The multiple-genomes browser of the IFB cloud
Marie-Laure Franchinard, Frederic Sapet, Sandra S. Derozier, Franck F. Samson, Valentin Loux, Jean-François Gibrat

To cite this version:

HAL Id: hal-02795851
https://hal.inrae.fr/hal-02795851
Submitted on 5 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.
The multiple-genomes browser of the IFB cloud

Marie-Laure Franchinard1, Frédéric Sapei2, Sandra Derozier1, Franck Samson1, Valentin Loux1, Jean-François Gibrat1

1 INRA, UR1404 Mathématiques et Informatique Appliquées du Genôme à l’Environnement, F-78000 Jouy-en-Josas, France
2 BIOGEMMA - Upstream Genomics Group, Centre de Recherche de Chappes, CS 90126, 63720 Chappes, France

Auteur à contacter: marie-laure.franchinard@inra.fr

IGV (Integrative Genomics Viewer) [1] is a very efficient genome browser written in Java, allowing users to visualize and explore a large variety of genomic data types, but limited to a single genome. However, to perform comparative genomic studies, it is very useful to be able to observe different types of data simultaneously on several genotypes.

As part of the BioDataCloud project, which received funding from the call for proposals "Cloud computing 3 - Big data" of Investments for the Future Initiative program, a collaboration between the INRA Migale platform and the Bioemma company was established to tackle this issue.


Motivation

According to the technical specifications set by Bioemma, a new feature has been added to IGV that allows users to jump to a new genome from different types of data (genes, regions in genomic sequence, genetic markers) selected by the user on the reference genome. This jump results in the opening of a new IGV window on these data, if they are available for the new genome. This window retains all IGV features and synchronizes simultaneously with the main window.

All jumps can be saved in an IGV session file allowing users to quickly restore already used genotypes and data or to share them with other. With this new feature, the user can now compare different genotypes with the reference genome and navigate between them synchronously while keeping the IGV performance.

Development

Deployment on the IFB cloud

The deployment of this new version of IGV in the "BioDataCloud-IGV" virtual machine (appliance) allows users to benefit from:

- Large material resources in terms of CPU, RAM, storage, network
- User-friendly environment for users as well as developers
- On-demand and flexible self service
- The software is thus readily available to the whole life sciences community

Browser limits

The number of jumps achievable and therefore the number of simultaneously observable genotypes depends only upon the available hardware capabilities and the availability of the corresponding resources:

- IGV is by default RAM guzzling and this new feature requires around 4NGo where N is the number of genomes to be compared
- Large storage capacities and efficient network are required to explore large genomic data