

SOFTWARE

Open Access



# EVR: reconstruction of bacterial chromosome 3D structure models using error-vector resultant algorithm

Kang-Jian Hua and Bin-Guang Ma\*

## Abstract

**Background:** More and more 3C/Hi-C experiments on prokaryotes have been published. However, most of the published modeling tools for chromosome 3D structures are targeting at eukaryotes. How to transform prokaryotic experimental chromosome interaction data into spatial structure models is an important task and in great need.

**Results:** We have developed a new reconstruction program for bacterial chromosome 3D structure models called EVR that exploits a simple Error-Vector Resultant (EVR) algorithm. This software tool is particularly optimized for the closed-loop structural features of prokaryotic chromosomes. The parallel implementation of the program can utilize the computing power of both multi-core CPUs and GPUs.

**Conclusions:** EVR can be used to reconstruct the bacterial 3D chromosome structure based on the contact frequency matrix derived from 3C/Hi-C experimental data quickly and precisely.

**Keywords:** Chromatin architecture, Hi-C, Prokaryotes, 3D genome, Structure modelling

## Background

Like eukaryotes, the prokaryotic genomes also fold with regular patterns. The spatial organization plays very important roles in many cellular activities such as transcriptional regulation [1], gene expression [2] and DNA replication [3]. Many approaches, e.g., FISH (Fluorescence in situ hybridization) [4] and 3C (chromosome conformation capture) technology [5] have been adopted to study the 3D architecture of prokaryotic chromosome. In FISH experiments, fluorescence is generated by binding a labeled probe to a specific DNA fragment, and the position of target DNA fragment in cell can be visually observed under a microscope [6]. However, due to the low throughput and limited resolution of FISH, it is difficult to be applied in the genome-wide study of chromosome structure. The 3C technology and its derivative technologies such as 4C (circular chromosome conformation capture) [7], 5C (3C-carbon copy) [8] and Hi-C (High-throughput/resolution chromosome conformation capture) [9] have significantly promoted the

research of three-dimensional genomics in the past decade [1]. In 3C technology, the genome is cross-linked by formaldehyde and then broken into fragments which are further ligated by enzymes and sequenced for subsequent analysis [10]. In the process of cross-linking, the probability that the fragments close to each other in space are connected is high. This feature is an important clue to study the spatial structure of chromosome [11]. In 2011, Umbarger et al. used 5C technology to generate a genome-wide DNA interaction map of *Caulobacter crescentus* with fragment size of approximately 13 kb [12], and established the 3D structure of chromosome using the integrated modeling platform (IMP) [13, 14]. Subsequently in 2013, Cagliero et al. obtained the chromosome interaction data of *E. coli* [15], but the final interaction map contained a large amount of noise, and it was difficult to use the data to reconstruct structures. In the same year, Le et al. produced the full-genome interaction maps of 10 kb resolution for *C. crescentus* using Hi-C and built its spatial structure with their own method [16]. With this research, they demonstrated the existence of CID (chromosomal interaction domains) boundaries and their roles in gene expression. In 2015, Marbouty et al. applied Hi-C to the chromosome

\* Correspondence: [mbg@mail.hzau.edu.cn](mailto:mbg@mail.hzau.edu.cn)

Hubei Key Laboratory of Agricultural Bioinformatics, College of Informatics, State Key Laboratory of Agricultural Microbiology, Huazhong Agricultural University, Wuhan 430070, China



research of *Bacillus subtilis* [17] and reconstructed 3D models with the ShRec3D program [18], and then they combined the 3D genome structure with super-resolution microscopy to unravel the role of chromosome 3D folding pattern in the regulation of replication initiation, chromosome organization and DNA segregation. In the meanwhile, Wang et al. also studied *B. subtilis* with Hi-C and a large number of chromosome interaction maps were generated at various physiological stages [19, 20], and their research revealed the mechanism of SMC condensin complexes in compacting and resolving replicated chromosomes. Recently, Liou et al. succeeded in applying Hi-C in the study of *E. coli* and obtained interaction maps in a high resolution of 5 kb, and the 3D chromosome structures were reconstructed with the ShRec3D program [18, 21] to demonstrate the role of nucleoid associated proteins (NAPs) in chromosome organizing.

Years of investigation have yielded many discoveries of bacterial chromosome features, such as cyclic pseudo-nucleation [22], supercoiled regions [23] and plectoneme structures [16], etc. With the knowledge of these features, scientists can build the structure of bacterial chromosome by combining experimental data with elaborately developed calculation methods. In 2017, a research work divided *E. coli* chromatin into plectoneme-abundant regions and plectoneme-free regions based on ChIP-chip data [24], and built its structure by using multi-scale modeling and Brownian dynamics approaches, resulting in a chromosome model with a resolution of 1NTB (nucleotide per bead) [25]. More recently, a lattice model of *E. coli* chromatin with resolution of 20 bp was constructed by giving the position and size of the plectoneme fragments [26] using a mesoscale modeling pipeline method [27], and the chromosome structure model of *C. crescentus* with 15 bp resolution was obtained by using Hi-C data combined with Monte Carlo sampling and molecular dynamics [28].

With the application of 3C technology, more and more chromosome interaction maps of prokaryotes emerge and the 3D models of chromosome structure are in great need for the studies of chromosome function and transcriptional regulation. Several software tools have been published for the reconstruction of chromosome 3D structures. These tools can be classified into two categories according to their modeling strategies: modeling based on restraints and modeling based on thermodynamics [29]. Restraint-based modeling is a commonly used method to derive 3D models from DNA interaction data. In restraint-based modeling, the DNA interaction data is first converted to the expected space distance, and then heuristic methods are used to optimize the structure under given restraint conditions, such as the gradient ascent method in MOGEN [30], Monte Carlo sampling in TADbit [31], multidimensional scaling algorithm in miniMDS [32], and shortest-path method in ShRec3D [18]. In these methods, ShRec3D [18]

has been applied to reconstruct the chromosome 3D structures of two model bacteria species (*B. subtilis* and *E. coli*) [17, 21], but it is sensitive to resolution and highly depends on data normalization.

In this study, a new algorithm was proposed specially for the chromosome structure reconstruction of prokaryotes, which used error vectors to guide structural optimization and added constraints of a circular genome shape on the implementation of program. This program can robustly deal with prokaryotic data and has high calculation speed by fully utilizing the computing power of multi-core CPUs and GPUs. With this algorithm, the 3D structures of the chromosomes of three prokaryotic model organisms (*C. crescentus*, *B. subtilis* and *E. coli*) were reconstructed and analyzed.

## Implementation

### Data source

The sequencing data from 3C/Hi-C experiments can be processed by pipelines such as HiC-Pro [33] or HiCUP [34] to obtain chromosome interaction information. In these pipelines, genomes are divided into equal-length segments and each segment is called a bin. The interaction information of bins is generally presented in the form of a matrix whose element is the number of interactions between two bins. Then the matrix will be normalized to remove some biases such as GC content, mappability, and fragment length, etc. [35], resulting in the interaction frequency (IF) matrix, which will be used to construct the spatial structure of chromosome (namely, determine the spatial coordinates of each bin).

