



**HAL**  
open science

## Caractérisation d'éléments transposables de type mariner chez les microalgues marines

Dorothee Hermann

► **To cite this version:**

Dorothee Hermann. Caractérisation d'éléments transposables de type mariner chez les microalgues marines. Sciences agricoles. Université du Maine, 2011. Français. NNT : 2011LEMA1021 . tel-00732952

**HAL Id: tel-00732952**

**<https://theses.hal.science/tel-00732952>**

Submitted on 17 Sep 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# UNIVERSITÉ DU MAINE

UFR Sciences et Techniques

**École Doctorale VENAM**

(Végétal – Environnement – Nutrition – Agroalimentaire – Mer)

## THÈSE

Présentée en vue d'obtenir le grade de

**Docteur de l'Université du Maine**

Spécialité : Biologie des Organismes

Présentée par

**Dorothee HERMANN**

## **Caractérisation d'éléments transposables de type *mariner* chez les microalgues marines**

**Soutenue le 16 Mars 2011 devant le jury composé de :**

<b>Florence ROULEUX-BONNIN</b> , Maître de conférences HDR, Université de Tours,	Rapporteur
<b>Emmanuelle LERAT</b> , Chargée de recherche CNRS, HDR, Université de Lyon 1,	Rapporteur
<b>Catherine BOYEN</b> , Directrice de recherche CNRS, Station Biologique de Roscoff,	Examineur
<b>Olivier PANAUD</b> , Professeur, Université de Perpignan	Examineur
<b>Annick MORANT-MANCEAU</b> , Professeur, Université du Mans,	Directrice de thèse
<b>Nathalie CASSE</b> , Maître de conférences, HDR, Université du Mans,	Co-Directrice de thèse
<b>Aurore CARUSO</b> , Maître de conférences, Université du Mans,	Co-encadrante de thèse

# REMERCIEMENTS

Je tiens tout d'abord à remercier les personnes qui ont accepté de faire parti du jury de thèse et qui ont évalué ce travail.

Merci à Mesdames Florence ROULEUX-BONNIN et Emmanuelle LERAT de me faire l'honneur de consacrer une partie de leur temps à rapporter ce travail au travers de l'examen de ce long mémoire. Je remercie vivement Madame Catherine BOYEN et Monsieur Olivier Panaud d'avoir accepté d'être examinateurs de ce travail.

En second lieu, je voudrais remercier celles qui m'ont fait confiance et qui m'ont permis de réaliser cette thèse de doctorat, ma directrice, co-directrice et co-encadrante : Madame et Mesdemoiselles Annick MANCEAU, Nathalie CASSE et Aurore CARUSO pour avoir encadré dans leurs spécialités respectives chacune des approches de ce travail et pour m'avoir beaucoup appris tant sur le plan scientifique que personnel. Merci pour toutes vos corrections et pour avoir supporté mes nombreuses questions sur tout, ainsi que mes anxiétés.

Je voudrais exprimer ma gratitude également aux directeurs des deux laboratoires de MMS qui m'ont accueilli, Messieurs Gérard TREMBLIN et Benoît CHENAIS.

Je n'oublie pas et j'exprime mes vifs remerciements et mon amitié sincère à Monsieur J. D. Rouault, qui a contribué à changer ma façon de voir et de faire la recherche et qui a également encadré une partie de ce travail, de même qu'à Madame Q. T. BUI qui m'a souvent conseillée et orientée sur un des aspects de cette thèse.

Je remercie tout particulièrement les « acteurs techniques » de la réalisation de ce travail. J'ai pleinement conscience que je leur dois beaucoup dans la réalisation de cette thèse, merci à Boris, Yohan (mon premier stagiaire et quel stagiaire !) et Elisabeth.

A toutes les personnes que je vais citer ci-dessous, vous avez tous contribué d'une manière ou d'une autre à la réalisation et à la clôture de ce travail, que ce soit par des conversations et des remarques constructives ou des aides informatiques etc. et je vous en remercie et je suis sûre que vous savez à quel point vous avez compté, je ne ressens donc pas le besoin de le détailler ici. Merci à Yann, Vincent, Jean Luc, Romain, Brigitte, Alain, Justine, Marie, Muriel, Laurence, Nhung, Sophie, Frédéric, Anne, Freddy, Fadoumo.

Je voudrais également exprimer ma gratitude aux membres du conseil de l'école doctorale VENAM et aux membres du conseil du collège doctoral de l'Université du Mans, qui m'ont fait confiance en tant que représentante des doctorants avec une dédicace spéciale aux membres du comité organisateur des Doctoriales 2010, pour m'avoir écouté, permis de

m'épanouir et de prendre des responsabilités au sein de la commission Entreprise. J'ai beaucoup appris dans cette expérience et cela m'a ouvert un nouveau champ de perspectives.

Pour leur soutien dans ma vie personnelle et au travail, je remercie de tout mon cœur : Boris, Aurélie, Anne-C., Will, Yves, Mafia, Berretta, Yann, mes parents, ma marraine, ma sœur et mon neveu, Georges, Vanessa, Katie, Mansour, Jean-Claude, Ludovic, Laurence, Jérôme, Laurent, Bérange, Romain, Marina, et tous ceux que j'oublie ici mais pas dans la vie !

Durant ces trois ans de recherche, j'ai bénéficié d'un financement de la région Pays de la Loire.

Pour finir je remercie les passionnés qui réalisent des œuvres d'art avec les diatomées et qui contribuent à faire connaître au grand public ces organismes extraordinaires. Les photos sont disponibles à cette adresse :

<http://bensonapbiology.concordcarlisle.wikispaces.net/The+Wooooooollly+Mammoth+Microbial+Diversity+Lab>.

# TABLE DES MATIERES

<b>TABLE DES ILLUSTRATIONS</b>	<b>8</b>
<b>LISTE DES TABLEAUX</b>	<b>11</b>
<b>LISTE DES ABREVIATIONS</b>	<b>12</b>
<b>INTRODUCTION GENERALE</b>	<b>13</b>
<b>CHAPITRE I</b>	<b>16</b>
<b>INTRODUCTION BIBLIOGRAPHIQUE</b>	<b>16</b>
<b>I Les diatomées</b>	<b>18</b>
I.1 Le rôle écologique des diatomées	18
I.2 La biologie des diatomées	18
I.3 La photosynthèse des diatomées	20
<i>I.3.1 L'appareil photosynthétique</i>	20
<i>I.3.2 La photosynthèse</i>	23
<i>I.3.3 La fluorescence chlorophyllienne</i>	25
I.4 Les stress	25
<i>I.4.1 Définition d'un stress</i>	25
<i>I.4.2 La réponse au stress</i>	26
<i>I.4.3 Effet du stress thermique sur le métabolisme des diatomées</i>	27
<i>I.4.4 Le stress oxydant</i>	28
I.5 Valorisation des diatomées	30
I.6 Génétique et génomique des diatomées	32
<b>II Les Éléments Transposables</b>	<b>34</b>
II.1 Définition	34
II.2 Classification des ET	34
<i>II.2.1 Les superfamilles de la classe I</i>	36
<i>II.2.2 Les superfamilles de la classe II</i>	40
<i>II.2.3 Les éléments non autonomes</i>	43
II.3 Place des ET dans les génomes : diversité et dynamique des ET	44
<i>II.3.1 Diversité des ET dans les génomes</i>	44
<i>II.3.2 Impact des ET sur la taille des génomes</i>	45
<i>II.3.3 Dynamique des ET dans les génomes</i>	46
<i>II.3.4 La transmission des ET</i>	49
II.4 Impact des ET sur les génomes et plasticité génomique	50
II.5 Influence des facteurs environnementaux sur les ET et le génome	51
<b>III La superfamille <i>Tc1-mariner</i></b>	<b>58</b>
III.1 La famille <i>MLE</i> ( <i>mariner</i> -like elements)	59
III.2 La famille <i>TLE</i> ( <i>Tc1</i> - like elements)	62
III.3 Caractéristiques structurales des <i>MLE</i> et des <i>TLE</i>	63
<b>CHAPITRE II</b>	<b>66</b>
<b>MATERIELS ET METHODES</b>	<b>66</b>

<b>I Matériel végétal et conditions de culture</b>	<b>67</b>
I.1 Origine géographique des microalgues	67
I.2 Les microalgues	67
I.3 Conditions de culture de référence	68
I.4 Stress thermiques	69
I.4.1 Conditions de stress thermiques	69
I.4.2 Détermination des températures de stress thermiques	69
I.4.3 Application de stress thermiques de courte durée	69
<b>II Paramètres physiologiques</b>	<b>71</b>
II.1 La croissance	71
II.1.1 Mesure de la densité cellulaire	71
II.1.2 Courbes de croissance	71
II.2 Les paramètres photosynthétiques	72
II.2.1 Extraction et dosage de la Chlorophylle a	72
II.2.2 Mesure de l'intensité photosynthétique	73
II.2.3 Mesure de la fluorescence de la chlorophylle a	73
II.3 Récolte de la biomasse algale	75
<b>III Techniques de biologie moléculaire</b>	<b>76</b>
III.1 Extraction des acides nucléiques	76
III.1.1 Extraction de l'ADN génomique	76
III.1.2 Extraction des ARN totaux	76
III.2 Traitement à la DNase	77
III.3 Contrôle de la qualité et de la quantité d'acides nucléiques	77
III.4 Transcription inverse	78
III.5 Amplification par PCR	78
III.6 Clonage des fragments d'ADN	81
III.7 Transformation bactérienne et clonage des fragments d'intérêt	81
III.8 Extraction plasmidique	82
III.9 Séquençage des fragments d'intérêt	82
<b>IV Approche bio-informatique</b>	<b>82</b>
IV.1 Analyse des séquences	82
IV.2 Recherche des ET dans le génome séquencé de <i>P. tricornutum</i> souche CCMP632/ CCAP1055.1	83
IV.2.1 Recherche des MLE dans le génome séquencé	83
IV.2.2 Recherche des MLE complets dans le génome séquencé	83
IV.3 Méthodes phylogénétiques	87
IV.4 Classification agrégative des ET	87
IV.4.1 Définition de la classification agrégative	88
IV.4.2 Matériel, ressources et données utilisées	88
IV.4.3 Méthode	89
IV.4.4 Le processus agrégatif UPGM-VM	90
<b>CHAPITRE III</b>	<b>93</b>
<b>RESULTATS ET DISCUSSION</b>	<b>93</b>
<b>I Recherche de MLE chez les microalgues marines</b>	<b>94</b>
I.1 Caractérisation moléculaire des MLE	94
I.2 Identification des MLE	96
I.3 Analyse des séquences de MLE	97
I.3.1 Analyse des séquences nucléiques	97
I.3.2 Analyse des séquences protéiques	99
I.4 Analyse phylogénétique des MLE de diatomées marines	107
I.5 Recherche des MLE dans le génome séquencé de <i>Phaeodactylum tricornutum</i> (souche CCMP632 /CCAP1055.1)	111
I.5.1 Détection bioinformatique des MLE dans la banque de données JGI	111

1.5.2 Recherche bioinformatique à l'aide des séquences signatures (Target Site Duplication) TSD dans le génome de <i>P. tricornutum</i>	111
1.5.3 Recherche bio-moléculaire de séquence MLE dans le génome de <i>P. tricornutum</i> .	114
I.6 Discussion	117
1.6.1 Détection de MLE chez les diatomées marines	117
1.6.2 Caractéristiques des MLE chez les diatomées marines	119
1.6.3 Polymorphisme et évolution des MLE chez les diatomées marines	123
1.6.4 Présence des MLE dans le génome séquencé de <i>P. tricornutum</i>	126
<b>II Positionnement des MLE de diatomées par rapport aux MLE de plantes et dans la superfamille <i>Tc1-mariner</i></b>	<b>131</b>
II.1 Position phylogénétique des MLE de diatomées par rapport aux MLE de plantes	131
II.2 Classification des MLE de diatomées dans la superfamille <i>Tc1-mariner</i>	133
II.2.1 Position des MLE végétaux dans la classification DTT491 de la superfamille <i>Tc1-mariner</i> .	135
II.2.2 Position et propriétés des groupes de MLE végétaux composant la famille <i>Chlorophyllis</i> dans la classification DTT491	139
II.2.3 Position des MLE des diatomées marines dans la classification DTT491.	141
II.3 Discussion	142
II.3.1 Position phylogénétique et évolution des MLE de diatomées marines par rapport aux MLE de végétaux terrestres	142
II.3.2 Classification des MLE de diatomées marines par rapport aux MLE de plantes dans la nouvelle famille <i>Chlorophyllis</i> de la superfamille <i>Tc1-mariner</i>	146
<b>III Impact de stress thermiques courts sur l'expression des MLE de trois diatomées marines</b>	<b>151</b>
III.1 Réponses physiologiques des trois diatomées soumises à des stress thermiques	151
III.1.1 Effets des stress thermiques sur la croissance	152
III.1.2 Effets des chocs thermiques sur l'intensité photosynthétique	154
III.1.3 Effets des chocs thermiques courts sur les paramètres de la fluorescence de la chlorophylle <i>a</i>	155
III.2. Expression des MLE en conditions de stress thermiques	158
III.2.1 Détection de l'expression des MLE et des gènes « contrôle » en conditions de stress thermiques chez <i>P. tricornutum</i>	159
III.2.2 Détection de l'expression des MLE et des gènes « contrôle » en conditions de stress thermiques chez <i>A. coffeaeformis</i>	160
III.2.3 Détection de l'expression des MLE et des gènes « contrôle » en conditions de stress thermiques chez <i>A. acutiuscula</i>	164
III.3 Identification des gènes MLE, <i>Blackbeard</i> et <i>Surcouf</i> exprimés dans les conditions expérimentales	166
III.4 Discussion	171
III.4.1 Conditions de stress thermiques et thermo-tolérance des diatomées	171
III.4.2 Estimation du stress subi par analyse des paramètres photosynthétiques	173
III.4.3 Expression des gènes marqueurs de la réponse aux stress et de l'expression génique	175
III.4.4 Expression des ET chez les diatomées	177
III.4.5 Induction des ET par le stress	179
<b>CONCLUSIONS ET PERSPECTIVES</b>	<b>185</b>
<b>BIBLIOGRAPHIE</b>	<b>192</b>
<b>ANNEXES</b>	<b>225</b>
Annexe 1 : Maintien des cultures de microalgues et axénie	226
Annexe 2 : Eau de Mer Artificielle (EMA)	227
Annexe 3 : Tableau récapitulatif des ET classés avec la méthode agrégative UPGM-VM dans la classification DTT491.	228
Annexe 4 : Etapes de la procédure d'agrégation de la méthode UPGM-VM	235

<b>Annexe 5 : Alignement des séquences de <i>MLE</i> de ~ 380 pb traduites <i>in silico</i> obtenues chez les 10 espèces de diatomées.</b>	<b>237</b>
<b>Annexe 6 : Les matrices d'identités</b>	<b>240</b>
<b>Annexe 7 : Méta-analyse des <i>MLE</i> de végétaux</b>	<b>242</b>
<b>Annexe 8 : Article soumis pour publication dans <i>Molecular Phylogenetics and Evolution</i>.</b>	<b>244</b>
<b>COMMUNICATIONS ET PUBLICATIONS</b>	<b>267</b>

# Table des Illustrations

Figure 1 : Frustule d'une diatomée pennée.....	19
Figure 2 : Multiplication clonale d'une diatomée par mitose.....	20
Figure 3 : A-Coupe dans des cellules de <i>Phaeodactylum tricorutum</i> renfermant deux chloroplastes observés en microscopie électronique (d'après Falkowski et Raven, 1997). B-Schéma de la structure d'un chloroplaste de diatomée avec des thylakoïdes empilés par trois. ....	22
Figure 4 : Schéma de l'organisation de la membrane d'un thylakoïde. ....	24
Figure 5 : Schéma des voies de dissipation de l'énergie lumineuse au niveau du photosystème II. ....	24
Figure 6 : Réactions catalysées par trois enzymes antioxydantes : superoxydes dismutase (SOD), catalase et ascorbate peroxydase. ....	29
Figure 7 : Cascade de réactions de détoxication des formes réactives de l'oxygène (ERO) produites lors d'un stress abiotique (d'après Mittler, 2002).....	30
Figure 8 : Schéma de la classification des ET selon les modes de transposition. ....	35
Figure 9 : Schéma général d'un rétrotransposon. ....	36
Figure 10 : Classification des éléments transposables (d'après Wicker et al., 2007). ....	39
Figure 11 : Schéma général d'un transposon. ....	40
Figure 12 : Principe de la transposition en <i>trans</i> des éléments non-autonomes à partir de la machinerie des éléments autonomes (d'après Feschotte et al., 2002).....	44
Figure 13 : Distribution des ET de classe II dans les génomes eucaryotes (d'après Feschotte et Pritham, 2007). ....	45
Figure 14 : Les trois phases du cycle de vie des ET dans les génomes (d'après Kidwell et Lisch, 2001). ...	47
Figure 15 : Schéma simplifié du mode de transfert vertical et horizontal des ET (d'après Eickbush et Malik, 2002). ....	50
Figure 16 : Modèle du rôle des ET dans les mécanismes d'adaptation au stress (d'après Capy et al., 2000). ....	52
Figure 17 : Changements phénotypiques liés à l'activité de <i>Tam3</i> dans le gène <i>niv</i> du muflier ( <i>Antirrhinum majus</i> ). ....	53
Figure 18 : Modèle écologique d'évolution des ET dans les génomes en relation avec les facteurs environnementaux selon Biéumont et Vieira (2006). ....	55
Figure 19 : Classification des éléments de la superfamille ITm en fonction des domaines de la triade catalytique (d'après Shao et Tu, 2001). ....	59
Figure 20 : Définition des amorces <i>MLE5A</i> et <i>MLE3A</i> par Feschotte et Wessler (2002). ....	60
Figure 21 : Représentation schématique des transposons de la famille des <i>MLE</i> . ....	63
Figure 22 : Représentation simplifiée de la structure des transposases des <i>TLE</i> , <i>MLE</i> animaux et des <i>MLE</i> végétaux. ....	65
Figure 23 : Dispositif expérimental utilisé pour appliquer les stress thermiques aux trois diatomées <i>Phaeodactylum tricorutum</i> , <i>Amphora coffeaeformis</i> et <i>Amphora acutiuscula</i> . ....	70
Figure 24 : Courbe de croissance théorique des diatomées cultivées en batch en fonction du temps. ....	72

Figure 25 : Schéma de la technique de fluorimétrie modulée pour mesurer la fluorescence chlorophyllienne d'une suspension de microalgues.....	74
Figure 26 : Evolution de la fluorescence chlorophyllienne d'une suspension de microalgues. ....	75
Figure 27 : Représentation schématique d'un transposons <i>MLE</i> végétal. ....	84
Figure 28 : Schéma des étapes de détection <i>in silico</i> des éléments à TIR dans le génome de <i>Phaeodactylum tricornutum</i> séquencé .....	86
Figure 29 : Définition d'un Clan au cours du processus d'agrégation (D'après Rouault et al., 2009).....	92
Figure 30 : Amplification des fragments de <i>MLE</i> chez 18 espèces de microalgues. ....	95
Figure 31 : Alignement des séquences consensus <i>MLE</i> obtenues chez les dix diatomées et traduites <i>in silico</i> .....	104
Figure 32 : Représentation en WebLogo de l'alignement des consensus traduits en acides aminés.....	105
Figure 33 : Analyse en Maximum de Parcimonie (MP) des séquences <i>MLE</i> consensus de diatomées traduites en protéines. ....	105
Figure 34 : Matrice d'identité des séquences consensus de <i>MLE</i> traduites <i>in silico</i> .....	106
Figure 35 : Amplification de fragments (~380 pb) de transposases <i>MLE</i> dans la souche de laboratoire NCC340 et la souche séquencée CCMP632 de <i>Phaeodactylum tricornutum</i> à l'aide des amorces <i>MLE5A</i> et <i>MLE3A</i> . ....	115
Figure 36 : Amplifications de fragments de transposases <i>MLE</i> (~ 170 pb) dans la souche de laboratoire NCC340 et la souche séquencée CCMP632 de <i>Phaeodactylum tricornutum</i> à l'aide des amorces spécifiques <i>Phaeo4</i> et <i>Phaeo7</i> .....	115
Figure 37 : Alignement des séquences <i>MLE</i> de 380 pb consensus obtenues avec la souche du laboratoire NCC340 et deux séquences de la souche CCMP632 de <i>Phaeodactylum tricornutum</i> traduites <i>in silico</i> ... ..	115
Figure 38 : Similarités des séquences de <i>MLE</i> de diatomées et des <i>MLE</i> de plantes. ....	122
Figure 39 : Représentation schématique des processus évolutifs supposés à l'origine des groupes phylogénétiques hétérogènes des <i>MLE</i> de diatomées. ....	126
Figure 40 : Position phylogénétique des <i>MLE</i> de diatomées marines par rapport aux <i>MLE</i> de plantes terrestres. ....	129
Figure 41 : Rosace obtenue à l'aide de la méthode UPGM-VM représentant la classification agrégative des <i>MLE</i> de diatomées et de végétaux terrestres par rapport aux ET de la superfamille <i>Tc1-mariner</i> . ....	134
Figure 42 : Structuration de la famille Chlorophyllis : classification des <i>MLE</i> de diatomées et de végétaux terrestres dans l'analyse DTT491. ....	138
Figure 43 : Structure de la sous-famille Dia, PAL et Sta.....	140
Figure 44 : Représentation schématique de l'évolution conjointe des diatomées et des plantes .....	144
Figure 45 : Courbes de croissance de <i>Phaeodactylum tricornutum</i> cultivée à 8, 16 et 24 °C.....	152
Figure 46 : Courbes de croissance de <i>Amphora coffeaeformis</i> cultivée à 8, 16 et 24 °C. ....	152
Figure 47 : Courbes de croissance de <i>Amphora acutiuscula</i> cultivée à 8, 16 et 24 °C. ....	152
Figure 48 : Intensité photosynthétique chez <i>Phaeodactylum tricornutum</i> , <i>Amphora coffeaeformis</i> et <i>Amphora acutiuscula</i> après 5 heures de stress à 4, 8, 16, 24 et 32 °C.....	155
Figure 49 : Evolution des rendements quantiques maximum (Fv/Fm) et effectif (φPSII) chez <i>Phaeodactylum tricornutum</i> à 4, 8, 16, 24 et 32 °C.....	156

Figure 50 : Evolution des rendements quantiques maximum (Fv/Fm) et effectif ( $\phi$ PSII) chez <i>Amphora coffeaeformis</i> à 4, 8, 16, 24 et 32 °C. ....	157
Figure 51 : Evolution des rendements quantiques maximum (Fv/Fm) et effectif ( $\phi$ PSII) chez <i>Amphora acutiuscula</i> à 4, 8, 16, 24 et 32 °C. ....	157
Figure 52 : Amplification des <i>MLE</i> et des gènes de contrôle par PCR sur ADNc et ADNg chez <i>Phaeodactylum tricornutum</i> cultivée à 16 °C et après 5 h de stress à 4, 8, 24 et 32 °C. ....	161
Figure 53 : Amplification des <i>MLE</i> et des gènes de contrôle par PCR sur ADNc et ADNg chez <i>Amphora coffeaeformis</i> cultivée à 16 °C et après 5 h de stress à 4, 8, 24 et 32 °C. ....	163
Figure 54 : Amplification des <i>MLE</i> et des gènes de contrôle par PCR sur ADNc et ADNg chez <i>Amphora acutiuscula</i> cultivée à 24 °C et après 5 h de stress à 4, 8, 16 et 32 °C.....	165
Figure 55 : Alignement des séquences nucléiques de <i>MLE</i> obtenues avec les ADNc et en ADNg chez <i>Phaeodactylum tricornutum</i> . ....	168
Figure 56 : Alignement des séquences nucléiques de <i>MLE</i> obtenues avec les ADNc et en ADN g chez <i>Amphora coffeaeformis</i> . ....	169
Figure 57 : Alignement des séquences nucléiques de <i>MLE</i> obtenues avec les ADNc et en ADNg chez <i>Amphora acutiuscula</i> . ....	170

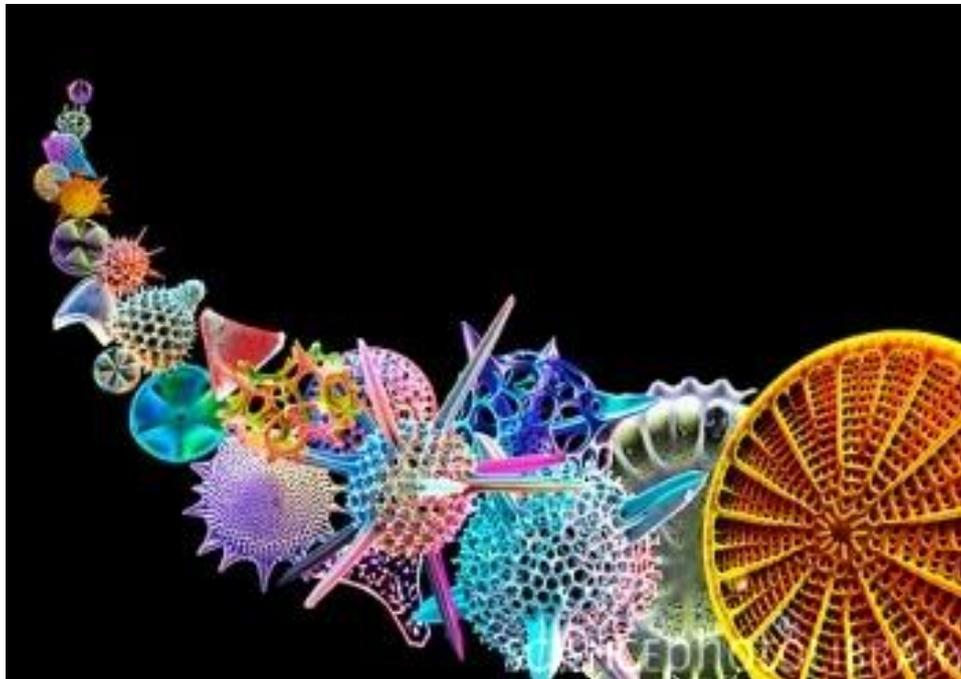
# LISTE DES TABLEAUX

Tableau 1 : Pourcentage d'ET dans les génomes modèles (modifié d'après Kidwell, 2002). .....	46
Tableau 2 : Réponse des ET aux stress abiotiques et biotiques.....	56
Tableau 3 : Les éléments potentiellement actifs appartenant à la famille <i>MLE</i> . .....	62
Tableau 4 : Liste des espèces utilisées, leur taxonomie, leur origine géographique et leur intérêt économique ou écologique. ....	67
Tableau 5 : Amorces dégénérées (Feschotte et Wessler, 2002) utilisées pour mettre en évidence la présence de <i>MLE</i> dans les génomes de microalgues.....	79
Tableau 6 : Amorces dégénérées ou spécifiques utilisées pour mettre en évidence la présence des gènes de ménage et de leurs transcrits.....	79
Tableau 7 : Amorces spécifiques utilisées pour mettre en évidence la présence des gènes de réponse au stress (oxydant et thermique) et leurs transcrits. ....	80
Tableau 8 : Amorces spécifiques utilisées pour la détection des éléments transposables et celle de leur transcrits chez les diatomées. ....	80
Tableau 9 : Liste des espèces testées, nombre de séquences clonées et nombre de séquences identifiées par espèce.....	97
Tableau 10 : Longueur des séquences <i>MLE</i> obtenues chez dix diatomées, nombre de cadre de lecture ouvert putatif (ORF), nom des consensus et des séquences et numéros d'accèsion. ....	100
Tableau 11 : Séquences obtenues par recherche bio-informatique au sein du génome de <i>Phaeodactylum tricornutum</i> .....	113
Tableau 12 : Propriétés de la superfamille <i>Tc1-mariner</i> (DTT) et de ses principales familles. ....	136
Tableau 13 : Propriétés de la famille Chlorophyllis et des groupes issus de diatomées marines (DTTChlDia-->) et de plantes terrestres (DTTChlPln-->) au sein de l'analyse DTT491.....	137
Tableau 14 : Position des <i>MLE</i> de diatomées marines dans la superfamille <i>Tc1-mariner</i> au sein de la classification DTT491. ....	141
Tableau 15 : Récapitulatif des conditions thermiques auxquelles des transcrits correspondant aux <i>MLE</i> ou aux rétrotransposons <i>Blackbeard</i> et <i>Surcouf</i> ont été obtenus chez <i>Phaeodactylum tricornutum</i> , <i>Amphora coffeaeformis</i> et <i>Amphora acutiuscula</i> . ....	166

# LISTE DES ABREVIATIONS

ADN : acide désoxyribonucléique  
ADNc : acide désoxyribonucléique complémentaire  
ADNg : acide désoxyribonucléique génomique  
AP : activité photosynthétique  
ARNm : acide ribonucléotique messenger  
BET : bromure d'éthidium  
CCAP : culture collection of algae and protozoa  
CCMP : Provasoli-Guillard national center for culture of marine phytoplankton  
Chl *a* : chlorophylle *a*  
Chl *c* : chlorophylle *c*  
CTAB : cetyl trimethylammonium bromide  
DEPC : diéthyl pyrocarbonate  
DO : densité optique  
DTT : DNA transposon, TIR group, *Tc1-mariner*  
EMA : eau de mer artificielle  
ERO : espèces réactives de l'oxygène  
ET : élément transposable  
F<sub>0</sub> : fluorescence minimale  
F<sub>m</sub> : fluorescence maximale  
F<sub>m</sub>' : fluorescence maximale à la lumière  
F<sub>v</sub> : fluorescence variable  
F<sub>v</sub>/F<sub>m</sub> : rendement quantique maximal du PSII  
F<sub>s</sub> : Fluorescence de base  
HSP : heat shock proteins  
IB : inference bayésienne  
IP : intensité photosynthétique  
Kpb : kilo paires de bases  
LTR : long terminal repeats  
Mpb : mega paires de bases  
MLE : mariner like element  
MP : maximum de parcimony  
NCC : Nantes culture collection  
NJ : neighbor joining  
ORF : open reading frame  
Pb : paire de bases  
φPSII : rendement quantique effectif du PSII  
PCR : réaction de polymérisation en chaîne  
PSI /PSII : photosystème I / phototsystème II  
Q-PCR : réaction de polymérisation en chaîne quantitative  
RT : reverse transcription  
sHSP : small heat shock proteins  
SOD : superoxyde dismutase  
TIR : terminal inverted repeats  
UPGMA : unweighted pair group method with arithmetic mean  
UPGM-VM : unweighted pair group method with variation metric  
UTR : untranslated region  
UV : ultraviolet

# INTRODUCTION GENERALE



Le phytoplancton marin est constitué de producteurs primaires tels que les algues, les microalgues et les cyanobactéries, qui sont à la base des réseaux trophiques. Parmi les microalgues, la classe des Bacillariophycées, appelées plus communément diatomées, est particulièrement diversifiée (Round, 1990) tant au niveau morphologique qu'au niveau de l'ornementation des frustules (enveloppe siliceuse). Les diatomées sont impliquées dans les cycles biogéochimiques (carbone, azote, silice, etc) et sont responsables de près de 50 % de la production primaire globale (Lopez et al., 2005; Armbrust, 2009). Leur présence dans tous les océans et leurs facultés adaptatives leur confèrent un rôle écologique majeur dans le milieu marin. De plus, certaines microalgues sont exploitées industriellement.

La biologie et la physiologie des diatomées commencent à être bien connues notamment grâce à leur culture généralement facile à mettre en œuvre au laboratoire et à leur capacité à se multiplier végétativement en un temps de doublement court (un à quelques jours). Avec le développement des techniques de biologie moléculaire et de la génomique, la connaissance des diatomées s'est largement accrue (Scala et Bowler, 2001; Falciatore et Bowler, 2002; Saade et Bowler, 2009; Maheswari et al., 2009). Durant cette dernière décennie, le génome de la diatomée *Thalassiosira pseudonana* a été entièrement séquencé en 2004 et un deuxième, celui de *Phaeodactylum tricornutum*, a été récemment séquencé (Armbrust et al., 2004; Bowler et al., 2008). Les informations génétiques qui en découlent ont permis de mieux comprendre l'évolution des organismes photosynthétiques (Archibald, 2009), les facultés d'adaptation des diatomées (Allen et al., 2008; Vardi et al., 2008; Nymark et al., 2009) et ont mis en évidence la présence d'éléments transposables (ET) (Armbrust et al., 2004; Bowler et al., 2008; Maumus et al., 2009). Les ET sont des séquences d'ADN capables de se déplacer dans le génome entraînant notamment des mutations. La découverte de ces éléments dans les années 1940 fut une grande avancée en génétique (McClintock, 1953; Shapiro, 1999; Feschotte et al., 2002). Depuis, ils ont été détectés dans tous les génomes étudiés. Ils forment chez certains organismes une part majeure des génomes, notamment chez des végétaux où ils peuvent constituer plus de la moitié du génome comme c'est le cas de l'orge et du maïs (SanMiguel et al., 1998; Kumar et Bennetzen, 1999). Les ET ont un rôle dans l'évolution des génomes et l'adaptation des organismes (McClintock, 1984; Capy et al., 1997; Kidwell et Lisch, 2001; Tenailon et al., 2010).

Les ET de type *mariner* (*Mariner-Like-Element* ou *MLE*) sont présents chez les animaux et les végétaux (Hartl et al., 1997a; Shao et Tu, 2001; Robertson, 2002; Feschotte et Wessler, 2002; Feschotte et Pritham, 2007). Ils ont aussi été mis en évidence chez quelques

microalgues marines dans le laboratoire Mer, Molécules, Santé à l'université du Mans. Ceci fut le point de départ de l'étude présentée ici.

**Dans le présent travail, nous nous proposons de répondre à trois questions :**

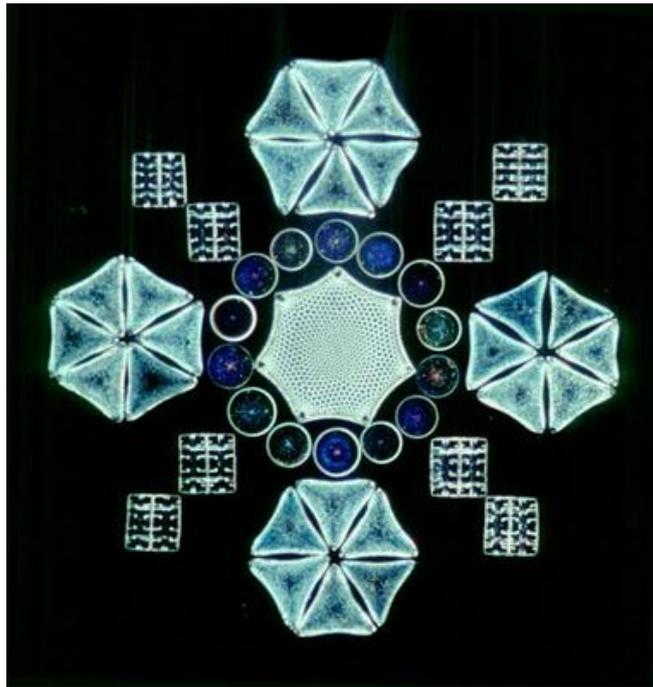
- Les *MLE* sont ils présents dans les génomes de microalgues et plus particulièrement ceux des diatomées marines ?
- S'ils sont présents, quelles sont leur propriétés et comment se classent- ils par rapport aux autres *MLE* de végétaux ?
- Sont-ils exprimés et peuvent-ils être induits par des stress thermiques chez les diatomées?

Pour répondre à ces questions, nous avons tout d'abord recherché, par une approche moléculaire, la présence des *MLE* chez 15 diatomées marines et trois microalgues appartenant à d'autres classes. Nous nous sommes ensuite attachés à comparer les *MLE* de diatomées avec les *MLE* de végétaux connus, par une approche bio-informatique. Enfin, l'expression des *MLE* de trois diatomées a été recherchée après l'application de stress thermiques.

# CHAPITRE I

## INTRODUCTION

### BIBLIOGRAPHIQUE



Les mers et les océans recouvrent environ 70 % de la planète Terre. Le milieu marin est constitué de nombreux écosystèmes (côtiers, pélagiques, benthiques...) dont le fonctionnement suscite encore de nombreuses interrogations. La composition et la structuration des écosystèmes marins dépendent en grande partie de leur biocénose et de l'ensemble des interactions entre les organismes vivants qui les composent. Ces interactions peuvent être largement perturbées par la variation des facteurs du milieu; lorsque ceux-ci sont altérés, il s'ensuit un déséquilibre de tout l'écosystème.

Dans les mers et les océans, le phytoplancton est le premier maillon des chaînes trophiques, il a une importance capitale pour le maintien de l'équilibre du milieu marin. Il est constitué d'un ensemble de microorganismes aquatiques, photosynthétiques, généralement unicellulaires, très diversifié de par la taille, la forme et la structure des organismes qui le composent (Reynolds, 1984; Falciatore et Bowler, 2002). La majeure partie de la fixation du carbone en milieu marin est due à ces microorganismes photosynthétiques (Field et al., 1998) qui composent le phytoplancton marin. Ceux-ci sont responsables de près de la moitié de la production primaire globale de la planète.

Le phytoplancton est composé de bactéries photosynthétiques (prochlorophytes et cyanobactéries) et de microalgues. Les microalgues sont divisées en trois groupes selon leur composition en pigments photosynthétiques : les Chlorophytes de couleur verte possédant de la chlorophylle *a* (Chl *a*) et *b*, les Rhodophytes de couleur rouge contenant de la Chl *a* et des phycobilines, et les Chromophytes de couleur brune, jaune mais aussi verte, bleue et rouge renfermant de la Chl *a* et *c*.

Parmi les Chromophytes, la classe des Bacillariophycées (ou diatomées) représente le groupe le plus diversifié au sein du phytoplancton marin avec plus de 250 genres et 100 000 espèces répertoriées (Round et al., 1990). Avec une répartition mondiale allant des biotopes dits stables aux environnements les plus extrêmes et assurant 40 % de la production primaire marine, elles sont considérées comme le groupe le plus prospère du phytoplancton (Field et al., 1998).

Grâce aux nombreuses particularités qu'elles présentent telles que la présence d'un frustule siliceux, leur répartition, leur diversité, leur sensibilité aux variations du milieu et la production de molécules d'intérêt pharmacologique ou industriel, les diatomées sont devenues des organismes modèles sur le plan écologique, physiologique, biochimique et génétique.

# I Les diatomées

## I.1 Le rôle écologique des diatomées

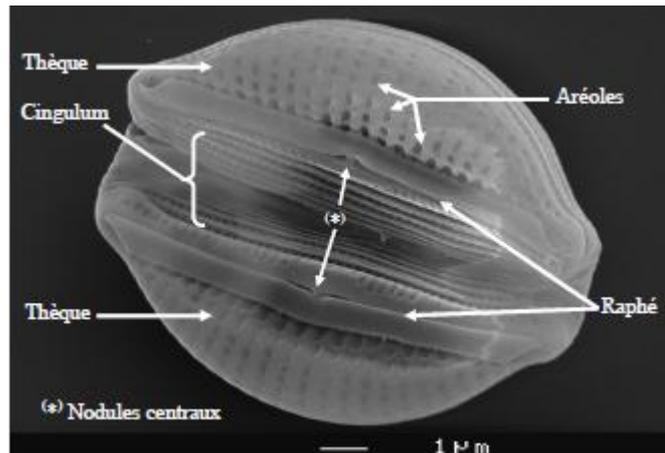
Les diatomées sont responsables de phénomènes de proliférations massives appelées blooms. Les diatomées s'agrègent entre elles et forment un biofilm à la surface de l'eau. Leur prolifération massive est due à l'allongement de la photopériode ainsi qu'à l'augmentation de l'éclairement qui favorise leur prolifération au printemps et, en automne, c'est la grande disponibilité de substances nutritives provenant des fonds qui permet le développement de ces biofilms.

Les diatomées sont impliquées dans les cycles biogéochimiques. Compte tenu de leur capacité à fixer de grandes quantités de CO<sub>2</sub> et de leur répartition mondiale, ces microalgues sont des acteurs clés dans le stockage du CO<sub>2</sub> atmosphérique. De même, leur capacité à produire une enveloppe siliceuse en utilisant le silicium dissous dans les océans les positionne au centre du cycle du silicium (Kemp et al., 2006).

Par ailleurs, la sensibilité des diatomées aux facteurs environnementaux en font également des bio-indicateurs de la qualité de l'eau et de l'état des écosystèmes (Smetacek, 1985; Wu, 1999).

## I.2 La biologie des diatomées

Les Bacillariophycées appartiennent à l'embranchement des Chromophytes (Van den Hoek et Johns, 1995). Ces microalgues présentent une caractéristique structurale remarquable par la présence d'une paroi cellulaire siliceuse appelée frustule. Celui-ci est composé de deux valves s'emboîtant l'une dans l'autre : l'hypothèque et l'épithèque, et de bandes ou ceintures (cingulum) qui lient les deux valves à la périphérie (Figure 1). Chaque valve est ornée de pores ou aréoles formant des motifs nanométriques dont l'organisation est utilisée comme critère de détermination des espèces. Les diatomées sont divisées en deux groupes selon la symétrie de leur frustule. On distingue les diatomées à symétrie radiale ou diatomées centriques et les diatomées à symétrie bilatérale ou diatomées pennées (Van den Hoek et Johns, 1995).

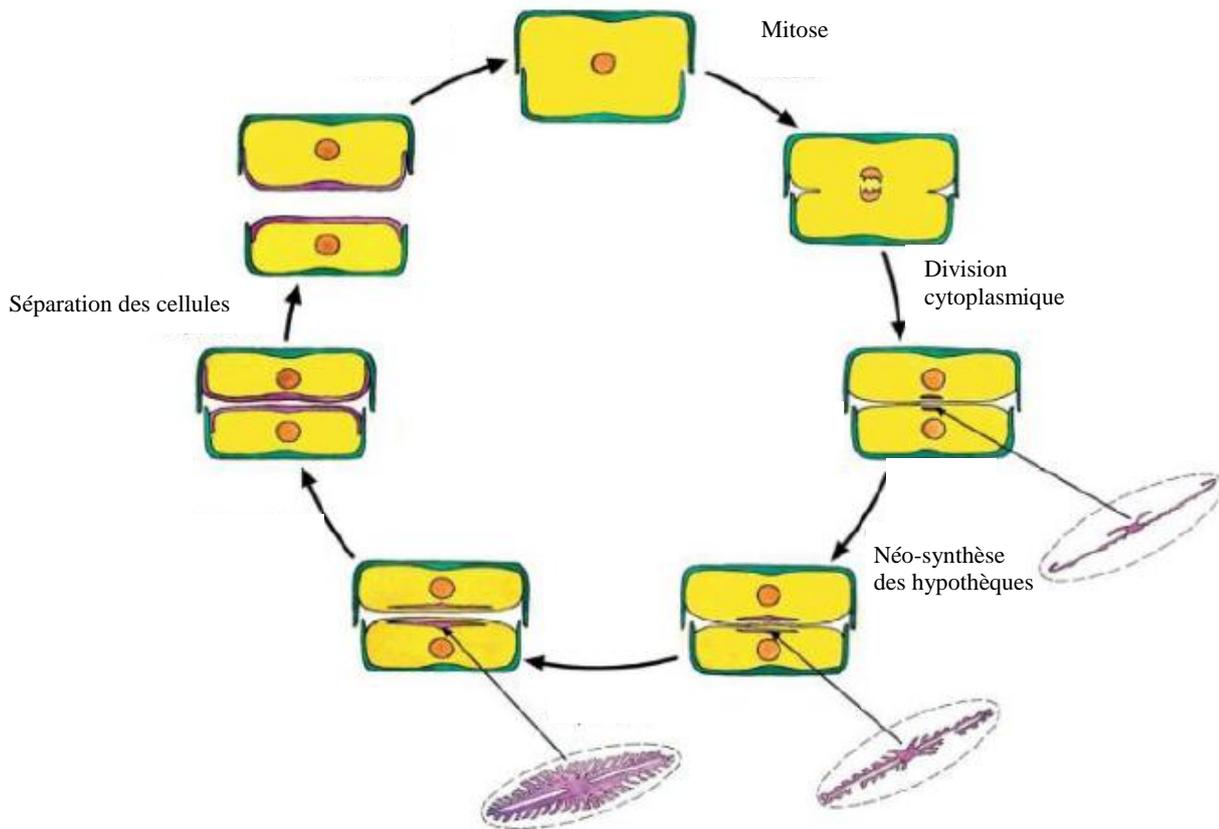


**Figure 1 : Frustule d'une diatomée pennée.**

Détail du frustule d'*Amphora acutiuscula* en vue cingulaire et en microscopie électronique à balayage.

Les diatomées se multiplient essentiellement par voie végétative, c'est-à-dire par mitoses successives, ou par reproduction sexuée. L'accroissement de la biomasse est réalisé par multiplication végétative des cellules, ce qui produit des populations monoclonales, tous les individus d'une population possèdent alors le même génotype. Toutefois, ce type de multiplication a pour inconvénient d'entraîner la diminution de la taille des cellules des diatomées (Drebes, 1977). En effet, chaque frustule est composé d'une valve légèrement plus petite (hypothèque) que l'autre. Lors de la bipartition, chacune des deux cellules filles garde une valve et synthétise l'hypothèque correspondante. La valve néoformée est donc légèrement plus petite que la valve matrice (Figure 2). Ainsi à chaque génération, il y a diminution de la taille des cellules.

Lorsque la taille des cellules devient critique, la reproduction sexuée permet la formation de zygotes de grande taille qui se multiplient ensuite par mitoses. Toutefois, la reproduction sexuée des diatomées n'est décrite que chez quelques espèces (Armbrust et al., 2004; Chepurnov et al., 2004, Mouget et al., 2009) car elle est difficile à observer et à caractériser.



**Figure 2 : Multiplication clonale d'une diatomée par mitose.**

Lors de la mitose, le contenu cellulaire est partagé lorsque le cytoplasme se divise. Chaque cellule fille conserve une valve de la cellule mère et synthétise une nouvelle valve, l'hypothèque (*en rose*). Lorsque les hypothèques respectives sont entièrement formées, les cellules arborant leur propre frustule se séparent. Une des deux cellules filles est de taille légèrement inférieure à la cellule mère (d'après Falciatore et Bowler, 2002).

### I.3 La photosynthèse des diatomées

La photosynthèse peut être définie simplement comme l'ensemble des réactions qui utilisent l'énergie lumineuse pour assurer la synthèse d'oses à partir de dioxyde de carbone et d'eau. La photosynthèse se déroule dans les chloroplastes et fait intervenir deux séries de réactions : les réactions photochimiques correspondant à la phase claire durant laquelle la scission de molécules d'eau libère du dioxygène ( $O_2$ ) et le cycle de Calvin correspondant à la phase sombre durant laquelle les premiers oses sont synthétisés.

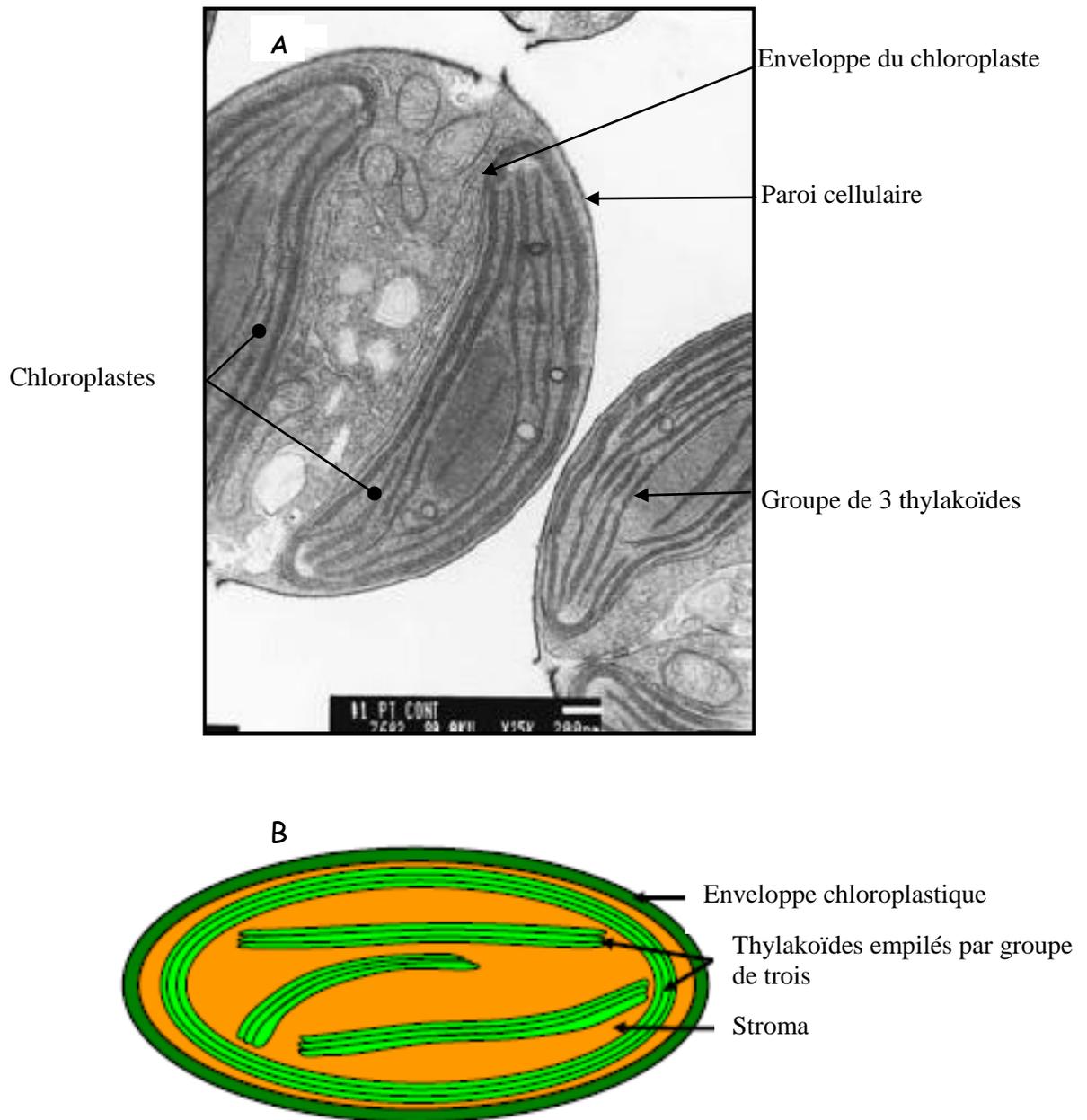
#### I.3.1 L'appareil photosynthétique

Chez les diatomées, les chloroplastes renferment de la Chl *a* et *c* et des pigments caroténoïdes ( $\beta$ -carotène, fucoxanthine, diadinoxanthine) leur donnant une couleur brune. Ces pigments sont localisés dans des membranes formant des replis délimitant des sacs clos

appelés thylakoïdes, disposés longitudinalement et groupés par trois à l'intérieur des chloroplastes (Figures 3A et B).

La membrane thylakoïdale renferme cinq complexes macromoléculaires dont la disposition hiérarchisée assure le bon déroulement des réactions (Figure 4) : les antennes collectrices de l'énergie lumineuse, deux photosystèmes (PS) nommés PSI et PSII, un complexe cytochrome b6/f et de l'ATP-synthase. Les antennes collectrices comprenant les pigments chlorophylliens et caroténoïdes captent l'énergie lumineuse et la transfèrent vers les centres réactionnels (photosystèmes) qui ont pour rôle la transformation de l'énergie lumineuse en énergie chimique (NADPH /H<sup>+</sup> et ATP). Cette énergie permettra ensuite la fixation du carbone inorganique.

Un photosystème est constitué d'un centre réactionnel composé de Chl *a*, de donneurs et d'accepteurs d'électrons entourés d'un ensemble complexe de pigments collecteurs formant l'antenne collectrice. La fonction de l'antenne collectrice est de capturer l'énergie des photons puis de réaliser le transfert de l'excitation au centre réactionnel (Papageorgiou et Govindjee, 2004; Farineau et Morot-Gaudry, 2006).



**Figure 3 : A-Coupe dans des cellules de *Phaeodactylum tricornutum* renfermant deux chloroplastes observés en microscopie électronique (d'après Falkowski et Raven, 1997). B-Schéma de la structure d'un chloroplaste de diatomée avec des thylakoïdes empilés par trois.**

### I.3.2 La photosynthèse

Lors de la phase claire, l'absorption d'un photon par une molécule de pigment fait passer celle-ci d'un état fondamental à un état excité instable correspondant à la délocalisation d'un électron sur une orbitale plus périphérique. Le retour à l'état fondamental est rapide et spontané. Il peut s'effectuer selon plusieurs voies : rapidement ( $10^{-12}$  à  $10^{-9}$  s) par transmission de l'excitation à une molécule voisine (cette voie est utilisée au sein de l'antenne collectrice), émission d'un photon de grande longueur d'onde (fluorescence) ou dispersion calorifique; plus lentement (supérieur à  $10^{-9}$  s) en permettant dans certaines conditions son éjection de la chlorophylle. Il y a alors séparation de charges (photochimie) mais seules les molécules de Chl *a* dans les centres réactionnels ont cette capacité. Les électrons transitent ensuite par différents transporteurs d'électrons du PSII au PSI (Figure 4) pour aboutir à la formation de NADPH /H<sup>+</sup>. La photolyse (ou dissociation) de l'eau permet de combler le trou électronique au niveau du PSII (avec dégagement d'O<sub>2</sub>) et la synthèse d'ATP réduit le gradient de protons entre la lumière du thylakoïde et le stroma.

Les produits de la phase claire, ATP et NADPH, sont ensuite utilisés pour fixer le CO<sub>2</sub> dissous (cycle de Calvin) dans le stroma. Cette réaction, catalysée par la ribulose bi-phosphate carboxylase-oxygénase (RuBisCO), permet la synthèse des premiers glucides à trois atomes de carbone (voie C3) (acide phosphoglycérique). Cette enzyme a également une fonction oxygénase correspondant à la photorespiration.

Plusieurs travaux ont montré que la fixation de carbone pouvait être en partie assurée par une voie alternative, la voie C4-like catalysée par la PEPcase (Phosphoenolpyruvate Carboxylase) dans le cytosol (Reinfelder, 2000; Riebesell, 2000, Rech et al., 2008).

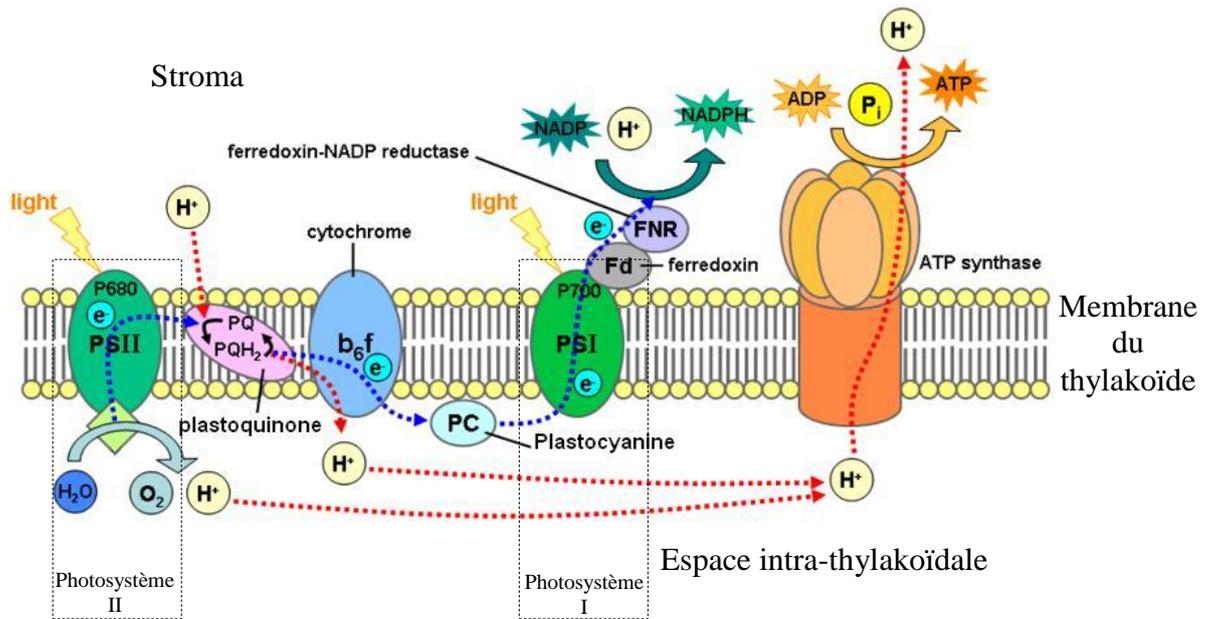


Figure 4 : Schéma de l'organisation de la membrane d'un thylakoïde.

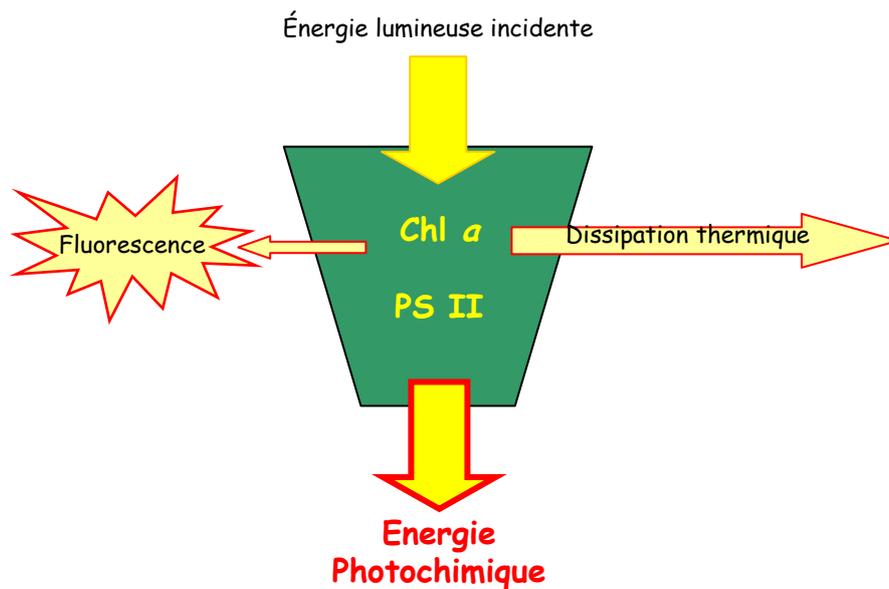


Figure 5 : Schéma des voies de dissipation de l'énergie lumineuse au niveau du photosystème II.

Lorsque l'énergie ne peut pas être transformée par la voie photochimique (transmission d'un électron au Photosystème I), elle peut être dissipée selon deux modes : la chaleur ou la fluorescence. Chl a : Chlorophylle a, PSII : Photosystème II.

### I.3.3 La fluorescence chlorophyllienne

La fluorescence correspond à l'émission d'énergie lumineuse par une molécule lorsqu'elle passe d'un état excité à un état fondamental. La Chl *a* a la propriété d'émettre de la fluorescence lorsqu'elle est éclairée. La fluorescence chlorophyllienne provient de l'excitation de la Chl *a* du PSII lors de la photosynthèse.

Les molécules de Chl *a* retournent ensuite à un état « désexcité ». Ce retour à l'état fondamental s'effectue selon trois voies complémentaires (Figure 5) :

- la photochimie : c'est la conversion de 90 à 95 % de l'énergie lumineuse en pouvoir réducteur. Il en résulte un flux d'électrons accompagné de la production d'ATP.
- la dissipation thermique : l'énergie non convertie par les processus photochimiques représentant environ 5 à 10 % de l'énergie lumineuse captée et dissipée sous forme de chaleur.

La fluorescence : entre 0.5 et 3 % de l'énergie lumineuse collectée est réémise sous forme quantifiable (Maxwell et Johnson, 2000).

Ces voies co-existent de manière compétitive. Lorsque l'énergie allouée à un processus augmente, l'énergie allouée aux deux autres processus diminue. Ainsi la mesure du taux de fluorescence de la Chl *a* permet de connaître l'efficacité des réactions photochimiques et d'estimer la dissipation de l'énergie sous forme de chaleur (Maxwell et Johnson, 2000).

La mesure des paramètres de fluorescence permet d'estimer l'efficacité photosynthétique des cellules et de renseigner sur l'état physiologique de l'appareil photosynthétique de l'organisme considéré (Hancke et al., 2008; Nymark et al., 2009; Guihéneuf et al., 2010). Les variations des paramètres de fluorescence reflètent plus spécifiquement les perturbations au niveau du photosystème II et les modifications de la photosynthèse.

Le taux de fluorescence varie dans les cellules en fonction de leur environnement. En effet de nombreux facteurs abiotiques ont un impact sur les fonctions physiologiques des organismes photosynthétiques.

## I.4 Les stress

### I.4.1 Définition d'un stress

De nombreuses définitions d'un stress ont été proposées (Hoffman et Parson, 1994; Bijlsma et Loeschke, 1997). Forbes et Callow (1997) considèrent comme stress tout facteur qui affecte le bon fonctionnement des organismes. L'organisme est considéré comme tolérant

lorsque la réponse physiologique qu'il développe lui permet de survivre, dans le cas contraire il est considéré comme sensible.

Les végétaux sont soumis à des stress abiotiques de nature physico-chimique et des stress biotiques de nature biologique dus à des organismes vivants (herbivores, bactéries, virus...). Chez les végétaux, l'impact de stress abiotiques ou environnementaux tels que les variations de la température, la salinité... a été étudié par de nombreux auteurs (Boyer, 1982; Holmberg et Bülow, 1998; Smirnoff, 1998; Cushman et Bonhert, 2000; Mahajan et Tujeta, 2005; Rajadurai et al., 2005). Les algues peuvent être soumises à deux types de stress (Davison et Pearson, 1996) : les stress limitants, dus à une réduction de la disponibilité d'une ressource et les stress perturbants, dus à une condition nuisible.

Généralement, le stress est perçu par les récepteurs présents à la surface de la membrane plasmique des cellules végétales. Le signal est ensuite transmis jusqu'au noyau via une cascade de réactions incluant des messagers secondaires tels que le calcium, le phosphate inositol et les espèces réactives de l'oxygène ou ERO (Mahajan et Tujeta, 2005).

#### 1.4.2 La réponse au stress

La perception du stress et la mise en place de la réponse s'effectuent suivant trois phases selon le stress et sa durée (Lichtenthaler, 1998; Selye, 1976).

- La phase de réponse correspond à la phase d'alarme durant laquelle une perturbation du fonctionnement de l'organisme par rapport aux conditions standards est observée. A ceci s'ajoute une augmentation des activités cataboliques, c'est la tolérance. Cette réponse consiste à utiliser ou développer des mécanismes permettant de compenser le préjudice. Ceci s'effectue par l'augmentation de la production des molécules de réponse au stress qui est accompagnée d'une diminution de l'activité métabolique (Bowler et al., 1992; Cushman et Bohnert, 2000; Mittler, 2002; Sun et al., 2002).

- La phase de résistance se met en place, lorsque la durée du stress augmente. Durant cette phase, des mécanismes adaptatifs et des processus de réparation sont mis en jeu. Un des mécanismes adaptatifs est l'évitement. Il consiste en la réduction de l'activité métabolique globale au minimum, c'est le cas de la dormance (végétaux) ou hibernation (animaux). Lorsque le stress est répété, cette phase peut induire une désensibilisation de l'organisme face au stress (Bowler et al., 1992, Holmberg et Bülow, 1998; Cushman et Bohnert, 2000; Mittler 2002; Mirón et al., 2003; Baniwal et al., 2004).

- Enfin la phase d'épuisement, lorsque le stress est trop long et/ou intense et qu'il entraîne des symptômes chroniques menant à la nécrose et la mort de l'organisme.

La réponse au stress met en jeu la régulation des gènes de réponse au stress. Ils peuvent être classés en deux groupes :

- les gènes à réponse précoce : dont l'induction suit très rapidement la perception du stress. Leur activation est également de courte durée.

- les gènes à réponse tardive : ils sont induits après un temps de latence, généralement après les gènes exprimés précocement.

Certains gènes impliqués dans la réponse aux stress abiotiques sont considérés comme des gènes marqueurs tels que les ceux codant les protéines chaperonnes ou HSP et les gènes de réponse au stress oxydant (Lin et al., 1984; Krishnan et al., 1989, Scandalios, 1993; Foyer et al., 1994; Cid et al., 1995; Holmberg et Bülow, 1998; Smirnoff, 1998; Cushman et Bohnert, 2000; Mittler, 2002; Baniwal et al., 2004, Mahajan et Tujeta, 2005; Perelman et al., 2006).

#### I.4.3 Effet du stress thermique sur le métabolisme des diatomées

La répartition mondiale des diatomées, incluant les environnements extrêmes depuis les mers tropicales jusqu'aux eaux polaires, dénote une capacité d'acclimatation et d'adaptation aux variations de température (Margalef, 1978; Blinn et al., 1989; Banzet et al., 1998; Falciatore et Bowler, 2002; Mock et Valentin, 2004; Morgan-Kiss et al., 2006).

La température est un facteur qui joue un rôle important sur le fonctionnement biologique des diatomées de par son impact sur l'activité enzymatique et la conformation des protéines (Raimbault, 1986; Geider, 1987; Davison, 1991; Suzuki et Takahashi, 1995; Kudo et al., 2000; Morgan-Kiss et al., 2006; Hancke et al., 2008), mais aussi sur la fluidité membranaire. L'arrangement structurel et la composition des lipides membranaires confèrent aux membranes leur fluidité qui augmente avec la température (Groth-Nard et Robert, 1993; Jiang et Gao, 2004). L'élévation de la température peut ainsi induire une augmentation de la perméabilité des membranes thylakoïdiennes pouvant modifier le flux d'électrons au cours de la photosynthèse (Rai et Gaur, 2001). De manière générale, le réchauffement du milieu a un effet positif sur la croissance des diatomées tant que la température imposée ne dépasse pas le seuil de tolérance (Goldmann et Carpenters, 1974). L'étude de la croissance des diatomées à différentes températures montre que généralement la température optimale la croissance maximale est proche de la température seuil au-delà de laquelle l'espèce ne peut survivre (William et Morris, 1982; Suzuki et Takahashi, 1995; Rajadurai et al., 2005). Toutefois, celle-ci reste variable selon les diatomées, leur environnement et leur capacité d'acclimatation (William et Morris, 1982; Mirón et al., 2003; Mock et Valentin, 2004; Rajadurai et al., 2005). Ceci dépend de la tolérance aux variations de température de l'organisme en question (Atkin

et Tjoelker, 2003). Lorsque l'activité photosynthétique est presque nulle et que l'activité métabolique est réduite au minimum, du fait des dommages causés aux enzymes et notamment celles du photosystème II, la diatomée est définie comme thermo-intolérante ou sensible (Davison, 1991; Lefebvre et al., 2007).

La croissance des diatomées, au même titre que les autres organismes photosynthétiques, est corrélée à l'intensité photosynthétique. Elle augmente avec la température jusqu'à une température optimale (Goldman et Mann, 1980; Geider, 1987; Hancke et al., 2008). Au delà de cette température, l'activité photosynthétique décroît brutalement.

La famille des HSP (Heat Shock Protein) ou protéines de choc thermique, est une famille de protéines dont la synthèse est accrue en réponse aux stress. Leur rôle est de prévenir les dommages liés à la dénaturation des protéines. Toutefois, elles interviennent aussi dans la transduction du signal notamment lors de stress dus à des hautes températures (Lindquist et Craig, 1988), mais aussi lors d'autres stress tels que le gel, la sécheresse, le stress oxydant...(Van Berkel et al., 1994; Banzet et al., 1998; Cheong et al., 2002). Ces protéines sont classées en cinq groupes selon leurs masses moléculaires : les HSP 100, 90, 70, 60, et les HSP de faible masse moléculaire ou small HSP (sHSP) (Vierling, 1991). Les sHSP sont des molécules chaperonnes qui se fixent aux protéines dénaturées ou dont la structure est altérée. Elles sont présentes dans les génomes des archéobactéries, des bactéries et des eucaryotes. Les HSP ont surtout été décrites chez les végétaux au sein desquels elles sont particulièrement diversifiées (Waters et Vierling, 1999; Ouyang et al., 2009) et présentes dans tous les compartiments cellulaires (Kotak et al., 2007). Leur rôle potentiel dans la thermo-tolérance a été démontré chez les plantes et les cyanobactéries (Nakamoto et al., 2000; Sun et al., 2002) où leur principale activité est la protection contre la dénaturation irréversible des protéines. La présence des sHSP a été mise en évidence récemment au sein des génomes de cinq espèces d'algues parmi lesquelles la diatomée modèle *Thalassiosira pseudonana* (Waters et Rioflorido, 2007). L'expression des gènes codant les sHSP est induite lors du stress oxydant qui accompagne généralement le stress thermique (Banzet et al., 1998; Sun et al., 2002).

#### 1.4.4 Le stress oxydant

Bien que l'oxygène soit une molécule nécessaire à la vie, il existe sous des formes toxiques appelées espèces réactives de l'oxygène (ERO) ou radicaux libres. En conditions

normales des ERO sont produites en quantités minimales dans les cellules mais leur production est amplifiée par de nombreux stress abiotiques (Mittler, 2002). Les ERO sont :

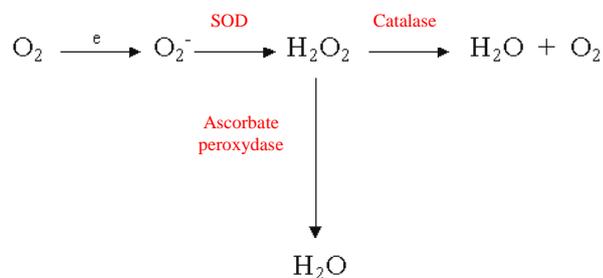
- l'anion superoxyde  $O_2^-$ , très dommageable notamment au niveau des lipides membranaires, et qui est dismuté en peroxyde d'hydrogène par la superoxyde dismutase (SOD).
- le peroxyde d'hydrogène,  $H_2O_2$ , toxique, il a la capacité de diffuser au travers des membranes.
- l'hydroxyde,  $HO\cdot$ , qui est l'espèce la plus toxique des ERO par sa capacité à oxyder tous types de molécules.

Les ERO sont des produits dérivés du métabolisme aérobie notamment de la photosynthèse et de la respiration. La photosynthèse est la source majeure de production des ERO dans le chloroplaste (Foyer et al., 1994; Nigoyi, 1999) au niveau des centres réactionnels des photosystèmes I et II. La photorespiration est également source d'ERO. Ces molécules sont également produites en signal d'alarme en quantité n'entraînant pas de dommages irrémediables. Selon Mittler et al. (2002) leur régulation est contrôlée par:

- des mécanismes modulant finement la quantité d'ERO produite pour l'effet alarme.
- les mécanismes de détoxication qui permettent d'éliminer rapidement ces molécules de la cellule lorsqu'elles sont présentes en quantité dommageable.

Ces mécanismes consistent à empêcher tout contact entre l'anion superoxyde et le peroxyde d'hydrogène (Bowler et al., 1992).

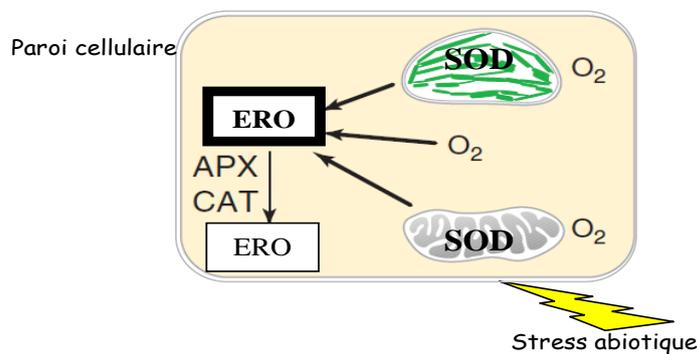
Ces derniers font intervenir la production d'enzymes antioxydantes qui ont pour rôle de transformer les ERO (Mittler, 2002; Zelko et al., 2002; Scandalios, 2005; Wolfe-Simon et al., 2005). Trois d'entre-elles sont communément étudiées chez les plantes mais aussi chez les diatomées (Bowler et al., 1992; Mittler, 2002) : les SOD, la catalase et l'ascorbate peroxydase (Figure 6).



**Figure 6 : Réactions catalysées par les trois enzymes antioxydantes : superoxyde dismutase (SOD), catalase et ascorbate peroxydase.**

Il existe quatre types de SOD qui sont différenciées selon leur cofacteur métallique : Cu/Zn SOD, Fe SOD, Ni SOD et Mn SOD. Cette dernière se retrouve chez les Rhodophytes et les diatomées (Wolfe-Simon et al., 2005) .

Les SOD sont les premières à agir chez les organismes photosynthétiques (Bowler et al., 1992; Scandalios, 1993) en permettant la dismutation de l'anion superoxyde en peroxyde d'hydrogène (Scandalios, 2005). Ce dernier est ensuite réduit en eau grâce à l'ascorbate peroxydase. Lorsque les SOD sont exprimées avec un niveau suffisant conjointement dans les mitochondries et dans les chloroplastes (Figure 7), l'organisme présente une meilleure tolérance au stress (Bowler et al., 1992).



**Figure 7 : Cascade de réactions de détoxication des formes réactives de l'oxygène (ERO) produites lors d'un stress abiotique (d'après Mittler, 2002).**

Les stress abiotiques induisent l'accumulation d'ERO produites dans les chloroplastes (en vert) et la mitochondrie (en gris). En réponse, des enzymes de détoxications telles que les SOD, la peroxydase et la catalase sont synthétisées. L'action de ces enzymes va permettre de diminuer la quantité importante d'ERO (en trait gras).

### I.5 Valorisation des diatomées

La diversité des diatomées (taxonomique et biochimique), leur position à la base des chaînes trophiques, leur taille et leur durée de génération indiquent que ces organismes peuvent être intéressants pour développer de nouvelles applications scientifiques et économiques. Ces applications en biotechnologie dans les domaines tels que l'aquaculture, l'alimentation ou les énergies renouvelables ont également permis d'étendre de nombreux axes de recherche. Les axes principaux tels que la biologie moléculaire, la chimie fine pour la production de composés bioactifs et le génie des procédés de production pour l'obtention de super-producteurs des molécules d'intérêt ou encore les nanotechnologies sont des secteurs en développement.

Les applications en biotechnologie ou bio-industries :

- l'aquaculture : les particularités des diatomées leur confèrent un rôle stratégique dans l'élevage larvaire en aquaculture. A la base des chaînes trophiques aquatiques, elles sont la

composante essentielle de l'alimentation des organismes marins d'élevage : poissons, larves et juvéniles des mollusques tels que les huîtres, les moules... (Muller et Feuga, 1997). Elles sont donc largement utilisées en aquaculture. De plus leur culture en conditions contrôlée est bien maîtrisée. Elles sont utilisées à tous les niveaux de production des écloséries-nurseries (géniteurs, élevage larvaire et post-larvaire), bien que mise au point dans les années 60, le fourrage phytoplanctonique reste la seule technique de production utilisée à ce jour en aquaculture des bivalves.

- l'alimentation humaine : les microalgues sont également produites et commercialisées comme source d'oméga-3 où elles entrent dans la composition des compléments alimentaires. En effet, elles sont capables de produire des acides gras poly-insaturés précurseurs d'oméga-3. Certaines espèces de diatomées peuvent produire des acides gras poly-insaturés en quantité aussi élevée que chez les poissons (Ward et Singh, 2005; Spolaore et al., 2006; Cardozo et al., 2007). L'avantage de l'utilisation des microalgues est de limiter l'épuisement des ressources par la pêche. De plus les diatomées bénéficient d'un temps de génération très court (1 à 2 jours) et leur culture en grands volumes est bien maîtrisée, ce qui représente un avantage considérable sur la pisciculture.

- les énergies renouvelables : les microalgues marines peuvent produire de grandes quantités de lipides pouvant représenter jusqu'à 80 % de leur masse de matière sèche (Christi, 2007). Ainsi depuis peu, des recherches sont menées afin d'utiliser les lipides de microalgues comme bio-carburant en complément de celui produit à partir du soja ou du maïs (Christi, 2007; Cadoret et Bernard, 2008).

- les nanotechnologies : les diatomées sont également un modèle d'étude pour les secteurs Recherche et Développement en nanotechnologie et en sciences de l'ingénieur. En effet, les chercheurs tentent de mettre en évidence les mécanismes de formation des frustules siliceux. Ceci permettrait de développer des applications dans l'industrie des semi-conducteurs, des nanopuces ou dans l'industrie pharmaceutique où les enveloppes siliceuses pourraient être utilisées comme capsules pour la diffusion de médicaments dans le sang.

Dans le domaine marin, les substances nouvelles ayant une activité potentiellement intéressante seraient 100 fois plus nombreuses que celles des organismes terrestres. Ainsi les axes de Recherche en développement sont :

- les technologies en biologie moléculaire, chimie fine et en génie des procédés : elles sont développées sur les diatomées afin d'identifier des souches d'intérêt, d'extraire les molécules ciblées telles que pigments protéiques ou caroténoïdes aux propriétés antioxydantes, enzymes

et autres substances bio-actives qui peuvent avoir des applications en santé et en cosmétique. Cela inclut également la manipulation génétique des souches sélectionnées pour accroître leur production de molécules et la mise en œuvre de procédés de conservation ou conditionnement favorables à la production des molécules ciblées.

## I.6 Génétique et génomique des diatomées

Le rôle clé du phytoplancton dans les écosystèmes ainsi que son intérêt alimentaire, nutritionnel, et technologique mettent en relief l'importance de comprendre les mécanismes évolutifs et génétiques des microalgues.

Pour comprendre les mécanismes sous-jacents aux facultés d'adaptation des microalgues et plus particulièrement des diatomées, il est nécessaire d'acquérir de l'information génétique. Actuellement, cela est possible assez facilement depuis le séquençage du premier génome complet de la plante modèle *Arabidopsis thaliana* en 2000. Ainsi les génomes chloroplastiques d'algues vertes (Chlorophytes), rouges (Rhodophytes), et des diatomées ont été séquencés (Oudot-Le Secq et al., 2007) ainsi que des génomes complets de Chlorophytes (Derelle et al., 2006 ; Merchant et al., 2007).

