Evolution of Overlapping Genes: Comparative Genomics of *Mycoplasma genitalium* and *Mycoplasma pneumoniae*

Yoko Fukuda [12] `t95775yf@sfc.keio.ac.jp`
Takanori Washio [13] `washy@sfc.keio.ac.jp`
Masaru Tomita [12] `mt@sfc.keio.ac.jp`

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

1 Introduction

Many overlapping genes have been identified in the genomes of procaryotes, bacteriophages, animal viruses, and mitochondria, some of which have been reported to have functional roles [1], but their evolutionary origin is not clearly understood. We systematically analyzed all overlapping genes in the genomes of two closely related species, *Mycoplasma genitalium* [2] and *Mycoplasma pneumoniae* [3]. In particular, careful comparisons were made for the homologous genes that are overlapped in one species but not in the other.

2 Materials and Methods

We first list all overlapping genes in their genomes according to the annotations in the databases [2, 3]. For each overlapping gene pair in one species, we aligned the sequence, using ClustalW [4], with the sequence of the homologous part of the other species. We then classified all the cases according to the three directional patterns as described in Fig. 1: "end-on", "uni-directional", and "head-on". For those genes that overlap in one species but not in the other, we made careful analyses in order to infer the cause of the overlapping. The putative causes of these events of overlapping were then classified into several types.

3 Results and Discussion

In the genome of *M. genitalium*, there are 155 overlapping gene pairs, according to the TIGR annotation. The genome of *M. pneumoniae*, on the other hand, contains 160 overlapping gene pairs. Many parts of genome sequences of these two species are homologous. There are 134 homologous overlapping gene pairs which exist in both species.

The other 21 and 26 overlapping gene pairs are found only in *M. genitalium* and *M. pneumoniae*, respectively. Comparative analysis of these gene sequences allows us to propose a model of how overlapping genes emerged in the course of evolution.

It was found that 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. More specifically, the loss of the stop codon took place as a result of the following events: deletion of the stop codon (72.1 %); point mutation at the stop codon (4.7%); and frame shift at the end of the coding region (7.0%). Overlapping genes, in a sense, might be thought of as the results of evolutionary pressure to minimize genome size. However, our analysis indicates that many overlapping genes, at

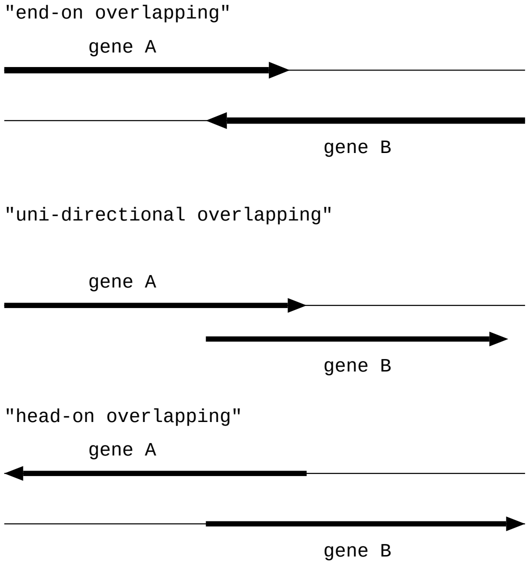

Figure 1: Three patterns of overlapping genes.

Table 1: summary of inferred causes of gene overlapping.

	deletion	point mutation	frame shift	unknown
end-on (→←)	8	1	2	0
uni-directional (→→)	24	1	1	7
total	32	2	3	7

least in the genomes of *M. genitalium* and *M. pneumoniae*, are merely due to incidental elongation of the coding regions.

Acknowledgements

This work was supported in part by a Grant-in-Aid for Scientific Research on Priority Areas 'Genome Science' from The Ministry of Education, Science, Sports and Culture in Japan.

References

[1] Normark, S., Bergstrom, S., Edlund, T., Grundstrom, T., Jaurin, B., Lindberg, F.P., and Olsson, O., Overlapping genes, *Annu. Rev. Genet.*, 17:499–525, 1993.

[2] Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G., Kelley, J.M., *et al.*,The minimal gene complement of Mycoplasma genitalium, *Science*, 270:397–403, 1995.

[3] Himmelreich, R., Plagens, H., Hilbert, H., Reiner, B., and Herrmann, R., Comparative analysis of the genomes of the bacteria Mycoplasma pneumoniae and Mycoplasma genitalium, *Nucleic Acids Res.*, 25:701–712, 1997.

[4] Higgins, D.G., Bleasby, A.J. and Fuchs, R., CLUSTAL V: improved software for multiple sequence alignment. *Computer Applications in the Biosciences (CABIOS)*, 8(2):189–191, 1992.