

Caractérisation du transcriptome des transporteurs de nutriments azotés de communautés phytoplanctoniques

Mise au point de Protocoles de Biologie Moléculaire



Lagunes
Palavasiennes
(34)



Liza DADU - Rapport de Stage 2013-2014

DUT Génie Biologique
Option Analyse Biologiques et Biochimiques
IUT de Montpellier



Laboratoire Ecosym – Montpellier
Tuteur : M. Daniel Grzebyk



REMERCIEMENTS

Je tiens tout d'abord à remercier chaleureusement **M. Daniel GRZEBYK et toute l'équipe du laboratoire Ecosym** pour leur accueil et les connaissances qu'ils m'ont apporté.

Je souhaite également remercier **M. Christophe BLANC et M. Didier CABRILLAC** pour m'avoir donné l'opportunité de faire une deuxième année de DUT Génie Biologique et **toute l'équipe pédagogique de la formation** pour la richesse de leurs enseignements, grâce à qui j'ai pu parfaitement appréhender mon sujet de stage et me familiariser avec les techniques utilisées.

Merci à tous mes proches pour leur aide et leur soutien de tous les jours.

SOMMAIRE

Introduction.....	6
I. Présentation De La Structure.....	7
II. Présentation Du Sujet.....	7
A. La biologie moléculaire : une discipline très prometteuse.....	7
B. Le phytoplancton : un groupe taxonomique très diversifié.....	9
C. Les lagunes Palavasiennes : un modèle écologique.....	11
D. Les transporteurs de nutriments azotés : des bioindicateurs ?.....	12
III. Matériel et méthodes.....	13
A. Prélèvements et matériel génétique utilisé.....	14
B. La Reverse transcription: Transformation du transcriptome en ADNc.....	15
C. Création d'une librairie de séquences et Blasts.....	15
1. Sélection des séquences de départ.....	15
1. Création de la librairie par BLASTs.....	16
D. Création et test d'amorces.....	17
1. Alignement des séquences de notre librairie et conception des amorces.....	17
2. Ajout de <i>Tails</i> pour les amorces AMT (TAMT1F1et R1).....	19
3. Vérifications sur les amorces.....	19
E. Les PCRs.....	20
1. Les Différentes approches successives de PCR (essais et mise au point) :.....	20
2. Purification des produits de PCR / des bandes du gel.....	22
F. Le Clonage : La séparation des séquences.....	22
1. Insertion des amplicons dans les plasmides et mise en culture sur boîte.....	23
2. Mise en culture liquide.....	23
3. Isolement de l'ADN et préparation pour le séquençage (quantification).....	24
G. Le séquençage.....	24
1. La S-PCR.....	24
2. Purification des produits de la PCR par précipitation à l'éthanol.....	25
3. Le Séquenceur à électrophorèse capillaire.....	25
IV. Résultats et Discussion.....	26
Conclusion.....	26
Références Bibliographiques.....	27

INDEX DES ILLUSTRATIONS

Figure 1 : Cycle biogéochimique du carbone dans les milieux marins (source : www.phytoplankton.net)	6
Figure 2 : Etang de Thau (photo Philippe Marsault)	6
Figure 3 : Bâtiment 24 de l'UM2 (photo L.D.)	7
Figure 4 : principe de la PCR	8
Figure 5 :	9
Colonie de Volvox sp.	9
Source : www.the-scientist.com	9
Figure 6: Longueurs d'ondes et pigments	9
Source : eric.lacouture.free.fr	9
Figures 7 : Illustration de la	9
Diversité phytoplanktonique	9
Figure 6 : (www.phytoplankton.net)	10
Figure 8 : Endosymbioses phytoplanktoniques	10
Figure 9 : Localisation de la zone d'étude et des stations	11
Figure 10 : Richesse spécifique des différentes stations (MOTUs = Molecular Operational Taxonomy Unit)	11
Figure 11 : Exemple de transporteur membranaire, ici l'ion Nitrate	12
Tableau 1 : Chronologie et Principe général de toutes les manipulations effectuées	13
Figure 12 : Protocole de filtration	14
Figure 13 : Extraction à l'azote liquide	14
Figure 14 : Protocole de purification : PureLink RNA mini kit (Life Technologies®)	14
Tableau 2 : Mélange des échantillons d'ARN	15
Figure 15 : Prime Script II 1st strand cDNA synthesis kit (takara)	15
Tableau 3 : Liste des séquences à la base de notre librairie	16
Tableau 4 : Résumé de la librairie pour AMT1	16
Tableau 5 : Résumé de la librairie pour NRT2	17
Figure 16 : Code des bases dégénérées	17
Figure 17 : alignement pour amt1 avec clustalX	18
Tableau 6: Conception d'Amorces NRT2 (astérisque = amorces commandées)	19
Figure 18 : principe de la "Tailed PCR"	19
Figure 19 : Thermocycleur pour les PCR	21
Figure 20 : Bac d'électrophorèse	21
Figure 21 : Photo d'un gel positif (sous UV)	21
Figure 23 : protocole du Kit TOPO TA	23
Figure 24 : protocole de repiquage et cultures liquides	23
Figure 22 : carte simplifiée du vecteur utilisé	23
Figure 26 : exemple de mesure du Nanodrop	24
Figure 25 : Principe du Nanodrop	24
Figure 28 : Précipitation à l'éthanol des Produits de S-PCR	25
Figure 27 : plaque à séquencer	25
Figure 29 : intérieur du séquenceur	25
Figure 30 : Exemple de Chromatogramme (N8)	25
Figure 31 : Taille des inserts	26

GLOSSAIRE DES ABRÉVIATIONS

A : Adénine	LB : milieu de culture bouillon lysogène
ADN : Acide Désoxyribonucléique	MA : milliards d'années
ADNc : ADN complémentaire	MEW : station de Méjean, zone ouest
AMT : gène codant le transporteur d'ammonium	MOTUs : Molecular Operational Taxonomy Unit
ARN : Acide RiboNucléique	NH₄⁺ : ion ammonium
ARNm : ARN messenger	NO₃⁻ : ion nitrate
ARNr : ARN ribosomique	NRT : gène codant le transporteur de nitrate
ARNt : ARN de transfert	Pb : paire de bases (unité de longueur des fragments)
AYR : station d'Ayrolle	PCR : Polymérase Chain Reaction
BLAST : Basic Local Alignment Search Tool	PRE : station Prevost
BLASTn : Blast de séquences nucléotidiques	PSM : Poste de Sécurité Microbiologique
BLASTp : Blast de séquences d'acides aminés	qPCR : PCR quantitative
C : Cytosine	RNases : ribonucléase
CO₂ : Dioxyde de carbone	RSL : Réseau de suivi Lagunaire
ddXTP : didésoxyribonucléotides modifiés (x = fluos)	Rtase : reverse transcriptase (polymérase)
dNTP : désoxyNucléotides TriPhosphate	Rt-PCR : PCR reverse transcriptase
Ecosym : Ecologie des systèmes marins côtiers	Sp : espèces
ETDA : Efflorescences toxiques et diversité algale	S-PCR : PCR de séquençage
EthOH : Ethanol	T : Thymine
G : Guanine	Tm : température d'hybridation
HPLC : High-performance liquid chromatography	U : Uracile
I : Inosine	UM2 : Université Montpellier II
IDT : Integrated DNA Technologies®	UMR : Unité Mixte de Recherche
INN : station d'Ingril, zone Nord	VIC : station de Vic-la-Guardiole
INS : station d'Ingril, zone Sud	

Abréviation des noms d'espèces :

Py	<i>Pyropia yezoensis</i>	Tp	<i>Thalassiosira pseudonana</i>	Mp	<i>Micromonas pusilla</i>
Cc	<i>Chondrus crispus</i>	Sc	<i>Skeletonema costatum</i>	Ot	<i>Ostreococcus tauri</i>
Pi	<i>Phytophthora infestans</i>	Db	<i>Ditylum brightwellii</i>	Ol	<i>Ostreococcus lucimarinus</i>
Cm	<i>Chaetoceros muellerii</i>	Aa	<i>Aureococcus anophagefferens</i>	Dv	<i>Dunaliella viridis</i>
Cf	<i>Cylindrotheca fusiformis</i>	Eh	<i>Emiliana huxleyi</i>	Cr	<i>Chlamydomonas reinhardtii</i>
Pt	<i>Phaeodactylum tricornutum</i>	Bp	<i>Bathycoccus prasinus</i>	Vc	<i>Volvox carteri</i>
Tw	<i>Thalassiosira weissflogii</i>	Msp	<i>Micromonas sp.</i>	Gt	<i>Guillardia theta</i>
Ht	<i>Heterocapsa triquetra</i>	Ao	<i>Alexandrium ostenfeldii</i>	Om	<i>Oxyrrhis marina</i>
Ss	<i>Symbiodinium sp</i>	Kb	<i>Karenia brevis</i>		
Ac	<i>Alexandrium catenella</i>	Pmar	<i>Perkinsus marinus</i>		
Am	<i>Alexandrium minutum</i>	Km	<i>Karlodinium micrum</i>		

INTRODUCTION

Les problématiques concernant l'écologie des écosystèmes marins sont des plus importantes à l'heure actuelle. En effet, l'eau représente 72% de la surface de la planète avec 93% d'eau salée et 3% d'eau douce. Les écosystèmes marins ont ainsi une très grande influence sur le fonctionnement de la biosphère et notamment sur la régulation du climat global.

Le phytoplancton marin, bien qu'il ne représente que 1% de la biomasse photosynthétique, est responsable de 45% de la production primaire planétaire (Field et al, 1998). Il produit ainsi la moitié du dioxygène mondial consommé par les organismes hétérotrophes. Il est également à la base des réseaux trophiques océaniques et de l'alimentation des oiseaux marins. Enfin, il joue un rôle climatique essentiel en pompant le CO₂ (gaz à effet de serre) de l'air (image ci-contre : le cycle du carbone). Il est donc important de mieux connaître la biologie de ces organismes pour mieux gérer et préserver les milieux.

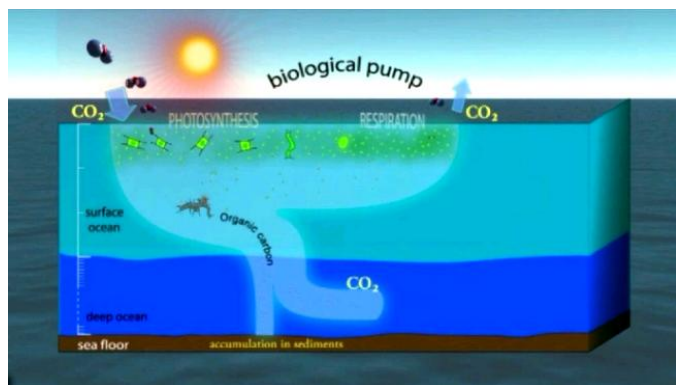


Figure 1 : : Cycle biogéochimique du carbone dans les milieux marins (source : www.phytoplankton.net)

Mon stage, effectué au sein du Laboratoire Ecosym de Montpellier, a visé à élaborer les protocoles de biologie moléculaire permettant d'étudier l'écophysiologie de la nutrition du phytoplancton marin. Ces études ont été effectuées sur des échantillons d'eau prélevés dans le complexe des lagunes palavasiennes, lesquelles constituent un véritable modèle spatio-temporel de restauration écologique. Ce rapport expose le travail que nous avons effectué avec mon tuteur de stage, M. Daniel Grzebyk.

