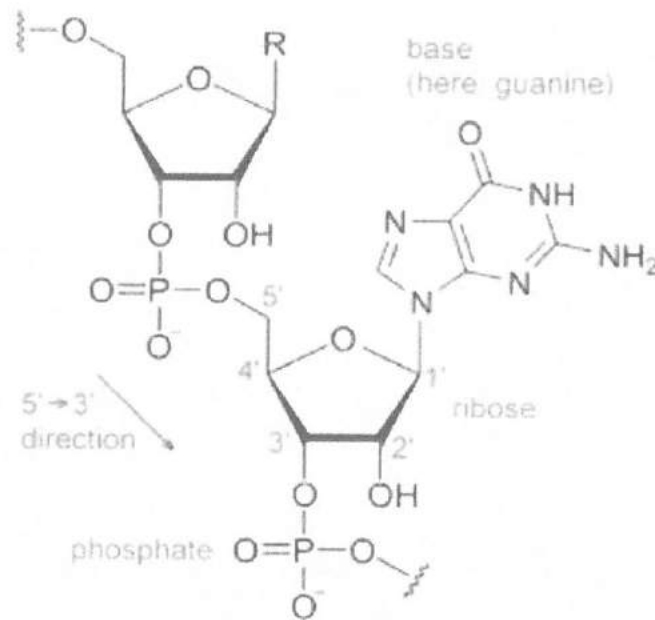


# Initiation à la génomique

IF I AM LOST IN TRANSLATION



**JUST BLAME MY RNA**

© WORDS & UNWORDS

**N°d'anonymat :**

**I.** (6 points) Citez les différentes étapes de la technique d'immunoprécipitation de la chromatine et expliquez en une phrase l'intérêt de chaque étape. Dans quel but cette technique est-elle utilisée ?

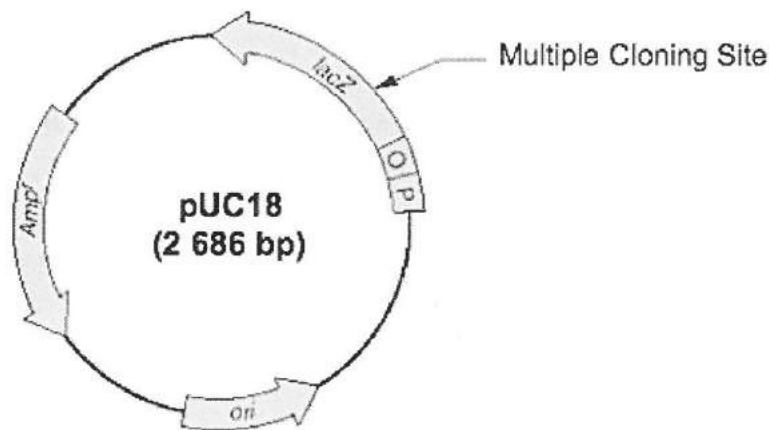
**II.** (5 points) Définissez le terme promoteur et décrivez les caractéristiques des promoteurs de gènes codant pour des protéines eucaryotes. Expliquez la raison pour laquelle il est difficile de les caractériser au niveau d'un génome.

**III.** (4 points) Définissez le terme transcriptome, décrivez brièvement sa composition et expliquez l'origine de sa complexité.

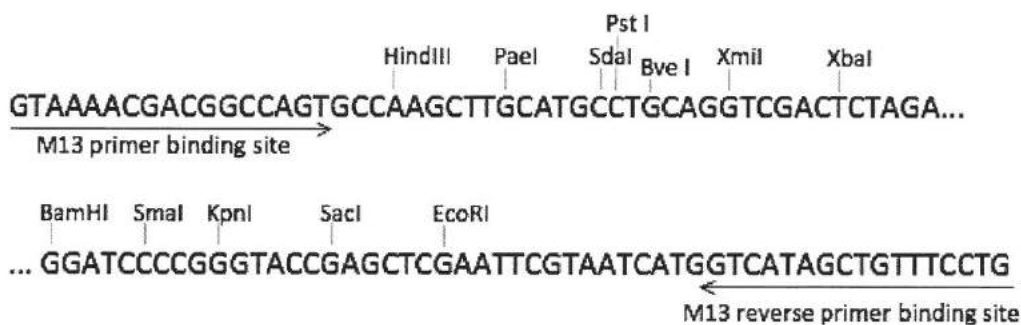
**IV.** (5 points) Montrez qu'un génome est une structure dynamique. Vous rédigerez un court texte de quelques lignes.

