

Biotechnologies

Analyse bioinformatique de séquences d'ADN

Niveau

• Terminale STL - Enseignement spécifique de biotechnologies

Thème du programme

• Initiation à la biologie moléculaire et au génie génétique
→ Du gène à la protéine

Situations pédagogiques

• Activité technologique en salle informatique

Liens internet

• <http://www.endmemo.com/bio/gc.php> • <http://www.ebi.ac.uk/Tools/msa/muscle/>
• <http://blast.ncbi.nlm.nih.gov> • <http://web.expasy.org/translate/>

Compétences B2i

• Domaine 1 : travailler dans un environnement numérique évolutif
• Domaine 3 : produire, traiter, exploiter et diffuser des documents numériques

Matériels TICE

• Un poste PC par binôme
• Une connexion internet
• Logiciel de traitement de texte et d'images

Mots clés

• Pourcentage de GC, identification de séquences, alignement de séquences, traduction de séquences, BLAST, MUSCLE (ClustalW).



Votre avis nous intéresse, merci de répondre à notre enquête concernant ce scénario.

Elève, cliquer [ici](#).

Professeur, cliquer [ici](#).

Activité technologique : Analyse bioinformatique de séquences d'ADN

Contexte de l'Activité technologique

Un patient découvre qu'il est porteur hétérozygote d'une mutation non répertoriée de la drépanocytose. Or, la drépanocytose est une pathologie dont la gravité dépend des mutations qui affectent le gène de la bêta globine.

Ainsi, pour évaluer le risque associé à cette mutation, plusieurs méthodes et techniques ont été mises en œuvre. D'abord, l'ADN du patient a été extrait puis purifié. Dans cet ADN, la séquence correspondant au gène codant la bêta globine a ensuite été amplifiée par PCR. Enfin, les résultats de PCR ont été envoyés à un laboratoire de séquençage afin d'obtenir les séquences exactes du gène de la bêta globine du patient.

Aujourd'hui, vous allez exploiter ces séquences à l'aide d'outils bioinformatiques pour évaluer, a priori, la gravité de la pathologie associée à la mutation de ce patient.

Objectifs

Étudier des séquences nucléotidiques et peptidiques pour comprendre leur fonction :

- Identifier des séquences en les cherchant dans des bases de données à l'aide de l'outil BLAST
- Traduire informatiquement des séquences nucléotidiques en séquences peptidiques à l'aide de l'outil ExPASy
- Comparer des séquences entre elles à l'aide de l'outil MUSCLE.

Modalités de travail

- L'ensemble des fichiers utiles sont contenus dans un dossier situé sur l'espace de partage de la classe sur le réseau (ou sur le Bureau de l'ordinateur).
- Les différents outils bioinformatiques à utiliser sont accessibles en ligne à l'aide de n'importe quel navigateur web comme Firefox ou Chrome.
- Le fonctionnement des outils utilisés est détaillé dans les annexes situées à la fin de ce document.
- Le compte-rendu de cette activité est à produire sous forme d'un document informatique. Ce document contiendra les réponses aux questions ainsi que des impressions d'écran lorsque cela est demandé. À la fin de la séance, le fichier produit devra être déposé sur la plateforme prévue à cet effet.

Activité n°1 : Trois séquences ?

Le laboratoire de séquençage nous a envoyé par e-mail le fichier "[ResultatSequencage.txt](#)" dont le contenu a été copié-collé à la fin de ce fichier. Il contient trois séquences distinctes notées "Séquence 1", "Séquence 2" et "Séquence 3". Or, nous avons envoyé un seul extrait d'ADN ; à quoi correspondent donc ces trois séquences ? Vous allez à présent utiliser des outils bioinformatiques pour identifier ces séquences.



Questions

1- À l'aide de l'[annexe 1](#), calculer les pourcentages de GC des trois séquences obtenues. Présenter une impression d'écran pour l'un de ces calculs.

