Rates of DNA Mutation in Genes and Inte-Gene Regions

Brian French, Nicole Markley, Laszlo Csonka, and Mark Daniel Ward

Department of Biology, Purdue University, West Lafayette, IN

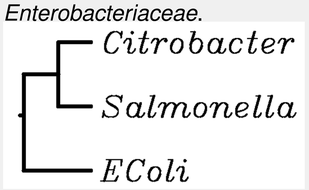
Abstract

The aim of our analysis was to compare the rates of evolution of DNA sequences of genes and intergenic regions. The chromosomes of the related genera of bacteria, Escherichia coli K-12, Salmonella enterica serovar Typhimurium, and Citrobacter koseri, contain extensive regions in which the order of genes of identical. The fact that the gene order is so conserved suggests that both the genes and intergenic spacers were derived by evolution from corresponding sequences in the last common ancestor of the three organisms, and it enabled us to compare the sequence changes in functional genes and in intergenic spacers across the three organisms.

Pairwise comparisons using BLASTN were used to identify regions that had shared genes. ClustalW was then used to identify single base mutations in both genes and inter-gene regions. Mutation frequencies were determined for all paired comparisons. This analysis indicated that the mutation frequencies are significantly higher in intergenic spacers do not have specific sequence-dependent functions, and therefore, they can accumulate mutations more liberally than genes, which are constrained by their function.

Background

- Genomes have both useful genetic information coded in genes and presumed useless inter-gene regions termed as gaps.
- Gene pairs between organisms can be isolated using a tool called **B**asic **L**ocal **A**lignment **S**earch **T**ool, commonly referred to as BLAST.
- *Salmonella Typhimurium LT2*, *Citrobacter Koseri*, and *Escherchia Coli MG1655* were chosen because they are closely related members of *Enterobacteriaceae*.



- The phylogeny tree built from 16s rRNA shows the relative relations of the three bacteria chosen to be analyzed, which is an explanation of why E. Coli and Citrobacter have smaller common regions than Citrobacter and Salmonella.

Motivation

- Mutation in genes has a double constrainment of DNA's built in error-handling as well as maintenance of gene function.
- Gaps are presumed to have no function, so they should mutate faster and more randomly.
- Choosing pairs of bacteria that were closely related could be used to infer rates of mutation in homologous genes and gap regions.
- Differing rates of mutation would suggest that gaps don't share genes' constrainment due to function.

Question

- Do gaps and genes mutate in the same way?
- Is this trend consistent in different pairs of bacteria?
- Do mutations show any form of dependence on other mutations?

Hypothesis

We hypothesize that genes will mutate at different rates than gaps because they have to code information essential to cell survival.

