

Exercices avec le logiciel 

Épreuve aMIG - Contrôle terminal - 20 juin 2006

J.R. Lobry, L. Duret, G. Perrière

11 mars 2008

Génomes de *Gallus gallus* et *Mycobacterium tuberculosis*

Contrôle terminal - 20 juin 2006

(Durée : 3 heures)

Documents autorisés

Envoyez (avant la fin de l'épreuve, heure de réception du mél faisant foi!)
votre compte-rendu au format PDF à :

{lobry, duret, perriere}@biomserv.univ-lyon1.fr.

Les trois problèmes sont complètement indépendants.

1 Problème 1 (J.R. Lobry)

Soit l'équation différentielle suivante :

$$\frac{dx}{dt} = a(t)x$$

On considère deux cas :

1. $a(t) = -\omega$ avec $\omega \in \mathbb{R}_+^*$ (autrement dit $\omega > 0$)
 - Donner l'expression de $x(t)$ avec la condition initiale $x(0) = x_0$ ($x_0 > 0$).
 - Donner la valeur de $\lim_{t \rightarrow \infty} x(t)$.
2. $a(t) = -e^{-t}$ (notez que $\forall t \in \mathbb{R} : e^{-t} > 0$)
 - Donner l'expression de $x(t)$ avec la condition initiale $x(0) = x_0$ ($x_0 > 0$).
 - Donner la valeur de $\lim_{t \rightarrow \infty} x(t)$.
3. Donner le code  permettant de produire la figure 1, avec $\omega = 1$ pour le cas 1).
4. Quelle propriété générale illustre la comparaison des cas 1) et 2) ci-dessus ?
5. On cherche à modéliser l'évolution au cours du temps du taux de G+C dans l'ensemble des génomes bactériens. La topologie et les longueurs des branches de l'arbre du vivant sont supposés être connus. Le modèle que vous allez choisir sera-t-il plutôt du type 1) ou bien du type 2) ? On ne demande pas de poser ni de résoudre le modèle, mais de justifier le choix à partir d'arguments biologiques. Il n'y a pas de réponse canonique à cette question.

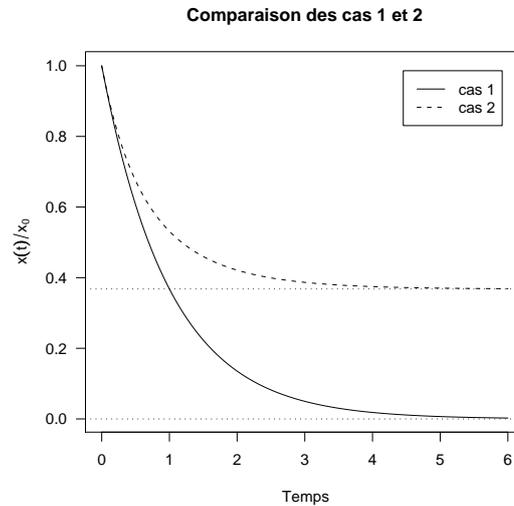


FIG. 1 – La figure à reproduire

2 Problème 2 (L. Duret)

Un fragment du génome du poulet (*Gallus gallus*) a été séquencé :

```
readLines("http://pbil.univ-lyon1.fr/R/donnees/chicken.fasta")[1:5]
[1] ">chicken"
[2] "GATAAAAAGTTTAACGTGATTTTTCAGTGATATAATTTCCATACAGGAAAAGTGTATGATA"
[3] "GATTTGAAAAAGGAATTATGTAATTCGTATTCATTTTTTTAAGGAGCAGATGATTTTAG"
[4] "GTAAGCTTCTAAGAGAGCTATTGAATAATCTATACGTTATAGACATACAGAGATTGGTT"
[5] "TTAGATGTTATTTTTCTGAAAAGATATACTAATGTGTAATAGAATACTTACTATGGCAAC"
```



Gallus gallus
(Linnaeus,
1758)

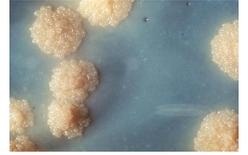
- Recherchez le(s) gène(s) contenu(s) dans ce fragment.
- Décrivez la structure (*i.e.* les positions des introns, des exons et des codons d'initiation et de terminaison) du (ou des) gène(s).
- Décrivez et justifiez votre démarche : pour chaque étape, expliquez son utilité, justifiez votre choix de logiciel, explicitez les paramètres.

3 Problème 3 (G. Perrière)

Etude de l'usage du code de la bactérie *Mycobacterium tuberculosis*, souche H37Rv.

Avant toute chose, calculez une AFC classique et une AFC intra-classe sur l'ensemble des parties codantes de *Mycobacterium tuberculosis* H37Rv récupérées dans la banque EMGLib. Quel est le nombre de facteurs qu'il vous semblerait pertinent de retenir pour chaque analyse? Comparez les résultats des deux AFC au niveau des deux premiers facteurs. Qu'en concluez vous relativement au biais affectant la composition en codons de *M. tuberculosis* H37Rv sur le premier facteur de l'AFC classique? Vérifiez cette hypothèse en regardant les annotations du petit groupe d'environ 50 gènes ayant un score extrême sur ce premier facteur.

Maintenant, en utilisant les résultats de l'AFC intra-classe, déterminez quels sont les autres facteurs affectant l'usage du code chez *M. tuberculosis* H37Rv. Pour ce faire, vous pouvez calculer les indices d'usage du code classiques (CAI, G+C, G+C3) pour l'ensemble des parties codantes.



Mycobacterium tuberculosis

Remarques importantes :

- Les bactéries du genre *Mycobacterium* utilisent le code génétique standard.
- Les noms des gènes codant pour des protéines ribosomales commencent tous par les lettres **rp**, et les dits noms sont tous rentrés en tant que **mots-clés** dans les banques de séquences. Il est donc possible de les récupérer en utilisant une requête *ad hoc*.