Correspondence

Flip-flop around the origin and terminus of replication in prokaryotic genomes
Paweł Mackiewicz, Dorota Mackiewicz, Maria Kowalczuk and Stanisław Cebrat

A response to **Evidence for symmetric chromosomal inversions around the replication origin in bacteria** by JA Eisen, JF Heidelberg, O White and SL Salzberg, *Genome Biology* 2000, **1:**research0011.1-0011.9.

Address: Institute of Microbiology, Department of Genetics, Wrocław University, ul. Przybyszewskiego 63/77, 51-148 Wrocław, Poland.

Correspondence: Stanisław Cebrat. E-mail: cebrat@microb.uni.wroc.pl

The problem of rearrangements in closely related prokaryotic genomes has been discussed several times recently [1-4]. A characteristic feature of these rearrangements is that many orthologs (genes coding for the same function in different genomes) stay at the same distance from the origin or terminus of replication, but they can be positioned on either of the two replichores (oppositely replicated halves of the genome [5], Figure 1). A specific picture is obtained when the positions of genes in one genome are plotted against the positions of their orthologs in a closely related genome (Figure 2a). A practical implication is that orthologous sequences found at the same distance from the origin or terminus of replication probably code for the same function. It is also important from the evolutionary point of view. The question is, why can genes be translocated between replichores but have their distance from the origin or terminus of replication conserved?

Tillier and Collins [4] have argued that a substantial proportion of rearrangements of gene order result from recombination sites that are determined by

the positions of replication forks. Their (plausible) theory is that replication forks are hot spots for recombination. Given that the two replication forks are at approximately the same distance from the origin (during the bi-directional replication), translocations are symmetrical about the origin-terminus axis. Thus, according to Tillier and Collins [4], specific constraints on the mechanisms of recombination are responsible for the observed bias in the frequency of finding particular rearrangement products. We argue that it is selection that may be mainly responsible for the observed bias, and the probability that a rearrangement product is viable depends on its topology.

The first aspect of topology that could lead to biased genome rearrangements is the distance of a gene from the origin of replication, as this determines the relative copy number of the gene in each cell of fast-growing cultures of bacteria. If the generation time is shorter than the replication period, the number of copies of genes lying near the origin is higher than the number of copies of genes lying near the terminus. Thus, selection pressure leads to the

optimal position of genes with respect to the distance from the origin of replication [6,7]. As a result, as well as observing a bias towards specific
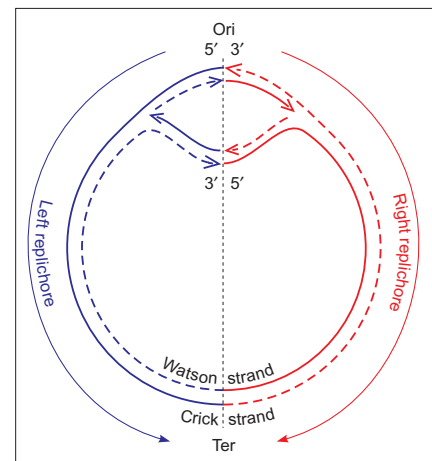

Figure 1
The topology of bi-directional replication of a circular prokaryotic chromosome. The continuous line is the DNA strand replicated as the leading strand; the dashed line is the DNA strand replicated as the lagging strand; Ori, the origin of replication; Ter, the terminus of replication. Ori and Ter divide the chromosome into two replichores, arbitrarily called left and right.

rearrangements, there is an asymmetry in the nucleotide composition of gene sequences and a biased amino-acid composition of proteins encoded by genes along the chromosome [8].

The second factor that may influence the likelihood of particular genome rearrangements is that the replication-associated mutational pressure is different for the leading and the lagging DNA strands. Different rates of accumulation of nucleotide substitutions on the leading and lagging strands suggest that there are qualitative and quantitative differences in the nucleotide com-

position of protein-coding sequences lying on different DNA strands ([9] and references therein). That is why sequences that have recently changed location from leading to lagging strand, or *vice versa*, are more prone to accumulating mutations. Thus, any inver-

