Comparative Study of Overlapping Genes in Bacterial Genomes

Yoko Fukuda[1,2]
yfukuda@sfc.keio.ac.jp
Takanori Washio[1,2]
washy@sfc.keio.ac.jp
Yoichi Nakayama[1,3]
ynakayam@sfc.keio.ac.jp
Masaru Tomita[1,3]
mt@sfc.keio.ac.jp

[1] Laboratory for Bioinformatics, Keio University, 5322 Endo, Fujisawa, Kanagawa 252-8520, Japan
[2] Graduate School of Media and Governance, Keio University
[3] Department of Environmental information, Keio University

Keywords: overlapping gene, comparative genomics, nucleotide substitution

1 Introduction

Overlapping genes are defined as a pair of adjacent genes whose coding regions are partly overlapping. Many overlapping genes have been identified in the genomes of procaryotes, bacteriophages, viruses, and mitochondria, but their evolutionary origins, i.e. how they have emerged, is not clearly understood. In a previous work, a comparative study in *Mycoplasma genitalium* and *Mycoplasma pneumoniae*, careful comparisons were made of homologous genes that were overlapped in one species but not in the other. It seems that the overlapping genes were generated primarily due to the loss of a stop codon in either gene, the absence of which resulted in elongation of the 3'end of the gene's coding region. The loss of the stop codon was, in most cases, result of one of the following events: deletion of the stop codon (64.4%); point mutation at the stop codon (4.4%); or frame shift at the end of the coding region (6.7%) [1].

In this work, we analyzed overlapping genes in genomes of *Haemophilus influenzae* and *Salmonella typhimurium* by comparing them with the genomes of *Escherichia coli*, usinf the method described in the previous work.

2 Method

There were 237 overlapping gene pairs in the genome of *H. influenzae*, and 806 gene pairs in the genome of *E. coli*. Of those, 108 adjacent gene pairs are conserved in these two species. Among those adjacent gene pairs, 42% are overlapped in both species; 45% are overlapped only in *E. coli*; and 13% are overlapped only in *H. influenzae*. The 57 gene pairs that are overlapped only in either species were aligned for further investigation.

3 Results and Discussion

Of the 57 gene pairs, 20 overlapping gene pairs were presumably caused by loss of the stop codon at the 3' end of either gene. Fifteen overlapping gene pairs, on the other hand, were presumably caused by simultaneous loss of the start codon at the 5'end of either gene (Fig. 1a). Eighteen pairs seem to have been caused by simultaneous loss of the start codon in one gene and the stop codon in the other (Fig. 1b). We could not determine the cause of overlapping for 7 gene pairs based on their sequences. These findings are summarized in Fig. 2, along with the previous result with *M. genitalium* and *M. pneumoniae*.

We also analyzed overlapping genes in the genome of *S. typhimurium*, whose sequence is only partly available at present. There are 135 overlapping gene pairs in *S. typhimurium*. Among those, 31 homologous gene pairs were found in *E. coli*. There are only 3 gene pairs that are overlapped in *S. typhimurium* but not in *E. coli*. Two were caused by loss of the start codons, one by loss of the stop codon.

Overlapping genes in mycoplasmas were generated primarily due to the loss of a stop codon in either gene. On the other hand, overlapping genes in *H. influenzae* and *E. coli* were due not only to loss of the stop codons but also to loss of the start codons. Biological and evolutionary significance of this difference is currently under investigation.

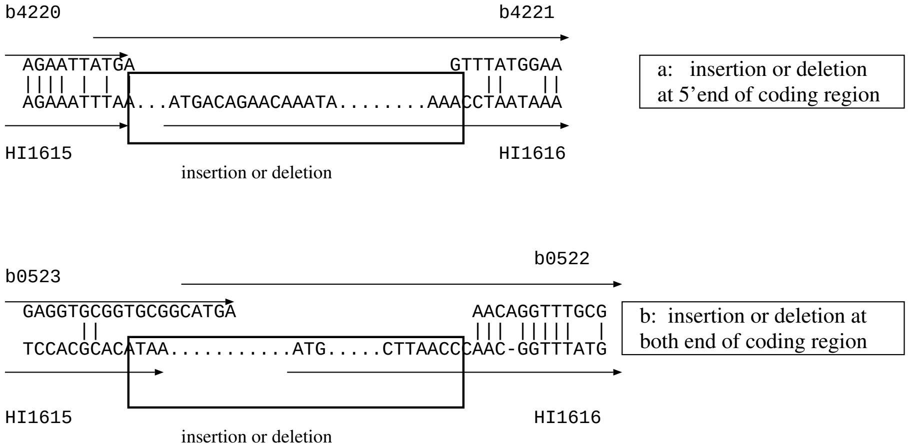

Figure 1: Cause of overlapping genes in *E. coli* and *H. influenzae*.

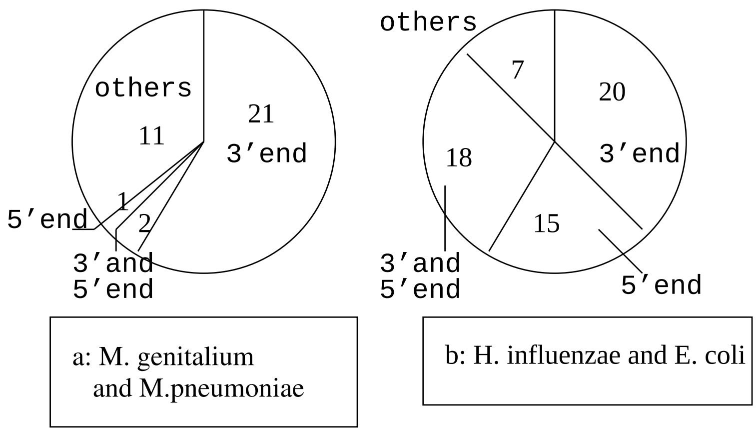

Figure 2: Mutations causing overlapping genes.

References

[1] Fukuda Y., Washio T. and Tomita M., Comparative study of overlapping genes in the genomes of Mycoplasma genitalium and Mycoplasma pneumoniae, *Nucleic Acids Res.*, 27(8):1847–53, 1999.