2- Comparez ces trois pourcentages et conclure quant à l'origine des trois séquences.

N.B : Des séquences ayant moins de 1% de différence dans leurs pourcentages de GC sont considérées comme étant issues du même organisme.

3- À l'aide de l'[annexe 2](#), utiliser l'outil BLAST pour chercher les trois séquences dans les bases de données. Présenter une impression d'écran des premiers résultats obtenus pour chacune des séquences recherchées.

Remarque : Dans le cadre de cette Activité, les séquences cherchées seront présentes dans les bases de données ; pour ces séquences, les pourcentages d'identité seront donc de 100%. Par conséquent, on considérera que les séquences ayant ce score de 100% d'identité avec la séquence cherchée seront effectivement la séquence cherchée.

4- Grâce aux résultats fournis par l'outil BLAST, déterminer quels gènes sont codés par ces trois séquences. Préciser le ou les organisme(s) dont ils sont issus. Faire une hypothèse sur la qualité des manipulations du laboratoire de séquençage.

Activité n°2 : Gène muté et gène non muté

Vous avez à votre disposition les séquences de la version mutée et de la version non mutée du gène de l'hémoglobine (dans le fichier "[ResultatSequencage.txt](#)"). Vous allez à présent évaluer la différence qui existe entre ces deux séquences nucléotidiques.



Questions

- 1- Comparer les 20 premiers nucléotides des deux séquences du gène de l'hémoglobine et mesurer le temps que vous mettez à faire cette comparaison. Présenter cette comparaison selon le même format que l'exemple de résultat de [l'annexe 3](#).
- 2- Calculer le temps qu'il vous faudrait pour comparer des séquences de 2000 nucléotides de long (c'est la taille approximative du gène de l'hémoglobine).
- 3- En vous aidant de [l'annexe 3](#), comparer à présent les deux séquences grâce à l'outil MUSCLE et présenter une impression d'écran du résultat obtenu.
- 4- Évaluer le temps que nécessite la comparaison par MUSCLE, comparer ce temps à celui de la question 2- et conclure sur l'efficacité de cet outil.
- 5- Identifier la nature de la mutation portée par le patient. Préciser s'il s'agit d'une insertion, d'une délétion, ou d'une substitution et indiquer combien de nucléotides sont affectés.

Activité n°3 : Du gène muté à la protéine mutée

Vous avez identifié la mutation portée par le patient à un niveau génétique, mais vous ne savez pas encore quelles conséquences cette mutation peut avoir sur le phénotype. Vous allez donc à présent traduire les gènes mutés et non mutés pour identifier les conséquences de la mutation sur l'hémoglobine.



Questions

- 1- À l'aide du code génétique fourni en [annexe 4](#), traduire les 20 premiers nucléotides du gène muté et mesurer le temps que vous mettez à faire cette traduction. Présenter cette traduction selon le même format que l'exemple de résultat de l'[annexe 5](#).
- 2- Calculer le temps qu'il vous faudrait pour traduire des séquences de 2000 nucléotides de long (c'est la taille approximative du gène de l'hémoglobine).
- 3- Sachant qu'un ORF ("Open Reading Frame" en anglais) commence toujours par un codon d'initiation AUG, préciser s'il y en a un dans les 20 premiers nucléotides que vous avez traduits à la main.
- 4- À l'aide de l'[annexe 5](#), utiliser l'outil ExPASy pour traduire les séquences des deux gènes et présenter une impression d'écran du résultat obtenu.
- 5- Évaluer le temps que nécessite la traduction par ExPASy, comparer ce temps à celui de la question 2- et conclure sur l'efficacité de cet outil.

N.B : Dans la suite, les séquences auxquelles on s'intéressera seront celles du troisième cadre de lecture dans le sens 5' → 3' ; c'est à dire les séquences notées "**5'3' Frame 3**".