We used the published chromosome interaction data of three bacteria species: normalized chromosome interaction frequency matrix (referred to as IF matrix) with 10 kb resolution of *Bacillus subtilis* (GEO accession number: GSE68418) [19], normalized IF matrix with 10 kb resolution of *Caulobacter crescentus* (GEO accession number: GSE45966) [16], and the raw IF matrix with 5 kb resolution of *Escherichia coli* (GEO accession number: GSE107301) [21].

### Algorithm

The EVR algorithm consists of the following steps: (i) transform IF matrix to the expected distance matrix  $D$  by Eq. 1; (ii) an initial conformation is generated by randomly assigning coordinates to DNA bins; (iii) calculate error vectors for each bin; (iv) a new conformation is obtained by moving bins according to the guidance of error vectors; (v) repeat steps (iii) and (iv) until the iteration stop condition is satisfied. The chromosome structure is the final conformation (coordinates of bins) after iteration (Fig. 1).

### Generation of the expected distance matrix

In cross-linked chromatin, the closer the spatial distance is, the higher the probability of cross-linking is. Therefore,

Eq. 1 is usually used to express the relationship between interaction frequency and the expected spatial distance [11]. In Eq. 1,  $F_{ij}$  is the interaction frequency of bin  $i$  and bin  $j$  ( $1 \leq i, j \leq N$  where  $N$  is bin number), and  $D_{ij}$  is the corresponding expected spatial distance. If  $F_{ij}$  is not equal to 0, a power function is used to convert  $F_{ij}$  into  $D_{ij}$ ; otherwise, the power function is invalid, and then the distance constraint of adjacent bins will be used to optimize the coordinates. Here, the power exponent  $\alpha$  is set to 0.5 by reference [21].

$$D_{ij} = \begin{cases} \left(\frac{1}{F_{ij}}\right)^\alpha & F_{ij} \neq 0 \\ \infty & F_{ij} = 0 \end{cases} \quad (1)$$

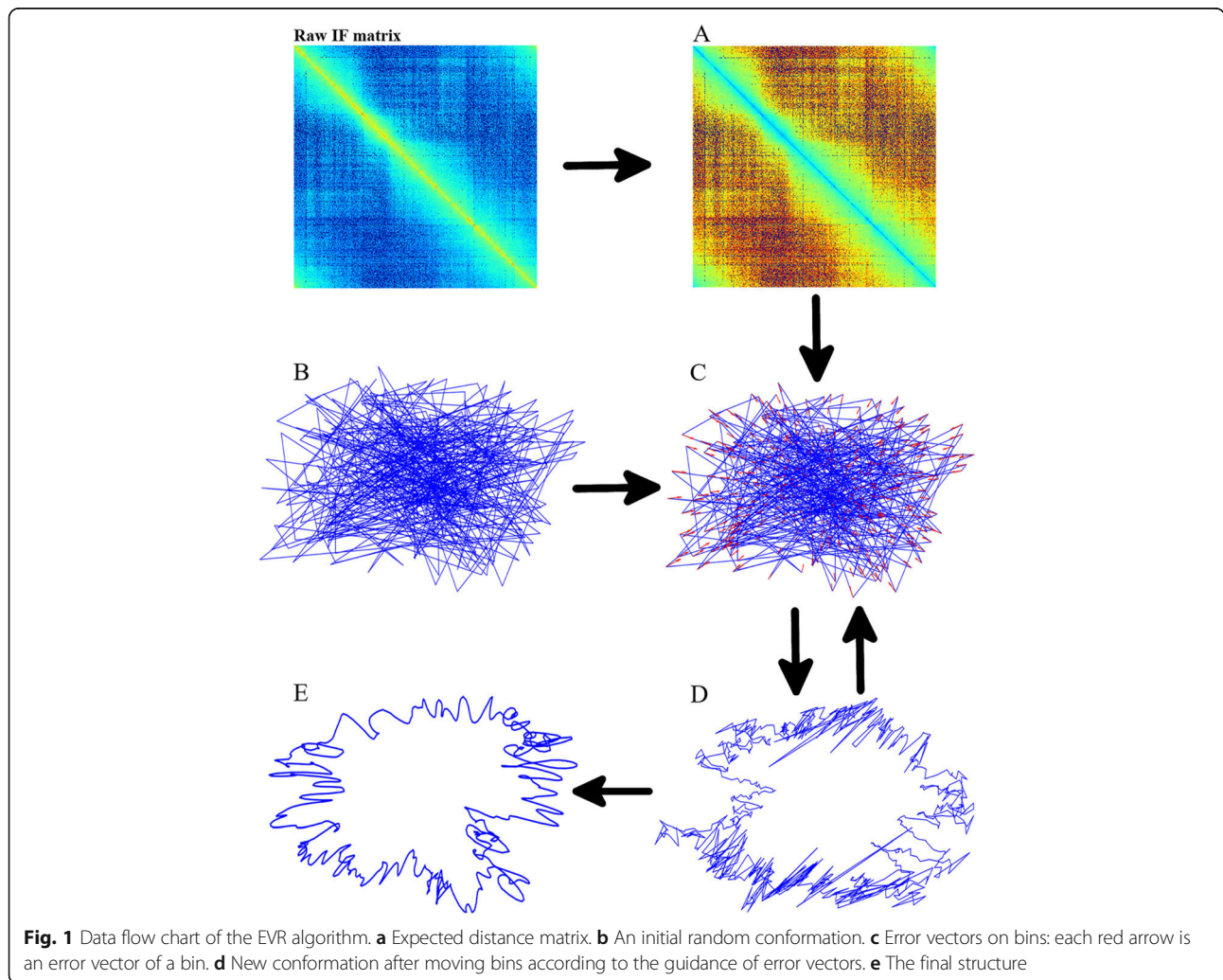
**Error vectors**

An initial chromosome conformation is generated by randomly assigning coordinates to all bins. With the distance matrix  $D$ , error vectors are calculated between every two bins. The error vectors are expressed as in Eq. 2 for bins that are not directly

adjacent on the linear genome. Here,  $\vec{P}_i$  and  $\vec{P}_j$  are the coordinates of bin  $i$  and bin  $j$  in vector form, respectively, and  $D_{ij}$  (Eq. 1) is the expected distance between the two bins.

$$\vec{e}_{ij} = \frac{\vec{P}_i - \vec{P}_j}{|\vec{P}_i - \vec{P}_j|} \cdot (|\vec{P}_i - \vec{P}_j| - D_{ij}) \quad (2)$$

