Université Mohamed KHEIDER Biskra
Faculté des Sciences Exactes et Sciences de la Nature et de le Vie
Département des Sciences de la Nature et de la Vie
1ère Année mastère : Biochimie Fondamentale et Appliquée, Microbiologie et Biologie
végétale.
Module : Bioinformatique
TP05 : Annotation des gènes et conception des amorces

Introduction

Dans ce TP vous allez prédire la présence éventuelle de gènes dans la séquence étudiée et, si vous en trouvez, caractériser leur structure exons-introns. On distingue deux types d'approches pour prédire des gènes :

-les approches comparatives : elles prédisent des gènes dans une séquence en cherchant des similarités avec des gènes déjà connus le programme "Blast" plus loin) ;

-les approches *ab initio* : elles ne comparent pas la séquence à des gènes connus, mais recherchent des éléments caractéristiques des gènes : cadre ouvert de lecture (=ORF), séquence promotrice consensus, séquences d'épissage... Cela nécessite de bien connaître ces éléments a priori.

1. La prédiction des exons et des introns

La séquence à analyser sera récupérée de la banque de données ''Nucleotide'' (nous choisissons la séquence de l'insuline de l'espèce : *Octodon degus*, sous format FAST).

```
>M57671.1 Octodon degus insulin mRNA, complete cds
GCATTCTGAGGCATTCTCTAACAGGTTCTCGACCCTCCGCCATGGCCCCGTGGATGCATCTCCTCACCGT
GCTGGCCCTGCTGGCCCTCTGGGGACCCAACTCTGTTCAGGCCTATTCCAGCCAGCACCTGTGCGGCTCC
AACCTAGTGGAGGCACTGTACATGACATGTGGACGGAGTGGCTTCTATAGACCCCACGACCGCCGAGAGC
TGGAGGACCTCCAGGTGGAGCAGGCAGAACTGGGTCTGGAGGCAGGCGGCCTGCAGCCTTCGGCCCTGGA
GATGATTCTGCAGAAGCGCGGCATTGTGGATCAGTGCTGTAATAACATTTGCACATTTAACCAGCTGCAG
AACTACTGCAATGTCCCTTAGACACCTGCCTTGGGCCTGGCCTGCTGCTCTGCCCTGGCAACCAATAAAC
CCCTTGAATGAG
```

-Pour utiliser le logiciel ''GENSCAN'' afin de prédire le nombre d'exons d'un gène, il faut taper l'URL suivant : http://argonaute.mit.edu/GENSCAN.html

- Dans la case vide, on va insérer notre séquence à analyser.

- On laisse tous les paramètres par défaut (sélectionnez l'organisme ''Vertebrate'', ensuite nous cliquons sur ''Run GENSCAN''

- Un résultat de recherche va apparaitre :

Abréviations :

Gn. Ex : indique le numéro du gène, numéro d'exon.

INIT : initial exon

INTR :internal exon

TERN :terminal exon

SNGL :Single Exon Gene

PROM:Promoteur

PLYA : Poly A Signal

S : le brin d'ADN (+/-)

BEGIN: le début de l'exon

END: la fin de l'exon

LEN : longueur de l'exon

FR : signifie (le cadre ouvert de lecture)

2. Conception d'amorces

Primer3 est un outil très utilisé pour dessiner des amorces pour PCR. Cette technique est utilisée pour de nombreux objectifs différents. Par conséquent, primer3 a de nombreux paramètres d'entrée différents que vous contrôlez et qui indiquent exactement à primer3 quelles caractéristiques constituent de bonnes amorces pour vos objectifs.

-Pour utiliser le logiciel ''Primer3'' afin de créer un couple d'amorces pour la PCR, il faut taper l'URL suivant : http://primer3.ut.ee

Suite :

Toujours on va analyser la séquence de l'insuline de l'espèce : *Octodon degus*, sous format FAST. C-à-d on va créer un couple d'amorces pour amplifier cette séquence (PCR), donc on insère la séquence, puis on clique sur le bouton **(Pick Primers).**

TH: Max Primer Hairpin | 24.0

Thermodynamic Template Alignments
☐ Use Thermodynamic Template Alignment
TH: Max Template Mispriming | 40.00
TH: Pair Max Template Mispriming | 70.00

Old Template Alignments
Max Template Mispriming | 12.00
Pair Max Template Mispriming | 24.00

Max #N's accepted | 0   Max Poly-X | 4
Inside Target Penalty | -1.0   Outside Target Penalty | 0   Note: you can set Inside Target Penalty to allow primers inside a target.
First Base Index | 1   CG Clamp | 0
Max GC in primer 3' end | 5
3' End Distance Between Left Primers | 3   3' End Distance Between Right Primers | 3
5 Prime Junction Overlap | 7   3 Prime Junction Overlap | 4   (Distance of the primer ends to one overlap position.)
Concentration of Monovalent Cations | 50.0   Salt Correction Formula | SantaLucia 1998 ▼
Concentration of Divalent Cations | 1.5   Concentration of dNTPs | 0.6
Annealing Oligo Concentration | 50.0   (Not the concentration of oligos in the reaction mix but of those annealing to template.)
Sequencing Spacing | 500   Sequencing Interval | 250
Sequencing Lead | 50   Sequencing Accuracy | 20

