

HGT and innovation of Genome Systems Complexity: Where do the selective pressures originate from?

Torbjörn Karfunkel^{1,*}, Devdatt Dubhashi¹, Malte Hermansson²

1. Department of Computing Science, Chalmers University of Technology, Gothenburg, Sweden

2. Department of Cell and Molecular Biology, Göteborg University, Gothenburg, Sweden

*E-mail: yehoshua@cs.chalmers.se

Introduction

The importance of Horizontal Gene Transfer (HGT) in the innovation of early genomic complexity has come to be increasingly appreciated. Goldenfeld and Woese [5] suggest that “gradual refinement through the horizontal sharing of genetic innovation would have led to the generation of a combinatorial explosion of genetic novelty”. In [12], it is argued that HGT may also be responsible for the universality and optimality of genetic codes. Going from the ancient to the modern, HGT is again poised to play a major role in the new field of *synthetic biology*. As observed in [6] and elsewhere, an important distinction between the engineering in synthetic biology and other engineering disciplines is that the products of this engineering are evolving. Thus, it is crucial to understand the mechanisms and selective forces involved in HGT.

Back-amelioration and origin of selective pressures

Genes transferred to a new environment will no longer be in a steady state, and will therefore start changing towards a new steady state, a process called *amelioration*. Genes in the process of amelioration will display characteristics intermediate to those of their original and their new environment. In order to use a mathematical model of amelioration, it is necessary to know the direction of amelioration. In some cases this is not at all trivial. When a gene enters a plasmid by transposition, and the plasmid is introduced into a bacterium by conjugation, the question is towards which genome will the gene ameliorate? The answer to this question depends on which mechanisms are responsible for the selective pressures that the gene experiences. It is not clear which these mechanisms are. Several possibilities have been discussed [2, 4, 7].

A model of amelioration, based entirely on codon-position-specific GC content was developed by Lawrence and Ochman [9]. We applied their *back-amelioration* technique twice for each coding region investigated, assuming amelioration towards the plasmid and towards the host organism, respectively. The assumption indicating an origin with the least deviation from the GC content distribution calculated by Muto and Osawa [10] was taken to be the better explanation (Figure 1).

Results and discussion

Looking at coding regions situated on transposable elements in 22 different plasmids from 18 different hosts, we found 33 cases where amelioration towards the host organism is more likely, and 54 cases better explained by amelioration towards the plasmid. Using the null

hypothesis that the direction of amelioration is in neither direction (giving equal probability for both outcomes), and applying a two-sided binomial test, the null hypothesis is rejected in favor of the hypothesis that amelioration goes towards the plasmid ($P \approx 0.0157$).

A possible explanation is that amelioration is shaping the genes in accordance with needs for secondary DNA structure. The potential to form stem-loops is widespread [1, 2, 3]. Stem-loops are known to be important for the control of copy number of the ColE1 plasmid [8, 11].

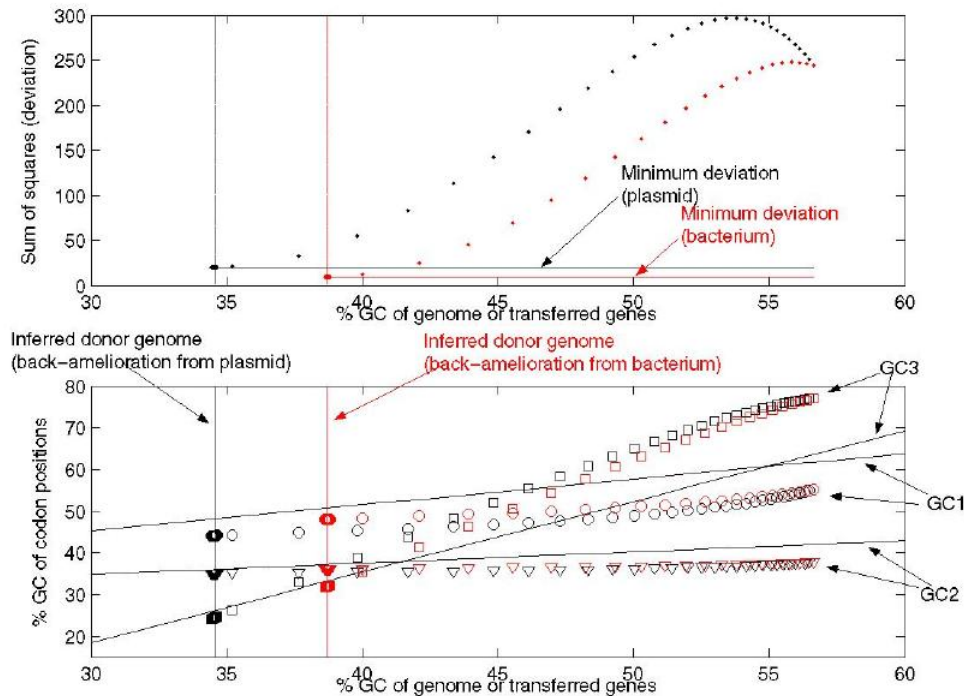


Figure 1: Top: Sum of squares deviation of a gene, from the Muto-Osawa GC lines [10], during back-amelioration. Bottom: GC content of the three codon positions.

References

- [1] I. H. Barrette, S. McKenna, D. R. Taylor, and D. R. Forsdyke "Introns resolve the conflict between base order-dependent stem-loop potential and the encoding of RNA or protein", *Gene* vol. **270**, pp. 181–189, 2001.
- [2] D. R. Forsdyke "A stem-loop "kissing" model for the initiation of recombination and the origin of introns", *Mol. Biol. Evol.* vol. **12**, pp. 949–958, 1995.
- [3] D. R. Forsdyke "Stem-loop potential in MHC genes: a new way of evaluating positive darwinian selection?", *Immunogenetics* vol. **43**, pp. 182–189, 1996.

- [4] D. R. Forsdyke and J. R. Mortimer “Chargaff’s legacy”, *Gene* vol. **261**, pp. 127–137, 2000.
- [5] N. Goldenfeld and C. Woese “Biology’s next revolution”, *Nature* vol. **445**, p. 369, 2007.
- [6] M. Heinemann and S. Panke, “Synthetic biology: putting engineering into biology”, *Bioinformatics* vol. **22**, no. **22**, pp. 2790–2799, 2006.
- [7] T. Ikemura, “Codon usage and tRNA content in unicellular and multicellular organisms”, *Mol. Biol. Evol.* vol. **2**, no. **1**, pp. 13–34, 1985.
- [8] F. A. Kolb, E. Westhof, C. Ehresmann, B. Ehresmann, E. G. H. Wagner, and P. Romby “Bulged residues promote the progression of a loop-loop interaction to a stable and inhibitory antisense-target RNA complex”, *Nucleic Acids Res.* vol. **29**, pp. 3145–3153, 2001.
- [9] J. Lawrence and H. Ochman “Amelioration of bacterial genomes: rates of change and exchange”, *J. Mol. Evol.* vol. **44**, pp. 383–397, 1997.
- [10] A. Muto and S. Osawa “The guanine and cytosine content of genomic DNA and bacterial evolution”, *PNAS* vol. **84**, pp. 166–169, 1987.
- [11] J. Tomizawa “Control of ColE1 plasmid replication: the process of binding of RNA I to the primer transcript”, *Cell* vol. **38**, pp. 861–870, 1984.
- [12] K. Vetsigian, C. Woese, and N. Goldenfeld, “Collective evolution and the genetic code”, *PNAS*, vol. **103**, no. **28**, pp. 10696–10701, 2006.