



HAL
open science

circHiC: circular visualization of Hi-C data and integration of genomic data

Ivan Junier, Nelle Varoquaux

► **To cite this version:**

Ivan Junier, Nelle Varoquaux. circHiC: circular visualization of Hi-C data and integration of genomic data. 2021. hal-02945275

HAL Id: hal-02945275

<https://hal.science/hal-02945275>

Preprint submitted on 16 Nov 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.

circHiC: circular visualization of Hi-C data and integration of genomic data

Ivan Junier and Nelle Varoquaux

CNRS, University Grenoble Alpes, TIMC-IMAG, Grenoble, France

Summary: Genome wide contact frequencies obtained using Hi-C-like experiments have raised novel challenges in terms of visualization and rationalization of chromosome structuring phenomena. In bacteria, display of Hi-C data should be congruent with the circularity of chromosomes. However, standard representations under the form of square matrices or horizontal bands are not adapted to periodic conditions as those imposed by (most) bacterial chromosomes. Here, we fill this gap and propose a Python library, built upon the widely used Matplotlib library, to display Hi-C data in circular strips, together with the possibility to overlay genomic data. The proposed tools are light and fast, aiming to facilitate the exploration and understanding of bacterial chromosome structuring data. The library further includes the possibility to handle linear chromosomes, providing a fresh way to display and explore eukaryotic data.

Availability and implementation: The package runs under Python 3 and is freely available at <https://github.com/TrEE-TIMC/circHiC>. The documentation can be found at <https://tree-timc.github.io/circHiC/>; images obtained in different organisms are provided in the gallery section and are accompanied with codes.

Contact: ivan.junier@univ-grenoble-alpes.fr, nelle.varoquaux@univ-grenoble-alpes.fr

Keywords: visualization, Hi-C

I. INTRODUCTION

Genome wide high-throughput chromosome conformation capture (Hi-C) methods (Lieberman-Aiden *et al.* 2009) and related techniques (Denker and De Laat 2016) allow to compute, up to an unknown global factor, contact frequencies between any two loci along a genome. Rationalization of chromosome structuring phenomena requires, in the first place, an appropriate visualization of these data. Several challenges along this line have been raised and (sometimes) solved. For instance, eukaryotic data can be browsed at different scales, under the form of square matrices (Durand *et al.* 2016, Kerpedjiev *et al.* 2018, Yardimci and Noble 2017).

Compared to eukaryotes, bacterial chromosomes are much smaller, yet they also harbor a complex multilayer organization (Lagomarsino *et al.* 2015). Most importantly, most bacterial chromosomes are circular. From the perspective of Hi-C visualization, this raises the ineluctable problem of representing a physical system with spherical symmetry using a two-dimensional Euclidean space. That is, the standard square matrix or horizontal band representations commonly used for Hi-C data are not appropriate for a system with periodic conditions as those associated with bacterial chromosomes. In particular, there is currently no tool, to the best of our knowledge, allowing to display Hi-C heat maps in a circular strip.

Here, we fill this gap by developing a Python library, `circHiC`, that efficiently displays bacterial Hi-C data accordingly. The library further includes the possibility to overlay genomic data. It also includes the possibility to handle linear chromosomes and, hence, can be used to

visualize eukaryotic data.

II. ALGORITHM

Let H be the input Hi-C matrix (size $N \times N$). Our algorithm consists in using a system of polar coordinates to project H onto a circular strip (Figure 1). Specifically, a circle of the strip corresponds to all contact frequencies between pairs of loci that are separated by a certain genomic distance, s . In this context, `circHiC` first consists in specifying the genomic distances s_{in} and s_{out} for the inner and outer circles – we note here that `circHiC` displays by default, in the inward and outward parts of the strip, contacts corresponding to $s \in [0, s_{in}]$ and to $s \in [0, s_{out}]$, respectively. We then consider, in each of these two parts of the strip, a linear relationship between the radius of a circle, r_s , and the corresponding genomic distance, s . Altogether, we eventually have $s = \left| -s_{in} + (s_{out} + s_{in}) \frac{r_s - r_{in}}{1 - r_{in}} \right|$ for the entire strip, with r_{in} the radius of the inner circle and where all the radii are normalized by the radius of the outer circle (i.e. $r_{out} = 1$). Next, any pair of loci can be written as $x - \frac{s}{2}$ and $x + \frac{s}{2}$, with x and s the respective associated middle locus and genomic distance – note that loci (i.e. indexes of H) are considered modulo N . The associated contact frequency is then properly positioned along its associated circle (with radius r_s) by specifying the polar angle of x . To this end, we use a clockwise, twelve o'clock origin angle θ_x that is linear in x : $\theta_x = \frac{\pi}{2} + \frac{2\pi(x-x_0)}{L}$. x_0 , which can be set by the user, corresponds to the middle locus of the pairs of loci whose contact frequencies are displayed vertically.

