NuST: analysis of the interplay between nucleoid organization and gene expression

Vittore F. Scolari1,2,3,*, Mina Zarei1,2,4, Matteo Osella1,2 and Marco Cosentino Lagomarsino1,2,5

1Genomic Physics Group, UMR 7238 CNRS Génomique des Microorganismes, 2Université Pierre et Marie Curie, 15 rue de L’École de Médecine, 75006, Paris, France, 3NCBS, Bangalore, India, 4Dip. Fisica, Università di Milano, Milano, Italy and 5Dip. Fisica, Università di Torino, Torino, Italy

ABSTRACT
Summary: Different experimental results suggest the presence of an interplay between global transcriptional regulation and chromosome spatial organization in bacteria. The identification and clear visualization of spatial clusters of contiguous genes targeted by specific DNA-binding proteins or sensitive to nucleoid perturbations can elucidate links between nucleoid structure and gene expression patterns. Similarly, statistical analysis to assess correlations between results from independent experiments can provide the integrated analysis needed in this line of research. NuST (Nucleoid Survey tools), based on the Escherichia coli genome, gives the non-expert access to the web server can be downloaded at http://www.lgm.upmc.fr/nust/downloads/, in order to allow the users to use NuST in systematic bioinformatic analyses.

Availability and Implementation: NuST is a web server (available at http://www.lgm.upmc.fr/nust). The website is implemented in PHP, SQLite and Ajax, with all major browsers supported, while the core algorithms are optimized and implemented in C. NuST has an extensive help page and provides a direct visualization of results as well as different downloadable file formats. A template Perl code for automated access to the web server can be downloaded at http://www.lgm.upmc.fr/nust/downloads/, in order to allow the users to use NuST in systematic bioinformatic analyses.

Contact: vittore.scolari@upmc.fr

Supplementary information: Supplementary data are available at Bioinformatics online.

Received on February 22, 2012; revised on April 12, 2012; accepted on April 16, 2012

The orchestration of coordinated global changes in the transcriptional program is at the basis of bacterial adaptation to environments and stresses. These changes depend on a large regulatory network mediated by specific binding of transcription factors, as well as on the physical organization of the chromosome, which affects the expression of large gene sets. The bacterial DNA is condensed in a compact DNA–protein complex called ‘nucleoid’, whose transcriptional activity depends on nucleoid-related factors, including the degree of supercoiling and the specific and non-specific binding of nucleoid associated proteins (NAPs) such as Fix, H-NS and HU. While the total level of supercoiling is controlled by enzymes like gyrase and topoisomerases (Travers and Muskhelishvili, 2005), NAPs like H-NS and Fix are believed to stabilize locally DNA loops (Luijsterburg et al., 2006). The nucleoid structure varies at different scales, from DNA supercoiled loops ~10 kb long (Postow et al., 2004) to large compartments organizing the genome in four macromdomains (Dame et al., 2011). The effects of transcriptional network and nucleoid on large-scale gene expression are coupled. For example, many NAPs are also specific transcription factors, and affect the expression of targets both directly and through the conformational changes that they can induce on the chromosome.

The complex interplay between chromosome organization and gene expression requires integration of data from different high- and low-throughput experiments with statistical analysis at multiple scales. The web server described here is an effort to fill this gap. Since many of the nucleoid structural features (e.g. supercoil domains and macromdomains) involve contiguous genomic regions, a typical hallmark of nucleoid-mediated regulation is aggregation along the chromosome of genes having specific properties. An extensive step-by-step documentation introduces analyses that can be performed using the web server and is divided into an introductory help page, and a ‘learn by example’ page guiding the user through the interpretation of the results and the choice of the parameters.

The input datasets are single column text files with one gene ID for each row. Standard gene IDs are Regulon DB database (Gama-Castro et al., 2011). For different gene IDs the server proposes synonyms. Sample datasets as well as the complete list of accepted synonyms can be obtained from the ‘Download’ page. Loaded data sets are stored in the ‘Personal’ part of the internal database and can be accessed for further analysis. They are deleted at the end of each anonymous session. A login (obtained sending an email to the administrator) allows to keep personal data for multiple sessions. The ‘Common’ part of the database contains datasets from the literature, organized by type of data and experimental technique.

*To whom correspondence should be addressed.
We would like to thank B. Sclavi, P. Cicuta, M. Cereda, M. Babu for useful feedback.

**ACKNOWLEDGEMENTS**

Funding: This work was supported by the International Human Frontier Science Program Organization, grant [RGY0606/2009-C].

Conflict of Interest: none declared.

REFERENCES