*L'usage des téléphones portables est interdit pendant toute la durée des épreuves. Les appareils doivent impérativement être éteints pendant les épreuves. Aucune calculatrice n'est autorisée pendant toute la durée de l'épreuve.*

Un fragment d'ADN qui présente des extrémités EcoRI (G'AATTC) a été inséré dans le site EcoRI du vecteur de clonage pUC18, représenté schématiquement ci-dessous.



### pUC18 Multiple Cloning Site Region



1. Quelle est la particularité des sites de restriction présents dans la cassette de clonage (Multiple Cloning Site Region) d'un vecteur de clonage ? Justifiez.
2. Quels types d'extrémités sont générés par l'enzyme EcoRI ? Combien de sens d'insertion du fragment dans le vecteur peut-on observer ? Justifiez votre réponse en 1 ou 2 phrases.

Le plasmide recombinant est digéré par les enzymes indiquées dans le tableau ci-dessous :

| Enzyme de restriction | Taille des fragments de restriction obtenus |
|-----------------------|---|
| EcoRI                 | 0,5 kb + 0,8 kb + 2,7 kb                    |
| BamHI                 | 0,8 kb + 3,2 kb                             |
| SmaI                  | 1 kb + 3 kb                                 |
| BamHI + SmaI          | 0,2 kb + 0,8 kb + 3 kb                      |
| EcoRI + SmaI          | 0,3 kb + 0,5 kb + 0,5 kb + 2,7 kb           |
| EcoRI + BamHI         | 0,3 kb + 0,5 kb + 0,5 kb + 2,7 kb           |
| EcoRI + BamHI + SmaI  | 0,2 kb + 0,3 kb + 0,3 kb + 0,5 kb + 2,7 kb  |

3. Quelle est la taille du fragment inséré ? Justifiez.

4. Etablissez la carte de restriction du plasmide recombinant. Justifiez succinctement vos propositions.

Afin de déterminer la séquence de ce fragment, une réaction de séquençage par la méthode de Sanger a été effectuée. L'amorce utilisée est le primer **M13 reverse**.

5. Ecrivez la séquence de l'amorce M13 reverse.

6. Ecrivez les séquences des 3 fragments les plus courts obtenus suite à l'élongation de l'amorce par la polymérase, dans le milieu réactionnel qui contient, entre autre, des ddGTPs,

---

*L'usage des téléphones portables est interdit pendant toute la durée des épreuves. Les appareils doivent impérativement être éteints pendant les épreuves.*

---

### **Partie 1 (5 points)**

- Définissez le terme cDNA.
- Comment est-il obtenu à partir d'ARN messagers polyadénylés ?
- Donnez une utilisation possible du cDNA.

### **Partie 2 (5 points)**

Les banques génomiques utilisées pour le séquençage complet d'un génome sont dites chevauchantes.

Définissez ce terme. En vous appuyant sur un schéma, démontrez qu'il est essentiel que les banques soient chevauchantes.

### **Partie 3 (10 points)**

1. Que représente la profondeur de séquençage d'un génome ?

Quelle est la profondeur de séquençage d'un génome de 65Mb pour lequel 1 500 000 clones ont été séquencés avec la méthode de Sanger, en appliquant le séquençage d'extrémités appariés (taille moyenne des séquences lues : 500 nucléotides).

Que pensez-vous d'une telle couverture ? Justifiez.