For bins that are directly adjacent on the linear genome (namely,  $|i - j| = 1, 1 \leq i, j \leq N$ ), error vectors are defined as in Eq. 3, where  $D_{max}$  and  $D_{min}$  are the allowed maximum and minimum distances between any two adjacent bins, respectively. In a conformation,  $D_{max}$  and  $D_{min}$  will be used to calculate the error vector between two adjacent bins if the distance of the two bins is less than  $D_{min}$  or greater than  $D_{max}$ ; otherwise, the expected distance  $D_{ij}$  is used. Owing to the fact that most prokaryotic chromosomes are circular, the first bin and the last bin are set as adjacent bins (namely,  $(i-j)\%N = 1$  if  $i = 1, j = N$  or  $(j-i)\%N = 1$  if  $i = N, j = 1$ , where % is



**Fig. 1** Data flow chart of the EVR algorithm. **a** Expected distance matrix. **b** An initial random conformation. **c** Error vectors on bins: each red arrow is an error vector of a bin. **d** New conformation after moving bins according to the guidance of error vectors. **e** The final structure

modulus operator), which means the error vectors as defined in Eq. 3 also exist between the first and the last bins.

$$\vec{e}_{ij} = \begin{cases} \frac{\vec{P}_i - \vec{P}_j}{|\vec{P}_i - \vec{P}_j|} \cdot (|\vec{P}_i - \vec{P}_j| - D_{min}) & \text{if } |\vec{P}_i - \vec{P}_j| < D_{min} \\ \frac{\vec{P}_i - \vec{P}_j}{|\vec{P}_i - \vec{P}_j|} \cdot (|\vec{P}_i - \vec{P}_j| - D_{ij}) & \text{if } D_{min} < |\vec{P}_i - \vec{P}_j| < D_{max} \\ \frac{\vec{P}_i - \vec{P}_j}{|\vec{P}_i - \vec{P}_j|} \cdot (|\vec{P}_i - \vec{P}_j| - D_{max}) & \text{if } |\vec{P}_i - \vec{P}_j| > D_{max} \end{cases} \quad (3)$$

For each bin, all its error vectors are summed up with certain weights to a resultant vector (EVR). As shown in Eq. 4,  $\vec{E}_j$  is the EVR for bin  $j$  which indicates the direction and distance that bin  $j$  should move in the next iteration step (Fig. 2a) and a weight  $w_{ij}$  determines the step size. When  $w_{ij}$  is large, the length of EVR is large, and the structure changes greatly compared with the previous one after one iteration step, and the optimization speed is fast, but the result may not converge. Conversely, when  $w_{ij}$  is small, the structural change in each iteration step is small, and the optimization speed is slow, but it is easier to converge. In our algorithm,  $w_{ij} = 1/N$ , where  $N$  is the number of bins.

$$\vec{E}_j = \sum_{i=0}^N w_{ij} \cdot \vec{e}_{ij} \quad (4)$$

An example of EVR calculation is shown in Fig. 2a, in which black arrows are error vectors that can be calculated according to Eq. 2 and Eq. 3 and red arrows are the resultant vectors calculated according to parallelogram law; these resultant vectors can guide the position optimization of bins.

### Iterative optimization

Starting with an initial random conformation, error vectors and resultant vectors can be calculated based on the expected distance matrix for all bins. The resultant vector of each bin indicates the direction and distance according to which the bin should move to change its position towards the target position. The chromosome structure is optimized by an iteration process during which all bins change their positions (coordinates) guided by their resultant vectors, resulting in the minimization of the sum of EVR lengths (Eq. 5). For example, the motion trajectories of the 3 bins  $P_{j1}, P_{j2}, P_{j3}$  are shown as light blue circles (Fig. 2a) and the distances of the 3 bins to their target positions decrease quickly during iteration (Fig. 2b).

$$F = \sum_{j=1}^N |\vec{E}_j| \quad (5)$$

$$\Delta F = |F_t - F_{t-1}| \quad (6)$$

In the iterative process, the value of  $F$  becomes smaller and smaller to approach a stable value. As shown in Eq. 6,  $\Delta F$  is defined as the difference between two successive  $F$  values of iterative steps  $t$  and  $t-1$ . When a level of  $\Delta F < 1E-5$  or the given maximum number of steps is reached, the iteration stops and an optimized conformation is obtained which is the final structure of a chromosome. A general issue in optimization is local optima; see Supplementary Information (Additional file 1) for the analysis of our algorithm in this respect.

### Programming

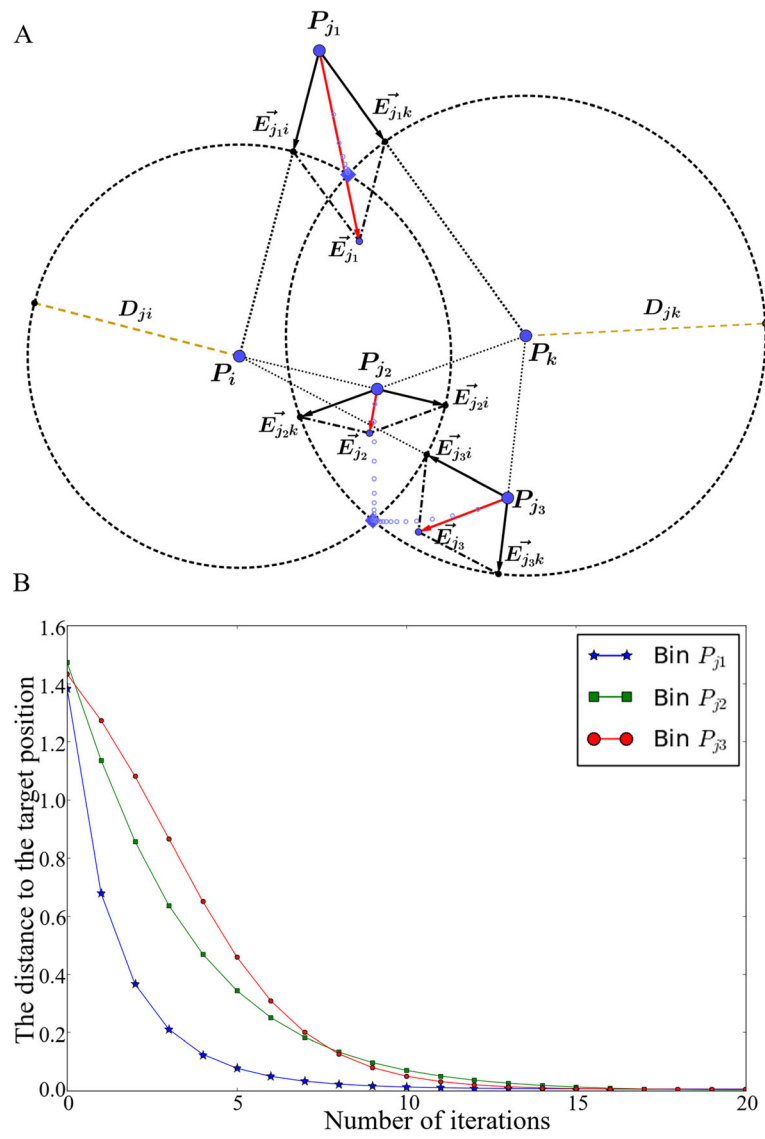
The main part of EVR program is implemented in Python with Numpy arrays for storing data. Two acceleration schemes are adopted in the iterative optimization part: Cython and OpenCL. In Cython scheme, the core code is written in Cython-C language and compiled into a dynamic library that can be accelerated by OpenMP. The Cython scheme can make full use of the computing power of multi-core CPUs. In OpenCL scheme, the core code is written in OpenCL-C language, and it can be compiled and invoked directly by the PyOpenCL library functions. The OpenCL scheme can use not only multi-core CPUs but also GPUs and other devices that support OpenCL to accelerate computing.

