La phylogénie moléculaire vise à reconstruire les relations de parentés évolutives entre les êtres vivants, par analyse comparée de leurs séquences génétiques respectives. Cette question peut être abordée dans un cadre probabiliste, en utilisant soit le principe du maximum de vraisemblance, soit l'inférence Bayésienne. Dans les deux cas, il est nécessaire de s'appuyer sur un modèle probabiliste. En pratique, on utilise généralement des modèles assez simples, postulant que le processus de substitution est Markovien, et qu'il est homogène au cours du temps, et le long des séquences. Toutefois, ces modèles ne sont pas satisfaisants: dans de nombreuses situations, ils engendrent des artefacts (erreurs systématiques), se traduisant par un positionnement statistiquement significatif, bien qu'erroné, d'espèces évoluant plus rapidement que la moyenne.

Je décrirai des méthodes semi-paramétriques permettant de relaxer à la fois ces deux hypothèses d'homogénéité, spatiale et temporelle. Dans la direction spatiale, les variations du processus de substitution le long de la séquence est modélisé par un processus de Dirichlet. Quant à la dimension temporelle, des variations le long des lignages sont introduites via un processus doublement stochastique. Les modèles ainsi obtenus ont été implémentés dans un cadre de Chaînes de Markov Monte Carlo, testés, et appliqués à des cas réels. Leur fit apparaît souvent meilleur, et leur robustesse est plus grande face aux artefacts phylogénétiques.