Modeling the impact of DNA sequence repeats on gene inversions and gene-strand bias in bacteria

Siva Subramanian Ayeraselvan, Malhar Atre, Sabari Sankar Thirupathy

IISER THIRUVANANTHAPURAM

Abstract

In bacteria, gene inversions switch genes from leading to the lagging strands of DNA replication and vice versa could influence the degree of gene-strand bias. I modeled a bacterial genome to examine how frequency and distribution of DNA sequence repeat (direct, inter-inverted, and intra-inverted repeats) affect gene inversions and, consequently, gene-strand bias. A null model with no selection pressure that allowed inversions without constraints greatly impacted gene-strand bias. The gene-strand bias stabilized at 50-50 (leading and lagging) irrespective of the initial parameters set. Imposing selection pressure onto the null model stabilizes the extant gene-strand bias.

Introduction

In bacterial genomes, repetitive DNA sequences are present across the genome,

(a) **Direct repeats** are identical sequences repeated in a head-to-tail fashion
(b) **inverted repeats** are identical sequences oriented in opposite directions.

The directionality of the repeats dictates their behavior. Among those repeats, inverted repeats are capable of **inversions**, defined as a type of chromosomal rearrangement where a segment of DNA is reversed and reinserted back into the same chromosome.

These inversions impact gene-strand bias and it refers to the uneven distribution of genes between the leading and lagging strands of DNA. intra-inverted repeats located within the same replicore and inter-inverted repeats located across two replicores. A replicore is a set of DNA sequences from the ori site (origin of replication) to ter site (termination) . Genes can be located on either the leading or lagging strand.

Studying the relationship between inversions and gene strand bias is important for understanding how bacteria adapt to changing environments and evolve new traits. The model is to mimic the natural bacterial genome, considering the bacterial genome as a circular genome with direct repeats, inverted repeats, leading strand genes, lagging strand genes, ori, and ter sites. Studying the distribution and characteristics of inverted repeats and the impact of inversions on gene-strand bias in bacterial genomes can provide the evolution of bacterial genomes, adaptation, and selection of the stable gene-strand bias.

