

L'outil bio-informatique est devenu indispensable pour la recherche : pour la gestion des données, pour la production et la diffusion de connaissances et pour l'interprétation des données.

En vue d'une analyse de l'expression de gènes en qPCR, l'outil bio-informatique est nécessaire et indispensable. Le sujet portera donc sur la mise au point d'un couple d'amorces spécifiques d'une séquence d'intérêt ainsi que sur le traitement des résultats du séquençage.

On travaillera sur l'organisme *Triticum aestivum* qui est l'espèce de blé la plus cultivée dans le monde. Cette espèce est sujette à une maladie racinaire appelée piétin-échaudage causée par un champignon tellurique pathogène, *Gaeumannomyces graminis*, entraînant d'importantes pertes de rendement.

Le support pour l'étude de l'expression de gènes sera l'ARN messenger car les ARNm sont des molécules monocaténares qui assurent la communication de l'information génétique entre le noyau et le cytoplasme. Après maturation, ils sont polyadénylés à leur extrémité 3'.

Toutefois, on travaillera sur de l'ADNc (après rétrotranscription des ARNm) car les ADNc sont moins sujets à la dégradation et peuvent être amplifiés et mesurés par la suite en qPCR.

Dans un premier temps, diverses étapes de vérification de la spécificité du couple d'amorces utilisées seront réalisées.

Parallèlement, on effectuera une rétrotranscription des ARNm en ADNc, puis une PCR, un clonage et enfin un séquençage sur l'ADN plasmidique extrait, afin de vérifier que la séquence amplifiée est bien celle attendue.

Pour mener à bien la recherche et la mise au point du couple d'amorces spécifiques de la séquence cible, on utilisera la banque de données du « National Center for Biotechnology Information » (NCBI ; <http://www.ncbi.nlm.nih.gov>), le logiciel NetPrimer (<http://www.premierbiosoft.com/netprimer>) ainsi que l'outil « Basic Local Alignment Search Tool » (BLAST ; <http://blast.ncbi.nlm.nih.gov/Blast.cgi>).

On va ici s'intéresser au gène codant l'Adénine PhosphoribosylTransférase de forme 1 (APT1). Cette protéine serait impliquée dans une voie métabolique très active lors de la phase de réponse de la plante au pathogène.

On recherche donc sur la banque de données du NCBI la séquence d'ARNm correspondante au gène d'intérêt chez l'espèce *Triticum aestivum*.

The screenshot shows the NCBI Nucleotide search interface. The search term 'APT1 Triticum aestivum' is entered in the search box. Below the search box, there are links for 'Save search', 'Limits', and 'Advanced'. The search results are displayed in a table with a red border. The first result is 'Triticum aestivum adenine phosphoribosyltransferase form 1 (APT1) mRNA, complete cds'. It is 845 bp linear mRNA. The accession number is U22442.1 and the GI number is 726304. There are links for 'GenBank', 'FASTA', 'Graphics', and 'Related Sequences'.

NCBI Resources How To

Nucleotide Nucleotide APT1 Triticum aestivum

Save search Limits Advanced

Display Settings: Summary

[Triticum aestivum adenine phosphoribosyltransferase form 1 \(APT1\) mRNA, complete cds](#)

845 bp linear mRNA

Accession: U22442.1 GI: 726304

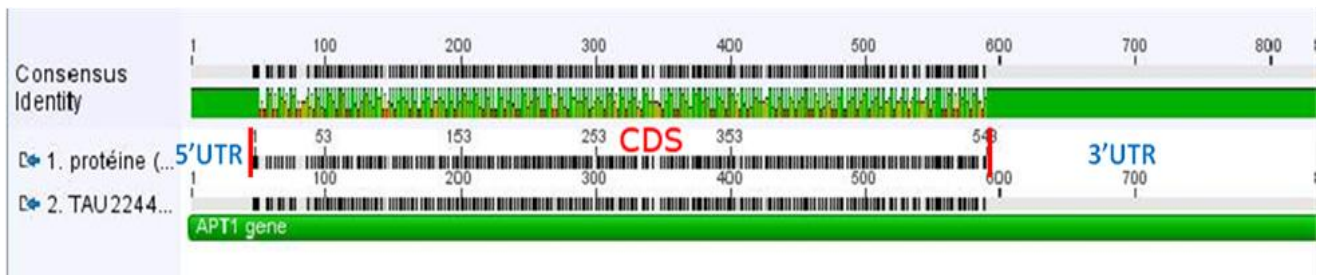
[GenBank](#) [FASTA](#) [Graphics](#) [Related Sequences](#)

Dans cette étude on part d'une matrice d'ARNm, or l'ARNm ne code pas entièrement pour la protéine. En effet, il existe des régions non codantes qui ne seront donc pas traduites et une région codante.

La séquence est totalement annotée (Coding DNA Sequence : CDS, 3' et 5' UnTranslated Region : 3'UTR et 5'UTR).

Lorsque la séquence n'est pas annotée, il faut alors dans ce cas trouver la séquence d'acides aminés de la protéine. Puis par le biais du code génétique, on passe d'une séquence d'acides aminés en une séquence d'acides nucléiques.

En alignant cette séquence obtenue avec celle de l'ARNm, on pourra observer la région CDS ainsi que les régions non codantes.



(Image obtenue à l'aide du logiciel Geneious™ lors de mon stage)

On remarque que la séquence nucléotidique de la protéine est similaire d'une région de l'ARNm qui correspond à la région CDS.

Les espaces en noir signifient que le nucléotide est identique pour les deux séquences.

Les espaces en blanc indiquent une différence entre les deux séquences mais cela est dû au code génétique qui est redondant. Il existe donc plusieurs codons pour un acide aminé ce qui explique ces différences.