Dans une première partie, je présenterai le contexte du stage à savoir la structure qui m'a accueillie, ses axes de recherches et le matériel dont elle dispose. Je développerai également le sujet de mon stage en décrivant brièvement le phytoplancton, ensuite en présentant la zone d'étude et son histoire ainsi que les gènes qui nous ont intéressés. Dans une troisième partie, je détaillerai les techniques de laboratoire que nous avons utilisé ainsi que les différents protocoles que nous avons testés. Enfin, je terminerai par les résultats obtenus avec les différents protocoles et je conclurai sur la méthode retenue.



Figure 2 : Etang de Thau (photo Philippe Marsault)

I. PRÉSENTATION DE LA STRUCTURE



Figure 3 : Bâtiment 24 de l'UM2 (photo L.D.)

Le laboratoire **Ecosym** (Ecologie des systèmes marins côtiers) – UMR 5119 de l'Université Montpellier II, est celui au sein duquel j'ai effectué mon stage.

Il a pour objectif d'étudier les « *effets des changements locaux et globaux liés à l'anthropisation sur les écosystèmes marins côtiers, les communautés, les populations et les organismes qui les composent* ».

Ce laboratoire comporte 8 équipes, dont l'équipe ETDA (Efflorescences toxiques et diversité algale) qui est celle de mon tuteur **M. Daniel Grzebyk**.

Cette équipe s'intéresse à l'écologie des communautés phytoplanctoniques dont elle étudie la diversité et les conditions de développement. L'objectif est de définir des bio-indicateurs des milieux côtiers. Les principales techniques qu'ils utilisent sont la cytométrie en flux, l'HPLC, la microscopie photonique et électronique, la spectrométrie de masse pour l'analyse élémentaire, l'élevage de macro-organismes (huîtres, poissons...) la culture cellulaire et microbiologique ainsi que la biologie moléculaire.

M. Grzebyk s'intéresse à l'écologie du phytoplancton en utilisant des approches moléculaires, en particulier pour aborder des questions liées à la diversité génétique des communautés. Mon stage s'est inscrit dans le cadre de ces recherches.

II. PRÉSENTATION DU SUJET

Cette partie, précisant le contexte de mon sujet d'étude, comporte quatre sections : la première est une brève introduction à la biologie moléculaire, la seconde une présentation du phytoplancton, la troisième section décrit la zone d'étude : les étangs palavasiens et, enfin, la dernière section aborde plus particulièrement les transporteurs de nutriments azotés, ainsi que les gènes les codants. Ces gènes seront le cœur de mon sujet de stage.

A. La biologie moléculaire : une discipline très prometteuse

La Biologie moléculaire consiste à étudier et à utiliser les mécanismes moléculaires impliquant les gènes et les protéines, notamment entre eux. Cette discipline se différencie de la génétique qui étudie la fonction des gènes, par ex les mutations. C'est une science qui est encore jeune puisqu'elle commence avec la découverte la structure et la fonction de **l'ADN** dans les années 50 par Watson et Crick. Dans les années 60, **l'ARN** est envisagé comme étant l'intermédiaire entre l'ADN (transcription) et les protéines (traduction). En réalité, il existe plusieurs sortes

d'ARN dont : les ARNm, Messenger qui sont les modèles pour les protéines, les ARNr, ribosomiques, constituant la machinerie de traduction (les ribosomes) et les ARN de transfert qui apportent les acides aminés lors de la traduction. Les ARNr ont des séquences très bien conservées dans le monde vivant. Les ribosomes sont constitués de 2 sous unités : une petite (ARN16s(bactéries)/18S(eucaryotes)) qui vérifie la correspondance codon ARNm/codonARNt et une Grande sous-unité (ARN23s (bactérie)/28s(eucaryotes)) qui attire les ARNt et transfère les Acides Aminés vers la chaîne peptidique (protéine).

Une technique, en particulier est très importante en biologie moléculaire : **la PCR**. Celle-ci est à la base de nombreuses technologies dont des amplifications en temps réel (qPCR), la reverse transcription (RT) et le séquençage.

Le principe de la PCR consiste à amplifier une séquence choisie et encadrée par deux amorces, qui seront les points de départ de l'élongation obtenue grâce à la Polymérase (ajout de bases à la suite les unes des autres). Trois cycles sont répétés (dénaturation, hybridation, élongation) afin d'obtenir de nombreux amplicons qui seront visibles sur gel. La figure ci-contre montre un cycle de PCR à répéter n fois (de 25 à 40 fois) dans une machine appelée thermocycleur.

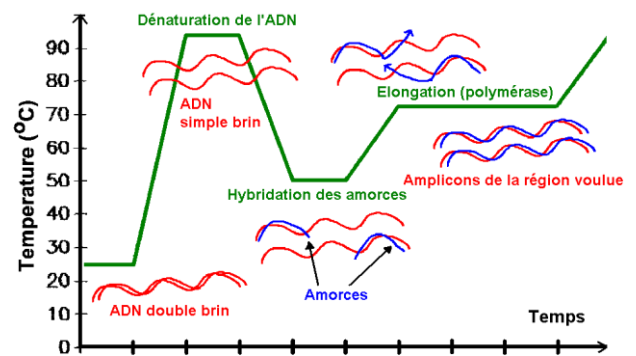


Figure 4 : principe de la PCR

Les amplicons peuvent être mis à migrer sur un **gel d'électrophorèse**. Ils se sépareront en fonction de leur taille, attirés par un champ électrique, transmis dans un tampon. La visualisation est rendue possible grâce à des intercalants de l'ADN, fluorescents aux UV (dans nos manipulations il s'agissait de SybrSafe).

Le séquençage selon la méthode de Sanger (années 70) consiste à effectuer une PCR (S-PCR) avec des bases modifiées pour arrêter l'élongation et pour être fluorescentes (une couleur par base : ddXTP). Dans les produits de S-PCR se trouve alors plusieurs amplicons de tailles différentes dont l'extrémité est terminée par une base fluorescente de couleur. La lecture se fait par électrophorèse : soit sur gel (méthode ancienne) soit par Electrophorèse capillaire qui revient au même sauf que le gel est dans un capillaire. Les amplicons se distribuent par taille et se rangent donc dans l'ordre de la séquence. La détection se fait grâce à un rayon laser et un capteur qui enregistre la longueur d'onde du ddXTP. Le document en sortie de l'appareil est appelé un chromatogramme.

Enfin, dans ce rapport, la notion de **métagénomique** sera souvent employée. Cette notion fait appel à un procédé méthodologique qui vise à étudier le contenu génétique mélangé d'un échantillon issu d'un environnement naturel complexe (par opposition à des souches pures cultivées en laboratoire). Le but de cette approche, via le séquençage direct de l'ADN présent dans l'échantillon, est d'avoir non seulement une description génomique du contenu de l'échantillon mais aussi un aperçu du potentiel fonctionnel d'un environnement.

Les prélèvements de phytoplancton que nous allons utiliser sont pour la plupart des échantillons naturels, donc des mélanges d'espèces différentes.

B. Le phytoplancton : un groupe taxonomique très diversifié

Le **Phytoplancton** est défini par trois critères : ce sont des organismes photosynthétiques, comme les algues (*phytos* : « plante » en grec) qui dérivent (Falkowski et Raven, 1997) dans la colonne d'eau (*plancton* : « qui flotte ») unicellulaires, solitaires ou coloniaux (par ex. *Volvox sp.*, voir photo ci-contre). C'est un groupe polyphylétique et il comprend aussi bien des organismes eucaryotes que procaryotes. Il existe donc une très grande diversité de taille (de 0,2µm à 200µm), de couleurs et de formes (plus de 25 000 formes dans 8 phyla).

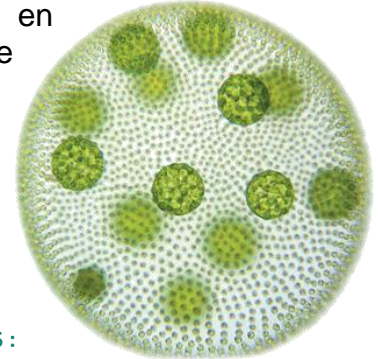


Figure 5 :
Colonie de *Volvox sp.*
Source : www.the-scientist.com

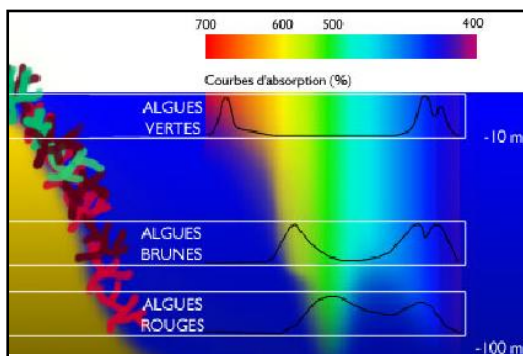


Figure 6: Longueurs d'ondes et pigments
Source : eric.lacouture.free.fr

La diversification des pigments photosynthétiques ont permis à ces organismes de coloniser toutes les niches écologiques de la colonne d'eau photique, principalement conditionnées par les longueurs d'ondes disponibles en fonction de la profondeur. (cf. schéma ci-contre).

La liste ci-dessous représente les différentes lignées du phytoplancton :

TAXONOMIE du Phytoplancton :

- **Empire Procaryotes :**
 - Subdivision: Cyanobactéries ou cyanophytes (algues bleues-vert)
 - Subdivision : Chloroxybactéries
- **Empire Eucaryotes :**
 - Règne : Protozoaires
 - Division : Euglènes (vertes)
 - Division : Dinophycées (rouges)
 - Règne : végétal (plantae)
 - Sous-règne : Biliphytae
 - Division : Glaucocystophytes (gris bleues)
 - Division : rhodophytes (rouges)
 - Sous-règne : viriplantidae (vertes)
 - Division : Chlorophytes
 - Division : Prasinophytes
 - Règne : Chromistes



Figures 7 : Illustration de la Diversité phytoplanctonique

- Sous-règne : Chlorenchia
 - Division : Chlorarachniophytes (vertes)
- Sous-règne : Euchromistes (rouges)
 - Division : Cryptophytes
 - Division : Haptophytes
 - classe : Prymnésiophycées
 - Division : Hétérokotes
 - Classe : Diatomées
 - Classe : Chrysophycées
 - Classe : Eustigmatophycées
 - Classe : Raphidophycées
 - Classe : Tribophycées

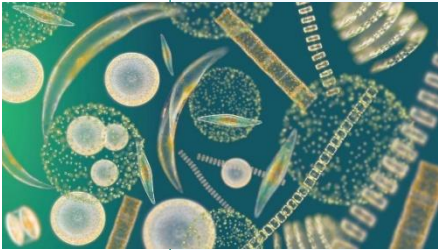


Figure 6 : (www.phytoplankton.net)

L'histoire évolutive du phytoplancton débute il y a 3.8 MA (précambrien) avec les premières cyanobactéries.

L'hypothèse endosymbiotique (Yoon et al 2004) émet l'idée que les chloroplastes des algues eucaryotes proviennent de l'internalisation de cyanobactéries.

Ceci a ensuite été confirmé par des analyses moléculaires (les chloroplastes ayant leur propre génome, comme les mitochondries). Il est largement admis que plusieurs endosymbioses (jusqu'à 3) ont permis la création de 5 des 8 clades eucaryotes actuels. Des mutations ont ensuite continué cette diversification

La figure ci-contre, tirée de l'article de Falkowski et al (2004) illustre ceci..

