

**Module Analyse de Séquences sous Unix**  
**ESIL Biotechnologie 2<sup>ème</sup> année**  
**Mai 2005**

*Durée: 1 heures*

*Documents autorisés*

D. Gautheret

1. Vérifiez que vous avez bien ~5Mb libres sur votre compte. Récupérez sur le serveur ftp du NCBI <ftp.ncbi.nlm.nih.gov> le génome de Rickettsia conorii au format Genbank.
2. Combien de gènes contient ce génome ?
3. Quels sont les dinucléotides et trinucleotides les plus fréquents dans ce génome ?
4. Extraire les CDS. Quels sont les codons les plus fréquents dans les gènes protéiques ? (programme cusp)
5. Extraire les 50 nucléotides en amont de chaque CDS (extractfeat)
6. La séquence TATA est-elle sur-représentée ? de combien par rapport à la fréquence attendue? (programme compseq)