Activité n°4 : Conséquences de la mutation

À présent que les séquences nucléotidiques ont été traduites en séquences peptidiques, vous allez comparer les protéines mutées et non mutées pour identifier leurs différences. Vous allez donc exploiter MUSCLE mais pour comparer des séquences peptidiques cette fois. Cela vous permettra de conclure sur le danger éventuel que représente cette mutation.



Questions

- 1- À l'aide de l'[annexe 3](#), comparer les séquences du premier ORF (surligné en rouge) du troisième cadre de lecture ("5'3' *Frame 3*") de chacun des deux gènes (muté et non muté). Présenter une impression d'écran du résultat obtenu.
- 2- Lister les différences et les ressemblances que vous observez entre la protéine mutée et la protéine non mutée.
- 3- Interpréter ces différences/ressemblances en termes de fonctionnement biologique et conclure sur la santé de personnes qui seraient homozygotes pour cette mutation.

Annexe n°1 : Protocole d'utilisation d'un outil de calcul de pourcentage de GC

- Adresse web : <http://www.endmemo.com/bio/gc.php>
- Apparence de l'outil :

The screenshot shows the ENDMEMO website interface for the DNA/RNA GC Content Calculator. The page title is "ENDMEMO" and the page content includes a search bar, a breadcrumb trail "Home > Biology > GC Content Calculation", a "PHP Tutorial" link, and the main heading "DNA/RNA GC Content Calculator". A large yellow text area is labeled "Please Paste the DNA/RNA Sequence". Below this area are two input fields: "GC Content:" with a percentage sign and "DNA Length:" with "bp". At the bottom, there are two buttons: "Calculate" and "Clear All". Annotations with arrows point to these elements: a red box labeled "Cadre où copier-coller la séquence à analyser" points to the yellow text area; a blue box labeled "Cadres de résultat" points to the two input fields; and a green box labeled "Bouton de lancement" points to the "Calculate" button.

- Protocole d'utilisation :
 - 1) Copier-coller la séquence à analyser dans le cadre prévu à cet effet.
 - 2) Cliquer sur le bouton de lancement "Calculate" situé au bas de l'écran.
 - 3) Lire les résultats dans les deux cadres situés au-dessus du bouton "Calculate".

Annexe n°2 : Principe de fonctionnement et utilisation de BLAST

- **Principe de l'outil :**

L'acronyme "BLAST" signifie en anglais "Basic Local Alignment Search Tool" ce qui pourrait être traduit par "Outil de Recherche par Alignement Local Basique". Cet outil permet en effet de rechercher, dans les bases de données, l'ensemble des séquences similaires à une séquence donnée.

Pour cela, cet outil va essayer d'aligner la séquence qu'on lui fournit aux différentes séquences contenues dans la base de données utilisée. Le résultat fourni par cet outil est donc une liste de séquences, présentes dans la base de données et rangées par ressemblance décroissante avec la séquence cherchée.

La méthode d'alignement est similaire à celle utilisée par ClustalW (Cf. annexe 3) et fournit donc différentes mesures de ressemblance des séquences entre elles. En outre, cette ressemblance peut être mesurée par le nombre de bases azotées qui sont parfaitement alignées ; c'est ce qu'on appelle le pourcentage d'identité.

- **Adresse web :**

http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome

- **Apparence de l'outil :**

- **Protocole d'utilisation :**

- 1) Copier-coller la séquence à chercher dans le cadre prévu à cet effet
- 2) Choisir la base de données à exploiter : cocher "Others (nr etc)" et choisir "Nucleotide collection (nr/nt)" dans le menu déroulant (cette base de données regroupe l'ensemble des ADN et ARN connus à l'heure actuelle).
- 3) Choisir le programme de recherche à exécuter : cocher "Highly similar sequences (megablast)" ; ce programme est optimal pour chercher des séquences très ressemblantes.
- 4) Cliquer sur le bouton de lancement "BLAST" puis attendre quelques instants les résultats sur la page qui se charge. La page s'actualisera automatiquement plusieurs fois avant d'afficher les résultats du BLAST.

