Poster H-61

Identification of replication origins in archaea using a moving window model along the double strands

Authors:
Yonghong Wang (*Department of Physics, Tianjin University, Tianjin, 300072, China; School of Computer Science and I*)
Xiaolei Liu (*Department of Physics, Tianjin University, Tianjin, 300072, China*)

Short Abstract: A new method has been developed to identify replication origins in prokaryotic genomes based on a moving-window model along double strands. Two curves corresponding to base distribution in two DNA strands, intersect at replication origins. The approach can be used to identify single and multiple replication origins in archaea.

Long Abstract:
The DNA replication is one of the fundamental steps in cell division process. The study on replication origin is very important to understand replication mechanism. In the three domains of life, replication process in bacteria starts from a single replication origin. The process in eukarya starts from multiple replication. In archaea, replication process starts from single or multiple replication origins. Almost all proteins involved in archaea replication have their homologous counterparts in eukarya [1-3]. Therefore, it is very important to analyze replication origins of archaea for understanding the replication mechanism in both bacteria and eukarya. With the rapid growth of completely sequenced prokaryotic genomes, identification of replication origins by computational methods becomes more and more important. In the past few years, computational approaches have been applied for identification of replication origin in prokaryotic genomes. Lobry used the DNA walk method to detect the replication origins [4]. Since then, several algorithms have been proposed for the identification of replication origins in prokaryotic genomes [5-9]. In general, the methods are mainly applied for identification of single origins. Recently, the Z curve method has been successfully applied in the identification of the single and multiple replication origins in archaea [10, 11]. With this method, prediction is mainly based on the asymmetry of compositions and oligomers between the leading strands and the lagging strands. Here a simple method is proposed for identification of replication origins in prokaryotic genomes. The method consists of two parts: a moving window model and the wavelet de-nosing technique. A sliding window is moved along the entire sequences. The contents of bases G and T in double strands are respectively calculated within a window. Two curves are formed to describe bases G and T distribution respectively along the main strand and the complementary strand. Taking the asymmetrical distribution of bases G and T around replication origins into account, the two curves intersect at some points close to the replication origins and termini. Moving one curve upwards or downwards makes the two curves completely symmetrical. The locations of the intersection points correspond to the replication origins and termini accurately. As the local composition fluctuates along DNA sequences, wavelet de-noising technique is used to reduce the noise signals. Using the present method, two archaea genomes have been analyzed, which are Halobacterium sp. NRC-1 and Pyrococcus abyssi GE5. In the archaea, the replication origins have been experimentally proven at the sequence level. Their genomic sequences were downloaded

from http://www.ncbi.nlm.nih.gov. In order to determine the replication origins, a large window is selected, such as 100kbp. The window moves along the double strands at 10 bp intervals. The wavelet de-noising technique includes two steps: wavelet decomposition and reconstruction. The Haar wavelet is applied to decompose the signal sequences until the 14th level. The signal sequences are reconstructed using the original low frequency coefficients and the high frequency coefficients modified by a threshold Cj. Actually the reconstructed signals are the same even though the signals are decomposed into more than 14 levels. Using the predicted criteria, replication origins for the two archaea are identified. For Pyrococcus abyssi, the two curves intersect at two points, the predicted replication origin lies at the position 1230kb. The prediction is consistent with the experiment evidence [12]. The two curves for Halobacterium sp. NRC-1 interact at four points, two replication origins are predicted. They lie at 910160bp and 1806200 bp, respectively. However, experiment evidence shows that Halobacterium sp. NRC-1 only has one replication origin, located in the range of 1807930-1809486bp. One of the prediction locations is consistent with the range. For the other location, Zcurve method also has provided similar prediction [11]. The characteristics around the location are consistent with those of most known replication origins. For prokaryotic organisms, the single replication origins can be successfully identified by computational methods. However, identification of multiple replication origins still remains a challenge. Clarification of replication mechanism can provide insight into the understanding of replication mechanism of eukarya.

References

[1] D. R. Edgell and W. F. Doolitte. Archaea and the origins of DNA replication proteins. Cell 89: 995-998, 1997.

[2] I. K. Cann and Y. Ishino. Archaeal DNA replication identifying the pieces to solve a puzzle. Genetics 152: 1249-1267, 1999.

[3] K. Bohlke, F. M. Pisani, M. Rosso, and G. Antranikian. Archaeal DNA replication spotlight on a rapidly moving field. Extremophiles 6: 1-14, 2002.

[4] J. R. Lobry. A simple vectorial representation of DNA sequences for detection of replication origins in bacteria. Biochimie 78: 323-326, 1996.

[5] A. Grigoriev. Analyzing genomes with cumulative skew diagrams. Nucleic Acids Res. 26: 2286-2290, 1998.

[6] J. Mrazek and S. Karlin. Strand compositional asymmetry in bacterial and large viral genomes. Proc. Natl. Acad. Sci. USA 95: 3720-3725, 1998.

[7] M. J. Mclean, K. H. Wolfe, and K. M. Devine. Base composition skew, replication orientation, and gene orientation in 12 prokaryote genomes. J. Mol. Evol. 47: 691-696, 1998.

[8] S. L. Salzberg, A. R. Kerlavage, and J. F. Tomb. Skewed oligomers and origins of replication. Gene 217: 57-67, 1998.

[9] P. Lopez, H. Philippe, H. Myllykallio, and P. Forterre. Identification of putative chromosomal origins of replication in archaea. Mol. Microbiol. 32: 883-886, 1999.

[10] R. Zhang and C.T. Zhang. Single replication origin of the archaeon Methanosarcina mazei revealed by the Z curve method. Biochem. Biophys. Res. Commun. 297: 396-400, 2002.

[11] R. Zhang and C. T. Zhang. Multiple replication origins of the archaeon Halobacterium species NRC-1. Biochem. Biophys. Res. Commun. 302: 728-734, 2003..

[12] W. V. Myllykallio, P. Lopez, P. Lopez-Garcia, R. Heilig, W. Saurin, Y. Zivanovic, H. Philippe, and P. Forteere. Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. Science 288: 2212-2215, 2000.