Supplementary Information

Benchmark of software tools for prokaryotic chromosomal interaction domain identification

Magnitov et al.

- Supplementary Figures S1-12
- Supplementary Tables S1-4
Supplementary Figure S1. Comparison of restriction fragment sizes and the quality of experimental contact maps. (A) Boxplot of the HindIII and HpaII restriction fragment sizes (in base pairs) and contact maps at 5 and 10 kb obtained using the respective restriction enzymes for the *B. subtilis* genome. (B) Boxplot of the BglII and NcoI restriction fragment sizes (in base pairs) and contact maps at 5 and 10 kb obtained using the respective restriction enzymes for the *C. crescentus* genome.
Supplementary Figure S2. (A) Normalized experimental (top row) and simulated (bottom row) Hi-C contact maps for *C. crescentus* generated with NcoI and BglII restriction enzymes at 10 and 5 kb resolutions. (B) The fraction of zeroed cells in experimental and simulated contact maps at different resolutions. (C) Venn diagrams showing the overlap in positions of zero-count columns and rows for experimental and simulated contact maps at 5 kb resolution.
Boxplots of the estimated restriction fragment sizes (in base pairs) and simulated contact maps for selected bacterial genomes with varying GC content. Generally, maps simulated with restriction enzymes with theoretically shorter restriction fragments have higher quality. When comparing two 4-cutter enzymes, MseI (TTAA recognition site) yields a better map when the genome has a lower GC content (third row), while HpaII (CCGG) exhibits the opposite trend. Similar results are observed for the HindIII (AAGCTT) and NcoI (CCATGG) restriction enzymes, which are 6-cutters that vary by the GC content of their recognition sites.
Supplementary Figure S4. Examples of the CIDs annotated for *E. coli* replicates at the 5 kb resolution, including the domains annotated in replicate 1 (in green) and replicate 2 (in blue).
Supplementary Figure S5. (A) Fraction of CID boundaries identified by each domain caller that were also identified by 0 (blue), 1 to 4 (green), 5 to 8 (orange) and 9 or more (red) other tools. (B) Boxplots of the total genome fraction covered by CIDs for all domain callers. Segmentation for each biological replicate used in the benchmark is represented by a dot (N = 16). Tools are ordered by median Jaccard Index value, from highest to lowest.
Supplementary Figure S6. Running time of the 29 TAD callers calculated for 5 and 10 kb contact matrices from *E. coli* pseudo-replicates. Tools are ordered by increasing average run time, from highest to lowest (ordered by increasing average value).
Supplementary Figure S7. Swarm plots for the Measure of Concordance values between the CID segmentations annotated for *E. coli* pseudo-replicate contact maps with varying coverage (1, 3, 5 and 10 millions of contacts) by each domain caller. Each dot represents a comparison between CID annotations obtained for two maps with different numbers of contacts. The black line represents the median MoC value for each tool.
Supplementary Figure S8. Venn diagrams showing the number of intersections between the CID boundaries annotated by each domain caller for *E. coli* pseudo-replicates with varying coverage (1, 3, 5 and 10 millions of contacts).
Supplementary Figure S9. Swarm plots for the Measure of Concordance values between the CID segmentations annotated for *E. coli* pseudo-replicate contact maps with different resolutions (3, 5, 10 and 15 kb) by each domain caller. Each dot represents a comparison between CID annotations obtained for two maps with different resolutions. The black line represents the median MoC value for each tool.
Supplementary Figure S10. Venn diagrams showing the percentage of overlap between the "extended" CID boundaries annotated by each domain caller for *E. coli* pseudo-replicates with different resolutions (3, 5, 10 and 15 kb).
Supplementary Figure S11. The association between the annotated domain boundaries and gene expression along the genome. From top to bottom: (i) normalized Hi-C contact map, (ii) a pile up of the domain boundaries predicted by each domain calling tool for both replicates, (iii) a consensus of these boundaries and (iv) gene expression track for (A) *C. crescentus*, (B) *B. subtilis* and (C) *E. coli*. 
Supplementary Figure S12. Gene expression levels at the CID boundaries annotated by lavaburst.modularity, HiCseg, Insulation Score and Directionality Index in *B. subtilis*, *C. crescentus* and *E. coli* datasets. Lines show the mean normalized RNA-seq expression levels in the vicinity of CID boundaries (the profile around boundaries annotated in replicate 1 are in blue, and those annotated in replicate 2 are in red) with a shaded area indicating the range of one standard deviation from the mean expression. The p-values for significance of the expression difference at the boundaries compared to the regions within CIDs are given in the subplots’ titles.
## Supplementary Table S1