Illustrations

Figure 1

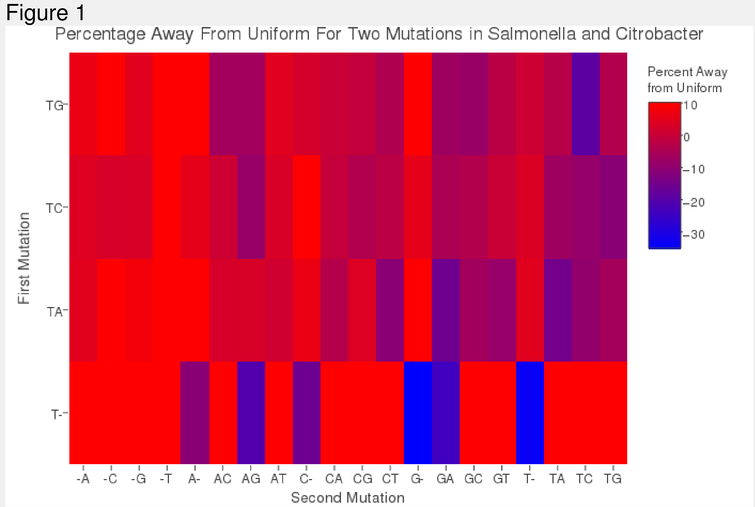


Figure 2

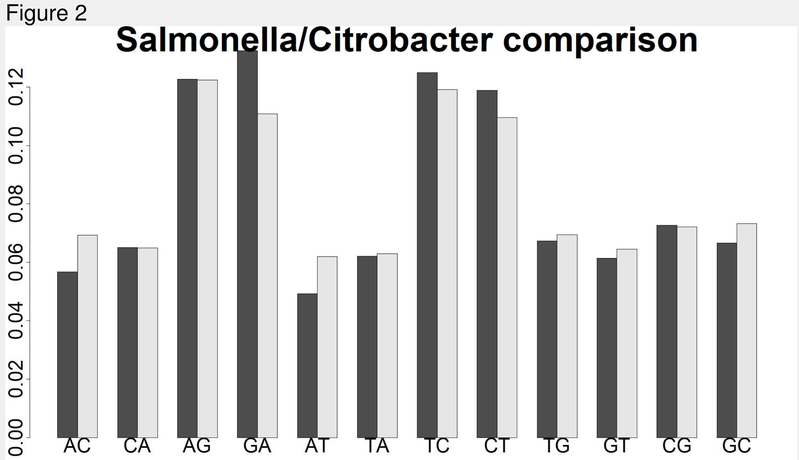
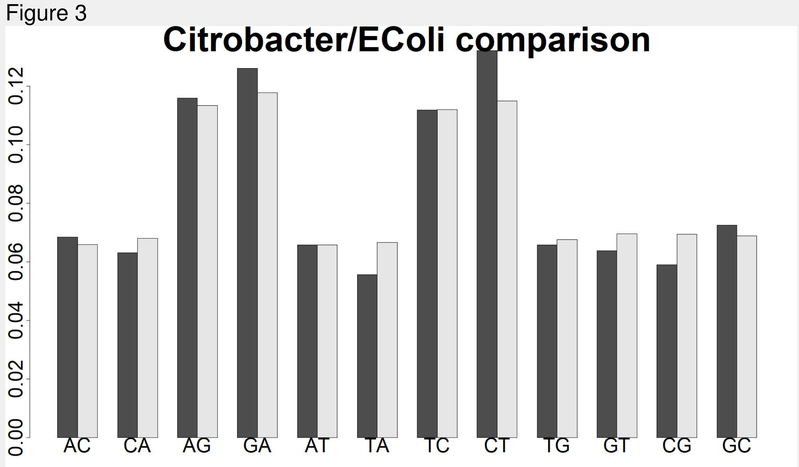


Figure 3



Procedure

Genomes were run through BLAST to identify matching regions for each pair comparison. BLAST matches were presumed to be homologous genes. Two matching regions with a non-matching region between them led the non-matching region to be labeled as a gap. Any gap longer than 500 bases was presumed to contain a gene that wasn't present in the compared genome and was therefore discarded.
Matched regions (both genes and gaps) were run through CLUSTALw and any base mismatch preceded and followed by at least 3 matching bases was presumed to be a mutation. Mutation rates were then tabulated, plotted, and chi-squared tested on mutation counts.

Conclusion

- Purine-¿purine and pyrimidine-¿pyrimidine mutations were more frequent than other possible single base-pair changes both in genes and in gaps, indicating that single-base changes were not random.
- Genes and gaps have statistically different mutation rates as shown by a chi-square test ($p < 2.2 * 10 \times 10^{-16}$) in all three gene pairs
- This suggests that genes and gaps have different mutation patterns
- When insertions and deletions are removed, there's a large drop in statistical significance, which indicates that a major part of the difference is gaps losing information.
- When insertions and deletions are removed, E. coli and Salmonella have the smallest change in significance, indicating that there are more base shift mutations in gaps than genes in that pair.
- In the case of two mutated bases next to each other, there is Markovian dependence in genes
- There were insufficient examples of two mutated bases that were adjacent to each other was in gaps to enable us to draw conclusions about their dependence.

Acknowledgments

This research is supported by NSF grant DMS #1246818 and NSF grant # IOS-1456829

Citations

Thompson, J. D., Gibson, T. J. and Higgins, D. G. 2002. Multiple Sequence Alignment Using ClustalW and ClustalX. Current Protocols in Bioinformatics. 00:2.3:2.3.1?2.3.22.
Stephen F. Altschul, Warren Gish, Webb Miller, Eugene W. Myers, David J. Lipman, Basic local alignment search tool, Journal of Molecular Biology, Volume 215, Issue 3, 1990, Pages 403-410, ISSN 0022-2836, http://dx.doi.org/10.1016/S0022-2836(05)80360-2.
(http://www.sciencedirect.com/science/article/pii/S0022283605803602) BLAST accessed through https://blast.ncbi.nlm.nih.gov/Blast.cgi

Dot Plot Representation of Salmonella/EColi Gene Locations