ARTwork, a bioinformatics solution dedicated to WGS data management and analysis in the context of foodborne pathogens surveillance and outbreak investigations

anses
agence nationale de sécurité sanitaire
alimentation, environnement, travail

Connaître, évaluer, protéger

UNIVERSITÉ DE ROUEN

Génomique Analyse Modélisation Risque

K. Durimel[1#], A. Felten[2], MY. Mistou[2*]

[1] Normandie Université, Anses, Laboratory for food safety, Maisons-Alfort (France).

[2] Anses, Laboratory for food safety, Maisons-Alfort (France).

Presenters (kevin.durimel@anses.fr)

BACKGROUND

To guarantee the **traceability** of **bioinformatics analysis**, it is mandatory to track all bionformatics processes. To ensure quick **data recovery**, a regularly saved database is a common solution. A multitude of software and cloud platforms already provide answers to meet these requirements [1] [2] [3].

Nevertheless, they are not suitable for use in the context of foodborne pathogens surveillance. We consequently designed **ARTwork**, an open source bioinformatics solution aiming to standardize **WGS data management** for the most common **foodborne pathogens** studied in the lab.

ARTwork : Assembly of Reads and Typing workflow

16 bioinformatics tools are implemented in ARTwork. A document-oriented database is also used to store information about each received WGS data. Finally, a user-friendly web application allows final users to query the database and easily retrieve data which can be used as a new starting point for downstream analyses.
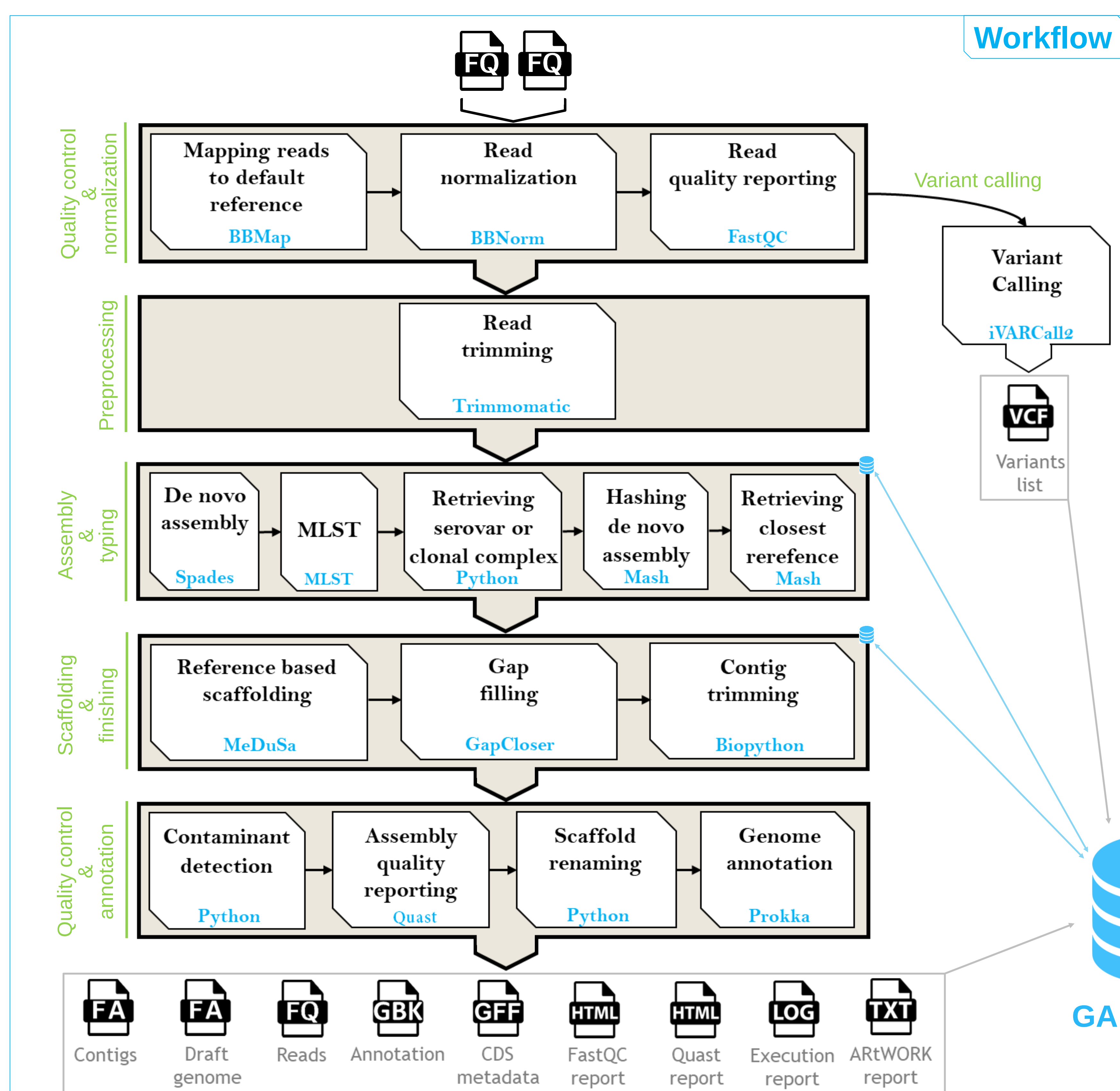
Workflow

