# 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).



-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 :

Sequence	/tmp/	/04_27_26	9-15:4	1:54.	fast	a : 46	7 bp :	58.54	4% C+G	: Isocho	re 4 (1	57 -	100	C+G%)	
Paramete	r matr	ix: Huma	anIso.	smat											
Predicte	d gene	es/exons	:												
Gn.Ex Ty	pe S .	Begin .	End	. Len	Fr Pl	n I/Ac	Do/T	CodRg	P	Tscr					
1.01 Sn	gl +	77	406	330	1	9 106	42	457	0.999	36.81					
1.02 Pl	yA +	449	454	6						1.05					
Suboptim	al exc	ons with	proba	bilit	:y > :	1.000									

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 : <u>http://primer3.ut.ee</u>

Duine on 2 mal									
PTIIIIEI3WED version 4.1.0 - Pic	k primers from a DNA sequence.	cautions							
select the Task for primer selection generic	T								
Template masking before primer design (availab	le species)	1							
Select species Example: Mus musculus Nuc	leotides to mask in 5' direction 1								
Primer failure rate cutoff < 0.1 Nuc	leotides to mask in 3' direction 0								
taste source sequence oerow (55, sumg of AC	NONE v		gnorea). FAS IA Ionnal ok, Frease IN-out undestrable	sequence (vector, ALUS	, LINES,				
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 <u>source sequence</u> with [ and ]: e.gATCT[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 <u>source sequence</u> with -: e.gATCTAC-TGTCAT means that primers must overlap the junction between the C and T.								
Excluded Regions	E.g. 401,768,3 forbids selection of primers in the 7 bases starting at 401 and the 3 bases at 68. Or mark the <u>source sequence</u> with < and >: e.gATCT <cccc>TCAT forbids primers in the central CCCC.</cccc>								
Pair OK Region List	See manual for help								