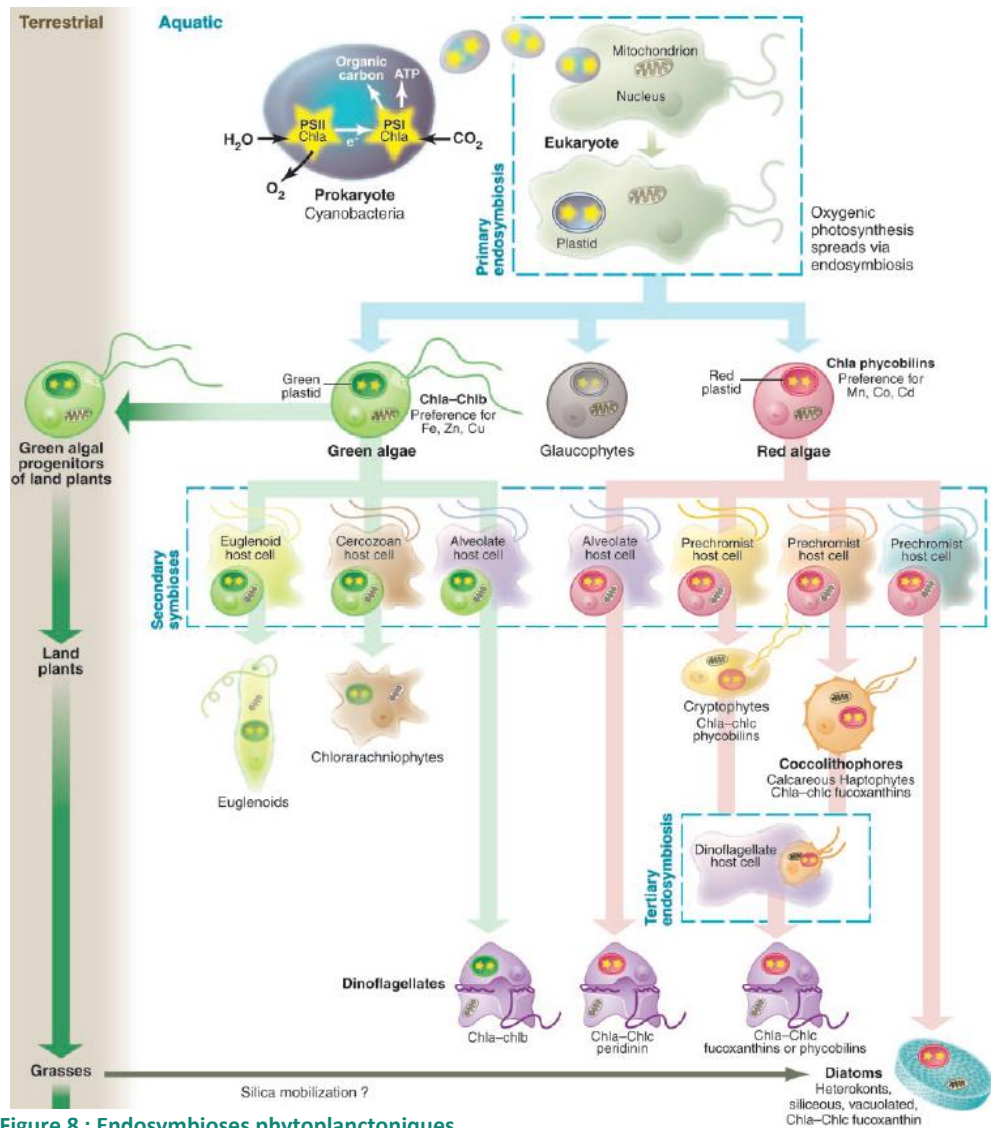


Figure 8 : Endosymbioses phytoplanctoniques

C. Les lagunes Palavasiennes : un modèle écologique

Le complexe des **lagunes palavasiennes** est constitué de 8 plans d'eau s'étalant du Canal du Rhône jusqu'à Sète. Le canal apporte les effluents du Rhône ainsi que celles du Lez par l'extrémité Est du complexe. Or, jusqu'en 2005, la station d'épuration de Montpellier rejetait les eaux traitées dans le Lez, ce qui entraînait un fort enrichissement en azote et en phosphore des lagunes (Dupré, 2002). Ainsi s'est installé un gradient d'eutrophisation décroissant d'Est en Ouest selon le RSL (Réseau de Suivi Lagunaire évaluant la qualité des milieux depuis 2000).

En 2005 a été installé un émissaire permettant de déverser les rejets plus loin, en mer. Le suivi RSL (Réseau de suivi Lagunaire) a observé une baisse de l'eutrophisation des lagunes. A l'aide de codes-barres d'ADNr 28S, l'équipe de M. Grzebyk et ses collaborateurs (Grzebyk et al, 2013) ont observé des différences dans la diversité génétique des communautés phytoplanctoniques entre 5 lagunes (cf carte). La lagune d'Ayrolle ne faisant pas partie du complexe palavasien est ici utilisée comme station de référence.

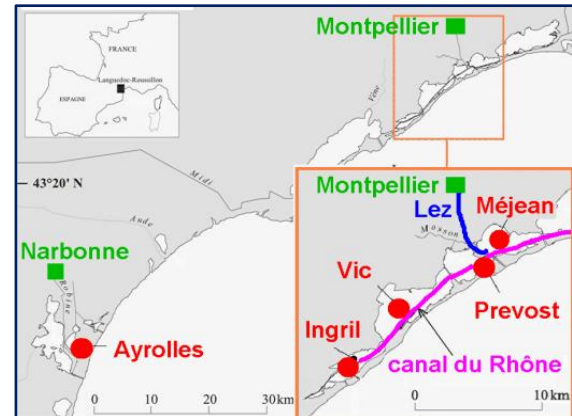


Figure 9 : Localisation de la zone d'étude et des stations

Le graphique ci-dessous est un résumé des résultats de cette étude : (INN et INS = Ingril Nord et Sud)

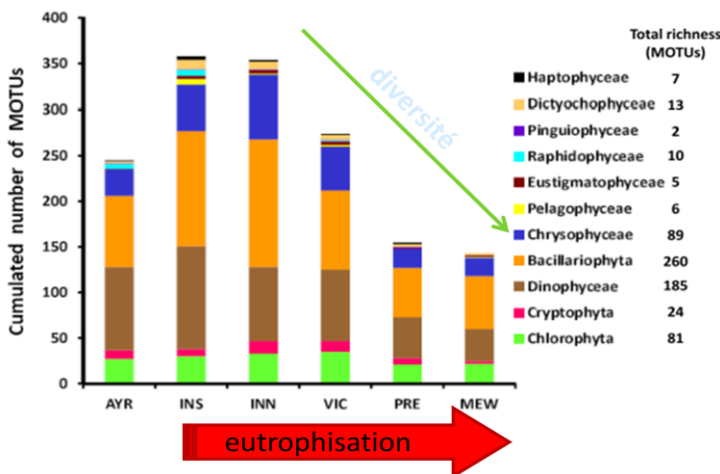


Figure 10 : Richesse spécifique des différentes stations (MOTUs = Molecular Operational Taxonomy Unit)

Les lagunes les moins eutrophisées seraient les plus diversifiées. La question qui se pose est de savoir si la diminution des apports nutritionnels influence la diversité taxonomique (plus d'espèces) mais également fonctionnelle, comme la conséquence d'une compétition accrue pouvant conduire à une réponse adaptative différente selon les organismes, afin d'exploiter au mieux tous les nutriments disponibles dans le milieu (de sorte que des organismes occupent des niches écologiques différentes).

L'étude de la transcription des gènes de transporteurs de nutriments azotés, qui est le sujet de mon stage, fait suite à cette étude de l'ARNr 28S. En effet, il est exploré la possibilité de faire de ce transcriptome des bioindicateurs de la réaction des organismes phytoplanctoniques aux changements environnementaux, notamment en ce qui concerne les apports nutritionnels.

D. Les transporteurs de nutriments azotés : des bioindicateurs ?

De nombreux travaux s'intéressent à l'utilisation des nutriments par les plantes (VonWittenstein et al 2014) et par le phytoplancton (Chung et al. 2003, Hung et al. 2013, Monier et al. 2012...), notamment de l'azote (N) et du phosphore (P) qui sont deux grands facteurs limitants de la croissance des organismes autotrophes.

Le nitrate (NO_3^-) est souvent la forme la plus abondante dans l'eau mais l'ammonium (NH_4^+) serait mieux utilisé par le phytoplancton. De par leur diversité, les organismes du phytoplancton ont des capacités différentes d'acquisition de ces nutriments. Il existe en particulier des travaux sur les gènes des transporteurs membranaires responsables de l'entrée de ces nutriments dans les cellules. Leur expression, étudiée dans plusieurs espèces, serait fonction des quantités de nutriments disponibles dans le milieu, ce qui en ferait de bons marqueurs de leur utilisation. Il s'agit du gène *nrt2* lié au transport du nitrate (Hildebrand & Dahlin 2000, Song & Ward 2007, Kang et al. 2011), d'un gène de transporteurs d'ammonium : gène *amt* (Hildebrand 2005). L'article de Kang et al (2011) fournissant des séquences d'amorces pour le gène NRT2, nous a inspiré pour la conception de nos propres amorces (cf. Matériel et méthodes partie D-b-1 p.19).

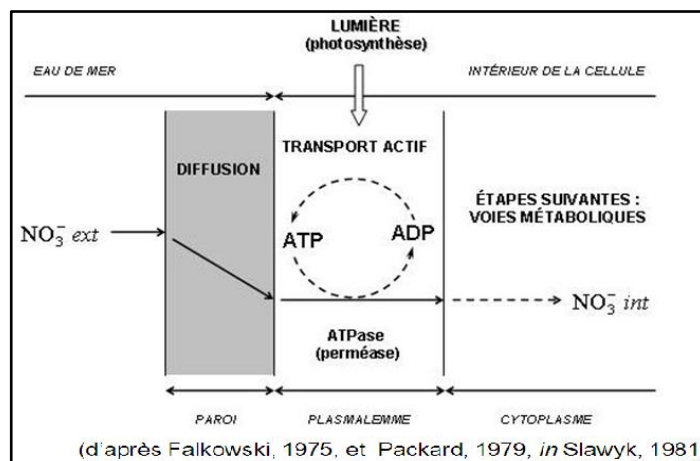


Figure 11 : Exemple de transporteur membranaire, ici l'ion Nitrate

Récemment, l'équipe de Von Wittgenstein (VonWittenstein et al 2014) a entrepris de construire des arbres phylogénétiques à partir de séquences de gènes de transporteurs d'ammonium (*amt1* ou *2*), de nitrate (*NRT1* ou *2*) et de peptides (*PTR*) et ce sur un panel exhaustif de végétaux, terrestres et aquatiques (dont certains appartenant au phytoplancton). Il ont ainsi constitué une librairie de séquences nucléotidiques à partir de données provenant de la littérature et de bases de données en ligne. Les séquences des organismes phytoplanctoniques de cette étude ont été à la base de nos travaux.

Le projet de M. Grzebyk est de pouvoir à terme, quantifier le niveau d'expression de ces gènes (grâce à la qPCR) afin d'en faire des bioindicateurs, caractéristiques de la réaction des organismes face aux changements de leur environnement.

