Authors: Daniella F Lato and G Brian Golding

Journal: Journal of Molecular Evolution

Corresponding Author Information:
G. Brian Golding
McMaster Univeristy
Department of Biology
1280 Main St. West
Hamilton, ON
Canada
L8S 4K1
Tel.: +905-525-9140
Email: golding@mcmaster.ca

Figure 1

Graphics Program: To create Figure 1, the TikZ package in LaTeX.

Caption: Schematic of the transformation used to scale the positions in the genome to the origin of replication and account for bidirectional replication. Circle (A) represents the original replicon genome without any transformation. Circle (B) represents the same replicon genome after the transformation. The origin of replication is denoted by "*oriC*" and the terminus of replication is denoted by "*ter*". The dashed line represents the two halves of the replicon separate by replication. The replicon genome in this example is 100 base pairs in length. Every 10 base pairs is denoted by a tick on the genome. The origin in (A) is at position 20 in the genome and is transformed in (B) to become position 1. The terminus is at position 60 in (A) and position 60 and 40 in (B). The terminus has two positions in (B) depending on which replicon half is being accounted for. If the replication half to the right of the origin is considered, the terminus will be at position 40. If the replication half to the left of the origin is considered, the terminus will be at position 60. Position 40 in (A) becomes position 20 in (B). Position 80 in (A) becomes position 40 in (B), because of the bidirectional nature of bacterial replication. "bp" denotes base pairs

Figure 2

Graphics Program: To create Figure 2, the TikZ package in LaTeX.

Caption: Schematic of the ancestral reconstruction of both the nucleotide and genomic position. Each horizontal row of rectangles represents three hypothetical bacterial genomes (a, b, c). The genomic position is indicated at the top of the diagram. The phylogenetic tree showing the relationship between all three bacteria is pictured on the right of the diagram. The light grey rectangle denotes the genomic segment of interest. In bacteria (a) and (b), this segment is located at genomic positions 1-3. In bacteria (c), this segment is located at genomic positions 7-9. Within this genomic region of interest there is a substitution where the nucleotides changed from C $\rightarrow$ A, this is highlighted in red and underlined. This substitution is at position $\underline{3}$ in bacteria (a) and (b), and in position $\underline{9}$ in bacteria (c). This is depicted by the values ($\underline{\mathrm{C}}_3$) and ($\underline{\mathrm{A}}_9$). The ancestral reconstruction process in this analysis can be seen at the inner nodes of the phylogenetic tree by the values ($\underline{\mathrm{C}}_3$). The most parsimonious reconstruction of the sequence and associated genomic position is having the value ($\underline{\mathrm{C}}_3$) present at the ancestor of bacteria (a) and (b). The ancestral node of all three bacteria would have a reconstruction of the sequence and associated genomic position of ($\underline{\mathrm{C}}_3$ / $\underline{\mathrm{A}}_9$). In this situation where there is a "tie" for two most parsimonious options, the option with the highest likelihood estimate would be chosen using maximum-likelihood methods (see[?] for more details). This would mean that in bacteria (c) there was a substitution from C $\rightarrow$ A which is also associated with a genomic position of 9.

Figures 3a - 3c

Graphics Program: To create Figures 3a - 3c, the R programming language and LaTeX were used.

Caption: The bar graphs show the number of substitutions along the genomes of *E. coli* (a), *B. subtilis* (b), and *Streptomyces* (c). For *E. coli* and *B. subtilis*, the distance from the origin of replication is on the x-axis beginning with the origin of replication denoted by position zero on the left, and the terminus indicated on the far right. For *Streptomyces* the origin of replication is denoted by position zero. The genome located on the shorter chromosome arm (to the left of the origin) has been given negative values, while the genome on the longer chromosome arm (to the right of the origin) has been given positive values. The origin of replication in the *Streptomyces* graph (c), has been visualized at position zero by a red vertical line. The y-axis of the graphs indicate the number of substitutions per 10,000 base pairs found at each position of the *E. coli* (a), *B. subtilis* (b), and *Streptomyces* (c) genomes. Each bar represents a section of the genome that spans 10Kbp. Outliers are represented in light grey bars.

Figures 4a - 4c

Graphics Program: To create Figures 4a - 4c, the R programming language and LaTeX were used.

Caption: The bar graphs show the number of substitutions along the replicons of *S. meliloti*: chromosome (a), pSymA (b), and pSymB (c). Distance from the origin of replication is on the x-axis beginning with the origin of replication denoted by position zero on the left, and the terminus indicated on the far right. The y-axis of the graph indicates the number of substitutions per 10,000 base pairs of the replicons of *S. meliloti*: chromosome (a), pSymA (b), and pSymB (c). Each bar represents a section of the genome that spans 10Kbp. Outliers are represented by light grey bars.

Figures 5a - 5c

Graphics Program: To create Figures 5a - 5c, the R programming language and LaTeX were used.

Caption:The graphs show the values of dN, dS, and ω along the genomes of *E. coli* (a), *B. subtilis* (b), and *Streptomyces* (c). For *E. coli* and *B. subtilis*, the distance from the origin of replication is on the x-axis beginning with the origin of replication denoted by position zero on the left, and the terminus indicated on the far right. For *Streptomyces* the origin of replication is denoted by position zero. The genome located on the shorter chromosome arm (to the left of the origin) has been given negative values, while the genome on the longer chromosome arm (to the right of the origin) has been given positive values. The origin of replication in the *Streptomyces* graph (c), has been visualized at position zero by a grey vertical line. The y-axis of the graph indicates the value of dN, dS, and ω found at each gene segment position of the *E. coli* (a), *B. subtilis* (b), and *Streptomyces* (c) genomes. Outliers are represented by light grey open circles. The average dN, dS, and ω values for each 10,000bp regions of the genome was calculated and represented by the dark brown points. A trend line represented in blue (using the `loess` method), was fit to these average values and the associated 95% confidence intervals for this line is represented by the grey ribbon around the blue trend line. For a complete list of outlier and zero value information, please see the Supplementary Material.

Figures 6a - 6c

Graphics Program: To create Figures 6a - 6c, the R programming language and LaTeX were used.

Caption: The graphs show the values of dN, dS, and ω along the replicons of *S. meliloti*, chromosome (a), pSymA (b), and pSymB (c). Distance from the origin of replication is on the x-axis beginning with the origin of replication denoted by position zero on the left, and the terminus indicated on the far right. The y-axis of the graph indicates the value of dN, dS, and ω found at each gene segment position of the chromosome (a), pSymA (b), and pSymB (c) of *S. meliloti*. Outliers are represented

by light grey open circles. The average dN, dS, and ω values for each 10,000bp regions (for the chromosome) and 50,000bp regions (for both pSymA and pSymB) of the replicons were calculated and represented by the dark brown points. A trend line represented in blue (using the `loess` method), was fit to these average values and the associated 95% confidence intervals for this line is represented by the grey ribbon around the blue trend line. For a complete list of outlier and zero value information, please see the Supplementary Material.