Depuis peu, deux génomes de diatomées sont entièrement séquencés et publiés : un génome de diatomée centrique *Thalassiosira pseudonana* et un génome de diatomée pennée *Phaeodactylum tricorutum*. Ces génomes sont de petite taille avec 32,4 Mpb pour *T. pseudonana* et 27,4 Mpb pour *P. tricorutum* (Armbrust et al., 2004, Bowler et al., 2008). Ces deux génomes sont composés de 10 402 et 11776 gènes respectivement pour *P. tricorutum* et *T. pseudonana*. L'analyse génomique de ces deux génomes a permis de découvrir :

- les gènes intervenant dans le cycle complet de l'urée, ce qui n'avait été décrit chez aucun eucaryote auparavant (Allen et al., 2006)
- des transferts génétiques d'origine endosymbiotique. Chez *T. pseudonana* quatre gènes codant six protéines sont communs avec ceux des plastes appartenant à des algues rouges (Rhodophytes). Ceci semble être le résultat de transferts génétiques entre génome nucléaire et génome chloroplastique chez les algues rouges. Les gènes communs entre Rhodophytes et diatomées proviendraient alors de l'endosymbiose secondaire à l'origine des diatomées (Armbrust et al., 2004). En effet, les diatomées dérivent d'un ancêtre ayant intégré secondairement le génome nucléaire d'une algue rouge (Rhodophyte).
- des transferts horizontaux entre bactéries et diatomées (Bowler et al., 2008). Sur les 10 402 gènes codant potentiellement des protéines dans le génome de *P. tricorutum*, 587 gènes apparaissent phylogénétiquement plus proches des gènes bactériens que des gènes des autres

algues (Bowler et al., 2008). De même chez *T. pseudonana*, 320 sur 11776 gènes pourraient provenir de transferts horizontaux depuis des génomes bactériens (Armbrust et al., 2004 ; Bowler et al., 2008).

Les génomes des deux diatomées comprennent également une part faible d'éléments mobiles aussi appelés éléments transposables. *P. tricornutum* contient une proportion de 6,4 % d'éléments transposables (ET) dans son génome, soit trois fois plus que *T. pseudonana* qui ne possède que 1,9 % d'ET. Ces ET appartiennent à différentes familles mais l'une d'entre elles est prépondérante, c'est la famille d'éléments nommée *Ty1-copia* qui représente 90 % des ET de *P. tricornutum* (soit 5,8 % du génome) et 58 % de *T. pseudonana*, (soit ~ 1 % du génome) (Armbrust et al., 2004; Bowler et al., 2008; Maumus et al., 2009). Au sein de cette famille d'ET, certains sont exprimés et régulés de manière différente en fonction des conditions de cultures des diatomées. Les auteurs suggèrent donc que l'expression des ET pourrait avoir un impact sur les capacités d'adaptation des diatomées (Maumus et al., 2009).

Associée à cette quête d'informations génétiques par le biais du séquençage, des banques de données de séquences exprimées (EST pour Expressed Sequences Tag) ont été développées à partir de cellules de diatomées cultivées dans 16 conditions différentes (Maheswari et al., 2005, 2009). L'ensemble des EST est disponible dans la Diatom EST database ([http : //avesthagen.sznbowler.com/](http://avesthagen.sznbowler.com/)).

Pour faire suite à ces découvertes, deux génomes de diatomées sont actuellement en cours de séquençage, une diatomée pennée psychrophile (adaptée aux climats froids) *Fragilariopsis cylindrus* et une diatomée produisant des peptides toxiques, *Pseudonitzschia multiseries*. L'analyse de ces génomes devrait permettre d'identifier les gènes responsables de l'adaptation dans un environnement extrême pour la première et de la production de toxines pour la seconde, mais aussi de préciser l'évolution de ces génomes à l'origine de la diversité des diatomées.

## II Les Éléments Transposables

Le génome a longtemps été considéré comme une succession de gènes, chacun associé à un locus invariant sur le chromosome. Cette vision statique a été bouleversée dans les années 1940, par Barbara Mc Clintock lorsqu'elle observa des phénotypes mutants instables chez le maïs (coloration des grains de l'épi) induits par des éléments génétiques appelés alors « éléments de contrôles ». Plus tard, McClintock démontra que ces mutations réversibles entraînent l'activation ou l'inactivation des gènes responsables de la coloration des grains de l'épi de maïs (McClintock, 1953). Elle émit, des lors, l'hypothèse que ces mutations pouvaient être induites en réponse à des chocs / événements affectant le génome (McClintock, 1984). Cette découverte a révolutionné la notion de génome qui de statique est devenu dynamique. Longtemps considérés comme de l'ADN égoïste car ils utilisent la machinerie de l'hôte pour leur prolifération (Doolittle et Sapienza, 1980), ces « éléments de contrôle » sont aujourd'hui appelés éléments transposables (ET) et sont reconnus comme une composante quantitative des génomes (Capy et al., 1997; Aziz et al., 2010). De nombreuses études ont démontré leurs implications dans l'évolution et notamment les processus adaptatifs des génomes (Shapiro, 1999; McDonald, 1995; Capy et al., 2000; Kidwell et Lisch, 2001; Schmidt et Anderson, 2006; Böhne et al., 2008).

### II.1 Définition

Les ET sont des courtes séquences d'ADN ayant la capacité de se déplacer dans un chromosome ou même d'un chromosome à l'autre, au sein du génome qui les héberge. Ils possèdent dans leur séquence le ou les gènes codant les enzymes nécessaires à leur déplacement encore appelé transposition. On distingue les ET autonomes qui sont mobiles de leur seul fait car ils possèdent des gènes codant des protéines fonctionnelles de transposition et les ET non-autonomes qui contiennent des gènes de transposition mutés ou délétés. Toutefois, ces-derniers peuvent être mobilisés par des éléments autonomes, ce processus est appelé mobilisation en *trans*.

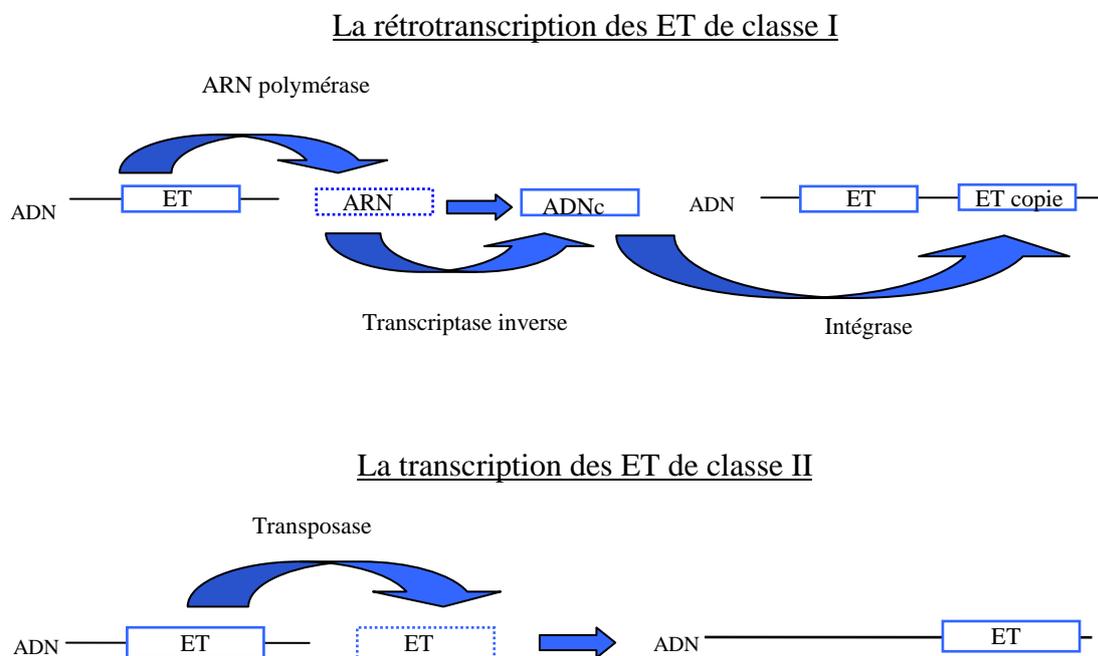
### II.2 Classification des ET

Les ET présentent une grande diversité de structures et de mécanismes de transposition, ce qui rend leur classification complexe. Plusieurs classifications se sont succédées, la première reposant sur le mécanisme de transposition a été proposée par Finnegan (1989). Ceci a donné lieu à la répartition des ET en deux grandes classes selon que

la transposition nécessitait un intermédiaire ARN (classe I ou rétrotransposon) ou non (classe II ou transposon).

Les ET de classe I ou rétrotransposons regroupent tous les éléments qui transposent selon un mécanisme qualifié de « copier-coller » (Figure 8). La transposition s'effectue au moyen de la production d'un ARN rétrotranscrit en ADN complémentaire (ADNc). L'ADNc correspond donc à une copie qui sera ultérieurement intégrée dans un locus accepteur du génome.

Les éléments de classe II ou transposons transposent directement, sans intermédiaire, selon un mécanisme qualifié de « couper-coller » (Figure 8). Ainsi l'élément présent au sein du site donneur est excisé et réinséré dans le génome au niveau d'un site accepteur.



**Figure 8: Schéma de la classification des ET selon les modes de transposition.**

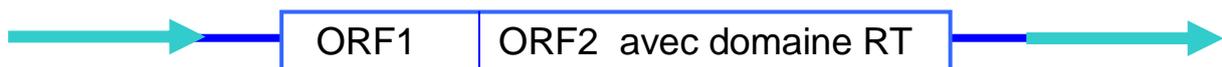
Devant la quantité de nouveaux ET découverts chaque année et la diversité de ces éléments, des classifications plus précises et permettant de classer tous les éléments découverts, ont été proposées. Ces dernières prennent en compte la structure des ET, certaines fonctions ou des spécificités d'insertion. La classification proposée par Finnegan (1989) a été affinée par Capy et al. (1997), en intégrant les niveaux subalternes, tel que l'ordre, la superfamille et la famille, aux deux classes initiales, selon leur fonction, leur structure

moléculaire et la présence de motifs signatures. Les classifications récentes de Wicker et al., (2007) et Kapitonov et Jurka (2008) proposent de classer les ET en fonction de l'organisation de leurs domaines fonctionnels mais aussi en fonction de leur site d'insertion ou TSD (Target Site Duplication) dans le génome. Ceci permet de classer notamment les ET non-autonomes dont on connaît mal le mécanisme de transposition tel que les MITE (Miniatures Inverse Transposable Elements) et les foldback (Cohen et al., 1985; Marzo et al., 2008) dans les superfamilles d'éléments autonomes de classe I et II. La classification de Wicker et al., (2007), repose sur l'homologie de séquence avec les ET autonomes permettant leur mobilisation en *trans*. La question de la classification n'est pas encore résolue car de nombreux cas restent ambigus : l'abondance d'éléments incomplets tronqués ou fossiles (dont seule une partie très conservée a été identifiée) et l'affluence d'ET *de novo* découverts *in silico* dus à l'émergence de séquençages massifs des génomes, pourraient remettre en cause les bases de ces classifications.

La classification qui sera détaillée ici repose sur la classification de Wicker et al., 2007 qui est limitée aux eucaryotes. Seules les familles les mieux connues seront détaillées et les éléments non-autonomes sont distingués dans des paragraphes séparés pour plus de clarté et de précisions sur leurs mécanismes.

### II.2.1 Les superfamilles de la classe I

(Figure 10, encadré en bleu, p34)



**Figure 9 : Schéma général d'un rétrotransposon.**

Les flèches correspondent à des répétition de même orientation, la séquence centrale est composée de plusieurs cadres de lecture ouverts (ORF), dont un contient un domaine codant la reverse transcriptase (RT).

Selon la classification publiée par Wicker et al. en 2007, les éléments de classe I sont divisés en quatre superfamilles d'éléments autonomes auxquelles est rattachée une superfamille d'éléments non-autonomes (les SINE, décrits en section II.2.3). La grande majorité des ET de classe I possède une structure comportant plusieurs ORF encadrées par des régions répétées de même orientation aux extrémités (Figure 9).

## Les rétrotransposons à LTR

Les rétrotransposons à LTR (Long Terminal Repeat) peuvent représenter une part importante des génomes le plus souvent chez les plantes (Bennetzen, et al., 2005). San Miguel et Bennetzen (1998) indiquent que l'activité des rétrotransposons serait responsable du doublement voire du quadruplement du génome du maïs. Chez le coton la composante principale du génome est constituée d'ET (> 79 %), parmi ceux-ci près de 62 % sont des éléments de classe I à LTR (Charles et al., 2008).

Les rétrotransposons à LTR ont une taille comprise entre 5 et 9 Kpb. Ils possèdent à leurs extrémités de longues séquences identiques, orientées dans le même sens d'où leur nom de Long Terminal Repeat et s'insèrent dans un site cible dont ils dupliquent 6 pb environ. Leur région centrale est composée de gènes codant plusieurs protéines. Le gène *gag* pour (group specific antigen) code la formation d'une capsid virale, le gène *pol* pour polyprotéine code les protéines nécessaires à la transposition et à l'insertion de la copie de l'élément dans le génome. Ces protéines sont la protéase (*Pr*), la reverse transcriptase (*RT*), et l'intégrase ou RnaseH (*Int*) (Figure 10). Ainsi, l'organisation de ces domaines définit cinq superfamilles parmi lesquelles celles des Rétrovirus et les éléments ERV. Ils comportent un domaine supplémentaire ENV, ce gène *env* code la formation d'une enveloppe qui permet l'infection d'autres cellules. L'organisation structurale des LTR rappelle celle des virus (Mc Clure, 1991) à tel point que la question d'une origine phylogénétique commune s'est posée (Lerat et Capy, 1999; Llorens et al., 2008).

Deux superfamilles principales constituent l'ordre des LTR : la superfamille des *Copia* ou *Ty1-copia* possèdent le gène de l'intégrase du côté 5' dans le module codant la transcriptase inverse. Cette superfamille est très représentée dans les génomes de plantes (Flavell et al., 1992; Vitte et Panaud, 2005; Kalendar et al., 2010) et chez les diatomées (Maumus et al., 2009). La deuxième superfamille regroupe les *ty3 et gypsy* éléments, le gène de l'intégrase se trouve du côté 3' du gène de la reverse transcriptase chez ces ET. Ces rétrotransposons sont distribués largement au sein des génomes de plantes, de vertébrés mais aussi de parasites (Bae et al., 2001; Böhne et al., 2008; Kalendar et al., 2010).

Les rétrotransposons non-LTR sont également nommés rétroposons et sont des éléments autonomes pour la grande majorité (les SINE qui sont non-autonomes sont intégrés dans cette catégorie mais nous les détaillerons dans le paragraphe II.2.3, dédié aux éléments non-autonomes).

### **Les DIRS-like éléments**

Leur nom DIRS-1 (Dictyocetelium Intermediate Repet Sequences) provient de l'hôte au sein duquel le premier élément a été découvert, l'amibe, *Dictyostelium discoideum* (Capello et al., 1984). Ils présentent une grande diversité d'hôte mais sont particulièrement présents dans les génomes de décapodes issus de sources hydrothermales (Poulter et Goodwin, 2005; Piednoël et Bonnivard, 2009). Les DIRS ne possèdent pas d'intégrase mais une tyrosine recombinase (YR) et ne génèrent pas de duplication au site d'insertion. Ils avaient été classés initialement avec les membres du groupe des *Ty3-gypsy* (Capy et al., 1997). Ils présentent deux couples de LTR, le premier est orienté en sens inverse du deuxième, la structure est alors, différente de celles des rétrotransposons à LTR. Leurs différences par rapport aux autres rétrotransposons leur ont permis d'être classés dans un groupe à part entière, en restant positionnés au sein des ET de classe I de par la phylogénie de leur domaine RT.

### **Les LINE**

Les ET de la superfamille des LINE (Long Intersperced Nuclear Elements), comprennent les ET les plus connus, les *LI* des mammifères, dont on compte un grand nombre de copies dans le génome humain (Kazazian et Moran, 1998). Ces ET comportent deux domaines ORF correspondant aux gènes *gag* et *pol* et ne présentent pas de LTR. Bien que les éléments de cette superfamille sont très diversifiés, deux types de structures ressortent. Le premier type est celui des R2-like qui présente une seule ORF avec un domaine RT et un domaine endonucléase (EN). Le deuxième type comporte deux ORF qui auraient des fonctions similaires aux gènes *gag* et *pol* bien que l'intégration s'effectue grâce à une endonucléase (APE).

### **Les PLE**

Les PLE ou Penelope-Like-Element ont été découverts récemment chez la drosophile (Evgen'ev et al., 1997; Evgen'ev et Archipova, 2005). Ces ET constituent un groupe très éparse, c'est pour quoi ils forment un groupe à part entière. Leur structure peut contenir une reverse transcriptase qui ressemble plus à une télomérase qu'aux RT des ET de classe I et les régions terminales peuvent être de même orientation ou de sens inverses.

Classification		Structure	TSD
Order	Superfamily		
<b>Class I (retrotransposons)</b>			
LTR	<i>Copia</i>	→ GAG AP INT RT RH →	4-6
	<i>Gypsy</i>	→ GAG AP RT RH INT →	4-6
	<i>Bel-Pao</i>	→ GAG AP RT RH INT →	4-6
	<i>Retrovirus</i>	→ GAG AP RT RH INT ENV →	4-6
	<i>ERV</i>	→ GAG AP RT RH INT ENV →	4-6
DIRS	<i>DIRS</i>	↔ GAG AP RT RH YR ↔	0
	<i>Ngaro</i>	→ GAG AP RT RH YR →	0
	<i>VIPER</i>	→ GAG AP RT RH YR →	0
PLE	<i>Penelope</i>	← RT EN →	Variable
LINE	<i>R2</i>	RT EN	Variable
	<i>RTE</i>	APE RT	Variable
	<i>Jockey</i>	ORF APE RT	Variable
	<i>L1</i>	ORF APE RT	Variable
	<i>I</i>	ORF APE RT RH	Variable
SINE	<i>tRNA</i>		Variable
	<i>7SL</i>		Variable
	<i>5S</i>		Variable
<b>Class II (DNA transposons) - Subclass 1</b>			
TIR	<i>Tc1-Mariner</i>	↔ Tase* ↔	TA
	<i>hAT</i>	↔ Tase* ↔	8
	<i>Mutator</i>	↔ Tase* ↔	9-11
	<i>Merlin</i>	↔ Tase* ↔	8-9
	<i>Transib</i>	↔ Tase* ↔	5
	<i>P</i>	↔ Tase ↔	8
	<i>PiggyBac</i>	↔ Tase ↔	TTAA
	<i>PIF-Harbinger</i>	↔ Tase* ORF2 ↔	3
	<i>CACTA</i>	↔ Tase ORF2 ↔	2-3
Crypton	<i>Crypton</i>	YR	0
<b>Class II (DNA transposons) - Subclass 2</b>			
Helitron	<i>Helitron</i>	RPA // Y2HEL	0
Maverick	<i>Maverick</i>	C-INT ATP // CYP POLB	6

#### Caractéristiques structurales



**Figure 10 : Classification des éléments transposables (d'après Wicker et al., 2007).**

Les ET eucaryotes de classe I et II sont divisés en sous-classes, ordres et superfamilles.

La taille du fragment dupliqué après insertion ou TSD pour Target Site Duplication est caractéristique de la plupart des superfamilles. LTR : long terminal repeat, TIR : terminal inverse repeat, RT : reverse transcriptase, RH : Ribonucléase H, Int : integrase, Pr : protéase, Env : gène de l'enveloppe, Gag : gène de la capsid, Tase : transposase (\* avec un motif DDE), ORF : cadre de lecture ouverte de fonction inconnue, HEL : hélicase, Y2YR : avec un motif YY, CYP : cystéine protéase, POLB : ADN polymérase B, RPA : protéine A de réplication, ATP : ATPase d'empaquetage, YR : tyrosine recombinase, C-INT : C-intégrase.

## II.2.2 Les superfamilles de la classe II

(Figure 10, encadré rouge)



**Figure 11 : Schéma général d'un transposon.**

Les flèches représentent les régions terminales inversées et la séquence centrale se compose d'un ORF codant la transcriptase.

Selon la classification publiée par Wicker et al. en 2007, les éléments de classe II sont divisés en deux sous-classes selon la spécificité de leur mécanisme de transposition.

La sous-classe 1 est composée des éléments ayant un mécanisme de transposition classique de type « couper-coller ». La sous-classe 2 comprend les éléments Hélitrons au mécanisme de cercle roulant (Kapitonov et Jurka, 2001) et *Maverick* dont le processus de transposition n'est pas connu (Pritham et al., 2007).

### **La Sous-classe 1 : les transposons à TIR**

Ce groupe réunit tous les éléments dont la structure est composée d'une région centrale comportant une ORF codant la transposase. Les extrémités du transposon sont composées par des régions terminales inversées et répétées ou TIR (Terminal Inversed Repeats) (Figure 11). Ces régions sont reconnues par la transposase qui excise le transposon au niveau des TIR et le réintègre dans un site receveur à un autre locus. Les éléments les mieux connus de cette sous-classe sont les éléments de type *Ac* du maïs découverts par Barbara McClintock (1984) et les deux éléments *P* et *mariner* découverts chez la drosophile (O'Hare et Rubin, 1983; Jacobson et al., 1986).

Cette sous-classe comprend les 10 superfamilles d'éléments suivantes : *Tc1-mariner* (Plasterk et al., 1999), *hAT* (McClintock, 1953), *Mutator* (Lisch, 2002), *P* (Bingham et al., 1982), *Merlin* (Feschotte, 2004), *Transib* (Kapitonov et Jurka, 2003), *piggyBac* (Fraser et al., 1985), *PIF-Harbinger* (Jurka et Kapitonov, 2001), *CACTA* (Kunze et Weil, 2002) et les ET de type *Crypton* qui sont classés dans une superfamille à part, du fait de l'absence de TIR et de gène codant la transposase (Goodwin et al., 2003).

Chaque superfamille est caractérisée par sa structure, la taille de la séquence dupliquée au niveau du site d'insertion ou TSD (Target Site Duplication) ainsi que des motifs de la transposase DDD ou DDE qui forment le cœur du site catalytique responsable de la transposition (Doak et al., 1994 ; Plasterk et al., 1999). Ces motifs signatures sont également

identifiés au cœur des séquences d'intégrases des rétrotransposons à LTR ce qui laisse supposer une origine commune (Capy et al., 1996).

Dans le paragraphe suivant seules 4 superfamilles seront développées : *Tc1-mariner*, *hAT*, *Mutator*, *P*, car elles sont à ce jour les plus répandues et les plus étudiées.

### **La superfamille *Tc1-mariner***

C'est la superfamille la plus répandue chez les organismes vivants, elle a été mise en évidence des protozoaires jusqu'à l'homme (pour revue : Plasterk et al., 1999; Shao et Tu, 2001). Elle est souvent associée aux Séquences d'Insertions (IS) bactériennes *IS630* formant alors la superfamille *ITm* pour *IS630-Tc1-Mariner* (Shao et Tu, 2001). Wicker et al., 2007 propose de la dénommer superfamille DTT (pour DNA transposon, TIRgroup, T*c1-mariner*). Une description plus détaillée de la superfamille *Tc1-mariner* fera l'objet de la partie III de ce chapitre, car les ET étudiés au cours du présent travail appartiennent à cette superfamille.

### **La superfamille *hAT***

Le nom de cette superfamille dérive des trois familles d'éléments : *hobo* chez la drosophile *Drosophila melanogaster* (McGinnis et al., 1983), *Activator-Dissociation* (*Ac/Ds*) chez le maïs, *Zea mays* (McClintock, 1953), et *Tam3* chez le muflier *Antirrhinum majus* (Coen et al., 1986). L'origine évolutive commune de ces trois familles a été mise en évidence par l'analyse de leurs séquences (Calvi et al., 1991). Les éléments de la superfamille *hAT* possèdent un gène codant une transposase de 500 à 800 acides aminés encadré par de courts TIR de 5 à 27 pb (Kempken et Windhofer, 2001). Les éléments actifs *Hermes* chez *Musca domestica* (Warren et al., 1994), *Tol2* chez le poisson *Oryzias latipes* (Koga et al., 1996) appartiennent à cette superfamille. Les études portant sur l'activité de ces deux éléments ont montré qu'ils pouvaient être utilisés comme outil de mutagenèse en système hétérologue et en cellules eucaryotes (Evertts et al., 2007; Kawakami, 2007).

### **La superfamille *Mutator***

Cette superfamille tire son nom de l'élément *Mutator* qui est le transposon de plante le plus actif et le plus mutagène découvert à ce jour (Alleman et Friiling, 1986; Jiang et al., 2004; Skibbe et al., 2009). Les éléments de la superfamille *Mutator* aussi dénommée *MULE* (*Mutator-like elements*) ont une longueur variant de 3 à 5 Kpb et possèdent de longs TIR de 100 à 200 pb. Ils possèdent également un gène codant une transposase de 700 à 900 acides aminés. La transposase de *Mutator* présente de fortes similarités avec le groupe bactérien

IS256. Les éléments de type *Jittery*, *MuDR* et *HOP* appartiennent à la superfamille *Mutator* (Lisch et al., 2001; Chalvet et al., 2003; Xu et al., 2004).

### **La superfamille des *P* éléments**

L'élément *P* a été décrit la première fois chez *Drosophila melanogaster* où sa mobilité est à l'origine du phénomène de dysgénèse dans la lignée germinale des mouches hybrides (Bingham et al., 1982). D'une longueur totale de 2,9 Kpb, les éléments *P* sont composés de TIR de 31 pb et d'un gène codant une transposase de 87 kDa (O'Hare et Rubin, 1983). Des éléments de la superfamille des éléments *P* ont été identifiés chez des métazoaires (Hammer et al., 2005) et chez l'algue unicellulaire *Chlamidomonas reinhardtii* (Jurka et al., 2005). L'élément *P* est actif et largement utilisé par les biologistes en tant qu'outil d'étude génomique et génétique de la drosophile (Rubin et Spradling, 1982; Ryder et Russell, 2003).

### **La Sous-classe 2**

#### **La superfamille *Helitron***

Les éléments *Helitron* décrits dans les génomes d'*Arabidopsis thaliana* et *Caenorhabditis elegans* où ils sont largement répandus (Kapitonov et Jurka, 2001) et récemment identifiés chez les champignons (Poulter et Goodwin, 2005; Hood, 2005), semblent se répliquer via un mécanisme de cercle roulant identique à celui du transposon bactérien *IS91* (Kapitonov et Jurka, 2001; Mendiola et De la Cruz, 1992). Ils ne génèrent pas de duplication du site d'insertion. D'une taille de 5,5 à 17 Kpb, les éléments *Helitron* ne possèdent pas de TIR. Ils codent une tyrosine recombinase similaire à celle trouvée dans des transposons *IS91* (Kapitonov et Jurka, 2001; Feschotte et Wessler, 2001).

#### **Les éléments *Maverick***

Les éléments *Maverick* ont été découverts chez le nématode *Caenorhabditis elegans* ou le poisson zèbre *Dario rerio* (Feschotte et Pritham, 2005). Ces éléments sont de longues séquences de 10 à 20 Kpb composées de TIR de 150 à 700 pb. Ils codent jusqu'à 11 protéines dont certaines de type viral et dont l'ordre des gènes peut varier suivant les éléments (Kapitonov et Jurka, 2006). Leur excision se fait sous la forme simple brin, l'élément est alors répliqué en ADN double brin puis réintégré au niveau de son site cible, il semblerait qu'ils réalisent des auto-répliquations (Pritham et al., 2007).

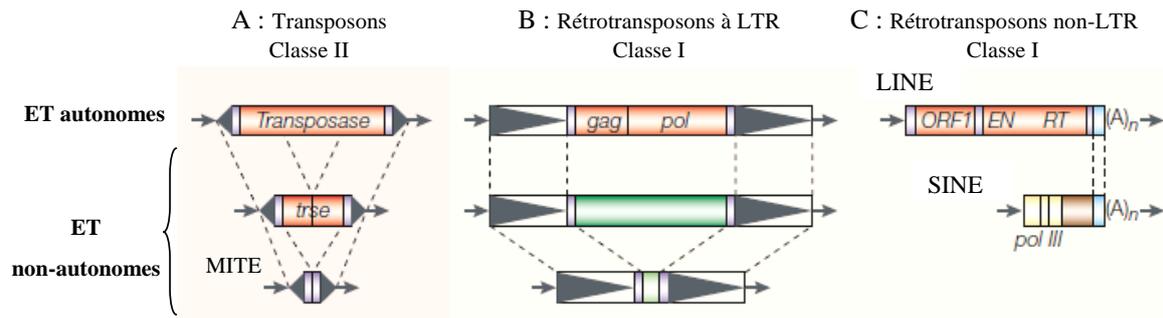
### II.2.3 Les éléments non autonomes

#### **Les SINE (Short Intersperced Nuclear Elements)**

Ces éléments sont abondamment présents dans le génome humain où ils sont très diversifiés (Lander et al., 2001). Non-autonomes et de petite taille, ils sont généralement inférieurs à 500 pb. Ils ne possèdent pas les gènes codant les enzymes permettant la transposition, ces derniers auraient été perdus au cours de l'évolution (Figure 12). Pour leur déplacement, ils semblent tributaires de la machinerie des LINE (Okada et al., 1997) et sont donc souvent classés avec les éléments de classe I (Hua-Van et al., 2005, Wicker et al., 2007). Ils dérivent des ARN de transfert (ARNt) aux exceptions près pour les familles *Alu* et *BI*. Les SINE ont une structure composite comprenant une région homologue aux ARNt, une région centrale dont la fonction reste inconnue bien que conservée au sein de la superfamille et une région variable riche en AT (Okada et al 1997). Les éléments les plus connus de cette superfamille sont les éléments *Alu*. Récemment Kramerov et Vassetzky (2005) ont démontré que les SINE *Alu* étaient mobilisés lors de transpositions accidentelles des transcrits de polymérase III et ne dépendaient pas de la machinerie des LINE comme il était supposé auparavant.

#### **Les MITE (Miniature Inverted Repeat Transposable Elements)**

Les MITE sont souvent intégrés au sein des superfamilles de transposons car les éléments de classe II autonomes sont capables de les déplacer (Feschotte et Mouches, 2000, Kidwell, 2002). Ils ont été décrits pour la première fois chez les plantes où il existe deux grandes familles, les éléments *Stowaway*-like éléments et les *Tourist*-like éléments (Bureau et Wessler, 1992, 1994). Ils ont ensuite été caractérisés chez les champignons (Yeadon et Catcheside, 1995; Dufresne et al., 2007), les amphibiens (Unsal et Morgan, 1995), les poissons (Izsvák et al., 1999) et l'Homme (Smit, 1996). Ils se caractérisent par leur petite taille qui est comprise entre 80 et 500 pb, TIR compris. Ils sont définis comme des versions tronquées et non autonomes des transposons de classe II car ils sont mobilisables *in trans* par les éléments autonomes complets et fonctionnels (Feschotte et al., 2002, 2003, 2005). Un faible nombre d'éléments autonomes peut être responsable de l'activation et de l'envahissement du génome par des formes non-autonomes. C'est le cas de l'élément *Osmar5* qui mobilise les MITE *Stowaway* chez le riz (Feschotte et al., 2005), ou des éléments *PIF-Harbinger* qui mobilisent les MITE *tourist* chez les plantes (Zhang et al., 2004), les insectes et les poissons (Jiang et al., 2004).



**Figure 12 : Principe de la transposition en *trans* des éléments non-autonomes à partir de la machinerie des éléments autonomes (d'après Feschotte et al., 2002).**

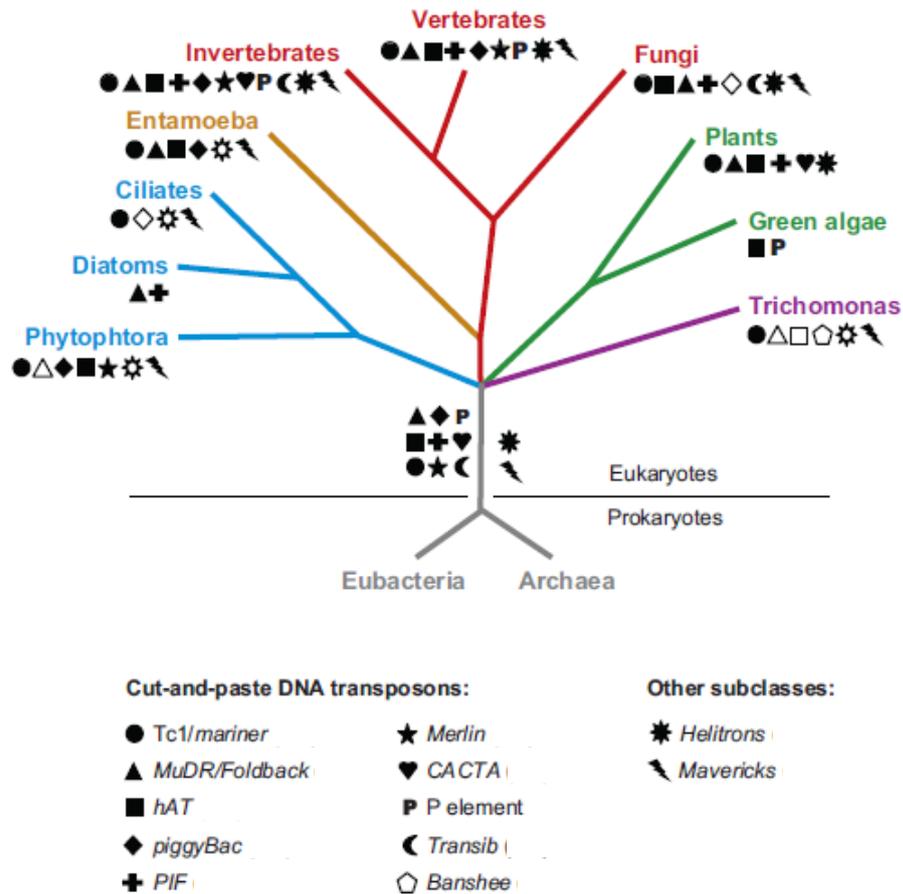
A -Les éléments non autonomes ayant conservé des TIR, peuvent être mobilisés par la transposase des éléments autonomes, même si il ne reste aucun vestige de cette séquence (c'est le cas de MITE, bien qu'ils ne dériveraient pas des éléments autonomes). B -Les éléments non-autonomes à LTR peuvent être mobilisés par les rétrotransposons à LTR même si leurs séquences centrales (en vert, pour les éléments non-autonomes) ne partagent aucune homologie. C : Les SINE non autonomes partagent une similarité au niveau de l'extrémité 3' poly-adenylée avec les rétrotransposons non-LTR ce qui suggère que les SINE seraient mobilisés par les LINE.

### II.3 Place des ET dans les génomes : diversité et dynamique des ET

La diversité des ET ne cesse d'augmenter depuis leur découverte, ceci s'explique par leur ubiquité dans les génomes. En effet, ils sont présents chez tous les êtres vivants chez lesquels ils ont été recherché, procaryotes et eucaryotes, animaux, végétaux ou champignons (pour revue : Kidwell, 2002; Feschotte et al., 2002).

#### II.3.1 Diversité des ET dans les génomes

Les génomes présentent une large variété d'ET (Figure 13) et dans des proportions différentes (Hua-Van et al., 2005; Feschotte et Pitham, 2007). Certaines superfamilles peuvent être prépondérantes et il peut y avoir une disproportion pour une classe ou une autre. Par exemple, les génomes végétaux et mammifères semblent dominés par les éléments de classe I (Kentner et al., 2003, Vitte et al., 2007). Cependant, la chauve souris, *Myotis lucifigus*, présentent une majorité d'ET de classe II, c'est aussi le cas des champignons ( Hua-Van et al., 2000; Daboussy et Capy, 2003; Ray et al., 2008; Pace II et al., 2008). Ces distributions ont des conséquences sur la diversité et la plasticité génomique de par leur mécanismes, leurs spécificités d'insertion etc. (Kidwell et Lisch, 1997; Bhöne et al., 2008; Venner et al., 2009; Tenaillon et al., 2010).



**Figure 13 : Distribution des ET de classe II dans les génomes eucaryotes (d'après Feschotte et Pritham, 2007).**

Cet arbre schématique représente la distribution des superfamilles d'ET de classe II (en pictogrammes) au sein des principaux groupes d'eucaryotes.

### II.3.2 Impact des ET sur la taille des génomes

Les ET appartiennent à la fraction des séquences répétées et dispersées dans les génomes. Leur quantité est variable selon les organismes considérés et ils peuvent représenter de 1,9 à 60 % de l'ADN génomique total d'un organisme (Tableau 1).

De nombreuses études ont postulé qu'ils pouvaient expliquer en grande partie l'accroissement des génomes (San Miguel et Bennetzen, 1998; Vitte et Panaud, 2005, Pace et al., 2008) car les génomes les plus grands, ceux des végétaux, sont composés en majorité d'ET, et les plus petits en présentent un faible pourcentage (Tableau 1). Toutefois, cette corrélation n'est pas linéaire (Kidwell, 2002). Les génomes compacts tel que les génomes procaryotes présentent une faible proportion d'ET. Toutefois, ceci peut s'expliquer par l'hypothèse de nombreuses périodes d'extinction d'ET au sein des lignées bactériennes dues à des remaniements génomiques fréquents chez ce type d'organismes (Wagner, 2006; Treangen et al., 2009).

**Tableau 1 : Pourcentage d'ET dans les génomes modèles (modifié d'après Kidwell, 2002).**

Espèces (Noms communs)	% d'ET par rapport à la taille du génome	Taille du génome (Mpb)
<i>Thalassiosira pseudonana</i> (Diatomée)	1,9	32
<i>Fugu rupribes</i> (Tetraodon)	2	400
<i>Saccharomyces cerevisiae</i> (Levure)	3	12
<i>Caenorhobditis elegans</i> (Nematode)	6	100
<i>Phaeodactylum tricornutum</i> (Diatomée)	6,4	27
<i>Dictyostelium discoideum</i> (Amibe)	10	34
<i>Arabidopsis thaliana</i> (Arabette)	14	125
<i>Oryza sativa</i> (Riz)	14	430
<i>Drosophila melanogaster</i> (Mouche)	15	180
<i>Anopheles quadrimaculatus</i> (Moustique)	16	245
<i>Mus musculus</i> (Souris)	40	3250
<i>Homo sapiens</i> (homme)	44	3000
<i>Hordeum vulgare</i> (Orge)	55	5000
<i>Zea mays</i> (Maïs)	60	2500

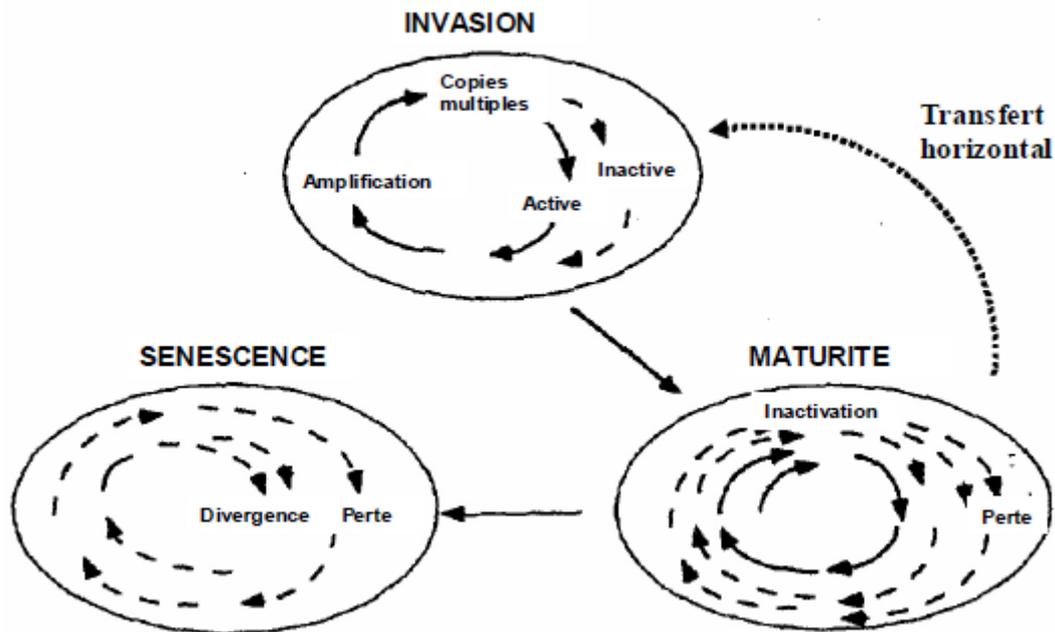
En fait le nombre d'ET est très probablement sous-estimé car cette estimation n'est possible que pour les génomes séquencés complètement et la détection des ET est réalisée par comparaison avec des banques de données, ainsi les ET inconnus ne sont pas détectés.

La prolifération des ET, notamment ceux de classe I, peut expliquer des différences de taille de génome entre les populations au sein d'une même espèce (Vieira *et al.*, 2002; Vitte et Panaud, 2005), cependant, les ET de classe II y contribuent également (Kidwell, 2002; Charles *et al.*, 2008, Aziz *et al.*, 2010). Ainsi les phases d'accumulation mais aussi de pertes ont un impact certain sur la composition génomique et les ET peuvent surpassés les gènes en quantité dans certains génomes, cela s'explique par le succès de leur dynamique de multiplication.

### II.3.3 Dynamique des ET dans les génomes

L'évolution des ET dans le génome hôte, suit une dynamique en trois phases qui sont successivement : la phase d'invasion, de maturité et la sénescence (Figure 14).

La première étape est la phase d'invasion du génome par un ET actif qui peut être le résultat d'un transfert vertical ou horizontal, elle s'effectue d'abord lentement jusqu'à l'augmentation exponentielle du nombre de copies de l'ET actif introduit.



**Figure 14 : Les trois phases du cycle de vie des ET dans les génomes (d'après Kidwell et Lisch, 2001).**

La phase d'invasion correspond à l'augmentation rapide du nombre de copies, elle est accompagnée de mutations entraînant l'inactivité de certaines copies. Ceci conduit à une phase de maturité où les amplifications sont contrebalancées par les pertes des ET (inactivation). Lorsque tous les éléments actifs sont perdus ou mutés (divergence des séquences), la balance gain/perte est déséquilibrée ce qui conduit à la phase de sénescence. Les éléments autonomes et actifs sont représentés en lignes continues et les éléments non-autonomes sont représentés en pointillés.

1) L'amplification exponentielle des ET est appelée burst de transposition. Ce phénomène qui constitue le point de départ de l'invasion des ET, est difficile à mettre en évidence car le nombre d'éléments actifs connus est restreint et les taux de transposition difficile à estimer pour la plupart des ET. Néanmoins, deux exemples : les éléments *Mu* (*Mutator*) atteignant près d'un événement par génération chez le maïs (Alleman et Freeling, 1986) et les éléments *mPing* qui peuvent former environ 56 copies par génération, confirment la capacité des ET à envahir le génome (Naito et al., 2006). Des exemples tels que celui de l'élément *hopi* chez *O. sativa* montrent qu'après plusieurs bursts récents au sein d'un génome, la population d'éléments ne contient plus d'éléments actifs suggérant la présence de phénomènes de régulation limitant la prolifération des ET (Vitte et Panaud, 2005).

2) La deuxième phase correspond à la « maturité » durant laquelle des systèmes de régulation de l'activité ainsi que l'apparition de mutations vont limiter la prolifération des éléments. D'après Kidwell et Lisch (2001), au cours de l'évolution, des mécanismes de régulation limitant les effets néfastes de la transposition, auraient été sélectionnés. Plusieurs types de mécanismes de régulation sont connus à l'heure actuelle :

- la régulation de la transposition par des mécanismes classiques de répression qui exercent un rétrocontrôle de la transposition. L'existence de facteurs capables de réprimer la

transcription des séquences d'insertions bactériennes *IS1* et *IS2* a été démontrée dans de nombreuses études (Machida et Machida., 1989; Zerbib et al., 1990; Hu et al., 1994). Un autre mécanisme de régulation de la transposition est l'autorégulation qui correspond à la surproduction de transposase dans le milieu. Ceci entraîne une inhibition de la production de la protéine par la formation d'oligomères de transposase en excès. L'autorégulation a été identifiée au sein de la superfamille *Tc1-mariner* (Lohe et Hartl, 1996; Zhang et al., 2001).

- la régulation épigénétique est impliquée dans la régulation génétique de manière générale. Ce mécanisme qui permet un contrôle de l'activité des ET par une méthylation de l'ADN, principalement sur les cytosines, entraînant ainsi une mise sous silence de l'ET, a été mis en évidence chez le maïs, *Zea mays*, pour l'élément *Mu1* (Martienssen et Baron, 1994) et chez *A. thaliana* pour l'élément *Tto1* (Slotkin et Martienssen, 2007). L'activité des ET peut être régulée via des ARN anti-sens qui ont la possibilité de former des dimères avec ceux produits par la transcription des ET. Cette régulation post-transcriptionnelle bloque la production des enzymes permettant la transposition qui ne sont alors pas produites, bien que l'ET soit exprimé (Abrusàn et al., 2006; Slotkin et Martienssen, 2007). Ce phénomène serait à l'origine de l'inhibition des éléments *Tc1* (Sijen et Plasterk, 2003), et *Botmar1* (Petit et al., 2007).

3) La phase de sénescence correspond à une diminution drastique du nombre de copies de l'ET considéré qui aboutit à la perte de cet ET à terme, voire de familles entières d'ET, au sein d'un génome, lorsque le taux de transposition est faible. Deux mécanismes conduisent à la phase de sénescence :

- L'inactivation verticale qui correspond à l'accumulation de mutations (substitutions ou insertions/ délétions) conduisant à la perte de fonctionnalité des ET (Lohe et al., 1995; Lohe et al., 1997) au cours des générations. La balance entre perte et gain est déséquilibrée et les gènes mutés ne sont plus transcrits et/ou ont perdu des domaines nécessaires à la mobilisation. Par exemple, lorsque la mutation/délétion a lieu dans la région catalytique, elle entraîne l'inactivation de l'élément (Lohe et al., 1997). Ce processus est considéré comme un des facteurs les plus influents sur la régulation de la dynamique des ET (Brookfield et Badge, 1997; Capy, 2004; Le Rouzic et Capy, 2005).

- La perte stochastique correspondant à des délétions de courts ou longs fragments d'ADN par dérive génétique peut entraîner l'élimination d'ET et donc participer à la sénescence (Lohe et al., 1995). Ce mécanisme a été étudié au sein de la famille *mariner* et permet d'expliquer en partie, la distribution sporadique des éléments *mariner*-like (*MLE*) au sein de différentes sous espèces de drosophiles (Lohe et al., 1995; Wang et al., 2005).

Cependant, ce concept est controversé, notamment par Biémont et al., 1994 qui considère la perte stochastique comme négligeable par rapport à l'inactivation verticale dans le cas de l'élément *P* chez la drosophile.

- La recombinaison est un phénomène qui correspond à l'appariement de séquences homologues et à l'excision de toute séquence d'ADN présente entre les deux sites homologues. Ce mécanisme a été mis en évidence chez les rétrotransposons à LTR, lorsqu'un très grand nombre de copies est présent dans le génome, il y a des recombinaisons entre les LTR identiques des rétrotransposon et la séquence codante peut être en partie ou en intégralité perdue. La recombinaison est un facteur majeur permettant de contenir l'expansion des ET (Biémont et al., 1997) et représente un phénomène important de la perte d'ADN au sein des génomes (Kalendar et al., 2010; Vitte et al., 2007).

#### II.3.4 La transmission des ET

Le mode de transmission « classique » des ET est le transfert vertical encore nommé mendélien (Lidholm et al., 1993; Lohe et al., 1995), c'est-à-dire la transmission des ET à la descendance via les lignées germinales (Figure 15). Depuis peu, le transfert horizontal des ET est pris en compte dans la balance gain/perte d'ET au sein des génomes. Ce mécanisme correspond au transfert d'informations génétiques entre espèces reproductivement isolées (Figure 15). Maruyama et Hartl (1991a) ont mis en évidence un transfert horizontal par l'analyse des séquences de *MLE* chez différentes espèces de drosophiles. Un taux d'identité supérieur à 95 % entre un *MLE* provenant de *Zaprionus tuberculatus* et *Mos1* isolé chez *Drosophila mauritiana*, deux espèces hôtes éloignées phylogénétiquement a mené les auteurs à la conclusion d'un transfert d'ET inter-espèces. Bien que ce phénomène soit bien caractérisé chez des espèces bactériennes, le transfert horizontal est toujours considéré comme un mécanisme marginal chez les eucaryotes (Koonin et Wolf, 2008), mais peut être démontré par l'incongruence phylogénétique (Diao et al., 2006). Concernant les ET, il a été particulièrement étudié chez les éléments *mariner* et *MLE* d'origine animale où il est soupçonné d'être à l'origine de la distribution en îlots (« patchy ») des *MLE*. (Robertson, 1993; Lohe et al, 1995; Robertson et Lampe, 1995). De nombreux exemples ont été mis en évidence avec d'autres ET tel que l'élément *P* entre les espèces *Drosophila willistoni* et *Drosophila melanogaster* (Silva et Kidwell, 2004), mais aussi chez les végétaux (Diao et al., 2006; pour revue Fortuné et al., 2008). C'est le cas de l'élément *RIRE1* dont la distribution au sein du genre *Oryza* est le résultat de nombreux transferts horizontaux entre huit espèces différentes (Roulin et al., 2008). Néanmoins, les ET transmis de manière verticale peuvent

être perdus au sein des espèces si bien qu'il est parfois difficile de privilégier l'une des deux hypothèses. De plus, les transferts horizontaux nécessitent des vecteurs qui permettraient de transmettre les ET d'une espèce à une autre. Les virus, les bactéries ou les champignons sont des vecteurs potentiels (Hartl et al., 1997a, Kidwell et Lisch, 2001), mais cette hypothèse n'a pas été vérifiée à ce jour (Won et Renner, 2003; Pace et al., 2008), bien que les parasites apparaissent comme des vecteurs de transfert crédibles (Gilbert et al., 2010).

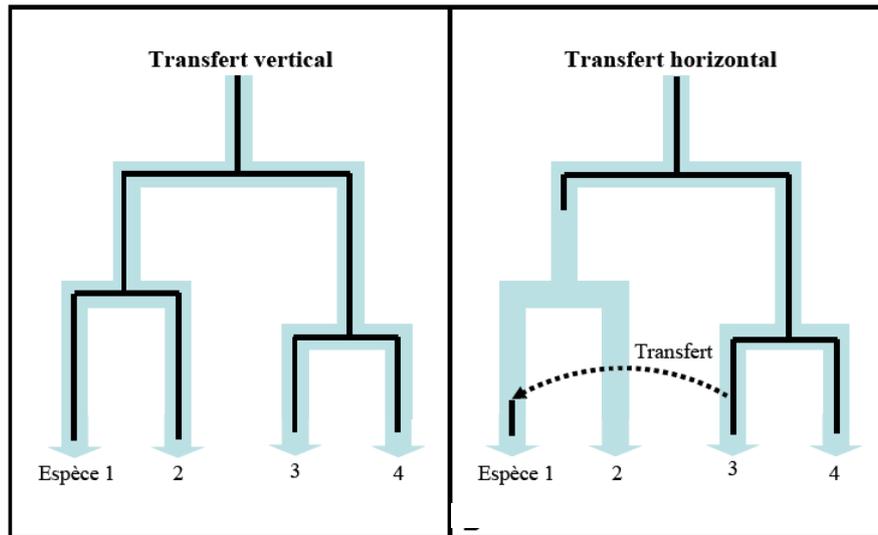


Figure 15 : Schéma simplifié du mode de transfert vertical et horizontal des ET (d'après Eickbush et Malik, 2002).

#### II.4 Impact des ET sur les génomes et plasticité génomique

La transposition des ET induit des modifications génomiques qui peuvent être dues à l'insertion et ou l'excision voire simplement à la cicatrice laissée par le déplacement. L'excision des transposons n'étant pas toujours précise, quelques paires de bases peuvent rester dans le site d'excision et la fonction initiale du gène en ce site n'est donc pas restaurée dans ce cas (Coen et al., 1989; Plasterk, 1999; Langin et al., 1995). Les modifications résultant des ET ont longtemps été classées en fonction des effets sur le génome hôte qui sont qualifiés de neutres, délétères ou bénéfiques (Kidwell et Lish, 2000). Or, les mutations délétères sont contre-sélectionnées (ils entraînent la non-viabilité des hôtes) et sont donc difficilement repérables (Venner et al., 2009). Dans les génomes eucaryotes, on observe une tendance à l'accumulation d'ET dans l'hétérochromatine plutôt que dans les séquences codantes (euchromatine). Ce fait n'est peut être que le reflet de la contre-sélection des ET insérés dans les gènes dit « stables », la sélection génomique cantonnerait en quelque sorte les ET dans des régions où les effets seraient neutres. Toutefois, avec le développement de

l'épigénétique, l'observation d'une corrélation entre la présence des ET dans l'hétérochromatine et l'hyperméthylation de cette fraction génomique, suggère que les ET accumulés n'auraient pas un effet neutre mais un rôle régulateur de l'expression génique (Slotkin et Martienssen, 2007).

En ce qui concerne les effets bénéfiques, ils sont sélectionnés et donc détectables. L'exemple le plus représentatif est le recrutement des ET par le génome, ce phénomène est baptisé « domestication ». Les ET recrutés ont un impact direct sur l'évolution des génomes, par la formation de nouveaux gènes chimériques composés d'une partie de l'ET et d'une partie d'ADN provenant de l'hôte. De nombreux exemples ont pu être mis évidence (pour revue : Feschotte, 2008; Sinzelle et al., 2009) comme les rétrotransposases d'éléments *pogo* qui ont été recrutées par les génomes d'insectes et de vertébrés et dont les protéines centromériques dériveraient (Casola et al., 2008). Des séquences codant des transposases entières sont également utilisées au bénéfice du génome hôte tel que le gène *Rag* qui dérive de la transposase d'une ancienne famille de transposons appelée *Transib* (Kapitonov et Jurka, 2005).

L'insertion d'ET dans les gènes ou les séquences régulatrices, peut induire l'inactivation du gène ou la modulation de l'expression des gènes. Ces insertions peuvent donc entraîner des changements phénotypiques. Le cas le plus représentatif est celui de *Tam3* qui s'insère dans la région régulatrice 5' du gène, *niv*, impliqué dans la synthèse des anthocyanines du mufler, *Antirrhinum majus*. Cette insertion induit une diminution de l'expression du gène *niv*. De fréquentes transpositions conduisent aux réarrangements chromosomiques et influent également sur la spécificité tissulaire de l'expression du gène *niv* (Lister et al., 1993; Haschida et al., 2003). Il est difficile de déterminer si ces modifications ont un effet bénéfique pour l'hôte, dans tous les cas elles induisent une évolution du génome de l'hôte.

## II.5 Influence des facteurs environnementaux sur les ET et le génome

Barbara Mc Clintock en 1984, faisait l'hypothèse que l'activité des ET pouvait être induite en réponse à des facteurs de stress génomiques. Ceux-ci nécessitant une restructuration de ce même génome pour survivre. Dans cette optique, l'activité des ET pourrait constituer une stratégie adaptative du génome face à des chocs tels que les stress environnementaux (pour revue : Wessler, 1996; Schmidt et Anderson, 2006).

Capy et al. en 2000, ont proposé un modèle de l'implication à court terme et long terme de l'activité des ET en réponse aux stress environnementaux (Figure 16). Les ET

feraient ainsi parti de la cascade de réactions face aux stress comme le propose Li et al. (1999) et pourraient avoir un rôle de signal d'alarme. L'induction des ET de manière concomitante avec l'induction des gènes de réponse aux stress peut être due à une homologie de séquence des promoteurs de ces gènes avec la séquence des ET. Cette caractéristique transmise à la descendance, fournirait aux organismes une meilleure adaptation face aux changements environnementaux. L'induction des ET en condition de stress représenterait donc une adaptation à long terme de la réponse aux stress.

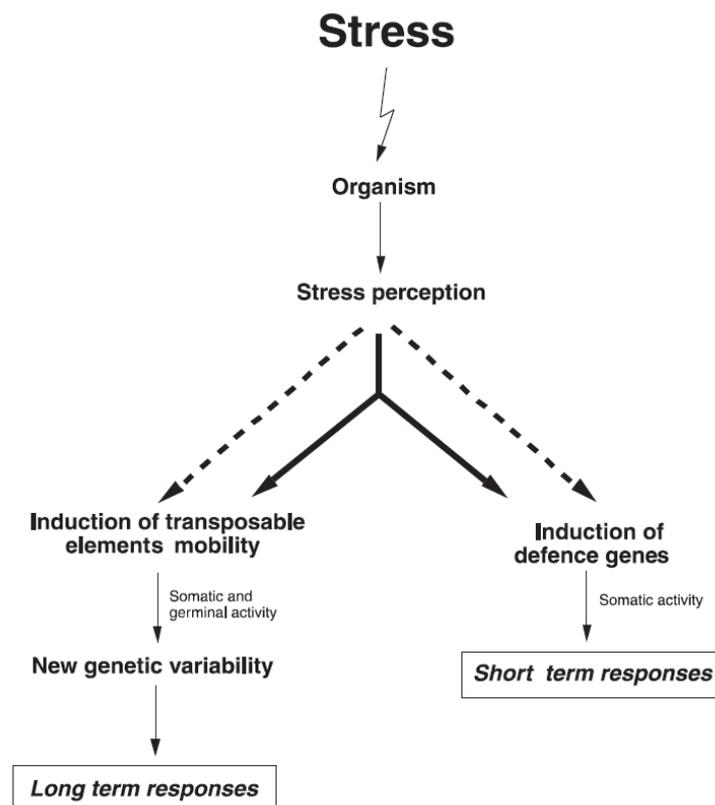
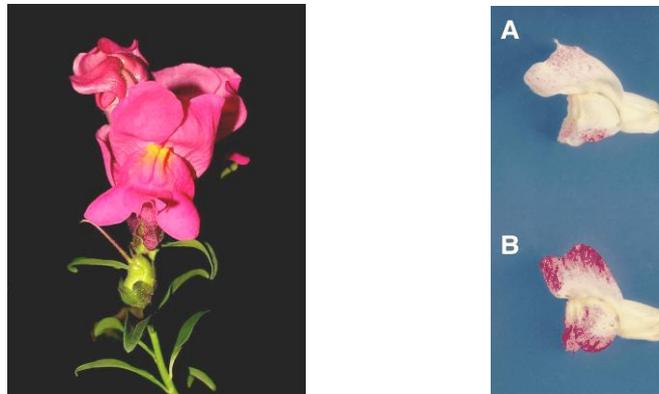


Figure 16 : Modèle du rôle des ET dans les mécanismes d'adaptation au stress (d'après Capy et al., 2000).

De nombreuses études ont été menées visant à rechercher l'impact des stress abiotiques (blésures, attaques chimiques...) et biotiques (attaques fongiques ou bactériennes) sur l'expression des ET et/ou l'activation de la transposition (Tableau 2), notamment, le facteur température a largement été étudié (Tableau 2).

L'étude de l'activité de *mariner* chez des populations de drosophiles d'origines géographiques différentes (France, Maroc et Congo) a permis de montrer que la température du site d'origine des drosophiles influence sur l'activité de l'élément *mariner* soumis à des variations thermiques (Chakrani et al., 1993). L'activité de l'élément *mariner* augmente avec

la température pour deux des trois populations testées, en ce qui concerne la troisième cette observation n'est pas significative suggérant que l'activation de *mariner* par des stress thermiques pourrait être dépendante d'une adaptation locale de l'hôte (Chakrani et al., 1993) ainsi que le propose le modèle de Capy et al. (2000). La transposition de l'élément *Tam3* au sein du génome du mufler, *Antirrhinum majus*, est plus importante à 15 °C qu'à 25 °C, l'activité de ce transposon est strictement dépendant de la température (Figure 17) (Coen et al., 1986; Hashida et al., 2003, 2006). Des effets similaires ont été reportés pour le transposon *Tn3* d'*Escherichia coli* (Kretschmer et Cohen, 1979). Dans un test de transposition *in vitro*, Lampe et al. (1998) ont également montré que l'efficacité de transposition de *Himar1* augmentait d'un facteur 10 entre 15 °C et 28 °C puis diminuait lors d'expositions à des températures plus élevées.



**Figure 17 : Changements phénotypiques liés à l'activité de *Tam3* dans le gène *niv* du mufler (*Antirrhinum majus*).**

La photo de gauche correspond au phénotype initial du mufler soit une couleur rose intense. Les photos de droite correspondent à des événements de transpositions de l'élément *Tam3* chez cette espèce. Lorsque *Tam3* est inséré dans le promoteur de gène *niv*, l'expression de celui-ci est inhibée, le phénotype associé est l'apparition d'une couleur rosée très pale (A). Lors de l'excision de l'élément, le phénotype initial est restauré, ceci est observable par des secteurs roses intenses (B). Ici sont représentées les lignées cultivées à 15 °C en A et 21 °C en B, la comparaison de ces deux phénotypes met en évidence que l'excision de *Tam3* est corrélée avec la diminution de température. Le passage de 21 à 15 °C induit une diminution de l'intensité de la couleur du mufler qui correspond à l'insertion de *Tam3*.

Les ET sont souvent induits par des chocs thermiques, aussi bien par des températures élevées que basses et cette induction peut être due à la présence de séquences similaires entre les promoteurs ou les régions régulatrices proches des ET et les promoteurs des protéines de choc thermique (Strand et McDonald, 1985; Zucker et al., 1984; Chakrani et al., 1993). C'est le cas notamment des éléments *OPOHIO1* et 2 des espèces fongiques du genre *Ophiostoma* qui présentent des motifs similaires à ceux des promoteurs de gènes codant les protéines de choc thermique (Bouvet et al., 2008). La température n'est pas le seul facteur induisant l'activation des ET.

Au sein du génome du tabac co-existent trois familles de *Tnt1* qui sont activées par des stress abiotiques différents tels que des traitements au méthyl jasmonate, à l'acide salicylique ou encore à l'auxine. Les promoteurs nommés BI et BII présents dans la région 3' UTR des ET sont responsables du différentiel d'activation des trois familles de *Tnt1* (Mhiri et al., 1997; Beguiristain et al., 2001). Ces régions promotrices présentent des motifs similaires à ceux impliqués dans l'activation des gènes de défense des plantes (pour revue : Grandbastien et al., 1997). Les éléments *Tnt1* seraient donc liés aux réponses de la plante face aux stress de manière générale (Grandbastien, 1998) car de nombreux stress induisent leur expression chez les Solanacées tels que : blessures, divers traitements avec des éliciteurs, formation de protoplastes, attaques fongiques, etc. (Pouteau et al., 1991; Casacuberta et Grandbastien, 1993; Mhiri et al., 1997; Grandbastien et al., 2005).

Le stress hydrique peut également induire une augmentation du taux de transposition comme c'est le cas de *BARE-1* chez l'orge, *Hordeum spontaneum* (Kalendar et al., 2000) ou encore les UV peuvent augmenter le taux de transposition du rétrotransposon *Remel* au sein du génome du melon, *Cucumis melo* (Ramallo et al., 2008).

De même, la mise en culture de tissus induit la transposition de l'élément *Tos17* dans le génome du riz (Cheng et al., 2006), de *Tto1* dans le tabac (Hiroshika et al., 2000) et des rétroéléments *MERE1* de la luzerne (Rakocevic et al., 2009). Plus précisément, dans le cas de *Tos17*, c'est la régénération depuis le cale qui entraîne des méthylations inhibant la transposition (Cheng et al., 2006).

Différents stress testés sur le champignon *Aspergillus oryzae* (pH, peroxyde d'hydrogène, UV, métallique et thermique) ont mis en évidence l'induction de l'expression du transposon *crawler* appartenant à la famille *Tc1-mariner* (Ogasawara et al., 2009). Les conditions induisant une augmentation de l'excision se sont révélées être des chocs thermiques et une addition de sulfate de cuivre (CuSO<sub>4</sub>) (Ogasawara et al., 2009). Les autres stress (pH, UV, stress métalliques...) ont induit l'excision du transposon, dans des proportions variables qui n'ont pas permis de mettre en évidence une augmentation significative due au stress (Ogasawara et al., 2009). Les ET peuvent donc être activés spécifiquement par certains stress et cela reste dépendant de l'hôte et de l'élément considéré (Grandbastien et al., 1997; Beguiristain et al., 2001; Schmidt et Anderson, 2006).

Récemment une étude portant sur l'activation des rétrotransposons de diatomées appartenant à la superfamille *Ty1-copia* nommés *CoDi* (*Ty1/Copia*-like elements from Diatoms) a été réalisée. Ces travaux ont mis en évidence une expression différentielle des ET en relation avec différents stress abiotiques tels que la limitation ou l'addition de fer ou silice

dans le milieu de culture, des traitements à l'aldehyde décadienal toxique (DD), des limitations en nitrates ou en dioxyde de carbone, etc. (Maumus et al., 2009). Les taux de *CoDi* transcrits sont significativement supérieurs par rapport au témoin chez la diatomée *P. tricornutum* soumise à l'addition d'une forte quantité de DD (5µg/mL) pendant 6 heures, mais une dose de 0,5 µg/mL n'induit pas une expression significative. Une induction de l'expression est également observée quand cette espèce est cultivée en condition de carence en nitrates ou soumise à un traitement UV pendant 1 heure sur des cultures adaptées à l'obscurité. Chez la diatomée *T. pseudonana*, les *CoDi* sont significativement plus exprimés en conditions limitantes en fer et en silice. Cette étude a également révélé que la majeure partie des *CoDi* induits par ces stress est constituée de transcrits des deux rétrotransposons nommés *Blackbeard* et *Surcouf*, qui sont particulièrement induits par le stress DD et la carence en nitrate (Maumus et al., 2009).

Il est évident que les facteurs environnementaux ont un impact sur les ET que ce soit l'activation ou la répression de ceux-ci mais les spécificités de la relation stress/ET suggèrent des mécanismes complexes de régulation (Li et Schmidt, 2001) et souvent ne permettent pas de définir si le stress a un effet direct sur les ET ou si le stress a un impact sur le statut épigénétique de ces éléments modifiant leur régulation (Figure 18) (Maldhung et Comai, 2004; Schmidt et Anderson, 2006; Maumus et al., 2009). Il est très probable que ces mécanismes soient des mécanismes importants dans l'évolution et l'adaptation des génomes (Figure 18) (Biémont et Vieira, 2006; Tenailon et al., 2010).

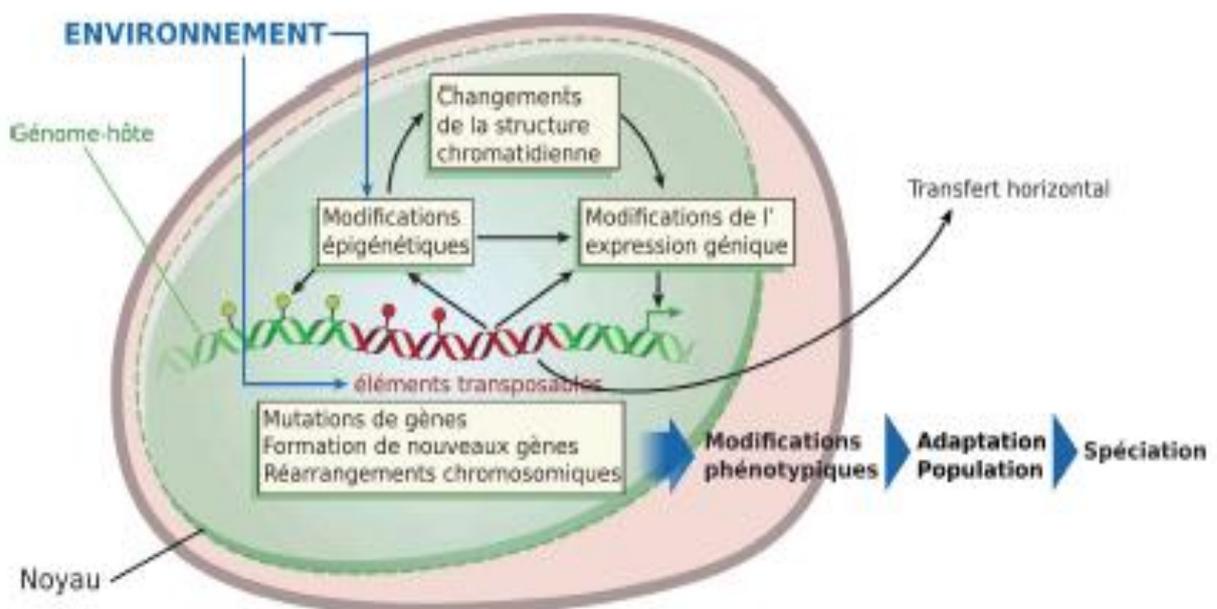


Figure 18 : Modèle écologique d'évolution des ET dans les génomes en relation avec les facteurs environnementaux selon Biémont et Vieira (2006).

**Tableau 2 : Réponse des ET aux stress abiotiques et biotiques.**

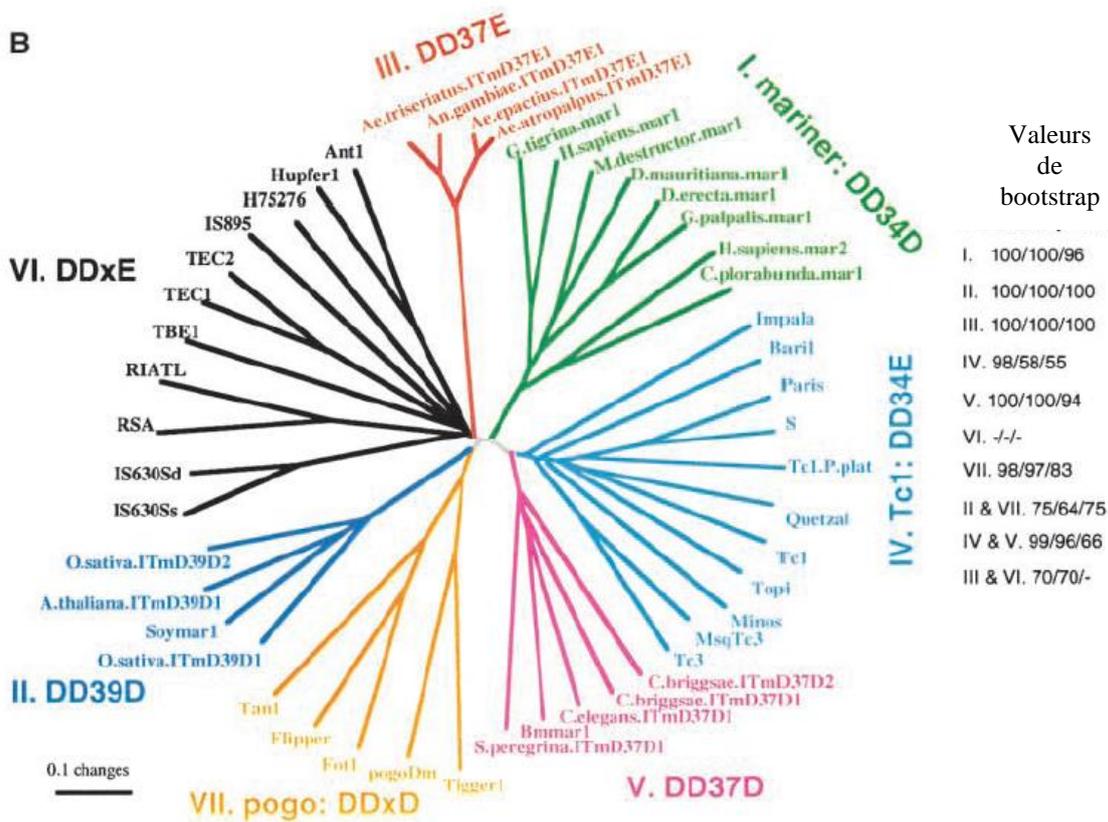
Le symbole + signifie une induction de l'expression ou bien une activation de la transposition, le symbole – correspond à une répression de l'expression ou a une inactivation. +/- correspond à une activation incomplète (excision seulement).

	Stress	ET / espèce	expression	répression	transposition	inactivation	Référence
<b>Température / abiotique</b>	Chaud	<i>Tn3</i> / bacterie			+		Kretschmer et Cohen, 1979
		<i>DIRS1</i> / amibe	+				Zuker et al., 1984
		<i>Copia</i> / drosophile	+			+	Strand et McDonald, 1985
		<i>B2</i> / hamster + souris	+				Fornace et Mitchell, 1986
		<i>Dm-412</i> / drosophile					Ratner et al., 1992
		<i>Bm1</i> / papillon	+				Kimura et al., 1999
		<i>Bm1</i> / cellules de souris	+				Li et al., 1999
		<i>Bm1</i> / ver à soie (larve de papillon)	+				Kimura et al., 2001a
		<i>Alu</i> / cellules eucaryotes	+				Li et Schmidt, 2001
		<i>LINE</i> / cellules eucaryotes				-	Li et Schmidt, 2001
		<i>Tam3</i> / gueule de loup				+	Kitamura et al., 2001
		<i>Maggy</i> / champignon	+				Ikeda et al., 2001
		Divers ET /corail	+				Desalvo et al., 2008
		<i>OPHIO</i> / champignon				+/-	Bouvet et al., 2008
		<i>Crawler</i> / champignon				+	Ogasawara et al., 2009
		<i>Copia + retrovirus</i> / algue brune	+				Pearson et al., 2010
		<b>Température / abiotique</b>	Froid	<i>ISH27</i> / bactérie			+
<i>Tnt1</i> / tabac + tomate	+						Mhiri et al., 1997
<i>OPHIO</i> / champignon	+					+	Bouvet et al., 2008
<i>MCIRE</i> / luzerne	+						Ivashuta et al., 2002
<i>Tam3</i> / geuele de loup	+					+	Hashida et al., 2003
<b>Biotique</b>	Infection fongique ou bacterienne	<i>Tnt1</i> / tabac + arabète + tomate	+				Pouteau et al., 1994; Melayah et al., 2001; Moreau-Mihri et al., 1996
		<i>Oare-1</i> / avoine	+				Casacuberta et Grandbastien, 1993; Mhiri et al., 1997 Kimura et al., 2001b

	Stress	ET /espèce	expression	répression	transposition	inactivation	Référence	
Abiotique	Milieu apauvri	<i>Tn4652</i> / bactérie				+	Ilves et al., 2001	
	Carence en nitrates	<i>Balckbeard</i> + <i>Surcouf</i> / diatomées	+				Maumus et al., 2009	
	UV		<i>IS10</i> / bacterie	+		+		Eichenbaum et Livneh, 1998
			<i>OPHIO</i> / champignon	+		+		Bouvet et al., 2008
			<i>Oare-1</i> / avoine					Kimura et al., 2001b
	Culture de tissus		<i>Tos17</i> + <i>Tos10</i> / riz				+	Hiroshika et al., 1996
			<i>Tto1</i> / Tabac				+	Hiroshika et al., 2000
			<i>MERE-1</i> / luzerne				+	Rakocevic et al., 2009
			<i>Tnt1</i> / luzerne				+	D'Erfurth et al., 2003
	Protoplastes		<i>LORE1</i> / Lotus	+				Madsen et al., 2005
			<i>TLC1</i> / tomate	+				Tapia et al., 2005
			<i>Tnt1</i> / tabac + arabète + tomate	+			+	Casacuberta et Grandbastien, 1993; Grandbastien et al., 1997
	Blessure		<i>Tto1</i> / tabac	+				Takeda et al., 1999
			<i>TLC1</i> / tomate	+				Tapia et al., 2005
	Signaux de défense (méthyl jasmonate, éliciteurs, acide salicylique...) + inducteurs de stress oxydant		<i>Oare-1</i> / avoine	+				Kimura et al., 2001
			<i>Tto1</i> / tabac	+				Takeda et al., 1998
			<i>TLC1</i> / tomate	+				Tapia et al., 2005
		<i>Bm1</i> / cellules de souris	+				Li et al., 1999	
		<i>Crawler</i> / champignon				+	Ogasawara et al., 2009	
		<i>Balckbeard</i> + <i>Surcouf</i> / diatomées	+				Maumus et al., 2009	
		<i>TLC1</i> / tomate	+				Tapia et al., 2005	
		<i>Oare-1</i> / avoine	+				Kimura et al., 2001b	
		<i>Tnt1</i> / tabac + arabète + tomate	+			+	Moreau-Mihri et al., 1996 ; Grandbastien et al., 1997 ; Mhiri et al., 1997 ; Vernhettes et al., 1997	
		<i>Maggy</i> / champignon	+				Ikeda et al., 2001	

### III La superfamille *Tc1-mariner*

La superfamille *Tc1-mariner* comprend les éléments apparentés à l'élément *mariner* isolé chez la drosophile (Jacobson et al., 1986) nommés *MLE* (*mariner*-like elements) et les éléments apparentés à l'élément *Tc1* isolé chez *C. elegans* (Emmons et al., 1983) nommés *TLE* (*Tc1*-like elements). Toutefois, elle peut être aussi référencée comme la superfamille *ITm* (pour *IS630-Tc1-mariner*) par intégration des transposons bactériens IS630 (Séquence d'insertion 630) présentant des similitudes structurales avec *Tc1-mariner* (Shao et Tu, 2001; Brillet et al., 2007). Largement distribués au sein des organismes, les éléments de cette superfamille présentent des similitudes de séquences des gènes codant les transposases et plus spécifiquement des motifs caractéristiques composés de trois acides aminés impliqués dans l'activité catalytique de l'enzyme (« triade catalytique »). Ces acides aminés peuvent être trois acides aspartiques (DDD) caractéristiques des *MLE* ou deux acides aspartiques et un acide glutamique (DDE) typiques des *TLE* (Doak et al., 1994; Plasterk et al., 1999). Une autre caractéristique de cette triade catalytique est la distance  $x$  entre le deuxième résidu aspartate et le troisième acide aminé (Plasterk et al., 1999; Shao et Tu, 2001). Shao et Tu (2001) proposent de classer les éléments en sept familles sur la base des caractéristiques de la triade catalytique (Figure 19) Chaque famille est nommée DDxD/E, ainsi les sept familles sont DD37E, DD37D (ou *mAT*), DD39D (*MLE* de plantes), DD41D, DD34E (ou *TLE*), DD34E bis (ou *Gambol*), DD34D (ou *MLE*). Cependant cette dénomination ne permet pas de distinguer les éléments de type *Gambol* des éléments de type *TLE* qui sont tous les deux définis par une triade DD34E. De plus, et la famille VI ou DDxE contient au moins trois types de triades différentes (J. D. Rouault, communication personnelle). Bien que ces familles soient bien définies, leurs positions les unes par rapport aux autres et leurs liens phylogénétiques restent non déterminés.



**Figure 19 : Classification des éléments de la superfamille *ITm* en fonction des domaines de la triade catalytique (d'après Shao et Tu, 2001).**

Les valeurs de bootstrap supportent les nœuds de chaque embranchement, mais la position des embranchements les uns par rapport aux autres reste non résolue.