2. Lors du séquençage d'un génome d'arthropode, 274 contigs ont été générés. Parmi ces derniers, les contigs *c026* et *c057* ont des longueurs de 100 846 pb et 38 657 pb respectivement. La lecture de la séquence de l'une des extrémités d'un clone correspond aux nucléotides 245 à 682 du contig *c026*; la lecture de la séquence de l'autre extrémité de ce même clone correspond aux nucléotides 102 à 677 du contig *c057*.

À partir de ces informations, et sachant que les fragments clonés ont une taille de 3kb, dessinez une carte de l'échelle contenant les contigs *c026* et *c057*, en indiquant la taille totale de l'échelle et la taille de la brèche entre les 2 contigs.

Comment procéderiez-vous pour obtenir la séquence complète de la brèche ?

3. Dans le cadre du séquençage de génomes humains, qu'appelle-t-on le séquençage d'exome ? Quel(s) avantage(s) y voyez-vous ?

-L'usage des téléphones portables est interdit pendant toute la durée des épreuves, y compris lors de la préparation des épreuves orales. Les appareils doivent impérativement être éteints pendant les épreuves. Ils ne peuvent donc pas être utilisés comme chronomètre ou calculatrice.

-L'usage d'une calculatrice est limité à des appareils non programmables ne comportant ni écran graphique et ni de caractères alphanumériques. Le prêt d'une calculatrice à un autre candidat est strictement interdit.

- les documents ne sont pas autorisés.

**1. (5 points)** Le gène codant pour une protéine eucaryote de 638 acides aminés, s'étend sur 67559 paires de bases (pb). La longueur cumulée des 16 exons est de 2947 pb. Commenter ces données. Justifier votre réponse.

**2. (4 points)** Les banques génomiques utilisées pour le séquençage complet d'un génome sont dites chevauchantes. Définissez ce terme. En vous appuyant sur un schéma, démontrez qu'il est essentiel que les banques soient chevauchantes.

**3. (9 points)** Un microbiologiste envisage de séquencer le génome d'une souche bactérienne isolée à partir d'un échantillon de sol. La bactérie isolée a un génome de 2,6 Mb environ. Il vous dit qu'il envisage de faire séquencer ce génome en utilisant la stratégie de séquençage dite clone par clone.

**a)** Pourquoi lui proposez – vous plutôt de suivre la stratégie appelée shot-gun ?

Une première étape de séquençage à partir d'une banque d'ADN constituée de plasmides contenant des inserts d'une taille moyenne de 3 kb a permis de générer 2,7 Mb de séquences qui se répartissent sur 230 contigs.

**b)** Définissez ce que l'on appelle la couverture ou profondeur de séquençage.

**c)** Déterminer la couverture de ce séquençage.

**d)** Qu'appelle – t- on un contig ?

Le microbiologiste revient vers vous et vous dit qu'il n'arrive pas à partir de ces séquences à reconstituer le génome en entier et que son gène favori n'est pas entièrement présent dans la séquence.

**e)** En vous appuyant sur la couverture de ce séquençage, expliquez-lui pourquoi il rencontre ses difficultés.

Un nouveau séquençage de la banque est entrepris qui conduit cette fois-ci à la genèse de 15 Mb.

**f)** Calculez la nouvelle couverture et expliquez pourquoi les problèmes précédents sont à présent résolus ?

**4. (2 points)** Un peu de culture générale : le 17 avril 2013, le journal Nature a publié un article relatant le séquençage complet du génome du Coelacanthe (*Latimeria chalumnae*). Quels sont les intérêts scientifiques que vous voyez pour le séquençage du génome d'une telle espèce.

L'usage des téléphones portables est interdit pendant toute la durée des épreuves, y compris lors de la préparation des épreuves orales. Les appareils doivent impérativement être éteints pendant les épreuves. Ils ne peuvent donc pas être utilisés comme chronomètre ou calculatrice.

L'usage d'une calculatrice est limité à des appareils non programmables ne comportant ni écran graphique et ni de caractères alphanumériques. Le prêt d'une calculatrice à un autre candidat est strictement interdit.

Au sein de chaque partie, répondez aux questions **dans l'ordre**.