sion of a gene within the replichore (Figure 3c) changes its sense strand from the leading to the lagging strand during replication, or *vice versa*, and increases the mutation rate of this gene [10,11]. If the inversion encompasses the origin or the terminus of replica-
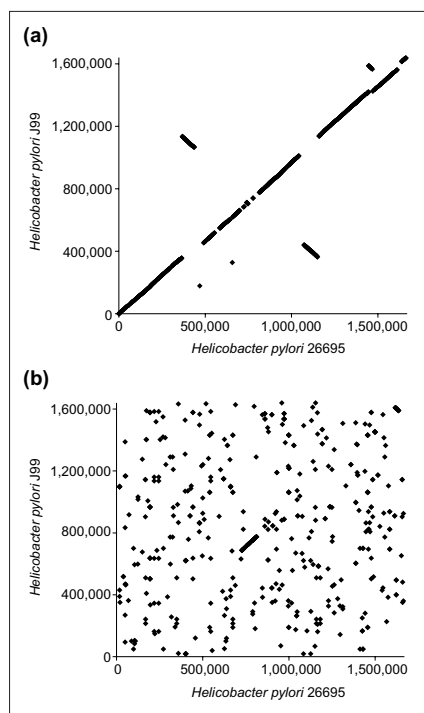

Figure 2
Plots of the relative positions of orthologs in the *Helicobacter pylori* J99 and *H. pylori* 26695 genomes. The values on the *x* and *y* axes represent the positions of genes on chromosomes, in base pairs. **(a)** The closest orthologs (best matches) that have not switched their positions between the leading and the lagging DNA strands. **(b)** All orthologs that have switched positions between the leading and the lagging DNA strands. The genome sequences and orthologs, extracted from the database of Clusters of Orthologous Groups ('COGs') [12], were obtained from the National Center for Biotechnology Information [13].
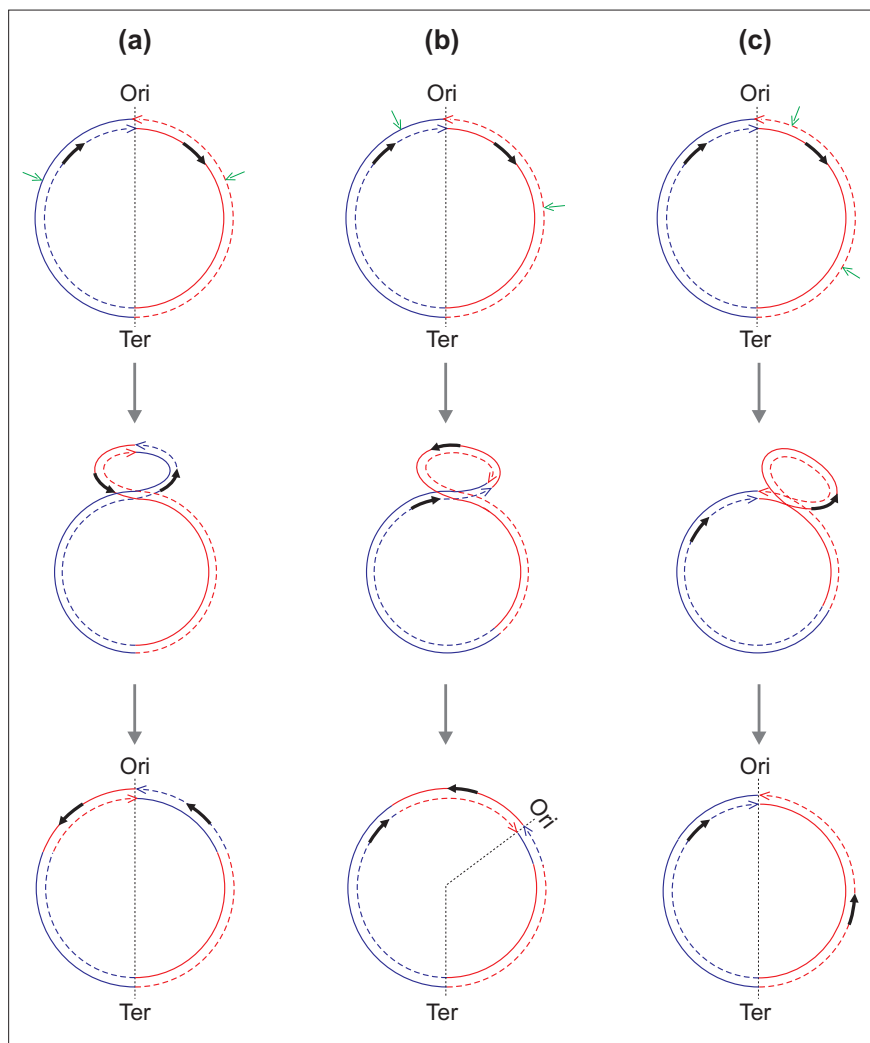