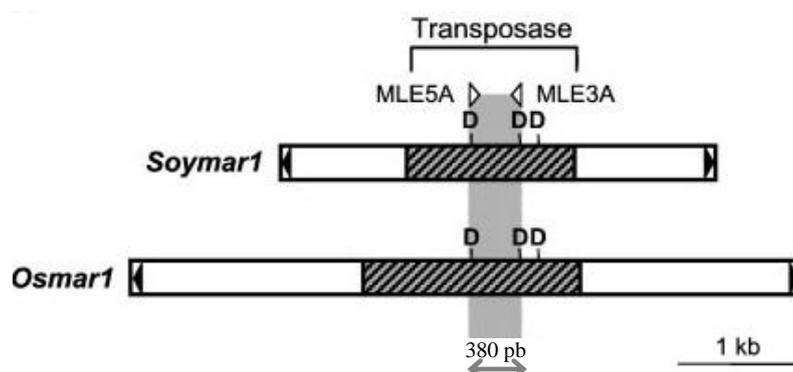
Dans le paragraphe suivant seules les familles les mieux caractérisées *MLE* (tous les éléments apparentés à l'élément *mariner*, incluant les *MLE* de plantes : DD34D et DD39D) et *TLE* seront traitées (soit les éléments DD34E) car les éléments qui font l'objet de ce travail sont des *MLE* et le modèle d'étude appartient au règne végétal.

### III.1 La famille *MLE* (*mariner*-like elements)

Le premier élément *mariner* a été découvert chez la drosophile *Drosophila mauritiana* par l'analyse d'une mutation instable dans l'allèle du gène *white* qui se traduit par un changement de coloration au niveau de l'oeil. Cet élément appelé *Dmmar* (pour *Drosophila mauritiana mariner*) est plus connu sous le nom de *Mos1* (Jacobson et al., 1986). Depuis, les éléments apparentés à l'élément *mariner* ont été appelés *mariner*-like elements (*MLE*). Les *MLE* ont été caractérisés chez un grand nombre d'espèces animales : insectes, nématodes, poissons, mammifères, crustacés (Plasterk et al., 1999; Robertson, 2002; Casse et al., 2006) ou encore chez le protozoaire *Trichomonas vaginalis* (Silva et al., 2005).

Les *MLE* d'origine animale sont répartis en neuf sous-familles désignées en fonction des espèces dans lesquelles les premiers membres ont été découverts : *mauritiana*, *cecropia*, *mellifera*, *capitata*, *irritans* et *elegans*, *briggsae*, *rosa*, *lineata* (Robertson, 1997, 2002). Récemment, une nouvelle sous-famille *marmoratus* a été mise en évidence par Bui et al. (2007, 2008) chez les crabes marins. Cette sous-famille compte un élément *mariner* potentiellement actif *Pacmmar1*, trouvé chez le crabe *Pachygrapsus marmoratus* (Delaurière et al., 2009).

Les études concernant les *MLE* de plantes sont peu nombreuses. La première étude a été menée par Jarvick et Lark en 1998. Les auteurs ont isolé, dans le génome du soja, un élément présentant des similitudes avec *mariner* qu'ils ont nommé *Soymar1* (soybean mariner élément 1). L'alignement des séquences protéiques montre des divergences par rapport aux autres *mariner* isolés chez la drosophile (*Mos1*), chez l'homme (*Himar1*) ainsi que chez des planaires, en effet près de la moitié de la séquence de la transposase diffère. Toutefois l'élément *Soymar1* présente également la triade catalytique DDxD typique des *MLE* et nécessaire à la transposition. Cette triade diffère par rapport aux *MLE* d'animaux puisque la distance x entre les deux derniers acides aspartiques est de 39 acides aminés au lieu de 34 (Jarvick et Lark, 1998). Cette caractéristique est partagée par le deuxième élément issu du végétal. En 2000, Tarchini et al. ont identifié chez le riz, l'éléments *Osmar1* (*Oryza sativa mariner élément 1*). Celui-ci présente 67 % d'homologie avec *Soymar1* et contient la triade catalytique DD39D, ainsi il est classé dans la famille *MLE*. Par la suite, l'étude des bases de données ont révélé deux autres transposons complets de type *MLE* dans le génome du riz *Osmar2* et *Osmar2b* (Feschotte et Wessler, 2002).



**Figure 20 : Définition des amorces *MLE5A* et *MLE3A* par Feschotte et Wessler (2002).**

Les amorces dégénérées *MLE5A* et *MLE3A* sont définies sur la base de deux motifs communs des éléments de type *mariner* *Soymar1* et *Osmar1*. Ces deux motifs correspondent aux deux premiers résidus aspartatés de la triade catalytique des plantes DD39D. Ils sont donc impliqués dans la fonctionnalité de la transposase.

L'alignement de *Soymar1*, *Osmar1* et *Osmar2* a révélé des motifs protéiques conservés utilisables pour définir des amorces PCR afin de rechercher des *MLE* dans les génomes de plantes (Figure 20). Ces amorces ont permis de mettre en évidence l'ubiquité des *MLE* chez les plantes (Feschotte et Wessler, 2002). En effet, ces auteurs ont mis en évidence des gènes codant une transposase de *MLE* chez 36 espèces d'Angiospermes. De la même façon, les *MLE* ont été détectés dans le genre *Beta* et un élément complet et potentiellement actif *Vulmar1* a été isolé chez la betterave (*Beta vulgaris mariner* élément 1) (Jacobs et al., 2004). De même, cinq *MLE* de plantes ont été découverts chez le pois, dont un élément complet potentiellement actif *Psmar1* (*Pisum sativum mariner* élément 1) (Macas et al., 2005). Les 21 *MLE Osmar* complets (dont sept potentiellement actifs) chez le riz (Feschotte et al., 2003) et les neuf éléments détectés *in silico* dans le génome d'*Arabidopsis thaliana*, présentent également une triade catalytique DD39D caractéristique des *MLE* de plantes. Les séquences des TIR de *MLE* de plantes montrent une forte homologie avec les séquences des TIR des MITE *Stowaway*. Ceci suggère que ces MITE peuvent être mobilisés par des transposases de *MLE* fonctionnelles. Les éléments *Stowaway* sont présents en grand nombre de copies dans les génomes de végétaux (22 000 copies dans le génome du riz), les MITE n'étant pas capables de transposer de façon autonome, l'envahissement des génomes suggère que des *MLE* actifs produisant la transposase sont présents dans les génomes végétaux (Feschotte et al., 2003; Yang et al., 2006). Toutefois la transposition d'un *MLE* dans un génome végétal n'a pas été démontrée à ce jour (Bureau et Wessler, 1994; Wessler et al., 1995; Turcotte et al., 2001; Feschotte et al., 2003; Macas et al., 2005). Bien que les MITE *Stowaway* ont été considérés comme des éléments *MLE* fortement délétés, ils ne semblent pas dériver des *MLE* et l'analyse des séquences ne révèle une homologie que sur les séquences TIR. Il semble donc que les éléments *Stowaway* ne dérivent pas des *MLE* mais sont apparus *de novo* (Feschotte et al., 2003; Yang et al., 2006).

Chez les animaux, seuls deux éléments naturels de la famille *MLE* sont actifs et autonomes, *Mos1*, isolé chez la drosophile *Drosophila mauritiana* (Jacobson et al., 1986) et *Famar1* isolé chez le perce-oreille *Forficula auricularia* (Barry et al., 2004). Toutefois les éléments *Himar1* et *Hsmar1* ont été réactivés par reconstruction de la séquence ancestrale et sont actifs (Lampe et al., 1997; Miskey et al., 2007). Actuellement 16 éléments de la famille *MLE* sont décrits comme potentiellement actifs, c'est-à-dire possédant un cadre de lecture ouvert mais dont l'activité n'a pas été démontrée. Ces éléments sont répertoriés dans le tableau 3 ainsi que leurs hôtes respectifs et les références bibliographiques associées.

**Tableau 3 : Les éléments potentiellement actifs appartenant à la famille *MLE*.**

	Élément	Espèce hôte	Référence
Animaux	<i>Dth-2</i>	<i>Dugesia tigrina</i> (vers plat)	Garcia-Fernandez et al., 1995
	<i>Ccmar 1</i>	<i>Ceratitis capitata</i> (mouche)	Gomulski et al., 1997
	<i>Crmar2</i>	<i>Ceratitis rosa</i> (mouche)	Gomulski et al., 2001
	<i>Mommar1</i>	<i>Metaseiulus occidentalis</i> (mite)	Jeyaprkash et Hoy, 1995
	<i>Mcmar1</i>	<i>Meloidogyne chitwoodi</i> (vers ronds)	Leroy et al., 2003
	<i>Tymar1</i>	<i>Trichomonas vaginalis</i> (protozoaire)	Silva et al., 2005
	<i>Mdmar1</i>	<i>Musca domestica</i> (mouche)	Liu, Pridgeon, et al., 2004
	<i>Bgmar1</i>	<i>Blatella germanica</i> (blatte)	Liu, Pridgeon, et al., 2004
	<i>Bytmar1</i>	<i>Bythograea therymydron</i> (crabe)	Halaimia-Toumi et al., 2004
	<i>Hvmar1</i>	<i>Heliothis virescens</i> (papillon)	Ren et al., 2006
	<i>Mboumar1</i>	<i>Messor bouvieri</i> (fourmi)	Palomeque et al., 2006
	<i>Pacmmar1</i>	<i>Pachygrapsus marmoratus</i> (crabe)	Delaurière et al., 2009
Végétaux	<i>Soymar1</i>	<i>Glycine max</i> (soja)	Jarvick et Lark, 1998
	<i>Osmar1</i>	<i>Oryza sativa</i> (riz)	Tarchini et al., 2000
	<i>Psmar1</i>	<i>Pisum sativum</i> (pois)	Macas et al., 2005
	<i>Vulmar1</i>	<i>Beta vulgaris</i> (betterave)	Jacobs et al., 2004
	<i>Osmar1</i> <i>Osmar5A</i> <i>Osmar9A</i> <i>Osmar17A</i> <i>Osmar19</i>	<i>Oriza sativa s.esp japonica</i> (riz)	Feschotte et al., 2003
	<i>Osmar5Bi</i> <i>Osmar15Bi</i>	<i>Oriza sativa s.esp indica</i> (riz)	Feschotte et al., 2003

Les éléments *Mos1* et *Himar1* sont très étudiés d'un point de vue biochimique et servent de modèle d'étude du mécanisme de transposition des éléments *mariner* (Lampe et al., 1997, 1998; Dawson et Finnegan, 2003; Augé-Gouillou et al., 2001a, 2001b; Brillet et al., 2007).

### III.2 La famille *TLE* (*Tc1*- like elements)

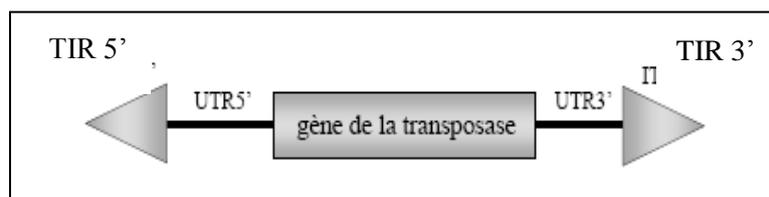
L'élément *Tc1* a été découvert chez *Caenorhabditis elegans* par observation d'un polymorphisme d'insertion entre deux souches *Bristol N2* et *Bergerac NO* (Emmons et al., 1979 et 1983). Les *TLE* ont ensuite été détectés dans un large spectre d'hôtes : insectes, mammifères, nématodes, poissons, amphibiens et champignons (Henikoff, 1992; Plasterk, 1999; Göttgens et al., 1999; Leaver, 2001; Hua-Van et al., 2000; Pocwierz-Kotuset al., 2007). La forte variabilité des *TLE* ne permet pas de mettre en exergue leurs relations phylogénétiques, ni de les classer en sous-familles (Sinzelle et al., 2005). De nombreux éléments naturels, actifs et autonomes ont été décrits, par exemple : l'élément *Tc1* (Collins et al., 1987), mais aussi les éléments *Impala* et *Fot1* isolés chez le champignon *Fusarium oxysporum* (Langin et al., 1995; Daboussi et al., 1992). De plus, des *TLE* actifs, *Sleeping Beauty* et *Frog Prince*, ont été obtenus par reconstruction d'une séquence consensus établie à partir de l'alignement de différentes copies isolées chez des salmonidés (Ivics et al., 1997) et

chez la grenouille *Rana pipiens* (Miskey et al., 2003). Les éléments *Minos* et *Sleeping Beauty* font l'objet de nombreux travaux en vue de leurs utilisations comme vecteur de transfert de gène (Miskey et al., 2005; Matès et al., 2007, Pavlopoulos et Averof, 2005).

### III.3 Caractéristiques structurales des *MLE* et des *TLE*

Les *MLE* ont une longueur allant de 1,3 Kpb chez les animaux jusqu'à 7,072 Kpb chez les plantes, et les *TLE* présentent des longueurs de 1,5 à 4,7 Kpb. Ils sont composés de TIR, de séquences non traduites ou UTRs (Untranslated Region) et du gène codant la transposase (Figure 21).

Le gène de la transposase a une structure simple avec un cadre de lecture unique. Ce gène ne possède pas d'introns chez les *MLE* issus d'animaux (Robertson, 2002) mais peuvent en contenir chez les *MLE* isolés chez les plantes (Feschotte et Wessler 2002; Feschotte et al., 2003). Chez les *TLE*, trois éléments possèdent des introns au sein de leur séquence : *Tc1*, *Tc3* et *Minos* (Vos et al., 1993; Colloms et al., 1994; Franz et al., 1994).



**Figure 21 : Représentation schématique des transposons de la famille des *MLE*.**

Le gène de la transposase est encadré par des séquences TIR (Terminale Inverted Repeat) et UTR (Untranslated Region).

- Les TIR des *MLE* ont une taille d'environ 20 à 40 pb, que ce soit chez les animaux ou les végétaux (Jarvick et Lark, 1998; Brillet et al., 2007; Jacobs et al., 2004; Macas et al., 2005; Feschotte et al., 2003). Les TIR des *MLE* possèdent un motif palindromique où chacune des deux parties du palindrome posséderait un site de liaison pour la transposase et un motif miroir dont le rôle serait de stabiliser les interactions protéines-TIR (Bigot et al., 2005). Les TIR des *TLE* sont hétérogènes en taille (de 20 à 700 pb) et en structure. Ils peuvent être constitués de plusieurs sites de liaison à la transposase (Cui et al., 2002; Izsvack et al., 2002) et présenter des répétitions directes, notamment chez les *TLE* d'amphibiens et de poissons (Ivics et al., 1996, 1997).

- Les UTR sont des régions non traduites situées en 5' et 3' du cadre de lecture codant la transposase. Ces régions peuvent atteindre plusieurs milliers de nucléotides et sont la principale cause de la variation de taille observée au sein des *TLE* (Sinzelle et al., 2005). Leur rôle est peu connu. Ils pourraient posséder les informations nécessaires à la transcription et à

la traduction de la transposase. Ainsi, ils pourraient permettre la fixation des facteurs de transcription (Palomeque et al., 2006; Leroy et al., 2003). Les TIR seuls sont suffisants pour que la transposase effectue la transposition (Augé-Gouillou et al., 2001a; Feschotte et al., 2003), cependant la présence d'au moins une partie des UTR permettrait une transposition plus efficace (Tosi et Beverley, 2000; Miskey et al., 2007).

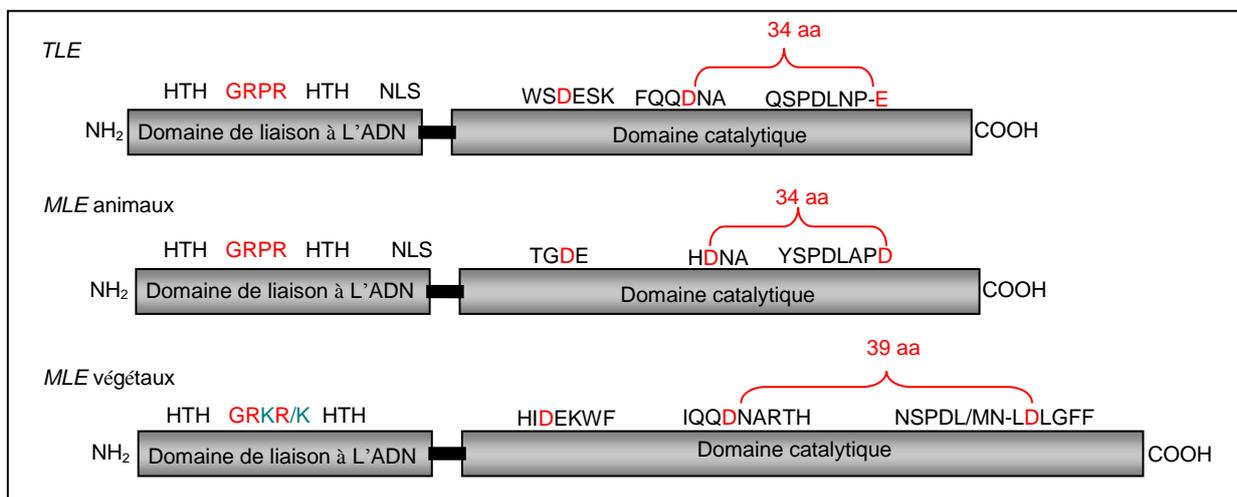
Les transposases ont une taille comprise entre 330 et 360 acides aminés chez les animaux et de 412 à 520 acides aminés chez les végétaux. Elles sont constituées de deux domaines fonctionnels : un domaine de liaison à l'ADN situé dans la région N-terminale et un domaine catalytique situé en région C-terminale. Les transposases possèdent un ou plusieurs signaux de localisation nucléaire (NLS) putatifs.

- Le domaine de liaison à l'ADN est situé dans les 150 premiers acides aminés de la transposase chez les *MLE* et les *TLE*. Il est constitué de deux motifs hélice-tour-hélice (HTH) séparés par un motif GRPR (Vos et plasterk, 1994). Ces domaines ont été déterminés par des analyses de prédiction de structure secondaire (Pietrokovski et Henikoff, 1997), et de cristallographie (Richardson et al., 2004) sur les *MLE* d'origine animale. La région N-terminale contient un ou plusieurs signaux de localisation nucléaire (NLS) permettant le transport de la transposase dans le noyau cellulaire. Pour la plupart des éléments, ce NLS a été défini *in silico* par analyse bioinformatique et n'est que putatif. En ce qui concerne les *MLE* de plantes, seuls les travaux de Yang et al. (2006) sur la transposition de *Osmar5* en système hétérologue ont permis de définir le domaine de liaison à l'ADN et les domaines hélice-tour-hélice ainsi que le mécanisme de transposition des *MLE* de plantes.

- Le domaine catalytique des transposases de type *MLE* est caractérisé par la triade catalytique DDxD. Le dernier D semble essentiel à l'activité de la transposase *Mos1*, son remplacement par un acide glutamique E (DD34E au lieu de DD34D) entraîne une perte de la fonctionnalité protéique (Lohe et al., 1997). Cependant dans le cas de *Hsmar1* le remplacement de l'acide aspartique (D) par l'asparagine (N) ne modifie pas l'activité *in vitro* de la transposase (Liu, et al., 2007). Pour assurer son rôle, la triade fixe un à deux cations divalents comme le magnésium ou le manganèse (Mizuuchi, 1997; Richardson et al., 2009). Les trois acides aspartiques sont inscrits dans les motifs conservés TGDE, HDNA et YSPDLAPxD des éléments isolés chez les animaux. Ce dernier motif a été décrit comme l'un des deux motifs signatures des *MLE*, avec le motif WVPHEL (Robertson, 1993), bien qu'aucun rôle ne lui ait encore été attribué. Les éléments isolés chez les végétaux présentent une triade catalytique légèrement plus longue. En effet, 39 acides aminés séparent les deux derniers acides aspartiques (Jarvick et Lark, 1998; Tarchini et al., 2000; Jacobs et al., 2004;

Macas et al., 2005). Les motifs conservés ont été identifiés comme étant **HIDEKWF** et **IQQDNARTH**. Des amorces *MLE5A* et *MLE3A* ont été définies à partir de ces motifs (Figure 22) (Feschotte et Wessler, 2002). Celles-ci ont permis d'amplifier des fragments de *MLE* d'environ 400 pb chez un grand nombre de végétaux (Feschotte et Wessler, 2002; Jacobs et al., 2004; Macas et al., 2005). Le troisième résidu aspartate de la triade catalytique semble apparaître dans un motif **NSPDL/MNxLDLGFF** (Figure 22). Toutefois, l'étude de la transposition de *Osmar5* en système hétérologue montre qu'une mutation du deuxième résidu aspartate de la triade catalytique induit une nette baisse de la transposition n'allant pas jusqu'à annuler totalement l'activité de la transposase (Yang et al., 2006). Ainsi les auteurs suggèrent que le motif **DD39D**, défini par la conservation des motifs, pourrait ne pas refléter le véritable centre réactionnel de la transposition des *MLE* chez les plantes. Ainsi, les motifs de la triade catalytique restent à définir par l'étude de la transposition d'autres *MLE* de plantes.

Le domaine catalytique des transposons *TLE* est défini par la triade **DD34E**, à l'instar des *MLE* les résidus aspartiques et l'acide glutamique sont insérés dans des motifs **WSDESK**, et **FQQDNA** pour les deux acides aspartiques de la triade et **QSPDLNPxE** pour l'acide glutamique. Des amorces définies sur la base de ces motifs ont permis de caractériser de multiples éléments *TLE* potentiellement actifs dans le génome du xénope (Sinzelle et al., 2005).

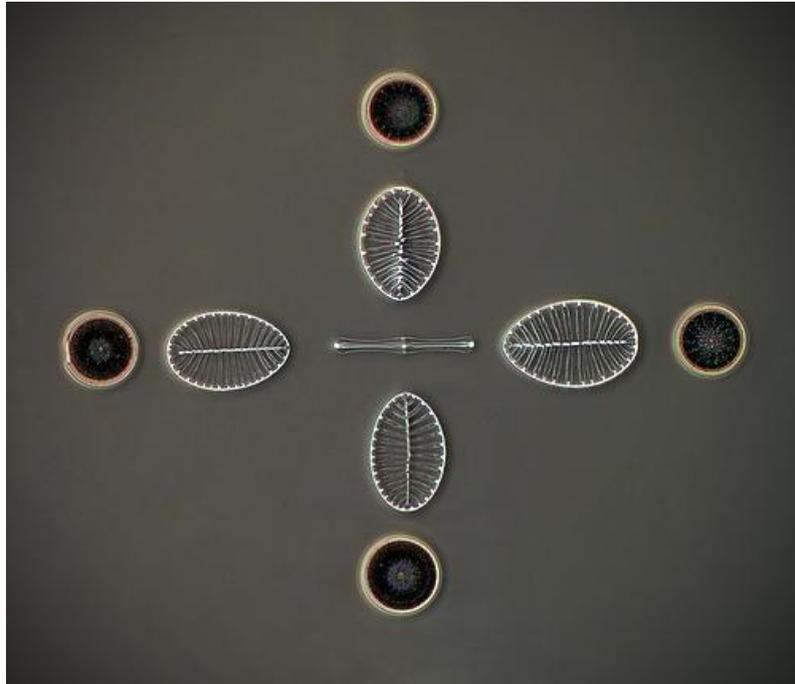


**Figure 22 : Représentation simplifiée de la structure des transposases des *TLE*, *MLE* animaux et des *MLE* végétaux.**

Les motifs HTH représentent les domaines hélice-tour-hélice du domaine de liaison à l'ADN, ils sont séparés par un motif commençant par GRxR conservé au sein de la superfamille et son suivi du signal de localisation nucléaire (NLS). Au sein du domaine catalytique sont représentés les motifs de la triade DDxD/E, les résidus aspartiques et glutamiques de la triade apparaissent en rouge. La structure des éléments *MLE* de végétaux a été définie par l'alignement manuel de *Soyamar*, *Vulmar1*, *Psmar1* et *Atmar1* en comparaison avec l'alignement des *MLE* du riz *Osmar* (Feschotte et al., 2003).

# CHAPITRE II

## MATERIELS ET METHODES



# I Matériel végétal et conditions de culture

## I.1 Origine géographique des microalgues

La majorité des microalgues sur lesquelles porte cette étude ont été collectées dans la baie de Bourgneuf (côte atlantique française) au climat tempéré. Les microalgues comprenant les diatomées ont été isolées par les collègues de Nantes et sont conservées en culture dans l'algothèque du Laboratoire Mer Molécules Santé (MMS) de Nantes et du Mans. Deux autres espèces diatomées proviennent de la côte sud-Est du Vietnam au climat tropical. Elles ont été collectées sur le site de Can Gio situé dans une mangrove, et ont été isolées par T.L.N. Nguyen-Deroche.

## I.2 Les microalgues

Une collection de 18 microalgues est entretenue et conservée au laboratoire dont 15 espèces de microalgues appartiennent à la classe des Bacillariophycées (ou diatomées). Parmi les 15 espèces de diatomées, 11 d'entre elles sont des diatomées pennales, les quatre autres espèces sont des centrales. Deux espèces de microalgues dites «vertes» appartenant respectivement aux classes des Prymnésiophycées et Pavlovophycées et une espèce de microalgue rouge (Rhodophycée) ont été ajoutées à cet échantillon. Toutes les microalgues ont été utilisées dans un premier temps pour la recherche des transposons. La liste des espèces étudiées, leur origine, leur intérêt écologique ou/et économique sont présentés dans le Tableau 4.

**Tableau 4 : Liste des espèces utilisées, leur taxonomie, leur origine géographique et leur intérêt économique ou écologique.**

Classe	Ordre	Espèce	Origine	Intérêt
Bacillariophycées (diatomées)	Pennales	<i>Amphora acutiuscula</i>	Côte vietnamienne	Espèce tropicale
	Pennales	<i>Amphora coffeaeformis</i>	Côte atlantique française	Espèce tempérée
	Pennales	<i>Cylindrotheca fusiformis</i>	Côte atlantique française	Peu silicifiée, frustule fin
	Pennales	<i>Cylindrotheca closterium</i>	Côte atlantique française	Peu silicifiée, frustule fin
	Pennales	<i>Entomoneis paludosa</i>	Côte atlantique française	Espèce cosmopolite, benthique
	Pennales	<i>Nitzschia laevis</i>	Côte atlantique française	Aquaculture
	Pennales	<i>Nitzschia palea</i>	Côte vietnamienne	Espèce tropicale

<i>Pennales</i>	<i>Phaeodactyllum tricornutum</i>	Côte atlantique française	Génome séquencé, benthique ou planctonique
<i>Pennales</i>	<i>Stauroneis constricta</i>	Côte atlantique française	Espèce cosmopolite
<i>Pennales</i>	<i>Haslea crucigera</i>	Côte atlantique française	Absence de marennine
<i>Pennales</i>	<i>Haslea ostrearia</i>	Côte atlantique française	Marennine responsable du verdissement des huîtres
<i>Centrales</i>	<i>Thalassiosira pseudonana</i>	Long island	Génome séquencé, planctonique
<i>Centrales</i>	<i>Thalassiosira weissflogii</i>	Long island	Planctonique
<i>Centrales</i>	<i>Odontella aurita</i>	Côte atlantique française	Produit des omega 3 Frustule avec épines
<i>Centrales</i>	<i>Skeletonema costatum</i>	Côte atlantique française	Aquaculture
<i>Rhodophycée</i>	<i>Porphyridium cruentum</i>	Côte atlantique française	Propriétés antioxydantes
<i>Prymnésiophycée</i>	<i>Isochrysis galbana</i>	Port Erin, Isle of Man, Irish Sea	Aquaculture; lipides
<i>Pavlovophycée</i>	<i>Pavlova lutheri</i>	Côte atlantique française	Aquaculture; lipides

### I.3 Conditions de culture de référence

Les microalgues sont cultivées dans une chambre de culture en conditions contrôlées. Toutes les cultures sont monoclonales et axéniques (c'est-à-dire dépourvues de micro-organismes contaminants) (Annexe 1). Les microalgues sont cultivées dans des fioles erlenmeyer renfermant de l'eau de mer artificielle (EMA) qui ne sont ni aérées, ni agitées (culture en Batch). L'EMA est préparée selon le protocole de Harrison et al. (1980) modifié par l'ajout de micro-éléments et vitamines avant autoclavage (Annexe 2). Toutes les espèces sont cultivées dans ce milieu à l'exception des deux espèces du genre *Amphora* qui sont cultivées dans un milieu EMA à teneur réduite en calcium (0,25 mM CaCl<sub>2</sub> au lieu de 9 mM) permettant de limiter l'adhésion des cellules entre elles et aux parois de l'erlenmeyer (Cooskey et al., 1981; Molino et al., 2008).

Les microalgues du littoral atlantique français adaptées à un milieu tempéré sont cultivées à une température de 16 ± 1 °C tandis que les espèces d'origine vietnamienne sont cultivées à une température de 24 ± 1 °C.

Les cultures sont réalisées sous un éclairage de  $100 \mu\text{mol}$  de photons  $\text{m}^{-2} \text{s}^{-1}$  avec une photopériode de 14 h. La mesure du niveau d'éclairage est effectuée avec une sonde quantamétrique  $4\pi$  (Walz, Allemagne) reliée à un quantamètre (Li-cor 189) et immergée dans le milieu de culture.

Le repiquage des souches est effectué tous les 10 jours en inoculant 5 mL de la culture mère dans 120 mL de milieu EMA neuf. Les algues sont ainsi maintenues en phase exponentielle de croissance.

## I.4 Stress thermiques

### I.4.1 Conditions de stress thermiques

Les stress thermiques ont été appliqués à trois diatomées : *P. tricornutum*, *A. coffeaeformis* usuellement cultivée à  $16^\circ\text{C}$  et *A. acutiuscula* cultivée à  $24^\circ\text{C}$ .

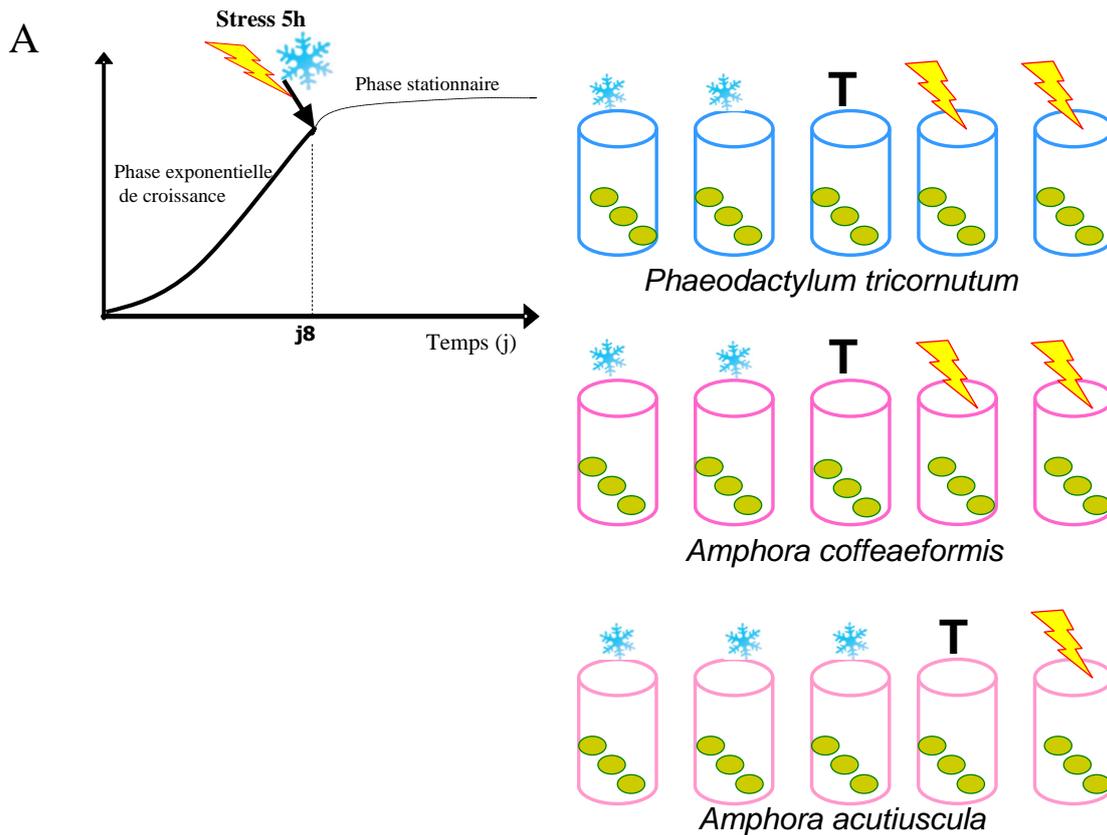
### I.4.2 Détermination des températures de stress thermiques

Après 8 jours de croissance en conditions standards, les trois espèces ont été mises en culture dans 125 mL d'EMA et placées pendant 10 jours à 8, 16 ou  $24^\circ\text{C}$ . Ces cultures ont servi au suivi de la cinétique de croissance permettant de déterminer les températures de stress.

### I.4.3 Application de stress thermiques de courte durée

A partir de ces cultures axéniques, pour chacune des espèces, cinq nouvelles cultures ont été réalisées dans 125 mL d'EMA, chacune avec une densité initiale de  $5.10^4$  cellules par mL. Pour accélérer la croissance, les cultures ont été placées sous un éclairage de  $300 \mu\text{mol}$  photons  $\text{m}^{-2} \text{s}^{-1}$  et une photopériode de 14h pendant 8 jours à la température habituelle de culture (témoin) :  $16^\circ\text{C}$  pour *P. tricornutum* et *A. coffeaeformis* et  $24^\circ\text{C}$  pour *A. acutiuscula*. Une vérification du niveau d'éclairage a été réalisée au début de chaque série d'essais. Les cultures n'ont été ni agitées ni aérées pour éviter toute contamination.

Au 8<sup>ième</sup> jour, soit en fin en phase exponentielle de croissance, des stress de courte durée 5h ont été appliqués aux trois espèces de diatomées. Les cultures ont été placées pendant 5h dans des bains-cryothermostatés (Figure 23). Les conditions de stress thermiques sont 4, 8, 24 et  $32^\circ\text{C}$  pour *P. tricornutum* et *A. coffeaeformis*,  $16^\circ\text{C}$  étant la condition témoin. Pour *A. acutiuscula*, originaire du Vietnam, la température de culture est  $24^\circ\text{C}$  et les conditions de stress sont 4, 8, 16 et  $32^\circ\text{C}$ .



**Figure 23 : Dispositif expérimental utilisé pour appliquer les stress thermiques aux trois diatomées *Phaeodactylum tricornutum*, *Amphora coffeaeformis* et *Amphora acutiuscula*.**

A : Schéma du protocole expérimental utilisé pour l'application des stress thermiques aux trois diatomées. Les flocons bleus symbolisent les stress basse température et les éclairs jaunes symbolisent les stress température élevée par rapport à la température de culture noté T.

B : Dispositif expérimental mis en place pour l'application des stress thermiques.

Ici les cultures de *Phaeodactylum tricornutum* sont soumises aux différents stress thermiques (4, 8, 16, 24 et 32 °C). Au premier plan, la culture de *P. tricornutum* soumise à une température de 8 °C pendant 5 h par immersion dans un bain thermostaté.

## II Paramètres physiologiques

### II.1 La croissance

#### II.1.1 Mesure de la densité cellulaire

Pour estimer la croissance des diatomées, les cellules ont été dénombrées à l'aide d'un hématimètre Neubauer. Avant chaque comptage, les cultures de diatomées ont été homogénéisées pendant 10 min à l'aide d'un agitateur magnétique. Les cellules d'*A. coffeaeformis* ont tendance à s'agglomérer et à adhérer aux parois des erlenmeyers (Molino et al., 2008) rendant le dénombrement difficile (Cooksey, 1981). Pour dissocier les cellules les unes des autres (rupture des liaisons hydrogènes au niveau des polysaccharides externes), 2 mL de culture homogène ont été mélangés avec une goutte de Tween80. Afin de différencier les cellules vivantes des cellules mortes, 40 µL de rouge neutre à 0,01 % (m/v) ont été ajoutés au 2 mL de culture de microalgues, le mélange a été laissé à température ambiante durant 15 min avant comptage (Crippen et Perrier, 1974). La densité cellulaire est calculée selon la formule suivante :

Densité cellulaire (Nb de cellules / mL) = Nombre moyen de cellules comptés x 10 000  
Toutes les densités cellulaires sont la moyenne de trois comptages différents, dans la zone quadrillée de volume 0,1 mm<sup>3</sup>.

#### II.1.2 Courbes de croissance

Les courbes de croissance des diatomées ont été établies en fonction du temps (jours). L'évolution de la densité cellulaire en fonction du temps lors d'une culture en « batch » se traduit par une courbe en forme de S (sigmoïdale) (Figure 24). La croissance se déroule en trois phases :

- une phase de latence
- une phase d'accélération ou phase exponentielle de croissance.
- une phase plateau où la croissance se stabilise après avoir atteint une densité maximale K, Cette dernière phase a lieu lorsque la quantité de nutriments disponible diminue. Le suivi des cultures a été réalisé pendant 8 à 10 jours jusqu'à obtention de la phase plateau.

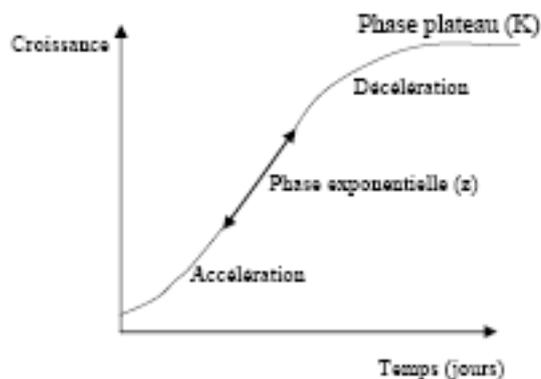


Figure 24 : Courbe de croissance théorique des diatomées cultivées en batch en fonction du temps.

## II.2 Les paramètres photosynthétiques

L'activité photosynthétique a été mesurée par oxymétrie (dégagement de dioxygène) et par fluorimétrie modulée (fluorescence de la Chl *a*) (Hancke et al., 2008) qui sont des techniques faciles à mettre en oeuvre et couramment utilisées au laboratoire (Maxwell et Johnson, 2000; Figueroa et al., 2003; Rech et al., 2003; Rech, 2004; Hancke et al., 2008).

### II.2.1 Extraction et dosage de la Chlorophylle *a*

Le dosage de la chlorophylle permettra ultérieurement de normaliser les mesures d'intensité photosynthétiques. Le dosage de la Chl *a* est effectué selon la méthode de Speziale et al. (1984). Suite à l'application des stress, un volume de 25 mL de culture a été prélevé pour extraire les pigments chlorophylliens. Afin de déterminer la teneur en Chl *a*, ce volume est filtré sous vide partiel sur un filtre GFC (Watman), retenant la biomasse qui peut être conservé à -20 °C. Le filtre est immergé dans 3 mL de diméthyl-formamide (DMF) dans une fiole en verre opaque permettant d'éviter la photolyse des pigments. L'ensemble est homogénéisé à l'aide d'un agitateur rotatif à température ambiante durant 1 h. Une centrifugation de 5 min à 4 °c et 10 000 g permet d'éliminer les débris cellulaires. La densité optique (DO) du surnageant aux longueurs d'onde de 630 nm, 664 nm et 750 nm est mesurée. La concentrations en Chl *a* ( $\mu\text{g mL}^{-1}$ ) est calculée comme suit :

$$\text{Chl } a = \frac{[11,47(DO_{664} - DO_{750}) - 0,4(DO_{630} - DO_{750})] \times Vs}{Vc}$$

DO<sub>630</sub>, DO<sub>664</sub> et DO<sub>750</sub> sont les densités optiques mesurées respectivement aux longueurs d'onde de 630 nm, 664 nm et 750 nm. Vs est le volume de DMF utilisé pour l'extraction (ici 3 mL) et Vc le volume de suspension d'algues utilisé (20 mL).

### II.2.2 Mesure de l'intensité photosynthétique

Afin de déterminer si les stress appliqués ont eu un impact sur l'activité photosynthétique des diatomées, des mesures de l'intensité de la photosynthèse ont été réalisées. Pour cela, il est nécessaire de déterminer la quantité de dioxygène (O<sub>2</sub>) libérée lors de la photosynthèse. L'émission de l'O<sub>2</sub> a été mesurée à l'aide d'un oxymètre composé d'un boîtier (Hansatech LTD, UK) de contrôle couplé à une électrode de Clark (Delieu et Walker, 1972) ou électrode à oxygène. L'électrode à oxygène est composée d'une anode en argent et d'une cathode en platine, recouvertes d'une membrane perméable à l'oxygène. Un volume de 1,5 mL de suspension de diatomées prélevé dans la culture est placé dans la cellule de mesure. Cette enceinte de mesure est partiellement fermée dans sa partie supérieure par un bouchon traversé par un capillaire réduisant au maximum le contact avec O<sub>2</sub> atmosphérique. La suspension est agitée par un micro barreau aimanté pour maintenir la teneur en O<sub>2</sub> homogène dans la cellule de mesure. Les diatomées sont maintenues à température constante (température de culture ou de stress) par circulation d'eau thermostatée dans la double paroi de la cellule de mesure. L'électrode à oxygène est connectée à un microgalvanomètre qui affiche l'intensité du courant électrique en mV, proportionnelle à la quantité d'O<sub>2</sub> présente dans la cellule de mesure (1 mg O<sub>2</sub> L<sup>-1</sup> = 100 mV à 25 °C). Avant chaque mesure, l'oxymètre a été calibré à la température à laquelle les diatomées ont été stressées. L'évolution de la teneur en O<sub>2</sub> dans la cellule de mesure est enregistrée en fonction du temps. La pente de la courbe qui en résulte permet de déterminer l'activité photosynthétique AP (exprimée en mg O<sub>2</sub> L<sup>-1</sup> min<sup>-1</sup>). L'intensité photosynthétique (IP) est ensuite calculée à l'aide de la relation ci-dessous :

$$IP = \frac{AP \times 60 \times 10^3}{32 \times [Chl a]}$$

IP, intensité photosynthétique en  $\mu\text{mol O}_2 \text{ h}^{-1} \text{ mg}^{-1} \text{ Chl } a$ ; AP, activité photosynthétique en  $\text{mg O}_2 \text{ L}^{-1} \text{ min}^{-1}$ ; [Chl a], teneur en Chl a ( $\text{mg L}^{-1}$ ); 32 est la masse molaire du O<sub>2</sub>.

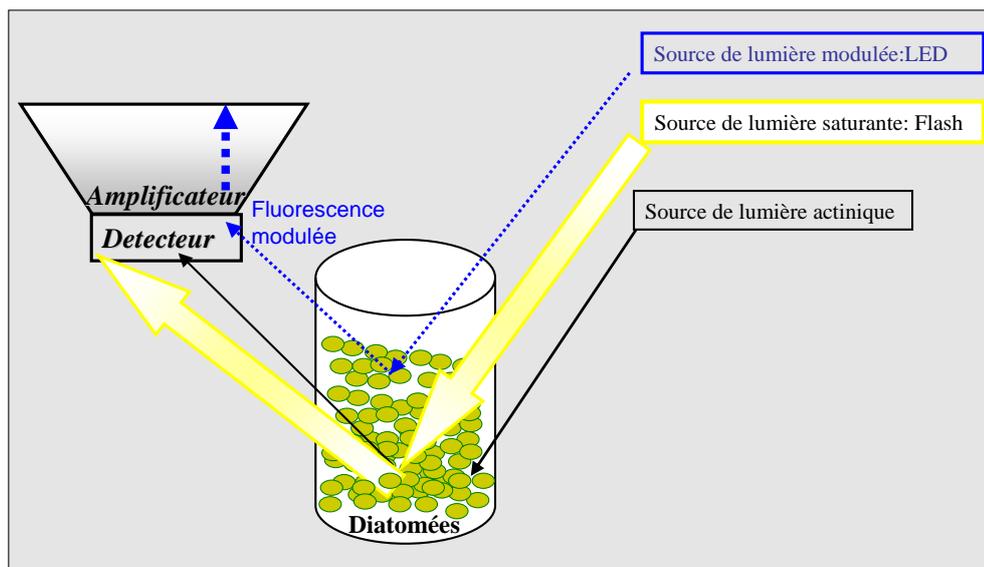
### II.2.3 Mesure de la fluorescence de la chlorophylle a

Afin d'étudier la réponse de l'appareil photosynthétique dans les différentes conditions thermiques, les paramètres de la fluorescence chlorophyllienne ont été mesurés par fluorimétrie modulée. Les paramètres de fluorescence sont mesurés *in vivo* directement sur la suspension de microalgues et sont le reflet inversé des événements photochimiques qui ont lieu au niveau du photosystème II (PSII). Le principe de la fluorimétrie modulée est basé sur

l'utilisation de plusieurs sources lumineuses permettant de détecter des variations de la fluorescence (Figure 25) :

- une source de lumière actinique blanche dont l'intensité est variable, est utilisée pour induire la photosynthèse.
- une source lumineuse de faible puissance (n'induisant pas la photosynthèse) émise sous forme d'un signal modulé avec une certaine fréquence, est utilisée pour mesurer la fluorescence. Cette lumière modulée ou lumière analytique est produite par des diodes électroluminescentes bleues (LED). Elle est couplée à un récepteur permettant de mesurer et d'enregistrer les variations du signal de fluorescence émis par les diatomées. Dans ces conditions, la lumière actinique induisant la photosynthèse n'influence pas la détection de la fluorescence.

Le fluorimètre FMS-1 (Hansatech) a été modifié par Rech et al. (2003) et utilisé avec succès pour effectuer des mesures avec des suspensions de microalgues marines peu concentrées. Il est couplé à un ordinateur permettant d'enregistrer les valeurs et de calculer les rapports correspondant aux paramètres de fluorescence.



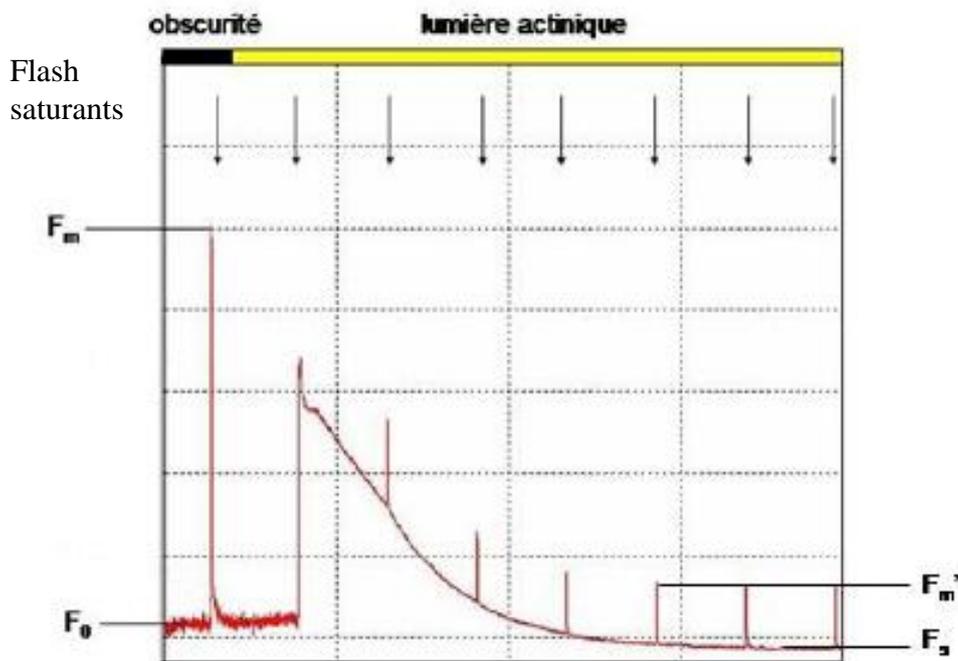
**Figure 25 : Schéma de la technique de fluorimétrie modulée pour mesurer la fluorescence chlorophyllienne d'une suspension de microalgues.**

La lumière modulée est émise par une LED, elle induit la fluorescence de la Chl<sub>a</sub> modulée qui seule, est amplifiée et mesurée par le système couplé au détecteur. La lumière actinique active la photosynthèse et le flash saturant permet de fermer tous les centres réactionnels du photosystème II et de mesurer la fluorescence maximale.

Les suspensions d'algues sont placées dans une cuve de spectrophotomètre sous agitation à l'aide d'un mini-barreau aimanté dans une cellule de mesure thermo-régulée au croisement des sources lumineuses.

Avant chaque mesure, les diatomées sont placées 10 min à l'obscurité. Pendant cette période d'adaptation tous les centres réactionnels du PSII sont ouverts. L'émission de lumière modulée va permettre de mesurer la-fluorescence minimale  $F_0$  (Figure 26). Un flash saturant de lumière blanche est alors envoyé conduisant à la fermeture des centres réactionnels du PSII qui deviennent alors inactifs. La fluorescence maximale  $F_m$  peut alors être enregistrée. A partir de ces deux paramètres, la fluorescence variable,  $F_v$  qui correspond à la fraction de l'énergie photonique absorbée convertible en énergie chimique, est calculée selon l'équation :  $F_v = F_m - F_0$ . Le calcul du rapport  $F_v/F_m$  correspond alors à une mesure du rendement quantique maximum des réactions photochimiques au niveau du PSII.

Lorsque l'échantillon est adapté à la lumière actinique, la fluorescence de base peut être mesurée  $F_s$  (Figure 26). L'émission d'un flash saturant, identique à celui utilisé pour la mesure de  $F_m$  permet d'obtenir  $F_m'$  la fluorescence maximale à la lumière, après une période d'adaptation à la lumière actinique. Il est alors possible de calculer le paramètre  $\phi_{PSII}$ , ou rendement quantique effectif du PSII d'après l'équation :  $\phi_{PSII} = (F_m' - F_s) / F_m'$ .



**Figure 26 : Evolution de la fluorescence chlorophyllienne d'une suspension de microalgues.**

La fluorescence minimale  $F_0$  est mesurée à l'obscurité. La fluorescence maximale  $F_m$  est ensuite obtenue après émission de flashes de lumière saturante.

A la lumière, la fluorescence minimale  $F_s$  est mesurée et la fluorescence maximale  $F_m'$  à la suite d'un flash saturant. Le système couplé à un ordinateur permet de calculer en temps réel le rapport  $F_v/F_m$  et le  $\phi_{PSII}$ . La fluorescence est exprimée en unités arbitraires.

### II.3 Récolte de la biomasse algale

Après 5 h de stress thermique, et prélèvement pour les mesures des paramètres physiologiques, les algues sont récoltées par centrifugation à 600 g. Pour éviter l'éclatement

des cellules, la centrifugation est réalisée avec une accélération et une décélération douce à la température du traitement thermique. La biomasse est récupérée dans des tubes Falcon de 50 mL et stockée dans au congélateur à -80 °C, jusqu'à l'extraction de l'ADN génomique ou des ARN totaux.

## III Techniques de biologie moléculaire

### III.1 Extraction des acides nucléiques

#### III.1.1 Extraction de l'ADN génomique

L'ADN génomique des diatomées a été extrait selon le protocole modifié de Doyle et Doyle (1990) à partir d'une culture axénique de 200 mL en phase exponentielle de croissance. Les cellules ont été collectées par centrifugation avec accélération et décélération lente à 4 °C pendant 10 min et congelées à -80 °C si besoin. Le culot cellulaire est broyé dans un mortier avec de l'azote liquide. Après broyage l'extrait est récupéré dans 6 mL de tampon d'extraction au CTAB ( $\beta$ - mercaptoéthanol 0,2 %; CTAB 2 %; EDTA 20 mM, pH 8; NaCl 1,4 M et Tris 100 mM, pH 8) préalablement chauffé à 60 °C. L'ensemble est incubé sous agitation régulière pendant 30 min à 60 °C. Après avoir ajouté 6 mL de chloroforme, le mélange est homogénéisé doucement par retournement durant 5 min puis centrifugé pendant 20 min à 5 500 g (4 °C). Enfin 4 mL d'isopropanol sont ajoutés à la phase aqueuse afin de précipiter les acides nucléiques. Après une incubation de 2 h à -20 °C, le mélange est centrifugé pendant 15 min à 5 500 g et le culot d'ADN subit un lavage à l'éthanol 70 %. Après centrifugation, le culot d'ADN est séché sous vide et resuspendu dans 80  $\mu$ l de tampon TE (EDTA 1 mM, pH 8 et Tris-HCl 10 mM, pH 8). Les ADN génomiques sont conservés à -20 °C.

#### III.1.2 Extraction des ARN totaux

L'extraction des ARN totaux a été effectuée sur les culots cellulaires des cultures de *P. tricornutum*, *A. coffeaeformis* et *A. acutiuscula* traitées par choc thermique pendant 5 h à 4, 8, 16, 24 et 32 °C.

L'ensemble du matériel utilisé est exempt de ribonucléase (RNase) suite à un lavage à l'éthanol 70% et un bain de H<sub>2</sub>O<sub>2</sub> à 3 % durant 15 min suivi de deux rinçages à l'eau DEPC 0,1% (Diéthyl Pyrocarbonate). Les culots sont broyés dans un mortier en présence d'azote liquide pour éviter une réactivation des enzymes dégradant les acides nucléiques et

principalement les ARN. Le broyat est divisé à raison d'environ 50 mg par tubes et conservé dans l'azote liquide puis stockés à  $-80\text{ }^{\circ}\text{C}$  pour de futures extractions.

1 mL de Tri Reagent ainsi que 50 $\mu\text{L}$  de bromoanisole (1-bromo-4-methoxybenzene) sont ajoutés au 50 mg de broyat. La solution est homogénéisée à la pipette puis mélangée au vortex et incubée sur glace pendant 15 min. Cette étape permet une dissolution des cellules et la libération des acides nucléiques ainsi qu'une dénaturation des ADN en pH acide. Une centrifugation de 15 min à  $4\text{ }^{\circ}\text{C}$  et 12 000 g permet de séparer les débris cellulaires des ARN restant en phase aqueuse. Afin de précipiter les ARN, 1 mL d'isopropanol est ajouté à la phase aqueuse et l'ensemble est incubé à température ambiante durant 10 min et une centrifugation à  $4\text{ }^{\circ}\text{C}$  et 12 000 g durant 5 min permet d'obtenir un culot d'acides nucléiques. Toute trace d'isopropanol est éliminée par un lavage à l'éthanol 70. Ce dernier est éliminé à la suite d'une centrifugation de 5 min à  $4\text{ }^{\circ}\text{C}$  et 6 000 g. Le culot d'ARN est séché à l'air libre et enfin resuspendu dans 30  $\mu\text{L}$  d'eau DEPC. Les ARN totaux sont ensuite conservés à  $-80\text{ }^{\circ}\text{C}$  pour toute analyse ultérieure.

### III.2 Traitement à la DNase

Un traitement à la DNase est réalisé sur les ARN totaux pour éliminer toute trace d'ADN contaminant, suivant les recommandations du fabricant soit : 2  $\mu\text{g}$  d'ARN sont ajoutés à un mélange réactionnel composé de 1 $\mu\text{L}$  de tampon 10 X RQ1 (400 mM Tris-HCl pH 8, 100 mM  $\text{MgSO}_4$ , 10 mM  $\text{CaCl}_2$ ), et 2 u de DNase RQ, complété à 10  $\mu\text{L}$  avec de l'eau DEPC. Une incubation de 30 min à  $37\text{ }^{\circ}\text{C}$  permet d'activer l'enzyme qui est ensuite inhibée par ajout de solution stop RQ1 (20 mM EGTA, pH 8) à raison de 1  $\mu\text{L}$  et incubation à  $65\text{ }^{\circ}\text{C}$  pendant 12 min.

### III.3 Contrôle de la qualité et de la quantité d'acides nucléiques

La quantité et la pureté des ADN et des ARN extraits sont déterminées par spectrométrie et par migration sur gel d'électrophorèse (deux méthodes complémentaires) afin de vérifier que les acides nucléiques ne soient pas dégradés et de déterminer les volumes à prélever pour les transcriptions inverse et pour les amplifications PCR. La spectrométrie permet de mesurer la DO à 260 nm (absorbance des acides nucléiques) et 280 nm (contamination protéique). Pour cela, 1  $\mu\text{L}$  de l'extrait d'acides nucléiques est dilué à 1% avec de l'eau ultra pure puis la DO est mesurée à 260 nm (une unité de DO correspondant à 50  $\mu\text{g mL}^{-1}$  d'ADN double brin et 40  $\mu\text{g/mL}$  pour les ARN). Un rapport DO 260 nm / DO 280 nm compris entre 1,8 et 2 pour l'ADN et autour de 1,6 pour les ARN indique une pureté correcte

des acides nucléiques. Une électrophorèse sur gel d'agarose 1 % analytique contenant du bromure d'éthidium ( $\text{BET } 4 \mu\text{g mL}^{-1}$ ) permet d'apprécier la qualité et d'estimer l'état de dégradation éventuel de l'ADN génomique et les ARN totaux sous UV. Les ARN ribosomaux sont majoritaires (~ 80 %) dans les ARN totaux et sont visualisés sous la forme de deux bandes principales (28S et 18S).

#### III.4 Transcription inverse

La transcription inverse a permis d'obtenir des ADN complémentaires (ADNc) des ARN messagers (ARNm) présents dans les ARN totaux. Les ADNc sont ainsi le reflet des séquences exprimées dans les cellules. 1  $\mu\text{g}$  d'ARN totaux sont ajoutés à 100  $\mu\text{mol}$  d'amorces oligodT modifié ( $5' - \text{GCTCGAGC}(\text{T})_{16}\text{VN} - 3'$ ) (Carpin et al., 1998) et le volume réactionnel est ajusté à 16  $\mu\text{L}$  avec de l'eau DEPC. Une incubation à 70 °C pendant 5 min permet de dénaturer les ARN, puis le milieu réactionnel est placé sur la glace. Aux 16  $\mu\text{L}$  précédents sont ajoutés : 5  $\mu\text{L}$  de tampon de réaction 5X (Tris HCl 250 mM pH 8,3 ; KCl 375 mM ;  $\text{MgCl}_2$  15 mM ; DTT 50 mM) ; 1,25  $\mu\text{L}$  de dNTP (10mM), 30 u RNAsin® inhibiteur (PROMEGA), 200 u de MMLV-RT (PROMEGA). Le milieu réactionnel est alors incubé 60 min à 42 °C et les ADNc synthétisés sont conservés à -20 °C.

#### III.5 Amplification par PCR

La technique de PCR (polymérisation en chaîne) a pour but d'amplifier en grand nombre une séquence particulière d'ADN. Les PCR réalisées sont en phase plateau, c'est à dire que la quantité de produit final amplifié a été limitée par les composants du milieu réactionnel. Ainsi, la quantité de produit final n'est pas proportionnelle à la quantité de matrice. Les résultats obtenus sont donc qualitatifs uniquement. Les amplifications PCR a été utilisée à des fins diverses :

- Détecter la présence des gènes de ménage (Acine et sous unité ribosomale 18S), des gènes marqueurs du stress oxydant (MnSOD) et du stress thermique (SHSP) et des éléments transposables dans le génome des diatomées (rétrotransposons *Blackbeard* et *Surcouf* ainsi que les *MLE*), la technique de PCR a été utilisée sur des matrices ADN génomiques.
- Mettre en évidence la présence de transcrits des gènes précédemment cités et des ET chez les diatomées, les PCR en point final ont été réalisées avec de l'ADN complémentaire. Elles n'ont donc pas permis de quantifier les transcrits mais de détecter la présence/absence de ceux-ci.

- Vérifier les clones bactériens transformés, des PCR sur colonies ont été réalisées.

Pour chaque PCR, un volume final de 25 ou 50 µL de milieu réactionnel a été utilisé comportant 5 ou 10 µL tampon Taq 5X (Promega, Madison, WI), du MgCl<sub>2</sub> 2,5 mM, des dNTP 0,2 mM, de la Taq polymérase à raison de 0,5 U pour 25 µl (Promega, Madison, WI), des amorces 1 µM. Pour les PCR sur ADN génomique, 100 à 150 ng d'ADN (pour 25 µL) ont été utilisés comme matrice. Respectivement 200 ng et 400 ng d'ADNc ont été utilisés comme matrice pour rechercher les gènes et les éléments transposables transcrits. Enfin pour vérifier l'insertion du gène d'intérêt dans les plasmides et dans les bactéries compétentes au cours du clonage une colonie a été prélevée pour servir de matrice.

Les programmes d'amplification comprennent 35 à 40 cycles avec une dénaturation initiale de 5 min à 94 °C. Chaque cycle est constitué d'une dénaturation de 1 minute à 94 °C, d'une hybridation d'1 min et d'une élongation de 30 secondes à 72 °C puis se termine par une élongation finale de 10 min à 72 °C.

La température d'hybridation est variable selon les amorces utilisées et sont répertoriées dans les Tableaux 5, 6, 7 et 8.

**Tableau 5 : Amorces dégénérées (Feschotte et Wessler, 2002) utilisées pour mettre en évidence la présence de *MLE* dans les génomes de microalgues.**

H = A/T/C, R = G/A, Y = T/C, D = G/A/T.

Transposon	Amorces	Séquence 5'→3'	T °	Taille du fragment amplifié
<i>MLE</i>	<i>MLE5A</i>	ATH GAT GAR AAR TGG TTC	47 °C	~380 pb
	<i>MLE3A</i>	GCA TTR TCY TGY TGD AT		

**Tableau 6 : Amorces dégénérées utilisées pour mettre en évidence la présence des gènes de ménage et de leurs transcrits.**

Y = T/C. *Pt* : *Phaeodactylum tricornutum* ; *Ac* : *Amphora coffeaeformis* ; *Aa* : *Amphora acutiuscula*

espèces	gène	amorces	Séquences 5'→3'	T °C	Taille du fragment amplifié
<i>Pt</i> <i>Ac</i> <i>Aa</i>	<i>Actine</i>	<i>ActF</i>	GGG GAA TTC ATG TTYG ARA CBT TCA ACG T	60°C	400 pb
		<i>Act R</i>	GGG GAA TTC GGA ARC GYT CGT TDC CGA T		
<i>Pt</i> <i>Ac</i> <i>Aa</i>	<i>18s</i>	<i>RibF</i>	GGG GGA GTA TGG TCG CAA GGC	70 °C	400 pb
		<i>RibR</i>	TCA GTG TAG CGC GCG TGC GGC		

**Tableau 7 : Amorces spécifiques utilisées pour mettre en évidence la présence des gènes de réponse au stress (oxydant et thermique) et leurs transcrits.**

*Pt* : *Phaeodactylum tricornutum* ; *Ac* : *Amphora coffeaeformis* ; *Aa* : *Amphora acutiuscula*.

Espèce	gène	Amorces	Séquence 5'→3'	T °C	Taille du fragment amplifié
<i>Pt</i> <i>Ac</i> <i>Aa</i>	<i>MnSOD</i>	<i>F-SODsp</i>	ATCTGTCGCCTAACTCG	48 °C	400 - 435 pb
		<i>R-SODsp</i>	TAACTCGTCGGTTCCTTCTG		
<i>Pt</i> <i>Ac</i> <i>Aa</i>	<i>sHSP</i>	<i>F-sHsp</i>	ATG GCG CTC AAC CGA TA	50 °C	400 pb
		<i>R-sHsp</i>	CTA GGC TTT CTT TTC TTC T		

**Tableau 8 : Amorces spécifiques utilisées pour la détection des éléments transposables et leur transcrits chez les diatomées.**

*Pt* : *Phaeodactylum tricornutum* ; *Ac* : *Amphora coffeaeformis* ; *Aa* : *Amphora acutiuscula*.

Espèce	Transposon	Amorces	Séquence 5'→3'	T °	Taille du fragment amplifié
<i>Pt</i>	<i>Phatrimar</i>	<i>Phaeo4</i>	GGC AGT GGC AAG GCC AAG GTT CAT TGAA	65 °C	170 pb
		<i>Phaeo7</i>	CGG TAC ACA GTT CGG TTC ACA TTC TTG TTC		
<i>Ac</i>	<i>Acomar</i>	<i>Ampho4</i>	ACT GCC ACC TGA CAA TGC AAG CAC AAG CG	53 °C	170 pb
		<i>Ampho7</i>	GTC TTT GTT TCC TGC ATG CCC ATG GGT CTC		
<i>Aa</i>	<i>Aacumar</i>	<i>AacutFc et</i>	TAC ATT GGC AAA GTG ACG TTC TTG	59 °C	190 pb
		<i>Aacut Re</i>	TCC TCT CGG TCT GTG TTT CGA T		
<i>Pt</i> <i>Ac</i> <i>Aa</i>	<i>Blackbeard</i>	<i>Bbd 5'-1</i>	AAA GCC CGA CTC GTT CGC CG	60 °C	270 pb
		<i>Bbd 3'-1</i>	ACG CCA AGC GGC ACC ACC CAT A		
<i>Pt</i> <i>Ac</i> <i>Aa</i>	<i>Surcouf</i>	<i>Scf 5'-1</i>	AAG GCA CGC CTT GTT GCT G	60 °C	274 pb
		<i>Scf 3'-1</i>	ACG CTG CGC CAC TGG ACT TT		

## **Visualisation et extraction des fragments amplifiés**

Les fragments d'ADN sont visualisés et séparés en fonction de leurs tailles par migration sur gel d'électrophorèse sur gel d'agarose 1% et à 0,5 % BET. Ils sont ensuite extraits du gel d'agarose en utilisant le kit d'extraction «WIZARD SV Gel and PCR clean up system (PROMEGA). Ce kit permet de purifier des fragments de tailles comprises entre 100 pb et 10 Kpb. L'ADN est ensuite élué des colonnes avec 50 µl d'eau ultra pure et stocké à -20 °C.

### III.6 Clonage des fragments d'ADN

Le clonage est une opération qui a pour but de reproduire en grand nombre un fragment d'ADN d'intérêt. Les fragments d'ADN purifiés ont été ligués dans des plasmides pGEM-T Easy (PROMEGA) à l'aide de la T4 DNA ligase. Le mélange réactionnel de ligation est composé de 1-3 µL de produit PCR purifié ; 5 µL de tampon de ligation 2X (60 mM Tris-HCl pH 7.8, 20 mM MgCl<sub>2</sub>, 20 mM DTT, 2 mM ATP et 10% polyéthylène glycol) ; 1 µL d'enzyme T4 DNA Ligase (3u/µl) ; 1 µL de vecteur pGEM-T Easy (50 ng/µl). Le volume réactionnel est complété à 10 µL avec de l'eau ultra pure. La ligation est réalisée pendant une nuit environ à 4°C.

### III.7 Transformation bactérienne et clonage des fragments d'intérêt

Les plasmides recombinés sont intégrés dans des bactéries compétentes permettant le clonage du plasmide et donc du fragment d'intérêt par multiplication bactérienne. Les bactéries ont acquis la chimiocompétence par un traitement chimique au CaCl<sub>2</sub>, MnCl<sub>2</sub>, MgCl<sub>2</sub> (Hanahan et Bloom, 1996) et sont stockées à -80 °C pendant 3 mois (durée maximale de compétence). La souche utilisée est *Escherichia Coli NEB5α*.

Le produit de ligation (10 µl) est inoculé dans 50 µL de bactéries compétentes. Une incubation de 30 min sur glace et un choc thermique de 42 °C pendant 3 min au bain-marie permettent la transformation par intégration des plasmides dans les bactéries. Celles-ci sont ensuite placées 2 min sur glace pour restaurer la rigidité de la paroi bactérienne. Les bactéries sont mises en culture dans du milieu de Luria Bertani liquide (LB : Tryptone 10 gL<sup>-1</sup> ; extrait de levure 5 gL<sup>-1</sup> ; NaCl 5 gL<sup>-1</sup> ; pH7,5) liquide à 37 °C, puis sont étalées sur boîte de Pétri contenant un milieu LB solide (LB liquide ; Agar 15 gL<sup>-1</sup>) auquel sont ajoutés du X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) à 2%, de l'IPTG (Isopropyl-β-D-thiogalactopyranoside) à 200 mg / mL et de l'ampicilline à 10 mg / mL. Les boîtes de Pétri

sont placées une nuit à l'étuve à 37 °C. Les colonies bactériennes ayant intégré le fragment d'intérêt sont discriminées par criblage blanc / bleu.

### III.8 Extraction plasmidique

La purification d'ADN plasmidique, est réalisée à l'aide du kit Wizard® *Plus* Minipreps DNA Purification System (PROMEGA), après la mise en culture d'une colonie bactérienne sélectionnée, dans 3 mL de LB liquide additionnés de 10 µg/mL d'Ampicilline durant une nuit à 37 °C. Le protocole d'extraction utilisé est le suivant : la culture est centrifugée 3 min à 10 000 g et à 4 °C, le culot est resuspendu dans 200 µl de solution de resuspension (Tris-HCl 50 mM, pH 7,5, EDTA 10 mM, ARNase A 100 µg/mL). Après avoir agité doucement, 200 µl de solution de lyse (NaOH 0,2 M, SDS 1 %) sont ajoutés et l'ensemble est incubé 5 min à température ambiante. La lyse est ensuite neutralisée par ajout de 200 µl de solution d'acétate de potassium 1,32 M et une centrifugation de 5 min à 10 000 g et à 4 °C permet de séparer le surnageant des débris cellulaires. Le surnageant contenant les plasmides clonés est filtré sur mini-colonne avec 1 mL de résine de purification. Après filtration le culot est purifié grâce à la solution de lavage (C<sub>2</sub>H<sub>3</sub>KO<sub>2</sub> 80 mM, Tris-HCl 8,3 mM, pH 7,5, EDTA 40 mM, éthanol 55 %). Les plasmides sont élués avec 50 µl d'eau ultra pure à température ambiante.

### III.9 Séquençage des fragments d'intérêt

Les plasmides contenant le fragment d'intérêt ont été séquencés par la Société Cogenics Beckman Coulter Genomics (<https://www.cogenicsonline.com>), les amorces utilisées pour la lecture des inserts sont SP6 et T7prom.

## **IV Approche bio-informatique**

### IV.1 Analyse des séquences

Les séquences amplifiées par PCR en ADN g et en ADNc ont été identifiées par Blastx sur les banques de données Genbank et Repbase pour les éléments transposables et par Blastn sur Genbank pour les gènes de ménage et marqueurs de stress. Des alignements et Une matrice d'identité ont été réalisés pour comparer les séquences entre elles grâce aux logiciels ClustalW version 1.8 (Thomson et al., 1994), GENEDOC (Nicholas et Nicholas, 1997) et Weblogo (Crooks et al., 2004). Les séquences ont été traduites *in silico*, grâce aux outils de

traduction des sites Infobiogène et de l'université de bourgogne : ([http : //www.u-bourgogne.fr/zebrafish-lab/tools/traduction](http://www.u-bourgogne.fr/zebrafish-lab/tools/traduction)).

#### IV.2 Recherche des ET dans le génome séquencé de *P. tricornutum* souche CCMP632/CCAP1055.1

Afin de déterminer si des séquences *MLE* incomplètes ou complètes étaient présentes et référencées dans le génome de *P. tricornutum*, une approche bioinformatique a été développée en collaboration avec J. D. Rouault et utilisée pour rechercher spécifiquement les *MLE* dans la souche séquencée CCMP632/CCAP1055.1 du génome de la diatomée.

##### IV.2.1 Recherche des *MLE* dans le génome séquencé

La recherche des *MLE* dans les génomes séquencés de diatomées a été réalisée à l'aide des programmes d'identification de séquences Blastx et Blastn. Ces programmes ont permis de rechercher des *MLE* dans les génomes séquencés de diatomées par homologie de séquences avec chacune des séquences identifiées par PCR dans la présente étude. Le génome séquencé de *P. tricornutum* souche CCAP1055.1 (Culture Collection of Algae and Protozoa) qui correspond également à la souche CCMP632 (Provasoli-Guillard National Center for Culture of Marine Phytoplankton) (De Martino et al., 2007) est disponible dans les banques de données, sur le site du Joint Genome Institute à l'adresse suivante :

[http : //genome.jgi-psf.org/cgi-bin/runAlignment?db=Phatr2&advanced=1](http://genome.jgi-psf.org/cgi-bin/runAlignment?db=Phatr2&advanced=1)

##### IV.2.2 Recherche des *MLE* complets dans le génome séquencé

La recherche de *MLE* complets dans le génome de *P. tricornutum* a complété cette analyse. Pour cela le génome de *P. tricornutum* a été téléchargé sur le site Joint Genome Institute : [http : //genome.jgi-psf.org/Phatr2/Phatr2.download.ftp.html](http://genome.jgi-psf.org/Phatr2/Phatr2.download.ftp.html). Pour cette analyse, ont été retenus : les fichiers « unmasked » de la version assemblée du génome en chromosomes, en organites (chloroplastes et mitochondries), en « scaffolds », en EST (Expressed Sequence Tag) et en bibliothèque de répétitions (éléments répétés et transposables) ont été retenus pour cette analyse.

Les fichiers suivants ont été téléchargés :

- Phatr2\_chromosomes\_assembly\_chromosome
- Phatr2\_assembly\_organelle.fasta
- Phatr2\_bd-unmapped\_assembly\_scaffolds
- Phatr2\_repeat

Dans le but de rechercher la présence de *MLE* dans les parties non-annotées du génome séquencé de *P. tricornutum* (tels que les scaffolds par exemples), une série de programmes rédigés en langage ADA par J.D. Rouault (Laboratoire Evolution Génome et Spéciation UPR CNRS 9034, Gif-sur-Yvette) ont permis d'automatiser la procédure de recherche des *MLE*. Le génome assemblé en chromosomes a été divisé en fichiers de données de type Fasta contenant la séquence d'un chromosome par fichier. Des fichiers ont également été réalisés pour les organites, les « scaffold » et les éléments répétés identifiés par Maumus et al. (2009). L'ensemble de ces fichiers formant une base de données locale (personnelle).

Pour rechercher dans le génome séquencé de *P. tricornutum* (base de données locale) la présence de séquences similaires à celles obtenues par PCR dans le présent travail, les séquences suivantes ont été confrontées à la base de données locale :

- Les 55 séquences nucléiques de *MLE* des diatomées marines
- les 55 séquences traduites *in silico* en protéines correspondantes
- Les 25 séquences consensus nucléiques
- Les 25 séquences consensus traduites *in silico*
- Enfin, les séquences de *MLE* complets de plantes : *Soymar1*, *Osmar1* et 2, *Vulmar1* et *Psmar1*

Les *MLE* sont classés dans la classification de Wicker et al., 2007 dans l'Ordre des transposons à TIR car ils présentent des séquences répétées inversées d'environ 30 pb aux deux extrémités. Les caractéristiques des *MLE* complets de végétaux ont été prises en compte pour rechercher les *MLE* complets dans le génome de *P. tricornutum* et sont les suivants : la longueur moyenne des TIR d'environ 30 pb, la longueur moyenne des *MLE* de végétaux variant en moyenne entre 3 et 5 Kpb, la duplication d'un dinucléotide TA lors de l'insertion de l'élément ou TSD pour Target Site Duplication (Figure 27).

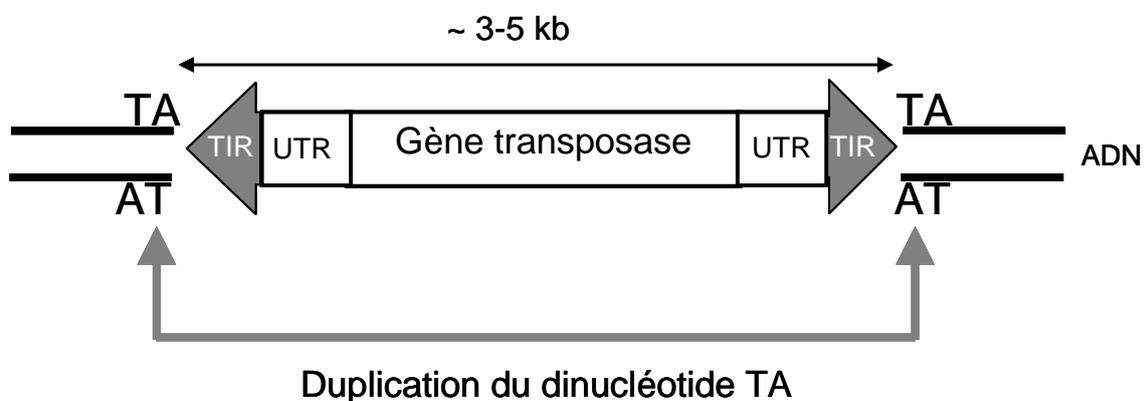


Figure 27 : Représentation schématique d'un transposons *MLE* végétal.

Le gène de la transposase est encadré par des séquences TIR (Terminale Inverse Repeat) et UTR (Untranslated Region). Il s'insère au sein d'un site cible qui est généralement un dinucléotide TA. Celui-ci est dupliqué durant l'insertion. Le dinucléotide TA est donc présent à chaque extrémité du transposon et représente une signature (TSD) de l'insertion du *MLE*.

Ainsi une nouvelle démarche a été mise en place pour identifier de nouveaux ET à TIR dans le génome de séquencé de *P. tricornutum*.

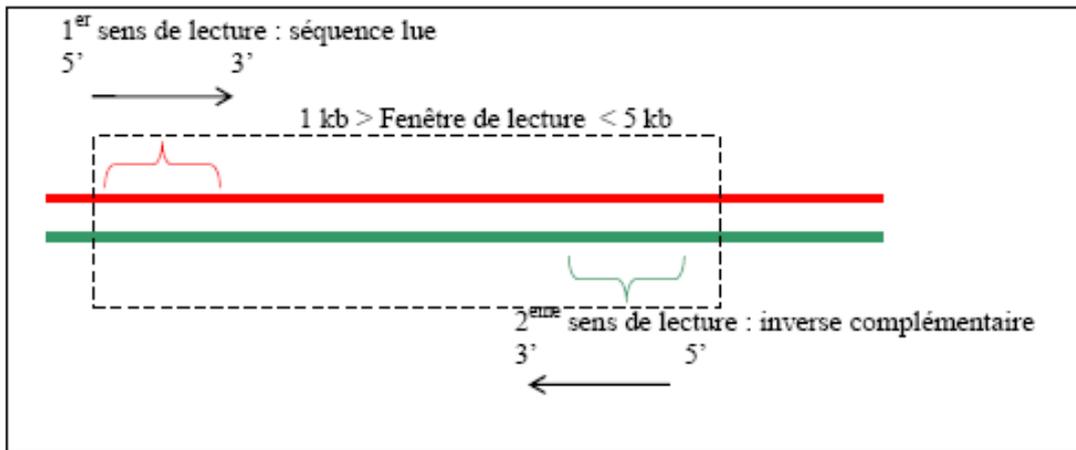
La démarche suivie est la recherche de séquences inversées répétées (TIR) commençant et finissant par un TA, le tout encadrant une séquence dont la longueur est comprise entre 1 Kpb et 5 Kpb dans le génome (Figure 28). Elle se compose de 6 étapes :

1. Recherche d'un premier TA
2. Recherche d'un second TA dans une fenêtre de 1 à 5 Kpb.
3. Extraction des TIR potentiels d'une longueur d'au moins 20 pb.
4. La séquence du second TIR est inversée et le complémentaire est réalisé
5. Comparaison des deux TIR. Si le nombre de différences est inférieur à un seuil fixé, on considère qu'on est en présence d'un ET potentiel.
6. Cette approche ne détecte pas les ET de longueur supérieure ou inférieure à la fenêtre considérée, les ET dont le TSD a muté (et n'est plus TA), les ET dont les TIR ont subi des insertions/délétions (problème de l'alignement direct).

