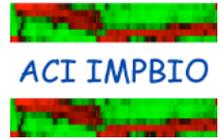


SuperTREE

(Projet de Recherche 2003)



Méthodes bioinformatiques pour l'analyse de super-matrices et la reconstruction de super-arbres, et étude de l'impact des données génomiques manquantes en phylogénie. [\[http://www.isem.univ-montp2.fr/PPP/PM/RES/Info/@Supertrees_ACI_IMPBio.php\]](http://www.isem.univ-montp2.fr/PPP/PM/RES/Info/@Supertrees_ACI_IMPBio.php)

CONTEXTE

L'explosion récente de la quantité de séquences d'ADN, ARN, et protéines disponibles chez une grande quantité d'organismes offre actuellement l'opportunité de mieux appréhender des questions complexes de biologie comparative, allant de la reconstruction de l'histoire évolutive des gènes et des espèces, jusqu'à l'identification des mécanismes qui en sont le moteur. Dans un nombre croissant d'études phylogénétiques, il apparaît cependant que la résolution claire d'un problème donné passe par la **combinaison** des informations apportées par de multiples segments génomiques, et dont le nombre peut excéder la centaine. Le présent projet se situe dans une problématique de combinaison de données, de manière directe sous forme de "**super-matrices**" de caractères (concaténation de gènes et de protéines), et de manière indirecte sous forme de "**super-arbres**" (phylogénies reconstruites à partir de lots d'espèces partiellement chevauchants).

OBJECTIFS

L'objectif du projet est la confrontation des deux principales approches de combinaison de données pour reconstruire "l'arbre du Vivant".

L'avantage des **super-arbres** réside dans la possibilité d'inclure tous les types de données (moléculaires, morphologiques, paléontologiques) et de constituer en quelque sorte un **bilan synthétique** des résultats de toutes les études sur lesquelles ils sont basés. En revanche, leur principal inconvénient réside dans l'embouteillage de deux étapes d'inférence phylogénétique, une pour construire les **arbres sources**, et l'autre pour reconstruire le super-arbre.

L'avantage des **super-matrices** réside quant à lui dans la possibilité de combiner directement — et de manière **synergique** — l'information phylogénétique apportée par chaque gène, et de calculer les indices de **fiabilité** des nœuds de l'arbre inféré. Les inconvénients sont en revanche des problèmes (i) de **temps de calcul** — conséquence du degré sans cesse grandissant du nombre de caractères et du nombre de taxons à analyser en une seule étape d'inférence d'arbre ; (ii) de combinaison sous des **modèles probabilistes** de gènes ou protéines aux comportements très variés, tant en termes de différence de composition en nucléotides ou acides aminés que de taux d'évolution et de processus mutationnels ; (iii) de gestion pour un certain nombre d'espèces de "trous", ou **données manquantes**, résultant du fait que certains gènes n'ont tout simplement pas été séquencés chez les espèces d'intérêt.

PROJET

- 1) Mesurer l'**impact des données manquantes** au sein des approches "multigéniques" et "multiprotéiques" sur l'exactitude et la précision des paramètres des modèles probabilistes ;
- 2) Proposer de **nouvelles méthodes** de construction de super-arbres à partir d'arbres sources ;
- 3) Visualiser **graphiquement** de très grandes phylogénies.

RESULTATS

- Impact modéré des données manquantes sur l'estimation des âges de divergence des eucaryotes par une approche bayésienne d'horloge moléculaire relâchée [1] ;
- Analyse par maximum de vraisemblance de supermatrices de caractères nucléotidiques et protéiques pour inférer la phylogénie des Rongeurs [3], des Primates (cf. Figure), des Animaux [5], et des Eucaryotes [2] ;
- Utilisation de grandes phylogénies pour l'estimation des taux de diversification des espèces au cours du temps en relation avec leurs traits d'histoire de vie [4] ;
- Reconstruction de super-arbres par compatibilité des arbres sources [6] ;
- Nouvelles fonctionnalités de l'éditeur d'arbres TreeDyn (F. Chevenet) ;
- SDM [*Super Distance Matrix*] : une nouvelle méthode de reconstruction de superarbres fondée sur une approche de distances [7].

CALENDRIER

2004	Janvier : Démarrage à Montpellier
2004	Avril : 1 ^{ère} table-ronde générale, avec visite d'H. Philippe (Montréal)
2005	Mars : 2 ^{ème} table-ronde générale
2005	Avril : visite E. Douzery à Montréal
2005	Automne : 3 ^{ème} table-ronde ; synthèse et perspectives.

LABORATOIRES IMPLIQUES

ISEM / UMR 5554 / Montpellier (Resp. proj. : Emmanuel Douzery)
LIRMM / UMR5506 / Montpellier (Resp. proj. : Vincent Berry)
IRD / UMR7503 / Montpellier (Resp. proj. : François Chevenet)
Laboratoire de Biochimie, Montréal, Québec, Canada
(Resp. proj. : Hervé Philippe).

MOTS-CLEFS

Phylogénie, arbres sources, super-arbres, super-matrices de distance, super-matrices de caractères, données manquantes, algorithmes d'inférence, simulations, validation biologique, eucaryotes, mammifères.

PRINCIPALES PUBLICATIONS

- [1] Douzery, E. J. P., E. A. Snell, E. Baptiste, F. Delsuc, and H. Philippe. 2004. The timing of eukaryotic evolution: Does a relaxed molecular clock reconcile proteins and fossils? *Proc. Natl. Acad. Sci. USA* 101:15386-15391.
- [2] Philippe, H., E. A. Snell, E. Baptiste, P. Lopez, P. W. H. Holland, and D. Casane. 2004. Phylogenomics of eukaryotes: impact of missing data on large alignments. *Mol. Biol. Evol.* 9:1740-1752.
- [3] Galewski, T., J.-F. Mauffrey, Y. L. R. Leite, J. L. Patton, and E. J. P. Douzery. 2005. Ecomorphological diversification among south american spiny rats (Rodentia; Echimyidae): a phylogenetic and chronological approach. *Mol. Phylogenet. Evol.* 34:601-615.
- [4] Paradis, E. 2005. Statistical analysis of diversification with species traits. *Evolution* 59:1-12.
- [5] Philippe, H., N. Lartillot, and H. Brinkmann. 2005. Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa and Protostomia. *Mol. Biol. Evol.*, 22:1246-1253.
- [6] Berry, V., and F. Nicolas. 2005. Maximum agreement and compatible supertrees.
- [7] Criscuolo, A., V. Berry, E. J. P. Douzery, and O. Gascuel. 2005. SDM: une méthode de distance rapide pour les études de phylogénomique. *JOBIM*.

UN SUPERARBRE DES PRIMATES