Figure 3
Consequences of inversions at different locations in a prokaryotic chromosome. The green arrows indicate sites of recombination. The black arrows represent a sense strand of a gene. Note that if the sense strand is lying on the leading DNA strand, the direction of transcription of the gene is the same as the direction of replication-fork movement. **(a)** A symmetrical inversion encompassing the origin of replication. After the inversion, the distances to the origin and the locations of the genes do not change with respect to the leading and lagging DNA strands. **(b)** The inverted region encompasses the origin but the origin is not located in the center of this region. As a result, the lengths of replichores change and the distances of the noninverted genes to the origin change, although the locations of the genes do not change with respect to the leading and lagging DNA strands. **(c)** An inversion within a replichore. The locations of genes within the inverted sequence change with respect to the leading and lagging DNA strands. Furthermore, genes located away from the center of the inverted region change their distance from the origin.

tion, the position of the gene (with respect to the direction of replication) is not changed (Figure 3a,b). In effect, the specific bias in the genome rearrangement is observed only for the 'closest' orthologs (top matches). The characteristic plot seen in Figure 2a shows orthologs that have not changed their position with respect to the leading or lagging behavior of the DNA strand. An inversion within a replichore does, however, change the position of the gene in this respect (Figure 3c). Very high mutation rates could eliminate a gene that has changed strand, unless the inversion is connected with a duplication. In the case of a duplication, the second copy of the gene can play a different role, and it could be allowed to diverge much faster, for example to generate a paralog. In Figure 2b we illustrate the relative positions of genes that have changed their positions with respect to the direction of replication; the characteristic diagonal line seen in Figure 2a, showing the highly biased orientation of rearrangements, has disappeared.

The third selection force that could lead to biased rearrangements might be the trend towards keeping both replichores the same size (see also [7]). If there is a selection pressure ensuring that the length of the two replichores in prokaryotic genomes stays almost the same, inversions symmetrical in respect to origin or terminus of replication ought to be preferred. Figure 3b shows how a recombination event encompassing the origin of replication but with the origin not in the center of the inverted fragment generates replichores of different length and changes the distances from the origin to genes lying outside the inverted sequence.

All of the above explanations do not exclude the possibility that there are hot spots of recombination connected with the replication forks, as suggested by Tillier and Collins [4], but we would like to stress that selection probably plays a very important role in producing the strange X image of translocation topology in closely related genomes.

Acknowledgements

Jonathan A Eisen responds:

I welcome the letter by Mackiewicz *et al.* relating to whole-genome X-alignments (which we refer to as X-files). I agree that selection is likely to be a contributing factor in the observations, on the basis of comparative genomics, that the distance a gene is from the origin of replication is maintained over evolutionary time [1,2,4]. Their suggestions for possible selective forces are all entirely reasonable, and it will be worth pursuing the contribution of each in future work. I would like to point out a few additional issues relating to the X-files, however. First, it is important to note that some very important work on this subject has been done using genetic approaches [6,7,14-20]. As pointed out in some of these studies and by Mackiewicz *et al.*, the presence of selection does not necessarily mean that mutation processes are not also an important contributing factor. It is likely that some type of mutation bias (such as strand switching during replication, as suggested by Tillier and Collins [4]) leads to a high frequency of inversions that are symmetric around the origin of replication. Many other inversions are also likely to occur. Thus, negative selection (such as selection against changes in replichore size or gene dosage, as suggested by Mackiewicz *et al.*) is likely to then cause the inversions that are observed over evolutionary time to be predominantly those that are symmetric around the origin of replication. What we now need from a scientific point of view is more information on frequencies and types of genome inversion that occur in the absence of selection, as well as information on the fitness differences between strains with different inversions.

In addition, I would like to comment on the suggestion that the observation of the X-alignment pattern can aid in making functional predictions for genes. Mackiewicz *et al.* suggest that if one finds homologous genes at the same distance from the origin of replication one can conclude that they have the same function. I would suggest this is not a good functional prediction criteria. As discussed previously [1], within individual genomes, pairs of paralogous genes are frequently found on both sides of the replication origin at equal distances (leading to a within-genome X pattern). We proposed that this is likely to be due to inversions that split tandemly duplicated paralogous genes. Because one or both of these genes may have diverged in function from that of a common ancestor, their position from the replication origin will not help in predicting their function when compared to other species. In addition, as orthologous genes do not always have the same function, even without the occurrence of tandem duplications, genome location alone is not likely to be a reliable predictor of gene function. Thus, while I believe the X-alignment pattern may reveal a great deal about mutation and selection pressures relating to inversions and genome position, I am not convinced that the function of a gene can be readily predicted by identifying homologous genes equidistant from replication origins.

