Detection of genetic transfers by phylogenetics : limits and interest of Likelihood-Based Tests of Topologies

Yoan Radondy, Nicolas Beaume and Catherine Dauga

Génopole - Institut Pasteur – 28, rue du Docteur Roux - 75015 PARIS, France
{yradondy,cdauga}@pasteur.fr

Abstract: *Horizontal gene transfers (HGT), potentially followed by recombination or replacement of resident homolog, is now recognized as a major force shaping evolutionary histories of prokaryotes. We proposed to identify HGT by using a pipeline including 4 Likelihood-Based tests of topologies to detect unusual evolutionary patterns of gene phylogenies. We performed simulations in order to assess the limits of tests on a large range of evolutionary situations. We highlighted that tree topologies, model complexity, as information level between sequences influenced the results of tests. At last, we pointed out the importance to have a strong phylogenetic signal in the reference topology to obtain a reliable detection of genetic transfers. Then, we showed that tests of topology, used with some caution, were powerful tools to detect horizontal transfers within two bacterial species, pathogens for humans.*

Keywords:. Phylogeny, Gene transfer, Likelihood-based tests of topology, *Chlamydia*, *Helicobacter.*

Horizontal (or lateral) gene transfers (HGT), potentially followed by recombination or replacement of resident homolog (orthologous replacement), is now recognized as a major force shaping evolutionary histories of prokaryotes [1] and eukaryotes [2].

Among methods for detecting instances of HGT, atypical nucleotide composition methods were interesting to detect recent HGT between distantly related species. Such methods failed when HGT involved genomes of similar GC contents, e.g. genomes of strains within species and, when gene transfers were ancient because acquired genes "ameliorate" progressively to have the compositional characteristics of their environment. We propose an alternative method based on the detection of unusual evolutionary patterns of gene phylogenies to identify HGT in all circumstances.

We developed a pipeline to compare tree topologies with different statistical tests based on likelihood between two sets of homologous genes. Our pipeline took in reference a first alignment, estimated the evolutionary model of data, build a Maximum Likelihood phylogenetic tree. This referent topology was compared with the tree topology obtained by the same process on a second alignment in which genes acquired by horizontal transfers are searched. The pipeline included four likelihood-based tests, the KH (Kishino & Hasegawa) test [3], the SH (Shimodarai & Hasegawa) test [4], the SOWH test [5] and the ELW (Expected Likelihood Weight) test [6].

To test our procedure, we analysed 2 sets of sequences coming from biological studies in which evidence of genetic transfers was highlighted by two distinct methods. The first set of sequences consisted in ribosomal protein genes of bacteria from different lineages [7]. The second set of

sequences concerned genes from a photosynthetic operon shared between five bacteria of the alpha branch of Proteobacteria [8]. In addition, we attempted to determine the limits of topology tests on data sets obtained by simulations using different evolutionary models and topologies found in real data. To measure the effectiveness of topology tests, we carried out our procedure on data sets obtained by simulation in using 29 evolutionary models. Three topologies corresponding to different phylodynamics were studied : a balanced tree, a tree with a caterpillar clade, and a tree including a long branch. Then, we applied our method to detect genetic transfer within two species, Helicobacter pylori [9] and Chlamydia trachomatis [10][11], both pathogens for humans.

The 4 tests of topologies showed different power for detecting HGT. The KH and SH tests yielded similar results on our biological and simulated data sets, although the KH test was known as inappropriate to test hypotheses a posteriori. Both KH and SH tests failed to detect transfer in several situations and seemed to be the more conservative tests. The SOWH test produced the highest rate of false positives. In addition to the impact of models on its results, we observed an effect of the tree building algorithms used (PAUP vs Phyml). Among tests studied, the ELW test showed the optimal sensitivity and specificity.

Furthermore, complexity of evolutionary models, tree topologies and phylogenetic information of data disrupted results of tests. Differences in model complexity between tested genes and referent genes had significant consequences on test results. If the evolutionary model of genes tested was more complex than those of referents, tests lacked sensitivity and HGT were not detected. It was also difficult to detect a gene acquired by transfer when it belongs to a phylogenetic tree with a long branch. At last, when topologies used in reference were based on low phylogenetic information, gene transfers could be wrongly detected with some tests.

Then, we performed the tests of topology on data from Chlamydia trachomatis, an intracellular living species that exhibits a low rate of genetic transfers and, from Helicobacter pylori, a species known for its very high level of genetic exchanges. In both cases, most of HGT previously described in literature were found. As expected, the use of a reference topology based on intergenic sequences instead of housekeeping genes improved the detection of genes acquired by transfers.

Acknowledgements

This work is supported by “Le Réseau National Genopole”.

References

[1] E.V. Koonin, K.S. Makarova, and L. Aravind. *Horizontal gene transfer in prokaryote : quantification and classification*. Ann Rev Microbiol., 55: 709-742, 2001.

[2] M. Mitreva, M.L. Blaxter, D.M. Bird and J.P. McCarter. *Comparative genomics of nematodes*. Trends Genet. 21: 573-581, 2005.

[3] H. Kishino, M. Hasegawa, *Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea*, J Mol Evol, 29(2):170-9, 1989.

[4] H. Shimodaira, M. Hasegawa. *Multiple comparisons of log-likelihoods with applications to phylogenetic inference*, Mol Biol Evol 16 (8): 1114-1116 1999

[5] N. Goldman, JP. Anderson, AG. Rodrigo, *Likelihood-based tests of topologies in phylogenetics*. Syst Biol, Dec;49(4):652-70, 2000.

[6] K. Strimmer, A. Rambaut, *Inferring confidence sets of possibly misspecified gene tree*,. Proc Biol Sci, 269(1487): 137–142, 2002.

[7] C. Brochier, H. Philippe, D. Moreira, *The evolutionary history of ribosomal protein RpS14: horizontal gene transfer at the heart of the ribosome*, Trends Genet, 16(12):529-33, 2000.

[8] N. Igarashi, J. Harada, S. Nagashima, K. Matsuura, K. Shimada, KV. Nagashima, *Horizontal transfer of the photosynthesis gene cluster and operon rearrangement in purple bacteria*, J Mol Evol, 52(4):333-41, 2001.

[9] SM. Devi, I. Ahmed, A.A. Khan, S.A. Rahman, A. Alvi, L.A. Sechi, N. Ahmed, *Genomes of Helicobacter*

pylori from native Peruvians suggest admixture of ancestral and modern lineages and reveal a western type cag-pathogenicity island, BMC Genomics 7(1), 197, 2006.

[10] J.P. Gomes, W.J. Bruno,A. Nunesl, N. Santos, C. Florindo, M.J. Borrego, D Dean, *Evolution of Chlamydia trachomatis diversity occurs by widespread interstrain recombination involving hotspots*, 17 (1): 50-60, Genome Res,2007.

[11] BW. Brunelle, GF. Sensabaugh, *The ompA gene in Chlamydia trachomatis differs in phylogeny and rate of evolution from other regions of the genome*, Infect Immun, 74(1):578-85, 2006.