

Why?

- Pathogen surveillance and diagnostic **methods** are constantly evolving. In particular, **sequencing technologies** offer unprecedented perspectives.
- More and more **sequence data from mollusc pathogens** are being generated.
- Pooling of available sequences, combined with standardised approaches to analyse them, is required for consistency in international studies.

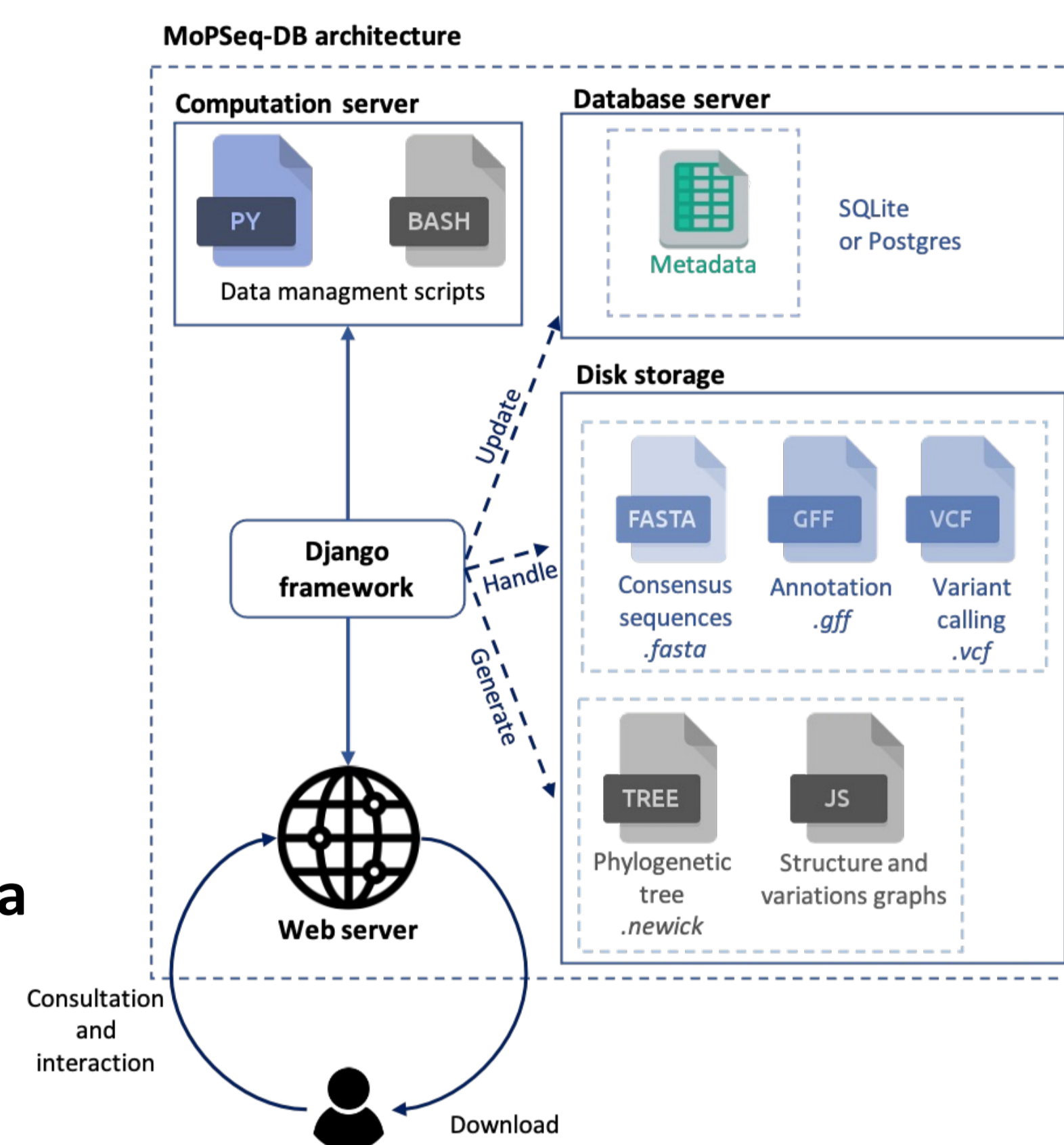
Our goal?

Develop a user-friendly web-application to :

- ➔ 1) Reference and harmonise genomic data related to **mollusc pathogens**
- ➔ 2) Give users opportunities to interactively visualise **genomic data**
- ➔ 3) Provide integrated **analysis tools**

How?