☑ Liberal Base   ☐ Show Debuging Info   ☐ Treat ambiguity codes in libraries as consensus
☐ Lowercase masking   ☐ Pick anyway   ☑ Print Statistics

[Pick Primers] [Download Settings] [Reset Form]

Objective Function Penalty Weights for Primers

Size   Lt 1.0   Gt 1.0
Tm   Lt 1.0   Gt 1.0

Primer3web version 4.1.0 - Pick primers from a DNA sequence.   disclaimer   code
                                                                          cautions

Select the Task for primer selection: generic ▼

Template masking before primer design (available species)
Select species | Example: Mus musculus   Nucleotides to mask in 5' direction | 1
Primer failure rate cutoff   < 0.1   Nucleotides to mask in 3' direction | 0

Paste source sequence below (5'->3', string of ACGTNacgtn -- other letters treated as N -- numbers and blanks ignored). FASTA format ok. Please N-out undesirable sequence (vector, ALUs, LINEs, etc.) or use a Mispriming Library (repeat library) NONE ▼

```
>M57671.1 Octodon degus insulin mRNA, complete cds
GCATTCTGAGGCATTCTCTAACAGGTTCTCGACCCTCCGCCATGGCCCCGTGGATGCATCTCCTCACCGT
GCTGGCCCTGCTGGCCCTCTGGGGACCCAACTCTGTTCAGGCCTATTCCAGCCAGCACCTGTGCGGCTCC
AACCTAGTGGAGGCACTGTACATGACATGTGGACGGAGTGGCTTCTATAGACCCCACGACCGCCGAGAGC
TGGAGGACCTCCAGGTGGAGCAGGCAGAACTGGGTCTGGAGGCAGGCGGCCTGCAGCCTTCGGCCCTGGA
GATGATTCTGCAGAAGCGCGGCATTGTGGATCAGTGCTGTAATAACATTTGCACATTTAACCAGCTGCAG
```

☑ Pick left primer, or use left primer below   ☐ Pick hybridization probe (internal oligo), or use oligo below   ☑ Pick right primer, or use right primer below (5' to 3' on opposite strand)

[Pick Primers] [Download Settings] [Reset Form]

Sequence Id | [   ]   A string to identify your output.
Targets | [   ]   E.g. 50,2 requires primers to surround the 2 bases at positions 50 and 51. Or mark the source sequence with [and]: e.g. ...ATCT[CCCC]TCAT.. means that primers must flank the central CCCC.
Overlap Junction List | [   ]   E.g. 27 requires one primer to overlap the junction between positions 27 and 28. Or mark the source sequence with -: e.g. ...ATCTAC-TGTCAT.. means that primers must overlap the junction between the C and T.
Excluded Regions | [   ]   E.g. 401,7 68,3 forbids selection of primers in the 7 bases starting at 401 and the 3 bases at 68. Or mark the source sequence with < and >: e.g. ...ATCT<CCCC>TCAT.. forbids primers in the central CCCC.
Pair OK Region List | [   ]   See manual for help.

Le résultat :

Un couple d'amorces apparait (l'amorce gauche et droite) aux niveaux de 2 extrémités de la séquence à amplifier. Aussi, les différentes caractéristiques de chaque amorce (TM, %GC, longueur, début, fin….).

Primer3 Output

```
PRIMER PICKING RESULTS FOR M57671.1 Octodon degus insulin mRNA, complete cds

Template masking not selected
No mispriming library specified
Using 1-based sequence positions
OLIGO            start  len      tm     gc%  any_th  3'_th  hairpin  seq
LEFT PRIMER        179   20   58.59   55.00   13.89   0.00     0.00  TGGCTTCTATAGACCCCACG
RIGHT PRIMER       383   20   59.02   55.00    0.00   0.00     0.00  CAAGGCAGGTGTCTAAGGGA
SEQUENCE SIZE: 432
INCLUDED REGION SIZE: 432

PRODUCT SIZE: 205, PAIR ANY_TH COMPL: 0.00, PAIR 3'_TH COMPL: 0.00

     1 GCATTCTGAGGCATTCTCTAACAGGTTCTCGACCCTCCGCCATGGCCCCGTGGATGCATC


    61 TCCTCACCGTGCTGGCCCTGCTGGCCCTCTGGGGACCCAACTCTGTTCAGGCCTATTCCA


   121 GCCAGCACCTGTGCGGCTCCAACCTAGTGGAGGCACTGTACATGACATGTGGACGGAGTG
                                                                  >>

   181 GCTTCTATAGACCCCACGACCGCCGAGAGCTGGAGGACCTCCAGGTGGAGCAGGCAGAAC
         >>>>>>>>>>>>>>>>>>

   241 TGGGTCTGGAGGCAGGCGGCCTGCAGCCTTCGGCCCTGGAGATGATTCTGCAGAAGCGCG


   301 GCATTGTGGATCAGTGCTGTAATAACATTTGCACATTTAACCAGCTGCAGAACTACTGCA


   361 ATGTCCCTTAGACACCTGCCTTGGGCCTGGCCTGCTGCTCTGCCCTGGCAACCAATAAAC
         <<<<<<<<<<<<<<<<<<<<

   421 CCCTTGAATGAG
```