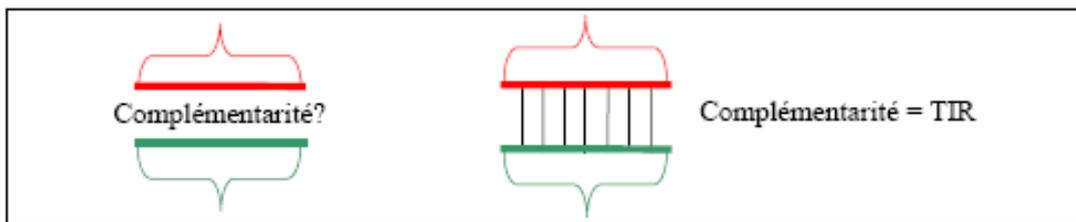
Le programme scanne ainsi les fichiers de données du génome et lit la séquence dans le sens 5'→3' dans une fenêtre de 1 à 5 Kpb. Dans cette fenêtre de lecture, le programme réalise la séquence inverse complémentaire et aligne les deux séquences. Lorsqu'une complémentarité d'au moins 20 pb est détectée (correspondant à un TIR de *MLE*), la séquence (incluant les TIR et les TA et la séquence centrale) est extraite pour analyse ultérieure. Le processus reprend jusqu'à la fin du fichier. La séquence est ensuite comparée aux séquences *MLE* de diatomées et par Blast sur les banques de données NCBI ainsi que Repbase.

Les séquences TIR des *MLE* sont fréquemment décrites comme imparfaites (Plasterk et al., 1999; Brillet et al., 2007). Les analyses ont été réalisées en intégrant la probabilité que les TIR soient imparfaits, c'est-à-dire qu'ils possèdent au moins une erreur ou une délétion l'alignement des deux séquences. Les résultats présentant des TIR imparfaits à plus de 9 erreurs n'ont pas été retenus.

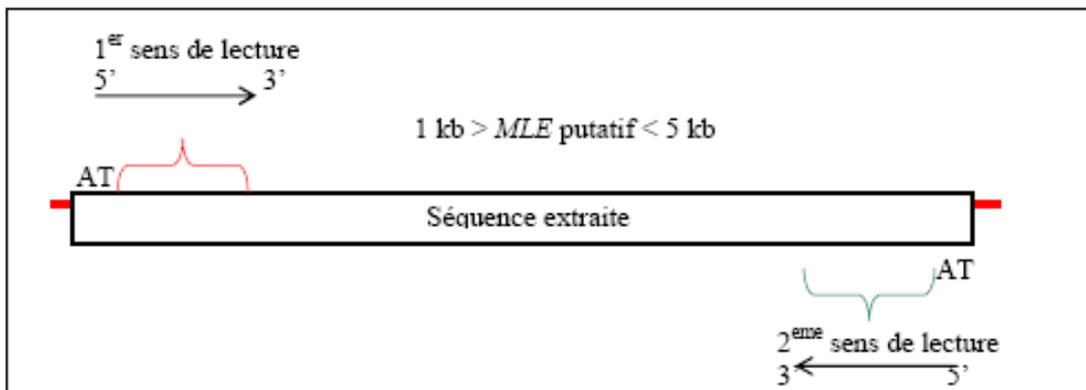
Etape 1 : lecture des séquences génomiques :



Etape 2 : détection des TIR :



Etape 3 : extraction de la séquence putative de MLE:



**Figure 28 : Schéma des étapes de détection *in silico* des éléments à TIR dans le génome de *Phaeodactylum tricorutum* séquencé.**

Après détection d'un TA, la première étape consiste en la définition de la fenêtre de lecture. La séquence présente dans cette fenêtre est lue dans le sens 5'-3' et la séquence inverse complémentaire est réalisée. Dans la deuxième étape, les deux séquences sont alignées afin de rechercher les régions terminales inversées (TIR). Lorsque que des TIR sont détectées et que des dinucléotides TA sont présents dans la région flanquante, la séquence entière (TIR, séquence centrale et séquences flanquantes contenant des TA) est extraite et ceci constitue la dernière étape.

### IV.3 Méthodes phylogénétiques

Afin de comparer les séquences *MLE* obtenues et de tenter de reconstruire l'histoire évolutive de ces séquences, plusieurs méthodes phylogénétiques ont été utilisées pour réaliser des arbres retraçant en partie les relations de parenté probables des *MLE* de diatomées marines mais également des *MLE* de plantes terrestres. Tous les arbres phylogénétiques ont été réalisés sur des séquences traduites *in silico* en acides aminés. Après alignement des séquences partielles de *MLE* grâce au logiciel ClustalW version 1.8 (Thomson et al., 1994), les séquences ont été alignées manuellement avec le logiciel GENEDOC (Nicholas et Nicholas, 1997).

La méthode de Neighbor Joining a été réalisée à l'aide du logiciel PAUP 4.0b08 (Swofford, 2002) avec le modèle MEAN (le modèle pour acide aminé de PAUP). La robustesse des arbres obtenus a été testée par le test de bootstrap effectué en 1000 réplifications. La méthode du maximum de parcimonie a été réalisée à l'aide du logiciel PAUP 4.0b08 (Swofford, 2002) avec la méthode recherche heuristique de l'arbre le plus parcimonieux : section et reconnection des arbres ou « Tree bisection and reconnection ». Le test de bootstrap a été effectué en 1000 réplifications avec addition de 10 séquences à chaque réplification.

La méthode de l'inférence bayésienne a été réalisée avec le modèle WAG+G déterminé par ProTest (Abascal et al., 2007) à partir du critère d'information Akaike. Le modèle a été implémenté dans le logiciel Mr Bayes version 3.1.2 (Ronquist et Huelsenbeck, 2003). Les analyses MCMC ont été conduites à l'aide de trois chaînes haute température et une chaîne basse température. Le programme a été lancé au moins deux fois indépendamment, avec le même jeu de données afin de vérifier la stabilité des résultats par une convergence de la déviation standard moyenne vers 0.

Les représentations des arbres ont été réalisées avec le programme Mega4 (Tamura et al., 2007).

### IV.4 Classification agrégative des ET

Une nouvelle méthode, UPGM-VM, a été développée par J. D. Rouault afin de classer des séquences de longueurs différentes. Cette méthode permet de comparer les séquences incomplètes avec les séquences complètes classées dans des familles définies dans la littérature (Rouault et al., 2009) ce qui n'est pas possible avec les méthodes de phylogénies et de classification classiques. Ainsi, la classification UPGM-VM a été sélectionnée pour classer les séquences *MLE* partielles issues de ce travail par rapport aux éléments (des séquences

complètes en majorité) appartenant à la superfamille *Tc1-mariner*. Cela permet plus particulièrement de classer les séquences partielles de *MLE* issues de végétaux et de diatomées (700 pb) avec les *MLE* végétaux complets (3 à 5 Kpb).

La méthode UPGM-VM bien qu'issue d'une collaboration sera détaillée dans les paragraphes suivants, car sa mise au point est très récente (Rouault et al., 2009) et la méthode est peu connue au moment de la rédaction de ce manuscrit.

#### IV.4.1 Définition de la classification agrégative

Une classification débute par l'intégration d'un ensemble de données ou items et par l'établissement de groupes selon des critères définis *a priori*. Les items sont répartis au sein des groupes avec la propriété pour chacun des items d'appartenir à un groupe et un seul (Bouvier et al., 2005). L'ensemble des groupes ainsi constitué forme la classification. Le problème de la classification est un cas particulier de la théorie de la catégorisation qui représente le processus par lequel les items sont perçus comme similaires et assemblables (Jordan et Russell, 1999). Il faut différencier la notion de classification de celle de classement. En effet, la classification est un processus qui est réalisé sans fixer *a priori* le nombre de classes ou groupes qui seront constitués par les items, à la différence du classement qui consiste à ranger des items inconnus parmi des groupes prédéfinis (par une classification antérieure par exemple) dont les propriétés sont connues.

Ici, les items considérés sont des séquences de *MLE* provenant de végétaux terrestres et marins ainsi que des transposons issus d'organismes animaux utilisés en groupe externe ou référent. Une classification a été établie à partir de ces séquences ainsi qu'un classement en groupes dont les propriétés seront définies dans le classement.

#### IV.4.2 Matériel, ressources et données utilisées

Notre ensemble de données se compose de 491 séquences complètes et incomplètes de transposons, les numéros d'accèsion GenBank de toutes les séquences sont répertoriés en Annexe 3.

Cet ensemble se décompose comme suit :

- 249 séquences *MLE* de végétaux soit :
  - 54 séquences obtenues expérimentalement chez 10 espèces de diatomées (55 séquences au départ, dont deux séquences identiques qui ont été confondues). Ces séquences ont été obtenues à l'aide des amorces *MLE5A* et *MLE3A* définies par Feschotte et Wessler (2002) et sont donc tronquées, elles mesurent entre 373 et 401 pb.

- 173 séquences de *MLE* courtes (<700 pb) de végétaux obtenus *in vivo* ou *in silico* par Feschotte et Wessler (2002) chez différentes espèces d'Angiospermes, par Jacobs et al. (2004) dans le genre *Beta*, et par Zhou et al. (2010) au sein des bambous.

- 22 séquences *MLE* dites longues (>700 pb) isolées du soja : *Glycine max* (Jarvik et Lark 1998), du riz, *Oryza sativa* (Tarchini et al. 2000; Feschotte et Wessler, 2002), de l'arabette, *Arabidopsis thaliana* (Feschotte et Wessler, 2002), de la betterave : *Beta vulgaris* (Jacobs et al. 2004), de la luzerne : *Medicago truncatula* et du pois : *Pisum sativum* (Macas et al. 2005).

Le groupe externe est composé de 242 séquences longues appartenant à différentes familles (*mariner*, *Fot*, *maT*, *Tc1*, *Gambol*, *Zator* ...) (Robertson, 1993; Robertson et Asplund, 1996; Robertson, 1997; Shao et Tu, 2001; Claudianos et al., 2002; Bui et al., 2008; Bao et al., 2009). Dans la mesure du possible, les séquences les plus longues possibles ont été retenues, TIR inclus lorsqu'ils étaient définis.

#### IV.4.3 Méthode

La méthode utilisée ici pour réaliser la classification des séquences de *MLE* de végétaux est une méthode développée par Rouault et Capy (2008) basée sur la classification ascendante hiérarchique (CAH).

Les méthodes hiérarchiques ascendantes (ou agrégatives) phénétiques, comme la méthode UPGMA (Unweighted Pair Group Method with Arithmetic Mean) consistent à calculer les distances entre tous les items et à constituer progressivement des groupes en fusionnant les sous-ensembles les plus proches (Lebeaux et Jambu, 1978). Le processus commence donc par la formation de  $n$  groupes de 1 item chacun ( $n$  correspondant au nombre d'items) et se finalise par la constitution d'un groupe de  $n$  items.

La méthode UPGMA consiste à calculer la matrice de distance entre tous les items à classer et à appliquer un algorithme de classification pour construire un arbre enraciné ou dendrogramme. A chaque nœud est associée une valeur moyenne qui se substitue aux valeurs individuelles pour la suite de la construction.

Les figures ont été réalisées à l'aide de Scilab et modifiées ou commentées par Paint.NET. Tous les programmes ont été écrits en langage de programmation ADA sur les plateformes WINXP et WINVISTA par Jacques Deric Rouault (LEGS UPR CNRS 9034, Gif-sur-Yvette).

#### IV.4.4 Le processus agrégatif UPGM-VM

La méthode développée par Rouault et Capy en 2008 est nommée UPGM-VM (pour **V**ariation de la **M**étrique). Elle se distingue de l'UPGMA (**U**nweighted **P**air **G**roup **M**ethod with **A**rithmetic mean) par deux aspects:

- Il n'y a pas de consensus ou de valeur moyenne associée à chaque nœud.

La distance entre deux groupes est calculée comme la distance moyenne entre tous les items de chacun des deux groupes. Donc seules les valeurs de distance de la matrice d'identité initiale sont utilisées. Ainsi chaque item est considéré pour calculer la distance du groupe par rapport aux autres groupes.

- La métrique est la fonction qui permet de calculer la distance de tous les items du groupe par rapport aux items des autres groupes. Cette métrique varie au cours du processus d'agrégation. En effet, au début du processus d'agrégation, la distance (mesure de l'éloignement entre deux items) est globale. Ainsi deux séquences sont identiques si et seulement si elles ont la même longueur et des nucléotides identiques. Une séquence présentant une délétion d'un seul nucléotide par rapport à la même séquence n'en présentant pas, sont considérées comme différentes. Au cours du processus d'agrégation, la distance devient locale, ainsi deux séquences identiques à quelques délétions près sont regroupées. Ceci permet de rassembler dans un même groupe une séquence complète ou mutée. L'utilisation de la distance locale revient à ignorer les insertions/délétions dans un alignement de séquence. Entre la distance globale et la distance locale, tous les intermédiaires sont possibles.

Au cours du processus d'agrégation de la méthode UPGM-VM, la métrique varie de globale à locale, ce qui permet d'intégrer dans l'analyse, des séquences de tailles très différentes (Rouault et al. 2009). Dans l'approche suivie ici, la méthode UPGM-VM est appliquée à des séquences nucléotidiques des transposons et non restreinte à la traduction protéique de l'ORF.

#### **La notion et la structure de Clan**

Le Clan est une propriété mathématique qui permet de définir un groupe comprenant des items partageant des propriétés communes qui ne sont pas partagées par les autres items de l'espace des éléments. Ceci aboutit à la stabilité du groupe formé. Le fait d'être un Clan est une propriété du groupe, et cela n'est pas lié à un niveau hiérarchique de la classification.

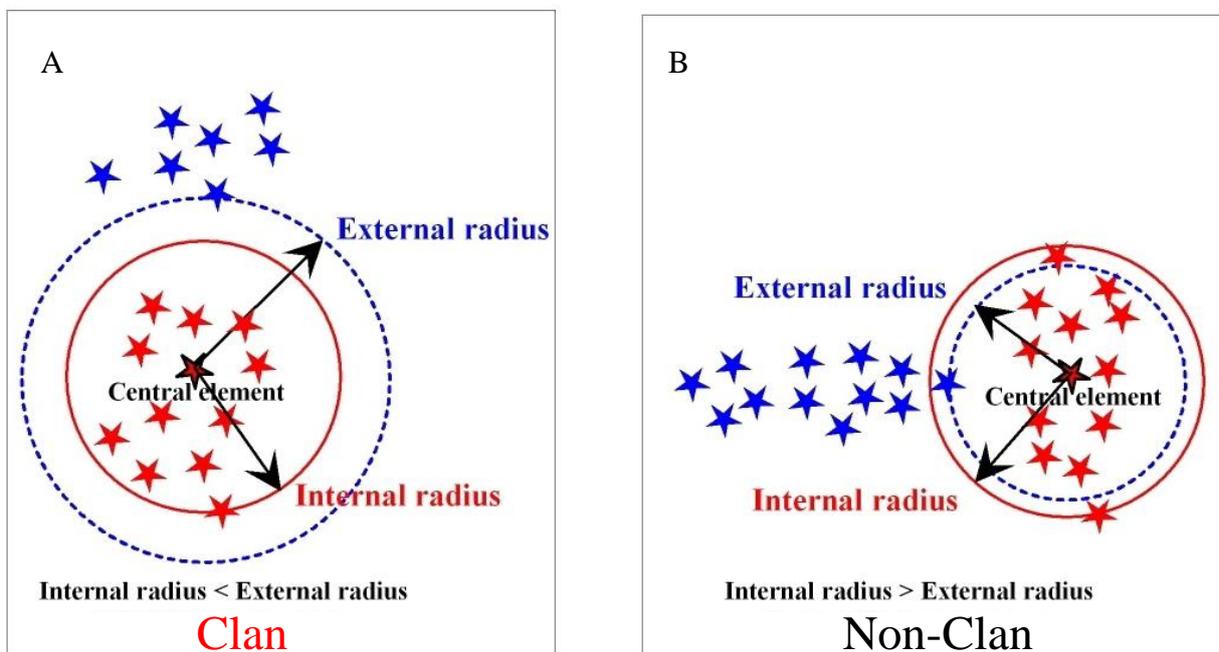
Dans l'approche mathématique UPGM-VM, l'espace des ET est pavé par des sphères de tailles variables. Chaque sphère intègre un groupe d'éléments. Selon les configurations, les sphères peuvent se recouvrir ou être disjointes.

Chaque groupe d'éléments transposables est caractérisé par trois paramètres qui sont directement calculés à partir de la matrice de distance deux à deux (Figure 29) :

1) L'élément central. C'est l'élément du groupe qui minimise la somme des distances à tous les autres éléments du groupe.

2) Le rayon interne. Il représente la distance de l'élément central à l'élément le plus éloigné du groupe.

3) Le rayon externe. Il est défini par la distance de l'élément central à l'élément extérieur au groupe qui est le plus proche.



**Figure 29 : Définition d'un Clan au cours du processus d'agrégation (D'après Rouault et al., 2009).**

Chaque groupe est caractérisé par l'élément central (*Central element*), le rayon externe (*External radius*) et le rayon interne (*Internal radius*). En A, le rayon externe est supérieur au rayon interne, le groupe a donc la propriété de Clan. En B, le rayon externe est inférieur au rayon interne, le groupe ne possède pas les propriétés d'un Clan.

Deux cas de figure sont possibles :

Le rayon interne est inférieur au rayon externe, dans ce cas le groupe est validé, c'est un Clan, ce qui signifie que ce groupe est parfaitement différencié et géométriquement bien défini par une sphère (élément central et rayon interne) (Figure 29).

Si le rayon interne est supérieur au rayon externe, alors le groupe est invalide ce n'est pas un Clan (Figure 29 B). Les éléments situés dans la sphère définie par le rayon externe sont correctement classés mais l'appartenance des éléments situés dans la couronne (espace entre

les deux rayons) à ce groupe est ambiguë. De manière générale, les premiers groupes formés par le processus d'agrégation sont des Clans, mais les derniers groupes ne sont pas des Clans. Il est donc nécessaire de rechercher les groupes intermédiaires ayant les propriétés de Clan constitués du maximum d'éléments de façon à optimiser la qualité du classement des séquences.

Chaque pas de la procédure (Annexe 4) a été programmé dans un des programmes informatiques suivants : ET 815, ET 816, ET 871, ET 872, ET 881, ET 890 par Jacques Deric Rouault (LEGS UPR CNRS 9034, Gif-sur-Yvette).

### **Terminologie de la classification UPGM-VM**

Dans la classification des ET, quatre niveaux hiérarchiques ont été définis : la superfamille, la famille, la sous-famille et la tribu en descendant dans la structure hiérarchiques jusqu'à ce que le groupe le plus petit ait la propriété de Clan. Suivant la terminologie de Rouault et al. (2009) ces noms sont des noms dédiés et notés avec une majuscule.

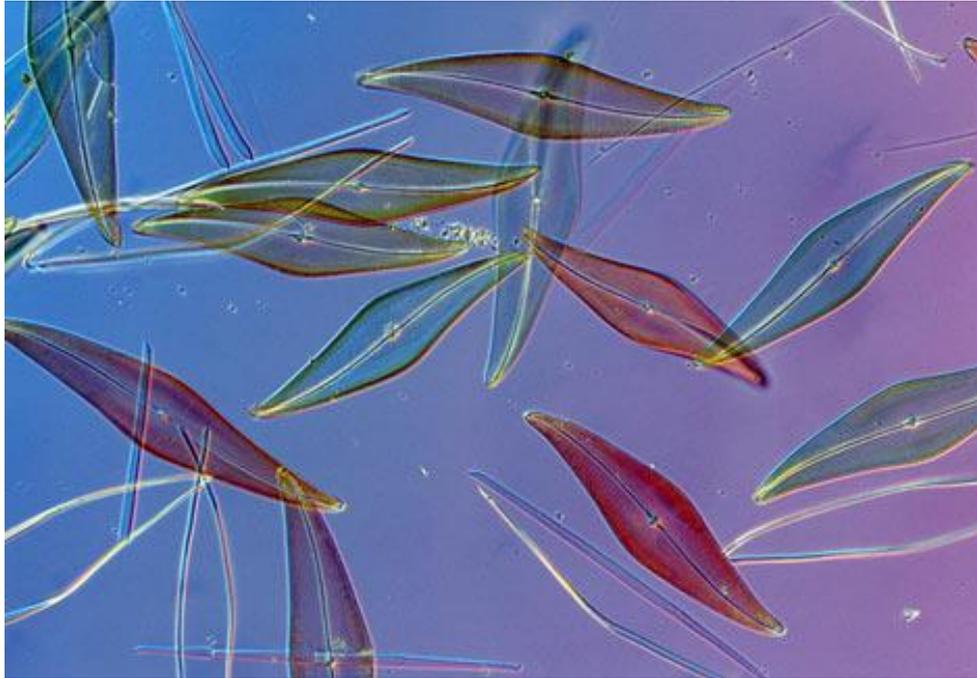
Wicker et al. (2007) ont associé à chacune des 30 superfamilles un code à 3 lettres. La superfamille *Tc1-mariner* est alors abrégée DTT (pour DNA transposons (class II) de l'ordre TIR et de la Superfamille T*c1-mariner*). Dans la nouvelle classification réalisée avec la méthode UGM-VM, la terminologie à 3 lettres a été étendue aux niveaux hiérarchiques inférieurs, en utilisant des mots de trois lettres, elle est alors appelée TESN pour Transposable Element Systemic Naming (Rouault et Capy, 2008 ; Rouault et al., 2009).

Par exemple DTTMarMauDIP désigne la tribu DIP (Diptera) de la sous-famille Mau (Mauritiana) de la famille Mar (*mariner*) de la superfamille DTT (*Tc1-mariner*) qui contient l'élément *Mos1*. Cette nomenclature très condensée permet de rappeler la classification systématique de chaque ET (Rouault et al., 2009).

Chaque séquence nucléotidique d'ET est intégrée dans la base de données personnelle selon le code NSI (Nucleotidic Sequence Identification). Chaque code NSI correspond à une séquence nucléotidique unique. Ainsi, une même séquence identifiée dans des hôtes différents sera affectée d'un code NSI unique. Le code NSI se compose de lettres (permettant une classification simple et une extension infinie de la base de donnée) séparant deux nombres, le premier, donne la longueur de la séquence et le second, le rang historique. Par exemple C451B51 correspond à la 51<sup>ème</sup> séquence de 451 pb.

# CHAPITRE III

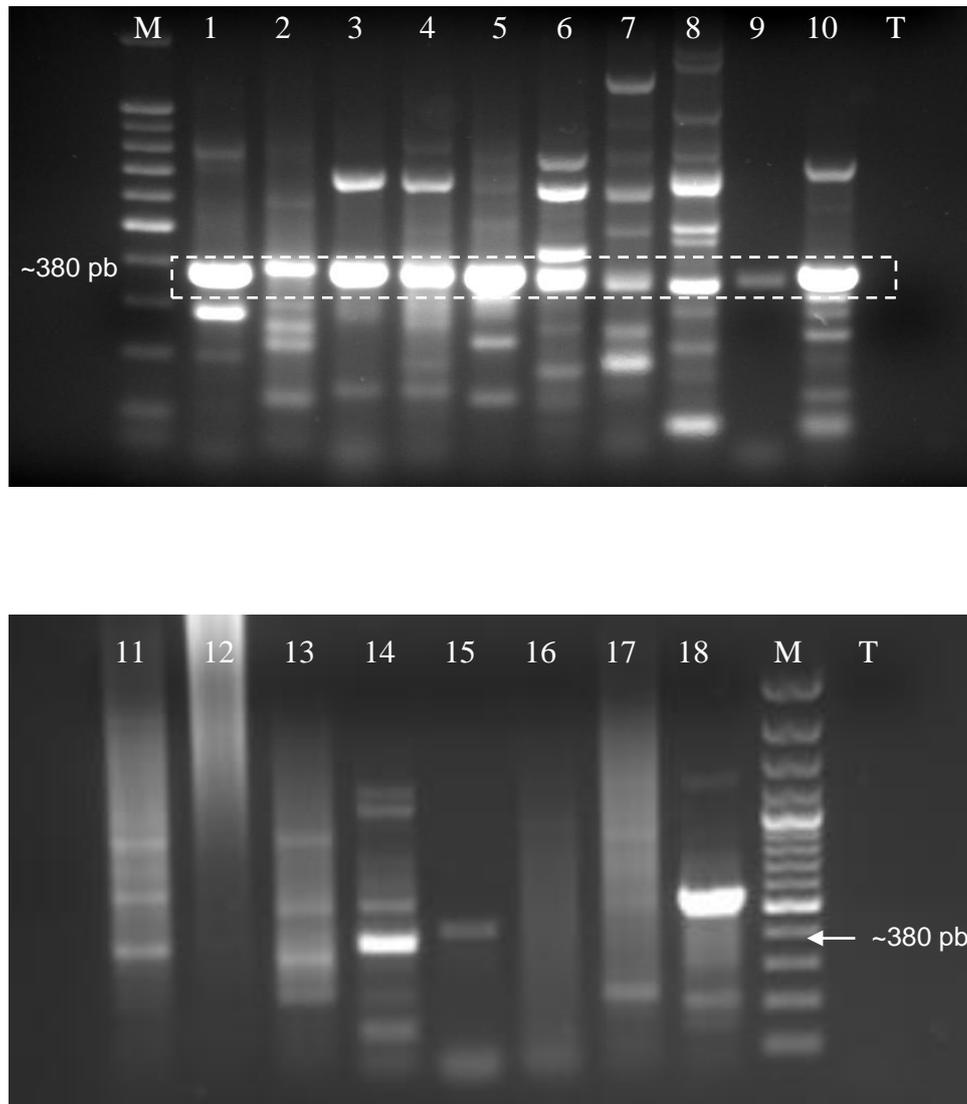
## RESULTATS ET DISCUSSION



# I Recherche de *MLE* chez les microalgues marines

## I.1 Caractérisation moléculaire des *MLE*

Une collection de 18 microalgues cultivées en conditions contrôlées est entretenue au laboratoire. Cette collection est composée de 15 espèces de diatomées, deux espèces de microalgues verte (une Prasinophycée et une Pavlovophycée) et une espèce de microalgue rouge (Rhodophycée). Les *MLE* ont été recherchés chez toutes les espèces de la collection en utilisant les amorces dégénérées *MLE5A* et *MLE3A*. Celles-ci permettent d'amplifier une région interne de 380 pb du gène codant la transposase. L'obtention de fragments de *MLE* avec des amorces dégénérées étant parfois difficile, cinq essais d'amplification par PCR au minimum ont été réalisés pour chaque espèce. Les résultats obtenus sont présentés dans la figure 30. Des fragments amplifiés de tailles différentes ont été obtenus chez toutes les espèces testées, excepté chez *S. constricta* (Figure 30, ligne 9) où un seul fragment d'environ 380 pb a été amplifié et aucune amplification n'a été obtenue chez *S. costatum* (Figure 30, ligne 17). Notre travail a été focalisé uniquement sur les fragments d'environ 380 pb. Tous les fragments de cette taille ont été clonés puis séquencés pour la recherche d'identification.



**Figure 30 : Amplification des fragments de *MLE* chez 18 espèces de microalgues.**

Les amplifications ont été obtenues avec les amorces *MLE5A* et *MLE3A*. Plusieurs fragments ont été obtenus pour chaque espèce exceptée chez *Stauroneis constricta*. Les amplifications encadrées correspondent aux fragments qui ont été séquencés et identifiés comme des *MLE*.

Ligne 1 : *Amphora acutiuscula*, Ligne 2 : *Amphora coffeaeformis*, Ligne 3 : *Entomoneis paludosa*, Ligne 4 : *Haslea crucigera*, Ligne 5 : *Nitzschia laevis*, Ligne 6 : *Nitzschia palea*, Ligne 7 : *Odontella aurita*, Ligne 8 : *Phaeodactylum tricorutum*, Ligne 9 : *Stauroneis constricta*, Ligne 10 : *Thalassiosira weissflogii*, Ligne 11 : *Cylindrotheca closterium*, Ligne 12 : *Cylindrotheca fusiformis*, Ligne 13 : *Haslea ostrearia*, Ligne 14 : *Isochrysis galbana*, Ligne 15 : *Porphyridium cruentum*, Ligne 16 : *Pavlova lutheri*, Ligne 17 : *Skeletonema costatum*, Ligne 18 : *Thalassiosira pseudonana*, M : marqueur de taille 100-bp DNA ladder plus, T : témoin négatif (l'ADN y a été remplacé par de l'eau ultra pure).

## I.2 Identification des *MLE*

Pour identifier des *MLE*, au total 88 séquences d'environ 380 pb ont été confrontées parallèlement aux deux bases de données, la base généraliste : NCBI ([http : //blast.ncbi.nlm.nih.gov/Blast.cgi](http://blast.ncbi.nlm.nih.gov/Blast.cgi)) et la base de données spécifique des éléments transposables : Repbase ([http : //www.girinst.org/censor/index.php](http://www.girinst.org/censor/index.php)). Les transposons étant des composants du génome extrêmement mutés même au sein d'un seul et même organisme, un Blastn au niveau nucléotidique n'était pas approprié. Dans cette étude, nous avons considéré une séquence identifiée comme *MLE* lorsque les comparaisons au niveau protéique par Blastx sur NCBI et RepeatMasking sur Repbase correspondent uniquement à des transposons apparentés aux éléments de type *mariner*, et ce, même si les scores de Blastx étaient faibles. De cette façon, 55 séquences ont été identifiées comme *MLE* hypothétiques chez 10 espèces de microalgues parmi les 18 testées (Tableau 9). Les scores de Blastx sur NCBI varient de 55 bits (pour les séquences de *S. constricta*) à 140 bits (pour les séquences de *T. weissflogii*). Les 55 séquences de *MLE* sont regroupées chez les 10 espèces de diatomées positives à la détection de *MLE* (Figure 30, ligne 1 à 9 et tableau 9), ce qui suggère une large distribution des *MLE* dans la classe des Bacillariophycées (diatomées).

Pour neuf des dix espèces de diatomées testées positivement par PCR, au moins 5 séquences de *MLE* ont été obtenues. Chaque séquence correspond à un clone PCR différent. Sept séquences ont été obtenues pour l'espèce *S. constricta* et six séquences ont été obtenues chez les espèces *A. acutiuscula*, *A. coffeaeformis*, *N. laevis* et *T. weissflogii*. Pour les autres espèces, cinq séquences ont été identifiées comme des *MLE* sauf chez *O. aurita* où seulement quatre séquences de *MLE* ont pu être obtenues.

Les amplifications des huit espèces regroupées en figure 30 ligne 11 à 18 ne correspondent pas à des *MLE*. Il est à noter que parmi ces huit espèces qui ne présentent pas de *MLE*, *P. cruentum*, *I. galbana* et *P. lutheri* n'appartiennent pas à la classe des diatomées. Ainsi dans la suite de cette étude, nous nous sommes intéressés à la caractérisation des *MLE* chez les diatomées uniquement.

**Tableau 9 : Liste des espèces testées, nombre de séquences clonées et nombre de séquences identifiées par espèce.**

En grisé, les espèces au sein desquelles aucun élément *MLE* n'a été mis en évidence.

	Espèces	Nombre de séquences de ~380 pb, clonées	Nombre de séquences correspondant à des <i>MLE</i>
<b>Diatomées</b>	1) <i>Amphora acutiuscula</i>	7	6
	2) <i>Amphora coffeaeformis</i>	8	6
	3) <i>Cylindrotheca fusiformis</i>	3	0
	4) <i>Cylindrotheca closterium</i>	2	0
	5) <i>Entomoneis paludosa</i>	6	5
	6) <i>Haslea crucigera</i>	5	5
	7) <i>Nitzschia laevis</i>	7	6
	8) <i>Nitzschia palea</i>	6	5
	9) <i>Odontella aurita</i>	6	4
	10) <i>Phaeodactyllum tricornutum</i>	9	5
	11) <i>Stauroneis constricta</i>	7	7
	12) <i>Thalassiosira pseudonana</i>	6	0
	13) <i>Thalassiosira weissflogii</i>	7	6
	14) <i>Haslea ostrearia</i>	1	0
	15) <i>Skeletonema costatum</i>	0	0
	16) <i>Porphyridium cruentum</i>	3	0
	17) <i>Isochrysis galbana</i>	3	0
	18) <i>Pavlova lutheri</i>	2	0
	<b>TOTAL</b>	88	55

### I.3 Analyse des séquences de *MLE*

#### I.3.1 Analyse des séquences nucléiques

#### **Longueur des séquences de *MLE***

Chez les diatomées, la longueur des séquences *MLE* varie entre les espèces de 372 pb chez *S. constricta* à 401 pb chez *A. coffeaeformis*, ce qui correspond à un intervalle de 29 pb.

Des variations de la taille des séquences entre 9 et 23 nucléotides ont été observées parfois au sein d'une même espèce. Cette variation intraspécifique concerne la moitié des espèces de diatomées étudiées. Chez *A. coffeaeformis*, la longueur des séquences varie de 378 pb à 401 pb, ce qui correspond à un écart de 23 pb. *Acomar4.1* est la seule séquence de 378 pb parmi les six obtenues pour cette espèce, la longueur des cinq autres séquences étant supérieure ou égale à 390 pb. De même, chez *E. paludosa*, les séquences varient de 373 à 390 pb, soit un intervalle de 22 pb. Une variation de 19 pb est observée entre les séquences de l'espèce *H. crucigera* dont les tailles varient de 378 à 397 pb. *S. constricta* présente seulement

deux longueurs de séquences, 372 et 381 pb, il en résulte une variation de 9 pb. Pour l'espèce *P. tricornutum*, une seule séquence diffère, *Phatrimar1.3* a une longueur de 380 pb contre 390 pb pour les autres séquences de cette espèce.

Chez quelques espèces, on observe des variations inférieures ou égales à 6 pb. Ainsi chez *A. acutiuscula*, la longueur des séquences varie de 375 à 379 pb. Chez *N. laevis*, on observe des longueurs de séquences de 378 pb et 384 pb. De manière similaire, chez l'espèce *T. weissflogii*, les séquences varient de 379 à 381 pb. Deux espèces présentent des séquences dont la taille ne fluctue que d'un nucléotide, *N. palea* dont les séquences sont de 387-388 pb et *O. aurita* de 396-397 pb.

Des variations d'environ 20 pb pourraient correspondre à la présence d'introns de petites tailles, toutefois le logiciel netgene2 ([http : //www.cbs.dtu.dk/services/NetGene2/](http://www.cbs.dtu.dk/services/NetGene2/)) communément utilisé pour la détection d'introns n'a pas permis de les mettre en évidence. Il paraît donc plus probable que ces fluctuations soient dues à la présence de différentes copies divergentes de *MLE*. Lorsque les variations sont inférieures ou égales à 6 pb, il peut s'agir d'évènements de mutations ponctuelles.

### **Comparaisons des séquences nucléotidiques et formation des consensus**

Chez chacune des espèces, les séquences nucléotidiques ont été comparées entre elles par alignement à l'aide du logiciel de bio-informatique ClustalW sur le serveur Infobiogen. Des matrices d'identités de séquences ont été réalisées pour chaque espèce (Annexe 6). Des consensus ont été réalisés sur la base du calcul des pourcentages d'identité partagés par les séquences. Lorsque le pourcentage d'identité est supérieur ou égal à 75 % entre deux séquences, l'appartenance à un seul consensus et donc à une seule et même famille génique est admise. Le consensus est donc le reflet de l'homogénéité des séquences qui le composent. Les consensus et les séquences qui les composent sont listés dans le Tableau 10. Un seul consensus représentant l'ensemble des séquences obtenues pour une espèce a été défini chez les trois diatomées, *A. acutiuscula*, *O. aurita* et *T. weissflogii*. Deux consensus différents ont été obtenus pour les espèces *N. palea*, *P. tricornutum* et *S. constricta*. Concernant les quatre espèces *A. coffeaeformis*, *E. paludosa*, *H. crucigera* et *N. laevis*, plus de trois consensus ont été définis. Chez ces quatre espèces, les séquences obtenues sont très différentes les unes des autres. *N. laevis* est la seule espèce pour laquelle il y a autant de consensus que de séquences amplifiées.

Ces résultats démontrent un fort polymorphisme de séquences au sein même des espèces. Ceci nous permet d'envisager la présence de copies divergentes de *MLE* chez les diatomées. Cette hypothèse expliquerait alors l'obtention de séquences à faible pourcentage

d'identité obtenu chez une même espèce. Chez les espèces *A. acutiuscula*, *O. aurita* et *T. weissflogii*, un seul consensus a été réalisé, ce qui suggère qu'un seul gène de la transposase appartenant à un seul et même transposon *MLE* a été amplifié. Ces données peuvent laisser supposer la présence d'un élément actif chez ces espèces dont les copies seraient majoritairement présentes et donc amplifiées de manière préférentielle. La divergence des séquences chez *N. laevis* nous conduit à penser, au contraire, qu'il y existe de nombreuses copies mutées de *MLE* chez cette espèce puisque chaque séquence apparaît comme ayant ses spécificités propres.

### I.3.2 Analyse des séquences protéiques

#### **Présence de cadres de lecture ouverts**

La majorité des séquences, soit 80 % ne présentent pas de codon stop dans la séquence partielle du gène de la transposase obtenue à l'aide des amorces *MLE5A* et *MLE3A*, ce qui se traduit par un cadre de lecture ouvert (ORF, pour Open Reading Frame) putatif (Tableau 10). Pour six des 10 espèces étudiées, la totalité des séquences *MLE* obtenues possède un ORF, il s'agit des espèces : *E. paludosa*, *H. crucigera*, *N. laevis*, *N. palea*, *P. tricornutum* et *T. weissflogii*. Chez les deux espèces du genre *Amphora*, la majorité de leurs séquences, soit 4 et 5 séquences respectivement pour *A. coffeaeformis* et *A. acutiuscula*, présente des ORF. Chez les espèces *O. aurita* et *S. constricta*, les *MLE* décelés ne présentent qu'une seule séquence à ORF.

Le fait que la plupart des séquences partielles de *MLE* obtenues chez les diatomées présentent des ORF putatifs, suggère l'existence potentielle d'éléments complets et intacts dans les génomes des diatomées. La recherche d'éléments complets dans le génome des diatomées fait l'objet de la section I.5 de ce chapitre.

**Tableau 10 : Longueur des séquences *MLE* obtenues chez dix diatomées, nombre de cadre de lecture ouvert putatif (ORF), nom des consensus et des séquences et numéros d'accension.**

\* indique la séquence *MLE* la plus longue et \*\* indiquent les séquences *MLE* les plus courtes.

Espèces	Longueur des séquences de <i>MLE</i>	Consensus	Nom des séquences	Présence d'un ORF ininterrompu	Numéros d'accension
<i>Amphora acutiuscula</i>	379	Aacutimar1	Aacutimar1.1	1	FN546185
	377		Aacutimar1.2	1	FN546186
	378		Aacutimar1.3	1	FN546187
	375		Aacutimar1.4	1	FN546188
	375		Aacutimar1.5	1	FN546189
	375		Aacutimar1.6	1	FN546190
<i>Amphora coffeaeformis</i>	396	Acomar1	Acomar1.1	1	FN546191
	396		Acomar1.2		FN546192
	396		Acomar1.3		FN546193
	401*	Acomar2	Acomar2.1	1	FN546194
	390	Acomar3	Acomar3.1	1	FN546195
378	Acomar4	Acomar4.1	1	FN546196	
<i>Entomoneis paludosa</i>	390	Entopalmar1	Entopalmar1.1	1	FN546197
	390		Entopalmar1.2	1	FN546198
	390	Entopalmar2	Entopalmar2.1	1	FN546199
	378	Entopalmar3	Entopalmar3.1	1	FN546200
	373	Entopalmar4	Entopalmar4.1	1	FN546201
<i>Haslea crucigera</i>	390	Hacrumar1	Hacrumar1.1	1	FN546202
	395		Hacrumar1.2	1	FN546203
	390		Hacrumar1.3	1	FN546204
	378	Hacrumar2	Hacrumar2.1	1	FN546205
	397	Hacrumar3	Hacrumar3.1	1	FN546206
<i>Nitzschia laevis</i>	378	Nilamar1	Nilamar1.1	1	FN546207
	378		Nilamar1.2	1	FN546208
	384	Nilamar2	Nilamar2.1	1	FN546209
	384	Nilamar3	Nilamar3.1	1	FN546210
	384	Nilamar4	Nilamar4.1	1	FN546211
	384	Nilamar5	Nilamar5.1	1	FN546212
<i>Nitzschia palea</i>	387	Nipamar1	Nipamar1.1	1	FN546213
	388		Nipamar1.2	1	FN546214
	387		Nipamar1.3	1	FN546215
	388		Nipamar1.4	1	FN546216
	387	Nipamar2	Nipamar2.1	1	FN546217
<i>Odontella aurita</i>	396	Odaumar1	Odaumar1.1	1	FN546218
	397		Odaumar1.2		FN546219
	396		Odaumar3.1		FN546220
	396		Odaumar4.1		FN546221
<i>Phaeodactylum tricornutum</i>	390	Phatrimar1	Phatrimar1.1	1	FN546222 FN546223
	390		Phatrimar1.2	1	
	380		Phatrimar1.3	1	
	390		Phatrimar1.4	1	
	390	Phatrimar2	Phatrimar2.1	1	FN546226

Espèces	Longueur des séquences de MLE	Consensus	Nom des séquences	Présence d'un ORF ininterrompu	Numéros d'accension
<i>Stauroneis constricta</i>	372**	Staucmar1	Staucmar1.1	1	FN546227
	372**		Staucmar1.2		FN546228
	372**		Staucmar1.3	1	FN546229
	372**		Staucmar1.4		FN546230
	381	Staucmar2	Staucmar2.1		FN546231
381	Staucmar2.2			FN546232	
381	Staucmar2.3			FN546233	
<i>Thalassiosira weissflogii</i>	381	Thaweissmar1	Thaweissmar1.1	1	FN546234
	379		Thaweissmar1.2	1	FN546235
	381		Thaweissmar1.4	1	FN546236
	381		Thaweissmar1.5	1	FN546237
	380		Thaweissmar1.6	1	FN546238
	381		Thaweissmar1.7	1	FN546239
<b>TOTAL</b>	<b>55</b>	<b>25</b>	<b>55</b>	<b>44</b>	<b>55</b>

Les consensus établis sur la base des identités nucléotidiques au sein de chaque espèce ont été traduits *in silico* en séquences protéiques. Un alignement de ces séquences traduites est présenté Figure 31. La taille des séquences protéiques varie de 124 à 132 acides aminés. Les motifs correspondants aux amorces *MLE5A* et *MLE3A*, soit IDEKWF et IQQDNA, sont présents à chaque extrémité, excepté dans la séquence consensus *Entopalmarcons4* qui présente dans la région terminale le motif tronqué IQQD.

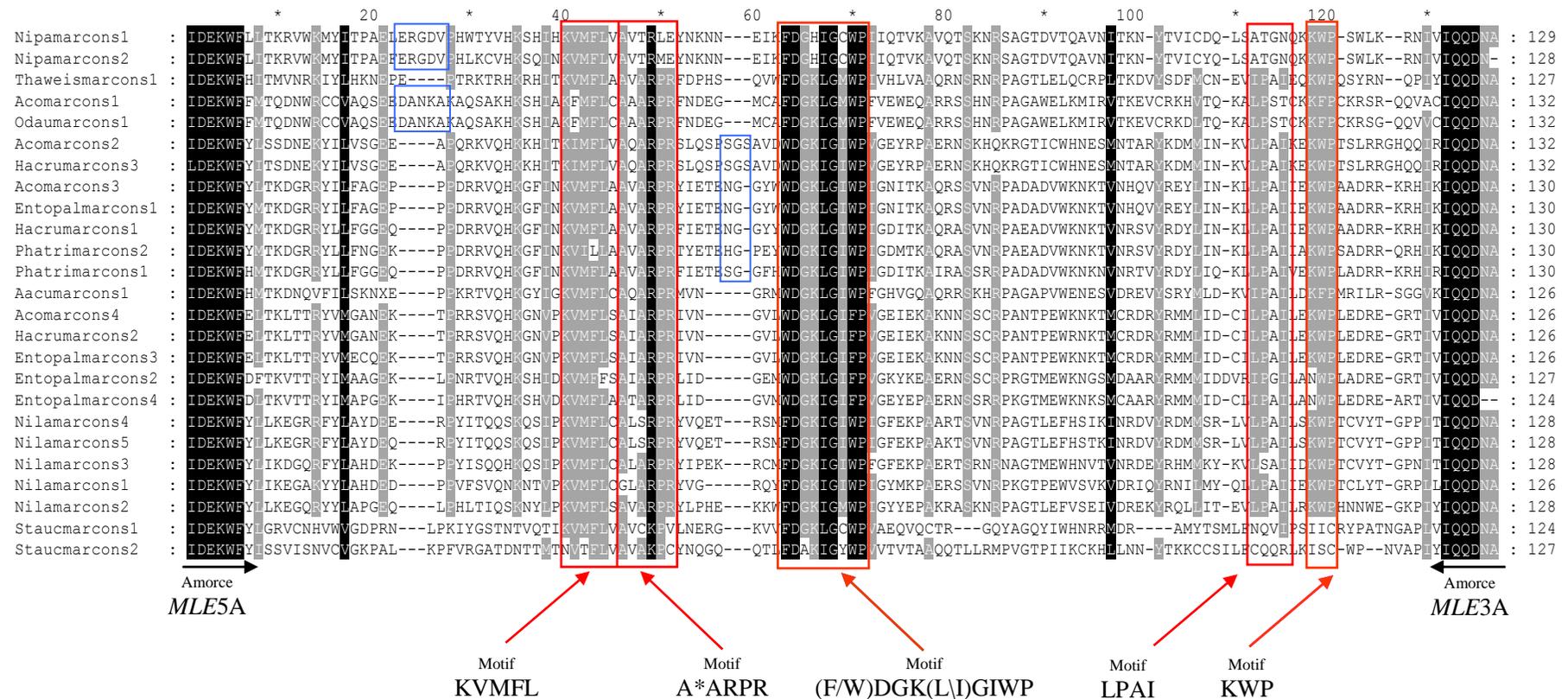
L'alignement montre une région centrale (entre les positions 41 et 72) relativement conservée comprenant trois motifs protéiques entre 5 et 9 acides aminés peu altérés entre les séquences et entre les espèces. Deux motifs sont conservés en partie terminale des consensus (entre les positions 112 et 121). Le premier motif bien conservé débute en position 40 et se compose de cinq acides aminés KVMFL. Ce motif est intact dans 20 des 25 séquences. Pour les cinq autres séquences, il n'y a qu'un acide aminé altéré sauf pour *Staucmarcons2* qui présente deux divergences. Le second motif est A\*ARPR à la position 46, il est parfaitement conservé dans 19 séquences. Chez deux espèces *N. palea* et *S. constricta* (soit quatre séquences consensus), ce motif est très altéré. Le motif le plus long est (F/W)GK(L/I)GIWP composé de sept acides aminés. Au début de ce motif, deux acides aminés peuvent être trouvés en première position soit la phénylalanine F soit le tryptophane W. De la même façon, deux acides aminés l'isoleucine I ou la leucine L sont présents alternativement en cinquième position. Ce motif est très conservé, seuls les consensus des espèces *S. constricta* et *N. palea* présentent des altérations de ce motif sur deux acides aminés maximum. Le motif LP AI en position 112 est totalement muté ou absent des consensus des espèces *S. constricta* et *N. palea*. Il est présent chez les 21 autres séquences mais est modifié pour *Acomarcon1*,

*Odaumarcons1* et *Nilamarcons3*. Le dernier motif est court et est composé seulement de trois acides aminés, KWP, il est présent dans 23 consensus.

L'alignement montre deux zones de divergence correspondant à des gaps pour 12 séquences sur les 25. En ce qui concerne les 13 autres séquences, des motifs variables imbriqués en position 23 pourraient correspondre à des introns (encadrés en bleu, Figure 31) pour *Nipamarcons1*, *Nipamarcons2*, *Acomarcons1* et *Odaumarcons1* et en position 56 pour *Acomarcons2*, *Hacrumarcons3*, *Acomarcons3*, *Entopalmarcons1*, *Hacrumarcons1*, *Phatrimarcons1* et 2.

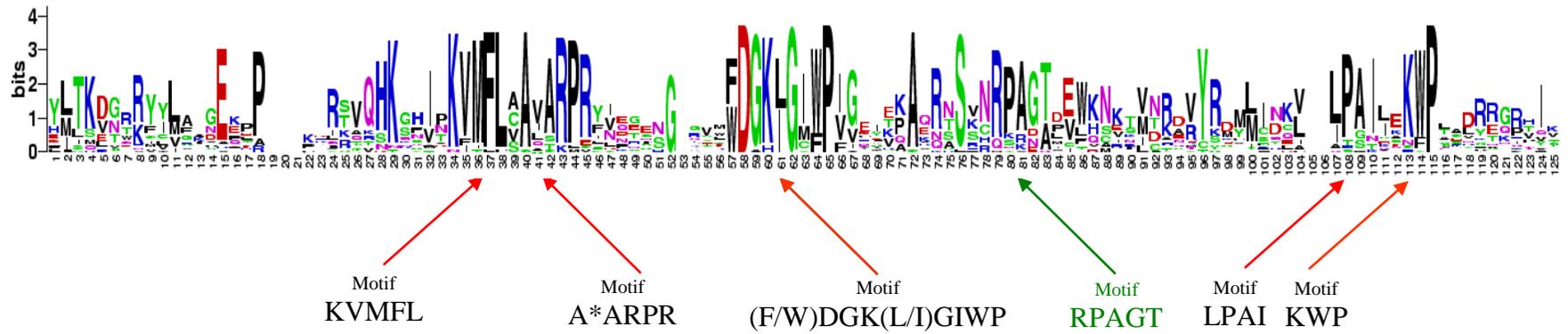
La part relative des différents motifs et la structuration de l'alignement autour de ces motifs conservés est mis en évidence dans la Figure 32. La représentation Weblogo ([http : //weblogo.berkeley.edu/](http://weblogo.berkeley.edu/)) affiche un alignement des séquences consensus traduites *in silico* auxquelles les amorces ont été retranchées. La taille de la lettre est proportionnelle à la fréquence de l'acide aminé à chaque position sur la séquence. Les motifs conservés apparaissent en lettres de grande taille ainsi que les acides aminés les plus fréquents. Les motifs conservés KVMFL, A\*ARPR, (F/W)GK(L/I)GIWP, LPAI et KWP chez les *MLE* de diatomées prédominent. Un nouveau motif RPAGT apparaît, relativement bien conservé en position 79-83, il est indiqué en vert sur la Figure 32.

Les séquences consensus de *MLE* traduites *in silico* sont polymorphes, toutefois les motifs sont conservés entre les 25 séquences *MLE* provenant de 10 espèces différentes de diatomées marines. Ceci suggère que la région catalytique est conservée chez ces organismes. Une analyse phylogénétique de ces séquences est réalisée afin de mettre en évidence la distribution et la variabilité des *MLE* entre les espèces de diatomées.



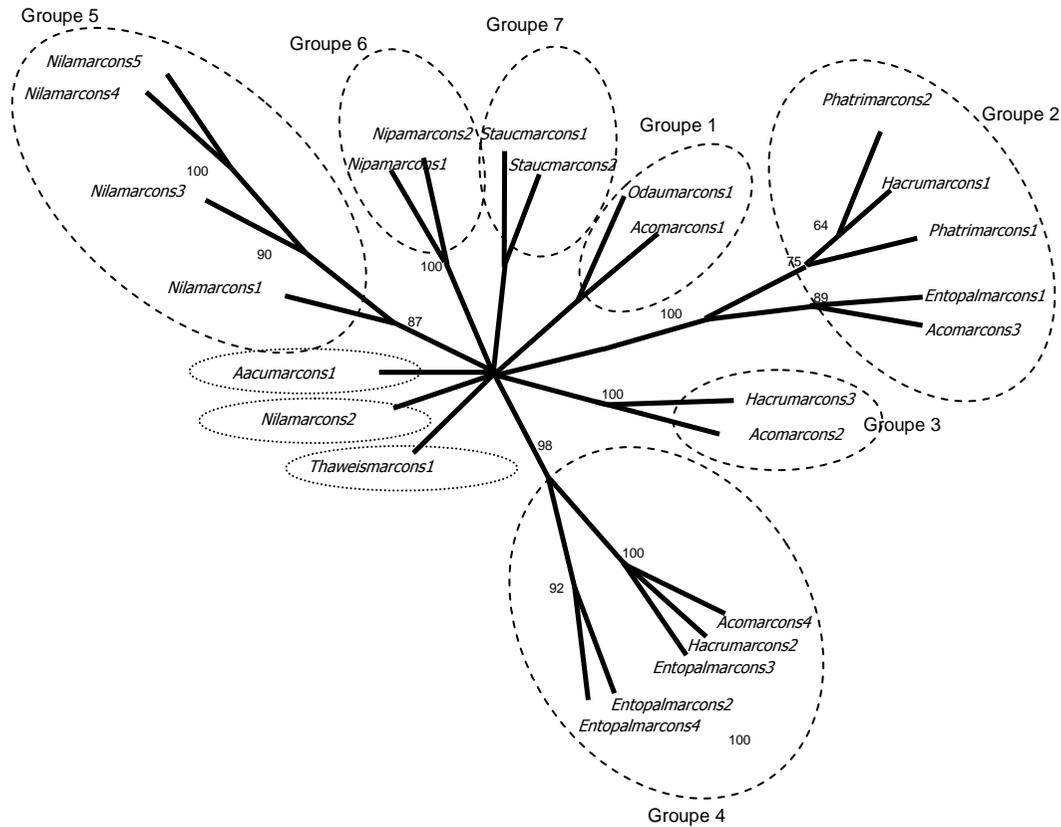
**Figure 31 : Alignement des séquences consensus MLE obtenues chez les dix diatomées et traduites *in silico*.**

Les séquences présentées dans cet alignement sont issues de la traduction *in silico* des consensus de MLE des diatomées marines. L'alignement présente des zones conservées (100 %) ombrées en noir ainsi que des zones de similitudes (80 %) en grisé. Les motifs conservés sont encadrés en rouge. Les motifs supplémentaires sont encadrés en bleu. Les amorces sont indiquées à chaque extrémité des séquences. A la fin de chaque séquence la longueur de la séquence en acide aminés est indiquée après le symbole :



**Figure 32 : Représentation en WebLogo de l'alignement des consensus traduits en acides aminés.**

Sont représentées les séquences consensus de *MLE* traduites *in silico*, obtenues chez les dix diatomées. Les amorces ont été retirées pour ne révéler que les motifs conservés. En abscisse est indiquée la position des acides aminés et en ordonnée, la fréquence de chaque acide aminé est indiquée en bits. La position des motifs conservés est indiquée par une flèche. Le motif RPAGT est indiqué en vert.



**Figure 33 : Analyse en Maximum de Parcimonie (MP) des séquences *MLE* consensus de diatomées traduites *in silico*.**

L'arbre consensus obtenu à partir de la majorité des arbres générés est représenté (PAUP). Les valeurs de bootstrap obtenues après 1000 répliques sont indiquées en pourcentage, seules les valeurs supérieures à 50 % sont données. Sept groupes sont constitués, sur la base des valeurs de bootstrap. Les groupes 1, 2, 3, 4 sont hétérogènes, et composés de séquences *MLE* provenant de différentes espèces de diatomées, et les groupes 5, 6, 7 sont des groupes homogènes et composés uniquement de séquences provenant d'une seule et même espèce.

	Groupe 1		Groupe 2				Groupe 3			Groupe 4				Groupe 5				Groupe 6		Groupe 7					
	<i>Acomarcons1</i>	<i>Odaumarcons1</i>	<i>Acomarcons3</i>	<i>Entoplamarcons1</i>	<i>Hacrumarcons1</i>	<i>Phatrimarcons1</i>	<i>Phatrimarcons2</i>	<i>Aacumarcons1</i>	<i>Acomarcons2</i>	<i>Hacrumarcons3</i>	<i>Acomarcons4</i>	<i>Hacrumarcons2</i>	<i>Entopalmarcons3</i>	<i>Entopalmarcons2</i>	<i>Entopalmarcons4</i>	<i>Nilamarcons4</i>	<i>Nilamarcons5</i>	<i>Nilamarcons3</i>	<i>Nilamarcons1</i>	<i>Nilamarcons2</i>	<i>Thaweismarcons1</i>	<i>Nipamarcons1</i>	<i>Nipamarcons2</i>	<i>Staucmarcons1</i>	<i>Staucmarcons2</i>
<i>Acomarcons1</i>	96		34	35	35	35	32	36	27	28	29	28	29	28	30	32	30	35	29	31	36	28	27	23	18
<i>Odaumarcon1</i>		96	34	35	35	35	32	36	26	27	30	29	30	28	30	32	30	35	30	32	36	27	26	24	18
<i>Acomarcons3</i>			99	89	81	80		49	45	45	51	50	50	45	47	43	42	47	46	45	41	28	28	26	20
<i>Entoplamarcons1</i>				90	82	81		50	44	45	50	50	50	45	46	42	41	46	45	44	41	27	28	25	20
<i>Hacrumarcons1</i>					88	88		50	44	45	50	50	50	43	45	43	43	46	45	45	41	26	27	26	19
<i>Phatrimarcons1</i>						78		46	42	43	47	46	47	42	45	42	41	45	44	44	38	26	27	22	19
<i>Phatrimarcons2</i>							49	43	44	51	50	51	45	48		40	40	43	41	44	39	26	27	22	18
<i>Aacumarcons1</i>								42	43	41	41	40	42	40		40	38	43	37	39	41	23	22	24	17
<i>Acomarcons2</i>									98	41	41	41	45	48		40	39	43	39	39	39	24	23	28	21
<i>Hacrumarcons3</i>										41	41	41	46	48		40	38	42	38	38	40	24	24	27	21
<i>Acomarcons4</i>											99	97	63	68		40	40	43	43	43	35	22	20	25	20
<i>Hacrumarcons2</i>												96	62	67		40	40	43	43	43	35	22	20	25	19
<i>Entopalmarcons3</i>													62	68		40	40	43	43	43	34	22	20	24	19
<i>Entopalmarcons2</i>														80		39	38	42	37	41	35	23	20	26	17
<i>Entopalmarcons4</i>															40	39	43	41	44	37	26	23	26	21	
<i>Nilamarcons4</i>																97	72	58	56	41	28	27	23	21	
<i>Nilamarcons5</i>																	71	57	56	40	28	27	23	21	
<i>Nilamarcons3</i>																		60	50	38	28	28	25	19	
<i>Nilamarcons1</i>																			54	33	28	28	24	21	
<i>Nilamarcons2</i>																				41	29	28	27	19	
<i>Thaweismarcons1</i>																					32	33	26	26	
<i>Nipamarcons1</i>																						92	19	19	
<i>Nipamarcons2</i>																							19	19	
<i>Staucmarcons1</i>																								24	
<i>Staucmarcons2</i>																									

**Figure 34 : Matrice d'identité des séquences consensus de *MLE* traduites *in silico*.**

L'homologie entre les séquences *MLE* consensus des diatomées est exprimée en pourcentage. Les séquences *MLE* consensus ont été traduites *in silico* et alignées à l'aide du programme ClustalW. La matrice a été réalisée avec le logiciel GENEDOC.

#### I.4 Analyse phylogénétique des *MLE* de diatomées marines

L'analyse phylogénétique des 25 séquences de *MLE* issues des diatomées marines est réalisée dans le but de positionner ces séquences les unes par rapport aux autres. Il ne s'agit pas ici de mettre en évidence des liens de parenté, car seule une partie de la séquence de la transposase a été isolée. Or les informations présentes dans le reste du domaine catalytique et dans le signal de localisation nucléaire (inconnues) par exemple peuvent fournir des informations à l'établissement de relations de parenté.

L'analyse phylogénétique présentée ici va permettre de visualiser la distribution des *MLE* chez les espèces de diatomées et de détecter le polymorphisme de séquences.

La méthode du Maximum de Parcimonie (MP) a été utilisée l'aide du programme PAUP (version 4.b1, Swofford 2002) pour positionner les séquences protéiques des *MLE* dans un arbre non enraciné. L'arbre consensus obtenu (sur la majorité des arbres générés) est présenté en Figure 33 (p 101). Les séquences sont composées de 135 caractères dont 126 sont des caractères informatifs.

En parallèle, une matrice d'identité a été réalisée à l'aide du logiciel GENEDOC après alignement par CLUSTALW, cette fois dans le but de quantifier le polymorphisme des séquences *MLE* des diatomées marines (Figure 34, p 102).

L'arbre réalisé avec la méthode du maximum de Parcimonie (MP) révèle la formation de sept groupes (Figure 33). Ces groupes sont supportés par de fortes valeurs de robustesse au niveau des nœuds 100, 100, 98, 87 et 100 %, respectivement pour les groupes 2, 3, 4, 5 et 6, excepté pour les groupes 1 et 7 qui présentent des valeurs de robustesse des noeuds inférieures à 50 % (Figure 33). Ces valeurs ont été obtenues après 1000 réplifications du test de bootstrap.

Les valeurs de robustesse des nœuds renseignent sur le degré de confiance statistique de la distribution de l'échantillon. Le test de bootstrap permet de vérifier que la distribution du groupe n'est pas réalisée au hasard. En d'autres termes, cette valeur permet d'estimer la validité d'un groupe produit par la méthode phylogénétique. On considère que les valeurs de bootstrap aux noeuds supérieures à 75 % valident la formation du groupe considéré.

Les groupes 1, 2, 3 et 4 sont composés de séquences *MLE* provenant d'espèces différentes, ils sont donc qualifiés de groupes hétérogènes, leur composition est telle que :

Le groupe 1 est constitué par la séquence consensus de *O. aurita*, *Odaumarcons1* et d'une des séquences consensus de *A. coffeaeformis*, *Acomarcons1*. Ce groupe présente des

valeurs faibles de bootstrap aux nœuds, inférieures à 50 %. Ces deux séquences protéiques partagent 96 % d'identité entre elles et moins de 36 % d'identité avec les 23 autres séquences (Figure 34). Elles sont donc réunies dans le groupe 1.

Le groupe 2 est supporté par une valeur de bootstrap de 100 %. Il est constitué par cinq séquences *MLE* qui partagent entre 88 et 99 % d'identité entre elles et qui proviennent de quatre espèces de diatomées différentes (*P. tricornutum*, *H. crucigera*, *E. paludosa*, et *A. coffeaeformis*). Ces séquences partagent moins de 50 % d'identité avec les 20 autres séquences *MLE* (Figure 34). Le groupe 2 peut être divisé en deux sous-groupes :

- Le premier sous-groupe est supporté par une valeur de nœud de 75 %, et il englobe deux séquences consensus de *P. tricornutum*, *Phatrimacons1* et *Phatrimarcons2* ainsi que la première séquence consensus de *H. crucigera*. Ces séquences ont un taux d'identité compris entre 78 et 88 %.
- Le deuxième sous-groupe présente une valeur de robustesse du nœud de 89 % et il est constitué par la première séquence consensus provenant de *E. paludosa*, *Entopalmarcons1*, et la troisième séquence consensus de *A. coffeaeformis*, *Acomarcons3*. Ces deux séquences partagent 99 % d'identité.

Les deux sous-groupes sont réunis au sein du groupe 2 du fait du fort pourcentage d'identité (90 %) entre *Hacrumarcons1* (premier sous-groupe) et *Entopalmarcons1* (deuxième sous-groupe).

Le groupe 3 n'est composé que de deux séquences provenant de *H. cucigera*, *Hacrumarcons3* et *A. coffeaeformis*, *Acomarcons2*. Ce groupe est supporté par une valeur au nœud de 100 %. Les séquences partagent 98 % d'identité quand elles ne présentent que des taux d'identité inférieurs à 45 % avec les 23 autres séquences, ce qui conforte la constitution du groupe (Figure 34).

Le groupe 4 est supporté par une valeur de robustesse de 98 % au nœud et les pourcentages d'identité entre les séquences qui le composent sont compris entre 62 et 99 %. Il peut être divisé en deux sous-groupes :

- Le premier sous-groupe est constitué des trois séquences consensus *MLE*, *Acomarcons4*, *Hacrumarcons2*, et *Entopalmarcons3* provenant de trois espèces *A. coffeaeformis*, *H. crucigera*, *E. paludosa* respectivement. Ce sous-groupe est supporté par des valeurs de nœud de 100 % et un fort pourcentage d'identité entre les séquences de 96 à 99 % (Figure 34).
- Le deuxième sous-groupe est composé par deux séquences provenant de l'espèce *E. paludosa*. Ces séquences sont *Entopalmarcons2* et *Entopalmarcons4* qui partagent 80 % d'identité. La valeur de nœud est de 92 %.

Les groupes 5, 6 et 7 sont constitués de séquences provenant d'une seule espèce, ils sont dits « homogènes » (Figure 33).

Le groupe 5 est composé par quatre des cinq séquences *MLE* consensus de l'espèce *N. laevis* et il est supporté par une valeur de nœud de 87 % et des pourcentages d'identité variant de 60 à 97 %. Ces séquences sont *Nilamarcons1*, *Nilamarcons3*, *Nilamarcons4* et *Nilamarcons5*. Les deux dernières séquences citées présentent un taux d'identité de 97 % et constituent un groupe terminal supporté également par une forte valeur de nœud de 100 %.

Le groupe 6 se compose des deux séquences *MLE* consensus de l'espèce *N. palea*, *Niapamarcons1* et *Nipamarcons2* qui sont identiques à 92 % et ne partagent pas plus de 30 % d'identité avec les 23 autres séquences *MLE* de diatomées.

Le groupe 7 est également constitué des deux séquences obtenues chez *S. constricta*, *Staucmarcons1* et *Staucmarcons2* qui ne partagent que 24 % d'identité entre elles. Ce groupe présente des valeurs de nœud inférieures à 50 %. Ceci semble cohérent avec le fait que *Staucmarcons1* partage un taux d'identité de 28 % avec *Acomarcons2* présente dans le groupe 3. Ce pourcentage d'identité est supérieur à celui que partagent les deux séquences *MLE* de *S. constricta*. La réunion de ces deux séquences semble être le résultat d'une variabilité importante des séquences de *S. constricta* par rapport aux autres séquences *MLE* de diatomées (Figure 34).

Les séquences isolées (Figure 33) sont :

- *Thaweismarcons1*, seule séquence consensus *MLE* de l'espèce *T. weissflogii*.
- *Aacumarcons1* qui est également la seule séquence consensus de l'espèce *A. acutiuscula*.
- *Nilamarcons2* de l'espèce *N. laevis*. Cette séquence ne s'insère donc pas dans le groupe 5 constitué des quatre autres séquences *MLE* consensus de *N. laevis*. Ceci suggère que l'échantillonnage en séquences chez cette espèce est sans doute trop faible. Avec l'ajout de séquences, *Nilamarcons2* pourrait sans doute s'insérer dans le groupe 5.

Les valeurs des nœuds à la base de chacune des séquences citées ci-dessus sont inférieures à 50 % et chacune de ces séquences présente un pourcentage d'identité inférieur à 50 % avec le reste des séquences *MLE* de diatomées (Figure 34). Ceci permet d'expliquer la topologie des séquences isolées dans l'analyse phylogénétique. En effet, ces séquences sont tellement divergentes des autres qu'elles ne peuvent être incluses dans aucun des groupes formés.

L'analyse phylogénétique a permis de mettre en évidence des groupes hétérogènes, c'est-à-dire composés de séquences *MLE* provenant d'espèces différentes (groupe, 1, 2, 3 et 4) et des groupes homogènes composés de séquences *MLE* provenant de la même espèce

(groupes 5, 6, et 7). Parmi les groupes homogènes, le groupe 5, bien qu'uniquement constitué de séquences provenant de *N. laevis* ne contient pas toutes les séquences de cette espèce, la séquence *Nilamarcons2* se retrouve isolée.

Les *MLE* de diatomées ne sont pas distribués par espèce ou par genres au sein des groupes. Les groupes hétérogènes présentent des valeurs de bootstrap aux nœuds supérieures à 90 % (excepté pour le groupe 1), ce qui confirme la validité de ces groupes. Ces groupes de séquences semblent donc avoir évolué de manière indépendante. En effet, les deux espèces *A. coffeaeformis* et *A. acutiuscula* appartiennent toutes deux au genre *Amphora*, de même les deux espèces *N. palea* et *N. laevis* appartiennent au genre *Nitzschia* mais leurs séquences sont polymorphes et distribuées dans des groupes différents. Toutefois, aucun groupe composé de séquences provenant d'un même genre n'a été observé. De plus, les taux d'identité permettent de confirmer, dans tous les cas, la formation de ces groupes hétérogènes. Nous pouvons en déduire que les *MLE* ne présentent pas de particularités intraspécifiques chez les différentes diatomées marines. Seul le groupe de séquences provenant de *N. palea* (groupe 6) semble porter des caractéristiques intrinsèques à l'espèce.

Les groupes 1, 7 et les séquences isolées semblent moins définies puisque les valeurs de robustesse des nœuds sont inférieures à 50 %. C'est le cas du deuxième consensus de *N. laevis*, *Nilamarcons2*, qui n'est pas inclus dans le groupe 5 et reste isolé. Bien que cette séquence présente un taux d'identité plus important avec les autres séquences de *N. laevis* (entre 50 et 56 %) par rapport aux autres séquences (<45 % d'identité), elle n'est pas incluse dans le groupe 5. Toutefois la valeur de bootstrap au nœud est inférieure à 50 %, ce qui laisse penser que sa place n'est pas clairement définie. Ces résultats signifient que les séquences présentent un polymorphisme tel qu'elles ne sont pas classables dans cet échantillon. Il est également possible que ces faibles valeurs de confiance soient le résultat d'une ambiguïté entre plusieurs groupes ; dans ce cas, elles sont isolées pour ne pas invalider les groupes préformés par la méthode de parcimonie.

## I.5 Recherche des *MLE* dans le génome séquencé de *Phaeodactylum tricornutum* (souche CCMP632 /CCAP1055.1)

Le génome de *Phaeodactylum tricornutum* est séquencé et disponible sur internet. Le clone de *P. tricornutum* CCAP1055/1 séquencé, a été isolé depuis une culture monoclonale de la souche CCMP632 (De Martino et al., 2007) et présente des éléments transposase-like qui ont été détectés par Maumus et al. (2009).

### I.5.1 Détection bioinformatique des *MLE* dans la banque de données JGI

La présence de *MLE* a été recherchée par Blastn et Blastx dans le génome de *P. tricornutum* CCAP1055.1 sur le site du Joint Genome Institute. Pour cela, les séquences consensus obtenues dans la partie I.3 de ce chapitre, pour chaque espèce ont été utilisées, sans donner de résultats positifs. Les 55 séquences de *MLE* isolées chez les 10 espèces de diatomées ont également été utilisées pour effectuer cette recherche en incluant les séquences caractérisées chez *P. tricornutum* NCC340 (souche cultivée au laboratoire). Les résultats des analyses ne montrent aucune homologie entre les séquences de *MLE* dont nous disposons et le génome de *P. tricornutum* CCMP632/CCAP1055.1.

Parallèlement, les séquences de trois *MLE* complets de plantes, *Soymar1* (Jarvik et Lark 1998) isolé chez le soja, *Osmar1* (Tarchini et al., 2000) trouvé chez le riz et *Vulmar1* (Jacobs et al., 2004) découvert chez la betterave, ont également été utilisés pour interroger les banques de données du génome de *P. tricornutum*.

Aucune similarité de séquence n'a été détectée entre les *MLE* des diatomées ou les *MLE* complets végétaux et le génome de la diatomée séquencée, ce qui signifie qu'aucun élément proche des *MLE* de plantes n'est présent dans le génome séquencé de *P. tricornutum* CCMP632/CCAP1055.1.

### I.5.2 Recherche bioinformatique à l'aide des séquences signatures (*Target Site Duplication*) *TSD* dans le génome de *P. tricornutum*

Les *MLE* sont des transposons à TIR (Wicker et al., 2007) dont la transposition montre des caractéristiques particulières, notamment la duplication d'un dinucléotide TA lors de l'insertion dans un nouveau locus. Ceci est nommé duplication du site cible ou TSD (*Target Site Duplication*).

Le génome de la souche CCAP1055.1 de *P. tricornutum* a été téléchargé au laboratoire et donc utilisé localement. Le génome de *P. tricornutum* se compose de 33 chromosomes, du génome chloroplastique, du génome mitochondrial, de 55 « scaffolds »

(ensemble de contigs orientés et ordonnés avec des gaps de longueurs connues) et d'une bibliothèque de séquences répétées. Ceci correspond à environ 30 Mpb.

Les *MLE* ont été recherchés suivant les caractéristiques de la structure générale des *MLE* de végétaux : ils possèdent des TIR d'environ 30 pb, leur séquence totale varie en moyenne entre 3-5 Kpb et ils s'insèrent dans un site cible (TSD) composé d'un dinucléotide TA qui est dupliqué et se retrouve à chaque extrémité du transposon.

Ainsi, toutes les séquences de 1 à 5 Kpb encadrées par des TIR (imparfaits jusqu'à 9 erreurs) inclus entre deux TA ont été extraites. Au total, 105 séquences ont été obtenues dont 98 dans les chromosomes et 7 séquences dans les « scaffolds » (Tableau 11). Aucune séquence TIR n'a été détectée dans les génomes chloroplastiques et mitochondriaux. L'ensemble des séquences TIR a été confronté aux 55 séquences *MLE* de diatomées sans trouver de similitudes. Les 105 séquences TIR de *P. tricornutum* ont également été confrontées aux deux banques de données NCBI et Repbase par Blastn et Blastx, et aucune correspondance n'a été trouvée par comparaison.

Les résultats obtenus par Blastx contre la base de données NCBI sont synthétisés dans le Tableau 11. Les correspondances obtenues ne font référence qu'à des fragments protéiques putatifs de *P. tricornutum*. Toutefois, aucune séquence n'a été identifiée comme étant un ET, excepté au niveau du chromosome 3. En effet, sur ce chromosome, une séquence de 1 340 pb encadrée par des TIR de 23 pb montre une correspondance avec un fragment du rétrotransposon *Ty1/copia* de *P. tricornutum* mis en évidence par Maumus et al. (2009).

Les méthodes bio-informatiques de détection des *MLE*, que ce soit par Blast ou par recherche des TIR et TSD, n'ont pas permis de mettre en évidence la présence de *MLE* dans le génome séquencé de *P. tricornutum*.

**Tableau 11 : Séquences obtenues par recherche bio-informatique au sein du génome de *Phaeodactylum tricornutum*.**

Les séquences ont été détectées par la méthode de recherche des TIR. Les caractéristiques des analyses réalisées sont : détection de séquences de 1 à 5 Kpb encadrées par des TIR d'au moins 20 pb (imparfaits jusqu'à 9 erreurs) et présentant un TA dupliqué. La longueur des séquences et des TIR est exprimée en pb. Les / représentent une absence d'identification.

Chromosome	Intervalle de longueur des séquences	Nombre de séquences	Longueur des TIR	Identification par Blastx
3	1000-2000	7	20-188	(1 Fragment de Rétrotransposon Ty1-copia)
	2000-5000	5	81-151	/
4	1000-2000	3	21-23	/
	2000-5000	2	60-216	/
6	1000-2000	0	/	/
	2000-5000	3	44-844	/
8	1000-2000	2	102-184	/
	2000-5000	1	100	/
9	1000-2000	1	660	/
	2000-5000	1	1722	/
10	1000-2000	4	23-120	/
	2000-5000	1	229	/
11	1000-2000	4	53-265	/
	2000-5000	2	80-232	/
12	1000-2000	6	95-190	/
	2000-5000	3	47-1190	/
13	1000-2000	1	335	/
	2000-5000	2	62-81	/
14	1000-2000	3	54-370	/
	2000-5000	3	35-900	/
15	1000-2000	1	100	/
	2000-5000	0	/	/
16	1000-2000	3	24-153	/
	2000-5000	1	78	/
17	1000-2000	1	55	/
	2000-5000	1	190	/
18	1000-2000	2	157-218	/
	2000-5000	1	560	/
19	1000-2000	2	112-843	/
	2000-5000	1	1450	/
20	1000-2000	0	/	/
	2000-5000	1	127	/
21	1000-2000	2	74-210	/
	2000-5000	1	550	/
23	1000-2000	0	/	/

	2000-5000	1	200	/
24	1000-2000	4	23-90	/
	2000-5000	1	135	/
25	1000-2000	3	56-190	/
	2000-5000	1	108	/
26	1000-2000	2	72-407	/
	2000-5000	1	32	/
27	1000-2000	2	106-188	/
	2000-5000	2	467-820	/
29	1000-2000	2	39-63	/
	2000-5000	2	39-179	/
30	1000-2000	1	170	/
	2000-5000	1	346	/
31	1000-2000	2	69-96	/
	2000-5000	2	84-192	/
Chloroplaste	1000-2000	0	/	/
	2000-5000		/	/
Mitochondrie	1000-2000	0	/	/
	2000-5000		/	/
Scaffolds	1000-2000	7	23-500	/
	2000-5000			/
Total		105		1

### I.5.3 Recherche bio-moléculaire de séquence MLE dans le génome de *P. tricornutum*.

Les résultats de la recherche de MLE dans le génome séquencé de *P. tricornutum* à l'aide des méthodes bio-informatiques étant négatifs, une vérification expérimentale par biologie moléculaire a été réalisée.

Le génome présent dans les banques de données est celui de la souche CCAP1055.1, elle-même issue d'un clone de *P. tricornutum*, souche CCMP632 (Bowler et al., 2008; Maumus et al., 2009). Il s'agit en fait de deux souches identiques enregistrées dans deux algorithmes différentes.

A l'aide des amorces dégénérées *MLE5A* et *MLE3A*, des PCR ont été réalisées pour détecter la présence de fragments *MLE* de 380 pb (Figure 35). De même, les amorces spécifiques *Phaeo4* et *Phaeo7* définies sur la séquence *Phatrimar1.3* ont été utilisées pour amplifier des fragments de 170 pb (Figure 36). Ces PCR ont été réalisées sur l'ADN extrait de la souche séquencée de *P. tricornutum* CCMP632 et de la souche de *P. tricornutum* NCC340 du laboratoire.



Les figures 35 et 36 présentent les résultats obtenus par PCR avec les deux souches de *P. tricornutum*, la souche NCC340 cultivée au laboratoire de l'université du Maine et la souche CCMP632.