III. MATÉRIEL ET MÉTHODES

Dans cette partie sont exposées en détail les manipulations qui ont été nécessaires dans mon travail. Les prélèvements d'eau sont tout d'abord purifiés, afin d'en extraire les ARN totaux. L'ARN est ensuite reverse-transcrit, afin de travailler avec de l'ADN (il est difficile de travailler sur l'ARN directement). Une fois les couples d'amorce conçues, différents protocoles de PCR sont testés, et, lorsque le résultat est satisfaisant (présence de bandes sur les gels), les produits de PCR sont séparés par clonage (transformation de plasmides). Enfin, ces plasmides sont extraits puis purifiés pour être séquencés.

Le tableau ci-dessous retrace le déroulement chronologique et le principe général des manipulations effectuées, de l'échantillon brut au résultat de séquençage. Les mots soulignés mettent en évidence les manipulations que je n'ai pas effectuées.

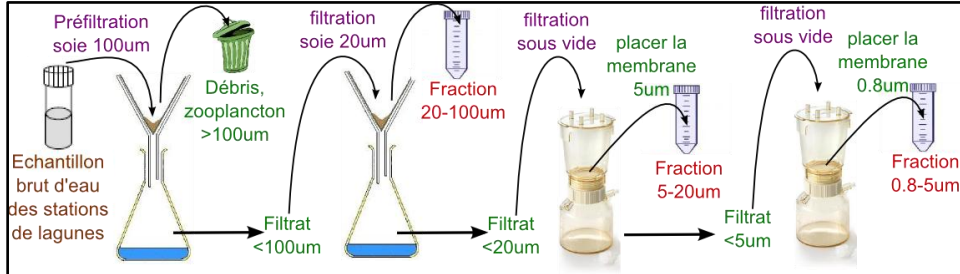
Tableau 1 : Chronologie et Principe général de toutes les manipulations effectuées

ECHANTILLONNAGE ET TRAITEMENT	
Manipulations	Objectifs/Principe
<ul style="list-style-type: none"> • <u>Prélèvement</u> • <u>Filtration sur soie et sous vide</u> 	<ul style="list-style-type: none"> • Prélèvements d'eau par station à différentes saisons (bateau) • Fractionnement correspondant au phytoplancton (0,2um-2um)
↓	
EXTRACTION DU TRANSCRIPTOME	
Manipulations	Objectifs
<ul style="list-style-type: none"> • <u>Isolement des ARNs totaux</u> • Reverse Transcription 	<ul style="list-style-type: none"> • Ne pas amplifier de séquences chromosomiques par la suite • Travailler avec de l'ADN (ici : ADN complémentaire)
↓	
RECHERCHE DU COUPLE D'AMORCES	
Manipulations	Objectifs
<ul style="list-style-type: none"> • Bibliographie / GenBank • Blasts et alignements de séquences 	<ul style="list-style-type: none"> • Recherche de séquences du gène d'intérêt • Positionnement des amorces sur les régions conservées
↓	
AMPLIFICATION DU GÈNE D'INTÉRÊT	
Manipulations	Objectifs
<ul style="list-style-type: none"> • PCR (+Visualisation sur gel) • Purification des produits de PCR 	<ul style="list-style-type: none"> • Observer la présence ou non du gène-cible • Amplifier le matériel pour les manipulations suivantes
↓	
CLONAGE	
Manipulations	Objectifs
<ul style="list-style-type: none"> • Transfection et culture sur boîte • Isolement des colonies et cultures liquides • Purification des plasmides 	<ul style="list-style-type: none"> • Obtenir une seule séquence par culture liquide pour le séquençage (car mélange métagénomique) • Amplification du matériel
↓	
SÉQUENÇAGE	
Manipulations	Objectif
<ul style="list-style-type: none"> • PCR de séquençage (Sanger) • Purification et <u>Séquenceur</u> 	Lire les séquences obtenues par le couple d'amorces choisi et évaluer la méthode (taux de réussite)
↓	
RÉSULTATS : BLASTS ET PHYLOGÉNIE	
Manipulations	Objectif
<ul style="list-style-type: none"> • Blasts sur les séquences obtenues 	<ul style="list-style-type: none"> • Vérifier que les séquences obtenues sont bien celles voulues (gène et espèces)

A. Prélèvements et matériel génétique utilisé

Le matériel génétique à la base des manipulations est de l'ARN total qui provient de filtrations et de purification de prélèvements en milieu naturel (et exceptionnellement de cultures). Voici les protocoles suivis :

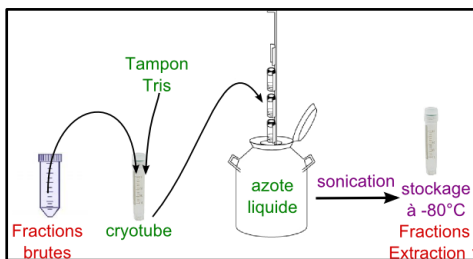
Figure 12 : Protocole de filtration



on obtient ici 3 fractions de cellules phytoplanctoniques :

**0,8-5µm ; 5-20µm
et 20-100µm**

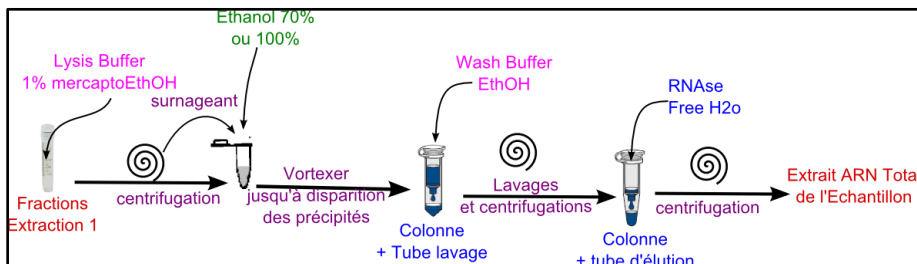
Figure 13 : Extraction à l'azote liquide



Le passage des échantillons à l'azote liquide permet de congeler ceux-ci en vue du stockage à -80°C mais également de lyser les cellules grâce aux cristaux de glace qui les font exploser. La sonication (lyse par ondes sonores). permet de terminer cette étape de lyse.

Le contenu des cellules (dont l'ARN) est alors en suspension dans le tube, prêt à être extrait.

Figure 14 : Protocole de purification : PureLink RNA mini kit (Life Technologies®)



Le tampon de lyse utilisé est de composition inconnue (kit). Il garde en suspension l'ARN lors de la centrifugation.

L'éthanol 70 ou 100% ensuite ajouté dans le surnageant précipite les brins d'ARN, qui seront retenus sur la membrane dans la colonne lors des lavages au tampon Wash Buffer. Enfin, l'ARN sera récolté dans un nouveau tube grâce à l'eau (exempte de RNase, pour ne pas abimer le matériel ainsi récolté) qui va décrocher les brins de la membrane pour l'éluier vers le tube. Les RNases étant présentes partout, les manipulations doivent être effectuées proprement.

Cet ARN total a ensuite été re-mélangé entre saisons et entre fractions de tailles sur les 3 stations du Méjean (MEW), d'Ingril Nord (INN) et, Ayrolles (AYR). Le but est d'augmenter la diversité spécifique et donc génétique dans les tubes, afin de renforcer la probabilité d'obtenir un résultat positif. Le tableau suivant rapporte les mélanges effectués :

Tableau 2 : Mélange des échantillons d'ARN

		mai			août			Mélange des 3 Tubes = tube « MIA »
tube M	MEW	20-100µm	5-20 µm	0,8-5 µm	20-100µm	5-20 µm	0,8-5 µm	
		mai			août			
tube I	INN	20-100µm	5-20 µm	0,8-5 µm	20-100µm	5-20 µm	0,8-5 µm	
		septembre	mai		août			
tube A	AYR	20-100µm	5-20 µm	0,8-5 µm	20-100µm	5-20 µm	0,8-5 µm	

B. La Reverse transcription: Transformation du transcriptome en ADNc

L'ARN total récolté ne peut être exploité directement pour la PCR. Celui-ci subit donc une reverse transcription (RT), utilisant une polymérase reverse Transcriptase permettant d'obtenir de l'ADN, complémentaire du brin d'ARN. Le protocole suivi est celui indiqué par le fabricant du kit utilisé dont le principe est résumé par la figure ci-dessous :

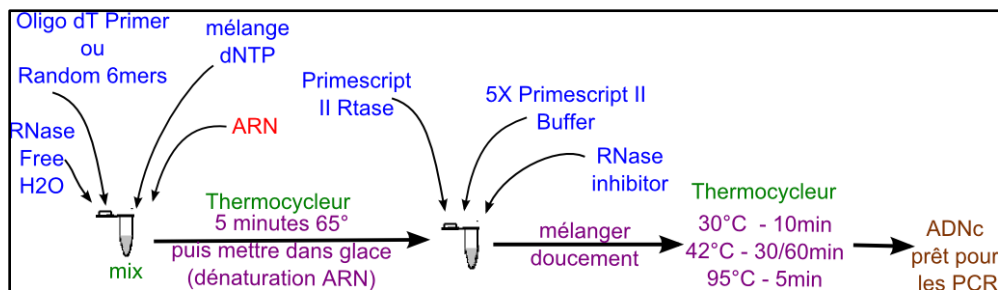


Figure 15 : Prime Script II 1st strand cDNA synthesis kit (takara)

Le principe du kit est de fournir à la polymérase « Primescript II Rtase » des amorces aléatoires (poly-Nx6 ou oligos dT) qui permettent l'élongation de la totalité du brin d'ARN. Des inhibiteurs de RNAses sont ici nécessaires.

L'ADN ainsi obtenu sera notre matrice pour les PCR. Cependant, pour effectuer l'amplification d'une région d'intérêt, il est tout d'abord nécessaire de localiser cette région pour ensuite concevoir les amorces encadrant celle-ci. L'étude portant sur un mélange de plusieurs organismes de différents groupes taxonomiques, il est indispensable de procéder à l'alignement des séquences afin de trouver les régions conservées correspondant au gène cible. Ceci est possible grâce à la création d'une librairie de séquences pour les gènes étudiés.

C. Création d'une librairie de séquences et Blasts

1. Sélection des séquences de départ

Les séquences protidiques de transporteurs d'ammonium (AMT1), de nitrate (NRT 2) récupérées dans le fichier Excel fourni par Wittgenstein et al (2014) ont été récoltées principalement par BLASTp (j'aborderai dans le chapitre suivant le principe des BLASTs) sur GenBank et Phytozome.net : 99 séquences pour AMT1 (33 espèces), 86 séquences pour AMT2 (30 espèces), 1097 séquences pour NRT 1/PTR (41 espèces) et 93 pour NRT2 (32 espèces).