- **Open-source** Python web application
- Relying on the **Django** framework
- Data storage:
 - ➔ **relational database** for basic information
 - ➔ pre-processed files stored **directly on disk**
- Application deployed using a **Docker** container
- MoPSeq-DB comes with **scripts to load/update data**



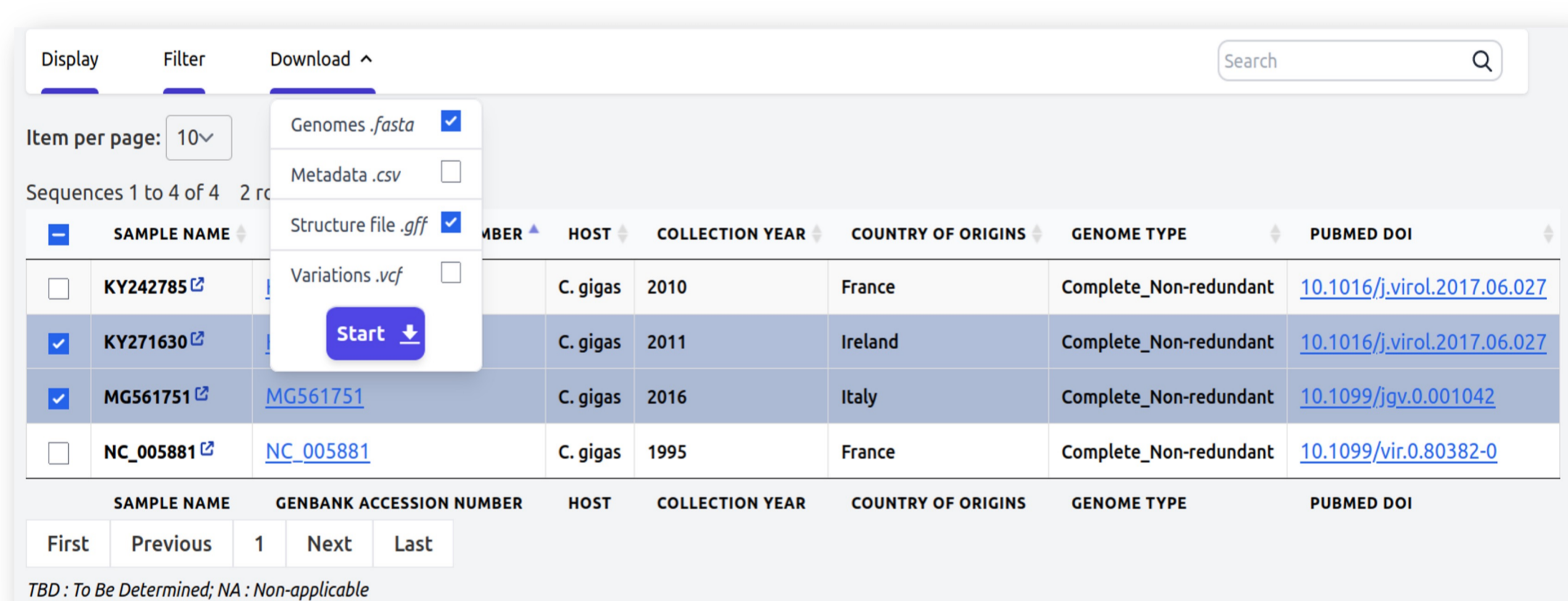
The interface needs to enable users :

- ➔ 1) To **navigate** through all the database
- ➔ 2) Get fully **interactive** views of genome structures, variants, phylogenetic trees
- ➔ 3) **Download** data in various formats (FASTA, GFF, VCF, Newick, PNG)

What?