### Partie 1

Votre laboratoire souhaite séquencer le génome de la bactérie *Streptococcus macedonicus*, dont la taille du génome est estimée à 2,1Mb.

Pour ce projet, le séquençage d'extrémités appariées est réalisé par la méthode de Sanger. Une profondeur théorique de séquençage de 5X a été prévue.

1. Quelle stratégie de séquençage est la plus couramment utilisée pour le séquençage de ce type de génome ?
2. Décrivez les étapes de cette stratégie.
3. Quels sont les avantages et les limites de cette stratégie ?
4. Que représente la profondeur de séquençage d'un génome ? Que pensez-vous de la profondeur théorique choisie ?
5. Que signifie « séquençage d'extrémités appariées » ? Quel est l'intérêt d'appliquer cette technique ?
6. Si on considère que la taille moyenne des lectures est de 800 nucléotides et que les fragments clonés ont une taille de 2000 pb, combien de fragments devront être séquencés pour atteindre la profondeur théorique souhaitée? Justifiez.

Lors de l'assemblage des lectures de ce génome, 307 contigs ont été générés. Parmi ces derniers, les contigs *cont002* et *cont215* ont des longueurs de 3 725 pb et 45 545 pb respectivement. La lecture de la séquence de l'une des extrémités d'un clone, de 725 nt de long, s'aligne sur les nucléotides 2 978 à 3 702 du contig *cont002*; la lecture de la séquence de l'autre extrémité de ce même clone, de 654 nt de long, s'aligne sur les nucléotides 44 870 à 45 523 du contig *cont215*.

7. À partir de ces informations, dessinez une carte de l'échelle contenant les contigs *cont002* et *cont215*, en indiquant la taille totale de l'échelle et la taille de la brèche entre les 2 contigs.
8. Le positionnement des contigs les uns par rapports aux autres vous aura permis d'obtenir 104 scaffolds. Afin de disposer d'une séquence complète du génome de *S. macedonicus*, proposez 2 méthodes qui vous permettraient de positionner ces 104 scaffolds les uns par rapport aux autres.



## Partie 2

Un fragment d'ADN issu de la digestion par l'enzyme *KpnI* (GGTAC/C) doit être inséré dans le vecteur pUC18 (2,7 kb, cf ci-dessous), linéarisé par cette même enzyme.

1. Quels types d'extrémités sont générés après coupure par l'enzyme *KpnI* ?

Le plasmide recombinant est digéré par les enzymes indiquées dans le tableau ci-dessous :

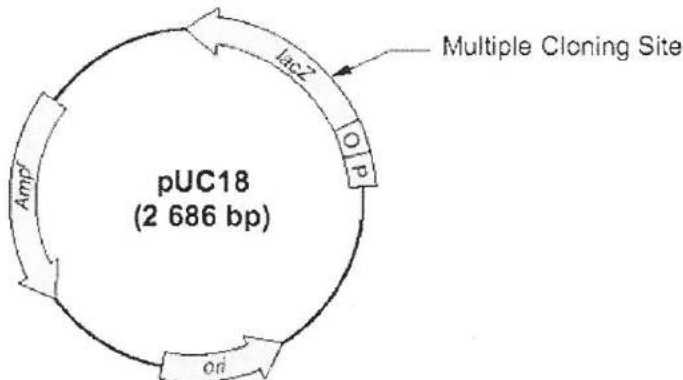
| Enzyme de restriction   | Taille des fragments de restriction obtenus    |
|-------------------------|--|
| BamHI                   | 1,2 kb + 3,8 kb                                |
| HindIII                 | 0,2 kb + 1,7 kb + 3,1 kb                       |
| EcoRI                   | 1,3 kb + 3,7 kb                                |
| HindIII + EcoRI         | 0,2 kb + 0,4 kb + 0,8kb + 0,9 kb + 2,7 kb      |
| BamHI + HindIII + EcoRI | 2 x 0,2 kb + 0,4 kb + 0,7 kb + 0,8 kb + 2,7 kb |

