

Plongements d'arbres dans des graphes orientés acycliques

Encadrants

V. Berry, LIRMM, Montpellier, vberry@lirmm.fr, 06 07 72 86 36, <https://www.lirmm.fr/~vberry/>

C. Scornavacca, ISEM, celine.scornavacca@umontpellier.fr, <https://sites.google.com/view/celinescornavacca>

Mots clefs : graphes, arbres, algorithmique combinatoire, probabilités.

Sujet du stage :

Le plongement d'arbres dans des DAG (*directed acyclic graphs*) permet de caractériser l'origine ancestrale des parties de génomes d'individus actuels. Cette problématique se retrouve par exemple pour les populations humaines mais aussi pour des plantes d'intérêt agronomiques comme le riz, les citrons, etc. Un arbre plongé représente l'évolution d'une partie homologue du génome de plusieurs individus et le DAG représente l'évolution des populations. Le plongement d'un arbre dans le DAG permet de savoir quels individus actuels descendent de telle ou telle population ancestrale (en fonction des arêtes du DAG par lesquelles le plongement passe). En associant une couleur à un petit nombre de populations ancestrales concurrentes, ce plongement nous permet de colorier les génomes des individus actuels (*chromosome painting*).

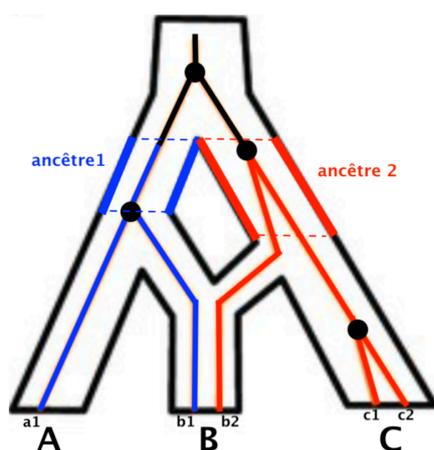


Figure : exemple de plongement d'un arbre dans un DAG. Les individus $b1$ et $b2$ de la population B n'ont pas le même ancêtre. Remarque : un autre plongement n'aboutit pas forcément aux mêmes conclusions.

On peut associer une probabilité à chaque plongement possible d'un arbre dans un DAG, cette probabilité est calculée analytiquement en fonction de paramètres associées aux arêtes du DAG (les largeurs et longueurs d'une arête représentent par exemple la taille de la populations associée et sa durée d'existence).

Dans le cadre de ce stage, on souhaite à la fois obtenir le plongement le plus probable d'un arbre dans un DAG, mais aussi une distribution des plongements les plus probables, ce qui permettra d'obtenir une mesure de confiance sur le fait que l'arbre emprunte ou pas telle ou telle arête du DAG, donc sur la couleur donnée à telle ou telle partie du génome d'un individu actuel.

On s'attachera aussi à optimiser le temps de calcul pour évaluer les plongements : quelles astuces algorithmiques est-il possible de mettre en œuvre pour réutiliser une partie d'un calcul précédent pour un nouveau plongement ? Est-il possible de passer d'un plongement à un plongement « voisin » en temps constant ? Un pré-calcul d'assignation de chaque partie de l'arbre à toute partie du DAG par programmation dynamique permet-il d'obtenir en temps polynomial la probabilité de tout plongement considéré ?

Santos J, Chebotarov D, McNally K, Bartholome J, Droc G, Billot C, Glaszmann J-C, *Genome Biology and Evolution*, **11** (5) : 1358-1373, 2019.

Berry V, Scornavacca C, Weller M, Proc. of *46th International Conference on Current Trends in Theory and Practice of Computer Science*, 2020.