### Results

First, we compared our EVR program with some published software tools for chromosome reconstruction, including: miniMDS [32], MOGEN [30] and ShRec3D [18]. All these software tools ran on multi-core CPUs with default parameter setting. The comparisons involve the speed, accuracy and sensitivity to noise level of calculations. Subsequently, our EVR program was applied to the chromosome structure reconstruction of 3 model species of prokaryotes with published 3C/Hi-C data. The obtained chromosome structures were analyzed in a biological background.

### EVR is fast

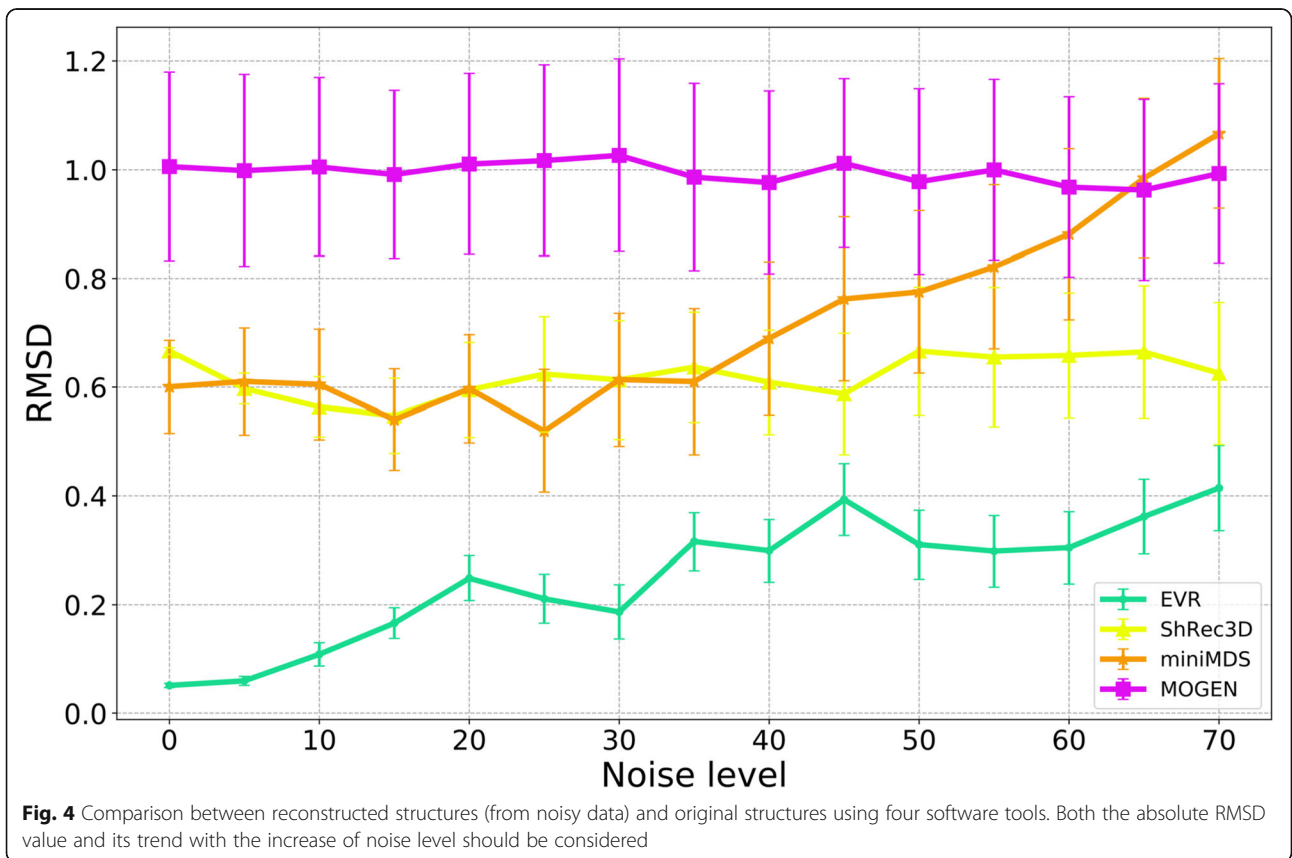
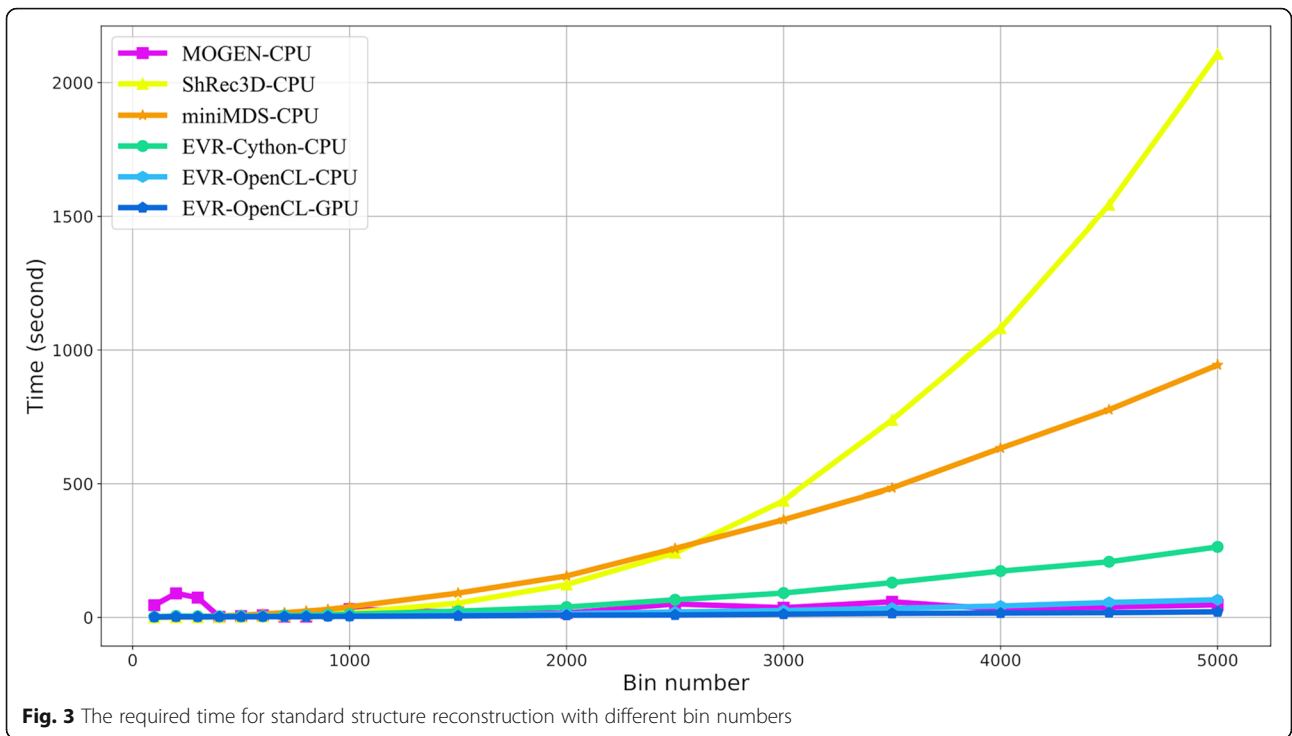
Among the compared software tools, miniMDS is written in Python and accelerated using the pypm library, and the computational complexity of a single iteration is  $O(N^2)$ ; MOGEN is written in java and can use multi-core CPUs for accelerating calculation and the computational complexity of a single iteration is  $O(N^2)$ ; ShRec3D is written in MATLAB, and the program itself is not parallelized, but the new version of MATLAB supports automatic parallelization. Because eigenvalue and feature vector decomposition are used in the algorithm, the computational complexity of



