



Multifiscale Complex Genomics



Project Acronym: MuG

Project title: Multi-Scale Complex Genomics (MuG)

Call: H2020-EINFRA-2015-1

Topic: EINFRA-9-2015

Project Number: 676556

Project Coordinator: Institute for Research in Biomedicine (IRB Barcelona)

Project start date: 1/11/2015

Duration: 36 months

Deliverable 3.5: Tailored Implementation of browser to support pilot projects in WP7, including the interface with the Yeast Genome viewer

Lead beneficiary: Institute for Research in Biomedicine (IRB Barcelona)

Dissemination level: PUBLIC

Due date: 31/10/2017

Actual submission date: 18/12/2017

Copyright© 2015-2018 The partners of the MuG Consortium



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 676556.

Document history

Version	Contributor(s)	Partner	Date	Comments
0.1	David Castillo	CNAG-CRG	25/11/2017	First Draft
0.2	Marc A. Marti-Renom	CNAG-CRG	29/11/2017	Comments First Draft
0.3	Mike Goodstadt	CNAG-CRG	30/11/2017	Minor revisions First Draft
1.0			18/12/2017	Approved by Supervisory Board





Table of Contents

1	INTRODUCTION	5
2	MULTI-CHROMOSOME VISUALIZATION	5
3	IMPORT 3D COORDINATES	6



Executive summary

In the development of Information Technology tools for the scientific community, the close collaboration between developers and end users is always important. MuG stimulates this cooperation by the coexistence of the technical programming of the platform and the scientific contribution of three pilot projects which guide the design and functionalities of the Virtual Research Environment (VRE) and its tools.

In the case of TADkit, the development team has faced the challenge of the visualization of chromatin strands of species with very different genome sizes. Hi-C like experiments of species with small genome sizes might explore changes in whole chromosomes while experiments of species with larger genome sizes are usually focused on limited regions. To accommodate these diverse representations of chromatin structures the MuG VRE needs a tailored implementation of the 3D browser.



1 INTRODUCTION

TADkit has originally focused on the representation of 3D conformations of limited regions of the genome. This is because the modelling of structures of species with large genome sizes is usually limited to the dimension of long-range interactions in the order of a few megabase pairs (Mbp). This is the case of pilot projects WP7.1 (Senescence) and WP7.3 (Transcription factor binding and DNA bending) which are devoted to the study of the Human genome.

WP7.2 (Whole Yeast Data) studies chromatin structural changes in Yeast under different conditions. Yeast has a relatively small genome (12Mb) which allows the visualization of such changes from nucleosome positions up to multiple chromosomes. Further development has been required to allow the rendering of multiple chromosomes at the same time in TADkit. This functionality is enhanced by the possibility of inspecting inter-chromosomal contacts in the interaction matrix of the 2D panel by the selection of two chromosomes.

A further improvement which will benefit all three pilot projects is the development of a module to import 3D coordinates to the current TADkit dataset. This new feature allows the visualization of structures generated with other modelling packages outside the VRE.

2 Multi-chromosome visualization

If the dataset of 3D conformations contains information of more than one chromosome, TADkit will give the possibility of visualizing them individually or all together. To select the chromosomes a set of checkboxes have been included in the “Dataset Settings” tab [Figure].

The inclusion of non-selected chromosomes will trigger the update of the 1D, 2D and 3D panels.