- **Apparence des résultats :**

Les résultats s'affichent sur une page relativement longue ; il faut par conséquent faire défiler cette page pour accéder aux différentes parties des résultats. Une page de résultats est organisé, de haut en bas, selon le schéma suivant :

- 1- Récapitulatif de la recherche :



→ Cette partie des résultats rappelle notamment quelle séquence est recherchée, dans quelle base de données et grâce à quel programme.

- 2- Résumé graphique de la recherche :



→ Dans ce résumé graphique, les traits de couleur correspondent à un alignement entre la séquence recherchée et une séquence de la banque. D'autre part, la couleur correspond au score d'alignement et la longueur à la taille de l'alignement.

Annexe n°3 : Protocole d'utilisation de MUSCLE pour comparer des séquences nucléotidiques et peptidiques

- Adresse web : <http://www.ebi.ac.uk/Tools/msa/muscle/>
- Apparence de l'outil :

- Protocole d'utilisation :

1) Copier-coller les deux séquences à comparer dans le cadre prévu à cet effet. Attention, pour cet outil, les séquences doivent être précédées d'un en-tête de la forme ">Sequence_n"; cela permet de distinguer les deux séquences. Pour éviter toute erreur, le plus simple est de copier-coller les en-têtes fournis dans le fichier "ResultatSequencage.txt" en même temps que les séquences.

2) Cliquer sur le bouton de lancement "Submit" situé au bas de l'écran puis attendre quelques instants le chargement des résultats sur la page. La page s'actualisera automatiquement plusieurs fois avant d'afficher les résultats de la comparaison.

- Exemple de résultat :

```
Sequence_1      ATCGGC GGATATTGAGGACCCCTAGGA 27
Sequence_2      ATC--CGGATATTGAGGACCTTAGGA 25
                ***  *****  *****
```

N.B : Les astérisques "*" symbolisent les alignements exacts ; l'absence d'astérisque marque donc une différence entre les deux séquences. Les tirets "-" symbolisent pour leur part les "gaps", c'est à dire les espaces créés pour que des séquences de tailles différentes puissent s'aligner correctement.