**Fig. 2** EVR calculation. **a** Illustration of error-vectors and error-vector results.  $P_{j1}$ ,  $P_{j2}$ , and  $P_{j3}$  are 3 bins whose positions (coordinates) need to be adjusted;  $P_i$  and  $P_k$  are 2 bins whose positions are fixed;  $D_{ji}$  and  $D_{jk}$  are the expected distances between these 3 bins to bins  $P_i$  and  $P_k$ ; the intersection points of the two dotted circles (centered at  $P_i$  and  $P_k$  with radii  $D_{ij}$  and  $D_{jk}$ , respectively) are the target positions of the 3 bins  $P_{j1}$ ,  $P_{j2}$ , and  $P_{j3}$ ; the closest intersection point is chosen as the target position for each of the 3 bins. **b** The distances between the 3 bins  $P_{j1}$ ,  $P_{j2}$ ,  $P_{j3}$  to their target positions decrease during iteration

ShRec3D is  $O(N^3)$ . The computational complexity of EVR's single iteration is also  $O(N^2)$ . Following convention [11], standard structures were generated for algorithm assessment. Here, toroidal spiral curves were generated and cut into fragments (bins) with bin numbers from 100 to 5000, and the IF matrixes of these structures were calculated to simulate a prokaryotic circular chromosome. Then, the above software tools were used to reconstruct the standard structures with these IF matrixes. See Supplementary Information (Additional file 1) for more details. The time efficiency of structure reconstruction is shown in Fig. 3.

Testing platform is Ubuntu 16.04 (64 bit) with Intel Core i5-2400 CPU@3.10 GHz and 16GB DDR3-1600 memory. The GPU platform is Geforce GTX 1050Ti. The comparison results show that EVR running on GPU with OpenCL used the least time, followed by EVR running on CPUs with OpenCL and MOGEN; meanwhile, their computation time is not sensitive to the bin numbers within a reasonable range. The running speed of EVR with Cython is slightly slower but still relatively fast. The computation time of miniMDS and ShRec3D shows a rapid rise with the increase of bin numbers.



### EVR is robust

To assess robustness, different levels of noise were added to the IF matrix generated from the structures of toroidal spiral curves, and then the noisy data were used in the 3D structure reconstruction by using the above four software tools. Noises were added to the IF matrix by randomly selecting values from the range of  $-0.5 \times P \times IF_{\max}$  to  $0.5 \times P \times IF_{\max}$ ; here,  $P$  is noise level and  $IF_{\max}$  is the maximum value in IF matrix. The resulting structure reconstructed from the noisy data was compared to the original noise-free structure and the RMSD (root-mean-square deviation) value was calculated. See Supplementary Information (Additional file 1) for more details. At each noise level, the above steps of adding noise, reconstructing structures, and comparing structures were repeated 100 times. The average RMSD values obtained are plotted against noise level as Fig. 4.

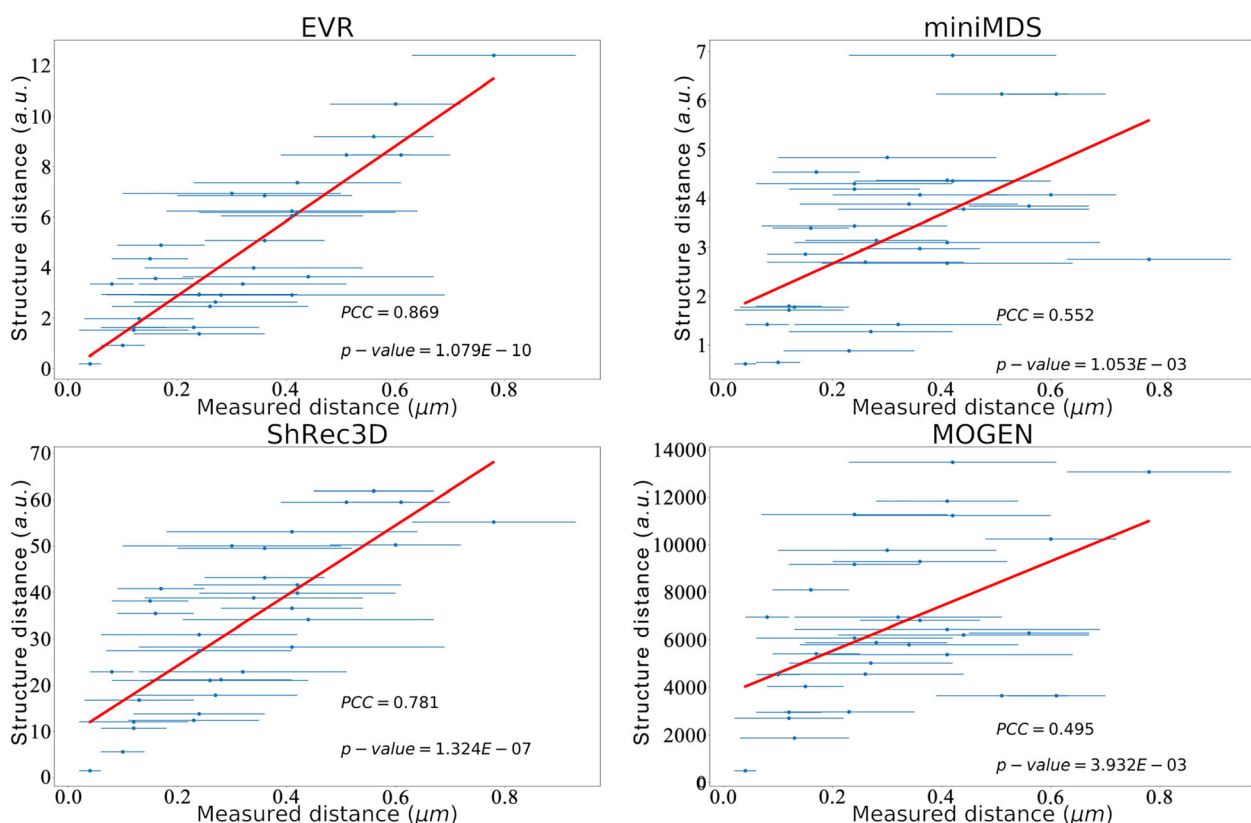
Since the IF matrix does not provide structural scale information, the scales of the structures obtained by different software tools may be not the same. When comparing structures, in order to ensure scale consistency, all the structures are scaled and compared with the original structure, and the structural alignment with the smallest RMSD value is selected for comparison. As shown in Fig. 4, our

