

Assessing the impact of horizontal gene transfer on the evolution of prokaryotes

Makarenkov, V.

*Department of Computer Science, Université du Québec à Montréal,
Case postale 8888, succursale Centre-ville,
Montreal (Quebec) H3C 3P8 Canada, email: makarenkov.vladimir@uqam.ca*

Badescu, D.

*McGill University, Faculty of Medicine, 1010 Sherbrooke Street West,
Montreal (Quebec) H3A 2R7 Canada, email: dunarel@mcgill.ca*

Horizontal Gene Transfer (HGT) is one of the major evolutionary processes affecting the evolution of prokaryotic species [1, 2]. Two known types of horizontal gene transfer are complete and partial transfers. Partial HGT can be viewed as a complete HGT followed by intragenic recombination and leading to the creation of a mosaic gene. The identification of the origins and the rates of horizontal gene transfers in the context of complete and partial HGT models, and for different phylogenetic families and ecological niches, is a very relevant and challenging problem. HGT has an important impact on microbial cooperation and bacterial virulence [3, 4]. There exist three well-known HGT mechanisms, i.e., transformation, transduction and conjugation, which allow DNA sequence acquisition either from the environment or directly from the donor species [5]. High prevalence of HGT in prokaryotes has been demonstrated by the discovery of pathogenicity islands and virulence attributes [6, 7]. The latter events are relatively recent, and have a clear ecological component associated with maintenance, expansion or change of microorganism's ecological niche [8]. Furthermore, bacteriophages, as gene transfer agents, stand as another compelling evidence of recent HGT [6].

We will present a novel bioinformatics framework designed to estimate and compare the rates of complete and partial HGT at different phylogenetic and ecological levels. Well-known methods of phylogenetic tree inference (e.g. RAxML [9]) and horizontal gene transfer detection (e.g. HGT-Detection [5]) will be used in our experiments. We support a “genome space” view of prokaryotic evolution, in which individual strains interact based on their

ecological niches and phylogenetic similarity. Our results suggest that partial HGTs are almost twice more frequent than the complete ones. Moreover, we show that partial HGTs, detected by the contemporary HGT detection algorithms, seem to be more recent than complete HGTs.

We estimated the probability that a contemporary prokaryotic allele or one of its direct ancestors (i.e. species located on the allele's lineage) have been ever affected by an HGT coming from another prokaryotic organism during its evolutionary history. We found that the majority of the existing prokaryotic alleles have not been affected by HGT. On the contrary, the majority of prokaryotic genes (i.e. a gene was represented by a multiple alignment of the corresponding alleles) have been affected multiple times by gene transfers during its evolutionary history: we found that 83% of the considered prokaryotic genes have been affected by at least one complete HGT and 96% - by at least one complete or partial. The comparison between complete and partial HGTs also highlights the fact that the ages of partial HGTs, which are more recent than complete transfers, can be detected with a better confidence.

Let us present the ten most frequent horizontal gene transfers among phylogenetic families for each selected bootstrap threshold of the HGT-Detection algorithm [5]. The most significant transfers are mapped into the phylogenetic tree of 111 prokaryotic species (see Figures 1). Circular tree views were selected for this presentation. We put together all of the 10 most significant transfers obtained for the 50%, 75% and 90% bootstrap thresholds. This resulted in 18 distinct transfers for the complete HGT. The obtained results confirm that the intragroup HGTs are very important for the process of the prokaryotic diversification. A majority of the highly-ranked intragroup HGTs were also ranked among the ten most frequent HGTs in general. The only notable exception is Actinobacteria.

