**Création d'un HMM pour prédire les promoteurs de type sigmaA de Bacillus subtilis**

Vous allez utiliser SHOW développé à l'INRA de Jouy en Josas pour construire votre HMM. Pour cela il faut prendre connaissance de sa syntaxe. Vous avez à votre disposition la [documentation](http://silico.biotoul.fr/enseignement/m1-mabs/BGPG/TD_HMM/show_doc.pdf) fournie avec le programme ainsi qu'un [extrait](http://silico.biotoul.fr/enseignement/m1-mabs/BGPG/TD_HMM/Show.pdf) commenté de cette syntaxe.  
Une fois la modélisation du HMM et son implémentation réalisée, il va falloir estimer les paramètres du modèle (probabilités d'émission et de transition). Pour cela un ensemble de séquence vous est fourni. Vous les trouverez dans l'[archive compressée](http://silico.biotoul.fr/enseignement/m1-mabs/BGPG/TD_HMM/seq.tar.gz) ci-jointe. Chaque séquence se trouve dans un fichier individuel ayant l'extension .dna.   
  
Vous ferez tourner l'étape d'estimation des probabilités de transition et d'observation de votre modèle et ensuite la prédiction en utilisant  l'algorithme de Baum-Welch puis l'algorithme de viterbi en utilisant SHOW que nous avons installé  sur notre [serveur](http://silico.biotoul.fr/cgi-bin/mobyle/portal.py#welcome) au travers de Mobyle. Pour cela, aller  dans annotation, puis dans genes.  
Pour estimer les probabilités choisir show\_emfit. Charger votre modèle. Vous obtiendrez plusieurs fichiers de résultats. Celui contenant votre modèle avec les nouvelles probabilités estimées se trouve dans la section Learn model. Dans further analysis vous pouvez directement lancer la prédiction sur vos séquences en lançant show\_viterbi.  
  
Les résulats du viterbi sont donnés sequence par sequence avec création pour chaque séquence d'un fichier .vit et sous un format à découvrir. Vous pouvez récupérer l'ensemble des fichiers sous forme d'archive en allant dans download en haut de la page.  
Pour pouvoir estimer le pouvoir prédictif du modèle, il va falloir parser ces fichiers.  
Le programme vous permettant de parser les résulats du viterbi vous est fourni [ici](http://silico.biotoul.fr/enseignement/m1-mabs/BGPG/TD_HMM/parse_viterbi.pl). Le code est à modifier pour donner le chemin correct de vos fichiers. Pour simplifier on pourra mettre les fichiers séquence .dna et les fichiers résultats .vit dans le même répertoire. Ce programme peut être largement amélioré et généralisé.