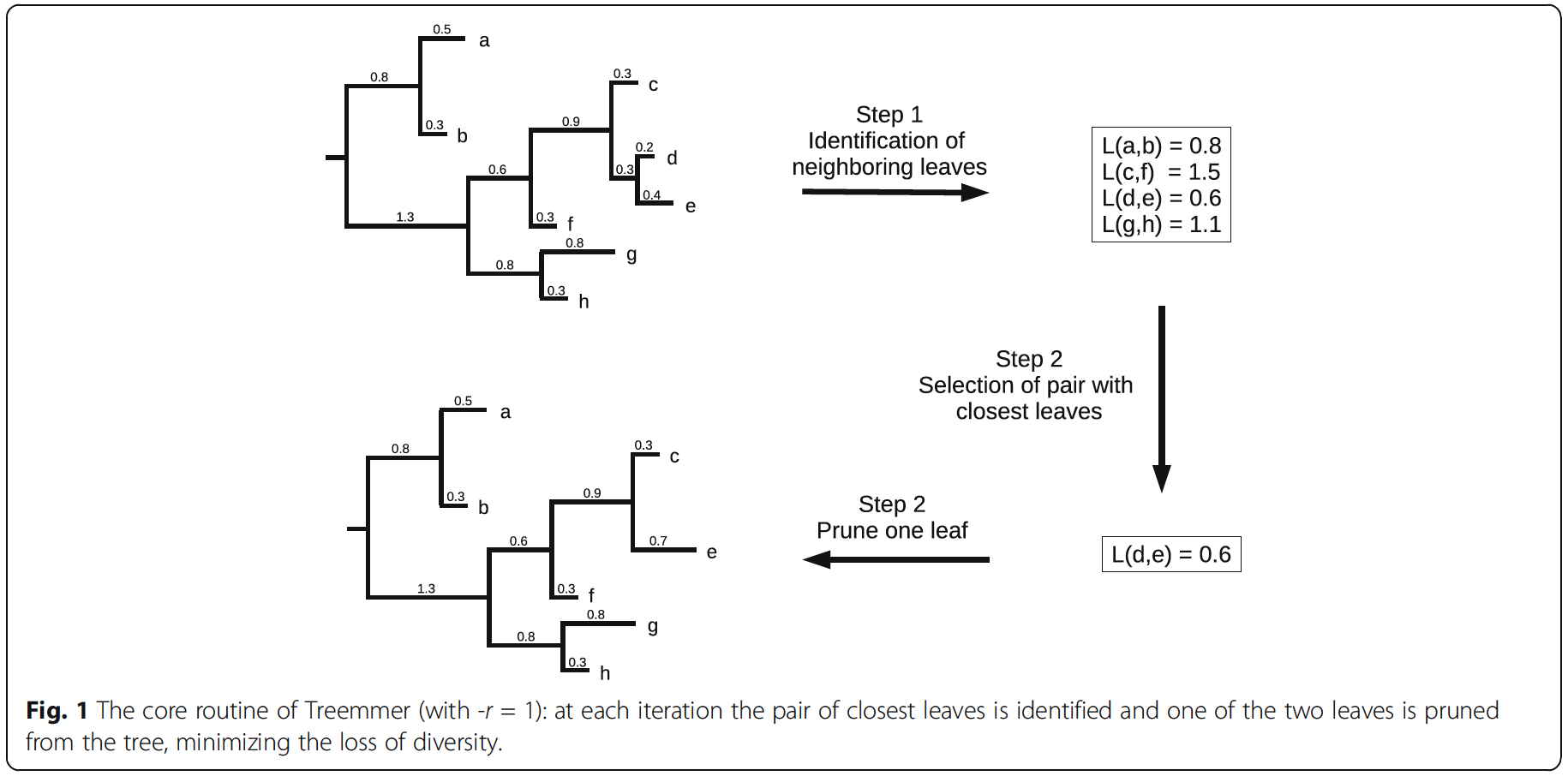
**Revue littérature pour simplifier les arbres**