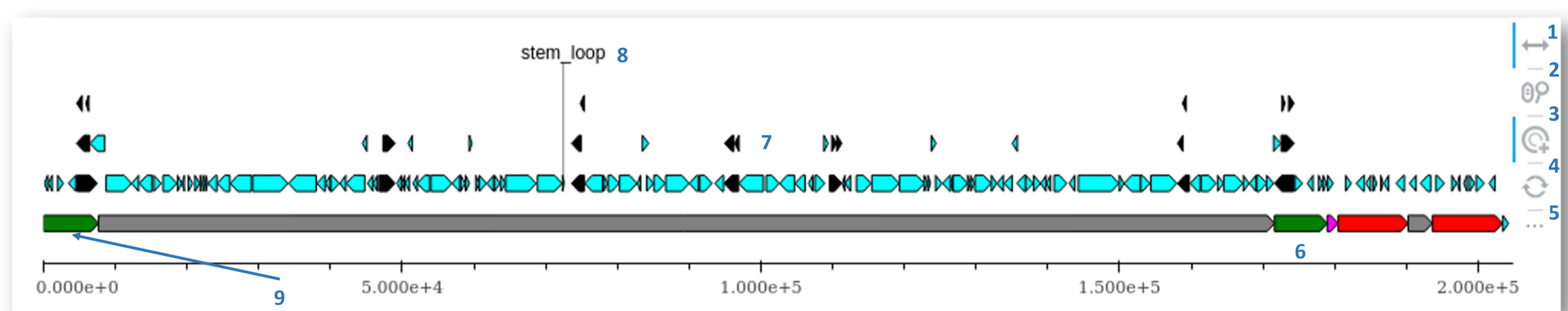
Data referencing and sharing



SAMPLE NAME	HOST	COLLECTION YEAR	COUNTRY OF ORIGINS	GENOME TYPE	PUBMED DOI
KY242785	C. gigas	2010	France	Complete_Non-redundant	10.1016/A.virel.2017.06.027
KY271430	C. gigas	2011	Ireland	Complete_Non-redundant	10.1016/A.virel.2017.06.027
MG561751	C. gigas	2016	Italy	Complete_Non-redundant	10.1099/ov.0.001042
NC_005881	C. gigas	1995	France	Complete_Non-redundant	10.1099/vir.0.80382-0

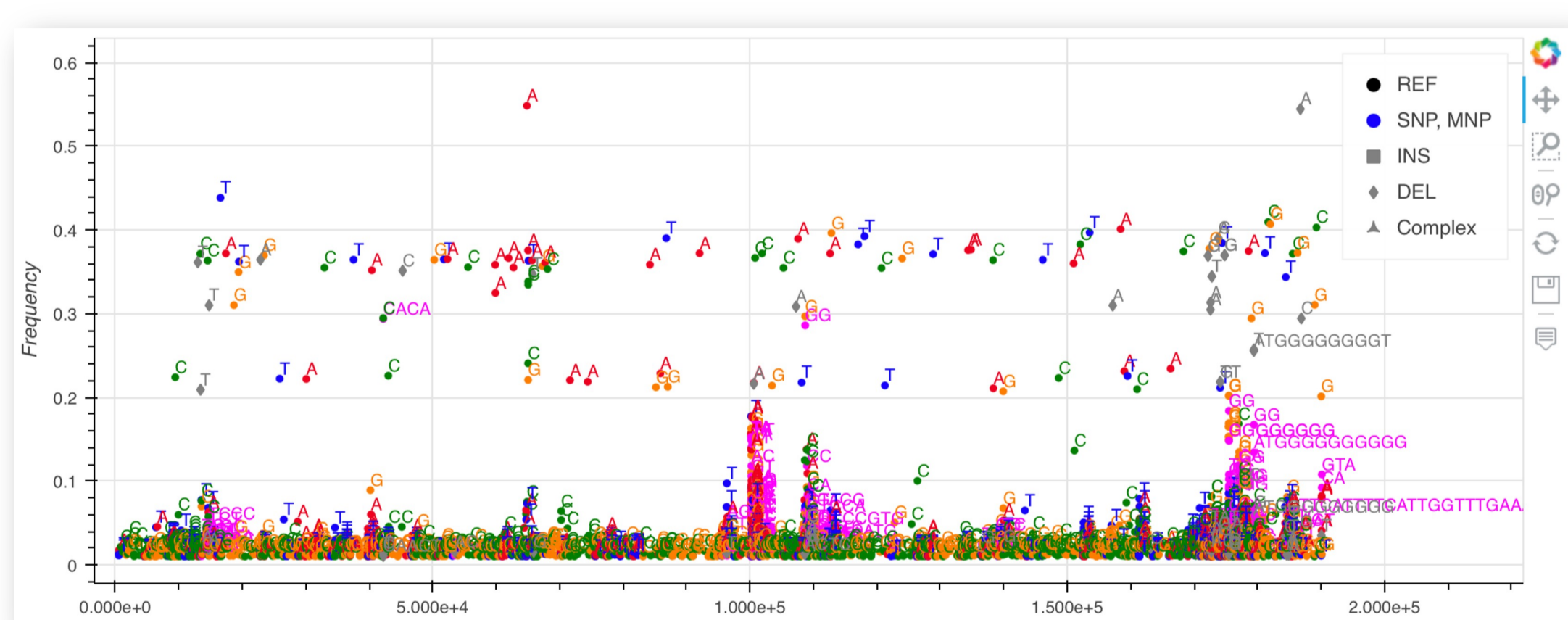
Simple but rich overview of the different genomes and pathogens referenced in the database, with the possibility to download the samples associated genomic files. Users have the opportunity to dynamically research for data, filter or order on attributes. Any attribute can be displayed or hidden.