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				
The second secon	Old Template Alignments				
	old rempilter mgimients				
TH: Max Template Mispriming	40.00	12.00			
TH. Dair May Template Misprinning	70.00 Max Template Mispriming	12.00			
TH. Pair Max Template Misprinning	Pair Max Template Misprimi	ng 24.00			
Max #N's accepted 0	Max Poly-X	4			
Inside Target Penalty -1.0	Outside Target Penalty	0 <u>Note: yo</u>	a can set Inside Target Penalty to all	ow 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 Primer	<u>rs</u> 3			
5 Prime Junction Overlap 7	3 Prime Junction Overlap	4 (Distanc	e of the primer ends to one overlap p	osition.)	
Concentration of Monovalent Cations 50.0	Salt Correction Formula	SantaLucia 1998	T		
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 t	hose annealing to template.)		
Sequencing Spacing 500	Sequencing Interval	250			
Sequencing Lead 50	Sequencing Accuracy	20			
Liberal Base Show Debuging	Info 💷 <u>Treat ambiguity codes in libraries</u>	as consensus			
🔲 <u>Lowercase masking</u> 🖾 <u>Pick anyway</u>	Print Statistics				
Pick Primers Download Settings Reset Fo	rm				
Objective Function Penalty Weights	for Primers				
Size Lt 1.0 Gt 1.0					
Tm Lt 1.0 Gt 1.0					
Primer3web version 410 -	Pick primers from a DNA sequence			disclaimer	code
Primer3web version 4.1.0 -	Pick primers from a DNA sequence.			disclaimer cautions	code
Primer3web version 4.1.0 - Select the Task for primer selection generic	Pick primers from a DNA sequence.			disclaimer cautions	code
Primer3web version 4.1.0 - Select the Task for primer selection generic Template masking before primer design (ava	Pick primers from a DNA sequence.	]		disclaimer cautions	code
Primer3web version 4.1.0 - Select the Task for primer selection generic Template masking before primer design (ava Select species Example: Mus musculus	Pick primers from a DNA sequence.	]		disclaimer cautions	code
Primer3web         version 4.1.0           Select the Task for primer selection         generic           Template masking before primer design (ava Select species Example: Mus musculus Primer failure rate cutoff         1	Pick primers from a DNA sequence.			disclaimer cautions	<u>code</u>
Primer3Web version 4.1.0 - Select the Task for primer selection generic Template masking before primer design (ava Select species Example: Mus musculus Primer failure rate cutoff <[0.1]	Pick primers from a DNA sequence.	- numbers and blanks	ignored). FASTA format ok. Please 1	N-out undesirable sequence (vector, ALI	Js, LINEs,
Primer3Web         version 4.1.0 -           Select the Task for primer selection generic	Pick primers from a DNA sequence.	- numbers and blanks	ignored). FASTA format ok. Please :	N-out undesirable sequence (vector, ALU	Js, LINEs,
Primer3Web         version 4.1.0 -           Select the Task for primer selection generic	Pick primers from a DNA sequence.	- numbers and blanks	ignored). FASTA format ok. Please :	N-out undesirable sequence (vector, ALU	Js, LINEs,
Primer3web         version 4.1.0           Select the Task for primer selection         generic           Template masking before primer design (non Select species Example: Mus musculus Primer failure rate cutoff < 0.1	Pick primers from a DNA sequence.	- numbers and blanks	ignored). FASTA format ok. Please ]	N-out undesirable sequence (vector, ALU	Js, LINEs,
Primer3web         version 4.1.0           Select the Task for primer selection         generic           Template masking before primer design (ava Select species Example: Mus musculus Primer failure rate cutoff         1           Primer failure rate cutoff         0.1         1           Paste source sequence below (5'~>3', string of etc.) or use a Mispriming Library (repeat library CATTCTRAGGCATTCTARACAGTTCTGRACCTCGTGGCCCTC GGATGCAGGCACTCTCTAGGCACATCTGTGGCGGGATT TGGAGGACCTCCAGGTGGGCCGGCAGGCATGGGCTGGGC	Pick primers from a DNA sequence.	- numbers and blanks	ignored). FASTA format ok. Please l	N-out undesirable sequence (vector, ALU	Js, LINEs,
Primer3web version 4.1.0 - Select the Task for primer selection generic Template masking before primer design (ava Select species Example: Mus musculus Primer failure rate cutoff < 0.1 2 Paste source sequence below (5'->5', string of etc.) or use a Mispriming Library (repeat libro VM57671.1 Octodon degus insulin mRNA, co cattrictagecattrictraAccastrictecacctres actoradescattrictraAccastrictecacctres coreseccreates accasaccasaccastricters datacctadescattrictrafaccatristecacted artistatescatter actagecates accasaccastricter tesasacctrictocasccasaccasacted actagecter datactates accasaccasaccasacted actagecter datasaccastricter accastricter actagecates datasatricter actagecates accasactas actagecter datasatricter actagecates accasactas actagecates datasatricter actagecates accasaccasacted actagecates datasatricter actagecates accasactas actagecates actagecates accasaccasaccasaccasaccasaccasaccasac	Pick primers from a DNA sequence.	- numbers and blanks	ignored). FASTA format ok. Please	N-out undesirable sequence (vector, ALU	Js, LINEs,