2. Quelle est la taille du fragment inséré dans le vecteur ? Justifiez votre réponse.

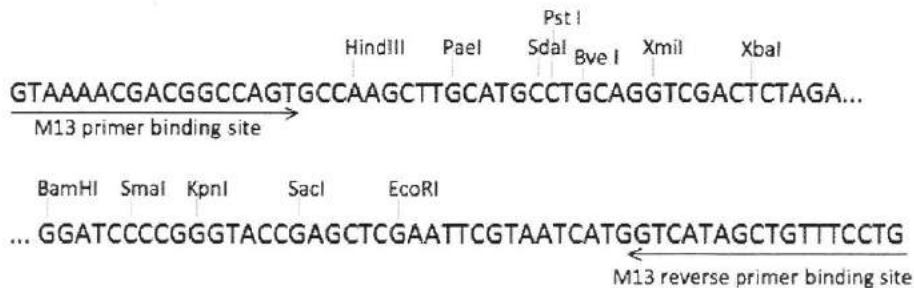
3. Établissez la carte de restriction du plasmide recombinant.

4. Donnez le nombre et la taille des fragments qui auraient été obtenus pour chacune des digestions considérées si le fragment *KpnI* avait été inséré en sens inverse dans le vecteur.

### Carte du vecteur pUC18 et séquence du MCS :



### pUC18 Multiple Cloning Site Region



Les sites de restriction présentés dans la cassette de clonage sont uniques pour le vecteur pUC18.

L'usage des téléphones portables est interdit pendant toute la durée des épreuves, y compris lors de la préparation des épreuves orales. Les appareils doivent impérativement être éteints pendant les épreuves. Ils ne peuvent donc pas être utilisés comme chronomètre ou calculatrice.

L'usage d'une calculatrice est limité à des appareils non programmables ne comportant ni écran graphique et ni de caractères alphanumériques. Le prêt d'une calculatrice à un autre candidat est strictement interdit.

Au sein de chaque partie, répondez aux questions **dans l'ordre**.

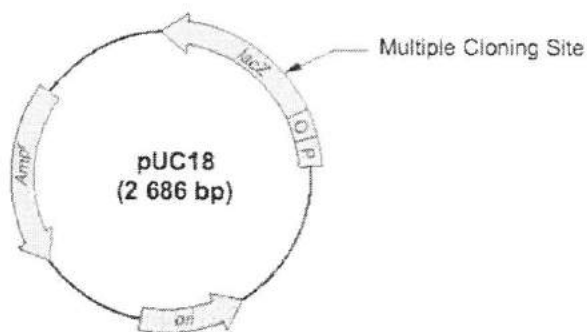
### Partie 1

Un fragment d'ADN de 550 paires de bases a été inséré dans le site *HindIII* (A'AGCTT) du vecteur de clonage pUC18 (cf ci-dessous).

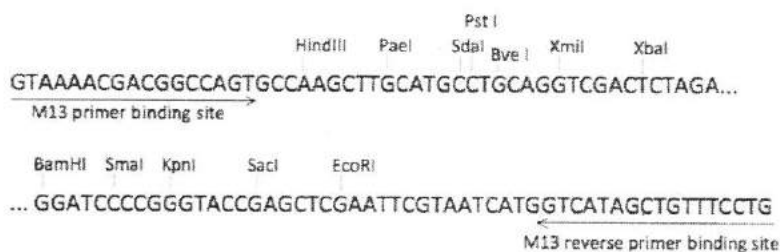
Afin de déterminer la séquence de ce fragment, une réaction de séquençage par la méthode de Sanger a été effectuée. L'amorce utilisée est le primer **M13 reverse**.

1. Quels types d'extrémités sont générés par *HindIII*.
2. Donnez le principe du séquençage de l'ADN par la méthode de Sanger.
3. Ecrivez la séquence de l'amorce M13 reverse.
4. Ecrivez les séquences des 4 fragments les plus courts obtenus dans le milieu réactionnel qui contient, entre autre, des ddATPs.