Jonathan A Eisen

The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA. E-mail: jeisen@tigr.org

References
1. Eisen JA, Heidelberg JF, White O, Salzberg SL: **Evidence for symmetric chromosomal inversions around the replication origin in bacteria.** *Genome Biol* 2000, **1:**research0011.1-0011.9.
2. Suyama M, Bork P: **Evolution of prokaryotic gene order: genome rearrangements in closely related species.** *Trends Genet* 2001, **17:**10-13.
3. Read TD, Brunham RC, Shen C, Gill SR, Heidelberg JF, White O, Hickey EK, Peterson J, Utterback T, Berry K, *et al.*: **Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39.** *Nucleic Acids Res* 2000, **28:**1397-1406.
4. Tillier ER, Collins RA: **Genome rearrangement by replication-directed translocation.** *Nat Genet* 2000, **26:**195-197.

5. Blattner FR, Plunkett G III, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, *et al.*: **The complete genome sequence of *Escherichia coli* K-12.** *Science* 1997, **277:**1453-1462.

6. Liu SL, Sanderson KE: **Rearrangements in the genome of the bacterium *Salmonella typhi*.** *Proc Natl Acad Sci USA* 1995, **92:**1018-1022.

7. Liu SL, Sanderson KE: **Highly plastic chromosomal organization in *Salmonella typhi*.** *Proc Natl Acad Sci USA* 1996, **93:**10303-10308.

8. Mackiewicz P, Gierlik A, Kowalczuk M, Dudek MR, Cebrat S: **How does replication-associated mutational pressure influence amino acid composition of proteins?** *Genome Res* 1999, **9:**409-416.

9. Frank AC, Lobry JR: **Asymmetric substitution patterns: a review of possible underlying mutational or selective mechanisms.** *Gene* 1999, **238:**65-77.

10. Tillier ER, Collins RA: **Replication orientation affects the rate and direction of bacterial gene evolution.** *J Mol Evol* 2000, **51:**459-463.

11. Szczepanik D, Mackiewicz P, Kowalczuk M, Gierlik A, Nowicka A, Dudek MR, Cebrat S. **Evolution rates of genes on leading and lagging DNA strands.** *J Mol Evol* 2001, **52:** 426-433.

12. Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, Kiryutin B, Galperin MY, Fedorova ND, Koonin EV: **The COG database: new developments in phylogenetic classification of proteins from complete genomes.** *Nucleic Acids Res* 2001, **29:**22-28.

13. **National Center for Biotechnology Information** [ftp://ftp.ncbi.nlm.nih.gov]

14. Segall AM, Roth JR: **Recombination between homologies in direct and inverse orientation in the chromosome of *Salmonella*: intervals which are nonpermissive for inversion formation.** *Genetics* 1989, **122:**737-747.

15. Schmid MB, Roth JR: **Selection and endpoint distribution of bacterial inversion mutations.** *Genetics* 1983, **105:**539-557.

16. Segall A, Mahan MJ, Roth JR: **Rearrangement of the bacterial chromosome: forbidden inversions.** *Science* 1988, **241:**1314-1318.

17. Mahan MJ, Roth JR: **Ability of a bacterial chromosome segment to invert is dictated by included material rather than flanking sequence.** *Genetics* 1991, **129:**1021-1032.

18. Mahan MJ, Roth JR: **Reciprocality of recombination events that rearrange the chromosome.** *Genetics* 1988, **120:**23-35.

19. Francois V, Louarn J, Patte J, Rebollo JE, Louarn JM: **Constraints in chromosomal inversions in *Escherichia coli* are not explained by replication pausing at inverted terminator-like sequences.** *Mol Microbiol* 1990, **4:**537-542.

20. Rebollo JE, Francois V, Louarn JM: **Detection and possible role of two large nondivisible zones on the *Escherichia coli* chromosome.** *Proc Natl Acad Sci USA* 1988, **85:**9391-9395.