Accession numbers of the publicly available Hi-C, 3C-seq and RNA-seq datasets (in Sequence Read Archive or European Nucleotide Archive) and reference genomes (in Genbank) used in the study.

| Organism                     | Hi-C datasets          | RNA-seq datasets       | Restriction enzyme | Genome          | Resolution | Used for domain-calling tools benchmark |
|------------------------------|------------------------|------------------------|--------------------|-----------------|------------|---------------------------------------|
| Bacillus subtilis            | rep1 - SRR2002580      | SRR922367              | HindIII            | NC_022898.1     | 10 kb      | Yes                                   |
|                              | rep2 - SRR2002581      |                        | HpaII              |                 |            |                                       |
|                              | rep1 - SRR2182094      |                        |                    |                 |            |                                       |
| Caulobacter crescentus       | rep1 - SRR824843       | SRR2818123             | BglII              | NC_011916.1     | 10 kb      | Yes                                   |
|                              | rep2 - SRR824844       |                        |                    |                 |            |                                       |
|                              | rep1 - SRR824846       |                        | NcoI               |                 |            | No                                    |
| Escherichia coli             | rep1 - SRR6354545,     | SRR2932231             | HpaII              | NC_000913.3     | 5 kb       | Yes                                   |
|                              | SRR6940991, SRR6490992, SRR6843204, SRR6843203 | |                    |                 |            |                                       |
|                              | rep2 - SRR6354546      |                        |                    |                 |            |                                       |
| Mycoplasma pneumoniae        | rep1 - ERR1413600      |                        |                    |                 |            |                                       |
|                              | rep2 - ERR1413595      |                        |                    |                 |            |                                       |
|                              | rep3 - ERR1413596      |                        |                    |                 |            |                                       |
|                              | rep4 - ERR1413597      |                        |                    |                 |            |                                       |
|                              | rep5 - ERR1413598      |                        |                    |                 |            |                                       |
|                              | rep1 - ERR1413593      |                        |                    |                 |            |                                       |
|                              | rep2 - ERR1413594      |                        |                    |                 |            |                                       |
| Sulfolobus acidocaldarius    | rep1 - SRR8699887      |                        | HindIII            | CP000077.1      | 15 kb      | Yes                                   |
|                              | rep2 - SRR869988     |                        |                    |                 |            |                                       |
|                              | rep3 - SRR8699889      |                        |                    |                 |            |                                       |
Supplementary Table S2

The description of the parameters for the domain calling procedure.