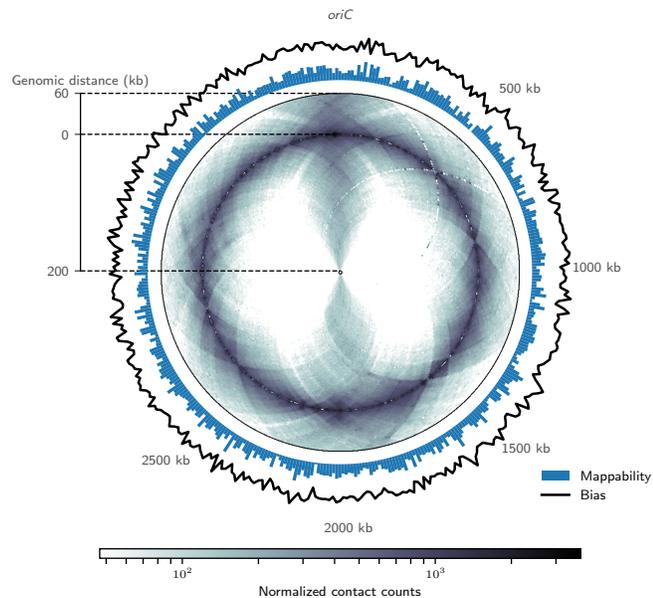


FIG. 1 Circular visualization of Hi-C data obtained in *C. crescentus* (Le *et al.* 2013) generated using `circHiC` in a default ‘reflect’ mode, using $s_{in} = 200$ kb and $s_{out} = 60$ kb. Two additional genomic data (the bias associated with the ICE normalization of data (Imakaev *et al.* 2012) in black lines and cumulative raw contact counts in black bars) are displayed outside the Hi-C data.

Using this framework, `circHiC` consists in filling up the entries of a matrix C (size $N_c \times N_c$) that are located in the circular strip centered in $(\frac{N_c}{2}, \frac{N_c}{2})$ with inner radius $\frac{N_c}{2}r_{in}$ and outer radius $\frac{N_c}{2}$, by considering the corresponding entries of H . To this end, we consider polar coordinates associated with the circular strip and we use, in place of the trigonometric angle, the above clockwise angle with twelve o’clock origin. Each entry (i_c, j_c) of the circular strip in C is thus associated with a coordinate (r_c, θ_c) with $r_c = \sqrt{\left(\frac{i_c}{N_c} - \frac{1}{2}\right)^2 + \left(\frac{j_c}{N_c} - \frac{1}{2}\right)^2}$ and θ_c such that $\frac{i_c}{N_c} - \frac{1}{2} = r_c \sin \theta_c$ and $\frac{j_c}{N_c} - \frac{1}{2} = r_c \cos \theta_c$. Given (r_c, θ_c) and following the discussion of the previous paragraph, we then set $C[i_c, j_c] = H[x_c - \frac{s_c}{2}, x_c + \frac{s_c}{2}]$ with $x_c = x_0 + \frac{(\theta_c - \frac{\pi}{2})L}{2\pi}$ and $s_c = \left| -s_{in} + (s_{out} + s_{in}) \frac{r_c - r_{in}}{1 - r_{in}} \right|$.

Note that, just as with square displays, data on each side of $s = 0$ are redundant (here up to $\min(s_{in}, s_{out})$), generating a reflection effect. This allows for instance to enhance the presence of chromosomal interaction domains as in the case of *Caulobacter crescentus* (Le *et al.* 2013) (Figure 1). In addition to this ‘reflect’ mode, `circHiC` provides a ‘distant’ mode to highlight contacts away from the main diagonal (see gallery in the online documentation). Note finally that N_c is larger or equal to N . We thus define the granularity of the circular projection by N/N_c , equal to 0.5 by default. The smaller the granularity is, the neater the circular display, at the cost of a longer computation.

III. IMPLEMENTATION

We built upon the popular visualization library Matplotlib (Hunter 2007), using core libraries of the scientific Python ecosystem (`numpy`, `pandas`, ...). `circHiC` relies on an object oriented approach, where a figure (the core object of the library) holds all the necessary elements for the visualization: the size of the genome, the orientation of the plot, and a number of information relating to the size and shape of the figure. A number of methods can then be executed to visualize the data: `plot_hic` provides support for transforming the square contact count matrix into a circular one, and `plot_lines`, `plot_bars`, ... allow to overlay relevant genomic information onto the contact count matrix.

IV. RESULTS

As an example, in Figure 1 we show a circular representation of the first high-resolution Hi-C data reported in a bacterial chromosome (Le *et al.* 2013) together with genomic data. In the gallery section of the online documentation, we provide additional examples using publicly available data, including the visualization of a linear human chromosome.

V. CONCLUSION

`circHiC` provides a useful, fast representation of Hi-C data that respects the circularity of bacterial chromosomes. The possibility to overlay genomic information aims at facilitating the exploration and understanding of chromosome structuring data. In practice, the build upon Matplotlib allows a great flexibility to generate complex figures.

REFERENCES

- Denker, A. and De Laat, W. (2016). The second decade of 3C technologies: detailed insights into nuclear organization. *Genes & Development*, **30**(12), 1357–1382.
- Durand, N. C. *et al.* (2016). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Systems*, **3**(1), 95–98.
- Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, **9**(3), 90–95.
- Imakaev, M. *et al.* (2012). Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nature methods*, **9**(10), 999–1003.
- Kerpedjiev, P. *et al.* (2018). HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome biology*, **19**(1), 125.
- Lagomarsino, M. C. *et al.* (2015). From structure to function of bacterial chromosomes: Evolutionary perspectives and ideas for new experiments. *FEBS Letters*, **589**(20 Pt A), 2996–3004.
- Le, T. B. K. *et al.* (2013). High-resolution mapping of the spatial organization of a bacterial chromosome. *Science*, **342**(6159), 731–734.
- Lieberman-Aiden, E. *et al.* (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**(5950), 289–293.
- Yardimci, G. G. and Noble, W. S. (2017). Software tools for visualizing Hi-C data. *Genome biology*, **18**(1), 26.