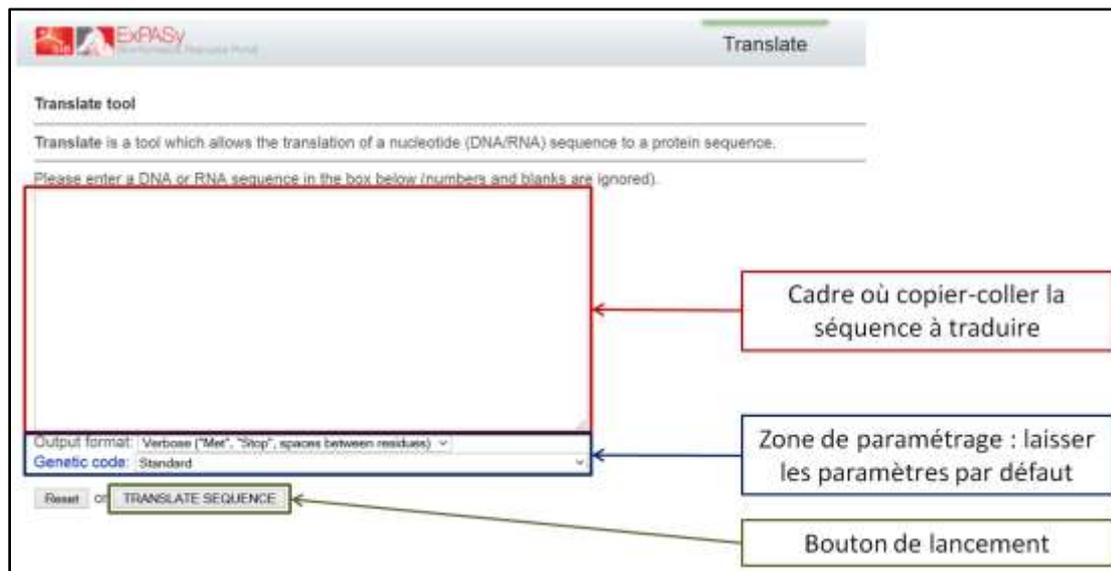
Annexe n°4 : Code génétique

		Deuxième nucléotide					
		U	C	A	G		
Premier nucléotide	U	UUU	UCU	UAU	UGU	U	
		UUC		UCC		UAC	UGC
		UUA	UCA	UAA	UAG	UGA	A
		UUG		UCG		Codon STOP	UGG
	C	CUU	CCU	CAU	CGU	U	
		CUC	CCC	CAC	CGC	C	
		CUA	CCA	CAA	CGA	A	
		CUG	CCG	CAG	CGG	G	
	A	AUU	ACU	AAU	AGU	U	
		AUC	ACC	AAC	AGC	C	
		AUA	ACA	AAA	AGA	A	
		AUG	ACG	AAG	AGG	G	
	G	GUU	GCU	GAU	GGU	U	
		GUC	GCC	GAC	GGC	C	
		GUA	GCA	GAA	GGA	A	
		GUG	GCG	GAG	GGG	G	

N.B : Le code à une lettre est noté entre parenthèses après chaque acide aminé.

Annexe n°5 : Protocole d'utilisation d'EspASy pour traduire des séquences nucléotidiques en séquences peptidiques

- Adresse web : <http://web.expasy.org/translate/>
- Apparence de l'outil :



- Protocole d'utilisation :
 - 1) Copier-coller la séquence à traduire dans le cadre prévu à cet effet. Attention, ici il ne faut **PAS** précéder la séquence d'un en-tête.
 - 2) Cliquer sur le bouton de lancement "TRANSLATE SEQUENCE" situé au bas de l'écran
 - 3) Les résultats présentent les traductions pour les 3 cadres de lectures possibles ("Frame 1/2/3") dans les deux sens de lectures possibles (5' → 3' et 3' → 5'), soit 6 traductions possibles en tout.
- Exemple de résultat :

Translate Tool - Results of translation

Open reading frames are highlighted in red. Please select one of the following frames - in the next page, you will be able to select your initiator and retrieve your amino acid sequence.

Sens 5'→3'	5'3' Frame 1	DRGLGRRRSPFRVFTLStopRASGGHHHLRQGPESStopIRHVGGAGGGSRRLRLGELStopIQSSHTLLREG R
	5'3' Frame 2	TEGWDEEDGHHSDSLHYEGRAVDITTSDDRDNKYGMetLARLAVEAGFDWVYYESKAHINOSVKA
	5'3' Frame 3	PRAGTKVTVTIPTSLEYIMetKGERWTRPPTTGTDTINTACWPQWRWKPASTDSTMetNPKLTYIAEStopRP
Sens 3'→5'	3'5' Frame 1	SAFTEQC MetStopALDS Stop StopTQSKPASTASLANMetPYLFRSLSEVVMetSTARPNStopCKDSEWStopF SSSQPSV
	3'5' Frame 2	RPSRSNVCELWIHSRPSRSLPPPAWPTCRIYSGPCR RW StopCPPLALHNVKTRRNGDLRPSPR
	3'5' Frame 3	GLHGA MetVYVSEGFIVDIVEAGFHRQPGQHAVFIPVYVGGGDVHRSPTIMetStopRLVG MetVTVFVPAUG

N.B : Les cadres ouverts de lecture, ou Open Reading Frame (ORF) sont surlignés en rouge

Document exploité au cours de ce TRAAM : "ResultatSequencage.txt"