Genomic structure and annotation



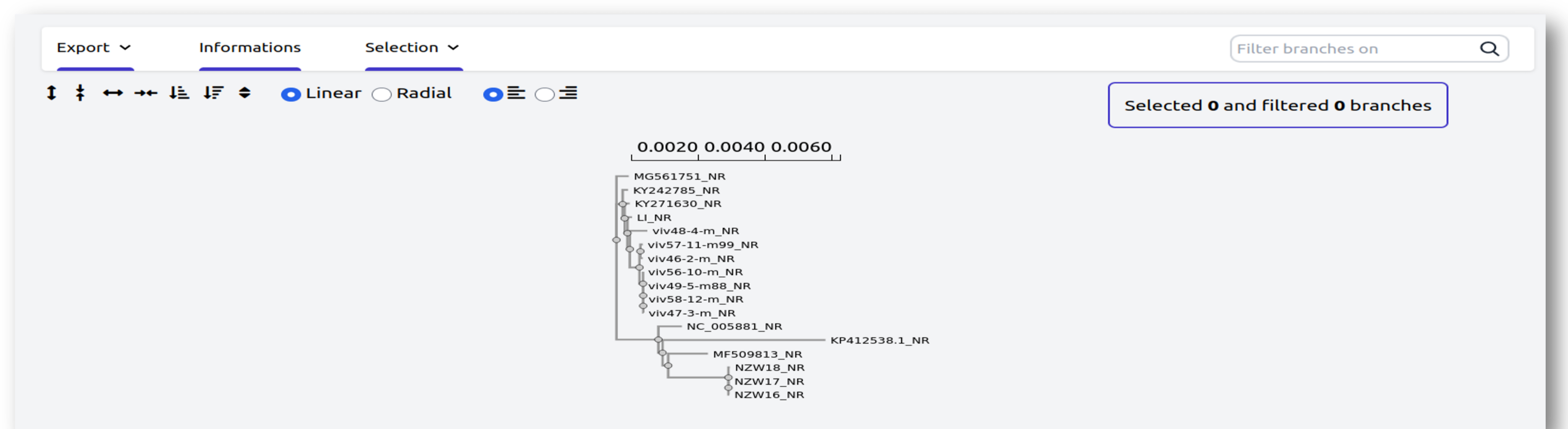
Example of a structure graph of an OshV-1 genome. Different interactivity options are given to users: 1: move horizontally on graph; 2: zoom on; 3: select and isolate; 4: revert to base image; 5: disable labels. Examples of displayed information: 6: label of a structural region; 7: representation of elements stored in a .gff file (e.g. CDS and functional annotation); 8: fixed labels; 9: homogenous colours for repeated regions e.g. here green for OSHV-1 Long Inverted Repeat (IRL) region.

Intra-individual variations



Intra-individual variation graphs. Different type of variation are visualized (SNP, MNP, INS, DEL and Complex). Variations detected in a sample were filtered to a frequency of 0.011%.

Phylogenetic placement



Rendered phylogenetic trees present different interactivity options for users to better handle data and visualise results. Users can zoom on or reorganise trees with different configurations to better fit their needs. It is possible to collapse or hide subtrees, select or filter different sequences or clades and re-root the tree. Downloading a tree file or a high quality picture of capturing the current annotation is also possible.

Conclusion

MoPSeq-DB

- ➔ Runs on all modern browser engines (Firefox, Chrome, Safari)
- ➔ Is highly **user-friendly**
- ➔ **Standardise** all sequence-based work related to **mollusc pathogens**
- ➔ Allows unprecedented opportunities to carry out **monitoring and diagnosis in near-real time**
- ➔ Meet the **FAIR** data principles: findability, accessibility, interoperability, and reusability (FAIR).

Availability

MoPSeq-DB source code and public web platform will be available at: <https://mopseq-db.ifremer.fr>

First release is expected in the first quarter of 2023.