Figure 1: *Global worfklow of ARTwork. Fastq files are taken as input in order to perform quality control of reads, read trimming, variant calling, de novo assembly, reference-based scaffolding, assembly quality control, genome typing and annotation, and track all metadata generated during the data processing lifecycle.*

Data validation

A large majority of the genomes assembled by ARTwork are **correct**, **complete** and **contiguous** according to the key metrics evaluated by **Quast** and recommended in the Assemblathon 2 study [4].
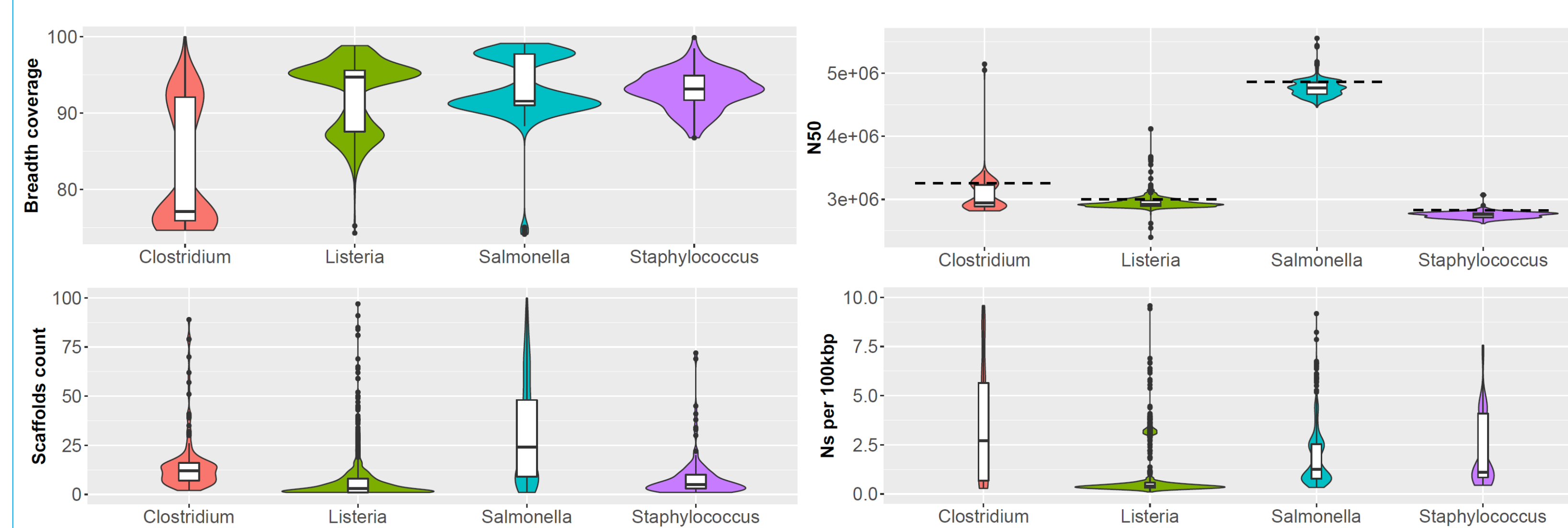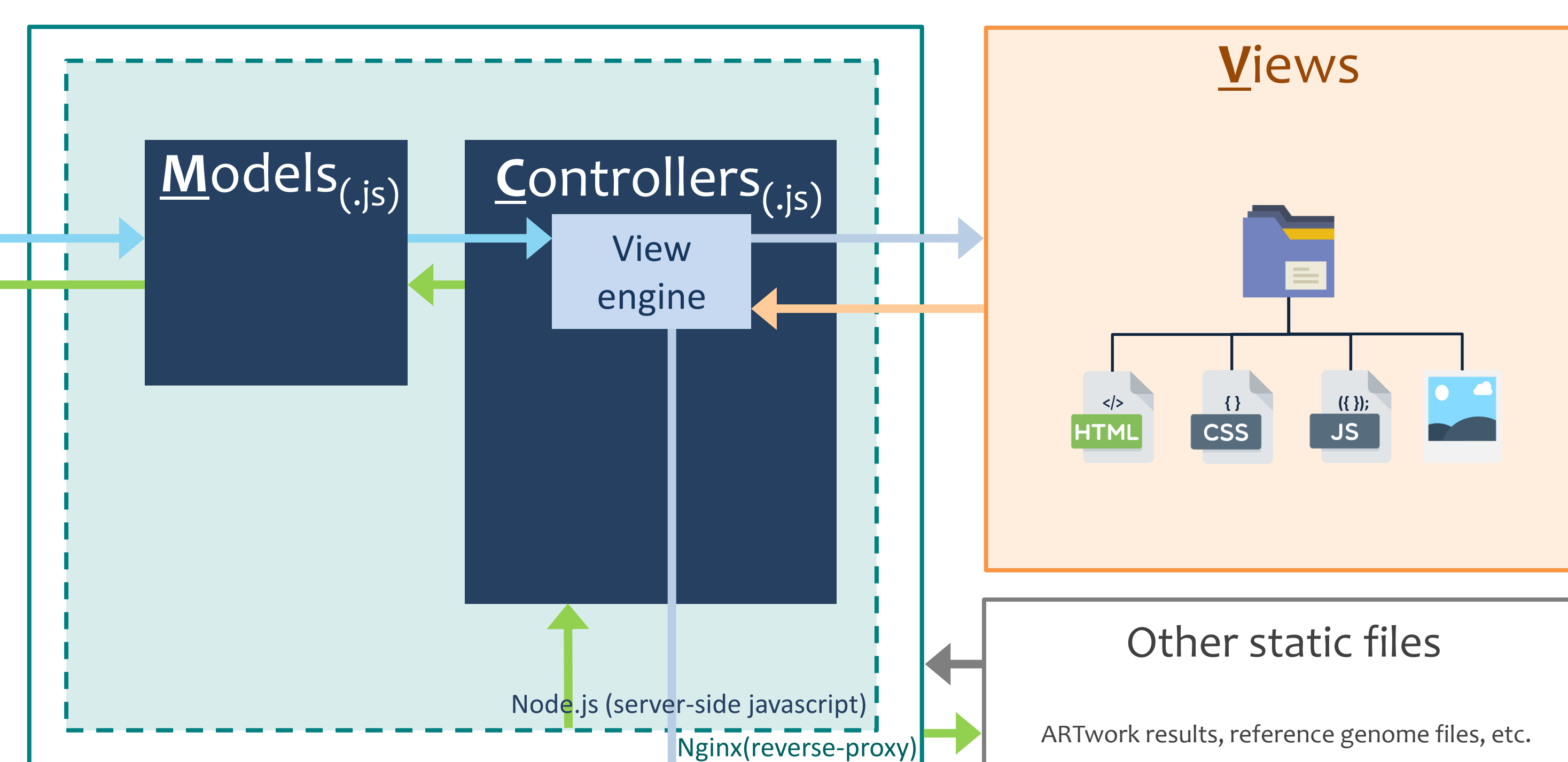

Figure 2 : *Quality metrics of the 2366 genomes assembled by ARTwork during the February 2017 – April 2018 period. n = 149 (Clostridium perfringens) ; 1 121 (Listeria monocytogenes) ; 940 (Salmonella enterica) ; 156 (Staphylococcus aureus)*

Figure 3 : *MVC pattern implemented in the ARTwork web app.*

Web application

WEB INTERFACE

The web application main page (*fig.4*) shows a global view of the GAMeRdb genomic database and is a starting point to the five "pathogen workpaces".

