



Année universitaire 2012/2013

Site :  Luminy  St-Charles  St-Jérôme  Cht-Gombert  Aix-Montperrin  Aubagne-SATIS

Sujet session de :  1<sup>er</sup> semestre -  2<sup>ème</sup> semestre -  Session 2

Durée de l'épreuve : ...1h.....

Examen de :  L1/L2/ L3 -  M1/ M2 -  LP -  DU

Nom diplôme : **BC + BCB + prepa**

Code Apogée du module : **BI4U2**

Libellé du module : **Bioinformatique appliquée**

Document autorisé :  OUI -  NON

Calculatrices autorisées :  OUI -  NON

## BIOINFORMATIQUE APPLIQUEE (BI4U2) EXAMEN DE SECONDE SESSION – JUIN 2013

### Consignes générales

1. Durée de l'épreuve : 1h.
2. Les documents écrits ne sont pas autorisés.
3. Les étudiants sont autorisés à se servir d'une calculatrice type « collègue ».
4. **Toutes les réponses aux questions doivent être rédigées sur ce questionnaire même.**
5. **Attention !** L'AMU a renforcé les exigences d'anonymisation des examens.

*Extrait de la Charte des examens de l'AMU : « L'anonymat est obligatoire pour les examens terminaux écrits. En tout état de cause, lorsque celui-ci est instauré, il doit être respecté par l'étudiant et contrôlé par le surveillant. **Les copies ne doivent comporter aucun signe distinctif. Dans le cas contraire, elles ne seront pas corrigées.** »*

Ceci nécessite un soin particulier pour assurer l'anonymat tout en évitant de perdre vos réponses.

- a. **Ne dégrafez pas** le questionnaire de la copie standard.
- b. Pour assurer l'anonymat des corrections tout en évitant de perdre des copies, indiquez votre nom et numéro de carte **uniquement sur la copie standard** agrafée à votre questionnaire.
- c. **Encollez** soigneusement le coin de cette copie standard.
- d. Assurez-vous de n'indiquer **aucun signe distinctif** (nom, numéro de carte) sur vos feuilles de questionnaires remplies.

**Question 1 (8 points)**

La figure 1 représente les résultats d'une recherche de similarité effectuée au moyen de l'outil BLAST, pour identifier dans le génome humain les régions codant pour l'enzyme acylphosphatase.

- a. Quelle est la longueur de la séquence requête (précisez l'unité). **(1 point)**

---

- b. Quelle modalité de BLAST a été utilisée et pourquoi ? **(1 point)**

---

---

- c. Comment interpréter les E-valeurs respectives des deux hits de la figure 1b ? **(2 points)**

---

---

- d. La figure 1b montre 2 alignements séparés, recouvrant des fragments différents de la protéine et du génome. Expliquez à quoi correspondent ces deux alignements respectifs. **(1 point)**

---

---

---

- e. Les deux alignements se chevauchent sur une partie de la séquence protéique (acides aminés 29 à 39). L'un de ces segments est un artefact. Lequel ? Expliquez pourquoi BLAST "se trompe" à cet endroit. **(3 points)**

---

---

---

---

---

Figure 1a: en-tête du résultat de BLAST

Your search is limited to records matching entrez query: txid9606 [ORGN].

[Edit and Resubmit](#) [Save Search Strategies](#) [Formatting options](#) [Download](#) [YouTube](#) [How to read th](#)

sp|P07311|ACYP1\_HUMAN Acylphosphatase-1 OS=Homo...

<b>Query ID</b>	Id 27521	<b>Database Name</b>	refseq_genomic
<b>Description</b>	sp P07311 ACYP1_HUMAN Acylphosphatase-1 OS=Homo sapiens GN=ACYP1 PE=1 SV=2	<b>Description</b>	NCBI Genomic Reference Sequences
<b>Molecule type</b>	amino acid	<b>Program</b>	TBLASTN 2.2.28+ <a href="#">Citation</a>
<b>Query Length</b>	99		

Figure 1b: sélection d'alignements retournés par BLAST

[Download](#) [GenBank](#) [Graphics](#) Sort by: E value [Next](#) [Previous](#) [Descriptions](#)

Homo sapiens chromosome 14 genomic scaffold, alternate assembly HuRef SCAF\_1103279188183, whole genome shotgun sequence  
Sequence ID: [ref|NW\\_001838113.2](#) Length: 26002911 Number of Matches: 2

**Range 1: 20219993 to 20220205** [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps	Frame
148 bits(374)	8e-39	Compositional matrix adjust.	71/71(100%)	71/71(100%)	0/71(0%)	+2

**Features:** [acylphosphatase-1 isoform a](#)  
[acylphosphatase-1 isoform b](#)

Query	29	AEGKKLGLVGVQNTDRGTVQSQQLQGPI SKVRHMQEWLETRGSPKSHIDKANFNNEKVL	88
		AEGKKLGLVGVQNTDRGTVQSQQLQGPI SKVRHMQEWLETRGSPKSHIDKANFNNEKVL	
Sbjct	20219993	AEGKKLGLVGVQNTDRGTVQSQQLQGPI SKVRHMQEWLETRGSPKSHIDKANFNNEKVL	20220172

Query	89	KLDYSDFQIVK	99
		KLDYSDFQIVK	
Sbjct	20220173	KLDYSDFQIVK	20220205

**Range 2: 20210100 to 20210210** [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#) [First Match](#)

Score	Expect	Method	Identities	Positives	Gaps	Frame
63.5 bits(153)	1e-09	Compositional matrix adjust.	32/39(82%)	33/39(84%)	2/39(5%)	+3

