



Genoscapist: online exploration of quantitative profiles along genomes via interactively customized graphical representations

Sandra Dérozier, Pierre Nicolas, Ulirke Mäder, Cyprien Guérin

► To cite this version:

Sandra Dérozier, Pierre Nicolas, Ulirke Mäder, Cyprien Guérin. Genoscapist: online exploration of quantitative profiles along genomes via interactively customized graphical representations. Bioinformatics, 2021, <10.1093/bioinformatics/btab079>. <hal-03131704>

HAL Id: hal-03131704

<https://hal.science/hal-03131704v1>

Submitted on 10 Feb 2021

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.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Genoscapist: online exploration of quantitative profiles along genomes via interactively customized graphical representations

Sandra Dérozier^{1,2,*}, Pierre Nicolas¹, Ulrike Mäder³, and Cyprien Guérin^{1,*}

December 22, 2020

*To whom correspondence should be addressed: sandra.derozier@inrae.fr, cyprien.guerin@inrae.fr.

¹ Université Paris-Saclay, INRAE, MaIAGE, 78350, Jouy-en-Josas, France; ² Université Paris-Saclay, INRAE, BioinfOmics, MIGALE bioinformatics facility, 78350, Jouy-en-Josas, France; and ³ Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany.

Abstract

Summary: Genoscapist is a web-tool generating high-quality images for interactive visualization of hundreds of quantitative profiles along a reference genome together with various annotations. Relevance is demonstrated by deployment of two websites dedicated to large condition-dependent transcriptome datasets available for *Bacillus subtilis* and *Staphylococcus aureus*.

Availability: Websites and source code freely accessible at <https://genoscapist.migale.inrae.fr>

1 Introduction

Analysis of quantitative genome-wide data, and in particular transcriptome data [Hör *et al.*, 2018], allows to constantly improve structural and functional annotation of bacterial genomes.

Albeit collected along the whole genome, transcriptome data are usually aggregated at gene level based on a predefined genome annotation in order to analyze differential expression. This aggregation masks local variations in signal intensity that are useful to understand technical biases or non-canonical biological mechanisms such as inefficient or early transcription termination [Bidnenko *et al.*, 2017]. More generally, examination of non-aggregated condition-dependent or strain-dependent transcription profiles along genomes is essential for delineation of transcription units and characterization of new genes such as antisense and other non-coding RNAs. Illustrative of this interest in profiles, websites (<http://genome.jouy.inra.fr/seb> and <http://genome.jouy.inra.fr/aeb>) were released with transcriptome-based reannotation studies of *Bacillus subtilis* [Nicolas *et al.*, 2012] and *Staphylococcus aureus* [Mäder *et al.*, 2016] to allow the community to further exploit the large-scale datasets generated for these studies. However, these widely used genome browsers are based on static images pregenerated using MuGeN [Hoebeke *et al.*, 2003] and thus only display a subset of transcription profiles, selected as representative of the diversity of the full dataset.

Since MuGeN, new genome browsers have been developed as independent applications such as IGV [James *et al.*, 2011] and JBrowse [Buels *et al.*, 2016], or as parts of broader projects, such as in Panoptes [Vauterin *et al.*, 2017]. Some allow visualization of multiple profiles stacked on one track but do not combine interactive profile selection and colouring with the easy access of an online tool, as useful to efficiently browse large sets of profiles. Here we present Genoscapist that answers this specific need. We designed Genoscapist as an independent application instead of a plugin to another -more generic-tool to have maximum freedom for the development and integration of graphical representation and browsing. As a proof of concept and to meet the demand of the respective scientific communities for customizable views of condition-dependent transcription profiles along genomes, we re-implemented the aforementioned websites dedicated to *B. subtilis* and *S. aureus*.

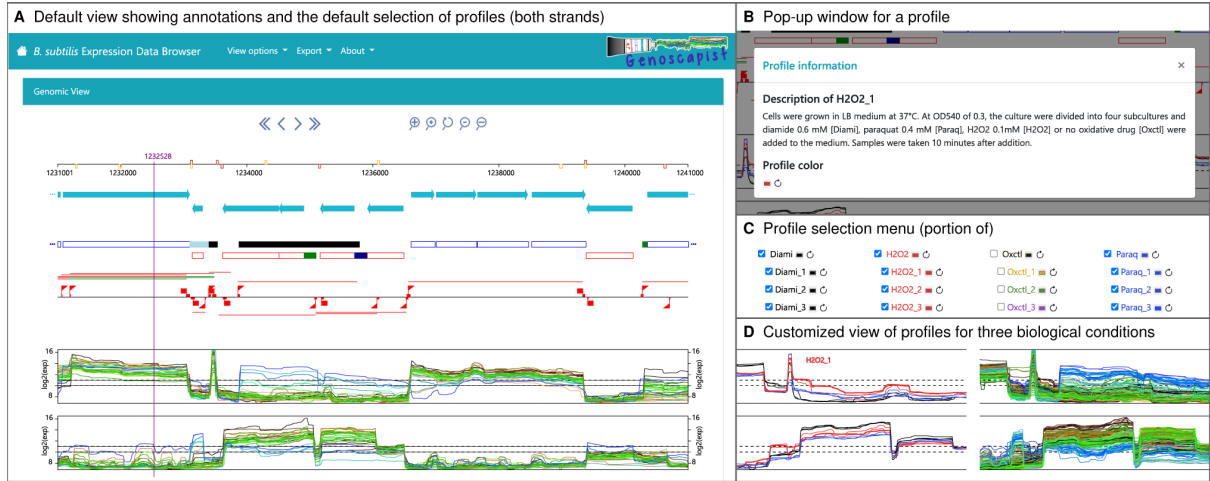


Figure 1: The genomic view of Genoscapist. Panels A to D illustrate interactive customization of the graphical representation of transcription profiles in a region of the *B. subtilis* genome.