>Sequence_1

```
CAACTCCTAAGCCAGTGCCAGAAGAGCCAAGGACAGGTACGGCTGTCATCACTTAGACCTCACCTGTG
GAGCCACACCCTAGGGTTGGCCAATCTACTCCCAGGAGCAGGGAGGGCAGGAGCCAGGGCTGGGCATAAA
AGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGACACAACCTGTGTTCACTAGCAACCTCAAACA
GACACCATGGTGACCTGACTCCTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGG
ATGAAGTTGGTGGTGAGGCCCTGGGCAGGTTGGTATCAAGGTTACAAGACAGGTTAAAGGAGACCAATAG
AACTGGGCATGTGGAGACAGAGAAGACTCTTGGGTTTCTGATAGGCACTGACTCTCTGCTATTGGT
CTATTTTCCCACCCTTAGGCTGCTGGTGGTCTACCCTGGACCCAGAGGTTCTTTGAGTCTTTGGGGAT
CTGTCCACTCCTGATGCTGTTATGGGCAACCCTAAGGTGAAGGCTCATGGCAAGAAAGTGCCTCGGTGCCT
TTAGTGATGGCCTGGCTCACCTGGACAACCTCAAGGGCACCTTTGCCACACTGAGTGAGCTGCACTGTGA
CAAGCTGCACGTGGATCCTGAGAACCTCAGGGTGAGTCTATGGGACCCTTGATGTTTTCTTTCCCTTCT
TTTCTATGGTTAAGTTCATGTCATAGGAAGGGGAGAAGTAACAGGGTACAGTTTAGAATGGGAAACAGAC
GAATGATTGCATCAGTGTGGAAGTCTCAGGATCGTTTTAGTTTCTTTTATTTGCTGTTCAACAATTGT
TTTTTTTTGTTAATTCTTGCTTTCTTTTTTTTTCTTCCGCAATTTTTACTATTATACTTAATGCCTT
AACATTGTGTATAACAAAAGGAAATATCTCTGAGATACATTAAGTAACTAAAAAAAAACTTTACACAGT
CTGCCTAGTACATTACTATTTGGAATATATGTGTGCTTATTTGCATATTCATAATCTCCCTACTTTATTT
TCTTTTATTTTAATTGATACATAATCATTATACATATTTATGGGTTAAAGTGTAAATGTTTTAATATGTG
TACACATATTGACCAAATCAGGGTAATTTTGCATTTGTAATTTAAAAAATGCTTTCTTTTAATATA
CTTTTTGTTTATCTTATTCTAATACTTTCCCTAATCTTTTCTTTTCCAGGGCAATAATGATACAATGTA
TCATGCCTCTTTGCACCATTTCTAAAGAATAACAGTGATAATTTCTGGGTTAAGGCAATAGCAATATTTCT
GCATATAAATATTTCTGCATATAAATTGTAAGTATGTAAGAGGTTTTCATATTGCTAATAGCAGCTACAA
TCCAGCTACCATTCTGCTTTTATTTATGGTTGGGATAAGGCTGGATTATTCTGAGTCCAAGCTAGGCC
TTTTGCTAATCATGTTACACCTCTTATCTTCTCCACAGCTCCTGGGCAACGTGCTGGTCTGTGTGCT
GGCCCATCACTTTGGCAAAGAATTCACCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTG
GCTAATGCCCTGGCCACAAGTATCACTAAGCTCGTTTTCTGCTGTCCAATTTCTATTAAAGTTCTTT
TGTTCCCTAAGTCCAACCTACTAACTGGGGGATATTATGAAGGGCCTTGAGCATCTGGATTCTGCCTAAT
AAAAACATTTATTTTTCATTGCAATGATGTATTTAAATTTTCTGAATATTTTACTAAAAAGGGAATGT
GGGAGGTCAGTG
```