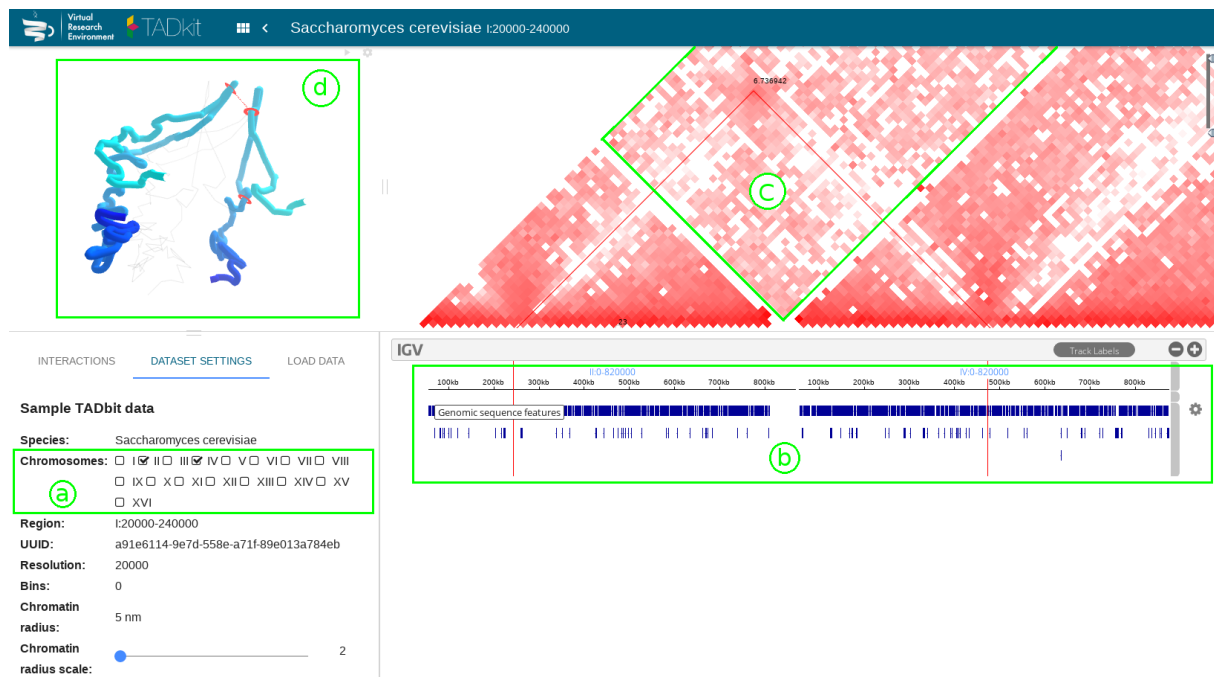


Figure 1: (a) Selection of chromosomes. (b) Split 1D tracks. (c) Inter-chromosomal interaction matrix. (d) Combined 3D structures.



The 1D panel will be restructured depending on whether one, two or more than two chromosomes are selected. In the case of one chromosome the panel will contain, as usual, the lineal features of the selected chromosome aligned with the 2D panel. If two chromosomes are selected the 1D panel will be split in two equally sized windows [Figure], one for each of the chromosomes. This conformation will allow the user to spot inter-chromosomal interactions by clicking in the interaction matrix.

If more than two chromosomes are selected, the panel will just display features of the first chromosome in the list leaving blank the remaining space below the interaction matrix of the other chromosomes.

The 2D panel will display the interaction matrix of the selected chromosomes in the order listed in the “Dataset Settings” tab. Inter-chromosomal contacts will also be represented [Figure].

The 3D panel will include the structures of the chromosomes selected [Figure]. Each one of them appear colored with a blue gradient to identify its start (light blue) and end (dark blue).

The “Interactions” tab will contain information about distance and features of two interacting loci both within and between the displayed chromosomes [Figure].