| Domain caller | Parameters selection | Reference |
|---------------|----------------------|-----------|
| Armatus       | We used Armatus v. 2.1 and varied the gamma parameter from 0.01 to 5.00 with a step of 0.01 to find the optimal CIDs segmentation of a genome. | [1] |
| Arrowhead     | The Juicer Tools v. 1.13.09 arrowhead module with default parameters was used to determine CIDs. The nested domains were then discarded to remove the hierarchy. | [2] |
| CaTCH         | From the CaTCH package we used the domain.call function with no parameters to annotate CIDs. | [3] |
| CHDF          | We executed CHDF with parameters Length, Number and Size equal to the dimensions of the corresponding contact map. | [4] |
| chromoR       | The function segmentCIM from the chromoR was used on the raw contact maps to find the domains. | [5] |
| ChromOSignt   | The ChromOSignt package was used in the detect mode with parameter pattern = "borders". | [6] |
| ClusterTAD    | We used Java implementation of ClusterTAD with the parameters window set to 10 and minimum CID size set to 4 bins. The search for optimal CID boundaries was performed among CID sets identified with different K values used for clustering. | [7] |
| deDoc         | We used deDoc(M) algorithm implementation with default parameters to identify CIDs. | [8] |
| Directionality_Index | We used the Directionality Index approach based on paired t-test. For the matrices with resolution higher than 10 kb, the number of bins considered for t-statistics calculation was set to 20, while for the lower resolution matrices it was set to 10. We then manually annotated CID boundaries using the directionality index profile and t-statistics threshold at significance level 0.1. | [9] |
| EAST          | The EAST script was executed with parameters maxW = 50 and minL = 4. | [10] |
| GMAP          | GMAP was executed on the raw contact maps with parameters dom_order set to 1, min_dp set to 4, max_dp set to 100, min_d set to 10 and icfr set to 0.75. | [11] |
| HiCEexplorer  | To identify CIDs with HiCEexplorer, we used its hicFindTADs module with minDepth parameter ranging from 3 to 20 matrix resolutions and delta from 0.005 to 0.1 with a step of 0.005. A correction for multiple testing was performed using the FDR method. | [12] |
| HiCseg        | The HiCseg was executed with parameters distrib = "G" and model = "D". We varied the nb_change_max parameter from 2 to 100 to find the optimal domains. | [13] |
| IC-Finder     | In order to identify CIDs, IC-Finder was executed with default parameters. | [14] |
| Insulation_Score | We used the original Perl implementation of the Insulation Score algorithm and varied the IS parameter from 5 to 15 resolution sizes with a step of 1 and the iDS parameter from 2 to 10 resolution sizes with a step of 2 in order to find the optimal segmentation of CIDs. | [15] |
| lava.armatus  | The Lavaburst package (https://github.com/nvictus/lavaburst) includes four domain calling algorithms: armatus, corner, modularity and variance. For each algorithm, we varied the gamma parameter in the range from 0.01 to 5.00 with a step of 0.01 to find the optimal CIDs. | [16] |
| lava.corner   | We used MrTADFinder v. 1.2 and varied the res parameter from 0.5 to 3.5 with a step of 0.1 to find the optimal segmentation of CIDs. | [17] |
| lava.modularity | We used MrTADFinder v. 1.2 and varied the res parameter from 0.5 to 3.5 with a step of 0.1 to find the optimal segmentation of CIDs. | [18] |
| lava.variance | We used MrTADFinder v. 1.2 and varied the res parameter from 0.5 to 3.5 with a step of 0.1 to find the optimal segmentation of CIDs. | [19] |
| MrTADFinder   | We used OnTAD with parameter minsz set to 4. For C.crescentus, B.subtilis and E.coli we then merged the annotated domains at levels 2, 3 and 4 into one set. | [20] |
| OnTAD         | We used OnTAD with parameter minsz set to 4. For C.crescentus, B.subtilis and E.coli we then merged the annotated domains at levels 2, 3 and 4 into one set. | [21] |
| spectral      | For spectral algorithm we varied the lambda threshold parameter from 0.01 to 1.00 with a step of 0.01 to find the optimal CIDs annotation. | [22] |
| SpectralTAD   | SpectralTAD was executed using a raw contact map as input data and a minimum CID size of 4 bins. | [23] |
| TADbit        | We used a raw contact map as input for TADbit and set the resolution parameter to correspond to each matrix. | [24] |
| TADpole       | We used TADpole with parameter resol corresponding to each matrix size. | |
### Supplementary Table S3

Optimal run parameters of the domain callers for each dataset as revealed by the Jaccard Index maximization analysis.