EVR algorithm always has the smallest RMSD values at different levels of noise, meaning that the structures reconstructed based on the noisy data are always very similar to the original structures based on the noise-free data, i.e., our EVR algorithm is robust to noise.

Since the initial conformation is randomly assigned in our algorithm, the dependency of the final structure on the initial conformation was also evaluated. As shown in Supplementary Information (Additional file 1: Figure S5), our EVR algorithm is not sensitive to the random initial conformations.

### EVR is accurate

To assess accuracy, the reconstructed chromosome structures of *E. coli* by using the four software tools were compared with the published experimental fluorescence microscopy data (Additional file 1: Table S1) [36]. The *E. coli* chromosome structure reconstruction was based on the recently published IF matrixes [21]. By mapping the fluorescence marker sites onto the reconstructed structures (Additional file 1: Figure S6), the 3D coordinates of these sites could be determined and the spatial distances between them were calculated and correlated with the experimentally measured distances in fluorescence



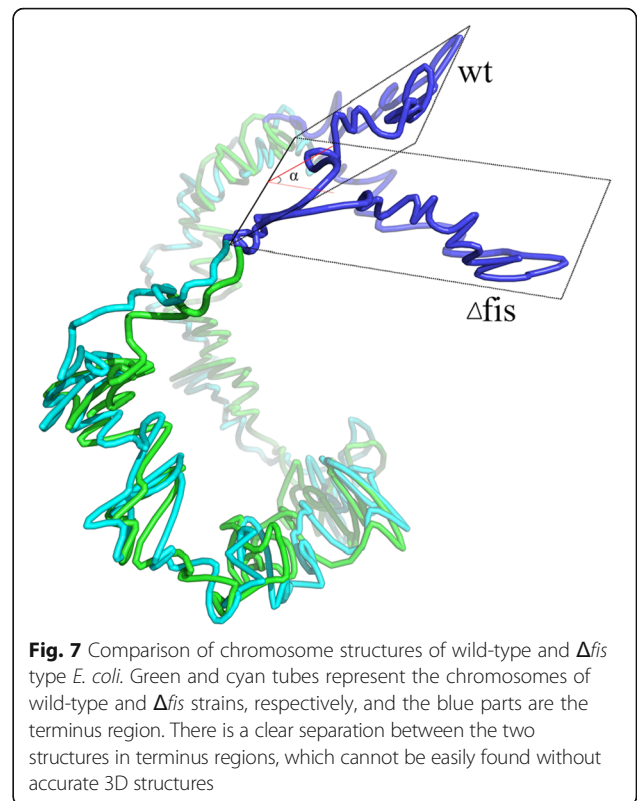
**Fig. 5** Correlation between the structure-based distance and the experimentally measured distance. The chromosome structures were reconstructed by using the four software tools based on recently published 3C data of *E. coli* [21] and the experimentally measured distances were compiled from literature [36]. See Supplementary Information (Additional file 1: Table S1) for more details

microscopy. Results show that there are high correlations (indicated by *PCC*: Pearson Correlation Coefficient) between the structure-based distances and the experimental distances (Fig. 5) for the EVR ( $PCC = 0.869$ ,  $p\text{-value} = 1.079E-10$ ) and ShRec3D ( $PCC = 0.781$ ,  $p\text{-value} = 1.324E-7$ ) software tools; The correlations for miniMDS ( $PCC = 0.552$ ,  $p\text{-value} = 1.053E-3$ ) and MOGEN ( $PCC = 0.495$ ,  $p\text{-value} = 3.932E-3$ ) are relatively lower but still significant, possibly because they are developed for eukaryotes, not suitable or optimized for prokaryotes. The significant and high correlation between the experimentally measured and structure-based distances indicate that the reconstructed structure is close to the real structure and that our EVR algorithm is accurate. More datasets from different species could be used to further test these algorithms when they become available.

**Application of EVR to prokaryotes**

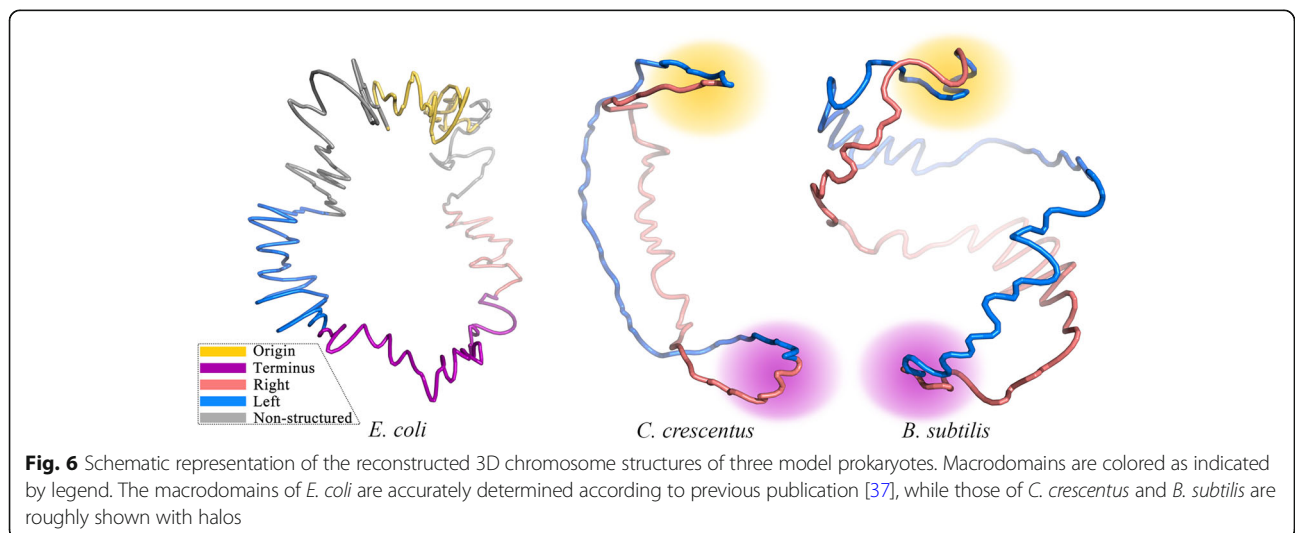
Our EVR algorithm was applied in the reconstruction of 3D chromosome structures of three prokaryotic model species *Escherichia coli* (MG1655), *Caulobacter crescentus* (CB15) and *Bacillus subtilis* (PY79) based on 3C data, and the obtained structures were divided into 4 macrodomains [1, 37]: origin and terminus of replication, left and right chromosomal arms (Fig. 6).

