Analysis of Translation Initiation Sites in the Complete Genome of *Mycoplasma genitalium* Suggests That This Bacterium Uses Alternative Signal for Translation Initiation

Rintaro Saito [12] **Masaru Tomita** [13]
`rsaito@mag.keio.ac.jp` `mt@sfc.keio.ac.jp`

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

1 Introduction

Although the Shine-Dalgarno (SD) sequence [5] is widely accepted as the signal sequence for ribosome-mRNA binding in procaryotic translation initiation, this sequence is not well conserved among genes within or among species. This is especially striking in *Mycoplasma genitalium* [2], where no obvious conserved sequence can be observed upstream of start codons[1]. To investigate the possibility of an alternative sequence pattern to account for the lack of SD sequences, we have conducted computer analyses of translation initiation sites (ORF start sites) in the complete genome sequence of *M. genitalium* [3].

2 Materials and Methods

Frequencies of all 64 triplets in within 500 bases upstream from start codons are computed. Then standardized frequencies (S_f) of each triplet at each position with respect to start codons were calculated according to the following formula:

$$S_f = \frac{n_o - np}{\sqrt{np(1-p)}} \sim N(0,1),$$

where n is the number of sequences to analyze, p is frequency of specific triplets that appear in the sequences, and n_o is the number of the specific triplets that were observed at specific position.

3 Results and Discussion

Triplets with significant frequencies are listed in Fig. 1(a). Frequency of triplet "TAA" was the most significant one. We have plotted frequency of "TAA" around start codons. An outstanding peak in the frequency of the triplet "TAA" was observed between positions -27 to -4 (Fig. 1(b)). The similar peak was still observed even when stop codons (TAA) of the preceding genes were eliminated from our analysis by only counting translation initiation sites with their preceding genes oppositely oriented (head-on genes) (Fig. 1(c)). The fact that the complementary triplet "TTA" does not exist in the 3'-terminus of 16S rRNA of *M. genitalium*[1] leads us to suspect the existence of an alternative mechanism for translation initiation in *M. genitalium* [4].

[1] The 3'-terminal sequence of *M. genitalium* 16S rRNA is "gggggtggatcacctc". Thus the predicted ribosome binding sequence would be "gaggtgatccaccccc".

(a)

Position	S_f	Triplet	Position	S_f	Triplet
−6	9.89	taa	−19	5.47	tta
−7	6.38	tta	−11	5.29	aaa
−12	5.90	taa	−15	5.19	taa
−6	5.87	ata	−4	5.17	agt
−3	5.59	gta			

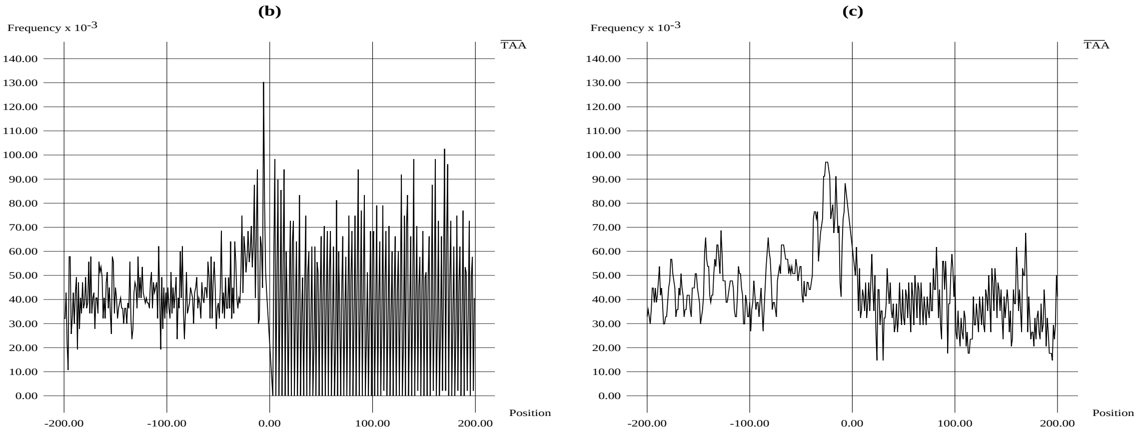

Figure 1: (a) Significant triplets that appears upstream of start codons and their positions and standardized frequencies. (b) Frequency of TAA triplets around start codons in *M. genitalium*. (c) Frequency of TAA triplets around start codons of head-on genes (Smoothed).

Acknowledgements

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

References

[1] Dalphin M.E. et al., The translational signal database, TransTerm, is now a relational database, *Nuc. Acid. Res.*, 26:335–337, 1998.

[2] Dybvig, K. and Voelker, L., Molecular Biology of Mycoplasmas, *Annu. Rev. Microbiol.*, 50:25–57,1996.

[3] Fraser, C.M. et al., The minimal gene complement of *Mycoplasma genitalium*, *Science*, 270:397–403, 1995.

[4] Loechel, S., Inamine, J.M., and Hu, P.-C., A novel translation initiation region from *Mycoplasma genitalium* that functions in *Escherichia coli*, *Nuc. Acid. Res.*, 19:6905–6911, 1991.

[5] Shine, J. and Dalgarno, L., The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: Complementarity to nonsense triplets and ribosome binding sites, *Proc. Natl. Acad. Sci. USA*, 71:1342–1346, 1974.