>Sequence_2

```
CAACTCCTAAGCCAGTGCCAGAAGAGCCAAGGACAGGTACGGCTGCATCACTTAGACCTCACCCCTGTG
GAGCCACACCTAGGGTTGGCCAATCTACTCCCAGGAGCAGGGAGGGCAGGAGCCAGGGCTGGGCATAAA
AGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGACACAACCTGTGTTCACTAGCAACCTCAAACA
GACACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGG
ATGAAGTTGGTGGTGGAGCCCTGGGCAGGTTGGTATCAAGGTTACAAGACAGGTTAAGGAGACCAATAG
AAACTGGGCATGTGGAGACAGAGAAGACTCTTGGGTTTCTGATAGGCACTGACTCTCTCTGCCTATTGGT
CTATTTTCCCACCCCTTAGGCTGCTGGTGGTCTACCCTGAACCCAGAGGTTCTTTGAGTCTTTGGGGAT
CTGTCCACTCCTGATGCTTATGGGCAACCCTAAGGTGAAGGCTCATGGCAAGAAAAGTCTCGGTGCCT
TTAGTGATGGCCTGGCTCACCTGGACAACCTCAAGGGCACCTTTGCCACACTGAGTGAGCTGCACTGTGA
CAAGCTGCACGTGGATCCTGAGAACTTCAGGGTGAGTCTATGGGACCCCTGATGTTTTCTTCCCCTTCT
TTTCTATGGTTAAGTTCATGTCATAGGAAGGGGAGAAGTAAACAGGGTACAGTTAGAAATGGGAAACAGAC
GAATGATTGCATCAGTGTGGAAGTCTCAGGATCGTTTTAGTTTTCTTTTATTGCTGTTTATAACAATTGT
TTTTTTTTGTTAATTCTTCTTTTTTTTTTTTTCTTCTCCGAATTTTTACTATTATACTTAATGCCTT
AACATTGTGTATAACAAAAGGAAATATCTCTGAGATACATTAAGTAACTTAAAAAAAAAATTTACACAGT
CTGCCTAGTACATTAATTTGGAATATATGTGTGCTTATTTGCATATTCATAATCTCCCTACTTTATTT
TCTTTTATTTTAAATTGATACATAATCATTATACATATTTATGGGTTAAAGTGAATGTTTTAATATGTG
TACACATATTGACCAATCAGGGTAATTTGCATTTGAATTTAAAAAATGCTTCTCTTTAATATA
CTTTTTGTTTATCTTATTCTAATACTTTCCCTAATCTCTTTCTTTAGGGCAATAATGATACAATGTA
TCATGCCTCTTTGCACCATTTCTAAAGAATAACAGTGATAATTTCTGGGTTAAGGCAATAGCAATTTTCT
GCATATAAATATTTCTGCATATAAATTGAACTGATGAAGAGGTTTCATATTGCTAATAGCAGCTACAA
TCCAGCTACCATTCTGCTTTTATTTTATGGTTGGGATAAGGCTGGATTATTCTGAGTCCAAGCTAGGCC
TTTTGCTAATCATGTTACATACCTTATCTTCTCCACAGCTCCTGGGCAACGTGCTGGTCTGTGTGCT
GGCCACTCTTTGGCAAAGAATTCACCCACAGTGCAAGGCTGCCTATCAGAAAAGGTTGGTGGTGGTGTG
GCTAATGCCTGGCCCAAGTATCACTAAGCTCGTTTTCTTGCTGTCCAATTTCTATTAAGGTTCTCTT
TGTTCCCTAAGTCCAACACTAACTGGGGGATATTATGAAGGGCCTTGAGCATCTGGATTCTGCCTAAT
AAAAAACATTTATTTTCTTGAATGATGATTTAAATTTTCTGAATATTTTACTAAAAAGGGAATGT
GGGAGGTCAGTG
```