Nous avons nous-mêmes utilisé les séquences sélectionnées (correspondant aux organismes phytoplanctoniques) comme base pour créer notre propre librairie, à partir de BLASTs sur GenBank. Le tableau ci-dessous liste les séquences que nous

avons utilisées pour notre travail. Nous avons sélectionné une séquence par groupe taxonomique avec un organisme susceptible d'être présent dans notre zone d'étude. L' «*accession number* » permet de les retrouver sur les bases de données en ligne (en noir : GenBank, en vert : Phytozome.net)

Tableau 3 : Liste des séquences à la base de notre librairie

Espèces	Abrév.	groupe taxonomique	gènes	accession number
<i>Chlamydomonas reinhardtii</i>	Cr	chlorophytes	amt1a	Cre03.g159250.t1.1
<i>Guillardia theta</i>	Gt	cryptophytes	amt1a	EKX32727
<i>Cylindrotheca fusiformis</i>	Cf	Hétérokontes diatomées	amt1a	AAK52491
<i>Ectocarpus siliculosus</i>	Es	Hétérokontes brunes	NRT2a	CBJ31727

1. Création de la librairie par BLASTs

Nous avons ensuite formaté les quatre séquences énoncées ci-dessus pour être reconnues par l'outil en ligne BLAST, en l'occurrence dans le format de fichier FASTA. Celui-ci consiste en un fichier .txt avec la forme suivante :

Titre de la séquence : > *Accession_Abréviation_espèce_gene_groupe* ←

Séquence :

```
MSDWKKYDITYDVPIDPKQQDKATTLKLYSFRRPHMRAFHFVWFGF
FMAFVSWFAFEIKADLGMTKNEVYNANISSVSSTILSRFVVGPLCDT
FGARITSSTLLLMGCIPTNSATDVAIIRFFIGVMHGAM
```

a) Le Principe du BLAST

Le BLAST (acronyme pour *Basic Local Alignment Search Tool*) est une méthode bio-informatique permettant de rechercher les régions similaires entre plusieurs séquences de nucléotides (BLASTn) ou d'acides aminés (BLASTp) d'aligner les régions conservées et de récupérer les séquences des régions alignées.

Ici, nous allons utiliser nos quatre séquences formatées FASTA en tant que requête afin d'élargir notre librairie avec davantage d'organismes correspondant à notre zone d'étude.

Ces 4 séquences, fournies par Von Wittgenstein et al (2014) seront à la base de toute la librairie. Celle-ci sera résumée sous forme de tableaux, où le nom des espèces est désigné par une abréviation. La correspondance des abréviations est donnée dans le glossaire en début de rapport (p.5).

b) Librairie de Séquences AMT 1

Le tableau ci-dessous rapporte le nombre de séquences sélectionnées pour être alignées, par espèce et par groupe taxonomique pour le gène AMT1.

Tableau 4 : Résumé de la librairie pour AMT1

Groupe	Diatomées	Pélagophycées	Haptophycées	Chlorophycées	Cryptophycées	Dinoflagellés
Espèces	Cf : 4 Pt : 10			Bp : 3 Msp : 3 Mp : 2 Ot : 1 Ol : 2 Dv : 1 Cr : 2 Vc : 3		Ht : 3 Ssp : 3 Ac : 9 Am : 2 Ao : 2 Kb : 1 Pmar : 1 Km : 1 Om : 7
+ Nombre de séquences	Tw : 1 Tp : 6	Aa : 4	Eh : 8		Gt : 6	
Total	21	4	8	17	6	29

c) Séquences NRT2

Ainsi que précédemment, voici le tableau résumant les séquences pour le gène NRT2 :

Tableau 5 : Résumé de la librairie pour NRT2

Groupe	Diatomées	Oomycètes	Rhodophycées	Chlorophycées	Dinoflagellés
Espèces + Nombre de séquences	Cf : 2 Pt : 4 Tw : 7 Tp : 4 Sc : 4 Db : 5 Cm : 4	Pi : 2	Py : 1 Cc : 1	Oi : 1	Am : 1
Total	30	2	2	1	1

D. Création et test d'amorces

Afin de créer des amorces, les séquences similaires de notre librairie précédemment créée (correspondant aux gènes d'intérêt) doivent être alignées pour être mises en évidence et ce grâce à deux logiciels : Dialign et ClustalX.

1. Alignement des séquences de notre librairie et conception des amorces

Le logiciel Dialign permet d'obtenir un fichier sous format FASTA avec l'alignement des régions similaires. Les « Gaps » ou régions vides dans les séquences peuvent provenir de mutations dans le génome ou de différents stades d'épissage de l'ARN séquencé.

Le logiciel ClustalX permet de visualiser le fichier d'alignement FASTA et de calculer le degré de similarité entre toutes les régions alignées (fort degré représenté par une astérisque). Ces régions conservées sont d'excellentes candidates pour recevoir les amorces. Cependant, les séquences n'étant pas exactement identiques, nous avons du avoir recours à des amorces dégénérées.

Les amorces dégénérées peuvent être de deux types : soit la base qui est trop variable est remplacée par une base universelle, l'inosine (code:Z), soit les fabricants d'amorces peuvent effectuer des mélanges d'amorces contenant les diverses combinaisons. Le tableau ci-contre fait état du code utilisé pour désigner ces bases dégénérées :

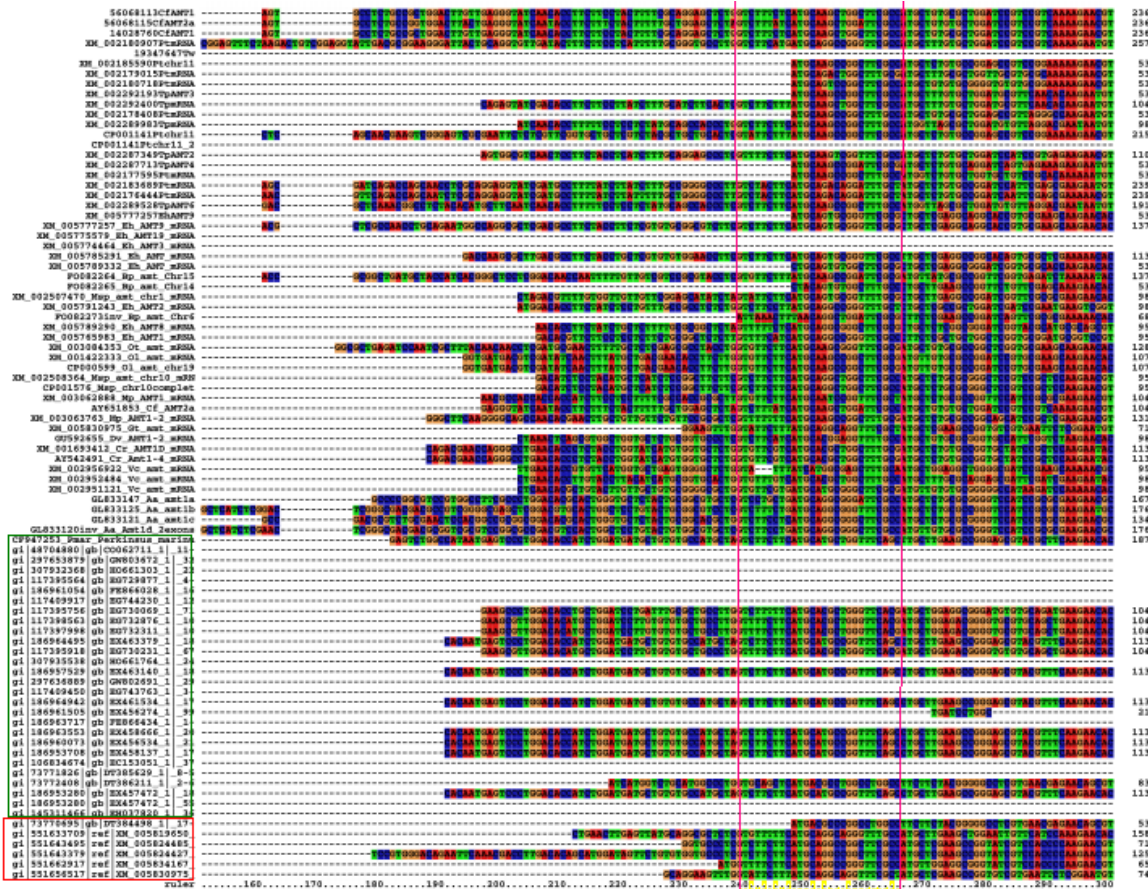
R	A	G		
Y	C	T		
M	A	C		
K	G	T		
S	C	G		
W	A	T		
H	A	C	T	
B	C	G	T	
V	A	C	G	
D	A	G	T	
N ou Z	A	C	G	T

Figure 16 : Code des bases dégénérées

Les deux sous parties suivantes présentent les régions que nous avons sélectionnées pour nos deux gènes d'intérêt et les amorces que nous avons codées :

a) *AMT1*

• Amorces sens : positionnement sur l'alignement



Amorces sens :

AMT1-F1 : **GTNTTYNTBATGCANDSNGGNTTYGC**

AMT1-F2i : **GTZTTYZTBATGCZZSZGGZTT**

Figure 17 : alignement pour amt1 avec clustalX

La première amorce commandée (AMT1-F1, position 240-266) était spécifique de tous les organismes sauf des dinoflagellés (encadré vert), c'est pourquoi lorsque nous avons fait notre deuxième commande (en même temps que celles pour NRT2 avec inosine) nous avons essayé des amorces AMT1 avec inosine et tronquées pour être spécifique de tous les organismes (AMT1-F2i).

• Amorces anti-sens :

Sur le même principe, voici les séquences des amorces anti-sens (position dans l'alignement de : 1315-1344). Elles sont complémentaire des séquences alignées.

AMT1-R1 : **TGNACNAYNCCNNWNCNGCRAARTC**

AMT1-R2i : **ACZAYZCCZZWZCCZGCRAARTC**

b) *NRT2*

Pour le gène *nrt2*, la démarche a été sensiblement différente puisque nous disposons des séquences des amorces fournies dans l'article de Kang et al (2011), celles-ci ont donc été adaptées pour correspondre à notre alignement.

Tableau 6: Conception d'Amorces NRT2 (astérisque = amorces commandées)

		Séquence de l'amorce	Séquence de la région ciblée sur l'ADN
Primers Kang et al (2011)	NrtF2*	GSZGGMTGGGGZAA MTZGGZGGZGG	Bis (car c'est une amorce sens)
	Nrtr7	GCW CCR RYR TTK CCZ CCZ GCZ CC	GG ZGC ZGG ZGG MAA YRY YGG WGC
Primer modifié	Nrtr7m*	GCW CCR RZR TTK CCT CCA GCZ CC	GG ZGC TGA ZGG MAA YZY YGG WGC

2. Ajout de Tails pour les amorces AMT (TAMT1F1et R1)

Les « tails » (sur le schéma en bleu et vert) sont des séquences indépendantes du gène d'intérêt qui sont ajoutées en bout des amorces spécifiques (violet et rouge) en 5'.

L'idée est d'obtenir des amplicons avec des extrémités de séquence fixée, ce qui augmente la fidélité d'hybridation lors d'éventuelles secondes PCR, rendant l'amplification plus efficace et non-sélective par rapport à la variabilité des régions génétiques cibles. Le principe est décrit par la figure ci-contre.

