



ABCDE



Scientific Report

First name / Family name

Murray PATTERSON

Nationality

Canadian

Name of the *Host Organisation*

INRIA

First Name / family name
of the *Scientific Coordinator*

Eric TANNIER

Period of the fellowship

01/02/2012 to 31/01/2013



I – SCIENTIFIC ACTIVITY DURING YOUR FELLOWSHIP

One of the goals stated in the ABCDE Research Training Programme for this period of the fellowship was to extend the algorithm of [2] to include horizontal gene transfers as detected in [1]. We now have a working implementation of this extended algorithm, which shows promising results when applied to datasets of cyanobacterial genome evolution (that involve horizontal gene transfer) generated by several methods, including that of [3]. This method is the first to reconstruct ancestral gene structures, explicitly handling horizontal transfers, which are very numerous in all unicellular organisms. This is important, since unicellular organisms represent more than 90% of the historical biodiversity on earth, and are suspected to allow for rapid evolution in microbes, in particular, resistance to antibiotics, so it is essential to understanding the evolution of life. The next immediate step is to apply this method to the many other datasets (cyanobacterial, and beyond) in order to better understand how genes are co-evolving in the presence of horizontal gene transfer.

After using the method above to reconstruct ancestral genome structures, given a dataset of genome evolution as inferred by a method such as that of [3], one way to assess the quality of this dataset is by how close the ancestral genome structure is to being linear (since it is now well-known that genes are linearly arranged, as a set of chromosomes, to make up a genome). This possibility to validate results in evolutionary studies is rare, because of course we have no access to the real evolutionary history (the ancestral genomes). Of course, because of many sources of error, these ancestral structures are rarely completely linear, and so one natural approach to this problem is to try to remove (or disregard) as few ancestral adjacencies as possible from the structure so that the remaining structure is linear. While there already exists a large body of work on this subject, mostly in the form of heuristics and optimization methods, for this (computationally hard) problem, we found an algorithm for a case of this problem that suits perfectly the case when the reconstructed ancestral genome is bacterial with plasmids. To our knowledge, this is the first method to linearize an ancestral genome structure to collection of chromosomes and plasmids, and moreover, it is an exact algorithm, rather than an optimization method or a heuristic. This result was published (cf. II - Publication(s) ...), and the application of this method on a *Yersinia Pestis* dataset (descendants of the bubonic plague) produced a good quality linearization.

References

- [1] J.P. Doyon, C. Scornavacca, K.Y. Gorbunov, G.J. Szollosi, V. Ranwez and V. Berry. An efficient algorithm for gene/species tree parsimonious reconciliation with losses, duplications and transfers. In Proceedings of *RECOMB Comparative Genomics* 2010.
- [2] S. Berard, C. Gallien, B. Boussau, G.J. Szollosi, V. Daubin and E. Tannier. Evolution of gene neighborhoods within reconciled phylogenies. *Bioinformatics* **28**(18):382-388, 2012.
- [3] G.J. Szollosi, B. Boussau, S. Abby, E. Tannier and V. Daubin. Phylogenetic modeling of lateral gene transfer reconstructs the pattern and relative timing of speciations. *PNAS* 2012, doi : 10.1073/pnas.1202997109



II – PUBLICATION(S) DURING YOUR FELLOWSHIP

J. Manuch, M. Patterson, R. Wittler, C. Chauve and E. Tannier. Linearization of ancestral multichromosomal genomes. In Proceedings of *RECOMB Comparative Genomics* 2012, BMC Bioinformatics **13**(19):S11 (2012).

III – ATTENDED SEMINARS, WORKHOPS, CONFERENCES

Combinatorial Pattern Matching (CPM), Helsinki, Finland, 2012
<<http://cpm2012.helsinki.fi/>>

RECOMB Comparative Genomics, Niteroi, Brazil, 2012 <<http://www.uff.br/recombcg/>>, where I presented the publication (cf. II - Publication(s) ...)

GDRE Comparative Genomics, Lyon, France, 2012 <<http://gdrera.univ-lyon1.fr/>>

Journées GTGC (Groupe de Travail sur Génomique Comparative), Lille, France, 2012, where I presented some material on the linearization method (cf. II – Publication(s) ...)

IV – RESEARCH EXCHANGE PROGRAMME (REP)

N/A