Rates of DNA Mutation in Genes and Inte-Gene Regions

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

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.

Introduction

Genomes have both useful genetic information 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 Dr. Csonka told me to use them–add real reason later

Motivation

Mutation in genes has a double constrainment of DNA's built in error-handling as well as maintaining function. We suspected that looking at both genes and gaps it could be determined how much similarity there would be in accumulation of mutations at the base level.

Research Question

We want to characterize something. The end.
TL;DR: Do mutations happen in genes and gaps at the same rate?

Illustrations

Figure 1

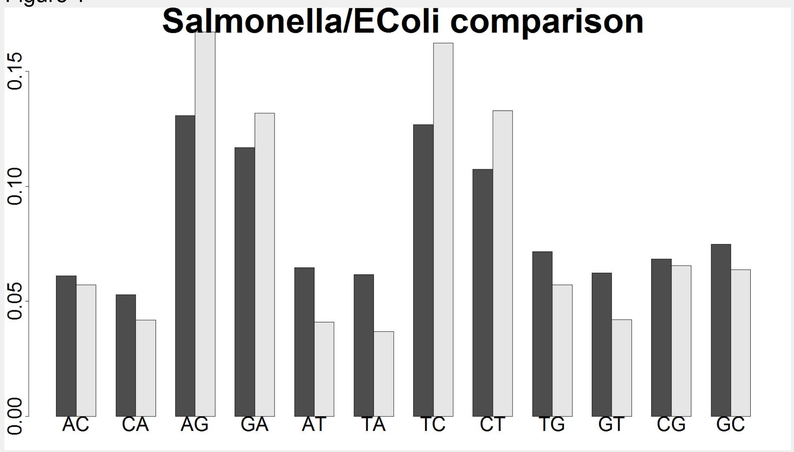


Figure 2

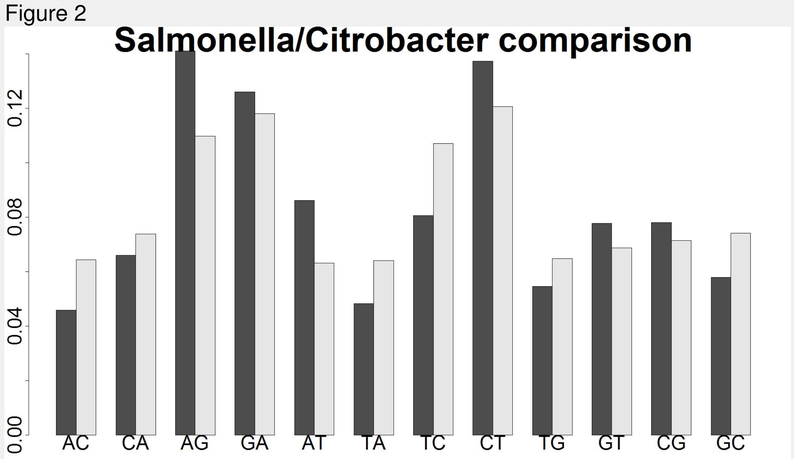
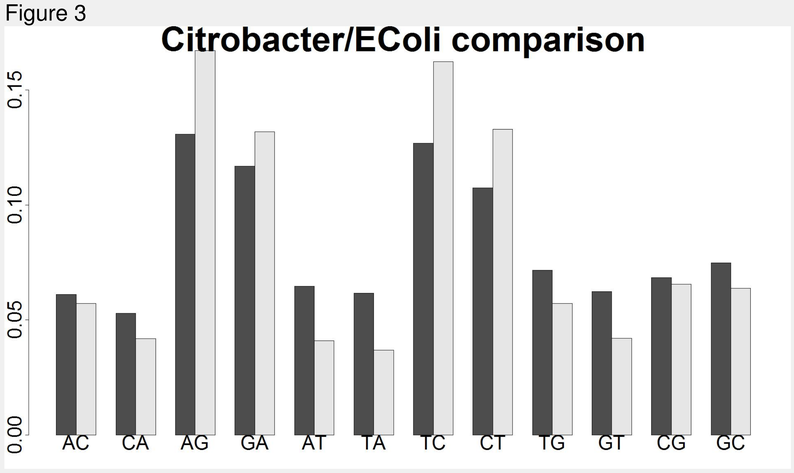


Figure 3



Application

Now I can discuss why that result matters to people.

Conclusion

Looking at the three figures off to the left, we have depictions of the three pairwise comparisons. The plots show the relative frequency of mutations in genes vs gaps. Gaps are shown in black, genes are in white. What immediately stands out is that the A to G/G to A and T to C/C to T mutation is clearly more frequent. Likely, this is because A and G are the two purine bases and T and C are the two pyrimidines.
Running statistical tests across the comparison, the gene and gap mutations are clearly from different distributions. Comparing the distribution of gene and gap mutations for Salmonella and E Coli with insertions and deletions left in, a chi-squared test returns a probability that they came from the same distribution of less than 2.2x10ᵃ16. When insertions and deletions are ignored, the likelihood that gene and gap mutations are from the same distribution rises to 1.766x10ᵃ13. Similar numbers arise from chi-squared testing the other two pairwise comparisons.
Really should say something about randomness testing, but I'm not 100% sure how to go about that.

Acknowledgments

The research of Student Name is supported by NSF grant DMS #1246818. That it totally not the right grant number