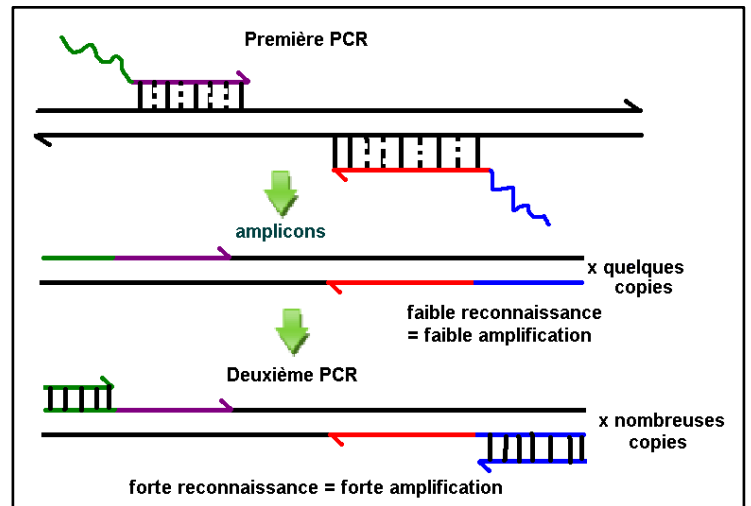


Figure 18 : principe de la "Tailed PCR"

Ces tails peuvent également avoir d'autres avantages, notamment lors du séquençage, car il peut être utile de localiser le début et la fin du produit de PCR au milieu de la séquence plasmidique.

Nous avons choisi pour les tails des séquences d'amorces universelles disponibles chez Eurofins : pEGFPC1for pour l'amorce sens et pCR3.1BGHRev pour l'antisens.

3. Vérifications sur les amorces

Bien que les amorces correspondent aux régions d'intérêt, il y a plusieurs vérifications à effectuer avant de les commander :

Tout d'abord, il est nécessaire de ne pas choisir des amorces incompatibles : la différence de température d'hybridation (ΔT_m) ne doit pas dépasser 5°C, elles ne doivent pas s'hybrider entre elles, surtout en 3' et ne doivent pas s'autohybrider au risque de former des structures secondaires et des amplicons de dimères d'amorces. Ceci a été vérifié grâce à l'outil en ligne OligoAnalyzer proposé par IDT-DNA.

Enfin, il a été calculé la taille prédictive des amplicons grâce aux alignements sur ClustalX, par simple soustraction des positions. Ce calcul est en effet préférable en vue du séquençage car, les diverses technologies n'offrent pas les mêmes possibilités de longueurs de séquençage.

Nous avons donc choisi des amorces produisant des amplicons de taille inférieure à 500pb, car le laboratoire dispose d'un séquenceur Sanger à Electrophorèse capillaire, ce qui nous évite des frais supplémentaires au cas où l'on doive faire appel à une entreprise extérieure pour le séquençage. Pour le gène AMT1, nous avons donc choisi des amorces produisant des amplicons de taille inférieure à 500 pb, de telle sorte que le protocole de séquençage puisse ultérieurement être adapté pour du séquençage à très haut débit de type Illumina MiSeq. En revanche pour NRT2, la taille estimée des amplicons (~700pb) ne serait actuellement accessible qu'à la technologie Roche 454.

Toutes nos amorces ont été commandées chez Eurofins genomics (AMT1F1, AMT1F2i, TAMT1F1, TAMT1R1, pEGFPC1for, pCR3.1BGHRev, NrtF2 et Nrtr7m).

E. Les PCRs

Lors de notre mise au point des techniques, nous nous sommes uniquement intéressés aux protocoles de PCR. Nous avons utilisés 3 kits différents : Les deux premiers étaient des kit Takara (PrimeSTAR GXL DNA polymerase / PrimeSTAR HS DNA polymerase) et le deuxième un kit Sigma (Jumpstart Taq DNA polymérase) avec une Taq Polymérase acceptant les bases inosines, ce qui n'était pas le cas des deux premières.

Nous n'avons pas testé toutes les possibilités, mais voici le résumé de notre démarche : (en gras, ce qui change)

1. Les Différentes approches successives de PCR (essais et mise au point) :

a) Essai de PCR n°1 : AMT 1, PCR classique et cultures pures

- Kit utilisé : Takara PrimeSTAR GXL DNA polymerase (Polymérase haute fidélité)
- Amorces :
 - AMT1-F1 et R1 (non-inosine)
 - contrôle positif grâce à des amorces sur ARNr 28S (cf partie II-C p.11)
- Matrice ADN : ADNc provenant de cultures pures d'*Alexandrium catanella* (tube A), de *Skeletonema costatum* (tube S) et de *Dunaliella tertiolecta* (tube D) : 1 uL pour 20uL final.
- PCR : classique, 1 session, comme indiqué sur le kit
- Résultat sur gel (agarose 1,5%) : pas de bande (contrôle positif validé)

b) Essai de PCR n°2 : AMT 1, « Tailed PCR », 2^e kit de PCR

- Kit utilisé : **PrimeSTAR HS DNA polymerase**
- Amorces :
 - **TAMT1-F1** et **TAMT1-R1** (amorces *Tailed*)
 - contrôle positif grâce à des amorces sur ARNr 28S
- Matrice ADN : cultures pures : tube A, S et D + **un échantillon naturel** de la station INS/août non mélangée (métagénomique).
- PCR : **2 sessions classiques** à la suite, sans rajout de matériel : la première aux Tm des séquences spécifiques seules, la deuxième à celui de la TAMT entière (plus haut).

- Résultat sur gel : pas de bande (contrôle positif validé), nous changeons donc la matrice (peu de diversité) et la méthode de PCR. Nous passons à un TouchDown.

c) Essai de PCR n°3 : AMT 1, « Tailed PCR » et TouchDown-PCR

- Amorces :
 - Trois tubes : TAMT1-F1 + TAMT1-R1 (TA), **couple précédent+tails seules (TT)**, AMT1-F1 et R1 (A).
 - **Pas de contrôle positif**
- Matrice ADN : **échantillon naturel** de la station **MeW/août** non mélangée
- PCR : 2 sessions à la suite, sans rajout de matériel : la première est en **TouchDown** (la température d'hybridation est diminuée à chaque cycle). Ceci permet de forcer l'hybridation tout en maintenant le maximum de spécificité. La deuxième session est classique.
- Résultat sur gel : pas de bande **sauf une faible pour le tube TT (et taille incorrecte)** combinant les 2 couples d'amorces : les Tails seules et les Amorces-Tailed.

d) Essai de PCR n°4 : AMT 1, « Tailed PCR », TD-PCR et kit Taq

- Kit utilisé : **Sigma Jumpstart Taq DNA polymérase**
- Amorces :
 - Trois tubes : TAMT1-F1 + TAMT1-R1 (TA), couple précédent + tails seules (TT), AMT1-F1 et R1 (A).
- Matrice ADN : **mélange M-I-A** (cf p. 15 tableau 2)
- PCR : 2 sessions à la suite, sans rajout de matériel : la première est en TouchDown, La deuxième session est classique, mais à une température d'hybridation basse (**55°C**).
- Résultat sur gel : **pas de bande**, même pour **TT**. Nous essayons donc de réutiliser les amplicons TA et TT comme base de la prochaine PCR.

e) Essai PCR n°5 : AMT1, 2° PCR emboîtée, Tails seules

- Amorces :
 - Trois tubes : tails seules (TA2, TT2) + un contrôle positif ARNr28S (tester la nouvelle polymérase Taq)
- Matrice ADN : **amplicons de la PCR précédente : TA et TT : 0,5uL**
- PCR : **classique, 1 session**, comme indiqué sur le kit
- Résultat sur gel : **bande nette pour TA2 et de taille correcte**. La piste de la combinaison TailedPCR + PCR classique avec Tails seules est retenue.



Figure 19 : Thermocycleur pour les PCR



Figure 21 : Bac d'électrophorèse

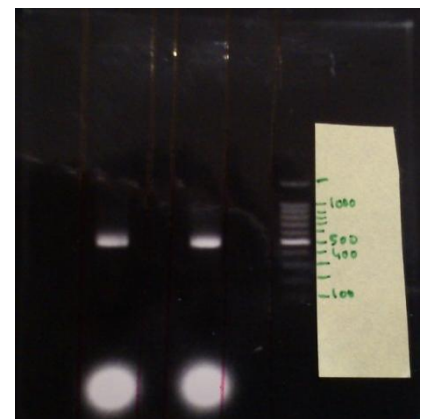


Figure 20 : Photo d'un gel positif (sous UV)

f) Essais PCR n°9 et 10: AMT1, NRT2 avec inosine (méthodes retenues)

- Matrice : Mix MIA
- AMT1 : **2 tubes**
 - **Amorces tailed + amorces inosine « TAAi3 »**
 - **Amorces tailed seules « TA3 »**
 - PCR : **2 sessions** à la suite, sans rajout de matériel : la première est en **TouchDown**, La deuxième session est classique, mais à une température d'hybridation basse.
- NRT2 : **5 tubes N20 « N20-1 » à «N20- 5»** (pour 100uL de produit PCR)
 - **NRT-F2 et NRTr7m**
 - PCR : **2 sessions** à la suite, sans rajout de matériel : la première est en **TouchDown**, La deuxième session est classique, mais à une température d'hybridation basse
- Résultats sur gel : **bande à 700pb pour N20 (taille correcte) et pas de bande pour les amplicons d'AMT**, comme attendu (se révèle à la deuxième PCR : essai n°11).

g) Essai PCR n°11 : deuxième PCR pour AMT1

- Matrice ADN : **amplicons précédents : TAAi3 et TA3 (0,5uL)**
 - Amorces : **tails seules (pEGFPC1for, pCR3.1BGHRev)**
- **4 tubes de 50uL** (duplicate pour obtenir 100uL de produit) **TAAi4 et TA4**
- PCR : **classique, 1 session**
- Résultat sur gel : **bandes nettes** pour les 2 produits de PCR. Les produits de PCR ainsi obtenus seront purifiés puis insérés dans des plasmides afin d'être clonés.

2. Purification des produits de PCR / des bandes du gel

Les produits de PCR, ainsi que les bandes excisées du gel, doivent être débarrassées des restes de la réaction de PCR (polymérase, amorces, dNTP, agarose...) afin d'être sûrs de n'insérer que les amplicons lors du clonage et ce sans perturbations par les anciens réactifs. Cette purification a été effectuée grâce à un kit (Purelink Quick Gel and PCR purification combo), dont le principe est très proche de celui utilisé pour purifier les ARN (cf. p.14 fig.13), avec un système de colonne retenant le matériel d'intérêt, des lavages et enfin une élution pour récupérer les amplicons.

F. Le Clonage : La séparation des séquences

L'objectif du clonage est de séparer les différents amplicons obtenus lors d'une même réaction de PCR, en vue du séquençage. En effet, les échantillons naturels étant un mélange d'espèces, les matrices des PCR sont différentes et ainsi les amplicons aussi. Le résultat lors du séquençage direct des amplicons serait alors complètement confus sans cette étape. Nous avons donc utilisé un kit (TOPO TA cloning kit for sequencing de Life Technologies), fournissant des plasmides TopoTA et des bactéries E.coli (TOP10, cryogénisées). La première étape consiste à insérer l'amplicon dans le plasmide, la deuxième à faire entrer le plasmide dans la bactérie (choc thermique), la troisième à séparer les bactéries pour n'avoir que des cultures pures sur une boîte de pétri avec un milieu solide sélectif (antibiotique) et enfin de repiquer les colonies pures dans un milieu liquide pour obtenir assez de matériel génétique pour la réaction de Sanger.