Primer3web version 4.1.0 - Select the Task for primer selection generic Template masking before primer design (ava Select species Example: Mus musculus Primer failure rate cutoff < 0.1 1 Paste source sequence below (5'->3', string of etc.) or use a Mispriming Library (repeat libra >M57671.1 Octodon degus insulin mRNA, co cattrictradsocrittorCataceTorGataceAttrictorAcceCos cattroctageCatTrictTacAcasTrictGataceCattrict GataGattrictageCatTrictTacAcasTrictGataceGattrict GataGattrictageCatTrictAcataceGataCettorGataceGattrictorGataceCattrict WB Pick left primer, or use left primer below	Pick primers from a DNA sequence.	- numbers and blanks	ignored). FASTA format ok. Please i ∭ Pick right primer, or use right pr	N-out undesirable sequence (vector, ALU	Js, LINEs,
Primer3web version 4.1.0 - Select the Task for primer selection generic Template masking before primer design (ava Select species Example: Mus musculus Primer failure rate cutoff < 0.1 1 Primer failure rate cutoff < 0.1 1 Paste source sequence below (5'->3', string of etc.) or use a Mispriming Library (repeat libra cattricra6aGactricrafacaBarticrasaccae actrocra6aGactricrafacaBarticrasaccae GartBarticra6aGactricrafacaTrifBarcaGatrifGBarcaBarticra GartBarticra6aGaccaftGartGartGBarcaBartGatrifGBarcaBarter GaraGattCra6aGacGaccaftGartGartGBarcaBarter Barbarticra6aCaftBartaCaftGBarcaBarter Barbarter Pibs/Berting Description (String Section 1) Pibs/Berting Description (Section 2) Pibs/Berting Description (Secti	Pick primers from a DNA sequence.	- numbers and blanks	ignored). FASTA format ok. Please i Pick right primer, or use right pr	N-out undesirable sequence (vector, ALU	Js, LINEs,
Primer3web         version 4.1.0           Select the Task for primer selection         generic           Template masking before primer design (ava Select species Example: Mus musculus Primer failure rate cutoff         1           Primer failure rate cutoff         0.1         1           Paste source sequence below (5'->5', string of etc.) or use a Mispriming Library (repeat libra ActCrATGAGGCATTCTAACAGGTUTCTGACCTCGG GCT6GCCTGGCCCTCTGGGACCAGCAGCAGCTCGGGCTGG GCT6GCCTGCTGGGCCCTCTGGGAGCCAGCAGCAGGGCTGG GATGATCTGCAGGGCATTCTAACAGTUTGGAACCAGGGCTGG GATGATCTGCGAGGCCGGCATTGTGGATCAGTGGGGCGG GT6GCCTGCTGGGCCCTGTGGGAGCCAGGCCTGGG GATGATCTGCGAGGCCGGCATTGTGGATCAGTGCTGG GATGATCTGCGAGGCGGCATTGTGGATCAGTGGGTGG GATGATCTGCGAGGCGGCATTGTGGATCAGTGGGCGGCTGG GATGATCTGCGAGGCGGCATTGTGGATCAGTGGGACCAGGCCTGG GATGATCTGCGAGGCGGCATTGTGGATCAGTGGGACCAGGCCTGG GATGATCTGCGAGGGCGGCATTGTGGATCAGTGGGACCAGGCCTGG GATGATCTGCGAGGCGGCATTGTGGATCAGTGGGACCAGGCCTGG GATGATCTGCGAGGCGGCATTGTGGATCGGCGGCATTGTGGATCGGGCGGCTGGGCGCGCGC	Pick primers from a DNA sequence.	), or use oligo below	ignored). FASTA format ok. Please i Ø Pick right primer, or use right pr	N-out undesirable sequence (vector, ALU	Js, LINEs,
Primer3web version 4.1.0 - Select the Task for primer selection generic Template masking before primer design (ava Select species Example: Mus musculus Primer failure rate cutoff < 0.1 Paste source sequence below (5'->3', string of etc.) or use a Mispriming Library (repeat library PM57671.1 Octodon degus insulin mRNA, occ cattrictadsGacttrictarAcaBottrictGacCatcristicad ActCrATGBAGGACTTICTARACABOTTCGGACCATCGTGGACCAGGACTGG GatTGATTCTGAGGACATTGTGACATGGATCAGGCTGG GatTGATTCTGAGGACGGGGCATTGTGGATCAGGGCTGG GatGattCTGGAGGACGGGGCATTGTGGATCAGGCTGG GatGattCTGGAGAGGCGGGCATTGTGGATCAGGCTGG GatGattCTGGAGAGGCGGGCATTGTGGATCAGGGCTGG GatGattCTGGAGAGGCGGGCATTGTGGATCAGGGCTGG GatGattCTGGAGAGGCGGGCATTGTGGATCAGGCTGG Pick Primers Download Settings Reset For Sequence Id	Pick primers from a DNA sequence.	- numbers and blanks	ignored). FASTA format ok. Please i	N-out undesirable sequence (vector, ALU rimer below (5' to 3' on opposite strand)	Js, LINEs,
Primer3Web version 4.1.0 - Select the Task for primer selection generic Template masking before primer design (ava Select species Example: Mus musculus Primer failure rate cutoff < 0.1 Pasts source sequence below (5'->3', string of etc.) or use a Mispriming Library (repeat libra PMS7671.1 Octodon degus insulin mRNA, occ actificataGaCATICTGACAGGTTCGACAGCACTGGTAGGACCAGGCATTGTGACAGGTTCGGACCAGGCATTGTGAGGACCTGGATCGGATCGGATGGGACCAGGCATTGTGAGGACCTGGGACCAGGCATTGTGAGGACCTGGATCGGATCGGATCGGATCGGATCGGATCGGATCGGATCGGATGGGACCAGGCCTGGATCGGA	Pick primers from a DNA sequence.	- numbers and blanks	ignored). FASTA format ok. Please i Pick right primer, or use right pri	disclaimer         cautions         N-out undesirable sequence (vector, ALU         rimer below (5' to 3' on opposite strand)	Js, LINEs,