Des amplifications sont observées avec les amorces dégénérées *MLE5A* et *MLE3A*, à la taille attendue de 380 pb, pour les deux souches de *P. tricornutum* testées. Le profil d'amplification diffère entre la souche NCC340 du laboratoire et la souche CCMP632 (Figure 35). Ces résultats indiquent que les génomes diffèrent entre les souches. L'amplification de fragments de *MLE* spécifiques obtenus avec les amorces *Phaeo4* et *Phaeo7* uniquement chez la souche du NCC340 de *P. tricornutum* suggère que l'élément *Phatrimar1.3* n'est pas présent dans la souche CCMP632.

Ces résultats nous permettent d'en déduire que la souche séquencée CCMP632 et la souche NCC340 cultivée au laboratoire n'ont pas la même composition en *MLE*.

Les séquences de 380 pb amplifiées avec les amorces *MLE5A* et *MLE3A* dans les deux souches de *P. tricornutum* CCMP632 et NCC340 (Figure 37) ont été séquencées. L'identification par confrontation aux bases de données NCBI et Repbase montre que ces séquences sont similaires aux *MLE* végétaux.

L'analyse et l'alignement de ces séquences *MLE* révèlent des similitudes avec les gènes codant les transposases, *Phatrimarcons1* et *2* mises en évidence au début de ce chapitre dans la souche *P. tricornutum* NCC340 du laboratoire. En effet, deux clones ont été séquencés pour la souche CCMP632, les séquences nucléiques obtenues nommées *PtCCMP632.mar1* et *PtCCMP632.mar2* présentent 99 % d'identité entre elles.

Les séquences protéiques traduites *in silico* de *PtCCMP632.mar1* et *PtCCMP632.mar2* n'ont que 60 % de similarité avec les séquences consensus protéiques *MLE* de *P. tricornutum* NCC340 (Figure 37). L'alignement des séquences *PtCCMP632.mar1* et *PtCCMP632.mar2* avec les séquences *Phatrimarcons1* et *Phatrimarcons2* met en évidence des régions conservées. Cinq des six motifs conservés chez les *MLE* de diatomées présentent de légères modifications chez les éléments identifiés dans la souche séquencée de *P. tricornutum* (Figure 37). En effet, le premier motif KVMFL devient KIMFL, le deuxième motif A\*ARPR devient AVARPR, le troisième motif (F/W)DGK(L/I)GIWP reste intact, le quatrième motif LPAI devient VLPAI pour la souche CCMP632 et le motif final KWP devient KFP. Toutefois, le sixième motif RPAGT n'est pas présent. Ces modifications légères ne concernent qu'un acide aminé dans chacun des motifs. Ceci confirme la conservation de ces sites qui pourraient avoir une fonction dans la transposase *MLE*. De plus, les séquences de

la souche CCMP632 ont produit des traductions *in silico* sans codon stop dès la première phase de lecture, comme obtenu usuellement avec les séquences de *MLE* des diatomées.

## I.6 Discussion

### I.6.1 Détection de *MLE* chez les diatomées marines

La recherche des *MLE* chez les diatomées a été réalisée à l'aide des amorces *MLE5A* et *MLE3A* définies par Feschotte et Wessler (2002). Ces amorces nous ont permis d'amplifier des séquences d'environ 380 pb chez 10 des 18 espèces de microalgues testées. Ces 10 espèces appartiennent toutes à la classe des Bacillariophycées, les autres classes ne semblent pas présenter de *MLE* ou bien elles sont tronquées ou mutées. Bien qu'elles soient dégénérées, les amorces *MLE5A* et *MLE3A* ont permis d'obtenir des séquences identifiées comme des *MLE* par la comparaison avec les banques de données révélant des similarités avec des fragments de gènes codant la transposase *MLE* végétale. L'utilisation de ces amorces pour déceler la présence de *MLE* dans les génomes des plantes et ici, dans les génomes de diatomées marines, permettent d'étendre notre connaissance des *MLE* végétaux. En effet, celle-ci reste pauvre en regard de la quantité de *MLE* décrits chez les animaux (Robertson, 1995; Shao et Tu., 2001; Brillet et al., 2007; Bui et al., 2007, 2008; Delaurière et al., 2009) avec un peu plus de 200 séquences de *MLE* identifiées chez des espèces végétales en grande majorité tronquées et restreintes à 380 pb alors que l'on connaît près d'un millier de séquences *MLE* identifiées chez les espèces animales dont près de 50 % sont des éléments complets.

En effet, depuis la définition de ces amorces en 2002, Feschotte et Wessler ont pu mettre en évidence la présence des *MLE* chez 37 espèces d'Angiospermes. Jacobs et al. (2004) les ont employées pour détecter des *MLE* chez quatre espèces du genre *Beta* dont la betterave, *Beta vulgaris*. De même chez le pois, *Pisum sativum*, des séquences *MLE* de 380 pb ont été mises en évidence à l'aide de ces amorces (Macas et al., 2005). Des *MLE* tronqués (380 pb) ont également été identifiés chez 44 espèces de bambous à l'aide de ces amorces (Zhou et al., 2010). Toutefois, ces amorces peuvent également donner des résultats aspécifiques. Ainsi, Feschotte et Wessler (2002) rapportent que les amorces dégénérées leur ont permis de détecter des amplifications chez 54 sur 91 espèces d'Angiospermes testées dont certaines n'ont pas révélées de correspondance avec les bases de données. De manière similaire, chez les 18 espèces de microalgues testées, des amplifications ont été observées, pourtant chez huit espèces, les amplifications n'ont pas montré de similitudes avec les séquences des bases de données. De plus, de nombreuses amplifications à des tailles

inférieures et supérieures à 380 pb sont observées chez les 10 espèces testées positivement pour la présence de *MLE*. Ces séquences ne présentent pas d'homologie avec des gènes de transposase *MLE*, comme le mentionnent également Feschotte et Wessler (2002) chez les Angiospermes.

L'ensemble des *MLE* de végétaux identifiés jusqu'ici avait été obtenu chez des Angiospermes. La mise en évidence des *MLE* chez les diatomées montre ainsi la constance des motifs sur lesquels ont été définies les amorces dans la lignée végétale. En effet, les amorces *MLE* ont été définies à partir de motifs conservés encadrant une partie de la triade catalytique DD39D putative des *MLE* de plantes. Ces motifs sont positionnés au niveau des deux premiers résidus aspartate (D) de cette triade. Ainsi, le fait que ces amorces permettent une amplification des *MLE* chez divers génomes végétaux indique que la triade catalytique est conservée dans la lignée végétale. Cependant, cela implique également que des *MLE* tronqués au niveau des deux premiers résidus aspartate de la triade catalytique ne seront pas détectés et amplifiés. De ce fait, l'absence d'amplification chez certaines espèces de diatomées ne signifie pas nécessairement que les *MLE* ne sont pas présents dans ces génomes. La présence de *MLE* détectée chez 75 % des espèces de diatomées testées, qu'elles soient de morphologie, de genres variés (centriques ou pennées) ou encore qu'elles aient des modes de vie différents (benthique, épipélique, planctonique), suggère que les *MLE* sont sans doute présents de manière générale dans le génome des Bacillariophycées.

Deux génomes de diatomées sont actuellement séquencés : le génome de la diatomée centrique *Thalassiosira pseudonana* (Armbrust et al., 2004), et celui de la diatomée pennée *Phaeodactylum tricorutum* qui vient d'être publié (Bowler et al., 2008). Le génome séquencé de *T. pseudonana* ne contient pas de *MLE* détectable, ce qui a été confirmé expérimentalement par notre approche moléculaire avec les amorces *MLE5A* et *3A*. Bien qu'aucun *MLE* n'ait été détecté que ce soit dans le génome ou dans les banques de transcrits (banque d'EST) de ces deux diatomées (Maumus et al., 2009), nous avons pu mettre en évidence des *MLE* par l'approche moléculaire chez *P. tricorutum*. Cependant, il faut noter que la souche de *P. tricorutum* séquencée et la souche présente au laboratoire ont des origines géographiques différentes. En effet, le génome séquencé est issu d'un clone de *P. tricorutum* dont la souche a été isolée dans la région de Blackpool (Royaume-Uni) en 1956 (De Martino et al., 2007). La souche de l'université du Maine, quant à elle, a été isolée dans les années 1990 dans la Baie de Bourgneuf (France). Cette origine différente ainsi que la culture prolongée au laboratoire de clones pourraient expliquer l'absence de *MLE* dans le génome séquencé. On connaît encore mal la variabilité (présence/ absence) des ET en

fonction des individus ou des populations d'une même espèce. Ainsi, le présent travail est le premier à présenter l'existence de *MLE* chez les diatomées marines.

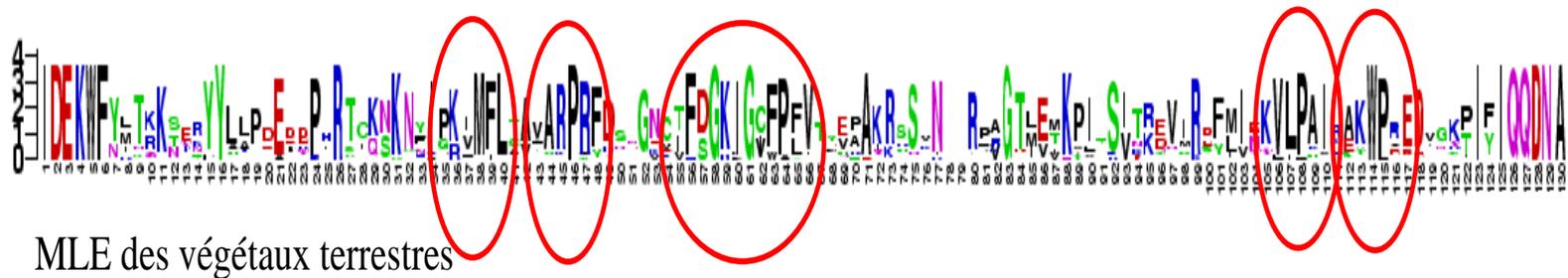
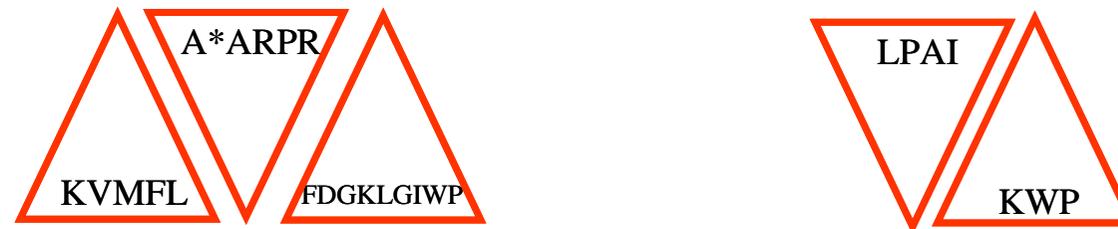
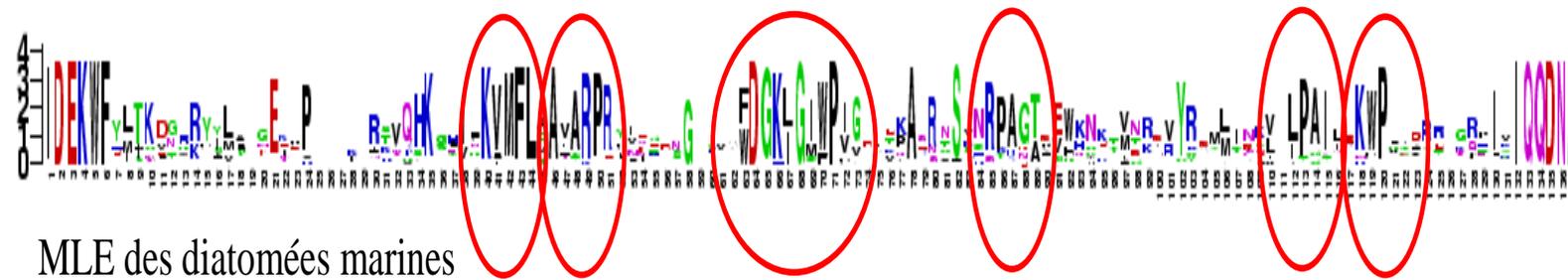
### 1.6.2 Caractéristiques des *MLE* chez les diatomées marines

Pour chacune des 10 espèces présentant des *MLE*, 5 amplifications de 380 pb minimum ont été séquencées, et au total 55 séquences ont été obtenues. Ces séquences présentent des variations au niveau nucléique (taille, mutations) ainsi qu'au niveau protéique (changement de cadre de lecture ou motifs supplémentaires), par exemple les séquences sont comprises entre 372 et 401 pb ou le motif DANKAKA supplémentaire chez deux espèces *O. aurita* et *A. coffeaeformis*. Ces variations peuvent être observées entre les espèces et entre les séquences issues d'une seule et même espèce. Pour chaque espèce, des séquences consensus ont été établies. La variabilité des séquences nous a conduit à établir plusieurs consensus *MLE* chez la plupart des diatomées (soit chez 7/10 espèces). Ceci témoigne de la diversité des *MLE* au sein des diatomées marines. Cette diversité pourrait être le résultat de l'existence de plusieurs *MLE* ancestraux dans les génomes hôtes ou par la présence d'un seul élément dont les copies auraient subi de nombreuses mutations au cours de l'évolution. Dans les deux cas, ces résultats s'expliqueraient par une invasion ancienne des *MLE* dans le génome des diatomées.

Cependant, cette explication doit être nuancée par le fait que chez certaines espèces telles que *A. acutiuscula*, *O. aurita* et *T. weissflogii*, les séquences *MLE* présentent une faible variabilité et un seul consensus a été réalisé suggérant l'existence d'un élément peu muté, récemment introduit et potentiellement actif dans le génome de ces trois diatomées. De plus, les *MLE* de diatomées présentent une faible proportion de codons stop dans les 55 séquences *MLE* obtenues, contrairement à la majorité des *MLE* décrits dans la littérature qui semblent être très mutés et ceci même dans la région catalytique (Hartl et al., 1997b; Robertson 1997; Capy et al., 1992). Bien que nous n'ayons pas accès à la séquence complète et que les séquences amplifiées soient courtes, l'absence de codon stop dans la région catalytique permet d'envisager une transcription potentielle du gène codant la transposase des *MLE* chez les diatomées (Lohe et al., 1995; Capy et al., 1992). En effet, 44 des 55 séquences obtenues présentent un ORF naturel : il n'a donc pas été nécessaire d'avoir recours à la reconstruction de séquence, comme cela est souvent cité dans la littérature. Pour exemple, Feschotte et Wessler (2002) indiquent que des réarrangements de séquences ont été nécessaires afin de retrouver le cadre de lecture dans les fragments *MLE* d'Angiospermes. De même, les séquences *MLE* tronquées de bambous présentent des insertions et délétions ayant introduit

des codons stop ayant conduit, sans doute dans la majorité des cas, à des séquences inactives (Zhou et al., 2010). Toutefois, chez le riz et le pois, des séquences *MLE* de 380 pb *Osmar* et *Psmar* présentant une ORF ont été mises en évidence. Les éléments complets correspondants ont été caractérisés par la suite et présentent, comme envisagé, des ORF sur l'intégralité de la séquence, ce qui dénote leur capacité à être activés (Feschotte et al., 2003; Macas et al., 2005). Ainsi, la présence de nombreux cadres de lecture ininterrompus chez les *MLE* des diatomées marines nous incite à penser que des éléments complets et potentiellement actifs seraient présents dans les génomes de diatomées et notamment chez les trois espèces *A. acutiuscula*, *O. aurita* et *T. weissflogii*. Les séquences *MLE* traduites *in silico* en protéines dévoilent la présence de motifs très conservés dans l'ensemble des séquences obtenues chez les diatomées marines. Les séquences protéiques *MLE* des diatomées présentent des motifs typiques des transposases de *MLE* tels que ceux contenant les deux premiers résidus aspartate de la triade catalytique (Doak et al., 1994; Vos et Plasterk, 1994; Jarvick et Lark, 1998; Feschotte et Wessler, 2002). La présence de ces motifs intacts dans les séquences *MLE* des diatomées marines est biaisée par l'utilisation des amorces *MLE5A* et *MLE3A* définies sur ces mêmes motifs. Cependant, parmi les motifs conservés dans les séquences *MLE* consensus des diatomées, certains sont également retrouvés chez les fragments *MLE* des végétaux. Les motifs A\*ARPR et LPAI sont retrouvés chez les *MLE* de riz et de bambous (Feschotte et al., 2003; Zhou et al., 2010) suggérant que ces deux motifs doivent jouer un rôle dans la fonctionnalité de la transposase *MLE* des végétaux. Les motifs conservés correspondent aux régions ayant permis l'identification des *MLE* de diatomées dans les banques de données. En effet, l'identification des séquences obtenues en PCR se base principalement sur leur traduction protéique *in silico* et la correspondance obtenue par Blast sur les banques de données. Remarquons que les Blast en nucléiques n'ont donné que très peu d'homologie, les Blast réalisés à partir des traductions *in silico* ont fourni de fortes similarités avec des *MLE* végétaux préalablement détectés et identifiés par Feschotte et Wesler (2002) et Zhou et al. (2010). Ceci semble indiquer que la pression de sélection agit principalement à un niveau protéique et que les transposases potentielles de *MLE* des diatomées marines seraient proches de celles des *MLE* des plantes terrestres, comme il est observé grâce au Weblogo (Figure 38). Les séquences *MLE* protéiques partagent en effet certains motifs avec les *MLE* de bambous tels que le motif RPAGT et KWP (Zhou et al., 2010), d'autres motifs comme KVMFL ou encore le motif long (F/W) DGK (L/I) GWP sont retrouvés chez les *MLE* du riz (Feschotte et al., 2003). Notons qu'aucun motif observé de manière conservé entre les séquences *MLE* de diatomées n'apparaît spécifique de celles-ci. En effet, les six motifs décelés chez les *MLE* de

diatomées sont également conservés dans les *MLE* des végétaux terrestres. Ces motifs conservés dans la quasi-totalité des séquences des *MLE* végétaux suggèrent que les séquences correspondantes doivent avoir un impact sur la fonctionnalité de la transposase et seraient ainsi moins soumises aux mutations (Maruyama et Hartl., 1991b). D'autre part, la variabilité importante dans les régions entre les motifs conservés observée ici est assez inattendue car les espèces dont les *MLE* sont issus appartiennent à une seule et même classe : celle des Bacillariophycées. Ainsi la variabilité observée semble être une spécificité des *MLE* de diatomées marines. Les relations phylogénétiques entre les *MLE* des diatomées marines et les *MLE* des plantes terrestres seront discutées ultérieurement. Au-delà des motifs conservés, les séquences *MLE* traduites en protéines montrent des régions de polymorphisme reflétant de nombreux événements de mutations. Ces événements sont typiques de l'évolution rapide de cette fraction génomique. De plus ce polymorphisme peut être observé au sein même des espèces incitant à se questionner sur l'évolution et la distribution des *MLE* dans les espèces de diatomées marines.



**Figure 38 : Similarités entre les séquences de MLE de diatomées et des MLE de plantes.**

Le weblogo des MLE de végétaux a été réalisé à partir des séquences MLE déterminées par Feschotte et Wessler (2002). Les motifs conservés aux MLE de diatomées marines et Angiospermes sont entourés en rouge sur les séquences et les motifs communs sont précisés dans les triangles rouges.

### I.6.3 Polymorphisme et évolution des MLE chez les diatomées marines

La comparaison des traductions *in silico* des consensus *MLE* indique comme pour les séquences nucléiques un polymorphisme interspécifique et intraspécifique. La structure des séquences au niveau du site catalytique des *MLE* de diatomées marines semble être constituée par une alternance de régions conservées et de régions variables. Parmi les séquences consensus *MLE* de diatomées, on distingue clairement des séquences consensus très similaires bien qu'elles proviennent d'espèces différentes. L'étude du taux d'identité et des relations phylogénétiques entre les séquences *MLE* des diatomées nous laisse penser que l'évolution des *MLE* serait indépendante de l'évolution des espèces de diatomées. L'analyse phylogénétique a produit des groupes très similaires à ceux observés dans la matrice d'identité. Les *MLE* des diatomées forment sept groupes dont cinq groupes sont bien isolés et soutenus par les valeurs de bootstrap comprises entre 87 et 100 %. Les groupes observés sont de deux types : intraspécifiques lorsque les séquences proviennent d'une même espèce et interspécifiques lorsqu'elles proviennent de différentes espèces. Des taux d'identités très élevés ont été observés entre des séquences *MLE* provenant d'espèces différentes, dépassant parfois ceux observés entre les *MLE* d'une même espèce. Prenons l'exemple des deux espèces *A. coffeaeformis* et *O. aurita*, les séquences *Acomarcons1* et *Odaumarcons1* présentent 96 % d'identité entre elles, tandis que le taux d'identité entre les quatre séquences consensus d'*A. coffeaeformis* varie de 27 à 34 %. Des résultats similaires ont été obtenus chez les animaux (Robertson, 1993; Capy et al., 1994; Bui et al., 2008) et chez les végétaux (Feschotte et Wessler, 2002; Feschotte et al., 2003; Jacobs et al., 2004, Zhou et al., 2010). Citons l'exemple des *MLE* du genre *Beta* (Jacobs et al., 2004) où aucun groupe monospécifique n'a été mis en évidence au cours de l'analyse des 40 séquences *MLE* isolées chez quatre espèces. La formation de groupes polyspécifiques suggère donc que plusieurs espèces contiennent un même *MLE*.

Par exemple chez les deux espèces du genre *Amphora*, un seul consensus a été mis en évidence chez *A. acutiuscula Aacumarcons1*, alors que chez *A. coffeaeformis*, quatre consensus ont été réalisés correspondant à quatre *MLE* différents. Les *MLE* de ces deux espèces du même genre sont pourtant très divergentes alors que chaque séquence *MLE* d'*A. coffeaeformis* trouve des homologues avec des espèces éloignées, notamment une des séquences de *A. coffeaeformis* (ordre des pénales) partage plus de 90 % d'homologie avec la séquence *MLE* consensus de *O. aurita* (ordre des centriques). Ces deux espèces diatomées appartiennent à deux ordres éloignés selon les critères morphologiques sur lesquels repose la classification des diatomées (Round et al., 1990) ou bien selon les informations moléculaires

et phylogénétiques (Medlin et al., 1993). Plusieurs hypothèses peuvent expliquer une distribution « patchy » ou sporadique caractéristique des *MLE* (Lohe et al., 1995) et une variabilité telle que celle observée chez les diatomées marines.

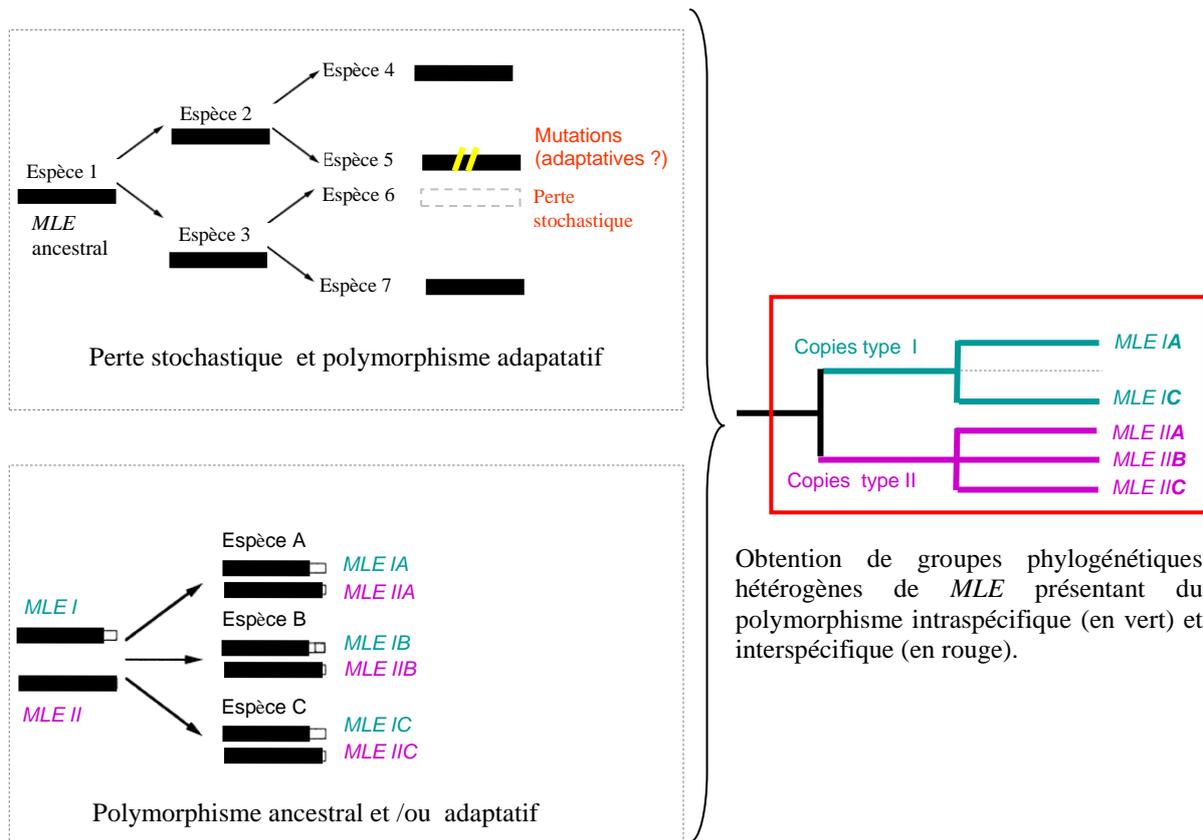
La première hypothèse correspond à l'alternance d'évènements de multiplication (burst de transposition) et d'élimination des ET ayant engendré des lignées de *MLE* différentes. Cette hypothèse n'est pas la plus parcimonieuse mais le concept de la perte stochastique a été largement discutée et étudiée dans le cas des *MLE* et constitue l'hypothèse la plus probable (Capy et al., 1994; Lohe et al., 1995; Kidwell et Lish, 1997; Lampe et al., 2001, Arkhipova et Meselson, 2005, Feschotte et Prittham, 2007). Cela suppose que différents *MLE* étaient présents de manière ancestrale dans les lignées de diatomées et que l'évolution a conduit à la perte des éléments chez certaines espèces (Lohe et al., 1995; Capy et al., 1994, 1997). La disparition d'ET au sein du génome d'une ou plusieurs espèces est le mécanisme supposé à l'origine de la distribution des *mariner* dans le genre *Drosophila*. Dans le cas des diatomées, organismes unicellulaires se reproduisant essentiellement par multiplication végétative, il semble évident que la dérive génétique soit un mécanisme à prendre en considération pour expliquer la distribution des ET (Nuzhdin et Petrov, 2003; Charlesworth et Wright, 2001).

Les pressions de sélection variées dues à l'adaptation à des environnements différents peuvent également conduire à une hypothèse évolutive des *MLE* de diatomées. L'ensemble des espèces que nous avons étudié est très hétérogène, il est constitué de diatomées provenant des côtes atlantiques françaises et du littoral vietnamien ainsi que d'écotypes différents, de diatomées benthiques ou planctoniques etc. L'hypothèse d'un polymorphisme adaptatif pourrait expliquer certains des résultats obtenus notamment dans le cas des *MLE* identifiés chez les espèces du genre *Amphora* et du genre *Nitzschia*. Les *MLE* d'*A. coffeaeformis*, originaire de la côte atlantique française aux eaux tempérées et d'*A. acutiuscula* originaire de la côte vietnamienne tropicale présentent peu de similarités. Les *MLE* de *A. coffeaeformis* sont très proches des *MLE* d'espèces éloignées telles que *P. tricornutum* ou *E. paludosa*. Ceci pourrait être le résultat de deux types de processus non exclusifs : la perte de copies communes et ancestrales et/ou la différenciation des *MLE* en réponse aux pressions environnementales différentes. De même, les *MLE* obtenus chez les espèces du genre *Nitzschia* qui forment deux groupes différents, le premier correspondant aux *MLE* de l'espèce française (*N. laevis*) et le deuxième correspondant aux *MLE* de l'espèce vietnamienne (*N. palea*). Dans ce cas, nous pouvons également imaginer que les variations de séquences qui conduisent à la formation de deux groupes distincts de *MLE* soient le résultat de pressions de

sélection dépendant du milieu. Afin de déterminer s'il existe des spécificités géographiques, il faudrait échantillonner plusieurs populations d'origines géographiques différentes d'une seule et même espèce de diatomées. De plus, l'adaptation environnementale ne serait certainement pas la seule à l'origine du polymorphisme des *MLE* dans les génomes de diatomées.

La troisième hypothèse qui permettrait d'expliquer la distribution des *MLE* chez les diatomées correspond à la possibilité de phase de ré-invasion par des ET présents chez des espèces non reliées phylogénétiquement mais proches géographiquement selon un mécanisme de transfert horizontal. Le domaine marin représente un environnement privilégié par sa richesse en microorganismes, vecteurs potentiels de transfert horizontal (Stewart et Sinigalliano, 1990; Jiang et Paul, 1998; Zhaxybayeva et al., 2006; Hildebrand, 1991; Sobecky et Hazen, 2009). Néanmoins en ce qui concerne les ET, ces évènements de transferts sont difficiles à mettre en évidence car il n'en existe que des preuves indirectes. Toutefois, les transferts horizontaux d'ET permettent d'expliquer le maintien des ET dans les populations (Kidwell 1992; Robertson, 1993, 2002; Handler and McCombs, 2000; Silva et Kidwell., 2004) notamment dans le cas des éléments *mariner* dans les populations de Drosophiles, suivant le modèle d'évolution de Hartl (1997a, 1997b). Ce mode de transferts des ET a également été décrit dans les génomes végétaux (Diao et al., 2006; Dooner et Weil, 2007; Cheng et al., 2009; Roulin et al., 2008, 2009). En outre, l'étude des éléments *Ty1-copia* de *P. tricornutum* suggère l'implication de transferts horizontaux à l'origine des groupes de rétrotransposons CoDi (Copia Diatom) (Maumus et al., 2009). Le groupe de CoDi6, composé de rétrotransposons issus de *P. tricornutum* et de *T. pseudonana*, apparaît phylogénétiquement plus proche des éléments *copia* que des autres groupes de rétrotransposons de diatomées. Nous ne pouvons donc exclure l'implication possible de ce type d'évènements dans la distribution des *MLE* chez les diatomées marines mais les données présentées ne soutiennent pas cette hypothèse. Les mettre en évidence serait très complexe car cela suppose que les espèces concernées soient isolées du point de vue de la reproduction (Fortune et al., 2008) et la reproduction sexuée n'est caractérisée que chez quelques espèces de diatomées (Chepurnov et al., 2004).

Ainsi, la distribution des *MLE* chez les diatomées marines nous apparaît comme étant sans doute à la fois le résultat de phénomènes de dérive génétique, de polymorphismes adaptatifs (Figure 39) et peut être de transferts horizontaux. L'intensification de la variabilité des séquences *MLE* observée chez les espèces de diatomées par l'apport de nouveaux éléments par transfert horizontal n'est pas à exclure.



**Figure 39 : Représentation schématique des processus évolutifs pouvant être à l'origine des groupes phylogénétiques hétérogènes des *MLE* de diatomées.**

#### 1.6.4 Présence des *MLE* dans le génome séquencé de *P. tricornutum*

Comme montré dans cette étude, les *MLE* semblent être largement distribués dans les génomes des diatomées marines. Des fragments *MLE* ont été mis en évidence chez *P. tricornutum* à l'aide des amorces dégénérées *MLE5A* et *3A* définies par Feschotte et Wessler (2002). Cette stratégie a deux inconvénients majeurs : i) elle ne permet pas de détecter la présence des *MLE* lorsqu'ils sont mutés en début et milieu de triade catalytique. ii) seules des fractions de *MLE* sont amplifiées et l'obtention d'éléments complets doit être abordée avec une autre méthode. La méthode la plus couramment utilisée lorsque le génome est séquencé est l'approche bioinformatique. Le génome de *P. tricornutum* souche CCMP632 a été séquencé récemment (Bowler et al., 2008) et l'annotation des ET dans ce génome rapporte qu'il est dépourvu de *MLE* (Maumus et al., 2009) mais contient des éléments désignés comme « transposase-like » qui n'ont vraisemblablement pas été identifiés. Le polymorphisme des *MLE* de *P. tricornutum* (souche de Bourgneuf) nous incite à penser que des *MLE* pourraient être cachés dans la fraction « transposase-like » du génome séquencé. En outre, le logiciel RepeatMasker (Smit et al., 1996-2004) utilisé chez *P. tricornutum* souche CCMP632, permet

de détecter les ET par confrontation avec Repbase, une base de données contenant des ET identifiés et répertoriés (Kohany et al., 2006). Or les *MLE* de diatomées et de végétaux présentant certaines dissimilarités, leur présence aurait pu échapper à la détection de Repeat Masker. Dans cette étude, nous avons recherché spécifiquement les *MLE* de *P. tricornutum* par détection des éléments TIR ayant des caractéristiques de longueur correspondant aux *MLE* végétaux et par comparaisons avec les séquences *MLE* identifiées chez les diatomées. Aucune séquence répondant à ces critères n'a été détectée dans le génome de *P. tricornutum* CCMP632, alors que des résultats positifs ont été obtenus avec cette souche par l'approche expérimentale. Les résultats contradictoires des deux approches, expérimentale et *in silico*, ne nous permettent pas de dégager une interprétation claire expliquant cette différence. Les annotations en ET sont très dépendantes de l'efficacité de séquençage et de l'assemblage. Il faut donc noter que si le génome contient uniquement des séquences tronquées de *MLE* localisées dans des répétitions en tandem ou dans des fragments non chevauchants, ces séquences peuvent être éliminées lors de l'assemblage du génome ou échapper à l'annotation bioinformatique (Bergmann et Quesneville, 2007).

En outre, l'approche moléculaire a mis en évidence la présence de séquences *MLE* chez la souche de *P. tricornutum* séquencée, différentes de celles obtenues dans la souche de Bourgneuf. Ceci corrobore l'hypothèse émise précédemment suggérant qu'il existerait un polymorphisme adaptatif des *MLE* au sein des espèces voire même entre les souches. Ce polymorphisme pourrait être relié à une origine géographique différente ou des conditions de cultures différentes de clones et pourrait être le résultat de forces adaptatives sous l'influence de l'environnement. Rechercher les polymorphismes d'insertions des *MLE* au sein des nombreuses souches de *P. tricornutum* collectées dans différents sites géographiques (De Martino et al., 2007) permettrait de tester cette hypothèse (Maumus et al., 2009) et de déterminer si des copies de *MLE* actives existent dans le génome de *P. tricornutum*.

Les principaux points mis en évidence dans la première partie de ce travail sont les suivants :

Des séquences de 380 pb correspondant en partie à la triade catalytique des *MLE* ont été détectées dans les génomes de 10 espèces de diatomées marines, parmi lesquelles la diatomée modèle *P. tricornutum*.

Ces séquences ont été identifiées comme des *MLE* de plantes et présentent des particularités :

- elles présentent des ORF naturels qui laissent supposer qu'il pourrait exister des copies actives de *MLE*
- les *MLE* de diatomées sont composés de régions conservées entre les espèces présentant des motifs communs avec les *MLE* de plantes, et de régions variables spécifiques des *MLE* de diatomées.

L'analyse phylogénétique révèle que le polymorphisme de séquences ne semble pas suivre le schéma évolutif des espèces de diatomées marines. Ce polymorphisme pourrait être adaptatif et semble résulter d'une longue évolution et de processus de transfert horizontaux.

Des fragments de *MLE* ont été caractérisés chez la souche séquencée de *P. tricornutum* par l'approche bio-moléculaire alors que l'approche bio-informatique n'a pas permis de détecter des éléments appartenant à la famille *Tc1-mariner*. La comparaison des *MLE* des deux souches (souche séquencée CCMP632 et souche MMS NCC340) de *P. tricornutum* semble révéler également un polymorphisme intraspécifique observable entre les souches de diatomées marines.

Dans la suite de ce travail, nous avons comparé l'évolution des *MLE* de diatomées avec les *MLE* de plantes au travers de l'analyse phylogénétique des séquences protéiques et nous avons classé les *MLE* de diatomées au sein de la superfamille *Tc1-mariner*.

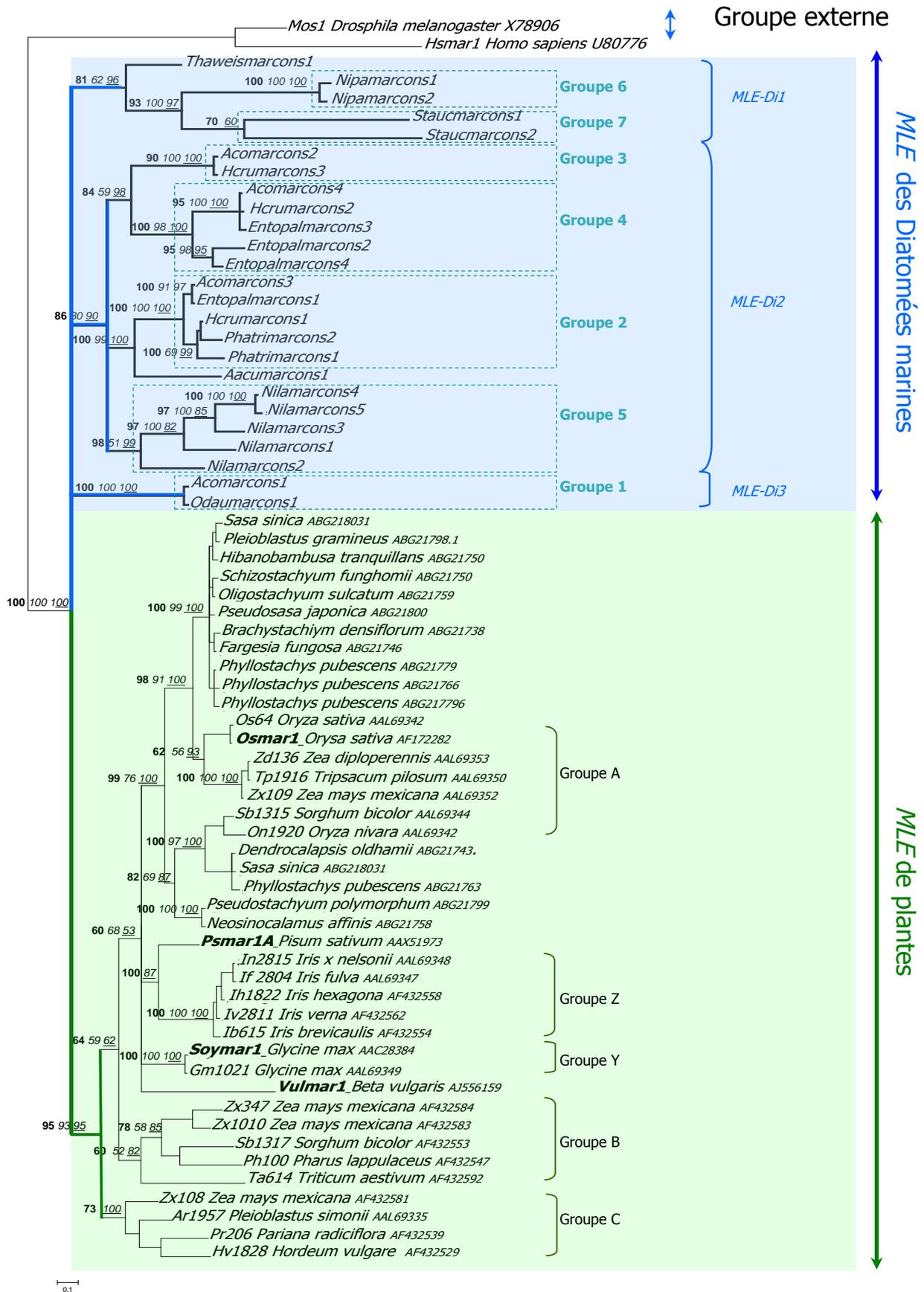


Figure 40: Position phylogénétique des MLE de diatomées marines par rapport aux MLE de plantes.

L'arbre phylogénétique a été réalisé avec les 25 séquences consensus de *MLE* de diatomées (*en bleu*) et 241 séquences *MLE* de plantes (*en vert*) traduites en acides aminés. Le groupe externe est composé de deux séquences *mariner* issues de *Homo sapiens* et de *Drosophila melanogaster*. Les noms en gras correspondent aux séquences extraites de *MLE* complets.

Les séquences *MLE* végétales forment un groupe polyphylétique qui se divise en quatre lignées dont trois appartiennent aux *MLE* de diatomées marines, la dernière étant la lignée des *MLE* de plantes terrestres.

Les trois lignées *MLE-Di1*, *MLE-Di2* et *MLE-Di3* forment trois clades monophylétiques. Au sein de ces lignées, les groupes mis en évidence dans la partie I-4 de ce chapitre sont conservés (encadrés en pointillés bleus).

L'arbre présenté a été obtenu par inférence bayésienne (les arbres consensus obtenu en NJ et MP ont une topologie similaire à celle présentée ci-dessus) et les groupes obtenus avec chaque méthode sont indiqués par les valeurs de bootstrap indiquées en caractères gras pour la méthode NJ, en italique pour la méthode MP et les valeurs de probabilité postérieure (IB) indiquées en pourcentage sont soulignées. Les valeurs <50 ne sont pas indiquées.

## II Positionnement des *MLE* de diatomées par rapport aux *MLE* de plantes et dans la superfamille *Tc1-mariner*

Les *MLE* identifiés chez les diatomées marines dans la partie III montrent clairement des similitudes avec les *MLE* de plantes. Afin de pouvoir les placer, nous avons opté pour deux approches bio-informatiques. La première approche est classique et basée sur la phylogénie moléculaire. Elle a été utilisée pour positionner les *MLE* de diatomées par rapport aux *MLE* de végétaux. La seconde approche est une nouvelle méthode de classification des ET développée par Jacques-Deric Rouault (LEGS UPR CNRS 9034, Gif-sur-Yvette) qui permet de classer ensemble tous les éléments ensemble qu'ils soient tronqués ou bien que leurs séquences soient complètes. Cette méthode a permis de classer les *MLE* des diatomées dans la superfamille *Tc1-mariner*.

### II.1 Position phylogénétique des *MLE* de diatomées par rapport aux *MLE* de plantes

La Figure 40 présente un arbre réalisé selon trois méthodes phylogénétiques différentes : Neighbor Joining (NJ), Maximum de Parcimonie (MP) et Inférence Bayésienne (IB). Les deux premières méthodes ont été réalisées avec 1000 réplifications du test de bootstrap et l'Inférence Bayésienne a été réalisée 4 fois (de manière indépendante) avec un million de générations.

Les analyses phylogénétiques ont été réalisées avec les séquences traduites en acides aminés des 25 séquences *MLE* consensus issues de diatomées marines et de 41 séquences partielles (~ 380 pb) de *MLE* de plantes disponibles dans les bases de données. Deux séquences de *mariner* issues d'animaux (*Mos1* de la drosophile et *Hsmar1* issu de l'homme) ont été utilisées pour enraciner ces arbres. Les arbres consensus obtenus à partir des trois méthodes présentent une typologie très similaire. Seul l'arbre bayésien est présenté et les groupes formés sont indiqués par les valeurs résultant des tests de bootstrap ou par la valeur de la probabilité postérieure (IB) (Figure 40).

L'arbre phylogénétique est divisé en quatre clades regroupés au sein d'un grand groupe de *MLE* végétaux, séparés des *MLE* animaux constituant le groupe externe (Figure 40).

Chacun de ces clades est validé par des valeurs de nœuds supérieures à 80 pour la majorité des méthodes. Les séquences *MLE* de diatomées (cadre bleu) distribuées dans trois clades, sont bien séparées des séquences *MLE* de plantes terrestres regroupées en un seul clade (cadre vert) (Figure 40).

Les analyses phylogénétiques des trois méthodes NJ, MP, et IB confirment la formation des groupes de *MLE* de diatomées numérotés de 1 à 7 obtenus précédemment dans l'arbre non raciné des *MLE* (Figure 33, p 101). Chaque groupe est retrouvé intact et supporté par des valeurs de nœuds élevées. Dans cette analyse, ils sont regroupés en trois lignées de *MLE* représentées (Figure 40) par les trois clades nommés *MLE-Di1*, 2, 3 (*MLE* de Diatomée) qui apparaissent distinctement séparés du quatrième clade constitué des *MLE* de végétaux terrestres.

Les trois premières branches de l'arbre sont constituées exclusivement de *MLE* issus de diatomées marines. Chacun de ces trois clades nommés *MLE-Di1*, *MLE-Di2*, *MLE-Di3* est soutenu par des valeurs de nœuds comprises entre 62 et 100 selon la méthode. Le clade *MLE-Di1* est un groupe hétérogène qui se divise en deux groupes homogènes préalablement obtenus en partie I ainsi que d'une séquence isolée. *MLE-Di1* composé de 5 séquences *MLE* est soutenu par les valeurs de bootstrap avec les méthodes NJ (81) et IB (96), toutefois, les valeurs de bootstrap de MP (62) sont moins convaincantes. Ceci s'explique par le fait que les séquences qui sont regroupées sont divergentes comme l'a montré l'arbre en MP des *MLE* de diatomées (Figure 33, p 101). La deuxième branche des *MLE* végétaux, *MLE-Di2*, est le plus grand groupe de *MLE* des diatomées, il comprend 18 des 25 séquences consensus et il se divise également en trois sous-branches hétérogènes (*MLE* provenant d'espèces de diatomées différentes) supportées par les valeurs de nœuds de plus de 80 %. La première sous-branche correspond au rassemblement des groupes 3 et 4 établis précédemment (§ III-A.4; p 87). Chacun de ces groupes est soutenu par des valeurs de nœuds supérieures à 90 avec le NJ et l'IB, de même que pour le *MLE-Di1* les valeurs de bootstrap de MP sont d'environ 50. La deuxième sous-branche correspond au groupe 2 établi précédemment et associé à la séquence *Aacumarcons1* (unique séquence consensus de l'espèce *A. acutiuscula*) qui apparaissait isolée dans la précédente analyse Figure 33, p 101. Enfin la troisième sous-branche correspond au groupe 6 de *MLE* de diatomées dans son intégralité. Le clade *MLE-Di3* ne correspond qu'à un seul groupe de *MLE* de diatomées, le groupe 1 qui ne comprend que 2 séquences. Ces deux séquences sont donc très éloignées des autres *MLE*.

Le clade des *MLE* de plantes est monophylétique et contient les 41 séquences de *MLE* préalablement identifiées chez diverses espèces de Gymnospermes (Feschotte et Wessler,

2002) et aussi de bambous (Zhou et al., 2010). Les 41 séquences partielles utilisées comportent 37 séquences obtenues par biologie moléculaire avec les amorces *MLE5A* et *MLE3A*. Les quatre autres séquences sont issues de *MLE* dont la séquence pleine longueur a été caractérisée. Seul le fragment entre les deux premiers D de la triade DD39D a été utilisé pour l'analyse phylogénétique. Dans ce clade, on retrouve les lignées A, Z, Y, B et C déterminées par Feschotte et Wessler (2002) au milieu desquelles s'insèrent les *MLE* de bambous (Figure 40). De plus, les séquences issues des *MLE* complets sont distribuées dans ce clade monophylétique, indiquant ainsi la proximité entre séquences partielles et complètes.

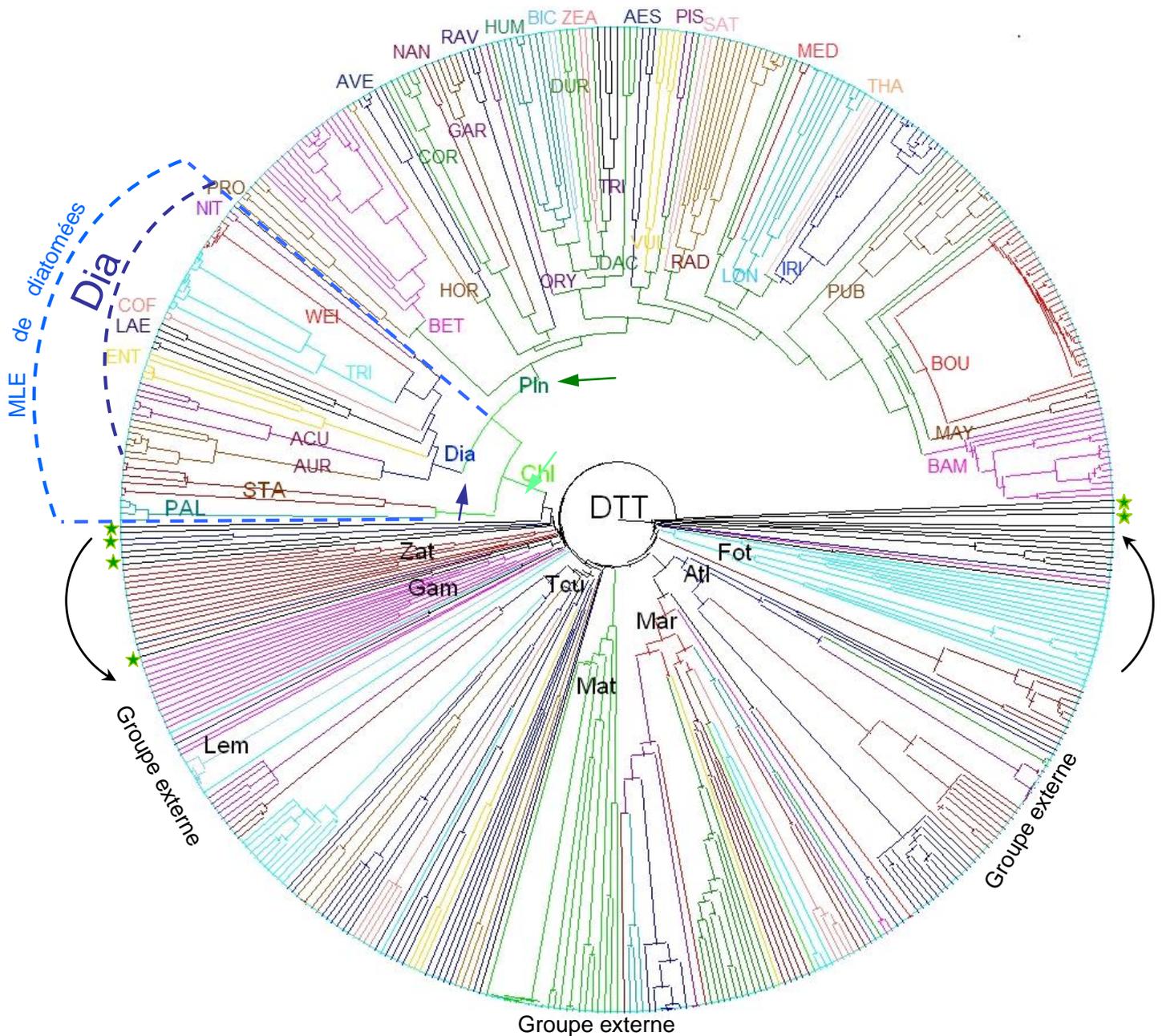
La présence d'un clade, regroupant tous les *MLE* de plantes terrestres ainsi que ceux des *MLE* des diatomées, indique que les *MLE* des végétaux ont une origine commune. Toutefois, la divergence en quatre clades avec un clade monophylétique exclusivement constitué de *MLE* provenant des espèces terrestres suggère que les *MLE* d'origine marine ont évolué indépendamment de ceux des espèces végétales terrestres. Ainsi les clades *MLE-Di* correspondent à des lignées marines et le dernier clade à la lignée terrestre des *MLE* végétaux.

Après avoir classé les *MLE* de diatomées et de plantes terrestres, ceux-ci ont été intégrés dans une méta-analyse qui rassemble toutes les séquences de *MLE* de végétaux ainsi que les éléments les plus représentatifs des familles composant la superfamille *Tc1-mariner*.

## II.2 Classification des *MLE* de diatomées dans la superfamille *Tc1-mariner*

La classification agrégative, nommée DTT491 (DNA Transposon Tc1-mariner), a été réalisée sur 491 séquences nucléotidiques d'ET comprenant 249 *MLE* de végétaux et 242 éléments de diverses familles issues de la superfamille *Tc1-mariner*, ces derniers constituant le groupe externe selon la méthode UPGM-VM.

Dans le but de déterminer la classification relative des *MLE* de diatomées par rapport aux *MLE* de plantes et dans la superfamille *Tc1-mariner*, les 54 séquences nucléiques (55 séquences au départ dont 2 séquences identiques) issues des 10 espèces de diatomées ont été intégrées au jeu de données de la classification pour la méthode UPGM-VM. Le résultat de cette classification est présenté en Figure 41.



**Figure 41 : Rosace obtenue à l'aide de la méthode UPGM-VM représentant la classification agrégative des MLE de diatomées et de végétaux terrestres par rapport aux ET de la superfamille *Tc1-mariner*.**

La rosace représente la position relative des MLE de diatomées (*en pointillés épais bleus*) parmi 491 séquences nucléiques d'ET en fonction de leur homologie de séquences. En groupe externe se positionnent de la gauche vers la droite, comme indiqué par les flèches noires, les familles déjà identifiées de *Tc1-mariner* : Zet (Zator), Gam (Gambol), Lem (Lemi), Tcu (Tc1), maT (Mat), Mar (*mariner*), Atl (Atlantis), Fot, ainsi que des séquences non classées (IS630, Pce1 à 4, hupfer1). La racine commune de la superfamille *Tc1-mariner* est nommée DTT (Dna Transposons Tc1-mariner).

Les MLE des diatomées marines et des végétaux terrestres sont regroupés dans la partie supérieure de la rosace dans la branche Chl (Chlorophyllis, *flèche verte*), nouvelle famille définie dans cette classification. Elle regroupe d'une part les MLE de diatomées (*en pointillés épais bleus*) formant les groupes Dia (pour Diatomis, *flèche bleue marine*), Sta (Stauroneis) et PAL (palea), d'autre part les MLE de plantes terrestres sont regroupés dans la grande majorité dans la branche Pln (Plantis, *flèche vert foncé*) à quelques exceptions près indiquées par des étoiles (*Atmar5*, 8, 9 et *Osmar3*, 4, 5).

Les *MLE* de diatomés formant le groupe Dia sont distribués en 10 groupes nommés en fonction des espèces de diatomées hôte les constituant : *de la gauche vers la droite*, NIT (Nitzschia), WEI (weisflogii), TRI (tricornutum), COF (coffeaeformis), LAE (laevis), ENT (entomoneis), ACU (acutiuscula), AUR (aurita).

Les *MLE* de plantes formant le groupe Pln sont distribués en 28 groupes : *de la gauche vers la droite*, PRO (procubens), BET (Beta), HOR (Hordeum), AVE (Avena), COR (cormar), NAN (nana), GAR (vulgaris), RAV (betterave), ORY (Oryza), HUM (humilis), BIC (bicolor), DUR (dura), ZEA (Zea), DAC (dactylon), TRI (Tripsacum), AES (aestivum), VUL (vulgaris), PIS (Pisum), SAT (sativum), RAD (radiciflora), MED (Medicago), Lon (longifolia), THA (thaliana), IRI (Iris), PUB (pubescens), BOU (Bambou), MAY (mays), BAM (Bambusa). Pour plus de détails voir Annexe3.

### II.2.1 Position des *MLE* végétaux dans la classification DTT491 de la superfamille *Tc1-mariner*.

La rosace de la Figure 41 présente les résultats de l'analyse DTT491. Cette analyse regroupe 478 items ou séquences de transposons de la superfamille *Tc1-mariner* et 13 éléments issus de procaryotes et en intégrant la nouvelle superfamille, Zator, apparentée aux *Tc1-mariner* (Bao et al., 2009). La moitié des séquences de cette analyse, soit 242 séquences, sont complètes (TIR compris) et jouent le rôle de groupe externe. Ces séquences appartiennent à différentes familles de la superfamille *Tc1-mariner* définies antérieurement (Shao et Tu., 2001; Rouault et al., 2009). L'autre moitié des données sont des séquences *MLE* de végétaux terrestres (plantes monocotylédones ou dicotylédones) auxquelles ont été ajoutées les séquences de diatomées marines obtenues dans le présent travail.

La branche Chl de la rosace se détache (*flèche vert clair*, Figure 41) et apparaît distante des autres branches correspondant aux familles de *Tc1-mariner* connues. Cette branche regroupe toutes les séquences *MLE* obtenues à partir des espèces de diatomées et la grande majorité des *MLE* de plantes. Comme nous l'avons observé dans l'analyse phylogénétique, les *MLE* des diatomées marines restent rassemblés avec les *MLE* de plantes dans un grand groupe séparé des *MLE* d'origine animale de la superfamille *Tc1-mariner*. Ce grand groupe correspond à une nouvelle famille nommée selon le code à trois lettres DTTChl. Ici, Chl signifie Chlorophyllis dans la mesure où tous les hôtes de ces *MLE* sont des organismes photosynthétiques.

**Tableau 12 : Propriétés de la superfamille *Tc1-mariner* (DTT) et de ses principales familles.**

Chaque groupe est caractérisé par son élément central dont le numéro d'accèsion dans Genbank et ses rayons interne et externe. Lorsque le rayon interne est inférieur au rayon externe de la sphère représentant la famille d'ET, le groupe est qualifié de Clan.

Ex : La famille Atlantis se compose de 38 séquences dont l'élément central a pour numéro d'accèsion AJ496129 dans Genbank, Le rayon interne de la sphère décrivant ces éléments est supérieur au rayon interne de la sphère, ainsi cette famille n'est pas un Clan.

Niveau hiérarchique	Groupes formés au cours la classification		Effectif	Élément central	Rayon interne	Rayon externe	Propriété
Famille	DTTAtl	( <i>Atlantis</i> )	36	AJ496129	0,722	0,647	
Famille	DTTChl	( <i>MLE végétaux</i> )	235	AF432529	0,610	0,611	Clan
Famille	DTTFot	( <i>Pogo-Fot</i> )	21	AF076632	0,714	0,618	
Famille	DTTGam	( <i>Gambol</i> )	16	AC005169	0,696	0,637	
Famille	DTTLem	( <i>Lemi</i> )	6	EF526301	0,004	0,382	Clan
Famille	DTTMar	( <i>Mariner</i> )	37	AC084518	0,712	0,663	
Famille	DTTMat	( <i>maT/mori</i> )	22	AY121784	0,625	0,634	Clan
Famille	DTTMsq	( <i>ITmDD37E</i> )	9	AF377999	0,216	0,609	Clan
Famille	DTTcu	( <i>Tc1</i> )	58	M93038	0,703	0,606	
Famille	DTTZat	( <i>Zator</i> )	13	D1733A1	0,706	0,570	
Superfamille	DTT	( <i>Tc1-mariner</i> )	491	U51183	0,704		

Le Tableau 12 présente les caractéristiques de la superfamille DTT et de ses principales familles. Certaines familles sont des Clans car le rayon externe est supérieur au rayon interne (DTTLem, DTTMat, DTTMsq). Cela semble concerner les familles présentant des effectifs peu élevés ainsi que la famille Chlorophyllis. Les valeurs de rayons de la famille Chlorophyllis avec un rayon externe très légèrement supérieur au rayon interne (respectivement, 0,611 et 0,610) semblent très compatibles avec les valeurs obtenues pour les autres familles. En effet les familles DTTAtl, DTTFot, DTTGam, DTTMar, DTTcu, DTTZat présentent des rayons internes de l'ordre de 0,70 (0,62 à 0,72) et des rayons externes d'environ 0,60 (0,56 à 0,66) (Tableau 12). Ceci confirme la définition de la nouvelle famille Chlorophyllis regroupant les *MLE* végétaux. Sa position clairement éloignée du centre de la rosace (DTT, Figure 41) confirme son appartenance à la superfamille *Tc1-mariner*.

Enfin, six éléments ne sont pas inclus dans la famille Chlorophyllis, bien qu'isolés à partir des plantes modèles *A. thaliana* (Atmar 5, 8, 9) et *O. sativa* (Osmar 3, 4, 5). Ces séquences apparaissent très différentes des autres séquences *MLE* végétales et se sont ainsi positionnées en périphérie de la famille DTTChl à proximité des séquences utilisées en groupe externe (Figure 41). Le groupe *Lemi* qui appartient à la famille *Tc1-mariner* mais qui n'a pas été décrit comme *MLE*, provient également des plantes (*A. thaliana*) et se situe en dehors de la famille Chlorophyllis.

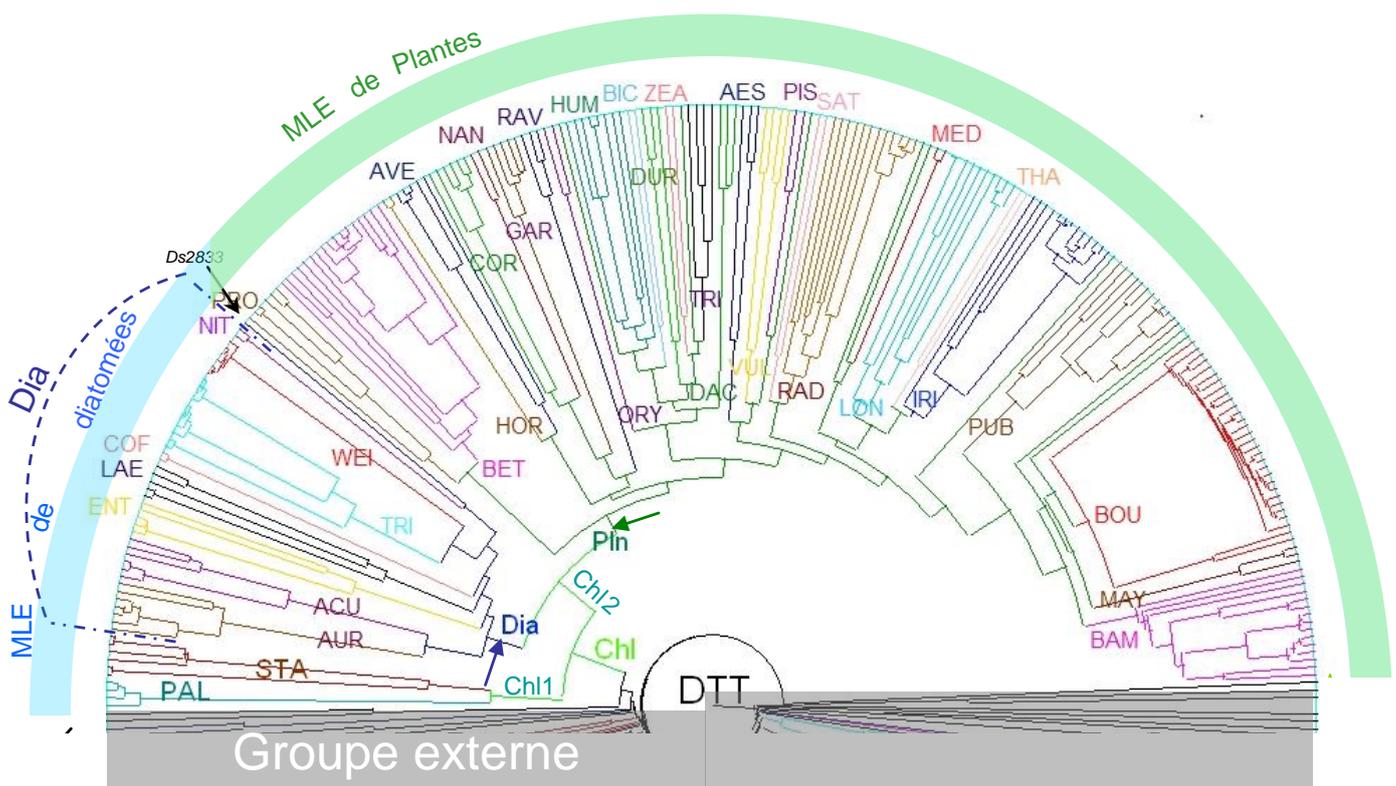
**Tableau 13 : Propriétés de la famille Chlorophyllis et des groupes issus de diatomées marines (DTTChlDia---) et de plantes terrestres (DTTChlPln---) au sein de l'analyse DTT491.**

Les propriétés des tribus formées par les *MLE* de diatomées sont colorées en bleu et celles de la famille Chlorophyllis, DTTChl, sont colorées en vert. Chaque groupe est représenté par son élément central (le numéro d'accèsion est indiqué) et ses rayons interne et externe. Les séquences longues (> 700 pb) sont indiqués par un \*.

Exemple 1 : La sous-famille DTTChlPAL est constituée de 4 séquences courtes de *MLE* de diatomées, son éléments central est disponible dans Genbank sous le numéro d'accèsion FN546214. Son rayon externe étant supérieur à son rayon interne, cette sous-famille est un Clan. Elle est nommée PAL et appartient à la famille Chlorophyllis et donc à la superfamille *Tc1-mariner*, DTT.

Exemple 2 : la Tribu BET appartenant à la sous-famille DTTChlPln est une tribu constituée de 16 séquences de *MLE* de plantes. Une séquence longue est incluse dans les 16 séquences courtes de cette tribu. On parlera alors de Clan au regard de ses rayons interne et externe.

	Niveau hiérarchique	Tribus formées au cours de la classification	Effectif	Élément central	Rayon interne	Rayon externe	Structure	
MLE de diatomées	Sous-famille	DTTChlPAL	4	FN546214	0,031	0,339	Clan	
	Sous-famille	DTTChlSta	6	FN546228	0,364	0,362		
	<b>clade</b>	<b>Chl1</b>	<b>= 10</b>	<b>FN546230</b>	<b>0,478</b>	<b>0,407</b>		
	Tribu	DTTChlDiaACU	6	FN546190	0,192	0,298	Clan	
	Tribu	DTTChlDiaAUR	7	FN546220	0,123	0,322	Clan	
	Tribu	DTTChlDiaCOF	2	FN546194	0,007	0,365	Clan	
	Tribu	DTTChlDiaENT	5	FN546196	0,269	0,329	Clan	
	Tribu	DTTChlDiaLAE	5	FN546212	0,333	0,344	Clan	
	Tribu	DTTChlDiaTRI	11	FN546226	0,269	0,383	Clan	
	Tribu	DTTChlDiaWEI	6	FN546236	0,011	0,339	Clan	
	Tribu	DTTChlDiaNIT	2	FN546209	0,016	0,225	Clan	
	Séquence isolée	DTTChlDia--- <i>Dianthus sp.</i>	1	Ds2833				
MLE de plantes / angiospermes	Tribu	DTTChlPlnAES	3	AF432592	0,297	0,334	Clan	
	Tribu	DTTChlPlnAVE	5	AF432517	0,268	0,313	Clan	
	Tribu	DTTChlPlnBAM	15	DQ528712	0,199	0,245	Clan	
	Tribu	DTTChlPlnBET	16+1*	AJ557185	0,281	0,327	Clan	
	Tribu	DTTChlPlnBIC	2	AF432549	0,272	0,282	Clan	
	Tribu	DTTChlPlnBOU	30	AF172282	0,245	0,265	Clan	
	Tribu	DTTChl PlnCOR	6	AJ557175	0,318	0,336	Clan	
	Tribu	DTTChlPlnDAC	3	AF432520	0,092	0,309	Clan	
	Tribu	DTTChlPlnDUR	3	AF432518	0,295	0,303	Clan	
	Tribu	DTTChlPlnHOR	2	AF432527	0,011	0,323	Clan	
	Tribu	DTTChlPlnHUM	6+2*	AF403142	0,271	0,280	Clan	
	Tribu	DTTChlPlnGAR	5	AJ557157	0,079	0,327	Clan	
	Tribu	DTTChlPlnIRI	9+4*	AF432558	0,269	0,277	Clan	
	Tribu	DTTChlPlnLON	5+5*	AF432571	0,292	0,293	Clan	
	Tribu	DTTChlPlnMAY	3	AF432580	0,073	0,253	Clan	
	Tribu	DTTChlPlnMED	2	AF435566	0,013	0,283	Clan	
	Tribu	DTTChlPlnNAN	2	AJ557174	0,063	0,303	Clan	
	Tribu	DTTChlPlnORY	2	AF432545	0,081	0,306	Clan	
	Tribu	DTTChlPlnPIS	1+1*	AY833553	0,009	0,266	Clan	
	Tribu	DTTChlPlnPRO	5	AJ557156	0,255	0,290	Clan	
	Tribu	DTTChlPlnPUB	12	AF432552	0,229	0,264	Clan	
	Tribu	DTTChlPlnRAD	12	AF432513	0,288	0,299	Clan	
	Tribu	DTTChlPlnRAV	3	AJ557159	0,075	0,320	Clan	
	Tribu	DTTChlPlnSAT	2	AF432529	0,305	0,323	Clan	
	Tribu	DTTChlPlnTHA	1+1*	AL161498	0,252	0,265	Clan	
	Tribu	DTTChlPlnTRI	4	AF432584	0,194	0,286	Clan	
	Tribu	DTTChlPlnVUL	4	AF432515	0,286	0,319	Clan	
	Tribu	DTTChlPlnZEA	2	AF432578	0,261	0,270	Clan	
		Séquences isolées Proche des tribus	DTTChlPln	8*				
		Sous-famille	DTTChlDia	45	FN546203	0,485	0,378	
	Sous-famille	DTTChlPln	180	AJ557187	0,671	0,478		
	<b>clade</b>	<b>Chl2</b>	<b>= 225</b>	<b>AF432590</b>	<b>0,597</b>	<b>0,551</b>		
	Famille	DTTChl : Chl1+Chl2	= 235	AF432529	0,610	0,611	Clan	



**Figure 42 : Structuration de la famille Chlorophyllis : classification des MLE de diatomées et de végétaux terrestres dans l'analyse DTT491.**

La demi-rosace représente un gros plan sur la classification DTT491 des éléments *Tc1-mariner* au niveau de la famille DTTChl nouvellement définie. Cette famille est composée des MLE de diatomées marines (*en pointillés épais bleus*) et de la majorité des séquences MLE de plantes terrestres.

La nouvelle famille Chl (Chlorophyllis, *branches colorées en vert clair*) définie dans cette classification se divise en 2 clades : Chl1 et Chl2. Le clade Chl1 est constitué de MLE de diatomées regroupées en 2 sous-familles PAL et Sta. Le clade Chl2 est également divisé en 2 sous-familles. La première sous-famille est nommée Dia (Diatomis, indiquée *en pointillés bleu marine*), elle est composée de 8 groupes (*branches colorées en bleu marine*) nommées en fonction des genres ou espèces de diatomées NIT (Nitzschia), WEI (weisflogii), TRI (tricornutum), COF (coffeaformis), LAE (laevis), ENT (Entomoneis), ACU (acutiuscula) et AUR (aurita). Ces 8 tribus regroupent 44 des 54 séquences MLE de diatomées et une séquence isolée Ds2833 provenant d'une plante terrestre : *Dianthus sp.*, indiquée par une flèche à la périphérie de DTTChlDia. Deux autres groupes Sta et PAL (*branche cyan et marron*) sont constitués exclusivement de MLE de diatomées mais ne sont pas inclus dans Dia. La seconde sous-famille Pln (Plantis) regroupe quasiment tous les MLE de plantes (179 séquences/186) à 6 exceptions près qui sont distribuées dans le groupe externe et celle issue de *Dianthus sp.*

### II.2.2 Position et propriétés des groupes de MLE végétaux composant la famille Chlorophyllis dans la classification DTT491

Il apparaît, à l'issue de l'analyse DTT491, que la famille Chlorophyllis présente une structure de Clan (Tableau 12). Cela implique qu'elle est très éloignée des autres familles et bien isolée dans l'ensemble des ET. Elle apparaît distante des autres familles de *Tcl-mariner*. La famille DTTChl se structure en deux clades :

➤ le 1<sup>er</sup> clade : *Chl1* est constitué de 10 séquences *MLE* issues de diatomées marines et n'a pas la propriété d'un Clan. Il regroupe les sous-familles Sta (constituée des séquences de *Stauroneis constricta*) et PAL (constituée des séquences de *Nitzschia palea*) (Tableau 14 et Figure 42). La sous-famille PAL a une structure de Clan tandis que la deuxième ne possède pas cette propriété. Ceci suggère que la sous-famille Sta n'est pas un groupe bien défini et qu'il est isolé des autres.

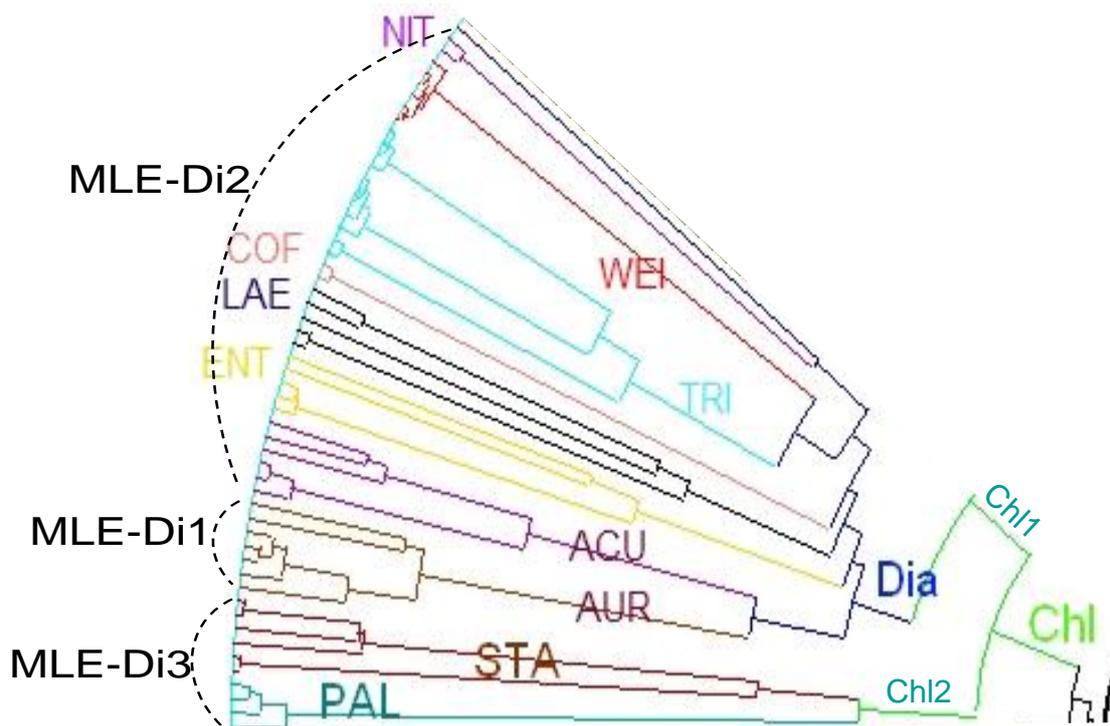
➤ le 2<sup>ème</sup> clade *Chl2* est constitué de 225 séquences. Il est subdivisé en deux sous-clades (Figure 42 et Tableau 13) : le premier regroupe 44 séquences *MLE* provenant de diatomées marines plus une séquence *MLE* issue d'une plante terrestre. Ce sous-clade est défini comme la sous-famille Diatomis (DTTChlDia). Le second comprend 180 séquences obtenues à partir de plantes terrestres, dont 17 séquences de longueurs supérieures à 700 pb (à ORF complète), il a été défini comme la sous-famille Plantis (DTTChlPln).

- La sous-famille Diatomis est composée de 8 tribus dont les noms reflètent la composition en espèces hôtes des *MLE* : ENT (Entomoneis), LAE (laevis), WEI (weissflogii), TRI (tricornutum), COF (coffeaformis), ACU (acutiuscula), AUR (aurita), NIT (Nitzschia) et d'un élément isolé Ds2833 de l'espèce terrestre *Dianthus sp* (Tableau 14 et Figure 42). Les huit tribus sont toutes des Clans, leurs rayons internes sont inférieurs à 0,31 et les rayons externes sont inférieurs à 0,44 (Tableau 13), ce qui définit des groupes compacts et isolés des autres. Les éléments *MLE* de cette sous-famille, provenant tous des diatomées à l'exception de Ds2833, apparaissent distincts des *MLE* des végétaux terrestres comme nous l'avons observé préalablement avec l'analyse phylogénétique. La séquence Ds2833 doit sans doute être très divergente par rapport aux autres *MLE* d'Angiospermes.

- La sous-famille Plantis est composée de 28 groupes nommés en fonction des principales espèces hôtes : BAM (Bambusa), MAY (Maïs), BOU (bambou), PUB (pubescens), VUL (vulgaris), ZEA (Zea), PRO (procumbens), TRI (Triticum), RAD (radiciflora), RAV (betterave), PIS (Pisum), ORY (Oryza), NAN (nana), MED (Medicago), LON (longifolia), IRI (Iris), GAR (vulgaris), HUM (humilis), HOR (Hordeum), DUR (dura), DAC (dactylon), COR (cormar), BIC (bicolor), BET (Beta), AVE (Avena), AES (aestivum).

Certains *MLE* ne rentrent pas dans les tribus définies mais sont très proches et se rattachent à la branche correspondante, c'est le cas des cinq tribus : BOU, HUM, PIS et TRI et MED (Figure 42, en vert).

Le fait que les *MLE* de diatomées soient compartimentés au sein de deux clades montre une divergence par rapport aux *MLE* de plantes terrestres qui sont eux mêmes rassemblés dans la sous-famille Plantis. Ceci confirme la distinction entre ces deux types de *MLE* différents.



**Figure 43 : Structure de la sous-famille Dia, PAL et Sta.**

Gros plan sur la classification des *MLE* de diatomées dans l'analyse DTT491. On peut noter que les lignées *MLE-Di1*, *MLE-Di2*, *MLE-Di3*, identifiées dans l'analyse phylogénétique précédente, sont conservées. Une exception est observable, ACU ici incluse dans la première branche de la sous-famille Dia est aussi incluse dans *MLE-Di2*. Cette branche et la position de ces séquences *MLE* sont ambiguës, elle met en évidence la divergence entre le groupe *MLE-Di2* de l'analyse phylogénétique et le groupe Dia de la classification. C'est la seule divergence entre les regroupements des séquences résultant de chacune des deux approches.

### II.2.3 Position des MLE des diatomées marines dans la classification DTT491.

**Tableau 14 : Position des MLE de diatomées marines dans la superfamille *Tc1-mariner* au sein de la classification DTT491.**

Répartition des séquences nucléotidiques *MLE* dans les groupes de diatomées formés dans la classification DTT491. \* indique les groupes constitués de séquences *MLE* issues d'une seule et même espèce.