Graphical representation of the genome model

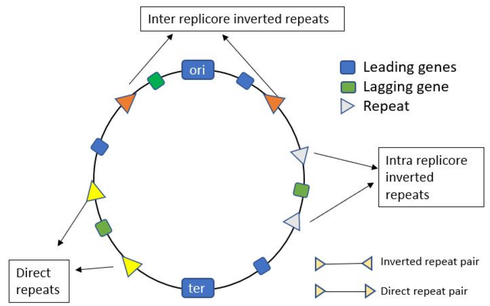

Fig 1: Graphical representation of the built bacterial genome model

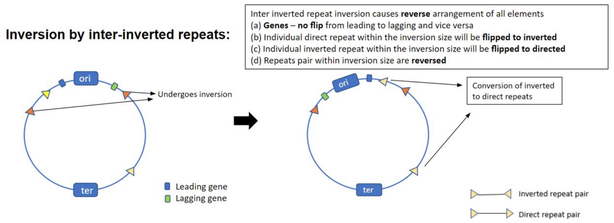

Fig 2: Inversion by inter-inverted repeat pair

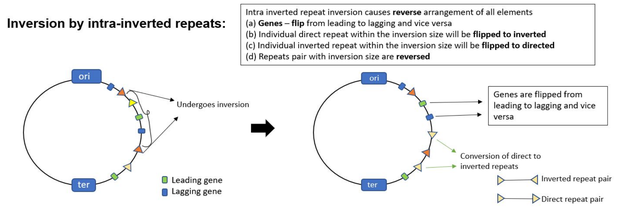

Fig 3: Inversion by intra-inverted repeat pair

Selection pressure

(a) Inversions causing drastic gene count imbalance are excluded i.e. when the ratio of ori-ter and ter-ori gene counts are within the threshold, the inversion happens.
(b) Introduction of the fitness values from the normal distribution for the individual leading and lagging genes
(c) Inversion disparity score for an inversion based on the leading and lagging genes between an intra-inverted repeat pair (IDS) is defined as the resulting sum of the genes determined by assigning positive to each leading gene and a negative to each lagging gene between the intra-inverted repeated pair. An inversion disparity limit (IDL) is set to determine if an inversion event can occur. If the IDS falls within the range of IDL, the inversion happens. If the score exceeds the inversion disparity limit, the inversion is considered to be disruptive and unlikely to occur. To find the optimized penalty limit, maximum likelihood estimator was used with a very small dataset of genes and repeats and extending it onto large datasets.

Results

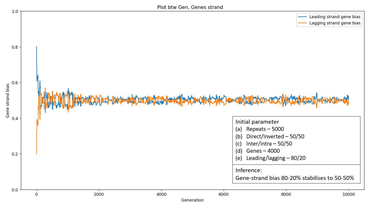

Fig 4: Null model results with specified parameter set with no selection constraints

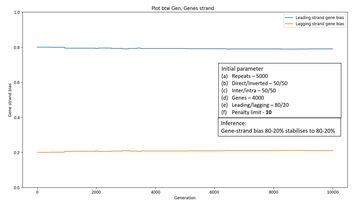

Fig 5: Model results with specified parameter set (Penalty limit - 10) and inference

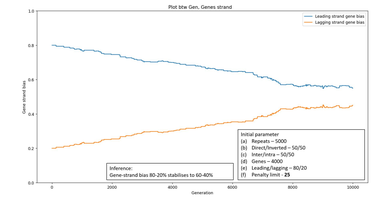

Fig 6: Model results with specified parameter set (Penalty limit - 25) and inference

Fig 7: Model results with specified parameter set (Penalty limit - 50) and inference

Interpretations:

Null model with no selection pressure stabilises the gene strand bias to 50-50% and this predicts the evolutionary aspects of the model as no selection is imposed on the inversions.

The model with selection pressure, imposed selection on the inversion based on gene imbalance and inversion score, the gene-strand bias is not affected. When the stringent inversion disparity score is low, gene strand bias remains contract and increasing the limit, allows larger inversions to taken happen, stabilizing the gene-strand bias.

Strengths, Discussion and Future Improvements of the model

(a) The model considers "selection constraints," implying that natural selection does not favor all inversions equally. Certain inversions may be more beneficial or less detrimental in specific evolutionary contents. These constraints could be influenced by factors such as the function of the genes involved, the impact on gene regulation, or the organism's overall fitness.
(b) This can significantly impact bacterial evolution and adaptation, as genes in specific regions may be essential for survival or confer advantageous traits.
(c) The model parameter can be changed; one can find how repeats number, genes, size, and location result in building different models with different selection pressure constraints, and their respective gene-strand biases can be analyzed.
(d) This model will become the most refined to study all parameters affecting gene-strand bias by incorporating GC skewness, gene gain-loss, and transcription replication machinery collisions. This model accounts for the role of inversions by inverted repeats on gene-strand bias.

References: Ussery, D. W., Wassenaar, T. M., & Borini, S. (n.d.). Word Frequencies and Repeats. *Computational Biology*, 137–150. https://doi.org/10.1007/978-1-84800-255-5_8; Achaz, G., Coissac, R., Netter, P., & Rocha, E. P. C. (2003, August 1). Associations Between Inverted Repeats and the Structural Evolution of Bacterial Genomes. Genetics. https://doi.org/10.1093/genetics/164.4.1279; van Belkum, A., Scherer, S., van Alphen, L., & Verbrugh, H. (1998, June). Short-Sequence DNA Repeats in Prokaryotic Genomes. Microbiology and Molecular Biology Reviews, 62(2), 275–293. https://doi.org/10.1128/mmbr.62.2.275-293.1998.