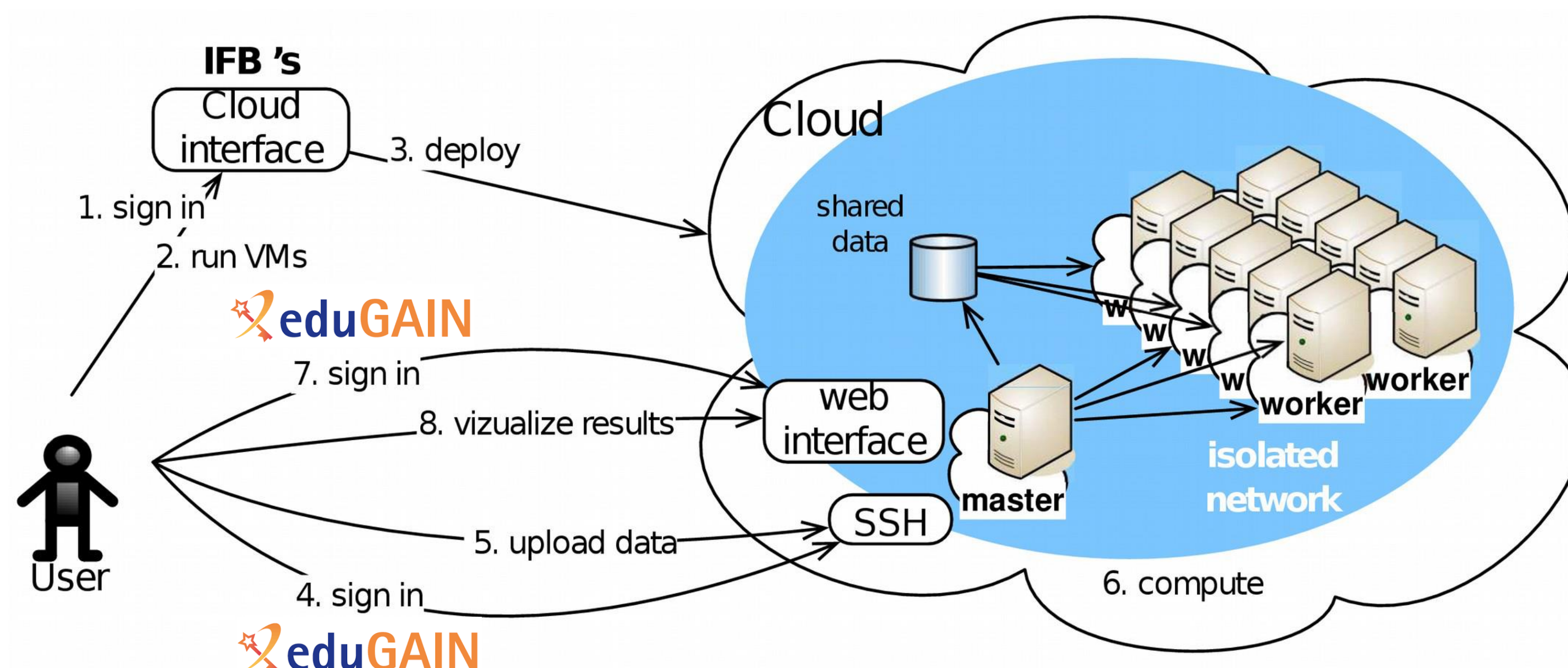
The 3 above views are interconnected:

- ☞ Transfer genes from a synteny to the orthologs table: evaluate their conservations in other species
- ☞ Check the genomic context of a gene from the orthologs table
- ☞ In short, transfer genes from one view to the other by double clicking on symbols

## Analyze your data with a VM, either locally or in the IFB cloud

The Insyght virtual machine (VM) is ready to use, you can customize the genomes to compare and various parameters. You can generate and analyze private data "in the cloud" as well by using the French Institute of Bioinformatics (IFB) cloud infrastructure : <https://biosphere.france-bioinformatique.fr/catalogue/>. The IFB appliances features:

- ☞ Authentication using academics federation for SSH (command line) and web access
- ☞ Access to a cluster to leverage parallel computing and speed up data generation
- ☞ automatic installation recipes (slipstream)
- ☞ multicloud deployment support (ifb-core-cloud, prabi-girofle, genouest-genostack, etc.).



Published online 23 September 2014

Nucleic Acids Research, 2014, Vol. 42, No. 21, e162  
doi: 10.1093/nar/gku867

**Insyght: navigating amongst abundant homologues, syntenies and gene functional annotations in bacteria, it's that symbol!**

Thomas Lacroix<sup>1,\*</sup>, Valentin Loux<sup>1</sup>, Annie Gendrait<sup>1</sup>, Mark Hoebeke<sup>2</sup> and Jean-François Gibrat<sup>1</sup>



Try the tool, check out our public datasets :

- ☞ 2692 complete bacteria from Ensembl Bacteria
- ☞ 210 reference archaea genomes from Uniprot
- ☞ coming soon : ~5000 complete reference bacteria from Uniprot