| Domain caller | C. crescentus      | B. subtilis       | E. coli          | M. pneumoniae HindIII | M. pneumoniae HpaII | S. acidocaldarius |
|---------------|--------------------|-------------------|------------------|-----------------------|---------------------|------------------|
| Armatus       | gamma = 0.16       | gamma = 0.01      | gamma = 0.07     | gamma = 0.09          | gamma = 0.12        | gamma = 0.01      |
| Arrowhead     |                    |                   |                  |                       |                     |                  |
| CaTCH         |                    |                   |                  |                       |                     |                  |
| CHDF          |                    |                   |                  |                       |                     |                  |
| chromoR       |                    |                   |                  |                       |                     |                  |
| Chromosight   |                    |                   |                  |                       |                     |                  |
| ClusterTAD    | K = 22             | K = 19            | K = 31           | K = 5                 | K = 3               | K = 16           |
| deDoc         |                    |                   |                  |                       |                     |                  |
| Directionality_index |            |                   |                  |                       |                     |                  |
| EAST          |                    |                   |                  |                       |                     |                  |
| GMAP          |                    |                   |                  |                       |                     |                  |
| HiCEXplorer   | window_size = 110000<br>cutoff = 0.03 | window_size = 70000<br>cutoff = 0.005 | window_size = 25000<br>cutoff = 0.005 | window_size = 45000<br>cutoff = 0.01 | - | window_size = 90000<br>cutoff = 0.01 |
| HiCseg        | Kmax = 14          | Kmax = 75         | Kmax = 79        | Kmax = 19             | Kmax = 21           | Kmax = 16        |
| IC-Finder     |                    |                   |                  |                       |                     |                  |
| Insulation_Score | IS = 130000     | IS = 110000       | IS = 75000       | IS = 15000             | IS = 18000           | IS = 90000       |
|              | IDS = 20000        | IDS = 60000       | IDS = 10000      | IDS = 30000            | IDS = 12000          | IDS = 60000      |
| lava.armatus  | gamma = 0.66       | gamma = 1.12      | gamma = 0.78     | gamma = 1.33           | gamma = 1.13         | gamma = 0.08     |
|              |                   |                   |                  |                       |                     |                  |
| lava.corner   | gamma = 0.81       | gamma = 0.95      | gamma = 1.15     | gamma = 1.01           | gamma = 0.99         | gamma = 0.65     |
| lava.modularity | gamma = 2.81     | gamma = 2.16      | gamma = 2.36     | gamma = 2.28           | gamma = 1.62         | gamma = 1.43     |
| lava.variance | gamma = 1.36       | gamma = 0.51      | gamma = 0.87     | gamma = 0.58           | gamma = 0.61         | gamma = 0.73     |
| MTADFinder    | res = 1.2          | res = 1.0         | res = 0.8        | res = 1.1              | res = 1.0            | res = 1.0        |
| OnTAD         |                    |                   |                  |                       |                     |                  |
| spectral      | lambda = 0.64      | lambda = 0.74     | lambda = 0.49    | lambda = 0.87          | lambda = 0.69        | lambda = 0.58    |
| SpectralTAD   |                    |                   |                  |                       |                     |                  |
| TADbit        |                    |                   |                  |                       |                     |                  |
| TADpole       | level = 11         | level = 5         | level = 6        | level = 15             | level = 1            | level = 5       |
| TADtool.directionality | window_size = 20000<br>cutoff = 0.0095 | window_size = 40000<br>cutoff = 0.6 | window_size = 30000<br>cutoff = 0.0 | window_size = 60000<br>cutoff = 0.0 | window_size = 51000<br>cutoff = 0.0 | window_size = 150000<br>cutoff = 0.0 |
| TADtool.insulation | window_size = 70000<br>cutoff = 0.0148 | window_size = 20000<br>cutoff = 0.0097 | window_size = 75000<br>cutoff = 0.0066 | window_size = 12000<br>cutoff = 0.0047 | window_size = 60000<br>cutoff = 0.0046 | window_size = 350000<br>cutoff = 0.0183 |
| TADtree       |                    |                   |                  |                       |                     |                  |
| TopDom        | window_size = 7     | window_size = 8    | window_size = 14  | window_size = 6        | window_size = 5      | window_size = 8   |
## Supplementary Table S4