Sous-famille	Tribus	Effectif	Séquences <i>MLE</i> du groupe	Hôte
DTTChlPAL*		4	<i>Nipamar11, Nipamar12, Nipamar13, Nipamar21,</i>	Diatomée
DTTChlSta*		6	<i>Staucmar11, Staucmar12, Staucmar13, Staucmar14, Staucmar21, Staucmar23</i>	Diatomée
<i>Chl1</i>		= 10		
DTTChlDia	ACU*	6	<i>Acutimar11, Acutimar12, Acutimar13, Acutimar14, Acutimar15, Acutimar16</i>	Diatomée
DTTChlDia	AUR	7	<i>Acomar13, Odaumar14, Acomar11, Odaumar11, Odaumar12, Acomar12, Odaumar13</i>	Diatomée
DTTChlDia	COF	2	<i>Hacrumar31, Acomar21</i>	Diatomée
DTTChlDia	ENT	5	<i>Entopalmar41, Entopalmar21, Acomar41, Entopalmar31, Hacrumar21</i>	Diatomée
DTTChlDia	LAE*	5	<i>Nilamar11, Nilamar12, Nilamar31, Nilamar41, Nilamar51</i>	Diatomée
DTTChlDia	TRI	11	<i>Acomar31, Entopalmar11, Phatrimar14, Entopalmar12, Phatrimar11, Phatrimar12, Phatrimar13, Phatrimar21, Hacrumar11, Hacrumar12, Hacrumar13</i>	Diatomée
DTTChlDia	WEI*	6	<i>Thaweismar11, Thaweismar12, Thaweismar14, Thaweismar15, Thaweismar16, Thaweismar17</i>	Diatomée
DTTChlDia	NIT*	2	<i>Nilamar21, Nilamar22</i>	Diatomée
DTTChlDia		1	<i>Ds2833</i>	Poaceae
DTTChlDia		=45		

#### **Structure des sousfamilles DTTChlDia, PAL et Sta**

Il apparaît ici que toutes les séquences issues de diatomées sont réparties en trois sous-familles (Dia, PAL et Sta) totalisant dix groupes. Ces deux sous-familles sont donc des groupes bien isolés et forment le premier clade de la famille Chlorophyllis. Ce résultat est cohérent avec ceux obtenus précédemment lors des analyses phylogénétiques des *MLE* de diatomées, les *MLE* issus de *Stauroneis constricta* et *Nitzschia palea* présentent des séquences très mutées et se positionnent toujours dans un ou plusieurs groupes isolés du reste des *MLE* de diatomées. Ils correspondent exactement à la première lignée de *MLE* décrite précédemment *MLE-Di1* (Figure 40, Partie II.1 de ce chapitre).

La sous-famille Diatomis se divise également en deux groupes qui pourraient représenter les deux lignées de *MLE-Di1* et *MLE-Di2* (Figure 43). La première se compose des tribus AUR et ACU, mais seul AUR correspond à la lignée *MLE-Di3* décrite

précédemment. Les *MLE* de la tribu ACU étaient inclus dans la lignée *MLE-Di2*. Cette lignée est retrouvée presque dans son intégralité avec la deuxième branche de la sous-famille DTTChlDia se composant de la majorité des séquences *MLE* de diatomées.

### **Position des *MLE* de diatomées marines dans la superfamille *Tc1-mariner***

Les *MLE* de diatomées marines sont rassemblés au sein de deux groupes indépendants dans la famille Chlorophyllis. Ce résultat est en accord avec les analyses phylogénétiques (partie II.1 de ce chapitre) et confirme que les *MLE* de diatomées marines présentent des divergences par rapport aux *MLE* de plantes terrestres. En outre, ils sont regroupés avec les *MLE* de plantes dans la famille Chlorophyllis parmi lesquels on compte 17 séquences à ORF complètes. Aucune séquence *MLE* de diatomée n'est proche des familles MAR (DD34D), maT (DD37D) ou Tcu (DD34E) suggérant que les séquences *MLE* de diatomées sont assez proches des *MLE* de plantes terrestres pour envisager qu'ils puissent appartenir également au groupe DD39D de la superfamille *Tc1-mariner*.

## II.3 Discussion

### II.3.1 Position phylogénétique et évolution des *MLE* de diatomées marines par rapport aux *MLE* de végétaux terrestres

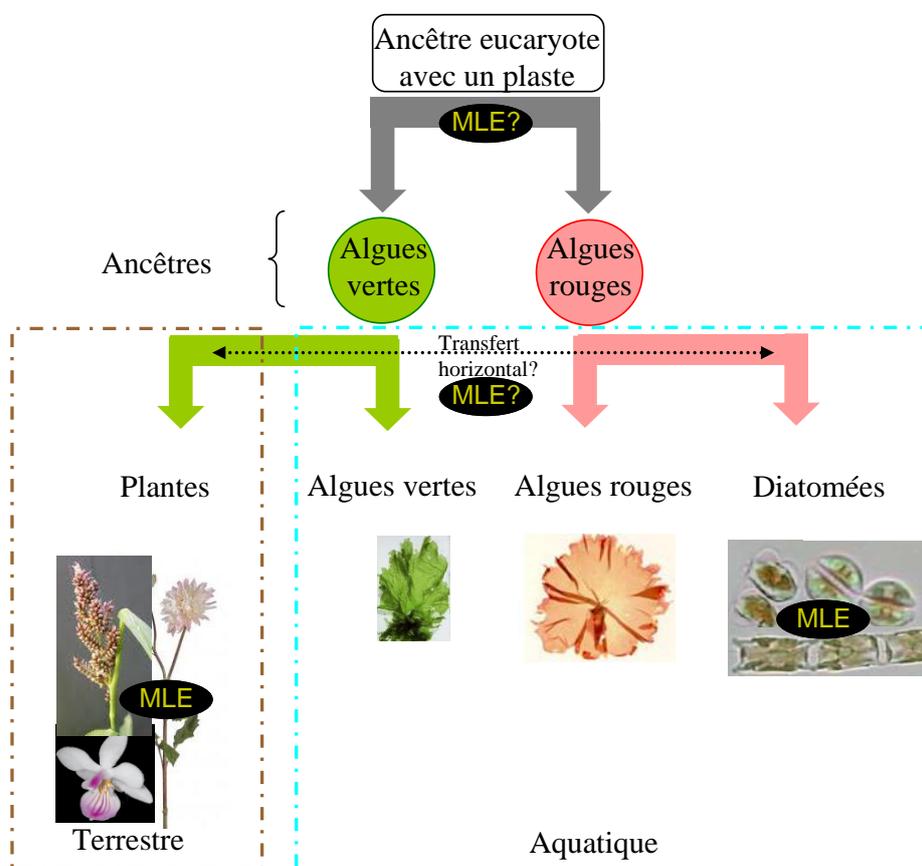
Chaque méthode de phylogénie a ses avantages et ses inconvénients qui conduisent à des erreurs ou des arbres qui peuvent ne pas correspondre à la véritable histoire évolutive des éléments étudiés (pour revue : Holder et al., 2003). Les arbres phylogénétiques ont été réalisés sur les fragments de gènes tronqués codant la transposase (de 380 pb) traduits en acides aminés selon trois méthodes de phylogénie différentes (NJ, MP, et IB). Les arbres consensus résultant des trois méthodes montrent une topologie similaire, ce qui valide les groupes formés et les positions relatives des *MLE*. Les *MLE* de diatomées sont distribués dans trois grands groupes *MLE-Di1*, 2 et 3. Aucune des lignées établies n'est homogène, ce qui traduit l'ancestralité de la lignée et les épisodes de diversification des *MLE* qui se sont produits avant la formation de chacune des espèces hôtes. Les analyses phylogénétiques révèlent que les lignées de *MLE* de diatomées sont distantes de la lignée unique des *MLE* de plantes terrestres. Celle-ci est composée des *MLE* d'Angiospermes terrestres (Feschotte et Wessler, 2002; Jacobs et al., 2004; Macas et al., 2005). Les 7 groupes de *MLE* de diatomées identifiés en partie III ont été rassemblés dans trois grands groupes *MLE-Di* 1 à 3 qui pourraient correspondre à trois lignées ancestrales. Des séquences isolées ont été rapprochées et incluses dans chacune des trois lignées témoignant de leur distance par rapport aux *MLE* de plantes

terrestres. De plus, chacune de ces lignées est monophylétique et se compose de *MLE* provenant d'espèces multiples. Ceci nous semble confirmer l'hypothèse de trois éléments ou familles d'éléments *MLE* de diatomées présents ancestralement, ayant divergé pour aboutir au polymorphisme de séquence observé dans la partie III. Chacune de ces lignées est indépendante l'une de l'autre, ce qui traduit une présence très ancienne des *MLE*, sûrement préexistante à la radiation évolutive des diatomées centriques et pennées ayant eu lieu au pré-Crétacé (Harwood et Gersonde, 1990).

Ces lignées *MLE* ont également évolué indépendamment de la lignée des végétaux terrestres, bien qu'elles soient regroupées finalement au sein d'un groupe monophylétique distinct des *MLE* animaux formant la racine des arbres phylogénétiques.

Les *MLE* de diatomées apparaissent donc clairement distincts des *MLE* isolés chez les végétaux terrestres. Comme Feschotte et Wessler (2002) l'ont montré, le groupe composé des *MLE* d'Angiospermes est monophylétique. L'analyse phylogénétique effectuée ici, comprenant également de nouveaux éléments *MLE* provenant du pois (Macas et al., 2005), de la betterave (Jacobs et al., 2004) et des diatomées marines, supporte également l'hypothèse de la monophylie des *MLE* de végétaux. Les *MLE* de diatomées ne s'insèrent à aucun moment dans un groupe avec des *MLE* de végétaux terrestres et les valeurs de confiance (bootstrap et probabilités postérieures) des trois méthodes soulignent la divergence des *MLE* de diatomées marines par rapport aux végétaux terrestres. Ceci suggère que les *MLE* des diatomées marines présenteraient des spécificités qui pourraient être le reflet de leur environnement. Comme nous l'avons déjà évoqué, le polymorphisme des séquences *MLE* chez leurs hôtes pourrait être en fait un polymorphisme adaptatif à défaut d'être le résultat d'événements de spéciation. Cette même question a été soulevée dans l'étude d'un *mariner* isolé chez un crabe marin issu des sources hydrothermales, *Bythmar1* (Halaimia-Toumi et al., 2004) et des crustacés littoraux (Bui et al., 2007). Les auteurs ont mis en évidence une spécificité dans la composition des séquences *MLE* en GC. Le taux de GC des *MLE* est supérieur chez les organismes hydrothermaux (Halaimia-Toumi et al., 2004) par rapport à celui des organismes littoraux lui-même supérieur à celui des organismes terrestres (Bui et al., 2007). Ces spécificités pourraient représenter des adaptations aux contraintes du milieu hydrothermal ou du milieu marin comme le suggère les auteurs. Les résultats présentés ici ne nous permettent pas d'accéder à cette information car nous n'avons mis en évidence qu'une portion de la séquence *MLE* des diatomées. Les spécificités des *MLE* de diatomées marines résident dans un polymorphisme important entre les motifs conservés des séquences. Ceci peut être également le résultat d'une invasion ancienne des diatomées par les *MLE* et l'accumulation

des mutations au cours de leur longue évolution comme les résultats précédents nous l'ont suggéré. La distance entre les *MLE* de diatomées marines et de végétaux terrestres peut s'expliquer par le fait que les diatomées se sont séparées de la lignée verte aboutissant aux plantes terrestres, il y a environ 245 millions d'années, période où l'endosymbiose secondaire aurait eu lieu conduisant à l'apparition des diatomées (Medlin et Kaczmarska, 2004). Bien que ces considérations soient issues d'estimations, le fait que les diatomées centriques aient été découvertes dans des sédiments fossiles datés du crétacé prouve que l'apparition des diatomées est antérieure à cette période (Harwood et Gersonde, 1990). Les diatomées sont âgées d'au moins 165 millions d'années, soit beaucoup plus tôt que les Angiospermes qui se sont diversifiés il y a environ 130 millions d'années (Falkowsky et Raven, 1997).



**Figure 44 : Représentation schématique de l'évolution conjointe des diatomées et des plantes.**

De ce fait, les lignées des *MLE* des diatomées et des plantes terrestres qui sont indépendantes pourraient correspondre au schéma évolutif de la lignée végétale car les diatomées et les plantes terrestres forment deux lignées évolutives indépendantes (Figure 44). La lignée végétale dériverait d'un ancêtre eucaryote primaire qui se serait diversifié en trois lignées des microalgues : les microalgues vertes dont certaines sont à l'origine des plantes terrestres et deux autres lignées parmi lesquelles celle des microalgues rouges à l'origine des

diatomées marines (Bhattacharya et Medlin, 1998, Falkowsky et al., 2004). Toutefois, si l'on considère que les *MLE* étaient présents très tôt dans l'ancêtre commun aux plantes et diatomées, il faut également considérer que les *MLE* ont évolué sur une période d'environ 300 millions d'années et dans ce schéma, on peut se demander quels sont les processus à l'origine de leur conservation dans les génomes des végétaux. Feschotte et Prittham (2007) suggèrent également que certaines superfamilles d'éléments telles que *Tc1-mariner* doivent avoir une origine très ancienne dans l'évolution des organismes vivants et que les principaux types de transposons pourraient être présents ancestralement avant la divergence des eucaryotes. Cela pourrait alors être le cas des *MLE*. Comme nous l'avons déjà évoqué, une alternative souvent utilisée pour expliquer la distribution sporadique ou l'absence dans certains clades des ET est le transfert horizontal. Celui-ci est supposé être à l'origine de nombreux gènes des Bacillariophycées. Ce mécanisme serait à l'origine de près de 5% des gènes de *P. tricornutum* qui proviendraient de bactéries (Bowler et al., 2008). De même, 56% des gènes qui composent le génome de *T. pseudonana* seraient d'origine fongique (Phytophthora), d'algues rouges et de cyanobactéries (Armbrust et al. 2004). Ainsi, nous pouvons supposer que des transferts horizontaux ont pu permettre la prolifération des *MLE* dans les organismes photosynthétiques mais aucune preuve n'est apportée ici, et la recherche des *MLE* chez différents types d'algues pourrait apporter une meilleure compréhension de l'évolution des *MLE* dans la lignée végétale.

Les méthodes de phylogénie classique ont permis de distinguer les *MLE* des diatomées par rapport aux *MLE* des végétaux mais les relations phylogénétiques de ces éléments sont étudiées sur une faible portion de la séquence alors que les quelques *MLE* végétaux connus présentent des tailles allant jusqu'à 7 Kpb. Les fragments de transposases de quelques éléments complets caractérisés par la triade catalytique DD39D tels que *Vulmar1* (Jacobs et al., 2004), *Psmar1* (Macas et al., 2005), *Osmar1* (Tarchini et al., 2000) et *Soymar1* (Jarvick et Lark, 1998) ont été intégrés dans l'analyse. Ces séquences se positionnent au milieu de la lignée des plantes terrestres avec les séquences d'éléments tronqués suggérant que tous les *MLE* de cette lignée pourraient appartenir au groupe des *MLE* de plantes aussi nommés ITmDD39D en raison de la structure du site catalytique (Jarvick et Lark, 1998; Shao et Tu, 2001). Les *MLE* de diatomées ne s'intégrant pas dans le clade des *MLE* d'Angiospermes nous ne pouvons pas conclure quant à la nature probable de la triade catalytique.

Des inconvénients sont apparus lors de l'analyse des *MLE* de diatomées marines par rapport aux *MLE* des végétaux terrestres ou aux autres éléments de la famille *Tc1-mariner* avec les approches de phylogénie :

- seules les séquences de longueur similaires peuvent être comparées. Ainsi si la longueur de deux séquences similaires diffère, il est impossible de les comparer sans qu'un biais soit ajouté sous forme de gap d'alignements.
- les séquences tronquées de *MLE* ont dû être traduites en acides aminés car les divergences importantes observées en nucléiques ne permettaient pas d'obtenir des arbres avec des fortes valeurs de confiance en bootstrap ou en probabilité postérieure.
- comparer des séquences de *MLE* végétaux avec celles des autres familles de la superfamille *Tc1-mariner* était irréalisable du fait de la difficulté d'alignement. En effet, les différentes familles de *Tc1-mariner* présentent de nombreuses divergences au niveau protéique (motifs, longueur...) et donc au niveau nucléique.

Nous avons donc fait une méta-analyse des *MLE* de végétaux. Celle-ci a été réalisée au moyen d'un nouvel outil de classification des ET, la méthode UPGM-VM décrite par Rouault et Capy (2008). Cette méthode innovante permet d'inclure les éléments de tailles variables et donc de comparer des séquences tronquées avec des séquences complètes. Nous avons ainsi pu comparer toutes les séquences nucléiques *MLE* présentes dans la littérature sans craindre les problèmes d'alignements dus aux longueurs différentes des séquences. Cette nouvelle méthode de classification nous a permis d'étudier la position des *MLE* de diatomées marines dans la superfamille *Tc1-mariner*. Elle n'a pas pour but d'établir une phylogénie mais de hiérarchiser les éléments en fonction de leurs similarités. Cette étude a révélé la formation d'une nouvelle sous-famille de *MLE* de diatomées incluse dans la famille redéfinie des *MLE* végétaux dans la superfamille *Tc1-mariner*.

### II.3.2 Classification des *MLE* de diatomées marines par rapport aux *MLE* de plantes dans la nouvelle famille *Chlorophyllis* de la superfamille *Tc1-mariner*

La nouvelle famille des *MLE* végétaux a été nommée *Chlorophyllis* car elle est constituée des *MLE* isolés chez les plantes terrestres mais également de l'ensemble des séquences identifiées dans le présent travail, chez les diatomées marines. Cette famille a été définie par son positionnement à l'intérieur de la superfamille *Tc1-mariner* et ses caractéristiques la placent comme un groupe distinct isolé des autres familles. Cette famille intègre les séquences appelées jusqu'ici *MLE* de plantes ou *Itm* DD39D telles que *Soymar1*, *Vulmar1*, les *Osmar* et *Atmar* (Shao et Tu, 2001; Feschotte et Wessler, 2002; Jacobs et al.,

2004; Brillet et al., 2007) ce qui renforce l'hypothèse évoquée précédemment, les séquences appartenant à cette famille seraient potentiellement caractérisées par la triade DD39D. Les *MLE* de végétaux semblent appartenir au groupe caractérisé par la triade DD39D. Néanmoins, les *MLE* de diatomées étant compartimentés au sein de trois autres sous-familles, nous ne pouvons pas en déduire que leur structure puisse être également caractérisée par la triade DD39D. Cette caractéristique reste donc à déterminer chez les *MLE* de diatomées.

De même que les Diatomis sont distincts des éléments Plantis issus de végétaux terrestres, les *MLE* d'hôtes animaux terrestres et marins sont divisés en deux familles. Les *MLE* obtenus à partir de crabes marins (Halaima- Toumi et al., 2004; Bui et al., 2008) forment la famille Atl ou Atlantis et la famille Mar comprend les *mariner* obtenus chez les animaux terrestres. Cette résultat a d'ailleurs été confirmée par les travaux de Rouault et al. (2009) qui décrivent la même topologie obtenue à partir d'une classification basée exclusivement sur des séquences longues (> 700 pb) d'éléments *mariner* d'origine animale. Toutefois ces deux familles sont reliées à la même branche au niveau supérieur de la même manière que les Plantis et les *MLE* de diatomées sont rattachés à la famille Chlorophyllis. Ces résultats sont cohérents avec la distance phylogénétique des *MLE* de diatomées marines et de plantes terrestres.

L'environnement aquatique ou terrestre de l'hôte semble être ce qui différencie les *MLE* de crabes marins et les *MLE* d'animaux terrestres. Les études sur les *MLE* végétaux et la fonctionnalité de ceux-ci sont rares (Yang et al., 2006, 2009). Nous n'avons donc pas pu découvrir la responsabilité des variations de séquences observées dans la séparation *MLE* de diatomées/*MLE* de plantes terrestres.

Au sein de la famille Chlorophyllis, les tribus des sous-familles Diatomis et Plantis ont été conservées entre les analyses de manière invariable qu'on ajoute ou qu'on enlève des éléments provenant de végétaux terrestres. Les structures de Clan des tribus contenant les éléments provenant de diatomées marines ou de végétaux terrestres, mais aussi des familles constituant le groupe externe n'ont pas varié à l'exception de Sta. La sous-famille Sta qui avait une structure de Clan a perdu ce statut lors de l'ajout de nouvelles séquences *MLE* de plantes. Cette sous-famille n'est donc pas bien définie, sans doute est-ce dû à un faible échantillonnage ou à l'extrême mutation des séquences qui la composent par rapport aux autres *MLE* de végétaux. Il apparaît également à l'issue de l'analyse DTT491 que la famille Chlorophyllis elle-même présente une structure de Clan. Cela implique qu'elle est très éloignée des autres familles d'ET et donc bien isolée dans l'espace des ET.

De plus, les séquences végétales qui ne sont pas incluses dans la famille Chlorophyllis sont exclusivement des séquences complètes, cette observation suscite une interrogation sur la possibilité qu'avec une amélioration de la quantité et de la qualité (séquences entières) de l'échantillon, cette famille se révèle moins distante et que la propriété de Clan ne soit pas conservée. En outre, les valeurs très voisines des rayons internes et externes de la famille Chlorophyllis indiquent que la structure de Clan risque d'être remise en cause par l'ajout de groupes de séquences intermédiaires qui se positionneraient entre Chlorophyllis et les familles voisines. L'origine de ce biais pourrait être l'échantillonnage ou la méthode de classification.

Les séquences intégrées dans la classification DTT491 sont des séquences connues et publiées de *MLE* obtenues à partir d'organismes photosynthétiques. Pourtant cet échantillon ne recouvre pas toute la diversité végétale potentielle puisque 10 espèces de diatomées marines sur 100 000 espèces connues à ce jour (Round et al., 1990) et seules des familles botaniques appartenant à l'embranchement des Angiospermes sont représentées. L'ajout de séquences provenant de cyanobactéries, d'algues rouges, de Bryophytes, de Ptéridophyte etc., pourraient remettre en cause les familles de transposons *MLE* définies dans la classification et notamment la famille Chlorophyllis. C'est ce que suggère la position de la seule séquence *MLE* de l'œillet qui apparaît intégrée à la sous-famille Diatomis. Remarquons que cet élément était déjà placé à l'extérieur du groupe monophylétique des *MLE* de plantes dans l'arbre de Feschotte et Wessler (2002). Il est probable que cette position soit un artéfact dû à l'isolement de cette seule et unique séquence *MLE* d'œillet. L'échantillonnage des éléments *Tc1-mariner* chez les organismes photosynthétiques est donc à améliorer pour avoir une idée réaliste de la diversité de cette superfamille. En effet, d'autres éléments issus de végétaux appartenant à la superfamille *Tc1-mariner* ne sont pas inclus dans Chlorophyllis tels que les éléments formant la famille nommée Lem ou Lemi issus de la plante modèle *A. thaliana* (Feschotte and Mouchès, 2000), des éléments *MLE* issus de cette même plante (*Atmar5*, 8 et 9) et des *MLE* isolés chez le riz (*O. sativa*) ne s'incluent dans aucune famille. La taille des séquences des éléments isolés est supérieure à 700 pb, ce qui soulève le problème lié à la taille des séquences.

Dans la famille Chlorophyllis, toutes les tribus composées d'éléments de diatomées et de nombreuses tribus issues d'éléments de plantes sont uniquement composées de séquences courtes (<700 pb). La partie commune de toutes ces séquences se situe entre les deux premiers résidus aspartates du domaine très conservé de la triade catalytique DDD (Plasterk, 1999, Doak et al., 1994). Bien que cette partie commune soit conservée sous forme de motifs protéiques, il a été possible de mettre à profit la variabilité nucléique observée pour constituer

des groupes : sous-familles ou tribus au sein de la classification. Parmi les 22 séquences longues (de 700 à 6444 pb) intégrées à cette l'analyse DTT491, 17 séquences se trouvent localisées dans la sous-famille *Plantis* et 14 de ces séquences sont intégrées dans des tribus structurées en Clan. Prenons l'exemple des sous-familles *Mauritiana* et *Cecropia* de la famille *mariner*, il apparaît que les TIR sont très conservées dans les tribus et très différents d'une tribu à une autre (J.D. Rouault, communication personnelle). Dans ce contexte, la présence de *MLE* végétaux complets ne peut que renforcer la formation des groupes déjà observés. Dans l'éventualité de disposer de séquences *MLE* allongées (> 401 pb) de diatomées marines, nous pouvons nous demander si la topologie observée ici (3 Sous-familles, *Dia*, *PAL* et *Sta*) sera conservée ou si elle sera restructurée pour ne former qu'un seul et même groupe. La classification agrégative résultant de la méthode UPGM-VM, à la différence des méthodes phylogénétiques, permet donc de classer des éléments de tailles différentes. Ce résultat est d'un grand intérêt pour étudier les familles d'ET peu décrites ou bien pour la classification d'éléments tronqués, de MITE, ou encore de séquences fossiles.

Cette classification comprenant l'intégralité des séquences *MLE* végétales à notre connaissance permet de synthétiser les connaissances récemment apportées sur les *MLE* végétaux. En effet, les séquences *MLE* d'Angiospermes intégrées dans cette analyse sont positionnées de manière parfaitement concordante avec les analyses phylogénétiques. En effet, les groupes formés avec les séquences issues des travaux de Feschotte et Wessler (2002) et de Jacobs et al. (2004), sont parfaitement retrouvés dans la classification DTT491 (Annexe 7). Ceci nous permet aussi de proposer, une nouvelle appellation « DTTChIPln » pour les *MLE* de plantes et DTTChIDia pour la plupart des *MLE* de diatomées. En effet, les éléments issus de végétaux ont été nommés de façon controversée sous le nom de *mariner* de plantes ou *MLE* de plantes (Feschotte et Wessler, 2002; Jacobs et al 2004, Macas et al., 2005) d'autres auteurs proposent de séparer totalement les familles d'éléments en fonction de la triade catalytique, tels que les *MLE* de végétaux soit nommé *ITm* DD39D (Shao et Tu 2001, Brillet et al. 2007). Le nom « *MLE* de plantes » apparaît comme un nom imprécis qui ne permet aucun rapprochement avec les autres familles de la superfamille *Tc1-mariner*. Le terme *ITm* DD39D n'est pas attribuable à la majorité des séquences tronquées obtenues chez les organismes photosynthétiques et qui ne comprennent que les deux premiers résidus aspartates de la triade catalytique. *ITm* DD39D fait référence à une structure commune basée sur la distance entre les deux derniers D de la triade. Il est donc impossible de classer les éléments de moins de 500 pb dans cette famille. La classification DTT491 procure l'énorme avantage de classer toutes les séquences.

Les principaux points mis en évidence dans la deuxième partie de ce travail sont les suivants :

- la comparaison des *MLE* des diatomées avec les *MLE* des Angiospermes a mis en évidence leur appartenance à un seul et même groupe monophylétique des *MLE* de végétaux.

- les séquences *MLE* de diatomées précédemment divisées en 7 groupes se sont rassemblées en trois lignées évolutives. Ces trois lignées sont monophylétiques et ont évolué indépendamment des *MLE* de plantes terrestres qui forment une seule lignée. Les différences entre les *MLE* de diatomées et de plantes terrestres pourraient donc être le résultat de l'évolution divergente des hôtes.

- les *MLE* des diatomées marines sont structurés en trois lignées indépendantes suggérant une invasion ancienne et une longue évolution des *MLE* chez les diatomées.

De la même manière, la classification des *MLE* de diatomées dans la famille *Tc1-mariner* (DTT) a permis de démontrer que les *MLE* de diatomées se classaient dans une même famille Chlorophyllis (DTTChl) avec les *MLE* d'Angiospermes. Chlorophyllis appartient à la superfamille *Tc1-mariner* mais se positionne de manière distante par rapport aux autres membres de cette superfamille. Chlorophyllis est divisée en trois branches dont une regroupe les *MLE* de plantes terrestres au sein de la sous-famille Plantis (DTTChlPln) et les deux autres branches étant constituées des *MLE* de diatomées marines regroupées en trois sous-familles Dia, PAL et Sta.

Les analyses phylogénétiques et la classification montrent que les *MLE* des diatomées sont divergents des *MLE* des plantes terrestres, toutefois on peut se demander si cette divergence est due aux environnements différents des hôtes ou bien si c'est le résultat de l'évolution séparée des diatomées et des plantes. Il est donc nécessaire d'améliorer la qualité de l'échantillonnage au niveau des *MLE* d'origine marine mais également au niveau des *MLE* de végétaux terrestres.

De plus les *MLE* de diatomées semblent appartenir au groupe des *MLE* DD39D, toutefois, pour confirmer cela il est nécessaire de trouver des copies complètes de *MLE* dans les génomes de diatomées. Dans la suite de ce travail, nous avons recherché l'expression de *MLE* qui caractériserait la présence potentielle de séquences complètes chez les diatomées.

### III Impact de stress thermiques courts sur l'expression des *MLE* de trois diatomées marines

Les séquences *MLE* mises en évidence dans la partie I de ce chapitre montrent des caractéristiques qui permettent d'envisager la présence de séquences complètes dans les génomes des diatomées car :

- les séquences traduites *in silico* présentent de nombreuses ORF partielles non-interrompues qui laissent supposer qu'elles peuvent être traduites et produire une transposase fonctionnelle.
- chez certaines espèces telles que celles du genre *Amphora*, les séquences sont facilement amplifiées par PCR ce qui indique la présence de nombreux éléments dans le génome.

La recherche de séquences complètes dans le génome séquencé de *P. tricornutum* avec des méthodes bio-informatiques n'ayant pas donné de résultat positif, nous avons choisi de rechercher l'expression des *MLE* par des techniques moléculaires. Ceci, en conditions standards et en soumettant les diatomées à différentes conditions thermiques potentiellement stressantes (le stress pouvant être un facteur d'induction des ET). Trois espèces ont été sélectionnées, la diatomée modèle *P. tricornutum*, ainsi que les deux espèces du genre *Amphora* qui sont adaptées à des climats différents. Les résultats présentés dans la partie suivante concernent l'étude préliminaire:

- de l'impact des conditions thermiques sur les paramètres physiologiques (croissance et photosynthèse) des trois espèces de diatomées.
- de la recherche de l'expression des *MLE* (étude qualitative) chez les diatomées en conditions standards et de stress thermiques.

#### III.1 Réponses physiologiques des trois diatomées soumises à des stress thermiques

L'impact de différentes conditions thermiques a été mesuré au niveau de la croissance et de l'activité photosynthétique chez les trois espèces de diatomées sélectionnées *Phaeodactylum tricornutum*, *Amphora coffeaeformis* et *Amphora acutiuscula*.

Les températures usuelles de culture pour *P. tricornutum* et *A. coffeaeformis* sont de 16 °C et de 24 °C pour *A. acutiuscula*. Les diatomées ont donc été cultivées pendant 8 jours à leurs températures usuelles de culture et en fin de phase exponentielle de croissance, elles sont ensuite placées pendant 5 h aux températures 4, 8, 16, 24 et 32 °C pour effectuer les chocs thermiques.

### III.1.1 Effets des stress thermiques sur la croissance

Afin de déterminer les températures stressantes pour chacune des espèces de diatomées, des mesures de croissance ont été effectuées. Pour cela, des cultures ont été réalisées pendant 8 jours aux températures 8, 16 et 24 °C. Les courbes de croissance de chaque espèce aux différentes températures sont présentées dans les Figures 45 à 47.

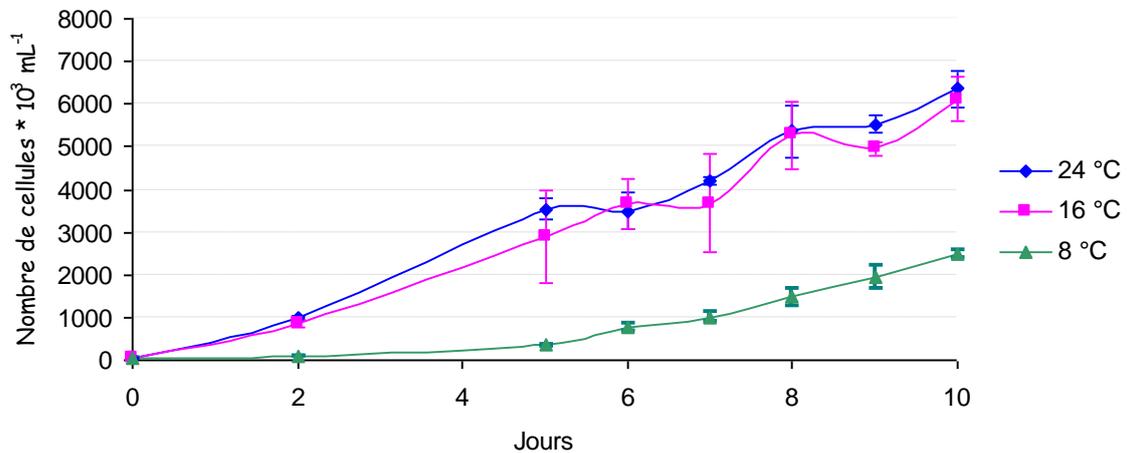


Figure 45 : Courbes de croissance de *Phaeodactylum tricornutum* cultivée à 8, 16 et 24 °C. La température standard de culture est 16°C.

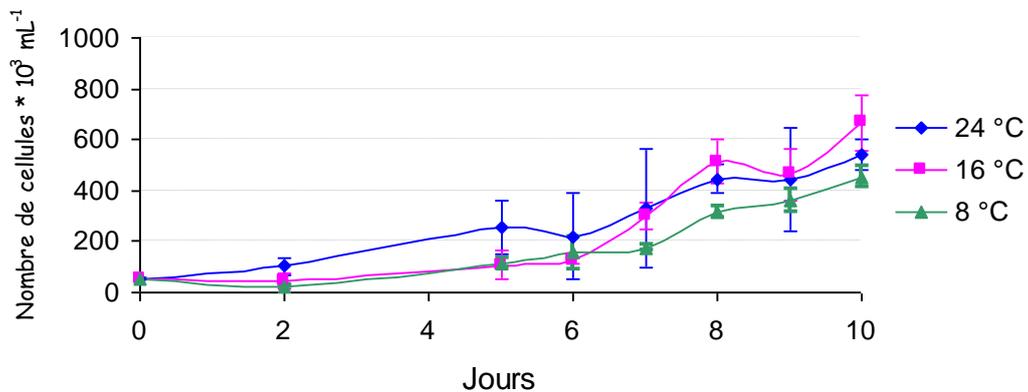


Figure 46 : Courbes de croissance d'*Amphora coffeaeformis* cultivée à 8, 16 et 24 °C. La température standard de culture est 16°C.

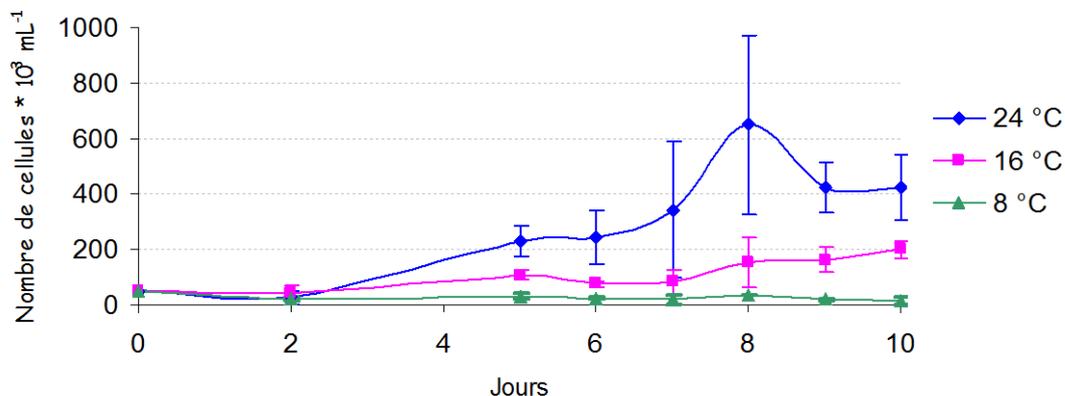


Figure 47 : Courbes de croissance d'*Amphora acutiuscula* cultivée à 8, 16 et 24 °C. La température standard de culture est 24°C.

Les courbes de croissance de *P. tricornutum* à 16 et 24 °C sont similaires et nettement supérieures à celle obtenue à 8 °C. En effet, après deux jours de culture à 16 et 24 °C la densité des suspensions de *P. tricornutum* est supérieure à 10<sup>6</sup> cellules/mL alors que les cultures exposées à une température de 8 °C n'atteignent cette densité cellulaire qu'après 6 jours de culture. A la fin des 8 jours de culture, on observe qu'à 8 °C la densité cellulaire est deux fois moins importante que celle des cultures de *P. tricornutum* à 16 et à 24 °C. Nous pouvons donc en déduire que *P. tricornutum* est stressée à 8 °C.

Pour *A. coffeaeformis*, la croissance mesurée aux trois températures ne diffère pas significativement excepté en fin de phase exponentielle (entre 6 et 10 jours) où à 16 °C *A. coffeaeformis* semble montrer une meilleure croissance.

La croissance d'*A. acutiuscula* à 24 °C est supérieure à celle enregistrée à 16 et à 8 °C dès 4 jours de culture. La température de 8 °C est particulièrement défavorable à la croissance de cette espèce. 8 °C représente donc une condition stressante pour *A. acutiuscula* car la densité cellulaire n'atteint pas 100 000 cellules/mL. Après une stabilisation de la densité cellulaire, une décroissance est observée après 8 jours de culture.

Ces résultats nous ont permis de déterminer des températures de chocs thermiques de 5 h. Les températures 16 et 24 °C permettent la croissance et son maintien chez les trois espèces. La température 16 °C entraîne un ralentissement de la croissance chez *A. acutiuscula* par rapport la croissance obtenue avec sa température de culture (24 °C). Ceci suggère que cette température représente un niveau de stress suffisant pour affecter la croissance de cette espèce. La condition 8 °C est apparue comme une température induisant un ralentissement et une diminution de la croissance cellulaire chez les espèces *P. tricornutum* et *A. acutiuscula*. Chez cette dernière, 8 °C représente une condition quasi létale. Cette condition thermique a donc été sélectionnée pour représenter un stress basse température. Toutefois dans le cas d'*A. coffeaeformis* et *P. tricornutum*, la croissance étant maintenue aux températures 8 et 24 °C et les chocs thermiques à appliquer étant de courtes durées, nous avons sélectionné des températures extrêmes pour effectuer des stress intenses et rapides. Les températures 4 et 32 °C ont été choisies respectivement pour induire un choc basse et haute températures. Ces deux conditions entraînent une diminution de la croissance chez les trois diatomées jusqu'à un niveau létal et le suivi de la croissance dans ces conditions n'est donc pas présenté ici.

### III.1.2 Effets des chocs thermiques sur l'intensité photosynthétique

Mesurer l'impact des expérimentations thermiques sur l'appareil photosynthétique permet de mettre en évidence une perturbation de la physiologie des diatomées. L'intensité photosynthétique (IP) des diatomées a été déterminée par la mesure du dioxygène dégagé pendant une heure rapporté à un mg de Chl *a*. Les résultats obtenus pour chaque espèce sont présentés dans la Figure 48.

Chez *P. tricornutum*, l'IP la plus élevée est enregistrée en condition témoin à 16° C. Les IP aux températures 8 et 24 °C sont plus faibles d'au moins 25 % par rapport au témoin (16 °C) soit respectivement 32% à 24°C et 69 % pour 8°C, contrairement à l'IP mesurée à 4 °C. Cette dernière est supérieure à la valeur obtenue après un choc thermique à 8 °C, ce résultat surprenant semble être due à un problème de calibration de l'oxymètre. La valeur d'IP la plus basse est enregistrée après 5 h de choc thermique à 32° C, en effet à cette température l'IP est nulle.

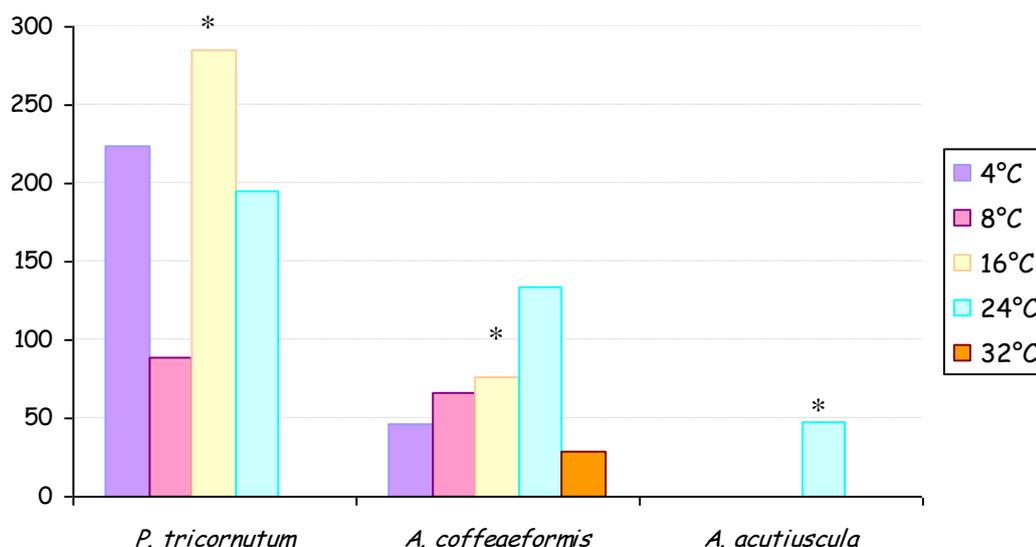
Chez *A. coffeaeformis*, la valeur de l'IP après 5h à 24 °C est la plus forte (134,35  $\mu\text{mold}'\text{O}_2 \text{ h}^{-1} \text{ mg}^{-1} \text{ chl}$ ). Elle est d'ailleurs doublée par rapport à l'IP mesurée à 4 et 32 °C. Les traitements thermiques 4, 8 et 32 °C induisent une diminution de l'IP par rapport au témoin 16 °C (75,85  $\mu\text{mold}'\text{O}_2 \text{ h}^{-1} \text{ mg}^{-1} \text{ chl}$ ) ce qui implique que ces traitements thermiques ont perturbé l'activité de photosynthèse des diatomées. Dans le cas du traitement à 24°C qui est supérieur à 16°C, on peut supposer que le stress a stimuler la photosynthèse chez *A. coffeaeformis*.

Chez *A. acutiuscula*, l'IP n'a pu être mesurée qu'à 24 °C, la condition témoin. Chaque choc thermique à 4, 8, 16 et 32 °C a induit une inhibition complète l'activité photosynthétique. Ici toutes les conditions apparaissent comme stressantes.

Les intensités photosynthétiques nulles enregistrée pour *P. tricornutum* après un choc thermique à 32 °C et pour *A. acutiuscula* à toutes les températures excepté chez le témoin, indiquent que l'activité photosynthétique permet de produire une quantité d'oxygène tout juste suffisante pour compenser la consommation d'O<sub>2</sub> par respiration (ainsi le bilan d'O<sub>2</sub> dégagé est nul). Cette réduction de l'IP indique que les diatomées sont en conditions stressantes.

Bien qu'il soit nécessaire de réaliser des réplicats afin de vérifier que les tendances observées sont significatives, la diminution de l'IP dans les conditions 8, 24 et 32 °C chez *P. tricornutum*, 4, 8 et 32 °C chez *A. coffeaeformis* et pour *A. acutiuscula* à 4, 8, 16 et 32 °C implique que ces températures représentent des conditions stressantes pour les diatomées.

Intensité photosynthétique  
( $\mu\text{mol O}_2 \text{ h}^{-1} \text{ mg}^{-1} \text{ chl } a$ )



**Figure 48 :** Intensité photosynthétique chez *Phaeodactylum tricornutum*, *Amphora coffeaeformis* et *Amphora acutiuscula* après 5 heures de traitement à 4, 8, 16, 24 et 32 °C. La température de culture (témoin) est indiquée par \*.

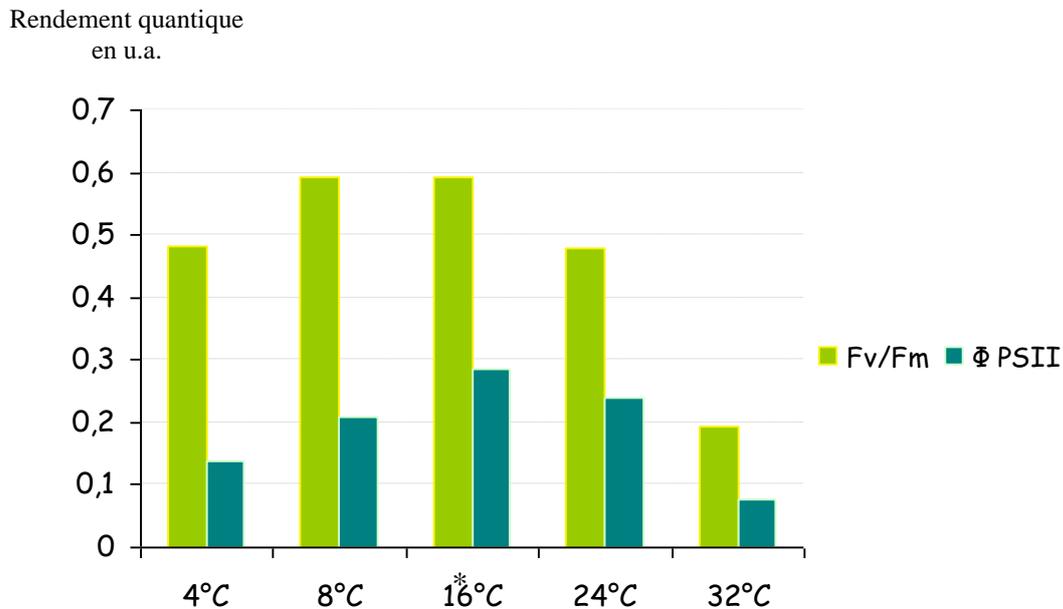
### III.1.3 Effets des chocs thermiques courts sur les paramètres de la fluorescence de la chlorophylle a

Mesurer les effets des chocs thermiques sur les paramètres de fluorescence permet de rendre compte de l'état fondamental des diatomées dans les conditions utilisées ici.

Les rendements quantiques maximum (Fv/Fm) et effectifs ( $\phi\text{PSII}$ ) de la Chl *a* ont été mesurés dans chaque condition thermique. L'évolution de ces rendements de fluorescence permet d'estimer l'état physiologique du photosystème II des diatomées.

Chez *P. tricornutum*, les plus fortes valeurs du rendement Fv/Fm sont enregistrées en condition témoin à 16 °C et après 5 h de stress à 8 °C (Figure 49). Les autres traitements thermiques entraînent une diminution de ce rendement qui est inférieur à 0,5 u.a. avec une diminution de moitié pour la condition 32 °C.

A la lumière, nous observons la même tendance en ce qui concerne le rendement quantique effectif ( $\phi\text{PSII}$ ). Le  $\phi\text{PSII}$  le plus élevé est obtenu en condition témoin et les autres traitements à 4, 8, 24 et 32 °C induisent une diminution du  $\phi\text{PSII}$  qui s'accroît avec l'écart thermique par rapport à 16 °C.

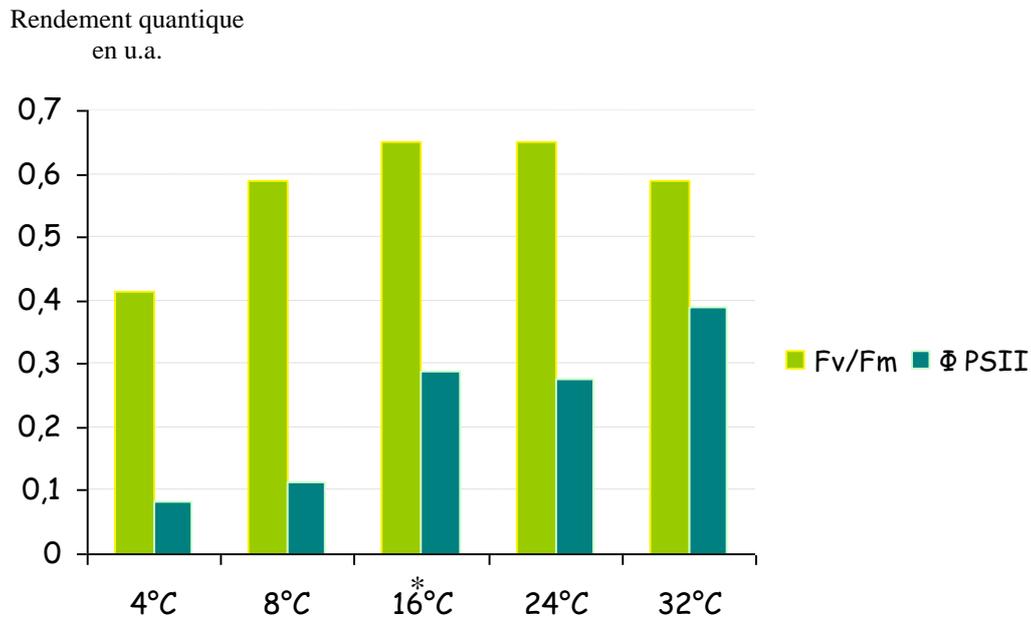


**Figure 49** : Evolution des rendements quantiques maximum (Fv/Fm) et effectifs ( $\phi$ PSII) chez *Phaeodactylum tricornutum* à 4, 8, 16, 24 et 32 °C

Les valeurs de Fv/Fm et  $\phi$ PSII sont exprimées en unités arbitraires (u.a). La température de culture (témoin) est indiquée par \*.

La valeur la plus élevée du rendement Fv/Fm de *A. coffeaeformis* est enregistrée pour le témoin à 16 °C et le traitement thermique à 24 °C (Figure 50). Les autres traitements thermiques induisent de faibles diminutions qui ne semblent pas être significatives puisque les rendements sont tous de l'ordre de 0,6 u.a. excepté à 4 °C où la diminution atteint 50 % par rapport au témoin.

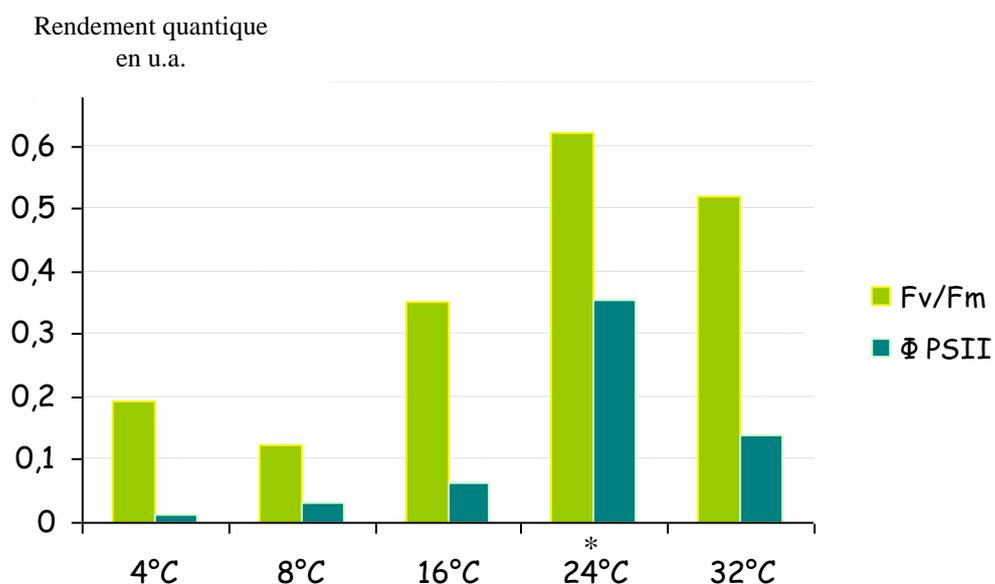
Le rendement quantique effectif ( $\phi$ PSII) augmente en fonction de la température, les valeurs à 16 et 24 °C et 32 °C sont doublées voire triplées par rapport aux  $\phi$ PSII à 4 et 8 °C.



**Figure 50 : Evolution des rendements quantiques maximum (Fv/Fm) et effectifs ( $\phi$ PSII) chez *Amphora coffeaeformis* à 4, 8, 16, 24 et 32 °C.**

Les valeurs de Fv/Fm et  $\phi$ PSII sont exprimées en unités arbitraires (u.a). La température de culture (témoin) est indiquée par \*.

Chez *A. acutiuscula*, les valeurs maximales enregistrées pour les deux rendements quantiques sont obtenues chez le témoin à 24 °C (Figure 51). Les autres traitements thermiques induisent une diminution de ces rendements quantiques qui sont inférieurs à 0,5 u.a. Les rendements quantiques effectifs quant à eux sont tous diminués de plus de 50 % par rapport au témoin.



**Figure 51 : Evolution des rendements quantiques maximum (Fv/Fm) et effectifs ( $\phi$ PSII) chez *Amphora acutiuscula* à 4, 8, 16, 24 et 32 °C.**

Les valeurs de Fv/Fm et  $\phi$ PSII sont exprimées en unités arbitraires (u.a). La température de culture (témoin) est indiquée par \*.

Les rendements de fluorescence  $F_v/F_m$  et  $\phi_{PSII}$  sont caractéristiques d'une espèce, en ce qui concerne le rendement quantique maximal à l'obscurité, chez les plantes il est de l'ordre de 0,8 u.a. et chez les algues et plus particulièrement les diatomées il est de l'ordre de 0,6-0,7 u.a selon l'espèce considérée (Guihéneuf et al., 2010). Un  $F_v/F_m$  inférieur de 0,1 par rapport au témoin indique la présence d'une perturbation de l'état du photosystème II. Toutefois les rendements de fluorescence de la Chl *a* présentés ici sont une étude préliminaire qui doit être confirmée par des réplicats et des tests statistiques afin de conclure que les traitements thermiques appliqués ont effectivement eu un impact physiologique sur l'appareil photosynthétique et représentent effectivement des stress. Néanmoins, à partir de ces premiers résultats nous pouvons penser que les conditions 4 et 32 °C représentent des stress pour *P. tricornutum*, alors que 8 °C qui soulève une ambiguïté car l'appareil photosynthétique semble peu affecté. Les traitements thermiques à 4 et 8 °C représentent un stress pour *A. coffeaeformis* et toutes les conditions hormis le témoin 24 °C sont stressantes pour *A. acutiuscula*.

### III.2. Expression des MLE en conditions de stress thermiques

Afin de vérifier si l'expression des MLE et des autres éléments transposables est induite par les conditions de stress thermiques, les ARN messagers ont été extraits à partir des culots de cultures des trois diatomées. Des RT-PCR ont été réalisées avec des amorces spécifiques des trois espèces *P. tricornutum*, *A. coffeaeformis* et *A. acutiuscula* pour détecter l'expression des MLE, après traitement à la DNase. L'expression des rétrotransposons *Blackbeard* et *Surcouf*, car ils sont inductibles dans certaines conditions de stress (traitements chimiques et culture en milieu apauvri en nitrate) chez *P. tricornutum* (Maumus et al., 2009), a également été étudiée. Cette analyse a été complétée par l'étude de gènes marqueurs de l'expression génique constitutive d'une part (Actine et sous-unité ribosomale 18S) pour vérifier la qualité des extraits et de gènes marqueurs de stress d'autre part (Superoxyde dismutase à manganèse et small Heat shock Protéins). Ces deux derniers gènes sont largement utilisés pour mettre en évidence la mise en place d'une réponse au stress chez les végétaux (Bowler et al., 1992; Scandalios et al., 1993; Waters et Rioflorido, 2007 Siddique et al., 2008).

Chaque fragment amplifié par PCR à la taille attendue a été cloné et séquencé. Les séquences ont été identifiées par Blastx et/ou Blastn sur les banques de données NCBI et/ou Repbase.

Dans la section suivante sont présentés les résultats des amplifications obtenues et des séquences d'ADNc identifiées pour les deux types d'éléments transposables recherchés (les résultats sont résumés dans le Tableau 15).

### III.2.1 Détection de l'expression des MLE et des gènes « contrôle » en conditions de stress thermiques chez *P. tricornutum*

#### **Expression des MLE**

Chez *P. tricornutum*, des amplifications de 170 pb, la taille attendue, ont été obtenues avec les amorces spécifiques *Phaeo 4* et *Phaeo 7* pour chaque traitement ainsi que pour la condition témoin 16 °C (Figure 52A). Les fragments obtenus ont été identifiés comme transposons *MLE*, pour chaque traitement de 5 h à l'exception de celui à 24 °C. Ces résultats indiquent donc que les *MLE* sont exprimés chez *P. tricornutum*.

#### **Expression des ET de « contrôle »**

Afin de vérifier que la composante transposable du génome des diatomées était exprimée, nous avons recherché l'expression des ET de contrôle tels que *Blackbeard* et *Surcouf*. Des amplifications d'ADNc du rétrotransposon *Blackbeard* ont été obtenues à la taille attendue de 270 pb pour chaque traitement thermique (Figure 52B). Les signaux sont plus intenses pour les traitements 8 °C et 32 °C. Pour *Surcouf*, de faibles amplifications sont présentes pour chaque traitement (Figure 52C).

Les séquences obtenues correspondent au rétrotransposon *Blackbeard* pour toutes les températures testées excepté à 4 °C et à *Surcouf* en condition témoin 16 °C. Ces résultats nous permettent de confirmer l'expression de ces éléments transposables dans le génome de la diatomée modèle dans nos conditions expérimentales.

#### **Expression des gènes de « contrôle »**

Des amplifications ont été obtenues pour toutes les conditions thermiques pour les gènes marqueurs de l'expression génique (actine et sous unité ribosomale 18S) ainsi que pour les gènes de réponse aux stress (MnSOD et sHSP). Pour l'actine, une amplification plus faible à 32 °C qu'aux autres températures est observée ce qui laisse supposer que ce gène n'a pas une expression constitutive (Figure 52D). Les signaux d'amplification relatifs au gène ribosomal 18S apparaissent à toutes les températures testées avec la même intensité (Figure 52E) comme attendu pour un gène constitutif. Le profil d'expression du gène de la MnSOD présente deux amplifications pour chaque traitement dont la taille est inférieure aux fragments obtenus avec l'ADN génomique (Figure 52F). La différence de taille entre les fragments d'ADNg et d'ADNc résulte de la présence d'un intron de 66 pb. Dans les PCR avec ADNc,

les deux fragments obtenus correspondent d'une part au fragment de 400 pb de la séquence répertoriée comme MnSOD de *P. tricornutum* dans les bases de données et d'autre part à une séquence codante de 436 pb. Outre ces 36 pb de différence, les deux séquences sont à 100 % identiques. La séquence de 36 pb ne correspond pas à un intron puisqu'elle est présente uniquement dans les ADNc. Ceci confirme que les fragments amplifiés sont effectivement des ADNc et non de l'ADN génomique contaminant. Les deux fragments obtenus avec l'ADNg correspondent sans doute à deux copies différentes du gène de la MnSOD. Ce qui est original dans la souche de *P. tricornutum* présente au laboratoire. En effet, dans le génome séquencé de *P. tricornutum* souche CCMP632/CCAP1055.1 une seule copie du gène la MnSOD est présent.

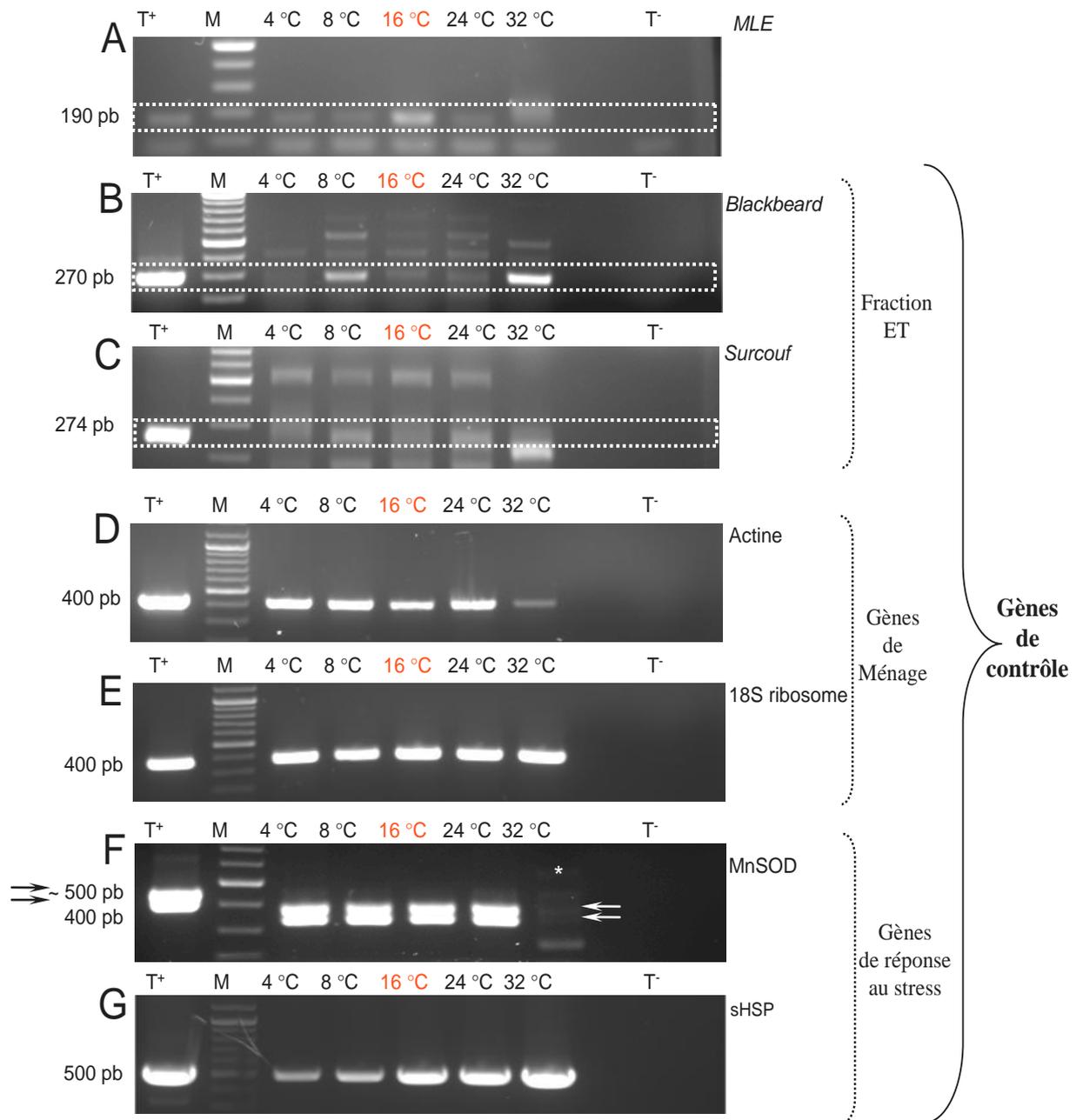
La méthode utilisée ici n'a pas permis de quantifier l'expression des gènes, toutefois le profil d'expression des gènes des sHSP semble s'intensifier avec l'élévation de température (Figure 52E) alors que le profil d'expression du gène 18S constant nous a permis de déterminer une quantité constante d'ARN messagers utilisés.

*A contrario*, les signaux d'intensité du gène de la MnSOD semblent constant ce qui paraît troublant mais peut s'expliquer par le manque de précision de l'expérience qui reste non quantitative.

### III.2.2 Détection de l'expression des MLE et des gènes « contrôle » en conditions de stress thermiques chez *A. coffeaeformis*

#### **Expression des MLE**

Les amorces spécifiques *Ampho4* et *Ampho7* ont permis d'amplifier des séquences de 170 pb correspondant à des transposons *MLE* pour deux des cinq conditions testées, correspondant aux températures 4 et 8 °C (Figure 53A). Aucune amplification n'a été détectée chez le témoin à 16 °C, ni après 5 h à 24 °C et les fragments amplifiés à 32 °C n'ont pas révélé de similitudes avec des *MLE*. Chez *A. coffeaeformis*, l'identification de *MLE* aux températures basses permet d'envisager que l'induction de la transcription des *MLE* soit due au stress.



**Figure 52 : Amplification des *MLE* et des gènes de contrôle par PCR sur ADNc et ADNg chez *Phaeodactylum tricornutum* cultivée à 16 °C et après 5 h de stress à 4, 8, 24 et 32 °C.**

Les fragments d'intérêt des éléments transposables *MLE*, *Blackbeard* et *Surcouf* sont encadrés en pointillé en A, B et C. Ces fragments de 170 pb (A) et d'environ 270 pb (B et C) ont été séquencés et identifiés pour chaque traitement thermique et pour chacun des gènes testés. Les gènes de ménage sont représentés en D et E et les gènes de réponse au stress en F et G.

T<sup>+</sup> : témoin positif (PCR réalisée avec de l'ADNg), M : marqueur de taille 100 pb. 4 °C, 8 °C, 16 °C, 24 °C, 32 °C : températures appliquées pendant 5 h. T<sup>-</sup> : témoin négatif (PCR réalisée sans ADN).

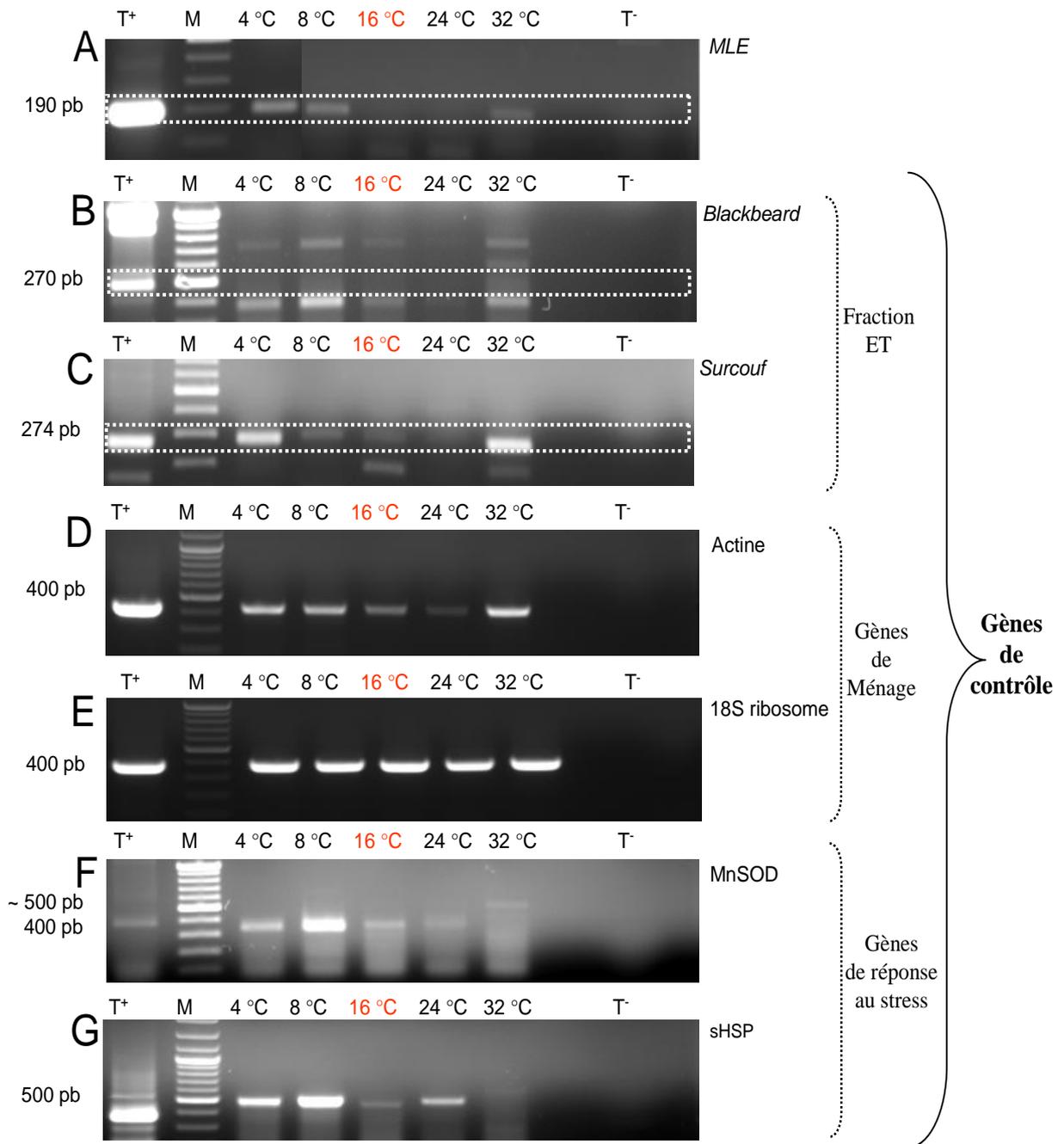
La température témoin est indiquée en rouge.

### **Expression des ET de « contrôle »**

L'amplification des rétrotransposons *Blackbeard* et *Surcouf* est détectée pour chaque condition thermique (Figures 53B et 53C). Les amplifications semblent toutefois plus faibles pour le témoin à 16 °C et après 5 h de traitement à 24 °C. Ce résultat est cohérent avec les courbes de croissance obtenues à 16 °C et 24 °C qui n'apparaissent pas significativement différentes (Figure 46). Aucune expression de *Blackbeard* n'a pu être mise en évidence chez *A. coffeaeformis* après séquençage des fragments, les séquences obtenues avec ADNc ne présentent pas de similitudes avec la séquence de *Blackbeard* présente dans les bases de données. Les fragments amplifiés pour *Surcouf* ne sont identifiés positivement qu'avec l'ADNg et la condition 32 °C. Ces résultats permettent de mettre en évidence la présence de *Blackbeard* et *Surcouf* ainsi que l'expression de *Surcouf* chez la diatomée *A. coffeaeformis* mais ne permettent pas de relier l'induction de cette expression avec les stress appliqués.

### **Expression des gènes « contrôle »**

En ce qui concerne le gène de l'actine et le gène ribosomal 18S, des amplifications sont observées pour chacune des conditions thermiques testées (Figures 53D et 53E) indiquant la bonne qualité des ADNc synthétisés. Ces gènes pourront être utilisés pour normaliser l'expression des ET et des gènes marqueurs de stress dans une étude quantitative ultérieure. Des profils d'expression différents ont été obtenus pour les gènes marqueurs du stress : MnSOD et sHSP. En effet les amplifications obtenues pour le gène de la MnSOD sont intenses pour les conditions thermiques 4 et 8 °C mais deviennent plus ténues aux autres températures (Figure 53F). A 32 °C, aucune amplification nette à la taille attendue n'est visible. Les amplifications obtenues correspondent, après séquençage, à deux copies du gène de la MnSOD avec ou sans séquence supplémentaire comme pour *P. tricornutum* bien que ces fragments d'environ 400 pb n'apparaissent pas comme des signaux d'amplifications distincts dans la Figure 53F. Le profil d'expression des sHSP diffère en fonction des températures de traitement. Les amplifications les plus intenses sont obtenues pour 4, 8 et 24 °C (Figure 53G). Les signaux sont faibles pour le témoin (16 °C) et à 32 °C.



**Figure 53: Amplification des *MLE* et des gènes de contrôle par PCR sur ADNc et ADNg chez *Amphora coffeaeformis* cultivée à 16 °C et après 5 h de stress à 4, 8, 24 et 32 °C.**

Les fragments d'intérêt des éléments transposables *MLE*, *Blackbeard* et *Surcouf* sont encadrés en pointillé en A, B et C. Ces fragments de 170 pb (A) et d'environ 270 pb (B et C) ont été séquencés et identifiés pour chaque traitement thermique et pour chacun des gènes testés. Les gènes de ménage sont représentés en D et E et les gènes de réponse au stress en F et G.

T<sup>+</sup> : témoin positif (PCR réalisée avec de l'ADNg), M : marqueur de taille 100 pb. 4 °C, 8 °C, 16 °C, 24 °C, 32 °C : températures appliquées pendant 5 h. T<sup>-</sup> : témoin négatif (PCR réalisée sans ADN).

La température témoin est indiquée en rouge.

### III.2.3 Détection de l'expression des MLE et des gènes « contrôle » en conditions de stress thermiques chez *A.acutiuscula*

#### **Expression des MLE**

Les fragments amplifiés par PCR à partir de l'ADNg et des ADNc de *A. acutiuscula* (Figure 54A) avec les amorces spécifiques *AacutFc* et *AacutRe* ont été identifiés comme des transposons *MLE* pour chaque traitement incluant la condition témoin 24 °C.

Ces résultats mettent en évidence l'expression des *MLE* chez la diatomée *A. acutiuscula* soumise ou non à un stress thermique de 5 h. Bien que la technique utilisée ici soit qualitative et ne permette pas de mettre en évidence une expression différentielle en fonction du traitement appliqué, nous remarquons que l'intensité du signal à 16 et 24 °C est nettement plus marquée qu'aux autres températures.

#### **Expression des ET de « contrôle »**

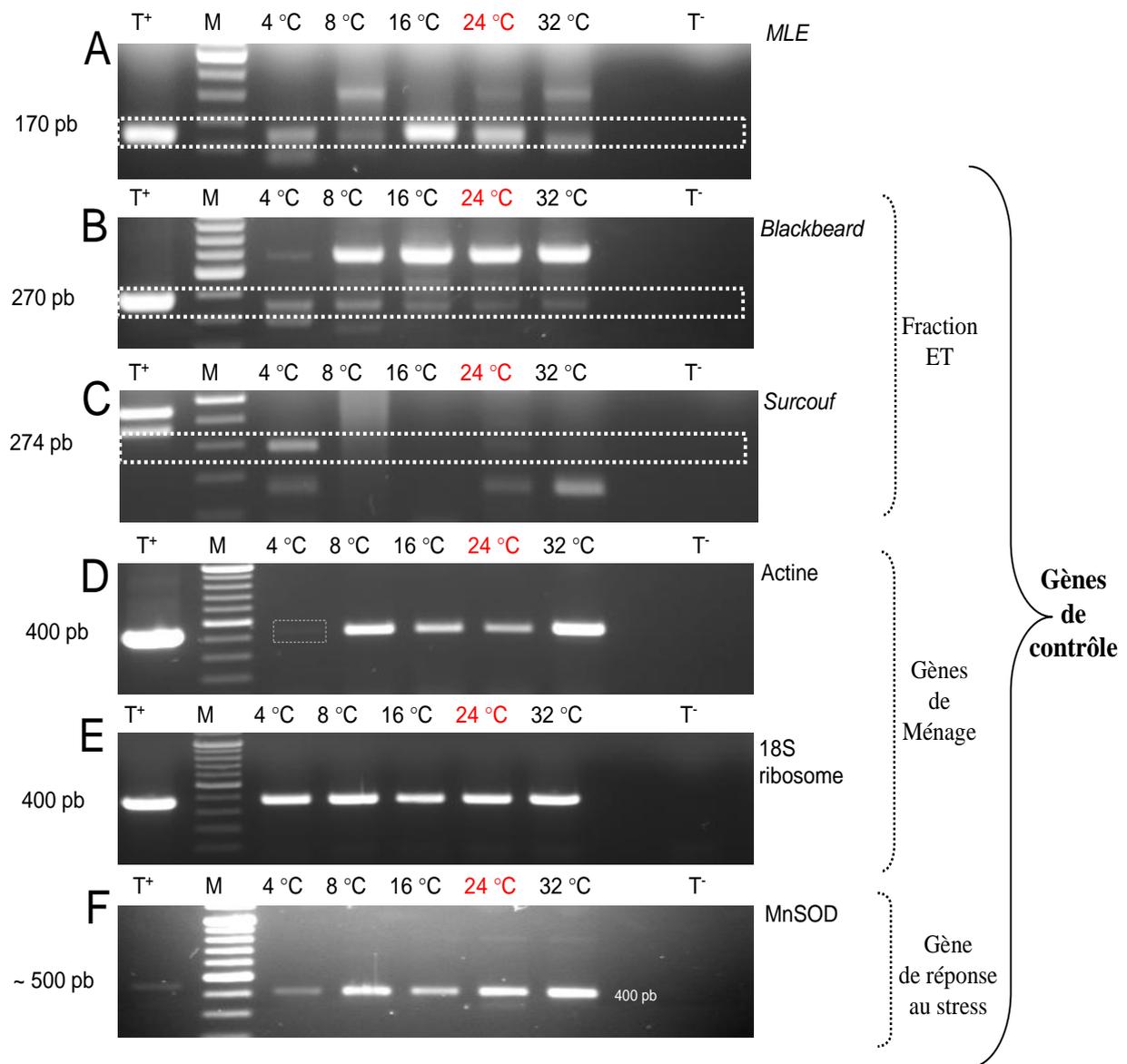
L'expression des rétrotransposons *Blackbeard* et *Surcouf* est présentée Figures 54B et 53C. Pour le rétrotransposon *Blackbeard*, une amplification de l'ADNc a été détectée pour chaque traitement (Figure 54B). Le signal à 4 °C apparaît moins intense qu'aux autres températures comme cela a été observé chez les deux autres diatomées *P. tricornutum* et *A. coffeaeformis*. Les séquences obtenues à 4, 16 et 32 °C ont été identifiées comme *Blackbeard* par comparaison avec les banques de données, ces résultats montrent que cet élément est présent et exprimé chez *A. acutiuscula* à 4, 16 et 32 °C. Concernant le rétrotransposon *Surcouf*, des amplifications à la taille attendue (environ 270 pb) ont été observées à 4 et à 24 °C (Figure 54C), mais aucune des séquences obtenues n'a été identifiée de manière positive. De plus, aucune amplification n'a été détectée dans le témoin positif ADNg, ceci permet d'envisager les deux hypothèses suivantes : i) cet élément est absent du génome d'*A. acutiuscula*, ii) les amorces utilisées qui ont été définies chez *P. tricornutum* et ne permettent pas l'amplification de *Surcouf* dans le génome d'*A. acutiuscula*.

#### **Expression des gènes de « contrôle »**

L'expression des gènes de l'actine et de la petite sous unité ribosomale 18S a été détectée pour toutes les conditions thermiques indiquant la bonne qualité des ARN et des ADNc, bien qu'il y ait une faible amplification observée à 4 °C (Figures 54D, 54E).

Les résultats de l'expression du gène de la MnSOD montrent un signal présent pour chaque température, toutefois l'amplification obtenue à 4 °C est moins intense que celles obtenues pour les autres traitements incluant le témoin à 24 °C (Figure 54F). Les amplicons obtenus à partir des ADNc sont de taille inférieure à celle des fragments amplifiés à partir de l'ADNg. Le gène de la MnSOD contient un intron de 66 pb qui est épissé à la fin de la

transcription. Les sHSP n'ont pu être étudiées chez cette espèce, aucune amplification n'ayant été obtenue avec les amorces définies à partir de la séquence de la sHSP de *P. tricornutum*.



**Figure 54 : Amplification des *MLE* et des gènes de contrôle par PCR sur ADNc et ADNg chez *Amphora acutiuscula* cultivée à 24 °C et après 5 h de stress à 4, 8, 16 et 32 °C.**

Les fragments d'intérêt des éléments transposables *MLE*, *Blackbeard* et *Surcouf* sont encadrés en pointillé en A, B et C. Ces fragments de 170 pb (A) et d'environ 270 pb (B et C) ont été séquencés et identifiés pour chaque traitement thermique et pour chacun des gènes testés. Les gènes de ménage sont représentés en D et E et les gènes de réponse au stress en F.