**Méthode Treemer**

**Article : « Treemmer: a tool to reduce large phylogenetic datasets with minimal loss of diversity »**

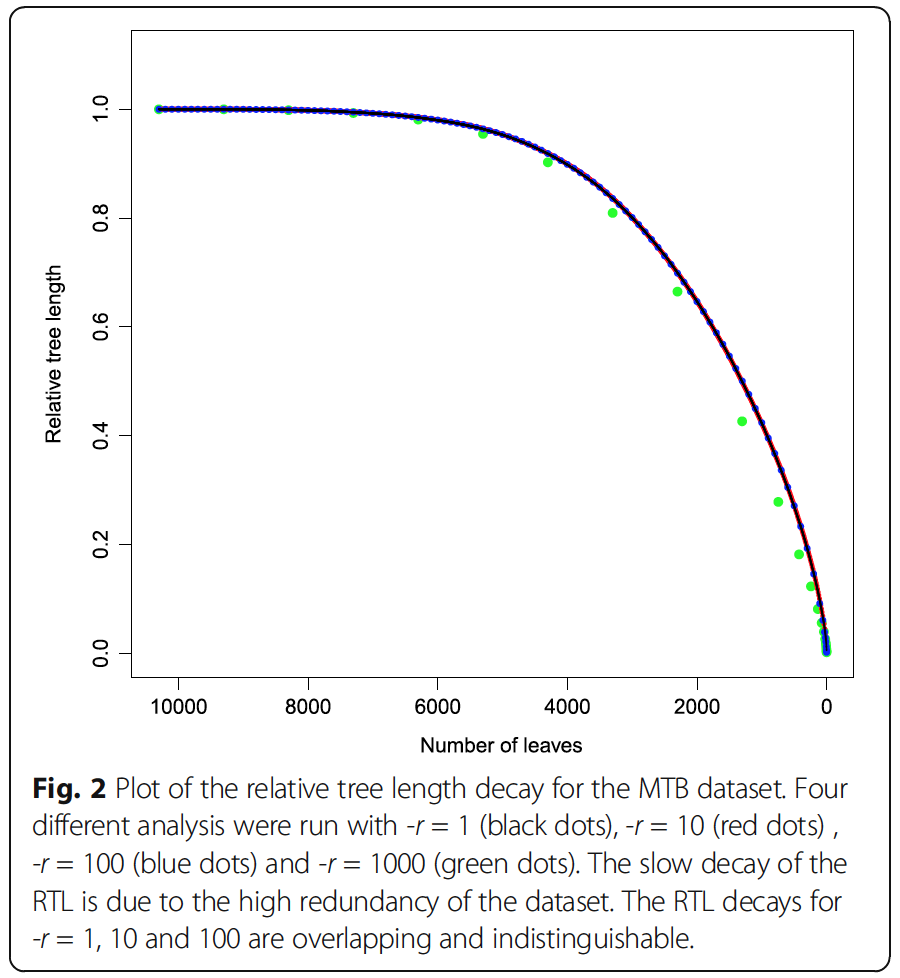
La méthode présentée dans l’article est la suivante (voir illustration de l’article ci-dessous) : pour chaque couple de feuilles voisines dans l’arbre, on détermine la distance séparant ces deux feuilles. On sélectionne ensuite le couple ayant la plus faible distance, et on supprime l’une des deux feuilles aléatoirement. On répète ensuite autant de fois que l’on veut cette deuxième étape, avant de répéter la première étape (calcul des distances), qui est l’étape computationnellement coûteuse.



Différentes options d’arrêt peuvent être envisagées :

-Fixation d’un nombre de feuilles final que l’on souhaite atteindre

-Fixation d’une jauge de RTL (Relative Tree Length) en dessous de laquelle on s’arrête : il s’agit d’une grandeur qui reflète la structure générale de l’arbre (voir figure ci-dessous)



On voit que même une division quasi par 2 du nombre de feuilles n’influe quasiment pas sur la structure générale de l’arbre : il est donc possible de grandement le simplifier sans influer sur la structure.

Adaptation à notre cas : RTL semble être un bon indicateur de structure de l’arbre, on pourrait (si ce n’est pas trop long computationnellement) calculer l’impact sur la RTL de la suppression de chacun des nœuds, et supprimer celui pour lequel cet impact est le plus faible. Par ailleurs, on peut fixer une jauge d’abondance à partir de laquelle on accepte de supprimer un nœud (on évitera par exemple de supprimer un nœud dont l’abondance est supérieure à x %)

On peut aussi combiner la notion de distance vue dans l’article, à celle d’abondance ainsi qu’à celle du nombre de descendants.

**Tree Pruner**

**Article : « Tree pruner: An efficient tool for selecting data from a biased genetic database »**

Cet article présente un logiciel de traitement manuel d’arbres phylogénétiques grâce auquel un utilisateur peut sélectionner les séquences qu’il veut conserver, ou au contraire celles qu’il veut voir disparaître (permet à un utilisateur expert de faire disparaître des biais de PCR qu’il observe dans les alignements de séquences, cf discussion du 09/04 sur les cas présentés par Marine Armand)

Inconvénients de la méthode à mon sens :

-utilisateur peut biaiser résultats en ‘faisant dire’ à l’arbre ce qu’il veut, et non la réalité

-Logiciel programmé en Java : nécessité de le réécrire sur python, mais on peut simplement s’inspirer des méthodes utilisées

**Tree Trimmer**

**Article : « Treetrimmer: a method for phylogenetic dataset size reduction »**

Cet article présente une méthode de simplification d’arbre phylogénétique basée sur l’utilisation d’Operationnal Taxonomic Unit (OTU) rentrées par l’utilisateur grâce à un fichier de support, en même temps que l’arbre à simplifier. Le principe est de ne garder, pour chaque OTU, qu’un nombre limitant de représentants. Ceci permet de simplifier l’arbre sans altérer sa structure générale de séparation des différentes OTU

Inconvénients :

-Nécessité de définir des OTU au préalable : Time consuming et relativement arbitraire

-Codé en Ruby