

Mahendra Mariadassou (INRA)

Robustesse des arbres phylogénétiques, influences des sites et des taxons.

## Résumé

Les arbres phylogénétiques sont une bonne façon de représenter les liens de parenté entre espèces et sont couramment utilisés dans de nombreux domaines de la biologie: génomique comparative, épidémiologie, biologie de la conservation, etc.

Pour toutes ces applications, l'arbre reconstruit doit être le plus proche possible du vrai arbre. A erreur de reconstruction donnée, l'arbre doit en particulier être le plus robuste possible. Une source majeure de non robustesse est la contamination du jeu de données par des données aberrantes.

En conséquence, détecter et isoler les données aberrantes constitue une stratégie qui permet de robustifier l'arbre.

Nous montrons ici comment les fonctions d'influence empirique permettent de détecter les données influentes, susceptibles d'être aberrantes, et comment supprimer les données les plus exceptionnelles permet de reconstruire un arbre robuste. L'application à deux jeux de données (mammifères à placentas et zygomycètes) montre que la reconstruction par maximum de vraisemblance n'est pas robuste aux données aberrantes et que la suppression des quelques données les plus influentes améliore sensiblement la robustesse de l'arbre reconstruit.