Les amorces sont ensuite dessinées à l'aide du programme Primer-BLAST (<http://www.ncbi.nlm.nih.gov/tools/primer-blast>) qui utilise l'algorithme Primer3 et l'outil BLAST qui permet de comparer une séquence à une banque de données afin de contrôler que les séquences des amorces ne s'hybrident pas sur d'autres régions du génome.

Les amorces sont dessinées dans la région codante (CDS) de la séquence d'intérêt selon plusieurs critères permettant une meilleure spécificité des amorces :

- température de fusion (T<sub>m</sub>) : 55°C à 60°C
- taux en G et C : 40 à 70%
- longueur des amorces : 16 à 24 nucléotides

On veille également à ce que la taille des amplicons soit inférieure à 200 pb pour une efficacité optimale de l'amplification et supérieure à 80pb afin de pouvoir les distinguer des éventuels dimères d'amorces.

Primer Parameters

Use my own forward primer (5'→3' on plus strand)

Clear

Use my own reverse primer (5'→3' on minus strand)

Clear

PCR product size

Min

80

Max

200

# of primers to return

5

Primer melting temperatures (T<sub>m</sub>)

Min

55.0

Opt

58.0

Max

60.0

Max T<sub>m</sub> difference

3

PCR Product T<sub>m</sub>

Min

Opt

Max

Primer Size

Min

16

Opt

20

Max

24

Primer GC content (%)

Min

40.0

Max

70.0

Primer pair 4

	Sequence (5'→3')	Template strand	Length	Start	Stop	T <sub>m</sub>	GC%
Forward primer	GCAGCCCAATGACCGAGTCC	Plus	20	392	411	57.69	65.00
Reverse primer	GGCCCTTCAGCTCTGGCAGT	Minus	20	534	515	58.69	65.00
Product length	143						

Products on potentially unintended templates

>U22442.1 Triticum aestivum adenine phosphoribosyltransferase form 1 (APT1) mRNA, complete cds

product length = 143

Forward primer 1 GCAGCCCAATGACCGAGTCC 20

Template 392 ..... 411

Reverse primer 1 GGCCCTTCAGCTCTGGCAGT 20

Template 534 ..... 515

Primer pair 5

	Sequence (5'→3')	Template strand	Length	Start	Stop	T <sub>m</sub>	GC%
Forward primer	GGGAACCCCTTGTGCCGCTG	Plus	20	440	459	58.45	65.00
Reverse primer	TTGTCCCGGCCCTTCAGCTC	Minus	20	541	522	58.70	65.00
Product length	102						

Products on potentially unintended templates

>U22442.1 Triticum aestivum adenine phosphoribosyltransferase form 1 (APT1) mRNA, complete cds

product length = 102

Forward primer 1 GGGAACCCCTTGTGCCGCTG 20

Template 440 ..... 459

Reverse primer 1 TTGTCCCGGCCCTTCAGCTC 20

Template 541 ..... 522

Les différents couples d'amorces sont ensuite testés avec le logiciel Net Primer (<http://premierbiosoft.com/netprimer>) afin de minimiser le risque de formation de dimères d'amorces et de structures en épingle à cheveux (« hairpin »).

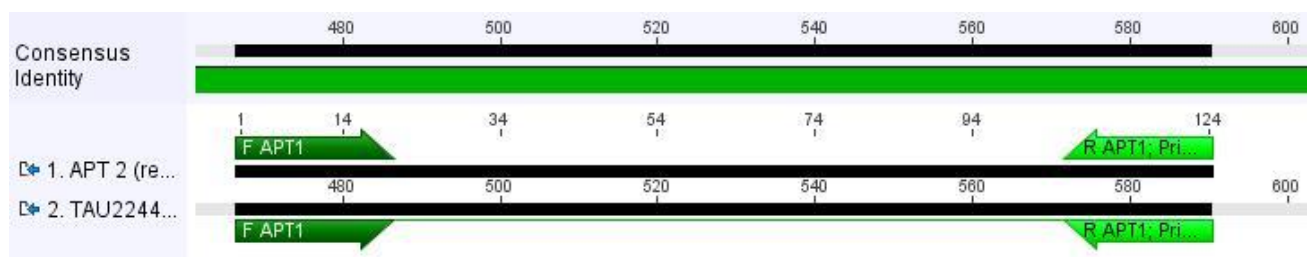
Le couple présentant le moins de ces structures est conservé.

Le couple d'amorces est ensuite commandé auprès de la société Eurofins MWG Operon (Ebersberg, Allemagne).

On réalise ensuite une PCR avec les amorces spécifiques sur les ADNc obtenus après rétrotranscription.

On clone les produits de PCR dans un plasmide (pGEM-T Easy, « TA Cloning ») qu'on insère ensuite dans des bactéries rendues compétentes au  $\text{CaCl}_2$ . Après culture, on sélectionne les clones recombinés (système blanc/bleu) et on effectue une PCR avec les amorces SP6/T7 présentes sur le plasmide. On extrait l'ADN plasmidique des produits de PCR qu'on envoie ensuite à la société GenoScreen de Lille pour être séquencé à partir des amorces universelles SP6 et T7 contenues dans le plasmide pGEM®-T Easy.

Les fragments séquencés sont ensuite alignés avec la séquence ayant servi à faire le dessin d'amorces. On vérifie ainsi que la séquence amplifiée correspond à la séquence cible.



(Image obtenue à l'aide du logiciel Geneious™ lors de mon stage)

On observe que la séquence du fragment séquencé (APT 2) correspond exactement à la séquence d'ARNm qui a servi de modèle (TAU22442).

À partir de là, on peut dire que les amorces sont spécifiques de notre séquence d'intérêt. On peut donc procéder à la suite des manipulations et à l'étude de l'expression de gènes en qPCR.