**Features:** [acylphosphatase-1 isoform a](#)  
[acylphosphatase-1 isoform b](#)

Query	1	MAEGNTLISVDYEIFGRVQGVFFRKHTQAEGRKLGVLGW	39
		MAEGNTLISVDYEIFGRVQGVFFRKHTQ G L +GW	
Sbjct	20210100	MAEGNTLISVDYEIFGRVQGVFFRKHTQVCG--LQALGW	20210210

[Download](#) [GenBank](#) [Graphics](#) Sort by: E value [Next](#) [Previous](#) [Descriptions](#)

**Question 2 (5 points)**

- a. Calculer les scores des deux alignements suivants à l'aide de matrices de substitution ci-dessous. **(1 point)**

	<b>A</b>	<b>T</b>	<b>C</b>	<b>G</b>
<b>A</b>	2	0	0	1
<b>T</b>	0	2	1	0
<b>C</b>	0	1	2	0
<b>G</b>	1	0	0	2

**Pénalités**

Ouverture de gap: 7

Extension de gap: 1

**Alignement 1**

Query	A	T	G	T	C	A	T	A	C	G	T
Subject	A	A	G	T	C	A	-	-	-	G	T
Score											

Score total :

**Alignement 2**

Query	A	T	G	T	C	A	T	A	C	G	T
Subject	A	A	G	T	C	-	-	A	-	G	T
Score											

Score total :

- b. Quel est l'intérêt d'utiliser des matrices de score pour l'alignement ? Quel alignement privilégieriez-vous basés sur les scores ? **(1 point)**

---



---

- c. Quel alignement privilégieriez-vous au point de vue biologique? Justifiez votre réponse. **(2 points)**

---



---

- d. Quel est l'intérêt général d'aligner deux séquences ? **(1 point)**

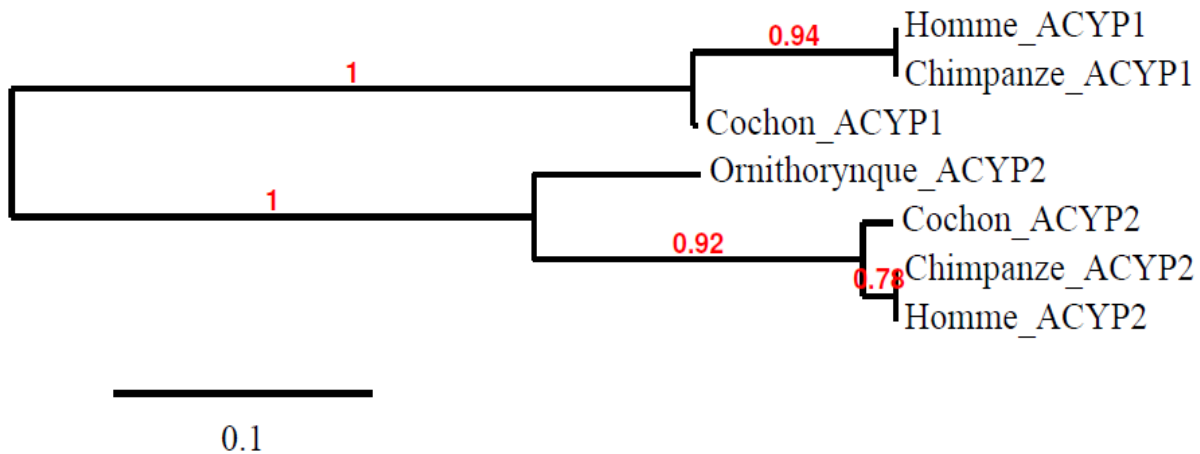
---



---

**Question 3 (7 points)**

**Figure 3: arbre moléculaire des homologues de l'acylphosphatase**



- a. La Figure 3 représente un arbre moléculaire des homologues de l'acylphosphatase chez 4 espèces de mammifères. S'agit-il d'un phylogramme ou d'un cladogramme ? Justifiez votre réponse en 1 ligne. **(1 point)**

- b. Sur base de vos connaissances générales en biologie, dessinez ci-dessous un cladogramme indiquant les relations entre les 4 espèces concernées. **(1 point)**



- c. La Figure 3 représente un arbre non-enraciné. Placez la racine dans l'arbre (sur la Figure 3), et justifiez votre choix. **(2 points)**

- d. Indiquez sur la Figure 3 les événements de spéciation (marquez-les d'un S) et duplication (marquez-les d'un D). Au cas où un événement pourrait se placer de façon équivalente à plusieurs endroits, n'en marquez qu'un seul de votre choix. **(1 point)**

- e. Pour des raisons indépendantes de l'analyse de séquences, on suppose que l'ornithorynque possède un gène ACYP1 fonctionnel, qui n'a apparemment pas encore été documenté dans notre base de données de séquences. Indiquez une position vraisemblable pour ce gène sur la Figure 3. Justifiez votre réponse. **(1 point)**

---



---



---

- f. Remplissez la Table ci-dessous en indiquant, pour chaque paire de séquences, le type d'homologie (O=Orthologie ; P=Paralogie ; I=Identité). **(1 point)**

	Homme ACYP1	Homme ACYP2	Chimpanzé ACYP1	Chimpanzé ACYP2	Cochon ACYP1	Cochon ACYP2	Ornithorynque ACYP2
Homme ACYP1							
Homme ACYP2							
Chimpanzé ACYP1							
Chimpanzé ACYP2							
Cochon ACYP1							
Cochon ACYP2							
Ornithorynque ACYP2							