T<sup>+</sup> : témoin positif (PCR réalisée avec de l'ADNg), M : marqueur de taille 100 pb. 4 °C, 8 °C, 16 °C, 24 °C, 32 °C : températures appliquées pendant 5 h. T<sup>-</sup> : témoin négatif (PCR réalisée sans ADN).

La température témoin est indiquée en rouge.

**Tableau 15 : Récapitulatif des conditions thermiques auxquelles des transcrits correspondant aux *MLE* ou aux rétrotransposons *Blackbeard* et *Surcouf* ont été obtenus chez *Phaeodactylum tricornutum*, *Amphora coffeaeformis* et *Amphora acutiuscula*.**

Les conditions thermiques témoins sont indiquées en gras et la présence (OUI) ou l'absence (x) des ET recherchés dans les génomes des trois diatomées est indiquée.

Gènes	Espèce	ADNc					ADNg
		4 °C	8 °C	16 °C	24 °C	32 °C	
<i>MLE</i>	<i>P. tricornutum</i>	OUI	OUI	OUI	OUI	x	OUI
	<i>A. coffeaeformis</i>	OUI	OUI	x	x	x	OUI
	<i>A. acutiuscula</i>	OUI	OUI	OUI	OUI	OUI	OUI
<i>Blackbeard</i>	<i>P. tricornutum</i>	x	OUI	OUI	OUI	OUI	OUI
	<i>A. coffeaeformis</i>	x	x	x	x	x	OUI
	<i>A. acutiuscula</i>	OUI	x	OUI	x	OUI	OUI
<i>Surcouf</i>	<i>P. tricornutum</i>	x	x	OUI	x	x	OUI
	<i>A. coffeaeformis</i>	x	x	x	x	OUI	OUI
	<i>A. acutiuscula</i>	x	x	x	x	x	x

### III.3 Identification des gènes *MLE*, *Blackbeard* et *Surcouf* exprimés dans les conditions expérimentales

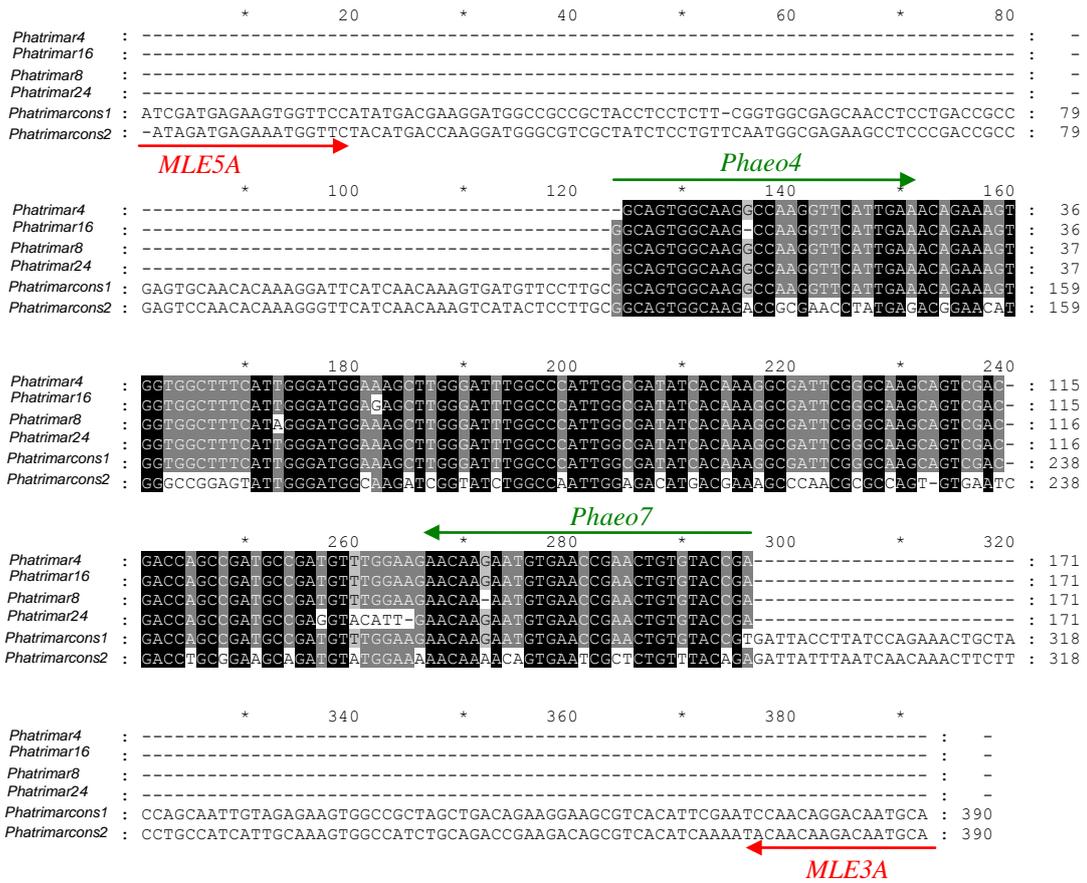
Les séquences *MLE* qui ont été obtenues à partir d'ADNc et identifiées pour chacune des espèces (Tableau 15) ont été alignées avec les séquences obtenues avec l'ADNg grâce au programme CLUSTALW. Les séquences d'ADNg utilisées pour l'alignement des *MLE* correspondent aux séquences consensus obtenues dans la partie I de ce chapitre. Ces comparaisons de séquences permettent d'indiquer si ce sont les *MLE* mis en évidence au début dans la partie I qui sont exprimés ou bien si ce sont des éléments *MLE* différents.

L'alignement entre les séquences de *MLE* obtenues avec ADNc chez *P. tricornutum* et les séquences de *MLE* obtenues avec l'ADNg est présenté en Figure 55. Il montre clairement une forte similitude de 80 à 100 % d'identité entre les séquences. Un taux d'identité plus fort est observé entre les séquences transcrites (*Phatrimar4* obtenue à 4 °C, *Phatrimar8* obtenue à 8 °C et *Phatrimar16* obtenue à 16 °C) et la séquence consensus *Phatrimarcons1* plutôt qu'avec *Phatrimarcons2*. Ceci n'est pas surprenant puisque les amorces *Phaeo4* et *Phaeo7* ont été définies à partir de la séquence consensus *Phatrimarcons1*. Les séquences *MLE* transcrites partagent plus de 95 % d'identité entre elles et entre 95 et 99

% d'identité avec la séquence *Phatrimarcons1*, ce qui indique qu'elles appartiennent à la famille d'élément *MLE* que *Phatrimarcons1*.

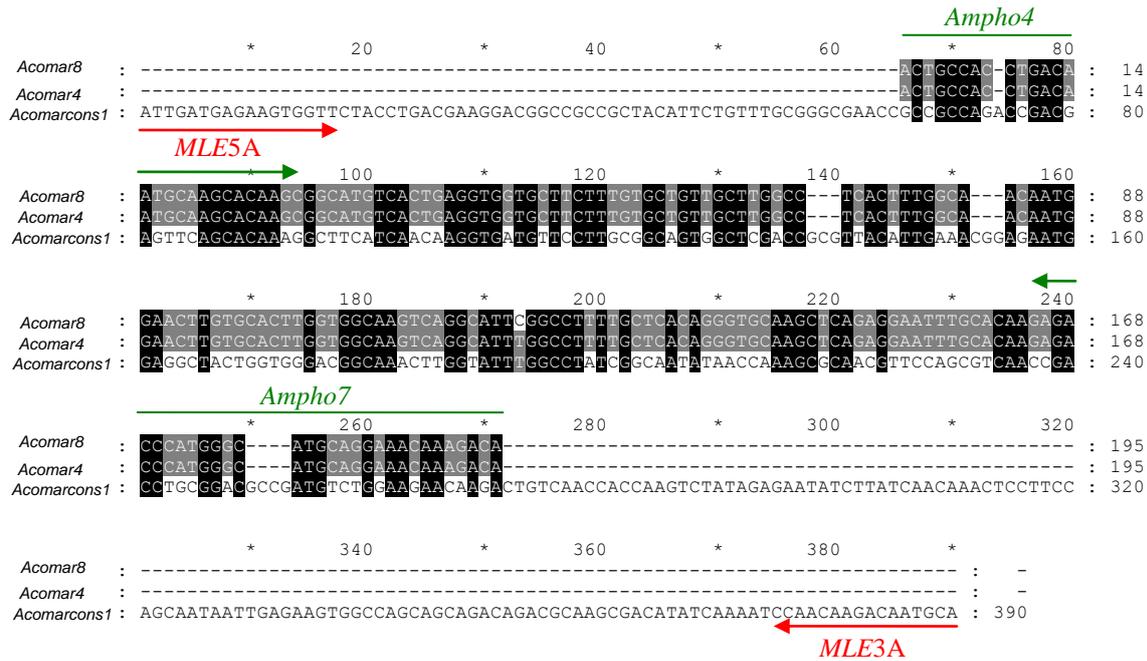
Les séquences *MLE* obtenues avec l'ADNc, nommées *Acomar4* et *Acomar8* en fonction des conditions thermiques 4 et 8 °C, partagent 99 % d'identité entre elles, ce qui indique que ce sont des transcrits d'un seul et même élément *MLE*. La Figure 56 présente l'alignement des séquences obtenues avec ADNc et avec la séquence ADNg *Acomarcons1* uniquement. Cette dernière est la séquence la plus similaire aux séquences obtenues en ADNc puisqu'elle présente 47 % d'identité avec les séquences *Acomar4* et *Acomar8*, alors que les similitudes avec les autres séquences consensus sont de seulement 15 % d'identité avec *Acomarcons2*, 35 % avec *Acomarcons3*, et 20 % avec *Acomarcons4*. Ces résultats impliquent que l'élément exprimé ici est différent de ceux mis en évidence dans la partie I de ce chapitre. Ce résultat semble surprenant car les amorces *Ampho4* et *Ampho7* qui ont permis d'amplifier les transcrits *MLE* d'*A. coffeaeformis* ont été définies sur la séquence d'*Acomarcons1*. Toutefois, ce résultat semble cohérent avec la variabilité des séquences *MLE* chez *A. coffeaeformis* préalablement observée dans la partie I de ce chapitre.

Les séquences ADNc obtenues chez *A. acutiuscula* sont identiques de 92 à 100 % entre elles. La séquence *Aacutimar24* est un peu divergente et ne partage que 92 % d'identité avec les autres transcrits de *MLE* obtenues quand les séquences *Aacutimar 4, 8, 16 et 32* sont identiques à 99 % et 100 % comme présenté dans la Figure 57. Ce résultat suggère que les séquences obtenues en ADNc proviennent de la transcription d'un seul et même élément *MLE* dans chaque condition testée. L'alignement nucléotidique avec l'unique séquence consensus obtenue avec l'ADNg d'*A. acutiuscula* montre une identité de 85 à 87 % avec les séquences transcrites *Aacutimar4, 8, 16, 24 et 32*. Ceci indique que de faibles variations sont observables entre la séquence du gène *Aacutimarcons1* et les séquences exprimées chez *A. acutiuscula* (Figure 57); néanmoins avec une identité nucléique d'au moins 85%, on peut supposer que l'élément exprimé chez *A. acutiuscula* appartient à la même famille de *MLE* que *Aacutimarcons1*.



**Figure 55 : Alignement des séquences nucléiques de *MLE* obtenues avec les ADNc et en ADNg chez *Phaeodactylum tricornutum*.**

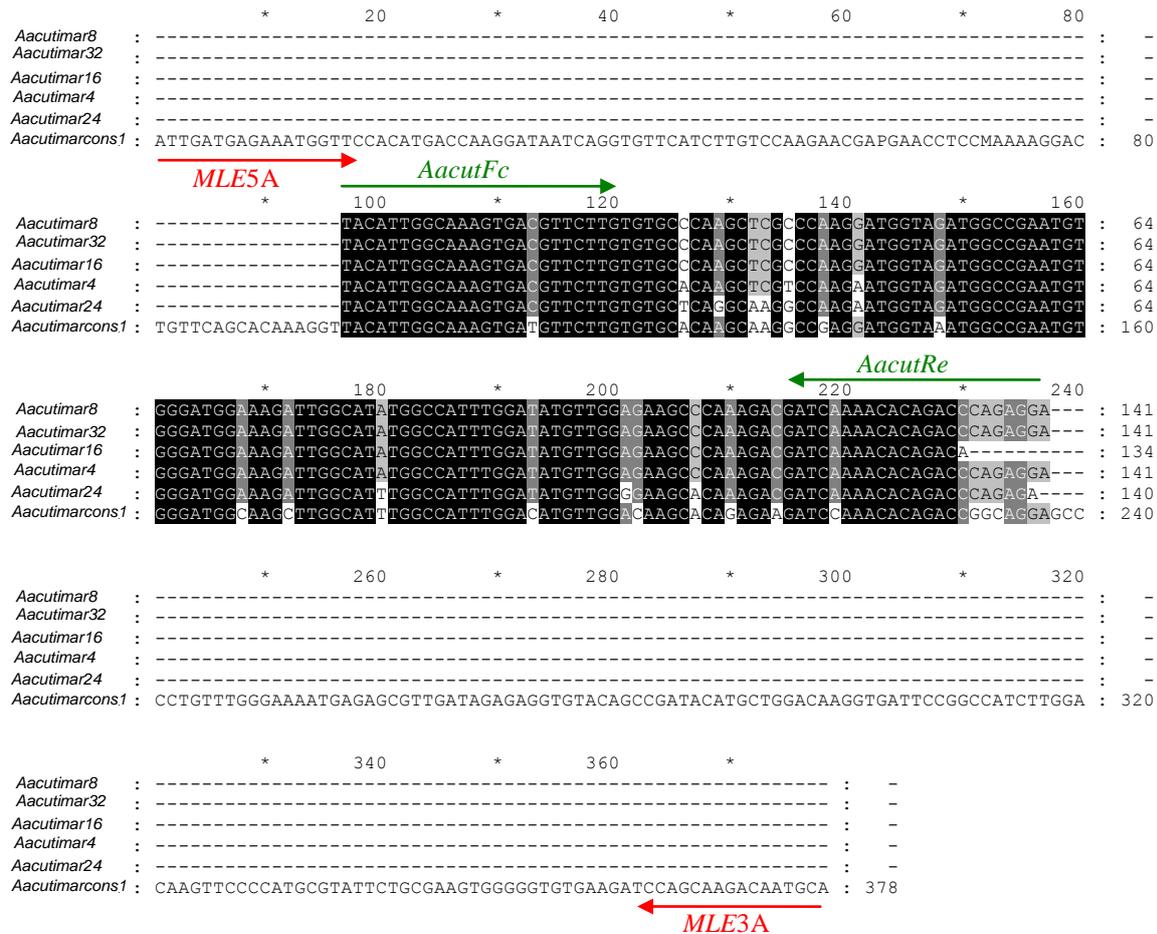
Les séquences *MLE* obtenues avec l'ADNg sont *Phatrimarcons1* et 2 et les séquences rétrotranscrites en ADNc sont *Phatrimar4*, 8, 16, 24 pour les conditions 4, 8, 16, 24 °C. En rouge sont indiquées les amorces *MLE5A* et *MLE3A* ayant permis l'amplification avec l'ADNg et en vert les amorces *Phaeo4* et *Phaeo7* qui ont permis d'amplifier les séquences des ADNc. Une identité de 80 % entre les séquences est indiquée en grisé et une identité de 100 % est indiquée en noir.



**Figure 56 : Alignement des séquences nucléiques de MLE obtenues avec les ADNc et en ADN g chez *Amphora coffeaeformis***

La séquence MLE obtenue avec l'ADNg est Acomarcons1 et les séquences rétrotranscrites en ADNc sont Acomar4 et 8 pour les conditions 4 et 8 °C respectivement. En rouge sont indiquées les amorces MLE5A et MLE3A qui ont permis l'amplification à partir de l'ADNg et en vert les amorces Amphi4 et Amphi7 qui ont permis d'amplifier les séquences d'ADNc. Une identité de 80 % entre les séquences est indiquée en grisé et une identité de 100 % est indiquée en noir.

En ce qui concerne les ET *Blackbeard* et *Surcouf*, nous observons le même type de résultat chez les trois diatomées. Ainsi à partir des ADNc de *P. tricornutum*, les séquences de *Blackbeard* en ADNc nommées *Blackbeard8*, *Blackbeard16*, *Blackbeard24* et *Blackbeard32* pour les conditions thermiques 8, 16, 24, et 32 °C respectivement partagent de 92 à 97 % d'identité entre elles. La comparaison des séquences obtenues avec l'ADNc et avec l'ADNg de *P. tricornutum* montre un pourcentage d'identité de 93 à 97 %. Il est ainsi probable que l'élément exprimé ici soit celui dont la séquence a été amplifiée avec l'ADNg. La séquence de *Surcouf* transcrite et nommée *Surcouf16* pour la condition témoin 16 °C partage 98 % d'identité avec la séquence ADNg chez *P. tricornutum*. Ceci suggère que l'élément transcrit ici est probablement le même que celui dont la séquence a été amplifiée à partir de l'ADNg.



**Figure 57 : Alignement des séquences nucléiques de MLE obtenues avec les ADNc et en ADNg chez *Amphora acutiuscula***

La séquence consensus MLE obtenue avec l'ADNg est *Aacutimarcons1* et les séquences rétrotranscrites en ADNc sont *Aacutimar4*, *8*, *16*, *24* et *32* pour les conditions thermiques 4, 8, 16, 24 et 32 °C respectivement. En rouge sont indiquées les amorces *MLE5A* et *MLE3A* qui ont permis l'amplification de l'ADNg et en vert les amorces *Ampho4* et *Ampho7* qui ont permis d'amplifier les séquences d'ADNc. Une identité de 80 % entre les séquences est indiquée en grisé et une identité de 100 % est indiquée en noir.

En ce qui concerne *A. coffeaeformis*, la séquence de *Surcouf* transcrite et nommée *Surcouf32* pour le traitement à 32 °C partage moins de 50 % d'identité avec la séquence ADNg. Ce résultat indique que l'élément exprimé est différent de la séquence qui a été amplifiée avec l'ADNg. Nous pouvons en déduire qu'il y a plusieurs éléments apparentés à *Surcouf* dans le génome de la diatomée *A. coffeaeformis*.

Pour *A. acutiuscula*, les séquences de *Blackbeard* transcrites obtenues nommées *Blackbeard4*, *16* et *32* respectivement pour les conditions 4, 16 et 32 °C partagent de 91 à 97 % d'identité avec la séquence ADNg. Par conséquent, le transcrit obtenu correspond à l'expression de l'élément *Blackbeard* dont la séquence a été amplifiée avec l'ADNg.

### III.4 Discussion

Dans cette partie nous avons mis en évidence l'expression des transposons *MLE* chez trois espèces de diatomées. L'expression des *MLE* traduit leur capacité potentielle à produire la transposase et ceci permet d'envisager qu'ils puissent être actifs et capables de se déplacer dans le génome des diatomées. La stratégie employée est basée sur la recherche de l'expression des *MLE* induite, ici, par des stress thermiques.

Les diatomées sont naturellement soumises à des variations des conditions du milieu qui peuvent être stressantes telle que la température qui s'avère être un paramètre fluctuant de l'environnement marin agissant sur le métabolisme des algues (Lobban et al., 1985; Rai et Gaur, 2001). De nombreux travaux ont étudié l'impact de la température sur la croissance des organismes photosynthétiques (pour revue : Davison, 1991) et la photosynthèse des microalgues (Suzuki et Takahashi, 1995; Rajadurai et al., 2005; Morgan-Kiss et al., 2006), facteur qui selon les espèces peut être limitant ou bien stimulant. La température peut induire des modifications de la composition membranaire, du contenu cellulaire en acides gras, de la conformation des protéines et des activités enzymatiques chez les algues (pour revue : Peschek et Zoder, 2001).

Les conditions environnementales et les facteurs de stress ont également un impact génomique (pour revue : Hoffman et Parsons, 1994) et peuvent avoir une influence sur l'activité des ET. En effet, de nombreuses études ont révélé une activation des ET suite à des stress tels que des variations des conditions environnementales (pour revue Capy et al., 2000; Schmidt et Anderson, 2006) et des stress thermiques que ce soit chez les animaux, les plantes et les champignons (Zuker et al., 1984; Ikeda et al., 2001; Hashida et al., 2003; Hey et al., 2008; Ogasawara et al., 2009).

#### III.4.1 Conditions de stress thermiques et thermo-tolérance des diatomées

Au cours de ce travail, nous avons appliqué des stress thermiques de courte durée et non létaux et nous avons estimé leur impact physiologique et moléculaire chez trois diatomées : *P. tricornutum* largement utilisée pour les études physiologiques et moléculaires (Mann et Myers, 1968; Ball, 2005; Sarthou et al., 2005; Vardi et al., 2006; Scala et Bowler, 2001; Allen et al., 2008; Nymark et al., 2009) et deux espèces du genre *Amphora*, *A. coffeaeformis* et *A. acutiuscula*. Ces deux dernières ont des origines géographiques différentes et sont adaptées à des environnements différents. La souche d'*A. coffeaeformis* est originaire de la côte atlantique française tempérée et celle d'*A. acutiuscula* provient du Sud-est vietnamien à climat tropical (Nguyen-Deroche et al., 2009).

Les résultats du suivi de croissance ont révélé des comportements de tolérance différents chez ces trois espèces par rapport au stress thermique. Les diatomées, en général, ont des comportements très différents en fonction de la température. Elles sont classées comme diatomées polaires (< 10 °C), tempérées (10-25 °C) ou tropicales (>20 °C) en fonction de la température moyenne de la zone où elles ont été collectées (Barsanti et Gualtieri, 2005; Lobban et al., 1985). Bien que la plupart des diatomées aient un comportement tolérant par rapport à la température et supportent des variations de  $\pm 15$  °C, les espèces peuvent être également classées en fonction de la gamme de variations thermiques supportée. Ainsi, certaines espèces sont dites « sensibles » ou thermo-intolérantes et ne peuvent pas s'acclimater à des variations de plus ou moins 10 °C (Patrick, 1977; Suzuki et Takahashi, 1995; Barsanti et Gualtieri, 2005). Bien que les diatomées soient stressées, elles ont la possibilité de s'acclimater lors d'une exposition prolongée à ces températures et peuvent présenter des capacités de tolérance aux variations de température (William et Morris, 1982; Yongmanitchai et Ward, 1991; Roush et al., 2004). Nos premiers résultats sont concordants avec ceux des études menées sur *P. tricornutum* qui ont montré que la croissance de cette diatomée fluctuait peu dans une gamme de températures comprises entre 10 et 25 °C (William et Morris, 1982; Yongmanitchai et Ward, 1991; Roush et al., 2004). Nous pouvons déduire que cette diatomée est tolérante par rapport aux variations de températures. Les espèces *A. coffeaeformis* et *A. acutiuscula* sont des espèces peu étudiées dans la littérature. Dans nos conditions d'expérimentation il apparaît que *A. coffeaeformis* est une diatomée thermo-tolérante puisque une variation de plus ou moins 8 °C par rapport à sa température de culture n'a pas induit une baisse de croissance. *A. contrario*, *A. acutiuscula* apparaît comme une diatomée sensible au facteur température.

Les stress appliqués aux diatomées marines sont des chocs thermiques, ils correspondent donc au passage des diatomées de leur température de culture à une température stressante pendant 5 h. Pour le choix de la durée du stress, les comportements tolérants des espèces *A. coffeaeformis* et *P. tricornutum* ont été pris en compte. En effet, la durée du choc thermique devait être suffisante pour qu'une réponse au stress soit mise en place chez les diatomées. Les travaux de Fouqueray et al. (2007) et Guihéneuf et al. (2010) portant sur l'effet des stress UV sur la photosynthèse des diatomées ont montré que des irradiations d'une durée de 5 h induisaient la mise en place d'une réponse au stress. Roush et al. (2003) ont mis en évidence chez la diatomée thermo-tolérante *Chaetoceros muelleri* une quantité significativement plus importante de protéines de réponse au stress après au moins 3 h de choc thermique. *A. contrario*, *P. tricornutum* ne présente pas de variation du taux de

protéines de choc thermique après la même durée de traitement. Les auteurs suggèrent qu'un temps de latence chez cette dernière est nécessaire avant la production de protéines de réponse au stress. Parallèlement, les études portant sur l'induction des ET par les chocs thermiques sont réalisées en moyenne avec des chocs variant de 1 et 5 h (Strand et McDonald, 1985; Bouvet et al., 2008; Desalvo et al., 2008; Ogasawara et al., 2009; Kimura et al., 2001a). Compte tenu des données de la littérature, la durée de stress de 5 h qui a été choisie pour les chocs thermiques et appliquée aux diatomées paraît convenable.

#### III.4.2 Estimation du stress subi par analyse des paramètres photosynthétiques

L'analyse des paramètres photosynthétiques, tels que l'intensité photosynthétique, les rendements quantiques maximum ( $F_v/F_m$ ) et effectif ( $\phi_{PSII}$ ) de la fluorescence de la Chl *a* reflètent l'intégrité de l'appareil photosynthétique (pigments, antennes collectrices, transporteurs d'électrons).

L'intensité de la fluorescence est liée à l'état des centres réactionnels du photosystème II (PSII), la fluorescence est minimale ( $F_0$ ) lorsqu'ils sont tous ouverts (oxydés) et maximale ( $F_m$ ) lorsqu'ils sont tous fermés (réduits). Le rapport  $F_v/F_m$  permet de mesurer l'efficacité photochimique maximale du PSII indiquant l'état physiologique du PSII. Quant à  $\phi_{PSII}$ , ce rendement renseigne sur l'efficacité du fonctionnement du PSII à la lumière et sur l'activité photosynthétique à un instant précis (Papageorgiou et Govindjee, 2004; Nymark et al., 2009). Ainsi ces rendements de fluorescence permettent d'estimer la réponse de l'appareil photosynthétique aux variations des conditions environnementales (Smillie et Heterington, 1990; Lichtenthaler, 1998; Nymark et al., 2009). Les travaux de Mouget et Tremblin (2002) et Rech et al. (2003) ont montré que la fluorimétrie modulée était une technique fiable pour mesurer les caractéristiques photosynthétiques des algues et des diatomées. Guihéneuf et al. (2010) l'ont également utilisée pour mettre en évidence que la prymnésiophycee *Pavlova lutheri* et la diatomée *Odontella aurita* étaient affectées par des radiations UV. En conditions non stressantes, les diatomées présentent des valeurs de  $F_v/F_m$  entre 0,6 et 0,7 selon l'espèce ciblée (Nymark et al., 2009; Guihéneuf et al., 2010), nos résultats sont en accord avec ces valeurs dans les conditions témoins chez les trois diatomées. Des études ont montré que les traitements haute et basse températures entraînent une modification du métabolisme et une réduction des mécanismes de photosynthèse en réponse au stress chez *P. tricorutum* (William et Morris, 1982; Kudo et al., 2000). Toutefois, la diminution des paramètres de fluorescence enregistrés aux températures choisies pour représenter un stress indique que le stress, même faible, est ressenti au niveau de l'appareil photosynthétique. Afin de mettre en

évidence de façon précise l'effet stressant des conditions d'expérimentation utilisées ici, des réplicats de ces expérimentations devront confirmer les premiers résultats obtenus. Des triplicats ont été réalisés et sont en cours d'analyse (données non présentées ici).

Les chocs thermiques appliqués peuvent induire également une diminution de l'intensité photosynthétique alors que les paramètres de fluorescence de la Chl *a* varient peu comme chez *A. coffeaeformis*, principalement à 32 °C. La diminution de l'activité photosynthétique est sans doute le résultat d'une baisse des activités enzymatiques induite par une dénaturation des protéines à cette température. Ceci confirme le comportement thermo-tolérant de cette espèce et suggère qu'elle possède des capacités d'acclimatation aux stress thermiques « chauds », comme décrit précédemment par Rajadurai et al. (2005) lors d'une exposition à 38 et 42 °C de courte (45 min) ou de longue (4 jours) durée. La diminution de la température induit une diminution des rendements Fv/Fm et  $\phi$ PSII chez cette espèce. Toutefois les résultats obtenus à 4 °C indiquent que l'activité photosynthétique reste intense bien que le fonctionnement du PSII soit diminué. Nous pouvons émettre l'hypothèse que le choc thermique à 4 °C chez *A. coffeaeformis* entraîne une rigidification des membranes des thylakoïdes ce qui peut diminuer le flux d'électrons. *A. coffeaeformis* apparaît donc comme tolérante aux températures élevées mais sensible aux basses températures.

Les résultats indiquent que *A. acutiuscula* est une diatomée très sensible aux chocs thermiques. Dans les conditions naturelles, cette espèce est adaptée à un environnement tropical et peu fluctuant en température, ses capacités d'acclimatation sont plus faibles que celles des diatomées adaptées à un environnement tempéré telles que *P. tricornutum* et *A. coffeaeformis* où les températures fluctuent selon les saisons (Lobban et al., 1985 ; Lefebvre et al., 2007). Cette espèce est donc considérée comme sensible aux stress. Ici, elle est sensible aux variations de température et Nguyen-Deroche et al. (2009) ont montré que cette diatomée était également sensible aux stress métalliques.

Les variations de l'intensité photosynthétique et des paramètres de fluorescence de la Chl *a* ont permis de montrer que les traitements appliqués ont effectivement induit des stress bien que ces premières interprétations nécessitent d'être corroborées par des tests statistiques ainsi qu'au niveau moléculaire, par l'étude de l'expression des gènes de réponse au stress notamment.

### III.4.3 expression des gènes marqueurs de réponse aux stress et de l'expression génique

Afin de mettre en évidence la réponse au stress thermique au niveau moléculaire, nous avons mis en évidence par RT-PCR l'expression de deux gènes induits précocement. Le stress thermique entraîne de nombreux changements métaboliques (respiration, photosynthèse et photorespiration notamment) induisant une augmentation de la formation d'espèces réactives de l'oxygène (ERO) (Mittler, 2002; Atkin et Joelker, 2003) qui peuvent endommager l'appareil photosynthétique (Bonhert et Sheleva, 1998). Le premier gène étudié est celui codant la superoxyde dismutase à manganèse, enzyme intervenant dans la régulation de la quantité de radicaux libres intracellulaires chez les végétaux et notamment les diatomées marines (Peers et Price, 2004; Ken et al., 2005; Wolfe-Simon et al., 2005). Le deuxième gène étudié est celui codant des protéines de choc thermique de petite taille, les small HSP (sHSP) qui interviennent dans la protection des enzymes et autres protéines (pour revue : Lindquist et Craig, 1988; Baniwal et al., 2004), dans la restauration de la conformation de celles-ci chez les plantes (Waters et al., 2008; Sun et al., 2002; Siddique et al., 2008) et chez les algues (Mannan et al., 1986; Nakamoto et al., 2000; Waters et Rioflorido, 2007). Les sHSP sont induites pendant des traitements thermiques chez les diatomées *P. tricornutum*, *Chaetoceros muelleri* et *Nitzschia alba* (Lai et al., 1988; Rousch et al., 2004) et confèreraient une thermo-tolérance à celles-ci notamment par la protection de l'appareil photosynthétique (Heckathorn et al.1998; Nakamoto et al., 2000).

Nos résultats montrent une expression des gènes de MnSOD quelles que soient les conditions thermiques (témoins et stress) chez les trois diatomées étudiées. Bien que le gène de la MnSOD soit déjà caractérisé chez la diatomée modèle *P. tricornutum*, pour les espèces du genre *Amphora* c'est la première fois que la présence et l'expression de ce gène sont mises en évidence. De même l'expression du gène des sHSP est détectée pour chaque condition chez *P. tricornutum* et chez *A. coffeaeformis* mais pas chez *A. acutiuscula*. Ceci suggère que le gène des sHSP de *A. acutiuscula* doit être assez divergent de ceux présents chez *P. tricornutum* et *A. coffeaeformis* car les amorces utilisées ici ont été définies à partir de la séquence de sHSP de *P. tricornutum*, souche séquencée CCMP632/CCAP1055.1 (Genbank). L'absence d'amplification des sHSP chez *A. acutiuscula* pourrait également indiquer un faible taux de transcription du gène codant les sHSP chez cette diatomée. Le taux de transcription des sHSP étant un facteur important de l'acquisition de la tolérance aux variations de température (Favatiér et al., 1997; Yildiz et Terzi, 2008; Taylor et al., 2009), cette hypothèse serait plausible puisque *A. acutiuscula* est adaptée à une température moyenne qui varie de 24 à 29 °C sur la côte vietnamienne et se révèle très sensible aux faibles variations de

température. Chez *A. coffeaeformis* nos résultats sont également les premiers à mettre en évidence la présence et l'expression des sHSP.

Les gènes codant la MnSOD et les sHSP sont exprimés dans toutes les conditions expérimentales qualifié abusivement de « constitutif » car la littérature montrent que ces gènes de réponse au stress qui sont en fait exprimés à un niveau minimal en conditions standards et surexprimés en conditions de stress (Ken et al., 2005; Siddique et al., 2008). Avec la méthodologie utilisée ici (soit des RT-PCR en phase plateau) nous n'avons pas pu vérifier une modification du taux de transcrits, puisque cette étude est une étape préliminaire à la quantification. Néanmoins de nombreuses études témoignent de la qualité de ces gènes pour l'étude de la réponse aux stress de manière quantitative. En effet, les traitements basses et hautes températures induisent une surproduction de SOD chez les organismes photosynthétiques (Peltzer et al., 2002) comme chez le thé (Vyas et Kumar, 2005) qui procurerait une tolérance aux basses températures aux organismes planctoniques tempérés tels que cyanobactéries, algues et microalgues (Perelman et al., 2006). Cette tolérance serait le résultat de la protection et/ou réparation des dommages que peut induire le stress thermique à l'appareil photosynthétique. Le rôle protecteur des MnSOD sur les PSI et II a été démontré récemment chez *T. pseudonana* (Wolfe-Simon et al., 2006). Les sHSP ont également été identifiées récemment dans le génome de la diatomée modèle *T. pseudonana* (Waters et Rioflorido, 2007). De nombreux exemples relatifs à la surproduction de sHSP en réponse aux stress sont présents dans la littérature (Banzet et al., 1998; Heckathorn et al., 1998; Sun et al., 2002; Ouyang et al., 2009). Chez les diatomées *P. tricornutum* et *Chaetoceros muelleri*, les sHSP ont été détectées en réponse à des chocs thermiques à 30 et 42 °C (Rousch et al., 2004). Ces auteurs ont observé un taux significativement plus élevé de sHSP chez *C. muelleri* par rapport à *P. tricornutum*. Les auteurs suggèrent que la forte capacité d'adaptation aux températures élevées de *C. muelleri* (croissance identique de 20 à 34 °C) serait liée à sa capacité de production de sHSP. *P. tricornutum* montre une thermo-tolérance restreinte (26-27 °C étant la limite de survie) qui serait liée à une production plus faible de sHSP (Rousch et al., 2004).

Les gènes des MnSOD et sHSP ayant été mis en évidence au cours de cette étude pourront être utilisés dans l'optique d'une continuité de ce travail avec des méthodes quantitatives. Cette étude préalable nous informe que ces gènes sont présents chez les espèces du genre *Amphora* et qu'ils peuvent représenter des témoins moléculaires de qualité de la réponse au stress.

Toutefois les gènes constitutifs de l'actine et la sous-unité ribosomale 18S choisis comme référence pour témoigner de la qualité des ARN et des ADNc sont soumis à caution. L'expression de l'actine semble inégale en fonction des conditions expérimentales ce qui ne lui permet pas d'être un gène référence de qualité. Ces résultats peuvent s'expliquer par la nature dégénérée des amorces déterminées par Leblanc et al. (1999) à partir de l'alignement de séquences isolées chez des diatomées marines, des algues brunes et des algues d'eau douce. De plus, de récents travaux montrent également que l'expression de ce gène peut être variable selon les expérimentations (Dheda et al., 2004; Libault et al., 2008; Nicot et al., 2005; Paolacci et al., 2009). Le gène ribosomal 18S quand à lui, apparaît comme un marqueur intéressant pour l'analyse en Q-PCR. Bien que les amorces soient dégénérées (Simon et al., 1992; Caruso et al., 2002), son profil d'expression est typique d'un gène constitutif chez les trois diatomées, cela étant sûrement dû à la grande quantité d'ARN ribosomaux (80 – 85 %) parmi les ARN totaux.

#### III.4.4 Expression des ET chez les diatomées

L'expression des *MLE* et des ET témoins (*Blackbeard* et *Surcouf*) a également été détectée par RT-PCR. Les résultats obtenus dans ce travail représentent la première étape vers l'étude de l'expression des ET et principalement des *MLE* par Q-RT-PCR en fonction des stress thermiques. Ainsi la méthode de RT-PCR utilisée ici a été exclusivement qualitative et elle avait pour objectifs :

- de mettre en évidence l'expression des *MLE* chez les diatomées marines et d'indiquer la présence d'éventuels éléments complets et potentiellement actifs du fait de leur expression dans le génome.
- de mettre en relation l'induction des ET en fonction du stress thermique

Les ET, *MLE* et Rétrotransposons de type *Ty1-copia* ont été mis en évidence chez les trois espèces de diatomées testées au cours de ce travail.

Dans cette troisième partie nous avons pu mettre en évidence l'expression des *MLE* chez les trois espèces de diatomées marines testées. A notre connaissance ces résultats sont les premiers à démontrer l'expression de *MLE* chez les microalgues. Ceci nous permet de supposer que des *MLE* potentiellement complets et actifs sont présents dans le génome des diatomées.

#### **Présence de *MLE* complets dans le génome des diatomées**

Le premier objectif de ce chapitre était de mettre en évidence l'expression des *MLE* chez les trois diatomées marines et de mettre en évidence la présence potentielle de *MLE*

complets ayant la possibilité de transposer. En cela, l'expression des *MLE* chez trois diatomées dans certaines conditions expérimentales est une information encourageante.

Les séquences *MLE* transcrites obtenues ont été alignées avec les séquences consensus *MLE* afin d'identifier le gène transcrit. Comme attendu, de forts taux d'identité ont été obtenus par comparaison des séquences *MLE* ADNc entre elles et avec les séquences consensus obtenues avec l'ADNg. Les *MLE* exprimés semblent correspondre à l'expression d'un seul et même élément chez *P. tricornutum* et *A. acutiuscula*. Ceci nous conduit à penser qu'il existerait au moins un *MLE* à cadre de lecture entier dans le génome de *P. tricornutum* et *A. acutiuscula*. Toutefois, chez *A. coffeaeformis*, les séquences *MLE* transcrites semblent diverger de la séquence consensus *Acomarcons1*. Ceci suggère qu'un élément différent est exprimé chez cette diatomée. Il semble donc probable que des éléments complets soient présents dans le génome des trois diatomées, toutefois, l'hypothèse d'une transcription en *cis* ne peut être exclue. En effet, il est possible que les séquences soient tronquées et que leur position génomique soit adjacente à un gène exprimé ou à un promoteur. Ainsi, dans l'hypothèse où les *MLE* seraient positionnés en aval d'un promoteur ou bien dans la séquence même d'un gène exprimé dans les conditions expérimentales appliquées ici, la transcription de ce même gène ou promoteur entraînerait la présence de transcrits des *MLE* même si leur séquence est tronquée (Slotkin et Martienssen, 2007).

### **Expression des rétrotransposons dans le génome des diatomées**

Au cours de ce travail, les rétrotransposons *Blackbeard* et *Surcouf* ont été utilisés comme marqueur de l'expression de la fraction ET du génome des diatomées. Récemment Maumus et al. (2009) ont montré que ces deux rétrotransposons étaient présents complets et exprimés dans le génome des diatomées modèles *P. tricornutum* et *T. pseudonana*. Ces éléments ont donc été étudiés ici comme ET référence de l'expression des ET chez les diatomées. Il est à noter que la présence et l'expression de ces rétrotransposons chez les espèces du genre *Amphora* constituent des résultats originaux. Ces résultats confirment la présence des rétrotransposons de type *Ty1-copia* dans les génomes des diatomées comme le supposent Bowler et al., 2008 et Maumus et al., 2009. De plus, Maumus et al. (2009) ont relié une surexpression de *Blackbeard* et *Surcouf* avec l'application de stress chimiques et une carence du milieu de culture en nitrates. L'induction de ces ET par des stress en ont fait des marqueurs intéressants pour notre étude bien que les stress thermiques réalisés par les auteurs ne semblent pas avoir d'incidence sur leur expression (Maumus et al., 2009).

### III.4.5 Induction des ET par le stress

Plusieurs études ont montré que les stress thermiques sont des facteurs souvent impliqués dans l'induction des rétrotransposons et transposons (Strand et McDonald, 1985; Paquin et Williamson, 1988; Mhiri et al., 1997; Fornace et Mitchell, 1986; Zuker et al., 1984; Hashida et al., 2003; Kimura et al., 2001b), mais à notre connaissance aucune étude n'a été menée sur l'induction des *MLE* en fonction des chocs thermiques. Il est à noter que l'effet de la température sur les *MLE* avait déjà suscité des questionnements. Deux études sur les populations naturelles ont mis en évidence une relation entre l'activité de l'élément *mariner* et la distribution géographique des drosophiles. Les taux d'excision de l'élément *mariner* sont variables en fonction du climat et de la température au site de reproduction (Chakrani et al., 1993; Capy et al., 2000). Ces auteurs ont émis l'hypothèse que la température serait un facteur environnemental capable d'induire la transposition de *mariner*. La seule étude, à notre connaissance, concernant l'effet de stress thermiques sur l'activation des éléments de la superfamille *Tc1-mariner* a été réalisée par Ogasawara et al. (2009) chez le champignon *Aspergillus oryzae*. Cette étude a mis en évidence que *Crawler* transposait en réponse à des stress thermiques de courte durée (4 h et 6 h) à 37 et 42 °C.

Seuls les profils d'expression des *MLE* obtenus avec *A. coffeaeformis* indiquent une possible activation des *MLE* par des stress basses températures. Ce n'est pas un cas isolé, Hashida et al. (2003), Yamagushi et al. (2004) et Bouvet et al. (2008), ont mis en évidence l'expression spécifique d'ET par des stress températures froides. Les réactivations des ET sont souvent le fait de mécanismes épigénétiques résultant de stress (Madlung et Comai, 2004). Les ET sont concentrés en général dans des régions de l'hétérochromatine hyperméthylée et donc indisponibles à la transcription. De nombreuses études ont montré que les ET pouvaient être activés lors de variations des conditions environnementales par déméthylation de l'hétérochromatine (pour revue : Federoff et al., 1995; Slotkin et Martienssen, 2007). Ainsi, nous pouvons supposer que les *MLE* d'*A. coffeaeformis* pourraient être induits par une modification épigénétique. De même que l'hypométhylation à basse température est à l'origine de la réactivation de *Tam3* chez le mufler (Hashida et al., 2003) et de *MCRIRE* chez la luzerne (Ivashuta et al., 2002), l'application de stress basses températures a pu entraîner une déméthylation des *MLE*, annulant ainsi le mécanisme de répression chez cette diatomée.

Bien que nous ne puissions confirmer l'induction des ET par les stress thermiques, nos résultats ne suggèrent pas non plus un mécanisme de répression des *MLE* et des rétrotransposons dans nos conditions expérimentales.

## La spécificité de d'induction des ET

Les profils d'expression variables obtenus avec les *MLE* ainsi que ceux des rétrotransposons chez les trois diatomées peuvent être également expliqués par le fait que l'induction des ET par le stress reste très spécifique. En effet, un ET peut être activé par un stress particulier dans une espèce et ne pas être induit par le même stress dans une autre espèce (Tapia et al., 2005; Kimura et al., 1999, 2001a; Maumus et al., 2009), ainsi une spécificité de stress au sein d'une espèce, d'une souche, voire d'un même individu peut être observée.

La variabilité des séquences dans un même génome implique que tous les éléments présents ne soient pas induits de la même façon et par les mêmes conditions, comme par exemple les éléments de type *Tnt1* actifs chez le tabac qui sont inductibles par une grande diversité de stress abiotiques (froid, blessure, addition de composés chimiques toxiques) et biotiques (attaques fongiques et bactériennes) (Mhiri et al., 1997; Grandbastien et al., 1997; Melayah et al., 2001). Beguiristain et al. (2001) ont montré que parmi les trois familles de *Tnt1* que l'on pouvait différencier dans le génome de *Nicotiana tabacum*, deux familles montrent des spécificités de réponse au stress en relation avec le promoteur interne à la région UTR 3', les trois familles ne sont pas induites par les mêmes stress. Le promoteur de la famille *Tnt1A* est exprimé spécifiquement par des éliciteurs et le méthyle jasmonate, messenger chimique des réactions de défense des plantes. Le promoteur de la famille *Tnt1C* est exprimé sous l'effet d'acide salicylique, un autre messenger chimique de la réaction de défense des plantes. Quant au troisième promoteur de la famille *Tnt1B*, aucune expression spécifique n'a été détectée (Beguiristain et al., 2001). Ainsi les éléments de type *Tnt1* peuvent être induits spécifiquement par différents stimuli et ces profils différents sont le reflet des variations de séquences des rétrotransposons.

Plusieurs couples d'amorces spécifiques ont été définis à partir de l'ensemble des séquences consensus de *MLE* de chacune des trois espèces de diatomées (trois séquences consensus pour *P. tricornutum*, quatre pour *A. coffeaeformis* et une seule pour *A. acutiuscula*, partie I de ce chapitre) mais leur utilisation ne nous a pas permis d'obtenir des amplifications à la taille attendue en RT-PCR. Les *MLE* transcrits chez *P. tricornutum* et *A. coffeaeformis* ont été obtenus avec des amorces spécifiques définies à partir d'une seule des séquences obtenues préalablement (partie I de ce chapitre) chez chacune des deux espèces de diatomées. Les amorces ayant permis l'amplification de transcrits chez *A. acutiuscula* ont été définies à partir de la seule séquence consensus *MLE* d'*A. acutiuscula* et ont permis d'obtenir des transcrits très similaires à la séquence consensus. Les séquences transcrites correspondent

respectivement à une seule copie de gène de la transposase *MLE* chez *P. tricornutum* et de même chez *A. coffeaeformis*. Ceci suggère que tous les transposons ne sont pas induits de la même manière que ce soit en condition témoin et en conditions de stress. En effet chez *P. tricornutum*, nous avons pu établir trois consensus différents; cependant, seule une séquence semble être exprimée chez cette diatomée. De la même manière chez *A. coffeaeformis*, une seule des séquences amplifiées en ADNg a été identifiée en ADNc. Chez *A. acutiuscula*, les amorces définies sur la séquence consensus ont permis d'obtenir des transcrits *MLE*, ce qui s'explique par le fait que toutes les séquences issues de cette espèce proviennent d'un seul et même élément représentée par un seul consensus. Nous suggérons que seules certaines séquences *MLE* sont exprimées et que la majorité des séquences *MLE* ne s'expriment dans les génomes de *P. tricornutum* et d'*A. coffeaeformis*. Ceci est sans doute le résultat de mutation dans le cadre de lecture ou peut être dû à la présence de promoteurs différents dans les séquences comme c'est le cas des éléments de *Tnt1* (Beguiristain et al., 2001). Toutefois, pour vérifier l'une ou l'autre de ces hypothèses, il est nécessaire d'obtenir les séquences complètes des *MLE* des diatomées.

Dans le cas précis des *MLE* d'*A. coffeaeformis* qui semblent être induits par le stress, pour démontrer une induction spécifique par le stress thermique il faudrait pouvoir examiner la séquence complète et rechercher si une homologie est détectée avec des séquences des gènes de réponse au stress. En effet, les ET qui sont inductibles par des stress thermiques possèdent au sein de leur séquence des régions homologues avec des facteurs de transcription des gènes de réponse au stress thermiques ou avec la région promotrice des gènes codant les HSP notamment (Zuker et al., 1984; Strand et McDonald, 1985; Beguiristain et al., 2001). Ces régions homologues sont situées dans les TIR ou des régions non traduites (UTR) encadrant le gène de la transposase. Citons l'exemple des éléments *Tnt1* chez le tabac qui possèdent au sein même des LTR des séquences homologues avec les promoteurs des HSP (Casacuberta et Grandbastien, 1993; Mhiri et al., 1997) et les éléments *Copia* chez la drosophile dont la région 5' UTR est également homologue des promoteurs des gènes HSP (Strand et McDonald, 1985). Ceci est également fortement suggéré chez les *MLE* de drosophiles (Chakrani et al., 1993, Capy et al., 2000). De plus, il faudrait également inspecter les séquences flanquantes des *MLE* car la proximité des ET et des gènes induits par le stress pourrait expliquer le rôle des ET dans la réponse au stress comme cela a été démontré pour l'activation des MITE *Castaway* chez le riz en conditions de stress basses températures (Yamasgushi et al., 2004).

Ainsi par la suite, la recherche d'éléments complets *MLE* chez les diatomées marines portera principalement sur les séquences obtenues et la diatomée *A. coffeaeformis* semble être un bon modèle pour tenter de quantifier l'induction des *MLE* en conditions de stress thermique.

### **Implication de la méthodologie employée : Durée du stress thermique**

L'expression variable des *MLE* et rétrotransposons ne semble pas corrélée avec l'expression des gènes de réponse au stress, sHSP et MnSOD, chez *P. tricornutum*, *A. coffeaeformis* et *A. acutiuscula*. Il est difficile de statuer sur cette éventuelle relation puisque la méthodologie utilisée n'était pas quantitative ainsi nous n'avons pas pu détecter de variation du profil d'expression des gènes de réponse au stress. La stratégie a été d'appliquer des stress de 5 h et de rechercher l'expression des ET dès la fin des chocs thermiques. Or les ET peuvent être induits de manière indirecte en relation avec les mécanismes de réparation des dommages ou de restauration des conditions physiologiques. L'induction des ET par le stress n'est pas nécessairement liée à l'activation des gènes de réponse au stress. A l'inverse, Kimura et al. (1999) ont montré l'activation du SINE *Bm1* du ver à soie en condition de stress de manière concomitante avec celle du gène codant les HSP70. Cependant, dans la majorité des cas, l'activation des SINE est plus tardive, atteignant un maximum de transcrits 24 h après le choc thermique. Les auteurs suggèrent que *Bm1* pourrait avoir un rôle dans la régulation cellulaire pendant la période de retour à la « normale » de l'organisme (Kimura et al., 2001a). Ainsi il serait intéressant de rechercher si une différence d'expression des *MLE* est détectée plus tardivement dans la réponse au stress, par exemple après l'application des chocs thermiques c'est-à-dire pendant la phase de retour à la normale ou encore pendant une durée de stress plus longue, participant ainsi à la cascade de réactions dans la réponse aux variations environnementales (Capy et al., 2000; Kalendar et al., 2000; Schmidt et Anderson, 2006; Grandbastien, 1998).

Par conséquent des expérimentations de stress thermiques longs (8 jours de stress) ont été initiées (les résultats sont en cours d'analyse). Nous envisageons également de refaire des stress courts (inférieurs à 5 h) avec des échantillonnages d'ARN et ceci de manière quantitative puisque pour deux des trois espèces l'expression des *MLE* a été détectée en condition de stress ainsi qu'en conditions témoin.

Les principaux points mis en évidence dans la deuxième partie de ce travail sont les suivants :

Les conditions de stress ont été déterminées par l'étude des paramètres photosynthétiques des trois diatomées en fonction des chocs thermiques : 4, 8, 16, 24 et 32 °C, 16 °C étant la température de référence pour *A. coffeaeformis* et *P. tricornutum* souche NCC340 et 24 °C pour *A. acutiuscula*. Nous avons déterminé que les trois espèces sélectionnées avaient un comportement différent par rapport aux chocs thermiques :

- *P. tricornutum* est une espèce tempérée et les chocs thermiques 4 et 32 °C représentent un stress intense pour cette espèce.
- *A. coffeaeformis* est une espèce thermo-tolérante pour laquelle, seules les conditions 4 et 8 °C induisent un stress intense.
- *A. acutiuscula* est une espèce sensible et tous les chocs thermiques testés représentent un stress.

L'expression du gène de réponse au stress MnSOD est détectée pour toutes les conditions chez les trois espèces étudiées alors que celle des sHSP n'a pas pu être détectée chez *A. acutiuscula*. La méthode n'étant pas quantitative, nous ne pouvons conclure sur la réponse moléculaire spécifique au stress en fonction des traitements.

Les ET (rétrotransposons et *MLE*) sont exprimés chez les trois espèces de diatomées.

L'expression des *MLE* chez *P. tricornutum*, *A. coffeaeformis* et *A. acutiuscula* a été détectée en conditions de stress et en condition témoin. De la même façon que pour les rétrotransposons, l'expression des *MLE* n'est détectée que dans certaines conditions.

- Chez *P. tricornutum* et *A. acutiuscula*, il y a une expression dans les conditions témoins et certaines conditions de stress. Ce résultat suggère que les *MLE* ne sont pas inductibles par le stress thermique chez ces deux espèces.
- L'expression des *MLE* chez *A. coffeaeformis* pourrait être induite par les stress basses températures. De plus, elle apparaît comme thermo-tolérante et les stress hautes températures semblent être bien supportés par cette espèce. Ceci suggère que les *MLE* sont exprimés en réponse au stress chez cette espèce.
- Les *MLE* sont exprimés de manière constitutive chez *A. acutiuscula* et les séquences transcrites présentent des taux d'identité élevés avec la séquence *MLE* consensus obtenues en ADN génomique chez cette espèce suggérant la présence d'un *MLE* récent et potentiellement actif.

De la même manière, nous avons mis en évidence la présence et l'expression des rétrotransposons *Blackbeard* et *Surcouf*, identifiés chez la souche séquencée de *P. tricornutum* (CCMP 632). Ces rétrotransposons ont été utilisés comme témoins de l'induction des ET par le stress mais leur expression s'est révélée variable selon les conditions ne permettant pas de conclure sur une induction par le stress thermique. Le rétrotransposon *Surcouf* n'a pas été détecté dans le génome de *A. acutiuscula*. Ces résultats pourraient suggérer une variabilité d'induction des ET en fonction de l'espèce et en fonction du stress appliqué ainsi qu'une régulation transcriptionnelle complexe des ET.

Les *MLE* et les rétrotransposons sont donc présents et exprimés chez les diatomées et présentent des profils d'expression variables selon l'espèce et selon son état physiologique (stressée ou non). Les résultats de cette première approche sont encourageants pour la suite de ce travail et permettent d'envisager l'étude de l'expression des *MLE* chez les diatomées de manière quantitative, ce qui permettrait d'améliorer la détection des *MLE* transcrits et de conclure sur l'induction de ceux-ci en conditions de stress.

De plus les expériences présentées dans cette partie sont actuellement reproduites afin de confirmer ces premiers résultats.

# **CONCLUSIONS ET PERSPECTIVES**

En première partie de ce travail de thèse, nous nous sommes attachés à mettre en évidence la présence de *MLE* dans les génomes de microalgues marines et plus particulièrement de diatomées. La détection des *MLE* a été réalisée au moyen de techniques de PCR avec les amorces dégénérées définies par Feschotte et Wessler (2002) qui leur ont permis d'amplifier des fragments de *MLE* chez les Angiospermes. Ces amorces permettent d'amplifier la région catalytique très conservée des *MLE*. Des fragments de 380 pb identifiés positivement comme *MLE* de plantes ont été amplifiés chez 10 diatomées sur 15 espèces étudiées; par contre, aucune amplification n'a été obtenue chez la Prasinophycée et la Rhodophycée. A notre connaissance, ces résultats constituent les premières détections de *MLE* chez des microalgues marines et donc les premiers *MLE* découverts chez les diatomées marines.

L'analyse des séquences de *MLE* de diatomées marines a révélé une similarité avec les *MLE* isolés chez les Angiospermes. Ces similitudes étant restreintes à des motifs, il est très probable qu'ils aient un rôle fonctionnel primordial dans la transposase du fait de leur conservation. En dehors des motifs conservés, le polymorphisme des *MLE* de diatomées marines est important. Les séquences peuvent être très divergentes tant au niveau nucléique que protéique et ce, même dans un même génome. Ainsi chez une même espèce, plusieurs copies de *MLE* ont pu être détectées montrant entre elles des divergences notables.

L'analyse des séquences de 380 pb de *MLE* des diatomées marines traduites en protéines a révélé la formation de groupes qui ne sont pas monospécifiques. Les espèces de diatomées échantillonnées appartiennent à divers genres et aux deux ordres des pennées et des centriques. Lors des analyses, les séquences *MLE* se regroupent indépendamment des espèces, des genres et de l'évolution présumée des ordres. Ces résultats suggèrent que l'évolution des *MLE* est complexe dans les génomes de diatomées marines et qu'elle peut être le résultat d'événements d'amplifications de certaines copies de *MLE*, d'extinction d'autres copies et éventuellement de transferts horizontaux entre espèces éloignées phylogénétiquement. **Pour confirmer ces interprétations, il serait judicieux d'étudier la phylogénie comparée des espèces de diatomées échantillonnées grâce au gène 18 S (Medlin et al., 1993) ou au gène chloroplastique *rbcL* (Ribulose 1,5-bisphosphate carboxylase) (Alverson et al., 2007) par rapport à celle des *MLE* obtenue ici. Cela permettrait de mettre en évidence d'éventuels transferts horizontaux. De plus, augmenter l'échantillonnage des diatomées centriques permettrait également de voir si ces diatomées ancestrales présentent une composition en *MLE* différente ou si les séquences *MLE* présentent un polymorphisme variable par rapport aux diatomées pennées plus récentes.**

Parmi les espèces de diatomées testées sont incluses les deux diatomées modèles dont le génome a été séquencé, *Thalassiosira pseudonana* et *Phaeodactylum tricorutum* (Armbrust et al., 2004; Bowler et al., 2008). Notre étude n'a pas révélé la présence de *MLE* chez la première espèce, toutefois nous avons identifié plusieurs copies de *MLE* chez la deuxième espèce, *Phaeodactylum tricorutum*. La souche dont nous disposons et chez laquelle les *MLE* ont été identifiés est la souche NCC340, elle a été isolée sur la côte atlantique française et elle est conservée au sein de l'allothèque du laboratoire MMS Nantes. Cette souche diffère de celle dont le génome a été publié qui est une souche anglaise CCMP632/CCAP1055.1 (De Martino et al., 2007; Bowler et al., 2008). Nous avons recherché par une méthode bioinformatique la présence des séquences *MLE* que nous avons identifiées dans le génome de la souche CCMP632/CCAP1055.1. Ceci n'a effectivement pas révélé l'existence de *MLE* dans le génome séquencé de *P. tricorutum*. Néanmoins, nous avons pu détecter expérimentalement la présence d'un fragment *MLE* de 380 pb dans cette souche séquencée, mais ce fragment semble différent de ceux obtenus chez *P. tricorutum* NCC340. Ce résultat surprenant soulève des questionnements sur les méthodes de séquençage et d'annotations ainsi que sur les méthodes de détection bioinformatique des ET. Les méthodes bioinformatiques doivent être couplées aux recherches expérimentales pour caractériser les ET dans les génomes séquencés. **Les fragments *MLE* divergents obtenus dans les deux souches suggèrent qu'il serait approprié de rechercher des *MLE* chez les différentes souches de *P. tricorutum* existantes (9 au total selon De Martino et al., 2007) et de comparer les *MLE* entre souches « écophénotypiques » (souche tropicale, souches aux morphotypes différents, souches sensibles au stress salin etc.). L'analyse des *MLE* au sein des différentes souches de cette espèce permettrait de mesurer l'évolution des *MLE* et d'identifier l'origine du polymorphisme des *MLE*. De plus il serait également intéressant, dans cette optique, de rechercher les *MLE* chez *Fragillariopsis cylindrus*, (une diatomée psychrophile, adaptée aux températures avoisinant zéro) et *Pseudonitzschia multiseriis* (une diatomée toxique et compétitive) dont les génomes complets ont été séquencés très récemment.**

La deuxième partie de ce travail a été consacrée à la comparaison des séquences *MLE* obtenues chez les diatomées marines avec les *MLE* des végétaux terrestres.

L'analyse phylogénétique des *MLE* végétaux, correspondant aux *MLE* de diatomées et aux *MLE* d'Angiospermes obtenus à partir des banques de données, a permis de mettre en évidence une origine commune des *MLE* des végétaux. En effet, toutes les séquences obtenues sont regroupées au sein d'un clade monophylétique bien distinct de la racine représentée par les *MLE* animaux. Le groupe des *MLE* des végétaux est subdivisé en quatre lignées évolutives. Trois de ces lignées sont exclusivement constituées de *MLE* de diatomées marines et la quatrième est une lignée de *MLE* de plantes qui regroupe tous les *MLE* d'Angiospermes terrestres. Les lignées de *MLE* semblent donc avoir divergé très tôt, il est possible que ce soit antérieurement à la radiation entre les microalgues vertes (ancêtre des végétaux) et des microalgues brunes. Ce résultat peut également s'expliquer par des événements de transferts horizontaux de manière assez ancienne car les séquences ont beaucoup divergé. **Pour valider l'une ou l'autre de ces hypothèses, il serait intéressant de pouvoir comparer les éléments complets des plantes terrestres avec des éléments complets issus des diatomées marines. Pour cela, il est nécessaire de rechercher la présence de *MLE* complets chez les diatomées marines car les séquences *MLE* identifiées dans cette étude ne représentent qu'une portion de la triade catalytique des *MLE* correspondant à une fraction de la séquence complète. Néanmoins, cette partie est la mieux conservée, ce qui laisse supposer une divergence croissante sur le reste de la séquence excepté aux niveaux des TIR qui sont généralement conservés du fait de leur rôle critique pour la transposition (Bigot et al., 2005).**

Les *MLE* des diatomées et ceux des plantes terrestres ont également été analysés à l'aide d'une nouvelle approche nommée classification agrégative UPGM-VM développée par le LEGS de Gif-sur-Yvette. Cette méthode innovante permet de comparer des séquences d'ET complets avec des ET incomplets, ce qui lui confère un avantage notable. Des résultats identiques à ceux obtenus en phylogénie classique ont été obtenus, les *MLE* des diatomées marines sont séparés des *MLE* des végétaux terrestres en deux sous-familles différentes. Toutefois celles-ci sont regroupées dans une seule et même famille dénommée Chlorophyllis, bien distincte des autres familles appartenant à la superfamille *Tc1-mariner*. La famille Chlorophyllis pourrait correspondre à la famille appelée DD39D dans les classifications précédentes (Shao et Tu, 2001; Feschotte et Wessler, 2002; Brillet et al., 2007). **Toutefois, pour confirmer cela il est nécessaire d'obtenir des *MLE* complets chez chacune des**

familles botaniques ou genre, ce qui n'est actuellement pas le cas. A ce jour, il n'existe que quelques éléments complets caractérisés par la triade DD39D chez les végétaux terrestres. Nous ne disposons pas d'informations sur la triade catalytique des diatomées. Cette classification nous permet d'envisager que les diatomées puissent posséder une triade catalytique DD39D, ainsi nous envisageons la recherche de *MLE* complets chez les diatomées en utilisant des amorces définies à partir des *MLE* complets de végétaux terrestres de la famille Chlorophyllis et notamment au niveau des TIR. Nous pouvons supposer que si les motifs catalytiques sont conservés entre les *MLE* des Angiospermes et les *MLE* des diatomées marines, il en sera de même en ce qui concerne les TIR (Jacobs et al., 2004; Macas et al., 2005; Feschotte et al., 2003). Parallèlement, la caractérisation des *MLE* complets est en cours au laboratoire MMS par des techniques de PCR inverse et de marche sur le gène.

La divergence des *MLE* entre les plantes terrestres et les diatomées marines soulève la question de l'évolution des *MLE* et des ET en général dans les génomes. Si le polymorphisme actuel est dû à la longue évolution dans chacune des lignées, on peut s'interroger sur les mécanismes qui permettent leur pérennité dans les génomes végétaux actuels. De plus, il est également possible que ces divergences soient dues aux contraintes différentes entre les milieux terrestres et marins. **Pour mettre ceci en évidence, il serait intéressant de rechercher des *MLE* chez des Angiospermes marins tels que la Posidonie par exemple, ou encore chez des organismes proches phylogénétiquement des diatomées tels que des algues brunes. Des clades moins évolués que les Angiospermes, comme les Gymnospermes, ont déjà été étudiés par Feschotte et Wessler (2002) sans résultats. Par ailleurs, il faudrait aussi rechercher les *MLE* chez les Bryophytes et les Ptéridophytes qui n'ont pas encore été étudiées. Une autre hypothèse de travail serait de rechercher les *MLE* chez les diatomées d'eau douce afin d'évaluer l'impact des environnements marin et d'eau douce sur la composition en *MLE* des génomes.**

Dans la troisième partie de ce travail, notre objectif était de caractériser l'expression des *MLE* des diatomées. Pour mettre en évidence une éventuelle expression, une stratégie basée sur l'application de stress thermiques a été mise en place, les stress étant des facteurs qui peuvent induire la transcription des ET (Grandbastien et al., 1998; Capy et al., 2000; Schmidt et Anderson, 2008). Quatre traitements thermiques de courte durée ont été appliqués à trois diatomées marines. L'état de stress des diatomées a été évalué par des mesures de la croissance et de l'activité photosynthétique. Chaque température, hormis la condition témoin, représente un stress plus ou moins intense pour chacune des espèces de diatomées testées *P. tricornutum* (diatomée tempérée), *Amphora coffeaeformis* (diatomée tempérée thermotolérante), *Amphora acutiuscula* (diatomée tropicale thermosensible). L'expression des *MLE* a été détectée chez les trois espèces de diatomées mais elle est variable. De plus, l'expression détectée ne semble pas être corrélée au stress excepté chez *A. coffeaeformis*. Toutefois, l'expression des *MLE*, même variable, permet d'envisager la présence d'éléments complets possédant des cadres de lecture ouverts, chez les diatomées. **La méthode utilisée ici n'est pas quantitative et ce travail représente un premier pas vers l'étude quantitative de l'expression des *MLE* des diatomées. La suite du présent travail est d'ores et déjà en cours puisque l'étude de l'expression des *MLE* chez les diatomées soumises à différents stress (thermiques, métalliques et salins) fait l'objet d'une thèse qui a débuté en septembre 2010. L'expression des *MLE* obtenus chez *A. acutiuscula* nous laisse supposer que cette diatomée contient des *MLE* complets, la recherche de *MLE* pleine taille chez cette diatomée fait également l'objet de la thèse débutée en septembre 2010.**

**Deux aspects supplémentaires de ce travail n'ont pas été développés dans ce manuscrit car ils étaient en cours de réalisation au moment de la rédaction :**

- des réplicats des stress courts (5 h) ainsi que des expériences en stress longs (8 j) ont été réalisés chez *P. tricornutum*. Les résultats sont en cours d'analyse et pourraient permettre de mettre en relation l'expression des *MLE* avec la durée du stress thermique.

- les *MLE* des végétaux continuent d'être collectés dans les banques de données et dans les génomes séquencés. La classification des *MLE* de végétaux terrestres fait l'objet d'une collaboration avec le LEGS de Gif-sur-Yvette. Nous avons collecté plus de 300 séquences au moment de la rédaction de ce manuscrit.

**Au cours de ce travail, nous avons pu répondre à nos trois objectifs : mettre en évidence la présence de *MLE* chez les diatomées marines, les classer et détecter leur expression en conditions de stress thermiques.**

**Durant ces trois dernières années, de nombreuses informations sur la génomique des microalgues ont été communiquées avec notamment le séquençage d'une dizaine de génomes de microalgues. Les données acquises devraient permettre de mieux comprendre leur évolution et leur adaptation à l'environnement marin. Ce travail exploratoire apporte également, dans cette optique, des informations sur la composition génomique des diatomées.**

# BIBLIOGRAPHIE

## A

- Abascal, F., R. Zardoya, et D. Posada. 2007. ProtTest : selection of best-fit models of protein evolution. *Bioinformatics* **21** (9) : 2104-2105.
- Abrusán, G., et H.-J. Krambeck. 2006. Competition may determine the diversity of transposable elements. *Theoretical Population Biology* **70** (3) : 364-375.
- Alleman, M., et M. Freeling. 1986. The *Mu* transposable elements of maize : evidence of transposition and copy number regulation during development. *Genetics* **112** : 107-119.
- Allen, A. E., J. LaRoche, U. Maheswari, M. Lommer, N. Schauer, P. J. Lopez, G. Finazzi, A. R. Fernie, et C. Bowler. 2008. Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation. *Proceedings of the National Academy of Sciences* **105** (30) : 10438-10443.
- Allen, A. E., A. Vardi, et C. Bowler. 2006. An ecological and evolutionary context for integrated nitrogen metabolism and related signaling pathways in marine diatoms. *Current Opinion in Plant Biology* **9** (3) : 264-273.
- Alverson, A. J., R. K. Jansen, et E. C. Theriot. 2007. Bridging the Rubicon : phylogenetic analysis reveals repeated colonizations of marine and fresh waters by thalassiosiroid diatoms. *Molecular Phylogenetics and Evolution* **45** : 193-210.
- Apt, K. E., et P. W. Behrens. 1999. Commercial developments in microalgal biotechnology. *Journal of Phycology* **35** (2) : 215-226.
- Archibald, J. M. 2009. The puzzle of plastid evolution. *Current Biology* **19** (2) : R81-R88.
- Arkhipova, I. R., et M. Meselson. 2005. Diverse DNA transposons in rotifers of the class Bdelloidea. *Proceedings of the National Academy of Sciences* **102** (33) : 11781-11786.
- Armbrust, E. V. 2009. The life of diatoms in the world's oceans. *Nature* **459** (7244) : 185-192.
- Armbrust, E. V., J. A. Berges, C. Bowler, B. R. Green, D. Martinez, N. H. Putnam, S. Zhou, A. E. Allen, K. E. Apt, M. Bechner, M. A. Brzezinski, B. K. Chaal, A. Chiovitti, A. K. Davis, M. S. Demarest, J. C. Detter, T. Glavina, D. Goodstein, M. Z. Hadi, U. Hellsten, M. Hildebrand, B. D. Jenkins, J. Jurka, V. V. Kapitonov, N. Kroger, W. W. Y. Lau, T. W. Lane, F. W. Larimer, J. C. Lippmeier, S. Lucas, M. Medina, A. Montsant, M. Obornik, M. S. Parker, B. Palenik, G. J. Pazour, P. M. Richardson, T. A. Rynearson, M. A. Saito, D. C. Schwartz, K. Thamatrakoln, K. Valentin, A. Vardi, F. P. Wilkerson, et D. S. Rokhsar. 2004. The genome of the diatom *Thalassiosira Pseudonana* : ecology, evolution, and metabolism. *Science* **306** (5693) : 79-86.
- Atkin, O. K., et M. G. Tjoelker. 2003. Thermal acclimation and the dynamic response of plant respiration to temperature. *Trends in Plant Science* **8** (7) : 343-351.

- Augé-Gouillou, C., M.-H. Hamelin, M.-V. Demattei, G. Periquet, et Y. Bigot. 2001a. The ITR binding domain of the *mariner Mos-1* transposase. *Molecular Genetics and Genomics* **265** : 58-65.
- Augé-Gouillou, C., M.-H. Hamelin, M.-V. Demattei, M. Periquet, and Y. Bigot. 2001b. The wild-type conformation of the *Mos-1* inverted terminal repeats is suboptimal for transposition in bacteria. *Molecular Genetics and Genomics* **265** : 51-57.
- Aziz, R. K., M. Breitbart, et R. A. Edwards. 2010. Transposases are the most abundant, most ubiquitous genes in nature. *Nucleic Acids Research* **38 (13)** : 4207-4217.

## B

- Bae, Y.-A., S.-Y. Moon, Y. Kong, S.-Y. Cho, et M.-G. Rhyu. 2001. *CsRn1*, a novel active retrotransposon in a parasitic trematode, *Clonorchis sinensis*, discloses a new phylogenetic clade of *Ty3/gypsy*-like LTR retrotransposons. *Molecular Biology and Evolution* **18 (8)** : 1474-1483.
- Ball, S. G. 2005. Eukaryotic microalgae genomics. The essence of being a plant. *Plant physiology* **137 (2)** : 397-398.
- Baniwal, S., K. Bharti, K. Chan, M. Fauth, A. Ganguli, S. Kotak, S. Mishra, L. Nover, M. Port, K. D. Scharf, J. Tripp, C. Weber, D. Zielinski, et P. von Koskull-Doring. 2004. Heat stress response in plants : a complex game with chaperones and more than twenty heat stress transcription factors. *Journal of Biosciences* **29 (4)** : 471-487.
- Banzet, N., C. Richaud, Y. Deveaux, M. Kazmaier, J. Gagnon, et C. Triantaphylides. 1998. Accumulation of small heat shock proteins, including mitochondrial HSP22, induced by oxidative stress and adaptive response in tomato cells. *The Plant Journal* **13 (4)** : 519-527.
- Bao, W., M. G. Jurka, V. V. Kapitonov, et J. Jurka. 2009. New superfamilies of eukaryotic DNA transposons and their internal divisions. *Molecular Biology and Evolution* **26 (5)** : 983-993.
- Barry, E. G., D. J. Witherspoon, et D. J. Lampe. 2004. A bacterial genetic screen identifies functional coding sequences of the insect *mariner* transposable element *Famar1* amplified from the genome of the earwig, *Forficula auricularia*. *Genetics* **166 (2)** : 823-833.
- Barsanti, L., et P. Gualtieri. 2005. *Algae, anatomy, biochemistry, and biotechnology*. Taylor et Francis (Eds), Boca Raton.
- Beguiristain, T., M.-A. Grandbastien, P. Puigdomenech, et J. M. Casacuberta. 2001. Three Tnt1 subfamilies show different stress-associated patterns of expression in Tobacco. Consequences for retrotransposon control and evolution in plants. *Plant Physiology* **127** : 212-221.
- Bennetzen, J. L., J. Ma, et K. M. Devos. 2005. Mechanisms of recent genome size variation in flowering plants. *Annals of Botany* **95** : 127-132.
- Bergman, C. M., et H. Quesneville. 2007. Discovering and detecting transposable elements in genome sequences. *Briefings in Bioinformatics* **8 (6)** : 382-392.

- Bhattacharya, D., et L. Medlin. 1998. Algal phylogeny and the origin of land plants. *Plant physiology* **116** : 9-15.
- Biéumont, C., et C. Vieira. 2006. Genetics : Junk DNA as an evolutionary force. *Nature* **443** : 521-524.
- Biéumont, C., A. Tsitrone, C. Vieira, et C. Hoogland. 1997. Transposable element distribution in *Drosophila*. *Genetics* **147** (4) : 1997-1999.
- Biéumont, C. 1994. Dynamic equilibrium between insertion and excision of *P* elements in highly inbred lines from an M' strain of *Drosophila melanogaster*. *Journal of Molecular Evolution* **39** : 466-472.
- Bigot, Y., B. Brillet, et C. Augé-Gouillou. 2005. Conservation of palindromic and mirror motifs within inverted terminal repeats of *mariner*-like elements. *Journal of Molecular Biology* **351** : 108-116.
- Bijlsma, R., et V. Loeschcke. 1997. *Environmental stress, adaptation, and evolution*. Birkhauser Press, Berlin.
- Bingham, P. M., M. G. Kidwell, et G. M. Rubin. 1982. The molecular basis of P-M hybrid dysgenesis : The role of the P element, a P-strain-specific transposon family. *Cell* **29** (3) : 995-1004.
- Blinn, D. W., R. Truitt, et A. Pickart. 1989. Response of epiphytic diatom communities from the tailwaters of Glen Canyon Dam, Arizona, to elevated water temperature. *Regulated Rivers : Research & Management* **4** : 91-96.
- Böhne, A., F. Brunet, D. Galiana-Arnoux, C. Schultheis, et J.-N. Volff. 2008. Transposable elements as drivers of genomic and biological diversity in vertebrates. *Chromosome Research* **16** : 203-215.
- Bohnert, H. J., et E. Sheveleva. 1998. Plant stress adaptations : making metabolism move. *Current Opinion in Plant Biology* **1** (3) : 267-274.
- Bouvet, G. F., V. Jacobi, K. V. Plourde, et L. Bernier. 2008. Stress-induced mobility of *OPHI01* and *OPHI02*, DNA transposons of the Dutch elm disease fungi. *Fungal Genetics and Biology* **45** (4) : 565-578.
- Bouvier, A., M. George, et F. Le Lionnais. 2005. *Dictionnaire des mathématiques*. PUF, France.
- Bowler, C., A. E. Allen, J. H. Badger, J. Grimwood, K. Jabbari, A. Kuo, U. Maheswari, C. Martens, F. Maumus, R. P. Otiillar, E. Rayko, A. Salamov, K. Vandepoele, B. Beszteri, A. Gruber, M. Heijde, M. Katinka, T. Mock, K. Valentin, F. Verret, J. A. Berges, C. Brownlee, J.-P. Cadoret, A. Chiovitti, C. J. Choi, S. Coesel, A. De Martino, J. C. Detter, C. Durkin, A. Falciatore, J. Fournet, M. Haruta, M. J. J. Huysman, B. D. Jenkins, K. Jiroutova, R. E. Jorgensen, Y. Joubert, A. Kaplan, N. Kroger, P. G. Kroth, J. La Roche, E. Lindquist, M. Lommer, V. Martin-Jezequel, P. J. Lopez, S. Lucas, M. Mangogna, K. McGinnis, L. K. Medlin, A. Montsant, M.-P. O.-L. Secq, C. Napoli, M. Obornik, M. S. Parker, J.-L. Petit, B. M. Porcel, N. Poulsen, M. Robison, L. Rychlewski, T. A. Ryneerson, J. Schmutz, H. Shapiro, M. Siat, M. Stanley, M. R. Sussman, A. R. Taylor, A. Vardi, P. von Dassow, W. Vyverman, A. Willis, L.