1. Insertion des amplicons dans les plasmides et mise en culture sur boîte

Les plasmides choisis (TA) sont spécialement conçus pour nos amplicons et pour le séquençage. En effet, La Taq polymérase utilisée (pour l'inosine) crée des embouts cohésifs collants T-A (thymine-Adénine), le vecteur devra alors aussi avoir ces embouts. Pour ce qui est des étapes de sélection, le plasmide comporte un gène de résistance à l'Ampicilline (sélection positive des bactéries transformées) mais également, la région d'insertion est au milieu d'un gène létalisant *ccdB*, qui permet une sélection négative des bactéries ayant intégré un plasmide vide. Ci-contre, la figure représente la carte (simplifiée) du vecteur en question.

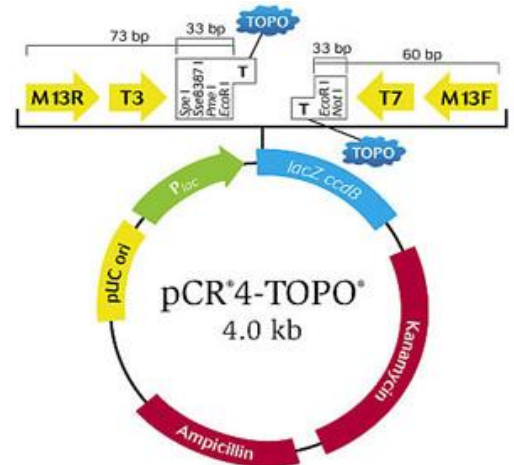


Figure 22 : carte simplifiée du vecteur utilisé

Ci-dessous, est décrit le protocole (simplifié) du kit (manipulations sous PSM) :

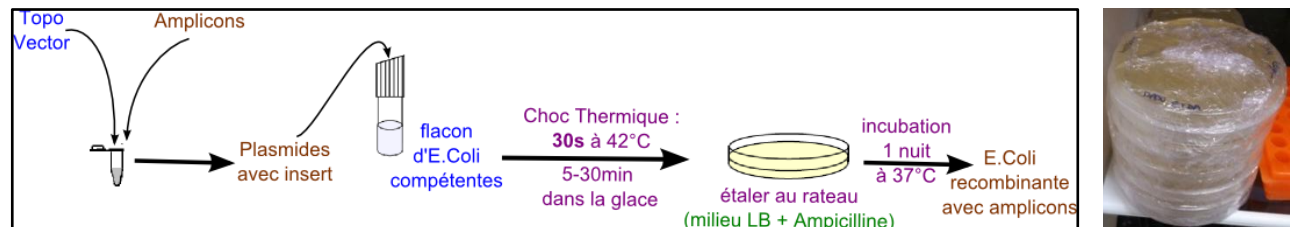


Figure 23 : protocole du Kit TOPO TA

Nous avons effectué en tout 9 boîtes : 3 volumes de suspension bactérienne différents multiplié par les 3 produits de PCR (N20, TAAi4 et TA4) à savoir 20, 40 et 90uL car le succès de la transformation et donc de la pousse reste imprévisible.

2. Mise en culture liquide

La mise en culture se fait par repiquage avec des cure-dents stériles (autoclavés) et inoculation dans du milieu LB liquide (5mL) contenant 75 µg/mL d'Ampicilline.

En tout, nous avons effectué 20 cultures pour N20, 18 pour TAAi4 et 18 pour TAAi4. Nous avons prévu une marge d'erreur de 22 tubes au cas où il aurait des problèmes de pousse (contamination, pas d'ensemencement, culture trop faible..).

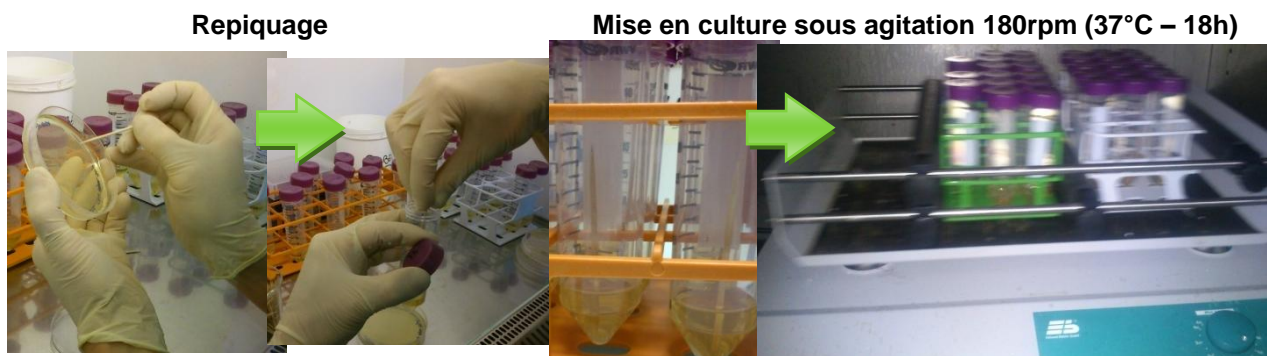


Figure 24 : protocole de repiquage et cultures liquides

3. Isolement de l'ADN et préparation pour le séquençage (quantification)

Après culture des colonies pures, il est nécessaire de procéder à une nouvelle extraction pour ne garder que les plasmides dans notre solution à séquencer. Nous choisissons les tubes présentant un trouble suffisant, indicateur de la pousse bactérienne (16 NRT2, 15 TA4 et 17 TAAi4 : 5% des tubes (8) n'ont pas poussé). Enfin, nous pourrions évaluer la quantité de matériel plasmidique dans nos extraits grâce à un appareil sophistiqué, le Nanodrop.

a) Extraction

L'extraction des plasmides ou miniprep est effectuée grâce à un kit (Purelink Quick Plasmid miniPrep), utilisant des colonnes d'extraction, dont le principe est similaire à celui utilisés pour l'isolement de l'ARN et des produits de PCR (cf p.14 fig.13).

b) Quantification du matériel plasmidique : Le NanoDrop

Le Nanodrop (Thermo scientific®) – photo ci-contre - est un spectrophotomètre de haute précision. En effet, il est capable de lire l'absorbance dans une goutte de 1 μ L d'une solution. Les solutions de plasmides sont donc ainsi mesurées une à une. En effet pour le mix de la réaction de S-PCR pour le séquençage, il est nécessaire d'avoir un minimum de 250ng/ul d'ADN pour obtenir une quantité suffisante détectable par le séquenceur.

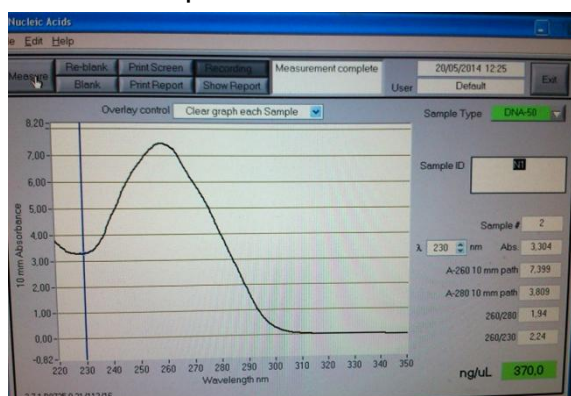
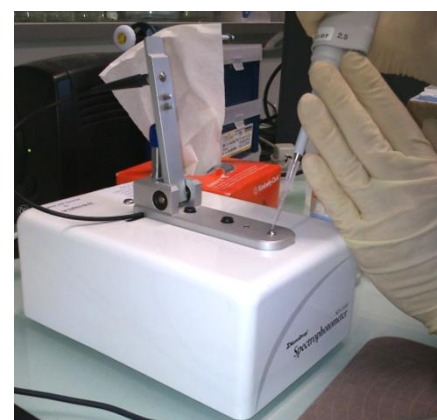


Figure 26 : exemple de mesure du Nanodrop

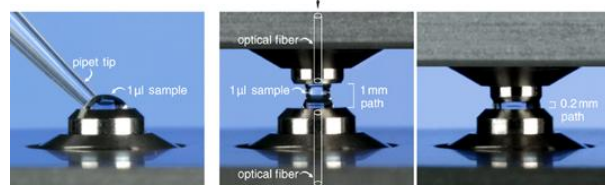


Figure 25 : Principe du Nanodrop

Sur nos 48 tubes purifiés, 5 n'ont pas la quantité suffisante. Le taux d'échec passe donc de 5 à 7% (8+5). Les autres sont alors utilisés comme matrice pour la PCR de séquençage, décrite ci-après.

G. Le séquençage

1. La S-PCR

La réaction de S-PCR repose sur le même principe que celui d'une PCR sauf qu'en plus des dNTP, il y a des ddXTP fluorescents, arrêtant l'élongation lorsque ceux-ci sont intégrés dans l'amplicon. Un mix déjà préparé est fourni par Life Technologies (BigDye Terminator V3.1 Cycle Sequencing Kit) contenant les ddXTP, la polymérase et les dNTPs). L'amorce utilisée est une de celles fournies dans le kit du vecteur (M13F, cf. fig. 21 p.23), choisie pour sa distance assez importante pour ne pas abimer le début de la séquence d'intérêt (bruit de fond) mais assez proche de l'insert pour ne pas avoir de trop grands amplicons à séquencer (limites du séquenceur). Les 48 tubes sont mis à précipiter, afin d'éliminer tous les réactifs de la S-PCR.

2. Purification des produits de la PCR par précipitation à l'éthanol

A chaque tube-produit de S-PCR sont ajoutés de l'acétate de sodium et de l'éthanol, qui permettent de neutraliser les phosphates de l'ADN (chargés positivement et donc solubles) et de précipiter par condensation en un réseau insoluble dans ces réactifs. Le surnageant, ne contient donc pas les grosses molécules d'ADN mais seulement les dNTP et les enzymes, à éliminer avant le séquençage. Le culot est ensuite ressolubilisé dans de l'eau mQ. Voici sur la figure ci-dessous, le protocole suivi :

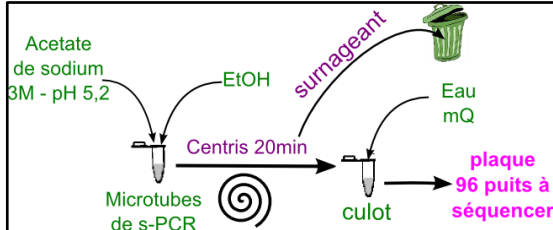


Figure 28 : Précipitation à l'éthanol des Produits de S-PCR



Figure 27 : plaque à séquençer

Les Tubes sont distribués dans une plaque 96 puits - cf. photo ci-dessus (48 puits utilisés : N20 : 18 puits, TA : 15 puits, TAAI : 15 puits) placée dans le séquenceur.

3. Le Séquenceur à électrophorèse capillaire

Le séquençage a été réalisé sur le plateau Séquençage-Génotypage du Centre Méditerranéen Environnement Biodiversité (Labex CeMEB) sur un séquenceur ABI3130XL, équipé de capillaires de 50 cm alimentés en polymère de type POP7.