>Sequence_3

```
ATAGAGTTTGATAATCGAGAGTTAATGATAACTTTTACAGTGAGGCATTTCAAACAATTTGAGGTGTTTT
CTGTTTATACAGATGTGAAAGCGAAGAAAAATTATACCGTATGTTTACTAGACTTTAGTCAATTTGGATAA
AGGTTTTGTGCAAAGTTTATGAAGATACAGGTTTCGTTTAAATATATGTTAAATTTTAACTAAAAAATTA
GAGTTTAGTGAATTTAAAAATTAAGTTTTTCAAATGATTGTCATGTCACGTTAAATTTGTTATACAAT
GTAGAAAATTTTAGTAAAGTAAAGTCTTATGTGGGACATTAATAAATTTGCGATAAATTTCAAAT
TATTGGTATGCTTACTATATTGAAAAAAGGGGGTTCCTATGAAAAACAATTTATTCTTTATTT
TAATATATTTCTTATTTTTAGGGATTGGATTAGTTTCTGTACTTCTGTATATTTGAAGGATTTA
GGATTAAGGTTAGTGACTTAGGAATGCTAGTTGCTGCTTTTGCATTATCACAAATGATTATTTACCAT
TTGGTGGGACACTAGCTGATAAATTTGGTAAAAAATTAATTTATGATCGGTTTAGTATTCTTTGCTGT
CTCTGAATTTATGTTGCGAGCCGGTCAAAGTTTTACCATTTAATCATTTACGTTGTTTAGGTGGCTTT
AGTGCAGGCATGGTCATGCCTGGTGAACAGGTATGATTGCAGATATTTCTCCAGGAGCTGATAAAGCTA
AAAATTTGGTTACATGTCGGCAATTTAATTCAGGTTTTATATTAGGACCTGGATTTGGAGGCTTTTT
AGCTGAAATTTACATAGATTACCTTTCTATGTTGCTGGAACATTAGGTGTTGTTGCATTATTATGTCA
GTTTTATTAATTCATAATCCTCAAAAAGCAACTACAGATGGATTCCACCAATATCAACCTGAATTTTCA
CTAAAATTAATTTGAAAGTATTTTACTCCAGTCATATTAACACTGTATTAGCATTGGTTTATCTGC
TTTTGAAACATTTTCTTATATACAGCTGACAAAAGTAAATATACTCCTAAAGATATTTGATAGCT
ATTATCGGTGGAGGCGTGTGGCGCATTATCCAAGTATTCTTCTTTGATAAATTTATGAAATATATGA
GTGAACCTAATTTTATTGCATGGTCTTACTATATTCAGCCATTGTTCTCGTTATGTTAGTCTTGCAA
CGTTTATTGGACGATTATGATTATTAGCTTTGTTGTTTTATAGGTTTTGATATGATTAGACCAGCTTTA
ACCAATTAATTTCTCGAATATAGCAGGCAAACGGCAAGGTTTTGCAGGTGGATTGAATTTCACTTTTACCA
GTATGGGTAATTTATAGTCTCTTGTAGCTGGTGCATTATTCGATGTTAATTTAGAGTTTCTTTATA
TATGGCTATTGCGGTTTCATTAAGTGAATTTATCATTATTTTATTGAAAAAGGACTTAAAGTACGCGCT
AAAGAAGCAAATTAATACGGGGCCCAACAAAGAGAATTTCCGCCGAGAAATTTACGGACAGAGCAAGTT
TTGGGGACGAGGTGGGACACAGTGTCAAGACGAATTTCTGTGCCACTTCTTTGTTTTATACTTTGTA
ATGTAAAGTTGATGTTATCTCAAATATAAGTTTTAAAGATTATGAGTTATTAGTGGTTTATGTAACAAA
AATAACAATTAATTAATGTTAGTGTACAAAATTAACAATCTATTTAAATTTGAATCTACATATCTTTT
ATTAAGTTGTTGTTAAGATAGAATAAATCTTAAGAGAAATG
```