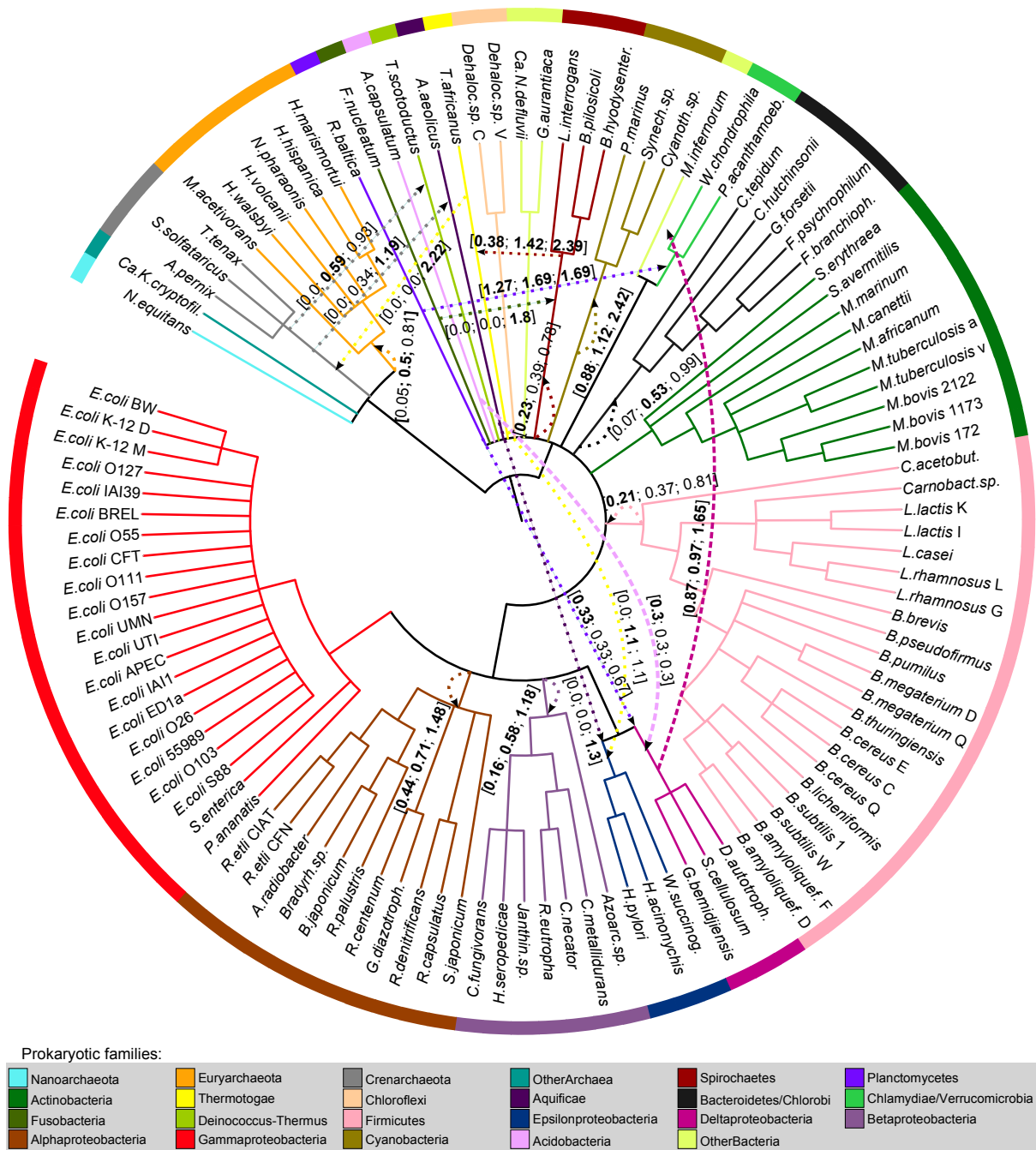


Figure 1. Phylogenetic network inferred for 111 prokaryotic species belonging to 23 different prokaryotic families, and including 18 most significant complete HGTs.

Here, the HGT rate is given for each of the three following HGT bootstrap confidence levels: 90%, 75% and 50%. Interval format is: [90%, 75%, 50%]. Arrows are colored according to the HGT source group. Values are boldfaced when they belong to the top ten list of the corresponding bootstrap confidence level.

References

1. Koonin E. V., Makarova K. S. and Aravind, L. (2001), Horizontal gene transfer in prokaryotes: quantification and classification, *Annual Reviews in Microbiology* **55**:709–742.
2. Wolf Y. I., Makarova K. S., Yutin N. and Koonin E. V. (2012), Updated clusters of orthologous genes for archaea: a complex ancestor of the archaea and the byways of horizontal gene transfer, *Biology Direct* **7**:46.
3. Nogueira T., Rankin D. J., Touchon M., Taddei F., Brown S. P. and Rocha E. P. (2009), Horizontal gene transfer of the secretome drives the evolution of bacterial cooperation and virulence, *Current Biology* **19**:1683–1691.
4. Takeuchi N., Kaneko K. and Koonin E. V. (2014), Horizontal gene transfer can rescue prokaryotes from Muller's ratchet: Benefit of dna from dead cells and population subdivision, *G3: Genes/ Genomes/ Genetics* **4**:325–339.
- Boc A., Philippe H. and Makarenkov V. (2010), Inferring and validating horizontal gene transfer events using bipartition dissimilarity, *Systematic Biology* **59**:195–211.
6. Koonin E. V. and Wolf Y. I. (2008), Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world, *Nucleic Acids Research* **36**:6688–6719.
7. Ochman H., Lawrence J. G. and Groisman E. A. (2000), Lateral gene transfer and the nature of bacterial innovation, *Nature* **405**:299–304.
8. Smillie C. S., Smith M. B., Friedman J., Cordero O. X., David L. A. and Alm E. J. (2011), Ecology drives a global network of gene exchange connecting the human microbiome, *Nature* **480**:241–244.
9. Stamatakis A., Ludwig T. and Meier H. (2005), RAxML-iii: a fast program for maximum likelihood-based inference of large phylogenetic trees, *Bioinformatics* **21**:456–463.