

Introduction à la Phylogénie moléculaire

M. Mariadassou & A. Bar-Hen

Université Paris Descartes, Paris 5, MAP5
45 rue des Saints Pères, 75270 Paris cedex 06

Résumé Cet article est une introduction au domaine de la phylogénie moléculaire et en particulier à la robustesse des arbres phylogénétiques. Nous commençons par une brève présentation historique du domaine avant de passer en revue les méthodes de reconstruction les plus populaires. Nous nous intéressons tout particulièrement à la méthode du maximum de vraisemblance. Cette méthode nécessite de construire un modèle probabiliste d'évolution des macromolécules biologiques mais fournit en contrepartie un cadre statistique propice à quantifier la variabilité de l'arbre estimé. Nous présentons tout d'abord les modèles d'évolution couramment utilisés, puis le calcul de la vraisemblance avant de montrer que la nature discrète de l'arbre rend caducs les outils traditionnels d'étude de la variabilité.

1 Le contexte de la phylogénie moléculaire

1.1 Origines du domaine

Les travaux précurseurs de Charles Darwin [4], sur lesquels a été bâtie la biologie évolutive moderne, ont radicalement changé notre compréhension de l'évolution. Darwin introduit dans son livre *De l'Origine des Espèces* la théorie de l'évolution, selon laquelle les espèces évoluent au fil des générations grâce au processus de sélection naturelle et que la diversité du vivant est obtenue grâce à l'accumulation graduelle de différences dans les sous-populations d'une espèce.

L'évolution peut être considérée comme un processus de branchement dans lequel des sous-populations d'une espèce se transforment par accumulation de différences avant de se détacher de leur espèce-mère pour former une nouvelle espèce ou s'éteindre. L'image d'arbre évolutif illustre bien le concept d'évolution et la formation de nouvelles espèces à partir d'espèces déjà existantes. Les liens de parenté qui unissent un groupe d'espèces sont communément représentés sous la forme d'arbres évolutifs, appelés "arbres phylogénétiques" ou encore "phylogénies".

Toutes les méthodes de reconstruction d'arbres phylogénétiques sont basées sur la même idée intuitive : étant donné que l'évolution intervient par accumulation de différences, deux espèces qui ont divergé récemment sont plus "semblables" que deux espèces dont la divergence est plus ancienne. La similitude entre espèces était mesurée par des critères de types morphologiques (à l'instar de la forme des os, du nombre de pattes ou du nombre de dents) jusque dans les années 50. La découverte de la structure de l'ADN par Watson et Crick en 1953 [15] et surtout les capacités de séquençage et d'analyse des molécules macrobiologiques qui ont rapidement suivi ont considérablement changé la donne en remplaçant avantageusement l'objet d'étude. Au lieu d'établir des liens de parenté à partir de critères morphologiques, pour certains fortement soumis à l'appréciation de l'expérimentateur et dont le nombre est généralement faible, les phylogénéticiens peuvent désormais s'appuyer sur des données moléculaires : des séquences génétiques (d'ADN) ou protéiques