### Carte du vecteur pUC18 et séquence du MCS :



### pUC18 Multiple Cloning Site Region



## Partie 2

Vous souhaitez séquencer le génome de *Apis mellifera*, un arthropode d'intérêt pour votre laboratoire.

La taille estimée de son génome est de 200 Mb.

1. Quelle stratégie de séquençage allez-vous mettre en place pour séquencer le génome de cet organisme ? Justifiez.
2. Décrivez les principales étapes de cette stratégie. Quels sont ses avantages ? Ses limites ?

Dans le cadre de ce projet, pour lequel le séquençage d'extrémités appariées a été réalisé, 1 123 873 clones contenant des inserts de 2 000 paires de bases ont été construits.

3. Que signifie « séquençage d'extrémités appariées » ? Quel est l'intérêt d'appliquer cette technique ?

4. Que représente la profondeur de séquençage d'un génome ?

Le séquençage de ce génome, réalisé en Sanger, a permis de générer des lectures de 800 nt de long en moyenne.

5. Quelle est la couverture théorique de ce génome ? Qu'en pensez-vous ?

Une fois la séquence du génome de *A. mellifera* reconstituée, vous souhaitez mettre en évidence les gènes localisés sur cette séquence.

6. Après avoir défini ce qu'est une ORF et quelles sont ses caractéristiques, précisez si la recherche d'ORF le long du génome de *A. mellifera* vous paraît pertinente pour la mise en évidence des gènes. Justifiez votre réponse.

8. Donnez si possible une estimation du nombre de gènes codés par ce génome, en justifiant votre réponse.

## Partie 3

Définissez la Génomique Comparative. Quels en sont les principaux objectifs ?

Mai 2012 session 1  
Durée 20 min

*L'usage des téléphones portables est interdit pendant toute la durée des épreuves, y compris lors de la préparation des épreuves orales. Les appareils doivent impérativement être éteints pendant les épreuves. Ils ne peuvent donc pas être utilisés comme chronomètre ou calculatrice.  
Aucune calculatrice n'est autorisée pendant toute la durée de l'épreuve.*

Définir le terme : épigénome.

Décrire les réactions enzymatiques impliquées dans la modification de la chromatine.

Mai 2012 session 2

Durée 20 min

*L'usage des téléphones portables est interdit pendant toute la durée des épreuves, y compris lors de la préparation des épreuves orales. Les appareils doivent impérativement être éteints pendant les épreuves. Ils ne peuvent donc pas être utilisés comme chronomètre ou calculatrice.  
Aucune calculatrice n'est autorisée pendant toute la durée de l'épreuve.*

- I.** (3 points) Qu'est ce qu'un promoteur. Qu'est ce que le +1 d'un gène, quelle démarche utiliseriez-vous pour localiser le +1 d'un gène au niveau d'un génome.
- II.** (3 points) Décrivez les promoteurs bactériens et donnez les caractéristiques des gènes codant pour les protéines bactériennes.

***L'usage des téléphones portables est interdit pendant toute la durée des épreuves. Les appareils doivent impérativement être éteints pendant les épreuves. Aucune calculatrice n'est autorisée pendant toute la durée de l'épreuve. Barème et durée : indications approximatives***

## **N°d'anonymat :**

### **1. Questions à réponses rapides : (6 points – 5 min)**

1-a : Quel est le premier génome non viral dont la séquence a été entièrement déterminée ?

1-b : Quel est le premier génome eucaryote séquencé ?

Question bonus (1pt) : indiquez ci-dessus les années de ces deux séquençages complets.

1-c : Quel est le rôle du ddATP dans la technique de séquençage de Sanger ?

1-d : Qu'est-ce qu'une séquence que l'on appelle Draft ?

1-e : Que représente la valeur C d'un génome ?

### **2. Le génome, une structure dynamique (3 points - 2 min)**

Citez trois arguments justifiant que le génome est une structure dynamique.

