ORIS: An interactive software tool for prediction of replication origin in prokaryotic genomes

Urminder Singh[1], **Kushal Shah**[2], **Suman Dhar**[3], **Vinod Kumar Singh**[1], **and Annangarachari Krishnamachari**[1]

1 School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi, India
2 Department of Electrical Engineering and Computer Science, Indian Institute of Science Education and Research (IISER), Bhopal - 462066, Madhya Pradesh, India **3** Special centre for Molecular Medicine, Jawaharlal Nehru University, New Delhi, India

Summary

A remarkable yet very precise event that takes place within the cell, is called DNA replication or chromosome copy (Alberts et al., 2013). The way it is done in eukaryotes is vastly different from prokaryotes yet something is common as well. The origin of replication is a small segment in the genome at which the DNA replication process is initiated. Bacterial genomes generally have a single origin of DNA replication whereas archaea and eukaryotes have multiple replication sites (Egan, Fogel, & Waldor, 2005; O'Donnell, Langston, & Stillman, 2013).

Experimental studies carried out on the model organism *S. cerevisiae* brought greater insights in understanding the complex replication process. A typical origin sequence in *S. cerevisiae* has a DNA segment called Autonomous Replication Sequence (ARS) and this further has an essential A element and multiple B elements and all these are required for proper origin function or firing. This kind of organizational structure may be present in other genomes but is not very obvious. Uncovering such unknown structures in the genome can improve our understanding of DNA replication and the origin of replication sites.

Hence, software tools should be developed incorporating many such useful contextual information for sequence analysis, with the hope that biologists will benefit and it may provide them a helping hand to search for origin like sequences in any genome of their choice.

There are a few online tools/databases which may help in finding the origin of replication sites for a few species (Frank & Lobry, 2000; Gao & Zhang, 2008). These tools are limited by the type of computational approaches they take and are not interactive.

Here, we present a standalone software tool, ORIS, for interactive analysis of prokaryotic genomes to find the origin of replication sites in the genome. An effort is made to develop a software suite that does context-based analysis which includes DSP (digital signal processing) based correlation measures, all types of skew measures and entropy-based methods for predicting the origin of replication (Beauchamp & Yuen, 1979; Shah & Krishnamachari, 2012). Also, features such as pattern matching, generating sequence logo, finding entropy distribution along the length of the genome, DNA bending profiles, and composition statistics are available as part of the analysis.

ORIS is a novel tool that lets the user interactively explore the whole genome sequence data using several computational methods and charts. ORIS is well suited for origin finding across bacteria, archaea and to some extent eukaryotes. ORIS was used for hypothesizing putative origin of replication sites in *Plasmodium falciparum*, which were later experimentally verified (Agarwal, Bhowmick, Shah, Krishnamachari, & Dhar, 2017).

We have tested and validated our tool, ORIS, by performing a case study, available as a supplementary document (https://github.com/urmi-21/ORIS). The method details are also included in the supplementary document. A user guide is available from https://github.com/urmi-21/ORIS.

ORIS is developed particularly for biologists and researchers who are working in the area of DNA replication. ORIS allows users with little or no programming background to interactively explore whole genome sequences and identify the putative origin of replication sites in the genome of interest. An obvious advantage of using ORIS is that all the computational methods implemented in ORIS are accessible through a simple and intuitive GUI but the user has to manually document/log the metadata associated with the results (e.g., the method used, method parameters, and input files). We plan to implement automatic logging in future releases of ORIS so that this information is easily recorded along with the results, which will make it, even more, easier to reproduce the results while minimizing user efforts.

Acknowledgements

References

Agarwal, M., Bhowmick, K., Shah, K., Krishnamachari, A., & Dhar, S. K. (2017). Identification and characterization of ARS-like sequences as putative origin(s) of replication in human malaria parasite Plasmodium falciparum. *The FEBS Journal*, *284*(16), 2674–2695. doi:10.1111/febs.14150

Alberts, B., Bray, D., Hopkin, K., Johnson, A. D., Lewis, J., Raff, M., Roberts, K., et al. (2013). *Essential cell biology*. Garland Science. doi:10.1201/9781315815015

Beauchamp, K. G., & Yuen, C. (1979). *Digital methods for signal analysis*. Routledge.

Egan, E. S., Fogel, M. A., & Waldor, M. K. (2005). Divided genomes: Negotiating the cell cycle in prokaryotes with multiple chromosomes. *Molecular Microbiology*, *56*(5), 1129–1138. doi:10.1111/j.1365-2958.2005.04622.x

Frank, A., & Lobry, J. (2000). Oriloc: Prediction of replication boundaries in unannotated bacterial chromosomes. *Bioinformatics*, *16*(6), 560–561. doi:10.1093/bioinformatics/16.6.560

Gao, F., & Zhang, C.-T. (2008). Ori-finder: A web-based system for finding oriCs in unannotated bacterial genomes. *BMC Bioinformatics*, *9*(1), 79. doi:10.1186/1471-2105-9-79

O'Donnell, M., Langston, L., & Stillman, B. (2013). Principles and concepts of DNA replication in bacteria, archaea, and eukarya. *Cold Spring Harbor Perspectives in Biology*, *5*(7), a010108. doi:10.1101/cshperspect.a010108

Shah, K., & Krishnamachari, A. (2012). Nucleotide correlation based measure for identifying origin of replication in genomic sequences. *Biosystems*, *107*(1), 52–55. doi:10.1016/j.biosystems.2011.09.003