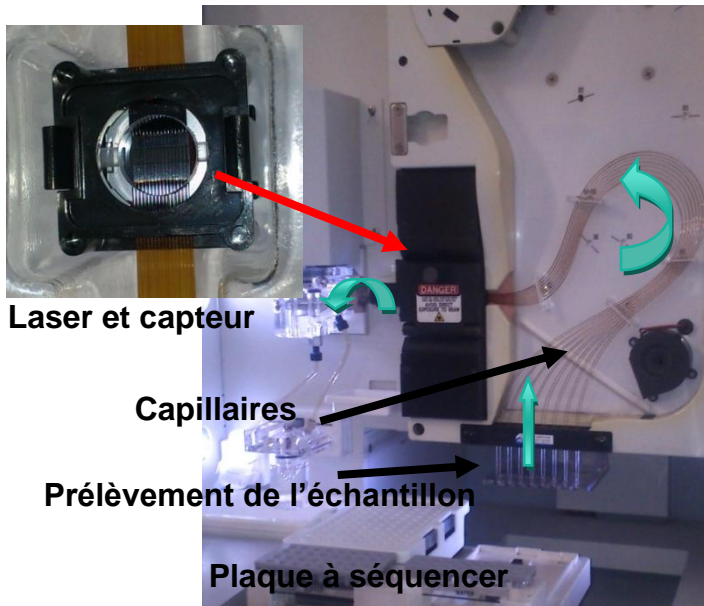


Figure 29 : intérieur du séquenceur

Les capillaires sont remplis avec un tampon de migration. Les acides nucléiques entament la migration grâce à un champ électrique. Ils sont attirés à travers le capillaire dans le même sens et, étant chargés de la même manière, se « rangent » en fonction de leur longueur. Les zones dénudées (cf. figure) du capillaire permet le passage du rayon, le capteur derrière enregistre la longueur d'onde.

Voici, ci-dessous, le type de résultat (chromatogramme) que l'on obtient en sortie d'appareil, lisible grâce au logiciel

SeqScanner. Ce logiciel permet de visualiser, de modifier et de récupérer les séquences obtenues. Les séquences sont ensuite inspectées pour discriminer celles correspondant au vecteur de celles correspondant aux amplicons insérés. Enfin, elles sont rassemblées dans un fichier Excel, Blastées sur GenBank pour vérifier qu'elles correspondent bien aux gènes-cibles *nrt* et *amt* et aux organismes attendus.

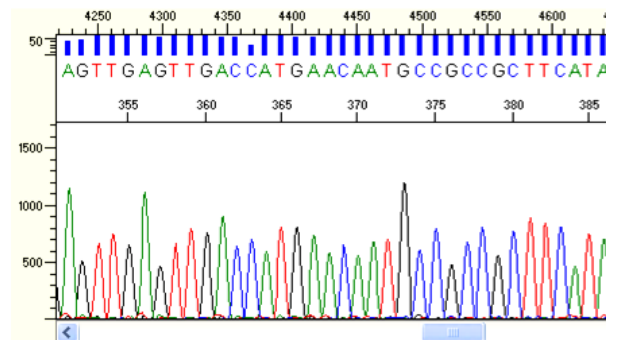


Figure 30 : Exemple de Chromatogramme (N8)

IV. RÉSULTATS ET DISCUSSION

Cette partie présente les résultats des méthodes de PCR, de purification retenue lors de notre mise au point du protocole (cf p.21-22). L'évaluation porte sur le taux de réussite de celle-ci.

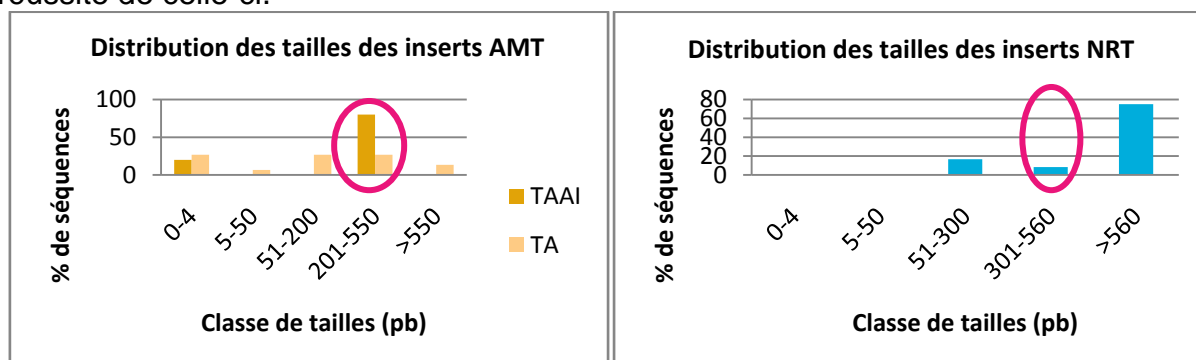


Figure 31 : Taille des inserts

Les graphes ci-dessus représentent les tailles des amplicons obtenus par nos trois stratégies de PCR. En rose sont encadrées les tailles attendues (AMT : 262-529pb ; NRT : 367-549pb) calculés sur les alignements (cf paragraphe). Pour ce qui est d'AMT, on observe que la méthode TAAi est plus efficace (80%) que la méthode TA (13%). Pour ce qui est de NRT, la méthode semble peu efficace, puisque es amplicons sont majoritairement de taille supérieure à celle attendue (75%).

Les résultats des BLASTs donnent peu de séquences référencées (**33% seulement des BLASTs sont positifs**). Pour TA, 2 séquences sur 15 (13%) sont positives (1 dans les tailles attendues et une de taille supérieure), pour TAAi, 5 positives sur 15 (33%) dont 4 dans les tailles prédites et 1 de taille inférieure, enfin pour NRT, la totalité des BLASTs positifs (9/12 soit 75%) sont de taille supérieure à celle attendue. Ceci confirme que la méthode TAAi est encore une fois meilleure que la méthode TA. Ces résultats doivent cependant être nuancés, car les tailles réelles dans notre échantillon peut être supérieure à celle de la base de données GenBank.

CONCLUSION

Mon stage a visé à évaluer des protocoles pour étudier l'expression des gènes amt et nrt d'échantillons naturels de phytoplancton. Le taux d'échec des cultures de bactéries transformées est de 5%, à ceci s'ajoute les tubes présentant trop peu de matériel ce qui fait passer le taux d'échec à 7%. La méthode TAAi est relativement bonne (33% de réussite) et sera à privilégier pour la suite de l'étude de l'expression du gène amt. Pour NRT, le taux de réussite est haut (75%), c'est une bonne méthode mais des tailles supérieures de 200pb à celles attendues restent à expliquer.

L'absence de correspondance des séquences dans la base de donnée genBank ne signifie pas que les séquences soient mauvaises. Il sera certainement nécessaire de créer notre propre librairie grâce au séquençage sur nos propres cultures pures algales.

La méthode semble correcte pour les deux gènes, bien qu'il reste encore des questionnements et un taux d'échec général assez important (80%). Elle doit donc être encore améliorée par d'autres essais. La poursuite de l'étude pourra alors se faire sur des cultures pures, également par l'utilisation de la qPCR, afin de quantifier station par station et saison par saison, l'expression de ces gènes.

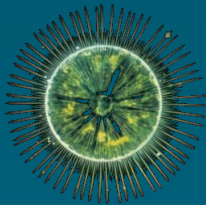
Ce stage m'a permis de réaliser des manipulations de biologie moléculaire, et ce, de l'échantillon, jusqu'au séquençage, ce qui a constitué un enseignement très riche.

RÉFÉRENCES BIBLIOGRAPHIQUES

- DUPRÉ N (2002). *Rôle des rejets des stations d'épuration dans l'eutrophisation des lagunes du Languedoc Roussillon. Exemple du complexe palavasien*. Rapport de stage de Diplôme d'Etudes Supérieures. Université Montpellier II. 202 pages.
- FALKOWSKI P.G. ET RAVEN J.A. (1997). *Aquatic photosynthesis*. Ed. Blackwell Science, 375 pages.
- FALKOWSKI P.G., KATZ M.E., KNOLL A.H. et al. (2004). *The evolution of modern eukaryotic phytoplankton*. Science, vol. 305, n° 5682, p. 354-360.
- FIELD C.B., BEHRENFELD M.J., RANDERSON J.T., et FALKOWSKI P. (1998). *Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components*. Science 10 July 1998: 281 (5374), p.237-240.
- GIBSON G., MUSE S.V., DOMENJOURD L., CUNIN R. (2004) (trad. Lionel Domenjoud), *Précis de génomique « 2 »*, Ed. De Boeck Université
- GRZEBYK D., BEC B., LASSERRE B., HATEY E., RIEUVILLENEUVE F., MALET N., AUDIC S., DE VARGAS C. (2013). *Diversité génétique du phytoplancton selon un gradient d'eutrophisation dans des lagunes côtières en voie de restauration écologique (DIVTROPHE_LAGUNES)*. Rapport d'activité APEGE INEE 2012, 6 pages.
- HILDEBRAND M. et DAHLIN K. (2000). *Nitrate transporters genes from the diatom *Cylindrotheca fusiformis* (Bacillariophyceae): mRNA levels controlled by nitrogen source and by the cell cycle*. J. Phycol. 36: p.702-713.
- KANG L.-K., WANG H.-F., CHANG J. (2011). *Diversity of phytoplankton nitrate transporter sequences from isolated single cells and mixed samples from the East China Sea and mRNA quantification*. Appl.Environ.Microb.77: p.122-130.
- SONG B. ET WARD B.B. (2007). *Molecular cloning and characterization of high-affinity nitrate transporters in marine phytoplankton*. Journal of phycology, 43(3), p.542-552.
- VON WITTGENSTEIN, NJJB, LE C.H., HAWKINS B.J. et al (2014). *Evolutionary classification of ammonium, nitrate, and peptide transporters in land plants*. BMC evolutionary biology, vol. 14, n°1, p. 11.

Caractérisation du transcriptome des transporteurs de nutriments azotés de communautés phytoplanctoniques

Mise au point de Protocoles de Biologie Moléculaire



Résumé

Ce rapport décrit le travail effectué lors de mon stage au sein du laboratoire ECOSYM, à savoir la mise au point de protocoles de biologie moléculaire en vue d'obtenir des empreintes génétiques d'organismes phytoplanctoniques exprimant des gènes (transcriptome) impliqués dans les processus d'absorption des nutriments minéraux azotés. Les deux gènes de transporteurs *nrt2* (nitrate) et *amt1* (ammonium) sont ici ciblés. Le développement des protocoles a commencé par la création d'une librairie de gènes correspondants identifiés dans les bases de données publiques (GenBank), puis par la construction d'alignements des protéines et la recherche de régions conservées pour le dessin d'amorces oligonucléotidiques. Ces amorces ont permis l'amplification par PCR de fragments géniques d'une longueur appropriée par rapport aux technologies de séquençage disponibles au laboratoire (Sanger et électrophorèse capillaire). Les protocoles ont été testés par des essais de PCRs sur des cultures algales et des échantillons de communautés naturelles. Les fragments amplifiés ont été clonés par insertion dans un plasmide et par transformation de bactéries *E. coli*, avant leur séquençage par la méthode de Sanger, pour séparer les différents amplicons. Enfin, pour vérifier leur homologie avec les gènes ciblés, nous avons finalement procédé à la comparaison des séquences obtenues avec celles des bases de données publiques (BLAST sur GenBank).

A terme, la quantification de l'expression de ces gènes par cette méthode pourrait constituer un éventuel bioindicateur, quantifiant la réponse de ces organismes aux conditions nutritionnelles du milieu et peut-être un moyen d'explorer les relations de compétition entre ceux-ci lors d'épisodes de raréfaction des nutriments.

Mots clés

diversité génétique et fonctionnelle, lagunes palavasiennes, mise au point de protocoles de biologie moléculaire, séquençage, communautés phytoplanctoniques, transcriptome, métagénomique, assimilation des nutriments azotés