From the pathogen workspaces (*fig.5*) , all the **data** and **metadata** availaible for the pathogens processed by ARTwork can be **dynamically explored** (*i.e* strains processed, reference genomes, serovar/cc/st distribution) and **downloaded** (*i.e* bioinformatic files).
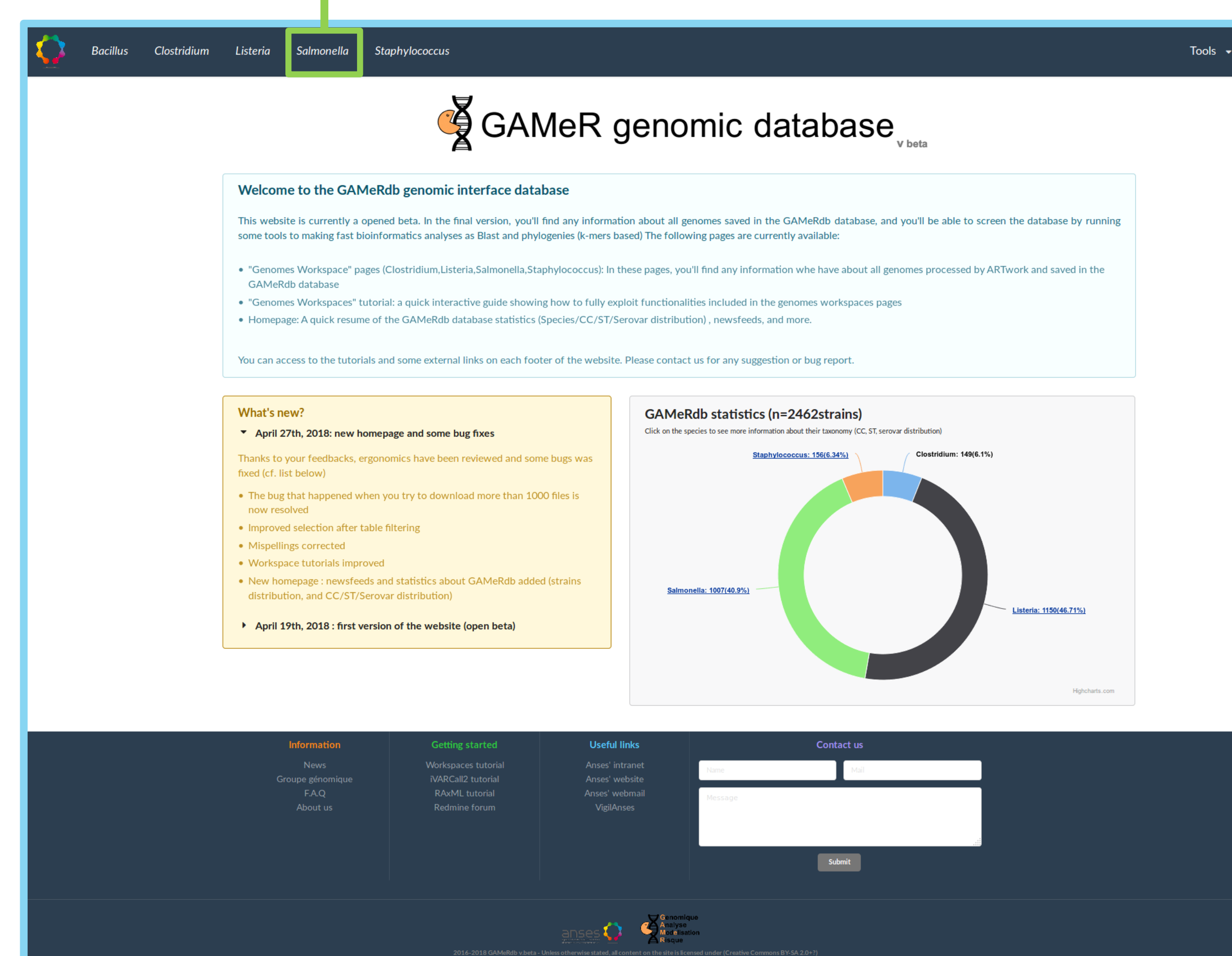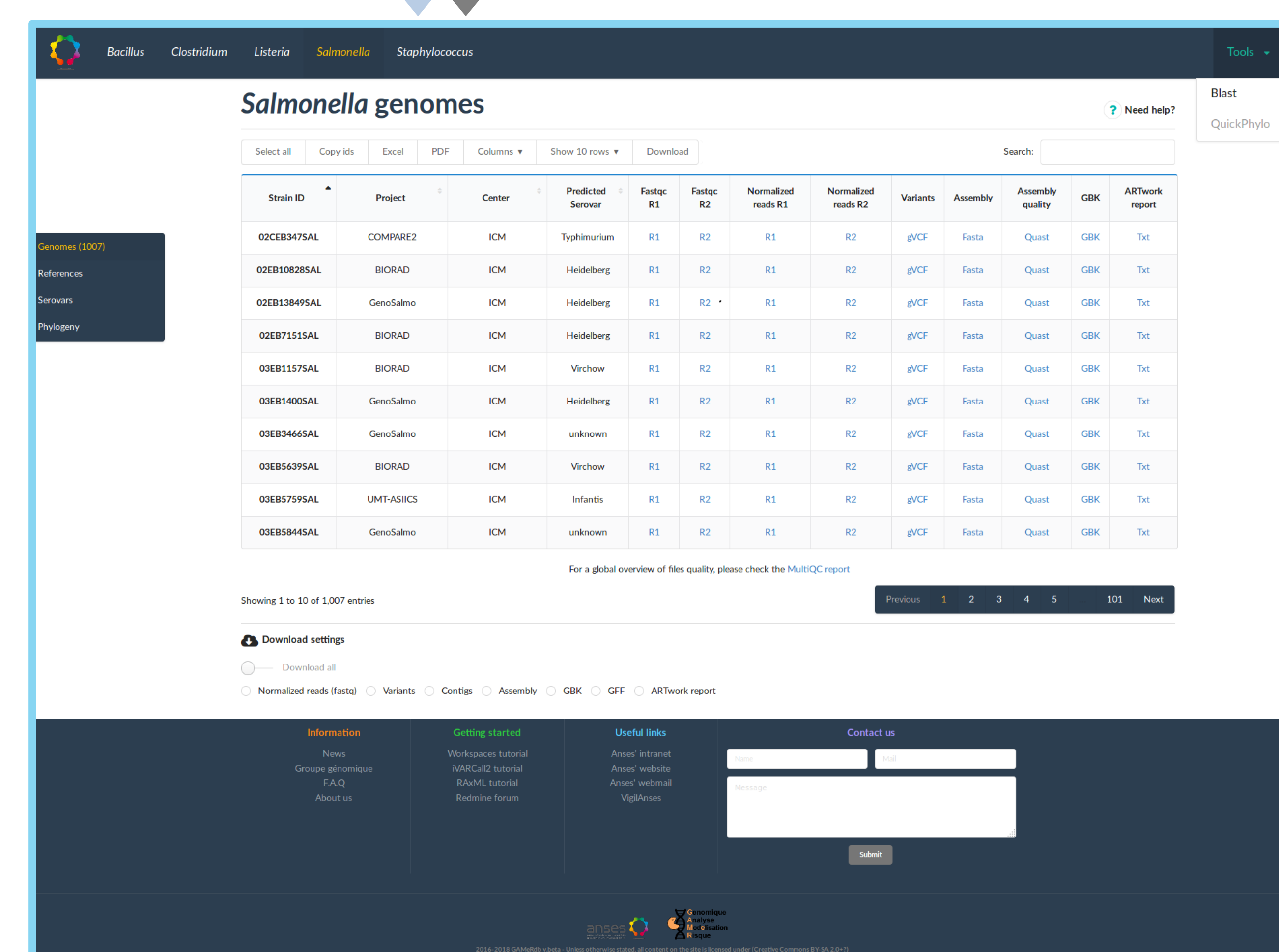

Figure 4: *Main page*

Figure 5: *Salmonella enterica genomes workspace.*

CONCLUSION

ARTwork and its **web application** strongly reduced delays between **bioinformatics data support**, **processing**, **exploration** and **recovery** for **subsequent analyses**.

Future improvements of this solution will incorporate lightweight bioinformatics tools directly usable from the web application interface.

REFERENCES

[1] Wendy A. Warr; Scientific workflow systems: Pipeline Pilot and KNIME. J Comput Aided Mol Des. 2012 Jul; 26(7): 801–804.

[2] Blankenberg D *et al*. Galaxy: a web-based genome analysis tool for experimentalists. Curr Protoc Mol Biol. 2010 Jan;Chapter 19:Unit 19.10.1-21. doi: 10.1002/0471142727.mb1910s89.

[3] http://www.applied-maths.com/bionumerics.

[4] Bradnam K et al. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. GigaScience, Volume 2, Issue 1, 1 December 2013, Pages 1–31

anses.fr