The comparative analysis of the results from this work and previously published benchmarks. N/A means the tool was not analysed in the corresponding study.

| Domain caller     | Reproducibility between replicates | Robustness to resolution and/or coverage | Concordance with other domain calling tools | Enrichment of biological features at the domain boundaries |
|-------------------|-----------------------------------|------------------------------------------|--------------------------------------------|----------------------------------------------------------|
|                   | Magnitov et al. | Forcato et al. | Magnitov et al. | Dali and Blanchett | Zufferey et al. | Magnitov et al. | Dali and Blanchett | Zufferey et al. | Magnitov et al. | Dali and Blanchett | Forcato et al. | Zufferey et al. |
| Armatus           | no                  | no                     | yes                          | yes                           | yes                          | yes                          | yes                          | yes                          | yes                          | yes                          | no                        |
| Arrowhead         | no                  | no                     | yes                          | yes                           | no                          | yes                          | yes                          | yes                          | yes                          | yes                          | yes                        |
| CaTCH             | yes                 | N/A                    | no                           | N/A                           | yes                          | no                          | N/A                           | yes                          | N/A                           | N/A                           | N/A                       |
| CHDF              | N/A                 | N/A                    | yes                          | N/A                           | yes                          | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | yes                       |
| chromoR           | N/A                 | N/A                    | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                       |
| Chromosight       | N/A                 | yes                    | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                       |
| ClusterTAD        | N/A                 | N/A                    | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | yes                       |
| deDoc             | no                  | N/A                    | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                       |
| Directionality Index | yes                | no                     | yes                           | yes                           | yes                          | yes                          | yes                           | yes                           | yes                           | yes                           | yes                       |
| EAST              | no                  | N/A                    | yes                           | N/A                           | no                           | yes                          | N/A                           | N/A                           | N/A                           | N/A                           | yes                       |
| GMAP              | no                  | N/A                    | yes                           | N/A                           | no                           | N/A                           | yes                           | N/A                           | N/A                           | N/A                           | yes                       |
| HiExplorer        | N/A                 | N/A                    | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | yes                       |
| HiCseg            | yes                 | yes                    | yes                           | yes                           | yes                          | yes                           | no                            | yes                           | yes                           | yes                           | yes                       |
| IC-Finder         | no                  | N/A                    | yes                           | N/A                           | yes                          | N/A                           | N/A                           | yes                           | N/A                           | N/A                           | yes                       |
| Insulation Score  | yes                 | yes                    | N/A                           | yes                           | N/A                           | no                            | yes                           | N/A                           | N/A                           | N/A                           | yes                       |
| lava.armatus      | N/A                 | yes                    | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                       |
| lava.corner       | N/A                 | yes                    | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                       |
| lava.modularity   | yes                 | yes                    | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                       |
| lava.variance     | N/A                 | yes                    | N/A                           | N/A                           | yes                          | no                            | N/A                           | N/A                           | N/A                           | N/A                           | N/A                       |
| MrTADFinder       | N/A                 | no                     | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                       |
| OnTAD             | N/A                 | N/A                    | N/A                           | N/A                           | yes                          | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                       |
| spectral          | no                  | N/A                    | no                            | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                       |
| SpectralTAD       | N/A                 | yes                    | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                       |
| TADbit            | yes                 | yes                    | yes                           | no                            | yes                          | no                            | yes                           | yes                           | yes                           | yes                           | yes                       |
| TADpole           | no                  | N/A                    | N/A                           | N/A                           | yes                          | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                       |
| TADtool.directionality | yes               | N/A                    | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                       |
| TADtool.insulation| N/A                 | N/A                    | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                       |
| TADtree           | no                  | yes                    | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | N/A                           | yes                       |
| TopDom            | N/A                 | yes                    | yes                           | yes                           | yes                           | yes                           | yes                           | yes                           | yes                           | yes                           | N/A                       |

**Supplementary Table S4.**
