

Détection des transferts de gènes, domaine d'application et limites de l'approche phylogénétique.

Yoan Radondy, Eulalie Pinto, Nicolas Beaume et Catherine Dauga

Plateforme 4, Génopole, Institut Pasteur, 28 rue du Dr Roux, 75724 Paris cedex 15, France

Les transferts de gènes horizontaux suivis par une recombinaison ou un remplacement homologue sont reconnus aujourd'hui comme une force majeure de l'évolution des micro-organismes. Les méthodes de détection de transferts de gènes basées sur la recherche d'une composition en bases atypique des séquences sont intéressantes pour mettre en évidence des transferts récents et entre espèces éloignées, mais elles ne sont pas utilisables quand :

- le transfert de gènes a eu lieu entre des génomes de composition en GC similaire, au sein d'une même espèce par exemple ;
- ou encore suite à des transferts anciens quand la composition en GC du gène transféré s'est harmonisée à celui de son environnement par « amélioration » progressive.

Nous proposons une approche alternative de détection des transferts horizontaux basée sur la comparaison d'arbres phylogénétiques. L'un de ces arbres reconstitue l'histoire évolutive du groupe de micro-organismes étudié, c'est l'arbre référent. L'autre représente l'histoire évolutive d'un gène étudié pour le même groupe de micro-organismes. La discordance entre les 2 arbres suggère la possibilité d'un ou plusieurs transferts au cours de l'évolution.

Nous avons développé un protocole de comparaison automatique d'arbres phylogénétiques issus de deux jeux de séquences homologues. Ce protocole intègre des tests statistiques de comparaison des topologies basés sur la vraisemblance : le test de Kishino & Hasegawa, de Shimodara & Hasegawa, le test SOWH et le test ELW (Expected Likelihood Weight) afin de mettre en évidence l'éventuelle non congruence entre les deux topologies.

Ces 4 tests de comparaison de topologies ont été évalués sur des jeux de données provenant d'études biologiques et sur une série de simulations jouant sur la topologie et le modèle d'évolution. Ces simulations nous ont permis aussi de mesurer l'influence du niveau d'information sur le résultat des tests et d'optimiser notre méthode. L'outil développé, *PhyT*³ met en évidence la non congruence entre 2 arbres phylogénétiques. Il est couplé à une approche originale de construction de l'arbre référent et à un test systématique de détection des longues branches.

Nous illustrons l'intérêt de notre méthode pour la détection de transferts à l'échelle de l'espèce, avec l'étude de gènes d'*Helicobacter pylori* et de *Chlamydia trachomatis* et à l'échelle génomique par l'analyse de bactéries de la branche *Gamma* des *Proteobacteria*.

Keyword : HGT, phylogeny, Likelihood-based Tree Topology Tests, *Helicobacter*, *Chlamydia*, *Proteobacteria*.