- 
- 
- 

### **3. Une banque d'ADN chevauchante (5 points – 10 min)**

Pourquoi une banque doit-elle être chevauchante dans le cadre d'un séquençage systématique. Vous argumenterez votre réponse par un schéma et vous expliquerez comment l'on peut utiliser une enzyme de restriction pour la fragmentation de l'ADN.

**4. Assemblage des séquences (6 points – 12 min)**

Expliquez en quoi la présence de séquences répétées peut poser un problème au moment de l'assemblage d'un génome. Vous pouvez vous baser sur un schéma pour justifier votre réponse. Citez 2 solutions pour palier à ce problème.

**L'usage des téléphones portables est interdit pendant toute la durée des épreuves. Les appareils doivent impérativement être éteints pendant les épreuves. Les calculatrices sont autorisées durant cette épreuve.**

**N°d'anonymat :**

## 1. Structure et contenu des génomes

1-a. Le génome humain : renseignez les informations suivantes (3 points)

Taille du génome humain :

Nombre de chromosomes :

Taille des régions codantes :

Taille des régions non-codantes relatives aux gènes (précisez les types d'éléments concernés) :

Taille des régions intergéniques :

Variabilité entre 2 individus non apparentés :

1-b. Quelles sont les principales différences, en termes de structure et de contenu, entre le génome de *Saccharomyces cerevisiae* et celui de l'homme. (3 points)

1-c. Que représente la valeur C d'un génome ? Pourquoi parle-t-on du paradoxe de la valeur C ? Appuyez-vous sur un exemple pour illustrer votre réponse. (2 points)

## 2. Séquençage de génome complet

Vous souhaitez séquencer le génome de *Lachancea quebecensis*, une levure d'intérêt pour votre laboratoire, dont la taille du génome est estimée à 11 Mb.

Dans le cadre de ce projet, pour lequel le séquençage d'extrémités appariées a été réalisé, 201 604 clones contenant des inserts de 2 kb ont été construits.

2-a. Que signifie « séquençage d'extrémités appariées » ? Quel est l'intérêt d'appliquer cette technique ? Justifier votre réponse à l'aide d'un schéma. (2 points)

2-b. Que représente la profondeur de séquençage d'un génome? (1 point)

Le séquençage de ce génome, réalisé en Sanger, a permis de générer des lectures de 800 nt de long en moyenne.

2-c. Quelle est la profondeur de séquençage appliquée ? Que pensez-vous de cette valeur ? Justifiez. (3 points)



L'assemblage de quelques lectures issues de ce projet vous est présenté ci-dessous.

|            |                            |                            |   |
|------------|----------------------------|----------------------------|---|
|            | ↓                          |                            | ↓ |
| lecture 1  | CTGATCGTAAACGTAGGCTTTT     |                            |   |
| lecture 2  | GTAGGCTTTTTCACCATGCGGGA    |                            |   |
| lecture 3  |                            | TGTGGGATGCAGTGATTTTACCCATG |   |
| lecture 4  | AGGCTTTTTCACCATGTG         |                            |   |
| lecture 5  |                            | CATGCGGGATGCAGTGATTTTACC   |   |
| lecture 6  | GTAGCCTTTTTCACCATGCG       |                            |   |
| lecture 7  | TGATCGTAAACGTAGGCTTTTTCACC |                            |   |
| lecture 8  |                            | TGTGGGATGCAGTGATTTTAC      |   |
| lecture 9  | GCTTTTTCACCATGCGGGATGCAGT  |                            |   |
| lecture 10 | GATCGTAAACGTAGGCTTTT       |                            |   |
| lecture 11 |                            | CGGGATGCAGTGATTTTACCCAT    |   |
| lecture 12 | GTAAACGTAGGCTTTTTCGA       |                            |   |
| lecture 13 | AGGCTTTTTCACCATGTGGGATGCA  |                            |   |
| lecture 14 | GTAAACGTAGGCTT             |                            |   |
| lecture 15 |                            | TTGCACCATGTGGGATG          |   |

2-d. Déduisez la séquence du contig qui découle de cet assemblage. Quelle est la couverture minimale du contig ? Quelle est sa couverture maximale ? Indiquez les résidus concernés en les soulignant. (2 points)

2-e. Qu'observez-vous au niveau des colonnes surmontées d'une flèche ? Quelles sont vos hypothèses pour expliquer ces phénomènes ? (2 points)

Vous souhaiteriez connaître la séquence située en aval de ce contig le long de la séquence d'ADN génomique.