2 Implementation

Genoscapist runs on an Apache webserver. It is written in HTML5/Javascript (client-side) and Python with Flask web framework (server-side). Using AJAX reduces response time by minimizing data transfers, by sending simultaneous server requests, and by taking advantage of the processing capability of the clients. Graphical rendering follows the Scalable Vector Graphics (SVG) Web standard.

Genome annotations and quantitative profiles are stored in a PostgreSQL database whose relational scheme is derived from FLAGdb++ [Dérozier *et al.*, 2011]. The database is typically created from a genome record (e.g. GenBank format) enriched by experimental data provided as tabulated (csv) files. Data integration is based on genomic coordinates on which indexes are built to optimize retrieval speed.

The source code, placed under the GNU General Public License, is available on a GitHub repository (<https://github.com/sderozier/genoscapist>).

3 Tool

As illustrated by the two websites <http://genoscapist.migale.inrae.fr/aeb> and <http://genoscapist.migale.inrae.fr/seb>, the main entry view consists of a list of transcribed loci (coding sequences, non-coding RNAs, 5', 3' and intra-operonic untranslated regions) hereafter collectively referred as segments, with basic information such as genomic coordinates and membership to co-expression clusters. Filters can be applied to shorten this list. Clicks on loci lead to segment-centric views from which genomic views can be accessed.

Segment-centric views present various information such as i) the genomic view centered on the segment, ii) links to entry views filtered by the co-expression clusters of the segment, iii) tables of conditions of highest and lowest expression for this segment, ii) links to other segments with highly (positively and negatively) correlated expression levels, iv) a figure of the expression level aggregated at segment-level across all profiles and v) functional annotation and direct link to an organism-oriented wiki [Zhu *et al.*, 2018, Fuchs *et al.*, 2018].

Genomic views present a default selection of expression profiles (Figure 1A) which can be customized by clicking on a profile line to change its colour (Figure 1B) or by using the “profile selection” menu (Figure 1C) to select profiles and change colours. Figures 1D and 1E show examples of customized views which can be centered on any selected segment and can also be applied when scanning the entire genome using the left/right arrows and forward/backward zooming (Figure 1A). Clicking on a segment activates the corresponding segment-centric view. Clicking elsewhere places a vertical mark and gives the exact position.

Taken together, segment-centric and genomic views provide an integrated framework to interactively explore datasets by navigating along relationships in the expression space and along the genome sequence.

The interface also allows to i) change the normalization method of the expression profiles, ii) add complementary expression data from a *rho*-null mutant impaired in factor-dependent transcription ter-

mination, iii) display all segment names, iv) export the genomic view as a SVG file, v) display a legend and vi) obtain a link to save and/or share a customized view.

4 Conclusions

Genoscapist allows interactive visualization of hundreds of quantitative profiles along a reference genome together with various annotations. Images saved in SVG format can directly be processed by vector graphics editors to assemble high-quality figures for scientific communication. Relevance is demonstrated by deployment of websites dedicated to condition-dependent transcriptomes of two important bacteria. Genoscapist can handle multi-chromosome reference genomes and the genomic view can readily display any type of quantitative profiles along genomes, such as those obtained for protein-DNA and protein-RNA interactions by immunoprecipitation.

Acknowledgements

The authors thank INRAE Migale (doi: 10.15454/1.5572390655343293E12) for hosting the websites, and Thomas Lacroix, Franck Samson, and the CATI SYSMICS group for sharing expertise.

Funding

PN and CG received support from ANR CoNoCo (ANR-18-CE12-0025).

References

- [Bidnenko *et al.*, 2017] Bidnenko, V. *et al.* (2017) Termination factor Rho: From the control of pervasive transcription to cell fate determination in *Bacillus subtilis*. *PLOS Genet.*, **13**, e1006909.
- [Buels *et al.*, 2016] Buels, R. *et al.* (2016) JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.*, **17**, 66.
- [Dérozier *et al.*, 2011] Dérozier, S. *et al.* (2011) Exploration of plant genomes in the FLAGdb++ environment. *Plant Methods*, **7**, 8.
- [Fuchs *et al.*, 2018] Fuchs, S. *et al.* (2018). AureoWiki - The repository of the *Staphylococcus aureus* research and annotation community. *Int. J. Med. Microbiol.*, **308**, 558-568.
- [Hoebeke *et al.*, 2003] Hoebeke, M. *et al.* (2003) MuGeN: simultaneous exploration of multiple genomes and analysis results. *Bioinformatics*, **19**, 859-864.
- [Hör *et al.*, 2018] Hör, J. *et al.* (2018). Bacterial RNA Biology on a Genome Scale. *Mol. Cell*, **70**, 785-799.
- [James *et al.*, 2011] James, T. *et al.* (2011) Integrative Genomics Viewer. *Nat. Biotechnol.*, **29**, 24-26.
- [Mäder *et al.*, 2016] Mäder, U. *et al.* (2016) *Staphylococcus aureus* transcriptome architecture: From laboratory to infection-mimicking conditions. *PLOS Genet.*, **12**, e1005962.
- [Nicolas *et al.*, 2012] Nicolas, P. *et al.* (2012) Condition-dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*. *Science*, **335**, 1103-6.
- [Vauterin *et al.*, 2017] Vauterin, P. *et al.* (2017) Panoptes: web-based exploration of large scale genome variation data. *Bioinformatics*, **33**, 3243-9.
- [Zhu *et al.*, 2018] Zhu, B. and Stülke, J. (2018) SubtiWiki in 2018: from genes and proteins to functional network annotation of the model organism *Bacillus subtilis*. *Nucleic Acids Res.*, **46**, D743-D748.