Primer3Web       version 4.1.0 -         Select the Task for primer selection generic         Template masking before primer design (ava         Select species       Example Mus musculus         Primer failure rate cutoff       0.1         Primer failure rate cutoff       0.1         PMS7671.1       Octodon degus insulin mRNa; cccs         Caracter cassocarter cranacharter stances caracter stances caracter cassocarter cranacharter stances caracter cassocarter cranacharter stances caracter stances caracter stances caracter cassocarter caracter stances caracter	Pick primers from a DNA sequence.	- numbers and blanks	ignored). FASTA format ok. Please Pick right primer, or use right pr as 50 and 51. Or mark the <u>source seq</u>	disclaimer         cautions         N-out undesirable sequence (vector, ALU         rimer below (5' to 3' on opposite strand)	Js, LINEs, TCAT
Primer3web       version 4.1.0 -         Select the Task for primer selection       generic         Template masking before primer design (ava         Select species       Example: Mus musculus         Primer failure rate cutoff       0.1         Primer failure rate cutoff       1         Paste source sequence below (5'-23', string of etc.) or use a Mispriming Library (repeat library Creptor)       1         >M57671.1 Octoon degus insulin mRNA, coccatter tradsaccatter tradsacatter tradsaccatter tradsacatter tradsaccatter tradsacter tradsacter tradsacter tradsacter tradsacter tradsacter tradsacter tradsaccatter tradsacter tradsacatter tradsacter tradsacter tradsacater tradsacter tr	Pick primers from a DNA sequence.	- numbers and blanks	ignored). FASTA format ok. Please 1 Pick right primer, or use right pr as 50 and 51. Or mark the <u>source seq</u> positions 27 and 28. Or mark the <u>sou</u>	disclaimer         cautions         N-out undesirable sequence (vector, ALU         rimer below (5' to 3' on opposite strand)         uence with [ and ]: e.gATCT[CCCC]         urce sequence with -: e.gATCTAC-TC	Js, LINEs, TCAT
Primer3web       version 4.1.0         Select the Task for primer selection       generic         Template masking before primer design (ava Select species Example: Mus musculus Primer failure rate cutoff       1         Primer failure rate cutoff       0.1       1         Paste source sequence below (5'->3', string of etc.) or use a Mispriming Library (repeat library Catter GAGGCATTECTAACAGGTETCEGACCAGCTEGGCCTGGCCCTCTGGGCCCTCTGGGCCCTCTGGGACCAGCAGGCCTGGATCAGTGGGACCAGGCCTGG GATGGTCTGGCAGCAGGCCGGCATTGTGGATCAGTGCTGG GATGGTCTCCAGGAGGCGGCATTGTGGATCAGGGCTGG GATGGTCTGCAGGAGGCGGGCATTGTGGATCAGTGCTGG GATGGTCTGCAGGAGGCGGGCATTGTGGATCAGTGCTGG GATGGTCTGCAGGAGGCGGGCATTGTGGATCAGTGCTGG GATGGTCTGCAGGAGGCGGGCATTGTGGATCAGTGCTGG GATGGTCTGCAGGAGGCGGGCATTGTGGATCAGTGCTGG GATGGTCTGCAGGAGGCGGGCATTGTGGATCAGTGCTGG GATGGTCTGCAGGAGGCGGGCATTGTGGATCAGTGCTGG GATGGTCTGCAGGAGGCGGGCATTGTGGATCAGTGCTGG GATGGTCTGCAGGAGGCGGGCATTGTGGATCAGTGCTGG GATGGTCTGCAGGAGGCGGGCATTGTGGATCAGTGCTGG GATGGTCTGCAGGAGGCGGGCATTGTGGATCAGTGCTGG GATGGTCTGCAGGAGGGCGGGCATTGTGGATCAGTGCTGG GATGGTCTGCAGGAGGCGGGCATTGTGGATCAGTGCTGG GATGGTCTGCAGGAGGGCGGGCATTGTGGATCAGTGCTGG GATGGTCTGCAGGAGGCGGGCATTGTGGATCAGTGCTGG GATGGTCTGCAGGAGGCGGGCATTGTGGATCAGTGCTGG GATGGTCTGCAGGAGGCGGGCATTGTGGATCAGGGCGGG GATGGTCTGCAGGAGGCGGGGCATTGGGACGGGGCGGAGGGGGGGG	Pick primers from a DNA sequence.	- numbers and blanks	ignored). FASTA format ok. Please i Pick right primer, or use right pr us 50 and 51. Or mark the <u>source seq</u> positions 27 and 28. Or mark the <u>sou</u> and T. arting at 401 and the 3 bases at 68. C	disclaimer         cautions         N-out undesirable sequence (vector, ALU         rimer below (5' to 3' on opposite strand)         uence with [ and ]: e.gATCT[CCCC]         arce sequence with -: e.gATCTAC-TC         yr mark the source sequence with < and 3	Js, LINEs, TCAT FTCAT ≈ c.g.

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 <u>start len te gcS any th 3' th hairpin ses</u> ULFT PRIMER 179 20 58.59 55.00 13.89 0.00 0.00 TGSCTTCTATAGACCCCACG RIGHT PRIMER 383 20 59.02 55.00 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 GCCAGCACCTGTGCGGCTCCAACCTAGTGGAGGGACTGTACATGACATGTGGACGGAGGG

241 TGGGTCTGGAGGCAGGCGGCCTGCAGCCTTCGGCCCTGGAGATGATTCTGCAGAAGCGCG

301 GCATTGTGGATCAGTGCTGTAATAACATTTGCACATTTAACCAGCTGCAGAACTACTGCA

421 CCCTTGAATGAG