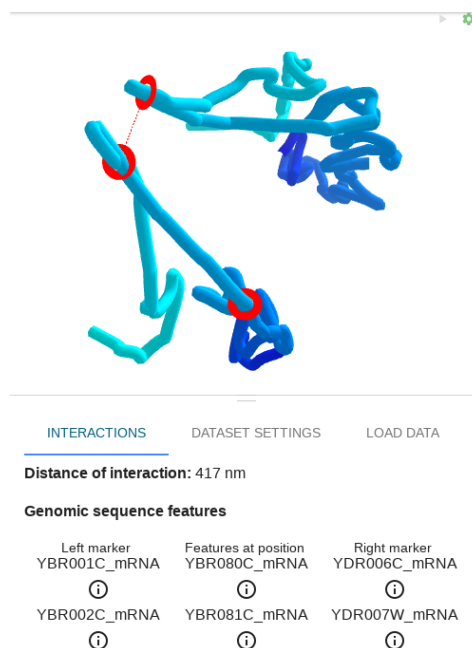
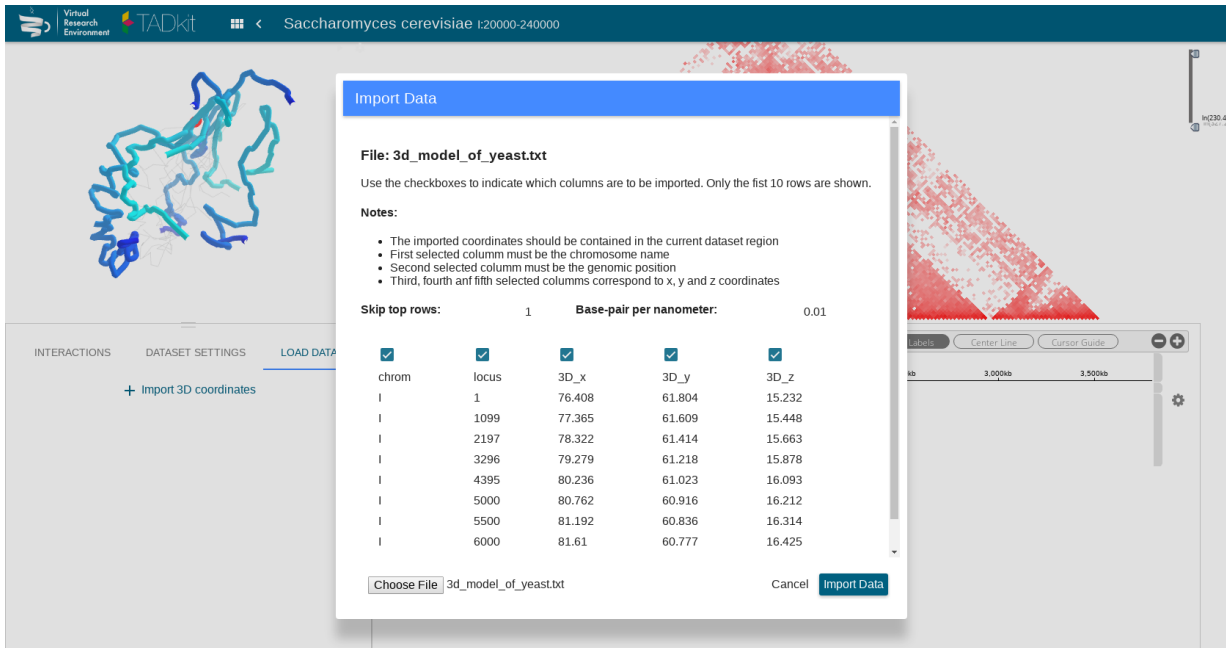


Figure 2: Two chromosomes and its interaction information

3 Import 3D coordinates

TADbit, included in the MuG VRE, computes and clusters models that best fit the interaction matrix of a limited region of the genome obtained in the 3C-based experiment and saves the resulting information in a JSON dataset that can be viewed in TADkit. To accommodate for visualizing 3D models built with other methods beyond TADbit, we equipped TADkit has been equipped with a module to import 3D coordinates from an external source. The new module can be accessed in the “Load Data”

tab of the Information panel. Clicking on the “Import 3D coordinates” button a form is presented to allow the user uploading the file with the coordinates [Figure].



Import Data

File: 3d_model_of_yeast.txt

Use the checkboxes to indicate which columns are to be imported. Only the first 10 rows are shown.

Notes:

- The imported coordinates should be contained in the current dataset region
- First selected column must be the chromosome name
- Second selected column must be the genomic position
- Third, fourth and fifth selected columns correspond to x, y and z coordinates

Skip top rows: 1 Base-pair per nanometer: 0.01

<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
chrom	locus	3D_x	3D_y	3D_z
I	1	76.408	61.804	15.232
I	1099	77.365	61.609	15.448
I	2197	78.322	61.414	15.663
I	3296	79.279	61.218	15.878
I	4395	80.236	61.023	16.093
I	5000	80.762	60.916	16.212
I	5500	81.192	60.836	16.314
I	6000	81.61	60.777	16.425

Choose File: 3d_model_of_yeast.txt Cancel Import Data

Figure 3: Import 3D coordinates form

The import of structures is limited to the species, chromosomes and genomic range of the loaded dataset. The resolution of the model in the file should be equal or higher than the resolution of the loaded dataset. If the resolution of the file is higher than the used in the dataset the coordinates will be averaged to adapt to lower resolution beads.

An important parameter to specify is the “Base-pairs per nanometer” which refers to an approximation of the fraction of base-pairs of chromatin contained in one nanometer.

In order to have high degrees of compatibility with possible external sources, the development team has chosen to support a simple Tab-Separated Values (TSV) format in which each line will represent a genomic location and its 3D coordinates.

The TSV file should at least contain the following columns in the appropriate order:

- Chromosome: a string representation of the chromosome name to which the loci belongs. The name should be consistent with the one used in the dataset.
- Locus: an integer with the genomic position in base-pairs.
- x coordinate: a floating point number with the x coordinate in nanometers.
- y coordinate: a floating point number with the y coordinate in nanometers.
- z coordinate: a floating point number with the z coordinate in nanometers.