

# Épreuve aMIG - Contrôle terminal - 14 juin 2005

P<sup>r</sup> Jean R. LOBRY

14 juin 2005

Génomomes de *Gallus gallus* et *Borrelia garinii*

Contrôle terminal - Durée : 3 heures

*Documents autorisés*

Envoyez (avant 12h, heure de réception du mél faisant foi!) votre compte-rendu au format PDF (incluant les résultats et discussions) à {lobry, duret, perriere}@biomserv.univ-lyon1.fr.

## 1 Question 1

Quels sont à votre avis les principales différences entre le métier d'ingénieur d'étude et celui d'ingénieur de recherche? Un paragraphe maximum.

## 2 Question 2

De quoi était il question dans la conférence du professeur Noboru Sueoka du vendredi 27 mai à 11 heures? Un paragraphe maximum.

## 3 Question 3

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

```
readLines("http://pbil.univ-lyon1.fr/members/lobry/ens/insa/poulet4.fasta")[1:5]
[1] ">poulet4"
[2] "GATAAAAGTTTAAACGTGATTTTTCAGTGATATAATTTCCATACAGGAAAAGTGTATGATA"
[3] "GATTTGAAAAAGGAATTATGTAATTCGATTCAATTTTTTTAAAGGAGCAGATGATTTTAG"
[4] "GTAAGCTTCTAAGAGAGCTATTTGAATAATCTATACGTTATAGACATACAGAGATTGGTT"
[5] "TTAGATGTTATTTTCTGGAAAGATATACTAATGTGTAATAGAATACTTACTATGGCAAC"
```

- Recherchez le(s) gène(s) contenu(s) dans ce fragment.
- Décrivez la structure (*i.e.* les positions des introns, des exons, des régions codantes) du (ou des) gène(s).
- Décrivez et justifiez chacune de vos étapes (*e.g.* pourquoi doit-on rechercher et masquer les séquences répétées?).

## 4 Question 4

Quels sont les facteurs qui régissent l'usage du code chez *Borrelia garinii*? On utilisera le jeu de données suivant :

```
load(url("https://pbil.univ-lyon1.fr/R/donnees/pps067.rda"))
names(pps067)
[1] "codons" "info" "rib"
```

La composante **codons** est un tableau (832 x 64) qui contient pour chacun des 832 gènes les fréquences des 64 codons.

```
dim(pps067$codons)
[1] 832 64
names(pps067$codons)
[1] "aaa" "aac" "aag" "aat" "aca" "acc" "acg" "act" "aga" "agc" "agg" "agt" "ata"
[14] "atc" "atg" "att" "caa" "cac" "cag" "cat" "cca" "ccc" "ccg" "cct" "cga" "cgc"
[27] "cgg" "cgt" "cta" "ctc" "ctg" "ctt" "gaa" "gac" "gag" "gat" "gca" "gcc" "gcg"
[40] "gct" "gga" "ggc" "ggg" "ggt" "gta" "gtc" "gtg" "gtt" "taa" "tac" "tag" "tat"
[53] "tca" "tcc" "tcg" "tct" "tga" "tgc" "tgg" "tgt" "tta" "ttc" "ttg" "ttt"
```

La composante **info** est un tableau (832 x 6) qui contient pour chaque gène :

- le numéro d'accèsion GenBank (**\$Acc**) ;
- le contenu en guanine et cytosine (**\$GC**) ;
- le contenu en guanine et cytosine en troisième position des codons (**\$GC3**) ;
- l'indice d'hydrophobicité de Kyte et Doolittle (**\$K.D**) ;
- l'indice d'adaptation des codons (CAI) (**\$CAI**) ;
- la localisation sur le chromosome (brins leading ou lagging) (**\$strand**).

```
dim(pps067$info)
[1] 832 6
names(pps067$info)
[1] "Acc" "GC" "GC3" "K.D" "CAI" "strand"
```

Enfin, la composante **rib** est un vecteur de caractères à 53 composantes contenant la liste des numéros d'accèsion GenBank pour les protéines ribosomales utilisées pour calculer le CAI.

```
length(pps067$rib)
[1] 53
pps067$rib[1:8]
[1] "CP000013.RPLI" "CP000013.RPSR" "CP000013.RPSF" "CP000013.RPSB" "CP000013.RPSA"
[6] "CP000013.RPLT" "CP000013.RPMI" "CP000013.RPME"
```