2-f. Donnez la séquence d'une amorce qui vous permettrait d'initier la réaction de séquençage. Justifiez. (2 points)

Sujet A. Friedrich. 10 points.  
Durée conseillée : 30 minutes.

***L'usage des téléphones portables est interdit pendant toute la durée des épreuves. Les appareils doivent impérativement être éteints pendant les épreuves. Les calculatrices sont autorisées durant cette épreuve.***

**N°d'anonymat :**

### **Séquençage d'un génome bactérien.**

Lors du séquençage d'*Anabaena circinalis*, une cyanobactérie dont la taille de génome est estimée à 7 Mb, le séquençage d'extrémités appariées a été réalisé sur 102 000 clones, contenant des inserts de 3 kb.

8 contigs ont été construits lors de l'assemblage des lectures.

Parmi ces derniers, les contigs C1 et C7 ont des longueurs de 1 350 780 pb et 531 756 pb, respectivement.

La lecture issue d'une extrémité du clone cl7914 correspond aux nucléotides 57 à 832 du contig C1; la lecture issue de l'autre extrémité de ce même clone correspond aux positions 72 à 642 du contig C7.

1- Que pouvez-vous dire par rapport à l'orientation des clones C1 et C7 ? Justifiez. (1,5 points)

2- Dessinez une carte du scaffold contenant les contigs C1 et C7, en précisant la taille de la brèche entre les 2 contigs. (3 points)

3- Quelle est la taille totale du scaffold ? (0,5 points)

4- Proposez une stratégie pour positionner l'ensemble des contigs les uns par rapport aux autres. (1,5 points)

5- Donnez une estimation du nombre de gènes codant pour des protéines que porte ce génome. Justifiez. (1 point)

Le gène codant pour la protéine Cas1, qui a une longueur de 350 acides aminés, a été localisé sur ce génome.

6- Qu'est-ce qu'une CDS et quelles sont ses caractéristiques ? (1,5 points)

7- Quelle est la longueur de la CDS du gène codant pour Cas1 ? (1 point)

Sujet E. Carbon. 10 points.  
Durée conseillée : 30 minutes.

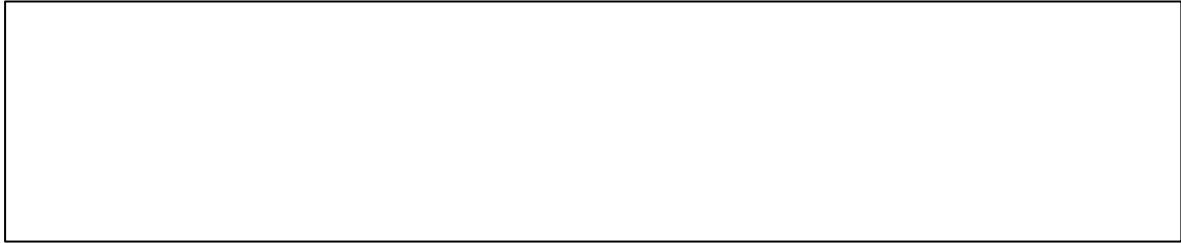
*L'usage des téléphones portables est interdit pendant toute la durée des épreuves. Les appareils doivent impérativement être éteints pendant les épreuves. Les calculatrices sont autorisées durant cette épreuve.*

**N°d'anonymat :**

1°) Comment définissez-vous la génomique fonctionnelle ? (1point)

2°) Décrivez un nucléosome. (2 points)

3°) Décrivez les réactions de modifications chimiques de la chromatine. (3 points)



4°) Au niveau du génome, qu'est ce qui peut influencer l'expression d'un gène ? (4 points)