After transforming the interaction matrix into a spatial structure, we can visually examine the structure from a global perspective and observe features that are difficult to find in the interaction matrix. For example, Fis is a very important nucleoid-associated protein in *E. coli*, and it plays a variety of roles in regulating DNA transactions and modulating DNA topology [38]. By comparing the interaction matrixes of wild-type and *fis*-deficient ( $\Delta fis$ ) *E. coli*, it can be seen that the interaction intensity is significantly



**Fig. 7** Comparison of chromosome structures of wild-type and  $\Delta fis$  type *E. coli*. Green and cyan tubes represent the chromosomes of wild-type and  $\Delta fis$  strains, respectively, and the blue parts are the terminus region. There is a clear separation between the two structures in terminus regions, which cannot be easily found without accurate 3D structures

reduced at the terminus region but conserved at other sites [21]. While converting these two IF matrixes into structures by using our EVR program and comparing them, the overall effect of Fis protein on the 3D structure is even more obvious (Fig. 7): compared to the wild type, there is a clear angle in the terminus region of the  $\Delta fis$  type, although other regions are substantially overlapping. Running many times of our software this angle persists and it cannot be found in



**Fig. 6** Schematic representation of the reconstructed 3D chromosome structures of three model prokaryotes. Macrodomains are colored as indicated by legend. The macrodomains of *E. coli* are accurately determined according to previous publication [37], while those of *C. crescentus* and *B. subtilis* are roughly shown with halos



the interaction map. Why the lack of *fts* causes the bending of the terminus region and what biological effects this bending will cause remain to be further investigated.

## Discussion

We have developed a software tool called EVR for the reconstruction of 3D chromosome structure models from the DNA interaction data measured by 3C/Hi-C experiments. This tool added constraints related to a circular shape in the reconstruction process and thus it is particularly suitable for the reconstruction of chromosome structures of prokaryotes. Assessed based on standard structures, our EVR program is very fast and robust compared with other software tools, and the reconstructed structure is accurately evaluated based on the fluorescence labeling data. Our EVR program was used in the reconstruction of the 3D chromosome structures of three model prokaryotic species with published DNA interaction data. The obtained structures provided consistent understanding on the 3D genomes of these species.

Limited by the resolution of 3C/Hi-C experiment data [39, 40], our EVR program as well as other programs can only establish the backbone structure models of the prokaryotic chromosomes at domain levels. Combined with other modeling approaches such as the thermodynamics-based methods used in a recent paper [25], more reliable and higher resolution chromosome models may be built to assist the fundamental research on the 3D genomics of prokaryotes. Another noticeable aspect is about the storage file format of the reconstructed chromosome structures. Currently, since the resulting structures contain only the spatial coordinate information of bins, they are usually saved in a simple PDB-file-like format such as the modified PDB file format [30] or “xyz” format [18]. With the development of 3D genomics, more specific format may be explored to accommodate more information such as domain partition, gene annotation and levels of transcription etc. in order to be readily integrated into a systems biology context.

## Conclusions

Our EVR software is optimized for closed-loop structures and can be used to reconstruct the bacterial 3D chromosome models based on the contact frequency matrix derived from 3C/Hi-C experimental data with high accuracy and robustness. Owing to its parallel implementation, this software tool also runs fast. The application of EVR to the model organism *E. coli* revealed a large structural variation between mutant and wild types, which provides clues for experimental investigation. In summary, EVR is a readily available software

tool for the studies of prokaryotic 3D genomes and related fields.

## Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12864-019-6096-0>.

**Additional file 1.** Supplementary materials for the evaluation of EVR algorithm.

## Abbreviations

1NTB: Nucleotide per bead; 3C: Chromosome Conformation Capture; 4C: Circular Chromosome Conformation Capture; 5C: 3C-Carbon copy; ChIP-chip: Chromatin immunoprecipitation on chip; CID: Chromosomal Interaction Domains; EVR: Error-Vector Resultant; FISH: Fluorescence in situ Hybridization; GEO: Gene Expression Omnibus; Hi-C: High-throughput/resolution Chromosome conformation capture; IF matrix: Interaction Frequency matrix; IMP: Integrated Modeling Platform; NAPs: Nucleoid Associated Proteins; PCC: Pearson Correlation Coefficient; RMSD: Root-Mean-Square Deviation; SMC: Structural Maintenance of Chromosomes

## Acknowledgements

We'd like to thank Yang Yi for his help in polishing the manuscript.

## Authors' contributions

BGM conceived and designed the study. KJH designed the algorithm and software supervised by BGM. KJH and BGM wrote the manuscript. All authors have read and approved the final manuscript.

## Funding

This work was supported by the National Natural Science Foundation of China (Grant 31570844), Project 2662016PY094 supported by the Fundamental Research Funds for the Central Universities, and the National Basic Research Program of China (973 project, Grant 2013CB127103). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

## Availability of data and materials

The source code and data of the program are available at <https://github.com/mbglab/EVR>.

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

Received: 25 March 2019 Accepted: 11 September 2019

Published online: 15 October 2019

## References

- Weng X, Xiao J. Spatial organization of transcription in bacterial cells. *Trends Genet.* 2014;30(7):287–97.
- Slager J, Veening JW. Hard-wired control of bacterial processes by chromosomal gene location. *Trends Microbiol.* 2016;24(10):788–800.
- Wang X, Rudner DZ. Spatial organization of bacterial chromosomes. *Curr Opin Microbiol.* 2014;22:66–72.
- Langer-Safer PR, Levine M, Ward DC. Immunological method for mapping genes on *Drosophila* polytene chromosomes. *Proc Natl Acad Sci U S A.* 1982;79(14):4381–5.
- Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science.* 2002;295(5558):1306–11.
- Bayani J, Squire JA. Fluorescence in situ Hybridization (FISH). *Curr Protoc Cell Biol.* 2004; Chapter 22:22.4.1–22.4.52.
- Zhao Z, Tavossidana G, Sjolinder M, Gondor A, Mariano P, Wang S, Kanduri C, Lezczano M, Sandhu KS, Singh U, et al. Circular chromosome conformation

- capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet.* 2006;38(11):1341–7.
8. Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, Honan TA, Rubio ED, Krumm A, Lamb J, Nusbaum C, et al. Chromosome conformation capture carbon copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* 2006;16(10):1299–309.
  9. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science.* 2009;326(5950):289–93.
  10. Lajoie BR, Dekker J, Kaplan N. The Hitchhiker's guide to hi-C analysis: practical guidelines. *Methods.* 2015;72:65–75.
  11. Zhang Z, Li G, Toh KC, Sung WK. 3D chromosome modeling with semi-definite programming and hi-C data. *J Comput Biol.* 2013;20(11):831–46.
  12. Umbarger MA, Toro E, Wright MA, Porreca GJ, Bau D, Hong SH, Fero MJ, Zhu LJ, Marti-Renom MA, McAdams HH, et al. The three-dimensional architecture of a bacterial genome and its alteration by genetic perturbation. *Mol Cell.* 2011;44(2):252–64.
  13. Russel D, Lasker K, Webb B, Velazquez-Muriel J, Tjioe E, Schneidman-Duhovny D, Peterson B, Sali A. Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS Biol.* 2012;10(11):e1001244.
  14. Le Dily F, Serra F, Marti-Renom MA. 3D modeling of chromatin structure: is there a way to integrate and reconcile single cell and population experimental data? *Wires Comput Mol Sci.* 2017;7(5):e1308.
  15. Cagliero C, Grand RS, Jones MB, Jin DJ, O'Sullivan JM. Genome conformation capture reveals that the *Escherichia coli* chromosome is organized by replication and transcription. *Nucleic Acids Res.* 2013;41(12):6058–71.
  16. Le TB, Imakaev MV, Mirny LA, Laub MT. High-resolution mapping of the spatial organization of a bacterial chromosome. *Science.* 2013;342(6159):731–4.
  17. Marbouty M, Le Gall A, Cattoni DI, Cournac A, Koh A, Fiche JB, Mozziconacci J, Murray H, Koszul R, Nollmann M. Condensin- and replication-mediated bacterial chromosome folding and origin condensation revealed by hi-C and super-resolution imaging. *Mol Cell.* 2015;59(4):588–602.
  18. Lesne A, Riposo J, Roger P, Cournac A, Mozziconacci J. 3D genome reconstruction from chromosomal contacts. *Nat Methods.* 2014;11(11):1141–3.
  19. Wang X, Le TB, Lajoie BR, Dekker J, Laub MT, Rudner DZ. Condensin promotes the juxtaposition of DNA flanking its loading site in *Bacillus subtilis*. *Genes Dev.* 2015;29(15):1661–75.
  20. Wang X, Brandao HB, Le TB, Laub MT, Rudner DZ. *Bacillus subtilis* SMC complexes juxtapose chromosome arms as they travel from origin to terminus. *Science.* 2017;355(6324):524–7.
  21. Liou VS, Cournac A, Marbouty M, Duigou S, Mozziconacci J, Espeli O, Boccard F, Koszul R. Multiscale structuring of the *E. coli* chromosome by nucleoid-associated and Condensin proteins. *Cell.* 2018;172(4):771–783 e718.
  22. Dorman CJ. Genome architecture and global gene regulation in bacteria: making progress towards a unified model? *Nat Rev Microbiol.* 2013;11(5):349–55.
  23. Boles TC, White JH, Cozzarelli NR. Structure of plectonemically supercoiled DNA. *J Mol Biol.* 1990;213(4):931–51.
  24. Grainger DC, Hurd D, Goldberg MD, Busby SJ. Association of nucleoid proteins with coding and non-coding segments of the *Escherichia coli* genome. *Nucleic Acids Res.* 2006;34(16):4642–52.
  25. Hacker WC, Li S, Elcock AH. Features of genomic organization in a nucleotide-resolution molecular model of the *Escherichia coli* chromosome. *Nucleic Acids Res.* 2017;45(13):7541–54.
  26. Goodsell DS, Autin L, Olson AJ. Lattice models of bacterial nucleoids. *J Phys Chem B.* 2018;122(21):5441–7.
  27. Johnson GT, Autin L, Al-Alusi M, Goodsell DS, Sanner MF, Olson AJ. cellPACK: a virtual mesoscope to model and visualize structural systems biology. *Nat Methods.* 2015;12(1):85–91.
  28. Yildirim A, Feig M. High-resolution 3D models of *Caulobacter crescentus* chromosome reveal genome structural variability and organization. *Nucleic Acids Res.* 2018;46(8):3937–52.
  29. Serra F, Di Stefano M, Spill YG, Cuartero Y, Goodstadt M, Bau D, Marti-Renom MA. Restraint-based three-dimensional modeling of genomes and genomic domains. *FEBS Lett.* 2015;589(20 Pt A):2987–95.
  30. Trieu T, Cheng J. MOGEN: a tool for reconstructing 3D models of genomes from chromosomal conformation capturing data. *Bioinformatics.* 2016;32(9):1286–92.
  31. Serra F, Bau D, Goodstadt M, Castillo D, Filion GJ, Marti-Renom MA. Automatic analysis and 3D-modelling of hi-C data using TADbit reveals structural features of the fly chromatin colors. *PLoS Comput Biol.* 2017;13(7):e1005665.
  32. Rieber L, Mahony S. miniMDS: 3D structural inference from high-resolution hi-C data. *Bioinformatics.* 2017;33(14):i261–6.
  33. Servant N, Varoquaux N, Lajoie BR, Viara E, Chen CJ, Vert JP, Heard E, Dekker J, Barillot E. HiC-pro: an optimized and flexible pipeline for hi-C data processing. *Genome Biol.* 2015;16:259.
  34. Wingett S, Ewels P, Furlan-Magaril M, Nagano T, Schoenfelder S, Fraser P, Andrews S. HiCUP: pipeline for mapping and processing hi-C data. *F1000Res.* 2015;4:1310.
  35. Yaffe E, Tanay A. Probabilistic modeling of hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. *Nat Genet.* 2011;43(11):1059–65.
  36. Espeli O, Mercier R, Boccard F. DNA dynamics vary according to macrodomain topography in the *E. coli* chromosome. *Mol Microbiol.* 2008; 68(6):1418–27.
  37. Niki H, Yamaichi Y, Hiraga S. Dynamic organization of chromosomal DNA in *Escherichia coli*. *Genes Dev.* 2000;14(2):212–23.
  38. Cho BK, Knight EM, Barrett CL, Palsson BO. Genome-wide analysis of Fis binding in *Escherichia coli* indicates a causative role for a–/AT-tracts. *Genome Res.* 2008;18(6):900–10.
  39. Liu M, Darling A. Metagenomic chromosome conformation capture (3C): techniques, applications, and challenges. *F1000Res.* 2015;4:1377.
  40. Bonev B, Mendelson Cohen N, Szabo Q, Fritsch L, Papadopoulos GL, Lubling Y, Xu X, Lv X, Hugnot JP, Tanay A, et al. Multiscale 3D genome rewiring during mouse neural development. *Cell.* 2017;171(3):557–572 e524.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