- S. Wyrwicz, D. S. Rokhsar, J. Weissenbach, E. V. Armbrust, B. R. Green, Y. Van de Peer, et I. V. Grigoriev. 2008. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **456 (7219)** : 239-244.
- Bowler, C., M. V. Montagu, et D. Inze. 1992. Superoxide dismutase and stress tolerance. *Annual Review of Plant Physiology and Plant Molecular Biology* **43** : 83-116.
- Boyer, J. S. 1982. Plant productivity and environment. *Science* **218 (4571)** : 443-448.
- Brillet, B., Y. Bigot, et C. Augé-Gouillou. 2007. Assembly of the *Tc1* and *mariner* transposition initiation complexes depends on the origins of their transposase DNA binding domains. *Genetica* **130 (2)** : 105-120.
- Brookfield, J., et R. Badge. 1997. Population genetics models of transposable elements. *Genetica* **100** : 281-294.
- Bui, Q.-T., N. Casse, V. Leignel, V. Nicolas, et B. Chénais. 2008. Widespread occurrence of *mariner* transposons in coastal crabs. *Molecular Phylogenetics and Evolution* **47 (3)** : 1181-1189.
- Bui, Q.-T., L. Delaurière, N. Casse, V. Nicolas, M. Laulier, et B. Chénais. 2007. Molecular characterization and phylogenetic position of a new *mariner*-like element in the coastal crab, *Pachygrapsus marmoratus*. *Gene* **396 (2)** : 248-256.
- Bureau, T. E., et S. R. Wessler. 1994. *Stowaway* : a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. *Plant Cell* **6 (6)** : 907-916.
- Bureau, T. E., et S. R. Wessler. 1992. *Tourist* : a large family of small inverted repeat elements frequently associated with maize genes. *Plant Cell* **4 (10)** : 1283-1294.
- C
- Cadoret, J.-P., et O. Bernard. 2008. La production de biocarburant lipidique avec des microalgues : promesses et défis. *Journal de la Société de Biologie* **202 (3)** : 201-211.
- Calvi, B. R., T. J. Hong, S. D. Findley, et W. M. Gelbart. 1991. Evidence for a common evolutionary origin of inverted repeat transposons in *Drosophila* and plants : *hobo*, *activator*, and *Tam3*. *Cell* **66 (3)** : 465-471.
- Cappello, J., S. M. Cohen, et H. F. Lodish. 1984. *Dictyostelium* transposable element *DIRS-1* preferentially inserts into *DIRS-1* sequences. *Molecular and Cellular Biology* **4 (10)** : 2207-2213.
- Capy, P. 2004. Structure et évolution des éléments transposables. *Journal de la Société de Biologie* **198 (4)** : 393-398.
- Capy, P., G. Gasperi, C. Biémont, et C. Bazin. 2000. Stress and transposable elements : co-evolution or useful parasites? *Heredity* **85 (2)** : 101-106.
- Capy, P., C. Bazin, D. Higué, et T. Langin. 1997. *Dynamics and evolution of transposable elements*. Landes Company, Austin.

- Capy, P., D. Anxolabéhère, et T. Langin. 1994. The strange phylogenies of transposable elements : are horizontal transfers the only explanation? *Trends in Genetics* **10** : 7-12.
- Capy, P., R. Vitalis, T. Langin, D. Higuët, et C. Bazin. 1996. Relationships between transposable elements based upon the integrase-transposase domains : is there a common ancestor? *Journal of Molecular Evolution* **42** (3) : 359-368.
- Capy, P., A. Koga, J. R. David, et D. L. Hartl. 1992. Sequence analysis of active *mariner* elements in natural populations of *Drosophila simulans*. *Genetics* **130** (3) : 499-506.
- Cardozo, K. H. M., T. Guaratini, M. P. Barros, V. R. Falcao, A. P. Tonon, N. P. Lopes, S. Campos, M. A. Torres, A. O. Souza, P. Colepicolo, et E. Pinto. 2007. Metabolites from algae with economical impact. *Comparative Biochemistry and Physiology Part C : Toxicology and Pharmacology* **146** (1-2) : 60-78.
- Caruso, A., D. Morabito, F. Delmotte, G. Kahlem, et S. Carpin. 2002. Dehydrin induction during drought and osmotic stress in *Populus*. *Plant Physiology and Biochemistry* **40** (12) : 1033-1042.
- Casacuberta, J. M., et M.-A. Grandbastien. 1993. Characterisation of LTR sequences involved in the protoplast specific expression of the tobacco *Tnt1* retrotransposon. *Nucleic Acids Research* **21** (9) : 2087-2093.
- Casola, C., D. Hucks, et C. Feschotte. 2008. Convergent domestication of *Pogo*-like transposases into centromere-binding proteins in fission yeast and mammals. *Molecular Biology and Evolution* **25** : 29-41.
- Casse, N., Q. T. Bui, V. Nicolas, S. Renault, Y. Bigot, et M. Laulier. 2006. Species sympatry and horizontal transfers of *mariner* transposons in marine crustacean genomes. *Molecular Phylogenetics and Evolution* **40** (2) : 609-619.
- Casse, N., E. Pradier, M.-V. Demattei, Y. Bigot, et M. Laulier. 2002. *Mariner* transposons are widespread genetic parasites in the genome of hydrothermal invertebrates. *Cahiers de Biologie Marine* **43** (3-4) : 325-328.
- Chakrani, F., P. Capy, et J. David. 1993. Developmental temperature and somatic excision rate of *mariner* transposable element in three natural populations of *Drosophila simulans*. *Genetics, Selection, Evolution* **25** (2) : 121-132.
- Chalvet, F., C. Grimaldi, F. Kaper, T. Langin, et M.-J. Daboussi. 2003. *Hop*, an active *Mutator*-like element in the genome of the fungus *Fusarium oxysporum*. *Molecular Biology and Evolution* **20** (8) : 1362-1375.
- Charles, M., H. Belcram, J. Just, C. Huneau, A. Viollet, A. Couloux, B. Segurens, M. Carter, V. Huteau, O. Coriton, R. Appels, S. Samain, et B. Chalhoub. 2008. Dynamics and differential proliferation of transposable elements during the evolution of the B and A genomes of wheat. *Genetics* **180** (2) : 1071-1086.

- Charlesworth, D., et S. I. Wright. 2001. Breeding systems and genome evolution. *Current Opinion in Genetics and Development* **11** (6) : 685-690.
- Cheng, X., D. Zhang, Z. Cheng, B. Keller, et H.-Q. Ling. 2009. A new family of *Ty1-copia*-like retrotransposons originated in the tomato genome by a recent horizontal transfer event. *Genetics* **181** (4) : 1183-1193.
- Cheng, C., M. Daigen, et H. Hirochika. 2006. Epigenetic regulation of the rice retrotransposon *Tos17*. *Molecular Genetics and Genomics* **276** (4) : 378-390.
- Cheong, Y. H., H.-S. Chang, R. Gupta, X. Wang, T. Zhu, et S. Luan. 2002. Transcriptional profiling reveals novel interactions between wounding, pathogen, abiotic stress, and hormonal responses in *Arabidopsis*. *American Society of Plant Biology* **129** : 661-677.
- Chepurnov, V. A., D. G. Mann, K. Sabbe, et W. Vyverman. 2004. Experimental studies on sexual reproduction in diatoms. *International Review of Cytology* **237** : 91-154.
- Chisti, Y. 2007. Biodiesel from microalgae. *Biotechnology Advances* **25** (3) : 294-306.
- Cid, A., C. Herrero, E. Torres, et J. Abalde. 1995. Copper toxicity on the marine microalga *Phaeodactylum tricornutum* : effects on photosynthesis and related parameters. *Aquatic Toxicology* **31** (2) : 165-174.
- Claudianos, C., J. Brownlie, R. Russell, J. Oakeshott, et S. Whyard. 2002. *maT*-A clade of transposons intermediate between *mariner* and *Tc1*. *Molecular Biology and Evolution* **19** (12) : 2101-2109.
- Coen, E. S., R. Carpenter, et C. Martin. 1986. Transposable elements generate novel spatial patterns of gene expression in *Antirrhinum majus*. *Cell* **47** (2) : 285-296.
- Cohen, J. B., B. Hoffman-Liebermann, et L. Kedes. 1985. Structure and unusual characteristics of a new family of transposable elements in the sea urchin *Strongylocentrotus purpuratus*. *Molecular and Cellular Biology* **5** (10) : 2804-2813.
- Collins, J., B. Saari, et P. Anderson. 1987. Activation of a transposable element in the germ line but not in the soma of *Caenorhabditis elegans*. *Nature* **328** : 726-728.
- Colloms, S. D., H. G. A. M. van Luenen, et R. H. A. Plasterk. 1994. DNA binding activities of the *Caenorhabditis elegans Tc3* transposase. *Nucleic Acids Research* **22** (25) : 5548-5554.
- Cooksey, K. E. 1981. Requirement for calcium in adhesion of a fouling diatom to glass. *Applied and Environmental Microbiology* **41** (6) : 1378-1382.
- Crippen, R. W., et J. L. Perrier. 1974. The use of neutral red and Evans blue for live-dead determinations of marine plankton. *Stain Technology* **49** : 97-104.
- Crooks, G., G. Hon, J.-M. Chandonia, et S. Brenner. 2004. WebLogo : a sequence logo generator. *Genome research* **14** : 1188-1190.
- Cui, Z., A. M. Geurts, G. Liu, C. D. Kaufman, et P. B. Hackett. 2002. Structure-function analysis of the inverted terminal repeats of the *sleeping beauty* transposon. *Journal of Molecular Biology* **318** (5) : 1221-1235.

Cushman, J. C., et H. J. Bohnert. 2000. Genomic approaches to plant stress tolerance. *Current Opinion in Plant Biology* **3** (2) : 117-124.

## D

Daboussi, M.-J., et P. Capy. 2003. Transposable elements in filamentous fungi. *Annual Review of Microbiology* **57** : 275-299.

Daboussi, M.-J., T. Langin, et Y. Brygoo. 1992. *Fot1*, a new family of fungal transposable elements. *Molecular and General Genetics* **232** : 12-16.

Davison, I. R., et G. A. Pearson. 1996. Stress tolerance in intertidal seaweeds. *Journal of Phycology* **32** (2) : 197-211.

Davison, I. R. 1991. Environmental effects on algal photosynthesis : temperature. *Journal of Phycology* **27** : 2-8.

Dawson, A., et D. J. Finnegan. 2003. Excision of the *Drosophila mariner* transposon *Mos1* : comparison with bacterial transposition and V(D)J recombination. *Molecular Cell* **11** : 225-235.

De Martino, A., A. Meichenin, J. Shi, K. Pan, et C. Bowler. 2007. Genetic and phenotypic characterization of *Phaeodactylum tricorutum* (Bacillariophyceae) accessions. *Journal of Phycology* **43** (5) : 992-1009.

Delauriere, L., B. Chénais, E. Pradier, Y. Hardivillier, S. Renault, et N. Casse. 2009. DNA binding specificity and cleavage activity of *Pacmmar* transposase. *Biochemistry* **48** (30) : 8.

Derelle, E., C. Ferraz, S. Rombauts, P. Rouze, A. Z. Worden, S. Robbens, F. Partensky, S. Degroeve, S. Echeynié, R. Cooke, Y. Saeys, J. Wuyts, K. Jabbari, C. Bowler, O. Panaud, B. Piégu, S. G. Ball, J.-P. Ral, F.-Y. Bouget, G. Piganeau, B. De Baets, A. Picard, M. Delseny, J. Demaille, Y. Van de Peer, et H. Moreau. 2006. Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proceedings of the National Academy of Sciences* **103** (31) : 11647-11652.

D'Erfurth, I., V. Cosson, A. Eschstruth, H. Lucas, A. Kondorosi, et P. Ratet. 2003. Efficient transposition of the *Tnt1* tobacco retrotransposon in the model legume *Medicago truncatula*. *The Plant Journal* **34** : 95-106.

Desalvo, M. K., C. R. Voolstra, S. Sunagawa, J. A. Schwarz, J. H. Stillman, M. A. Coffroth, A. M. Szmant, et M. Medina. 2008. Differential gene expression during thermal stress and bleaching in the Caribbean coral *Montastraea faveolata*. *Molecular Ecology* **17** (17) : 3952-3971.

Dheda, K., J. F. Huggett, S. A. Bustin, M. A. Johnson, G. Rook, et A. Zulma. 2004. Validation of housekeeping genes for normalizing RNA expression in real-time PCR. *Biotechniques* **37** : 6.

Diao, X., M. Freeling, et D. Lisch. 2006. Horizontal transfer of a plant transposon. *PLoS Biol* **4** : e5.

- Doak, T. G., F. P. Doerder, C. L. Jahn, et G. Herrick. 1994. A proposed superfamily of transposase genes : transposon-like elements in ciliated protozoa and a common "D35E" motif. *Proceedings of the National Academy of Sciences* **91** (3) : 942-946.
- Doolittle, W. F., et C. Sapienza. 1980. Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284** (5757) : 601-603.
- Dooner, H. K., et C. F. Weil. 2007. Give-and-take : interactions between DNA transposons and their host plant genomes. *Current Opinion in Genetics and Development* 17 (6) : 486-492.
- Doyle, J. J., et Doyle J.L. 1990. Isolation of plant DNA from fresh tissue. *Focus* 12 : 13-15.
- Drebes, G. 1977. Sexuality. *In : The biology of diatoms*, D. Werner (Ed). Blackwell Scientific, Berkley, pp. 252-283.
- Dufresne, M., A. Hua-Van, H. Abd el Wahab, S. B. M'Barek, C. Vasnier, L. Teyssset, G. H. J. Kema, et M.-J. Daboussi. 2007. Transposition of a fungal miniature inverted-repeat transposable element through the action of a *Tc1-like* transposase. *Genetics* **175** : 441-452.

## E

- Eichenbaum, Z., et Z. Livneh. 1998. UV light induces IS10 transposition in *Escherichia coli*. *Genetics* 149 (3) : 1173-1181.
- Eickbusch, T. H., et H. S. Malik. 2002. Origins and evolution of retrotransposons. *In : Mobile DNA II*, N. L. Craig, R. Craigie, M. Gellert, et A. M. Lambowitz (Eds). American Society of Microbiology Press, Washington, pp. 1111-1146.
- Emmons, S. W., L. Yesner, K. Ruan, et D. Katzenberg. 1983. Evidence for a transposon in *Caenorhabditis elegans*. *Cell* 32 : 55-65.
- Emmons, S. W., M. R. Klass, et D. Hirsh. 1979. Analysis of the constancy of DNA sequences during development and evolution of the nematode *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences* **76** (3) : 1333-1337.
- Evertts, A. G., C. Plymire, N. L. Craig, et H. L. Levin. 2007. The *Hermes* transposon of *Musca domestica* is an efficient tool for the mutagenesis of *Schizosaccharomyces pombe*. *Genetics* **177** (4) : 2519-2523.
- Evgen'ev, M. B., et I. R. Arkhipova. 2005. *Penelope*-like elements - a new class of retroelements : distribution, function and possible evolutionary significance. *Cytogenetic and Genome Research* **110** (1-4) : 510-521.
- Evgen'ev, M. B., H. Zelentsova, N. Shostak, M. Kozitsina, V. Barskyi, D.-H. Lankenau, et V. G. Corces. 1997. *Penelope*, a new family of transposable elements and its possible role in hybrid dysgenesis in *Drosophila virilis*. *Proceedings of the National Academy of Sciences* **94** : 196-201.

## F

- Falciatore, A., et C. Bowler. 2002. Revealing the molecular secrets of marine diatoms. *Annual Review of Plant Biology* **53** : 109-130.
- Falkowski, P. G., M. E. Katz, A. H. Knoll, A. Quigg, J. A. Raven, O. Schofield, et F. J. R. Taylor. 2004. The evolution of modern eukaryotic phytoplankton. *Science* 305 (5682) : 354-360.
- Falkowski, P. G., et J. A. Raven. 1997. *Aquatic Photosynthesis*. Blackwell Scientific, Oxford.
- Farineau, J., J. F. Morot-Gaudry, et J. F. Soussana. 2006. *La photosynthèse, processus physiques, moléculaires et physiologiques*. INRA, Paris.
- Favatier, F., L. Bornman, L. E. Hightower, E. Gånther, et B. S. Polla. 1997. Variation in hsp gene expression and Hsp polymorphism : do they contribute to differential disease susceptibility and stress tolerance? *Cell stress and Chaperones* 2 : 141-155.
- Feschotte, C. 2008. Transposable elements and the evolution of regulatory networks. *Nature Reviews Genetics* 9 (5) : 397-405.
- Feschotte, C., et E. J. Pritham. 2007. DNA transposons and the evolution of eukaryotic genomes. *Annual Review of Genetics* **41** : 331-368.
- Feschotte, C., et E. J. Pritham. 2005. Non-mammalian c-integrases are encoded by giant transposable elements. *Trends in Genetics* **21 (10)** : 551-552.
- Feschotte, C., M. T. Osterlund, R. Peeler, et S. R. Wessler. 2005. DNA-binding specificity of rice *mariner*-like transposases and interactions with *Stowaway* MITEs. *Nucleic Acids Research* **33 (7)** : 2153-2165.
- Feschotte, C. 2004. *Merlin*, a new superfamily of DNA transposons identified in diverse animal genomes and related to bacterial *IS1016* insertion sequences. *Molecular Biology and Evolution* **21 (9)** : 1769-1780.
- Feschotte, C., L. Swamy, et S. R. Wessler. 2003. Genome-wide analysis of *mariner*-like transposable elements in rice reveals complex relationships with *Stowaway* miniature inverted repeat transposable elements (MITEs). *Genetics* **163 (2)** : 747-758.
- Feschotte, C., N. Jiang, et S. R. Wessler. 2002. Plant transposable elements : where genetics meets genomics. *Nature* **3 (5)** : 329-341.
- Feschotte, C., et S. R. Wessler. 2002. *Mariner*- like transposases are widespread and diverse in flowering plants. *Proceedings of the National Academy of Sciences* **99** : 280-285.
- Feschotte, C., et S. R. Wessler. 2001. Treasures in the attic : rolling circle transposons discovered in eukaryotic genomes. *Proceedings of the National Academy of Sciences* **98 (16)** : 8923-8924.
- Feschotte, C., et C. Mouchès. 2000. Evidence that a family of miniature inverted-repeat transposable elements (MITEs) from the *Arabidopsis thaliana* genome has arisen from a *pogo*-like DNA transposon. *Molecular Biology and Evolution* **17 (5)** : 730-737.
- Field, C. B., M. J. Behrenfeld, J. T. Randerson, et P. Falkowski. 1998. Primary production of the biosphere : integrating terrestrial and oceanic components. *Science* **281 (5374)** : 237-240.

- Figueroa, F., R. Conde-Alvarez, et I. Gomez. 2003. Relations between electron transport rates determined by pulse amplitude modulated chlorophyll fluorescence and oxygen evolution in macroalgae under different light conditions. *Photosynthesis Research* **75** (3) : 259-275.
- Finnegan, D. J. 1989. Eukaryotic transposable elements and genome evolution. *Trends in Genetics* **5** : 103-107.
- Flavell, A. J. 1992. Ty1-copia group retrotransposons and the evolution of retroelements in the eukaryotes. *Genetica* **86** : 203-214.
- Forbes, V. E., et P. Calow. 1997. Response of aquatic organisms to pollutant stress : Theoretical and practical implications. *In : Environmental stress, adaptation, and evolution*, R. Bijlsma et V. Loeschke (Eds). Birkhauser Press, Berlin, pp. 25-41.
- Fornace, A. J., et J. B. Mitchell. 1986. Induction of B2 RNA polymerase III transcription by heat shock : enrichment for heat shock induced sequences in rodent cells by hybridization subtraction. *Nucleic Acids Research* **14** (14) : 5793-5811.
- Fortune, P., A. Roulin, et O. Panaud. 2008. Horizontal transfer of transposable elements in plants. *Communicative and integrative biology* **1** : 1 - 4.
- Fouqueray, M., J.-L. Mouget, A. Morant-Manceau, et G. Tremblin. 2007. Dynamics of short-term acclimation to UV radiation in marine diatoms. *Journal of Photochemistry and Photobiology B : Biology* **89** : 1-8.
- Foyer, C. H., M. Lelandais, et K. J. Kunert. 1994. Photooxidative stress in plants. *Physiologia Plantarum* **92** (4) : 696-717.
- Franz, G., T. G. Loukeris, G. Dialektaki, C. R. Thompson, et C. Savakis. 1994. Mobile *Minos* elements from *Drosophila hydei* encode a two-exon transposase with similarity to the paired DNA-binding domain. *Proceedings of the National Academy of Sciences* **91** (11) : 4746-4750.
- Fraser, M. J., J. S. Brusca, G. E. Smith, et M. D. Summers. 1985. Transposon-mediated mutagenesis of a baculovirus. *Virology* **145** : 356-361.

## G

- Garcia-Fernández, J., J. R. Bayascas-Ramirez, G. Marfany, A. M. Munoz-Marmol, A. Casali, J. Baguna, et E. Salo. 1995. High copy number of highly similar *mariner*-like transposons in planarian (*Platyhelminthe*) : evidence for a trans-phylo horizontal transfer. *Molecular Biology and Evolution* **12** (3) : 421-431.
- Gaur, J. P., et L. C. Rai. 2001. *in Algal adaptation to environmental stresses : Physiological, Biochemical and Molecular Mechanism*. (L.C. RAI et J.P. GAUR, Eds. Springer Verlag Pub., Heidelberg : 363.
- Geider, R. J. 1987. Light and temperature dependence of the carbon to chlorophyll a ratio in microalgae and cyanobacteria. *New Phytologist* **106** : 1-34.

- Gilbert, C., S. Schaack, J. K. Pace II, P. J. Brindley, et C. Feschotte. 2010. A role for host-parasite interactions in the horizontal transfer of transposons across phyla. *Nature* **464 (7293)** : 1347-1350.
- Goldman, J. C., et R. Mann. 1980. Temperature-influenced variations in speciation and chemical composition of marine phytoplankton in outdoor mass cultures. *Journal of Experimental Marine Biology and Ecology* **46** : 29-39.
- Goldman, J. C., et E. J. Carpenters. 1974. A kinetics approach to the effect of temperature on algal growth. *Limnology and Oceanography* **19 (5)** : 756-766.
- Gomulski, L. M., C. Torti, M. Bonizzoni, D. Moralli, E. Raimondi, P. Capy, G. Gasperi, et A. R. Malacrida. 2001. A new basal subfamily of *mariner* elements in *Ceratitis rosa* and other Tephritid flies. *Journal of Molecular Evolution* **53 (6)** : 597-606.
- Gomulski, L. M., C. Torti, A. R. Malacrida, et G. Gasperi. 1997. *Ccmar1*, a full-length *mariner* element from the mediterranean fruit fly, *Ceratitis capitata*. *Insect Molecular Biology* **6 (3)** : 241-253.
- Goodwin, T. J. D., M. I. Butler, et R. T. M. Poulter. 2003. *Cryptons* : a group of tyrosine-recombinase-encoding DNA transposons from pathogenic fungi. *Microbiology* **149 (11)** : 3099-3109.
- Göttgens, B., L. M. Barton, D. Grafham, M. Vaudin, et A. R. Green. 1999. *Tdr2*, a new zebrafish transposon of the *Tc1* family. *Gene* **239 (2)** : 373-379.
- Grandbastien, M.-A., C. Audeon, E. Bonnard, J. M. Casacuberta, B. Chalhoub, A.-P. P. Costa, Q. H. Le, D. Melayah, M. Petit, C. Poncet, S. M. Tam, M.-A. van Sluys, et C. Mhiri. 2005. Stress activation and genomic impact of *Tnt1* retrotransposons in *Solanaceae*. *Cytogenetic and Genome Research* **110 (1-4)** : 229-241.
- Grandbastien, M.-A. 1998. Activation of plant retrotransposons under stress conditions. *Trends in Plant Science* **3 (5)** : 181-187.
- Grandbastien, M.-A., H. Lucas, J.-B. Morel, C. Mhiri, S. Vernhettes, et J. Casacuberta. 1997. The expression of the tobacco *Tnt1* retrotransposon is linked to plant defense responses. *Genetica* **100** : 241-252.
- Groth-Nard, C., et J. M. Robert. 1993. Les lipides des diatomées. *Diatom research* **8 (2)** : 281-308.
- Guihéneuf, F., M. Fouqueray, V. Mimouni, L. Ulmann, B. Jacquette, et G. Tremblin. 2010. Effect of UV stress on the fatty acid and lipid class composition in two marine microalgae : *Pavlova lutheri* (Pavlovophyceae) and *Odontella aurita* (Bacillariophyceae). *Journal of Applied Phycology* **22 (5)** : 629-638.

## H

- Halaimia-Toumi, N., N. Casse, M. V. Demattei, S. Renault, E. Pradier, Y. Bigot, et M. Laulier. 2004. The GC-rich transposon *Bytmar1* from the deep-sea hydrothermal crab, *Bythograea*

- thermydron* may encode three transposase isoforms from a single ORF. *Journal of Molecular Evolution* **59** : 747-760.
- Hammer, S. E., S. Strehl, et S. Hagemann. 2005. homologs of *Drosophila P* transposons were mobile in zebrafish but have been domesticated in a common ancestor of chicken and human. *Molecular Biology and Evolution* **22** (4) : 833-844.
- Hanahan, D., et F. R. Bloom. 1996. Mechanism of DNA transformation. Pages 2448-2459 in F. C. Neidhardt, R. C. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger, editors. *Escherichia coli* and *Salmonella* : cellular and molecular biology. American Society for Microbiology, Washington, D.C.
- Hancke, K., T. B. Hancke, L. M. Olsen, G. Johnsen, et R. N. Glud. 2008. Temperature effects on microalgal photosynthesis-light responses measured by O<sub>2</sub> production pulse-amplitude-modulated fluorescence and <sup>14</sup>C assimilation. *Journal of Phycology* **44** : 501-514.
- Handler, A. M., et S. D. McCombs. 2000. The *piggyBac* transposon mediates germ-line transformation in the oriental fruit fly and closely related elements exist in its genome. *Insect Molecular Biology* **9** (6) : 605-612.
- Harrison, P. J., R. E. Waters, et F. J. R. Taylor. 1980. A broad spectrum artificial sea water medium for coastal and open ocean phytoplankton. *Journal of Phycology* **16** : 28-35.
- Hartl, D. L., A. R. Lohe, et E. R. Lozovskaya. 1997a. Modern thoughts on an ancient mariner : function, evolution, regulation. *Annual Review of Genetics* **31** : 337-358.
- Hartl, D. L., E. R. Lozovskaya, D. I. Nurminsky, et A. R. Lohe. 1997b. What restricts the activity of *mariner*-like transposable elements? *Trends in Genetics* **13** : 197-201.
- Harwood, D. M., et R. Gersonde. 1990. Lower Cretaceous diatom from ODP Leg 113 site 693 (Weddell Sea), part 2, resting spores, Chrysophycean cysts, and endoskeletal dinoflagellates and notes on the origin of diatoms. In. *Proceeding Ocean Drilling Program of Sciences Results*. P. F. Barker and J. P. Kennett (Eds), pp 403-426.
- Hashida, S.-N., T. Uchiyama, C. Martin, Y. Kishima, Y. Sano, et T. Mikami. 2006. The temperature-dependent change in methylation of the *Antirrhinum* transposon *Tam3* is controlled by the activity of its transposase. *Plant Cell* **18** : 104-118.
- Hashida, S.-N., K. Kitamura, T. Mikami, et Y. Kishima. 2003. Temperature shift coordinately changes the activity and the methylation state of transposon *Tam3* in *Antirrhinum majus*. *Plant Physiology* **132** (3) : 1207-1216.
- Heckathorn, S. A., C. A. Downs, T. D. Sharkey, et J. S. Coleman. 1998. The small, methionine-rich chloroplast heat-shock protein protects photosystem II electron transport during heat stress. *Plant physiology* **116** : 439-444.
- Henikoff, S. 1992. Detection of *Caenorhabditis* transposon homologs in diverse organisms. *The New Biologist* **4** (4) : 382-388.

- Hey, P., G. Robson, M. Birch, et M. BroMLEy. 2008. Characterisation of *Aft1* a *Fot1/Pogo* type transposon of *Aspergillus fumigatus*. *Fungal Genetics and Biology* **45** (2) : 117-126.
- Hildebrand, M., D. K. Corey, J. R. Ludwig, A. Kukel, T. Y. Feng, et B. E. Volcani. 1991. Plasmids in diatom species. *Journal of Bacteriology* **173** (18) : 5924-5927.
- Hirochika, H., H. Okamoto, et T. Kakutani. 2000. Silencing of retrotransposons in *Arabidopsis* and reactivation by the *ddm1* mutation. *Plant Cell* **12** (3) : 357-369.
- Hirochika, H., K. Sugimoto, Y. Otsuki, H. Tsugawa, et M. Kanda. 1996. Retrotransposons of rice involved in mutations induced by tissue culture. *Proceedings of the National Academy of Sciences* **93** (23) : 7783 - 7788.
- Hoffmann, A. A., et P. A. Parsons. 1994. *Evolutionary genetics and environmental stress*. Oxford University Press, Oxford.
- Holder, M., et P. O. Lewis. 2003. Phylogeny estimation : traditional and Bayesian approaches. *Nature Reviews Genetics* **4** (4) : 275-284.
- Holmberg, N., et L. Bülow. 1998. Improving stress tolerance in plants by gene transfer. *Trends in Plant Science* **3** (2) : 61-66.
- Hood, M. E. 2005. Repetitive DNA in the automicetic fungus *Microbotryum violaceum*. *Genetica* **124** : 1-10.
- Hu, S.-T., J.-H. Hwang, L.-C. Lee, C.-H. Lee, P.-L. Li, et Y.-C. Hsieh. 1994. Functional analysis of the 14 kDa protein of insertion sequence 2. *Journal of Molecular Biology* **236** (2) : 503-513.
- Hua-Van, A., A. Le Rouzic, C. Maisonhaute, et P. Capy. 2005. Abundance, distribution and dynamics of retrotransposable elements and transposons : similarities and differences. *Cytogenetic and Genome Research* **110** (1-4) : 426-440.
- Hua-Van, A. I., J.-M. Davière, F. Kaper, T. Langin, et M.-J. Daboussi. 2000. Genome organization in *Fusarium oxysporum* : clusters of class II transposons. *Current Genetics* **37** (5) : 339-347.
- II
- Ikeda, K., H. Nakayashiki, M. Takagi, Y. Tosa, et S. Mayama. 2001. Heat shock, copper sulfate and oxidative stress activate the retrotransposon *MAGGY* resident in the plant pathogenic fungus *Magnaporthe grisea*. *Molecular Genetics and Genomics* **266** (2) : 318-325.
- Ilves, H., R. Horak, et M. Kivisaar. 2001. Involvement of sigma s in starvation-induced transposition of *Pseudomonas putida* transposon *Tn4652*. *Journal of Bacteriology* **183** (18) : 5445-5448.
- Ivashuta, S., M. Naumkina, M. Gau, K. Uchiyama, S. Isobe, Y. Mizukami, et Y. Shimamoto. 2002. Genotype-dependent transcriptional activation of novel repetitive elements during cold acclimation of alfalfa (*Medicago sativa*). *The Plant Journal* **31** (5) : 615-627.
- Ivics, Z., P. B. Hackett, R. H. Plasterk, et Z. Izsvák. 1997. Molecular reconstruction of *Sleeping Beauty*, a *Tc1*-like transposon from fish, and its transposition in human cells. *Cell* **91** (4) : 501-510.

- Ivics, Z., Z. Izsvák, A. Minter, et P. B. Hackett. 1996. Identification of functional domains and evolution of *Tc1*-like transposable elements. *Proceedings of the National Academy of Sciences* **93** (10) : 5008-5013.
- Izsvák, Z., D. Khare, J. Behlke, U. Heinemann, R. H. Plasterk, et Z. Ivics. 2002. Involvement of a bifunctional, paired-like DNA-binding domain and a transpositional enhancer in *Sleeping Beauty* transposition. *Journal of Biological Chemistry* **277** : 34581-34588.
- Izsvák, Z., Z. Ivics, N. Shimoda, D. Mohn, H. Okamoto, et P. B. Hackett. 1999. Short inverted-repeat transposable elements in teleost fish and implications for a mechanism of their amplification. *Journal of Molecular Evolution* **48** : 13-21.

## J

- Jacobs, G., D. Dechyeva, G. Menzel, C. Dombrowski, et T. Schmidt. 2004. Molecular characterization of *Vulmar1*, a complete *mariner* transposon of sugar beet and diversity of *mariner*- and *En/Spm*-like sequences in the genus *Beta*. *Genome* **47** : 1192-1201.
- Jacobson, J. W., M. M. Medhora, et D. L. Hartl. 1986. Molecular structure of a somatically unstable transposable element in *Drosophila*. *Proceedings of the National Academy of Sciences* **83** (22) : 8684-8688.
- Jarvik, T., et K. G. Lark. 1998. Characterization of *Soymar1*, a *mariner* element in soybean. *Genetics* **149** (3) : 1569-1574.
- Jeyaprakash, A., et M. A. Hoy. 1995. Complete sequence of a *mariner* transposable element from the predatory mite *Metaseiulus occidentalis* isolated by an inverse PCR approach. *Insect Molecular Biology* **4** : 31-39.
- Jiang, N., C. Feschotte, X. Zhang, et S. R. Wessler. 2004. Using rice to understand the origin and amplification of miniature inverted repeat transposable elements (MITEs). *Current Opinion in Plant Biology* **7** (2) : 115-119.
- Jiang, H., et K. Gao. 2004. Effects of lowering temperature during culture on the production of polyunsaturated fatty acids in the marine diatom *Phaeodactylum tricorutum* (Bacillariophyceae). *Journal of Phycology* **40** (4) : 651-654.
- Jiang, S. C., et J. H. Paul. 1998. Gene transfer by transduction in the marine environment. *Applied and Environmental Microbiology* **64** (8) : 2780-2787.
- Jordan, M., et S. Russell. 1999. Categorization. *In The MIT Encyclopedia of the Cognitive Sciences*. The MIT press, Cambridge, Massachusetts. pp 104-106.
- Jurka, J., V. Kapitonov, A. Pavlicek, P. Klonowski, O. Kohany, et J. Walichiewicz. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research* **110** : 462 - 467.
- Jurka, J., et V. V. Kapitonov. 2001. *PIFs* meet *Tourists* and *Harbingers* : a superfamily reunion. *Proceedings of the National Academy of Sciences* **98** (22) : 12315-12316.

## K

- Kalendar, R., A. J. Flavell, T. H. N. Ellis, T. Sjakste, C. Moisy, et A. H. Schulman. 2010. Analysis of plant diversity with retrotransposon-based molecular markers. *Heredity* **in press** : 1-11.
- Kalendar, R., J. Tanskanen, S. Immonen, E. Nevo, et A. H. Schulman. 2000. Genome evolution of wild barley (*Hordeum spontaneum*) by *BARE-1* retrotransposon dynamics in response to sharp microclimatic divergence. *Proceedings of the National Academy of Sciences* **97 (12)** : 6603-6607.
- Kapitonov, V. V., et J. Jurka. 2008. A universal classification of eukaryotic transposable elements implemented in Repbase. *Nature* **9 (5)** : 411-412.
- Kapitonov, V. V., et J. Jurka. 2006. Self-synthesizing DNA transposons in eukaryotes. *Proceedings of the National Academy of Sciences* **103 (12)** : 4540-4545.
- Kapitonov, V. V., et J. Jurka. 2005. RAG1 core and V(D)J recombination signal sequences were derived from *Transib* transposons. *PLoS Biol* **3 (6)** : e181.
- Kapitonov, V. V., et J. Jurka. 2003. Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. *Proceedings of the National Academy of Sciences* **100** : 6569-6574.
- Kapitonov, V. V., et J. Jurka. 2001. Rolling-circle transposons in eukaryotes. *Proceedings of the National Academy of Sciences* **98 (15)** : 8714-8719.
- Kazazian, H. H., et J. V. Moran. 1998. The impact of *L1* retrotransposons on the human genome. *Nature genetics* **19** : 19-24.
- Kempken, F., et F. Windhofer. 2001. The *hAT* family : a versatile transposon group common to plants, fungi, animals, and man. *Chromosoma* **110** : 1-9.
- Ken, C.-F., T. M. Hsiung, Z.-X. Huang, R. H. Juang, et C.-T. Lin. 2005. Characterization of Fe/Mn superoxide dismutase from diatom *Thalassiosira weissflogii* : cloning, expression, and property. *Journal of Agricultural and Food Chemistry* **53 (5)** : 1470-1474.
- Kentner, E. K., M. L. Arnold, et S. R. Wessler. 2003. Characterization of high-copy-number retrotransposons from the large genomes of the Louisiana iris species and their use as molecular markers. *Genetics* **164 (2)** : 685-697.
- Kidwell, M. G. 2002. Transposable elements and the evolution of genome size in eukaryotes. *Genetica* **115** : 49-63.
- Kidwell, M. G., et D. R. Lisch. 2001. Perspectives : transposable elements, parasitic DNA, and genome. *Evolution* **55** : 1-24.
- Kidwell, M. G., et D. R. Lisch. 2000. Transposable elements and host genome evolution. *Trends in Ecology and Evolution* **15 (3)** : 95-99.

- Kidwell, M. G., et D. Lisch. 1997. Transposable elements as sources of variation in animals and plants. *Proceedings of the National Academy of Sciences* **94** (15) : 7704-7711.
- Kidwell, M. G. 1992. Horizontal transfer. *Current Opinion in Genetics and Development* **2** (6) : 868-873.
- Kimura, R. H., P. V. Choudary, et C. W. Schmid. 1999. Silk worm *Bm1* SINE RNA increases following cellular insults. *Nucleic Acids Research* **27** (16) : 3380-3387.
- Kimura, R. H., P. V. Choudary, K. K. Stone, et C. W. Schmid. 2001a. Stress induction of *Bm1* RNA in silkworm larvae : SINEs, an unusual class of stress genes. *Cell stress Chaperone* **6** : 263-272.
- Kimura, Y., Y. Tosa, S. Shimada, R. Sogo, M. Kusaba, T. Sunaga, S. Betsuyaku, Y. Eto, H. Nakayashiki, et S. Mayama. 2001b. *OARE-1*, a *Ty1-copia* retrotransposon in oat activated by abiotic and biotic stresses. *Plant and Cell Physiology* **42** (12) : 1345-1354.
- Kitamura, K., S. Hashida, T. Mikami, et Y. Kishima. 2001. Position effect of the excision frequency of the *Antirrhinum* transposon *Tam3* : implications for the degree of position-dependent methylation in the ends of the element. *Plant Molecular Biology* **47** (4) : 475-490.
- Koga, A., M. Suzuki, H. Inagaki, Y. Bessho, et H. Hori. 1996. Transposable element in fish. *Nature* **383** (6595) : 30.
- Kohany, O., A. Gentles, L. Hankus, et J. Jurka. 2006. Annotation, submission and screening of repetitive elements in Repbase : RepbaseSubmitter and Censor. *BMC Bioinformatics* **7** : 474.
- Koonin, E. V., et Y. I. Wolf. 2008. Genomics of bacteria and archaea : the emerging dynamic view of the prokaryotic world. *Nucleic Acids Research* **36** (21) : 6688-6719.
- Kotak, S., J. Larkindale, U. Lee, P. von Koskull-Döring, E. Vierling, et K.-D. Scharf. 2007. Complexity of the heat stress response in plants. *Current Opinion in Plant Biology* **10** (3) : 310-316.
- Kramerov, D. A., et N. S. Vassetzky. 2005. Short retroposons in eukaryotic genomes international review of cytology. *International Review of Cytology* **247** : 165-221.
- Kretschmer, P. J., et S. N. Cohen. 1979. Effect of temperature on translocation frequency of the *Tn3* element. *Journal of Bacteriology* **139** (2) : 515-519.
- Krishnan, M., H. T. Nguyen, et J. J. Burke. 1989. Heat shock protein synthesis and thermal tolerance in wheat. *Plant physiology* **90** : 140-145.
- Kudo, I., M. Miyamoto, Y. Noiri, et Y. Maita. 2000. Combined effects of temperature and iron on the growth and physiology of the marine diatom *Phaeodactylum tricornutum* (Bacillariophyceae). *Journal of Phycology* **36** (6) : 1096-1102.
- Kumar, A., et J. L. Bennetzen. 1999. Plant Retrotransposons. *Annual Review of Genetics* **33** : 479-532.
- Kunze, R., et C. F. Weil. 2002. The hAT and CACTA superfamily of plant transposons. *In : Mobile DNA II*, N. L. Craig, R. Craigie, M. Gellert, et A. M. Lambowitz (Eds). American Society of Microbiology Press, Washington, pp. 565-610.

## L

- Lai, Y.-K., C.-W. Li, C.-H. Hu, et M.-L. Lee. 1988. Quantitative and qualitative analyses of protein synthesis during heat shock in the marine diatom *Nitzschia alba* (Bacillariophyceae). *Journal of Phycology* **24** (4) : 509-514.
- Lampe, D. J., K. K. O. Walden, et H. M. Robertson. 2001. Loss of transposase-DNA Interaction may underlie the divergence of *mariner* family transposable elements and the ability of more than one mariner to occupy the same genome. *Molecular Biology and Evolution* **18** (6) : 954-961.
- Lampe, D. J., T. E. Grant, et H. M. Robertson. 1998. Factors affecting transposition of the *Himar1* mariner transposon *in vitro*. *Genetics* **149** : 179-187.
- Lampe, D. J., M. E. A. Churchill, et H. M. Robertson. 1997. A purified mariner transposase is sufficient to mediate transposition *in vitro*. *The EMBO journal* **16** (13) : 4153.
- Lander. 2001. Initial sequencing and analysis of the human genome. *Nature* **409** (6822) : 860-921.
- Langin, T., P. Capy, et M.-J. Daboussi. 1995. The transposable element *impala* a fungal member of *Tc1-mariner* superfamily. *Molecular and General Genetics* **246** : 19-28.
- Le Rouzic, A., et P. Capy. 2005. The first steps of transposable elements invasion : parasitic strategy vs. genetic drift. *Genetics* **169** (2) : 1033 - 1043.
- Leaver, M. J. 2001. A family of *Tc1*-like transposons from the genomes of fishes and frogs : evidence for horizontal transmission. *Gene* **271** (2) : 203-214.
- Lebeaux, M. O., et M. Jambu. 1978. *Classification automatique pour l'analyse des données. Tome I : Méthodes et algorithmes*. Dunod, Paris.
- Leblanc, C., A. Falciatore, M. Watanabe, et C. Bowler. 1999. Semi-quantitative RT-PCR analysis of photoregulated gene expression in marine diatoms. *Plant Molecular Biology* **40** (6) : 1031-1044.
- Lefebvre, S., J. L. Mouget, P. Loret, P. Rosa, et G. Tremblin. 2007. Comparison between fluorimetry and oximetry techniques to measure photosynthesis in the diatom *Skeletonema costatum* cultivated under simulated seasonal conditions. *Journal of Photochemistry and Photobiology* **86** (2) : 131-139.
- Lerat, E., et P. Capy. 1999. Retrotransposons and retroviruses : analysis of the envelope gene. *Molecular Biology and Evolution* **16** (9) : 1198-1207.
- Leroy, H., P. Castagnone-Sereno, S. Renault, C. Augé-Gouillou, Y. Bigot, et P. Abad. 2003. Characterization of *Mymar1*, a *mariner*-like element with large inverted terminal repeats (ITRs) from the phytoparasitic nematode *Meloidogyne chitwoodi*. *Gene* **304** : 35-41.
- Li, T.-H., et C. W. Schmid. 2001. Differential stress induction of individual *Alu* loci : implications for transcription and retrotransposition. *Gene* **276** (1-2) : 135-141.

- Li, T.-H., J. Spearow, C. M. Rubin, et C. W. Schmid. 1999. Physiological stresses increase mouse short interspersed element (SINE) RNA expression *in vivo*. *Gene* **239** (2) : 367-372.
- Libault, M., S. Thibivilliers, D. D. Bilgin, O. Radwan, M. Benitez, S. J. Clough, et G. Stacey. 2008. Identification of four soybean reference genes for gene expression normalization. *The Plant Genome* **1** : 44-54.
- Lichtenthaler H. K. 1998. The stress concept in plants : an introduction. *Annals of the New York Academy of Sciences* **851** : 187-198.
- Lidholm, D. A., A. R. Lohe, et D. L. Hartl. 1993. The transposable element *mariner* mediates germline transformation in *Drosophila melanogaster*. *Genetics* **134** (3) : 859-868.
- Lin, C. Y., J. K. Roberts, et J. L. Key. 1984. Acquisition of thermotolerance in soybean seedlings. *Plant physiology* **74** : 152-160.
- Lindquist, S., et E. A. Craig. 1988. The heat-shock proteins. *Annual Review of Genetics* **22** : 631-677.
- Lisch, D. 2002. *Mutator* transposons. *Trends in Plant Science* **7** (11) : 498-504.
- Lisch, D. R., M. Freeling, R. J. Langham, et M. Y. Choy. 2001. *Mutator* transposase is widespread in the grasses. *Plant Physiology* **125** (3) : 1293-1303.
- Lister, C., D. Jackson, et C. Martin. 1993. Transposon-induced inversion in *Antirrhinum* modifies *nivea* gene expression to give a novel flower color pattern under the control of *cycloidearadialis*. *Plant Cell* **5** (11) : 1541-1553.
- Liu, N., J. W. Pridgeon, H. Wang, Z. Liu, et L. Zhang. 2004. Identification of *mariner* elements from house flies (*Musca domestica*) and german cockroaches (*Blattella germanica*). *Insect Molecular Biology* **13** (4) : 443-447.
- Llorens, C., M. Fares, et A. Moya. 2008. Relationships of *gag-pol* diversity between *Ty3/Gypsy* and retroviridae LTR retroelements and the three kings hypothesis. *BMC Evolutionary Biology* **8** : 276.
- Lobban, C. S., P. J. Harrison, et M. J. Duncan. 1985. *The physiological ecology of seaweeds*. Cambridge University Press, Cambridge.
- Lohe, A., D. DeAguiar, et D. Hartl. 1997. Mutations in the *mariner* transposase : The D,D(35)E consensus sequence is nonfunctional. *Proceedings of the National Academy of Sciences* **94** (4) : 1293-1297.
- Lohe, A. R., et D. L. Hartl. 1996. Autoregulation of *mariner* transposase activity by overproduction and dominant-negative complementation. *Molecular Biology and Evolution* **13** (4) : 549-555.
- Lohe, A. R., E. N. Moriyama, D. A. Lidholm, et D. L. Hartl. 1995. Horizontal transmission, vertical inactivation, and stochastic loss of *mariner*-like transposable elements. *Molecular Biology and Evolution* **12** : 62-72.

M

- Macas, J., A. Koblizkova, et P. Neumann. 2005. Characterization of *Stowaway* MITEs in pea (*Pisum sativum* L.) and identification of their potential master elements. *Genome* **48** : 831-839.
- Machida, C., et Y. Machida. 1989. Regulation of *ISI* transposition by the *insA* gene product. *Journal of Molecular Biology* **208** (4) : 567-574.
- Madlung, A., et L. Comai. 2004. The effect of stress on genome regulation and structure. *Annals of Botany* **94** (4) : 481-495.
- Madsen, L. H., E. Fukai, S. Radutoiu, C. K. Yost, N. Sandal, L. Schauser, et J. Stougaard. 2005. *LORE1*, an active low-copy-number *TY3-gypsy* retrotransposon family in the model legume *Lotus japonicus*. *The Plant Journal* **44** (3) : 372-381.
- Mahajan, S., et N. Tuteja. 2005. Cold, salinity and drought stresses : an overview. *Archives of Biochemistry and Biophysics* **444** (2) : 139-158.
- Maheswari, U., T. Mock, E. V. Armbrust, et C. Bowler. 2009. Update of the Diatom EST Database : a new tool for digital transcriptomics. *Nucleic Acids Research* **37**(suppl 1) : D1001-D1005.
- Maheswari, U., A. Montsant, J. Goll, S. Krishnasamy, K. R. Rajyashri, V. M. Patell, et C. Bowler. 2005. The Diatom EST Database. *Nucleic Acids Research* **33** (suppl 1) : D344-D347.
- Mann, J. E., et J. Myers. 1968. Photosynthetic enhancement in the diatom *Phaeodactylum tricorutum*. *Plant physiology* **43** (12) : 1991-1995.
- Mannan, R. M., M. Krishnan, et A. Gnanam. 1986. Heat shock proteins of *Cyanobacterium anacystis nidulans*. *Plant and Cell Physiology* **27** (2) : 377-381.
- Margalef, R. 1978. Life-forms of phytoplankton as survival alternatives in unstable environment. *Oceanologica Acta* **1** : 493-509.
- Martienssen, R., et A. Baron. 1994. Coordinate suppression of mutations caused by Robertson's *Mutator* transposons in maize. *Genetics* **136** (3) : 1157-1170.
- Maruyama, K., et D. L. Hartl. 1991a. Evidence for interspecific transfer of the transposable element *mariner* between *Drosophila* and *Zaprionus*. *Journal of Molecular Evolution* **33** (6) : 514-524.
- Maruyama, K., et D. L. Hartl. 1991b. Evolution of the transposable element *mariner* in *Drosophila* species. *Genetics* **128** (2) : 319-329.
- Marzo, M., M. Puig, et A. Ruiz. 2008. The foldback-like element *Galileo* belongs to the P superfamily of DNA transposons and is widespread within the *Drosophila* genus. *Proceedings of the National Academy of Sciences* **105** (8) : 2957-2962.
- Matès, L., Z. Izsvák, et Z. Ivics. 2007. Technology transfer from worms and flies to vertebrates : transposition-based genome manipulations and their future perspectives. *Genome Biology* **8** (suppl1).
- Maumus, F., A. Allen, C. Mhiri, H. Hu, K. Jabbari, A. Vardi, M.-A. Grandbastien, et C. Bowler. 2009. Potential impact of stress activated retrotransposons on genome evolution in a marine diatom. *BMC Genomics* **10** : 624.

- Maxwell, K., et G. N. Johnson. 2000. Chlorophyll fluorescence : a practical guide. *Journal of Experimental Botany* **51 (345)** : 659-668.
- McClintock, B. 1984. The significance of responses of the genome to challenge. *Science* **226 (4676)** : 792-801.
- McClintock, B. 1953. Induction of instability at selected loci in maize. *Genetics* **38 (6)** : 579-599.
- McClure, M. A. 1991. Evolution of retroposons by acquisition or deletion of retrovirus-like genes. *Molecular Biology and Evolution* **8 (6)** : 835-856.
- McDonald, J. F. 1995. Transposable elements : possible catalysts of organismic evolution. *Trends in Ecology and Evolution* **10 (3)** : 123-126.
- McGinnis, W., A. W. Shermoen, et S. K. Beckendorf. 1983. A transposable element inserted just 5' to a *Drosophila* glue protein gene alters gene expression and chromatin structure. *Cell* **34** : 75-84.
- Medlin, L. K., et I. Kaczmarek. 2004. Evolution of the diatoms : V. Morphological and cytological support for the major clades and a taxonomic revision. *Phycologia* **43 (3)** : 245-270.
- Medlin, L. K., D. M. Williams, et P. A. Slims. 1993. The evolutions of the diatom (Bacillariophyta). I. Origin of the group and assessment of the monophyly of its major divisions. *European Journal of Phycology* **28** : 261-275.
- Melayah, D., E. Bonnivard, B. Chalhoub, C. Audeon, et M.-A. Grandbastien. 2001. The mobility of the tobacco *Tnt1* retrotransposon correlates with its transcriptional activation by fungal factors. *The Plant Journal* **28 (2)** : 159-168.
- Mendiola, M. V., et F. De la Cruz. 1992. *IS91* transposase is related to the rolling-circle-type replication proteins of the pUB110 family of plasmids. *Nucleic Acids Research* **20 (13)** : 3521.
- Merchant, S. S., S. E. Prochnik, O. Vallon, E. H. Harris, S. J. Karpowicz, G. B. Witman, A. Terry, A. Salamov, L. K. Fritz-Laylin, L. Maréchal-Drouard, W. F. Marshall, L.-H. Qu, D. R. Nelson, A. A. Sanderfoot, M. H. Spalding, V. V. Kapitonov, Q. Ren, P. Ferris, E. Lindquist, H. Shapiro, S. M. Lucas, J. Grimwood, J. Schmutz, P. Cardol, H. Cerutti, G. Chanfreau, C.-L. Chen, V. Cognat, M. T. Croft, R. Dent, S. Dutcher, E. Fernández, H. Fukuzawa, D. González-Ballester, D. González-Halphen, A. Hallmann, M. Hanikenne, M. Hippler, W. Inwood, K. Jabbari, M. Kalanon, R. Kuras, P. A. Lefebvre, S. p. D. Lemaire, A. V. Lobanov, M. Lohr, A. Manuell, I. Meier, L. Mets, M. Mittag, T. Mittelmeier, J. V. Moroney, J. Moseley, C. Napoli, A. M. Nedelcu, K. Niyogi, S. V. Novoselov, I. T. Paulsen, G. Pazour, S. Purton, J.-P. Ral, D. M. Riano-Pachon, W. Riekhof, L. Rymarquis, M. Schroda, D. Stern, J. Umen, R. Willows, N. Wilson, S. L. Zimmer, J. Allmer, J. Balk, K. Bisova, C.-J. Chen, M. Elias, K. Gendler, C. Hauser, M. R. Lamb, H. Ledford, J. C. Long, J. Minagawa, M. D. Page, J. Pan, W. Pootakham, S. Roje, A. Rose, E. Stahlberg, A. M. Terauchi, P. Yang, S. Ball, C. Bowler, C. L. Dieckmann, V. N. Gladyshev, P. Green, R. Jorgensen, S. Mayfield, B. Mueller-Roeber, S.

- Rajamani, R. T. Sayre, P. Brokstein, I. Dubchak, D. Goodstein, L. Hornick, Y. W. Huang, J. Jhaveri, Y. Luo, D. Martinez, W. C. A. Ngau, B. Otilar, A. Poliakov, A. Porter, L. Szajkowski, G. Werner, K. Zhou, I. V. Grigoriev, D. S. Rokhsar, et A. R. Grossman. 2007. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* **318** (5848) : 245-250.
- Mhiri, C., J.-B. Morel, S. Vernhettes, J. M. Casacuberta, H. Lucas, et M.-A. Grandbastien. 1997. The promoter of the tobacco *Tnt1* retrotransposon is induced by wounding and by abiotic stress. *Plant Molecular Biology* **33** (2) : 257-266.
- Mirón, A. S., M. C. C. García, A. C. Gómez, F. G. Camacho, E. M. Grima, et Y. Chisti. 2003. Shear stress tolerance and biochemical characterization of *Phaeodactylum tricorutum* in quasi steady-state continuous culture in outdoor photobioreactors. *Biochemical Engineering Journal* **16** (3) : 287-297.
- Miskey, C., B. Papp, L. Matés, L. Sinzelle, H. Keller, Z. Izsvák, et Z. Ivics. 2007. The ancient *mariner* sails again : transposition of the human *Hsmar1* element by a reconstructed transposase and activities of the *SETMAR* protein on transposon ends. *Molecular and Cellular Biology* **27** (12) : 4589-4600.
- Miskey, C., Z. Izsvák, K. Kawakami, et Z. Ivics. 2005. DNA transposons in vertebrate functional genomics. *Cellular and Molecular Life Sciences* **62** (6) : 629-641.
- Miskey, C., Z. Izsvák, R. H. Plasterk, et Z. Ivics. 2003. The *Frog Prince* : a reconstructed transposon from *Rana pipiens* with high transpositional activity in vertebrate cells. *Nucleic Acids Research* **31** (23) : 6873-6881.
- Mittler, R. 2002. Oxidative stress, antioxidants and stress tolerance. *Trends in Plant Science* **7** : 405-410.
- Mizuuchi, K. 1997. Polynucleotidyl transfer reactions in site-specific DNA recombination. *Genes to Cells* **2** : 1-12.
- Mock, T., et K. Valentin. 2004. Photosynthesis and cold acclimation : molecular evidence from a polar diatom. *Journal of Phycology* **40** : 732-741.
- Molino, P. J., O. M. Hodson, J. F. Quinn, et R. Wetherbee. 2008. The quartz crystal microbalance : a new tool for the investigation of the bioadhesion of diatoms to surfaces of differing surface energies. *Langmuir* **24** (13) : 6730-6737.
- Moreau-Mhiri, C., J.-B. Morel, C. Audeon, M. Ferault, M.-A. Grandbastien, et H. Lucas. 1996. Regulation of expression of the tobacco *Tnt1* retrotransposon in heterologous species following pathogen-related stresses. *The Plant Journal* **9** (3) : 409-419.
- Morgan-Kiss, R. M., J. C. Priscu, T. Pockock, L. Gudynaite-Savitch, et N. P. A. Huner. 2006. Adaptation and acclimation of photosynthetic microorganisms to permanently cold environments. *Microbiology and Molecular Biology Reviews* **70** : 222-252.

- Mouget, J. L., G. Gastineau, O. Davidovich, P. Gaudin, et N. A. Davidovich. 2009. Light is a key factor in triggering sexual reproduction in the pennate diatom *Haslea ostrearia*. *FEMS Microbiology Ecology* **69** (2) : 194-201.
- Mouget, J.-L., et G. Tremblin. 2002. Suitability of the Fluorescence Monitoring System (FMS, Hansatech) for measurement of photosynthetic characteristics in algae. *Aquatic Botany* **74** : 219-231.
- Muller-Feuga, A. 1997. *Microalgues marines : les enjeux de la recherche*. Ifremer (Ed).

## N

- Naito, K., E. Cho, G. Yang, M. A. Campbell, K. Yano, Y. Okumoto, T. Tanisaka, et S. R. Wessler. 2006. Dramatic amplification of a rice transposable element during recent domestication. *Proceedings of the National Academy of Sciences* **103** (47) : 17620-17625.
- Nakamoto, H., N. Suzuki, et S. K. Roy. 2000. Constitutive expression of a small heat-shock protein confers cellular thermotolerance and thermal protection to the photosynthetic apparatus in cyanobacteria. *FEBS Letters* **483** (2-3) : 169-174.
- Nguyen-Deroche, T. L. N., H. T. LE, B. Trang Viet, Y. Rince, G. Tremblin, et A. Morant-Manceau. 2009. Effects of copper on growth and photosynthesis in marine diatoms : a comparison between species from two different geographical areas. *Cryptogamie. Algologie* **30** (2) : 13.
- Nicholas, K., H. Nicholas, et D. Deerfield. 1997. GeneDoc : analysis and visualization of genetic variation. *Embnet News* **4** : 14.
- Nicot, N., J.-F. Hausman, L. Hoffmann, et D. Evers. 2005. Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. *Journal of Experimental Botany* **56** (421) : 2907-2914.
- Niyogi, K. K. 1999. Photoprotection revisited : genetic and molecular approaches. *Annual Review of Plant Physiology and Plant Molecular Biology* **50** : 333-359.
- Nuzhdin, S. V., et D. A. Petrov. 2003. Transposable elements in clonal lineages : lethal hangover from sex. *Biological Journal of the Linnean Society* **79** : 33-41.
- Nymark, M., K. C. Valle, T. Brembu, K. Hancke, P. Winge, K. Andresen, G. Johnsen, et A. M. Bones. 2009. An integrated analysis of molecular acclimation to high light in the marine diatom *Phaeodactylum tricornutum*. *PLoS ONE* **4** (11) : e7743.

## O

- Ogasawara, H., H. Obata, Y. Hata, S. Takahashi, et K. Gomi. 2009. Crawler, a novel *Tc1/mariner*-type transposable element in *Aspergillus oryzae* transposes under stress conditions. *Fungal Genetics and Biology* **46** (6-7) : 441-449.

- O'Hare, K., et G. M. Rubin. 1983. Structures of *P* transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. *Cell* **34** : 25-35.
- Okada, N., M. Hamada, I. Ogiwara, et K. Ohshima. 1997. SINEs and LINEs share common 3' sequences : a review. *Gene* **205 (1-2)** : 229-243.
- Oudot-Le Secq, M.-P., J. Grimwood, H. Shapiro, E. Armbrust, C. Bowler, et B. Green. 2007. Chloroplast genomes of the diatoms *Phaeodactylum tricorutum* and *Thalassiosira pseudonana* comparison with other plastid genomes of the red lineage. *Molecular Genetics and Genomics* **277 (4)** : 427-439.
- Ouyang, Y., J. Chen, W. Xie, L. Wang, et Q. Zhang. 2009. Comprehensive sequence and expression profile analysis of Hsp20 gene family in rice. *Plant Molecular Biology* **70 (3)** : 341-357.

## P

- Pace, J. K., C. Gilbert, M. S. Clark, et C. Feschotte. 2008. Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. *Proceedings of the National Academy of Sciences* **105 (44)** : 17023-17028.
- Palomeque, T., J. Antonio Carrillo, M. Munoz-Lopez, et P. Lorite. 2006. Detection of a *mariner*-like element and a miniature inverted-repeat transposable element (MITE) associated with the heterochromatin from ants of the genus *Messor* and their possible involvement for satellite DNA evolution. *Gene* **371 (2)** : 194-205.
- Paolacci, A., O. Tanzarella, E. Porceddu, et M. Ciaffi. 2009. Identification and validation of reference genes for quantitative RT-PCR normalization in wheat. *BMC Molecular Biology* **10** : 11.
- Papageorgiou, G. C., et Govindjee. 2004. *Chlorophyll fluorescence : a signature of photosynthesis*. Kluwer Academic Publishers, Dordrecht.
- Paquin, C. E., et V. M. Williamson. 1988. Effect of temperature on Ty transposition. *In : Eukaryotic transposable elements as mutagenic agents*, M. E. Lambert, J. F. McDonald, et I. P. Weinstein (Eds). Cold Spring Harbor Laboratory Press, New York, pp. 235-244.
- Patrick, R. 1977. Ecology of freshwater diatom and diatom communities. *In : The biology of diatoms*. Blackwell scientific, Oxford, pp. 284-332.
- Pavlopoulos, A., et M. Averof. 2005. Establishing genetic transformation for comparative developmental studies in the crustacean *Parhyale hawaiiensis*. *Proceedings of the National Academy of Sciences* **102 (22)** : 7888-7893.
- Pearson, G., G. Hoarau, A. Lago-Leston, J. Coyer, M. Kube, R. Reinhardt, K. Henckel, E. Serrao, E. Corre, et J. Olsen. 2010. An expressed sequence tag analysis of the intertidal brown seaweeds *Fucus serratus* (L.) and *F. vesiculosus* (L.) (Heterokontophyta, *Phaeophyceae*) in response to abiotic stressors. *Marine Biotechnology* **12** : 195-213.
- Peers, G., et N. M. Price. 2004. A role for manganese in superoxide dismutases and growth of iron-deficient diatoms. *American Society of Limnology and Oceanography* **49 (5)** : 10.

- Peltzer, D., E. Dreyer, et A. Polle. 2002. Differential temperature dependencies of antioxidative enzymes in two contrasting species : *Fagus sylvatica* and *Coleus blumei*. *Plant Physiology and Biochemistry* **40** (2) : 141-150.
- Perelman, A., Z. Dubinsky, et R. Martinez. 2006. Temperature dependence of superoxide dismutase activity in plankton. *Journal of Experimental Marine Biology and Ecology* **334** (2) : 229-235.
- Peschek, G., et R. Zoder. 2001. Temperature stress and basic bioenergetic strategies for stress defence. *In : Algal adaptation to environmental stresses*, L. C. Rai et J. P. Gaur (Eds). Verlag Springer, New York, pp. 203-258.
- Petit, A., F. Rouleux-Bonnin, M. Lambele, N. Pollet, et Y. Bigot. 2007. Properties of the various *Botmar1* transcripts in imagoes of the bumble bee, *Bombus terrestris* (Hymenoptera : Apidae). *Gene* **390** (1-2) : 52-66.
- Pfeifer, F., et U. Blaseio. 1990. Transposition burst of the *ISH27* insertion element family in *Halobacterium halobium*. *Nucleic Acids Research* **18** (23) : 6921-6925.
- Piednoel, M., et E. Bonnivard. 2009. *DIRS1*-like retrotransposons are widely distributed among Decapoda and are particularly present in hydrothermal vent organisms. *BMC Evolutionary Biology* **9** : 86.
- Petrokovski, S., et S. Henikoff. 1997. A helix-turn-helix DNA-binding motif predicted for transposases of DNA transposons. *Molecular and General Genetics* **254** (6) : 689-695.
- Plasterk, R. H. A., Z. Izsvák, et Z. Ivics. 1999. Resident aliens : the *Tc1-mariner* superfamily of transposable elements. *Trends in Genetics* **15** (8) : 326-332.
- Pocwierz-Kotus, A., A. Burzynski, et R. Wenne. 2007. Family of *Tc1*-like elements from fish genomes and horizontal transfer. *Gene* **390** (1-2) : 243-251.
- Poulter, R. T. M., et T. J. D. Goodwin. 2005. *DIRS-1* and the other tyrosine recombinase retrotransposons. *Cytogenetic and Genome Research* **110** (1-4) : 575-588.
- Pouteau, S., E. Huttner, M. A. Grandbastien, et M. Caboche. 1991. Specific expression of the tobacco *Tnt1* retrotransposon in protoplasts. *The EMBO journal* **10** (7) : 1911-1918.
- Pritham, E. J., T. Putliwala, et C. Feschotte. 2007. *Mavericks*, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. *Gene* **390** (1-2) : 3-17.

## R

- Rai, L. C., et J. P. Gaur. 2001. *Algal adaptation to environmental stresses. Physiological, biochemical and molecular mechanisms*. Verlag Springer, New York.
- Raimbault, P. 1986. Effect of temperature on nitrite excretion by three marine diatoms during nitrate uptake. *Marine Biology* **92** (2) : 149-155.
- Rajadurai, M., E. H. Poornima, S. V. Narasimhan, V. N. R. Rao, et V. P. Venugopalan. 2005. Phytoplankton growth under temperature stress : Laboratory studies using two diatoms from a tropical coastal power station site. *Journal of Thermal Biology* **30** (4) : 299-305.

- Rakocevic, A., S. Mondy, L. Tirichine, V. Cosson, L. Brocard, A. Iantcheva, A. Cayrel, B. Devier, G. A. Abu El-Heba, et P. Ratet. 2009. *MERE1*, a low-copy-number *Copia*-type retroelement in *Medicago truncatula* active during tissue culture. *Plant Physiology* **151** (3) : 1250-1263.
- Ramallo, E., R. Kalendar, A. Schulman, et J. Martinez-Izquierdo. 2008. *Remel*, a *Copia* retrotransposon in melon, is transcriptionally induced by UV light. *Plant Molecular Biology* **66** : 137-150.
- Ratner, V. A., S. A. Zabanov, O. V. Kolesnikova, et L. A. Vasilyeva. 1992. Induction of the mobile genetic element *Dm-412* transpositions in the *Drosophila* genome by heat shock treatment. *Proceedings of the National Academy of Sciences* **89** (12) : 5650-5654.
- Ray, D. A., C. Feschotte, H. J. T. Pagan, J. D. Smith, E. J. Pritham, P. Arensburger, P. W. Atkinson, et N. L. Craig. 2008. Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*. *Genome Research* **18** (5) : 632-639.
- Rech, M., A. Morant-Manceau, et G. Tremblin. 2008. Carbon fixation and carbonic anhydrase activity in *Haslea ostrearia* (Bacillariophyceae) in relation to growth irradiance. *Photosynthetica* **46** : 56-62.
- Rech, M., J.-L. Mouget, et G. Tremblin. 2003. Modification of the Hansatech FMS fluorometer to facilitate measurements with microalgal cultures. *Aquatic Botany* **77** : 71-80.
- Reinfelder, J. R., Kraepiel, Anne M. L., Morel, et Francois M. M. 2000. Unicellular C4 photosynthesis in a marine diatom. *Nature* **407** (6807) : 996-999.
- Ren, X., Y. Park, et T. A. Miller. 2006. Intact *mariner*-like element in tobacco budworm, *Heliothis virescens* (Lepidoptera : Noctuidae). *Insect Molecular Biology* **15** (6) : 743-748.
- Reynolds, C. S. 1984. Phytoplankton periodicity : the interactions of form, function and environmental variability. *Freshwater Biology* **14** (2) : 111-142.
- Richardson, J. M., S. D. Colloms, D. J. Finnegan, et M. D. Walkinshaw. 2009. Molecular architecture of the *Mos1* paired-end complex : the structural basis of DNA transposition in a eukaryote. *Cell* **138** (6) : 1096-1108.
- Richardson, J. M., L. Zhang, S. Marcos, D. J. Finnegan, M. M. Harding, P. Taylor, et M. D. Walkinshaw. 2004. Expression, purification and preliminary crystallographic studies of a single-point mutant of *Mos1 mariner* transposase. *Acta Crystallographica* **60** (5) : 962-964.
- Riebesell, U. S. 2000. Carbon fix for a diatom. *Science* **407** (6807) : 959-960.
- Robertson, H. M. 2002. Evolution of DNA transposons in eukaryotes. *In : Mobile DNA II*, N. L. Craig, R. Craigie, M. Gellert, et A. M. Lambowitz (Eds). American Society of Microbiology Press, Washington, pp. 1093-1110.
- Robertson, H. M. 1997. Multiple mariner transposons in flatworms and hydras are related to those of insects. *Journal of Heredity* **88** (3) : 195-201.

- Robertson, H. M., et M. L. Asplund. 1996. *Bmmarl* : a basal lineage of the mariner family of transposable elements in the silkworm moth, *Bombyx mori*. *Insect Biochemistry and Molecular Biology* **26 (8-9)** : 945-954.
- Robertson, H. M. 1995. The *Tcl-mariner* superfamily of transposons in animals. *Journal of Insect Physiology* **41 (2)** : 99-105.
- Robertson, H. M., et D. J. Lampe. 1995. Recent horizontal transfer of a *mariner* transposable element among and between *Diptera* and *Neuroptera*. *Molecular Biology and Evolution* **12 (5)** : 850-862.
- Robertson, H. M. 1993. The *mariner* transposable element is widespread in insects. *Nature* **362 (6417)** : 241-245.
- Ronquist, F., et J. P. Huelsenbeck. 2003. MrBayes 3 : Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19 (12)** : 1572-1574.
- Rouault, J. D., et P. Capy. 2008. Classification of transposable elements based upon nucleotidic distances and an aggregation process. *Journées Ouvertes Biologie Informatique Mathématiques. Lille, France, 30 juin-2 juillet.*
- Rouault, J.-D., N. Casse, B. Chénais, A. Hua-Van, J. Filée, et P. Capy. 2009. Automatic classification within families of transposable elements : application to the *mariner* family. *Gene* **448 (2)** : 227-232.
- Roulin, A., B. Piegu, P. Fortune, F. Sabot, A. D'Hont, D. Manicacci, et O. Panaud. 2009. Whole genome surveys of rice, maize and sorghum reveal multiple horizontal transfers of the LTR-retrotransposon *Route66* in *Poaceae*. *BMC Evolutionary Biology* **9** : 58.
- Roulin, A., B. Piegu, R. A. Wing, et O. Panaud. 2008. Evidence of multiple horizontal transfers of the long terminal repeat retrotransposon *RIRE1* within the genus *Oryza*. *The Plant Journal* **53 (6)** : 950-959.
- Round, F. E., R. M. Crawford, et D. G. Mann. 1990. *The diatoms : biology and morphology of the genera*. Cambridge University Press, Cambridge
- Rousch, J. M., S. E. Bingham, et M. R. Sommerfeld. 2004. Protein expression during heat stress in thermo-intolerant and thermo-tolerant diatoms. *Journal of Experimental Marine Biology and Ecology* **306 (2)** : 231-243.
- Rousch, J. M., S. E. Bingham, et M. R. Sommerfeld. 2003. Changes in fatty acid profiles of thermo-intolerant and thermo-tolerant marine diatoms during temperature stress. *Journal of Experimental Marine Biology and Ecology* **295 (2)** : 145-156.
- Rubin, G., et A. Spradling. 1982. Genetic transformation of *Drosophila* with transposable element vectors. *Science* **218 (4570)** : 348-353.
- Ryder, E., et S. Russell. 2003. Transposable elements as tools for genomics and genetics in *Drosophila*. *Briefings in Functional Genomics and Proteomics* **2** : 57-71.

## S

- Saade, A., et C. Bowler. 2009. Molecular tools for discovering the secrets of diatoms. *BioScience* **59** (9) : 757-765.
- Sanmiguel, P., et J. L. Bennetzen. 1998. Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. *Annals of Botany* **82** (suppl1) : 37-44.
- Sarthou, G., K. R. Timmermans, S. Blain, et P. Tréguer. 2005. Growth physiology and fate of diatoms in the ocean : a review. *Journal of Sea Research* **53** (1-2) : 25-42.
- Scala, S., et C. Bowler. 2001. Molecular insights into the novel aspects of diatom biology. *Cellular and Molecular Life Sciences* **58** (11) : 1666-1673.
- Scandalios, J. G. 2005. Oxidative stress : molecular perception and transduction of signals triggering antioxidant gene defenses. *Brazilian Journal of Medical and Biological Research* **38** : 995-1014.
- Scandalios, J. G. 1993. Oxygen stress and superoxide dismutases. *Plant physiology* **101** : 7-12.
- Schmidt, A. L., et L. M. Anderson. 2006. Repetitive DNA elements as mediators of genomic change in response to environmental cues. *Biological Reviews* **81** (4) : 531-543.
- Selye, H. 1976. Forty years of stress research : principal remaining problems and misconceptions. *Canadian Medical Association Journal* **115** : 53-56.
- Shao, H., et Z. Tu. 2001. Expanding the diversity of the *IS630-Tc1-mariner* superfamily : discovery of a unique DD37E transposon and reclassification of the DD37D and DD39D transposons. *Genetics* **159** (3) : 1103-1115.
- Shapiro, J. 1999. Transposable elements as the key to a 21st century view of evolution. *Genetica* **107** : 171-179.
- Siddique, M., S. Gernhard, P. von Koskull-Döring, E. Vierling, et K.-D. Scharf. 2008. The plant sHSP superfamily : five new members in *Arabidopsis thaliana* with unexpected properties. *Cell Stress and Chaperones* **13** (2) : 183-197.
- Sijen, T., et R. H. A. Plasterk. 2003. Transposon silencing in the *Caenorhabditis elegans* germ line by natural RNAi. *Nature* **426** (6964) : 310-314.
- Silva, J. C., F. Bastida, S. L. Bidwell, P. J. Johnson, et J. M. Carlton. 2005. A potentially functional *mariner* transposable element in the protist *Trichomonas vaginalis*. *Molecular Biology and Evolution* **22** : 126-134.
- Silva, J. C., et M. G. Kidwell. 2004. Evolution of *P* elements in natural populations of *Drosophila willistoni* and *D. sturtevantii*. *Genetics* **168** (3) : 1323-1335.
- Simon, L., M. Lalonde, et T. D. Bruns. 1992. Specific amplification of 18S fungal ribosomal genes from vesicular-arbuscular endomycorrhizal fungi colonizing roots. *Applied Environmental Microbiology* **58** : 291-295.

- Sinzelle, L., Z. Izsvák, et Z. Ivics. 2009. Molecular domestication of transposable elements : From detrimental parasites to useful host genes. *Cellular and Molecular Life Sciences* **66** (6) : 1073-1093.
- Sinzelle, L., N. Pollet, Y. Bigot, et A. Mazabraud. 2005. Characterization of multiple lineages of *Tc1*-like elements within the genome of the amphibian *Xenopus tropicalis*. *Gene* **349** : 187-196.
- Skibbe, D. S., J. F. Fernandes, K. F. Medzihradzky, A. L. Burlingame, et V. Walbot. 2009. *Mutator* transposon activity reprograms the transcriptomes and proteomes of developing maize anthers. *The Plant Journal* **59** (4) : 622-633.
- Slotkin, R. K., et R. Martienssen. 2007. Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics* **8** (4) : 272-285.
- Smetacek, V. S. 1985. Role of sinking in diatom life-history cycles : ecological, evolutionary and geological significance. *Marine Biology* **84** (3) : 239-251.
- Smillie, R. M., et S. E. Hetherington. 1990. Screening for stress tolerance by chlorophyll fluorescence. *In : Measurement techniques in plant science*, Y. Hashimoto, P. J. Kramer, H. Noanami, et B. R. Strain (Eds). Academic Press Incorporation, San Diego, pp. 229-261.
- Smirnoff, N. 1998. Plant resistance to environmental stress. *Current Opinion in Biotechnology* **9** (2) : 214-219.
- Smit, A., R. Hubley, and P. Green. 1996-2004. RepeatMasker Open-3.0. [http : //www.repeatmasker.org](http://www.repeatmasker.org).
- Smit, A. F. 1996. The origin of interspersed repeats in the human genome. *Current Opinion in Genetics and Development* **6** (6) : 743-748.
- Sobecky, P. A., et T. H. Hazen. 2009. Horizontal gene transfer and mobile genetic elements in marine systems. *Horizontal Gene Transfer* **532** : 435-453.
- Speziale, B. J., S. P. Schreiner, P. A. Giammatteo, and J. E. Scxhindler. 1984. Comparison of N, N-dimethylformamide, dimethyl sulfoxide, and acetone for extraction of phytoplankton chlorophyll. *Canadian Journal of Fisheries and Aquatic Sciences* **41** : 1519-1522.
- Spolaore, P., C. Joannis-Cassan, E. Duran, et A. Isambert. 2006. Commercial applications of microalgae. *Journal of Bioscience and Bioengineering* **101** (2) : 87-96.
- Stewart, G. J., et C. D. Sinigalliano. 1990. Detection of horizontal gene transfer by natural transformation in native and introduced species of bacteria in marine and synthetic sediments. *Applied and Environmental Microbiology* **56** (6) : 1818-1824.
- Strand, D. J., et J. F. McDonald. 1985. *Copia* is transcriptionally responsive to environmental stress. *Nucleic Acids Research* **13** (12) : 4401-4410.
- Sun, W., M. Van Montagu, et N. Verbruggen. 2002. Small heat shock proteins and stress tolerance in plants. *Biochimica and Biophysica Acta* **1577** : 1-9.
- Suzuki, Y., et M. Takahashi. 1995. Growth responses of several diatom species isolated from various environments to temperature. *Journal of Phycology* **31** (6) : 880-888.

Swofford, D. 2003. PAUP. Phylogenetic Analysis Using Parsimony (and Other Methods). Version 4.

## T

Takeda, S., K. Sugimoto, H. Otsuki, et H. Hirochika. 1999. A 13-bp cis-regulatory element in the LTR promoter of the tobacco retrotransposon *Tto1* is involved in responsiveness to tissue culture, wounding, methyl jasmonate and fungal elicitors. *The Plant Journal* **18** (4) : 383-393.

Takeda, S., K. Sugimoto, H. Otsuki, et H. Hirochika. 1998. Transcriptional activation of the tobacco retrotransposon *Tto1* by wounding and methyl jasmonate. *Plant Molecular Biology* **36** (3) : 365-376.

Tamura, K., J. Dudley, M. Nei, et S. Kumar. 2007. MEGA4 : molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24** (8) : 1596-1599.

Tapia, G., I. Verdugo, M. Yanez, I. Ahumada, C. Theoduloz, C. Cordero, F. Poblete, E. Gonzalez, et S. Ruiz-Lara. 2005. Involvement of ethylene in stress-induced expression of the *TLC1.1* Retrotransposon from *Lycopersicon chilense* Dun. *Plant Physiology* **138** (4) : 2075-2086.

Tarchini, R., P. Biddle, R. Wineland, S. Tingey, et A. Rafalski. 2000. The complete sequence of 340 kb of DNA around the rice *Adh1-Adh2* region reveals interrupted colinearity with maize chromosome 4. *Plant Cell* **12** (3) : 381-392.

Taylor, N. L., Y.-F. Tan, R. P. Jacoby, et A. H. Millar. 2009. Abiotic environmental stress induced changes in the *Arabidopsis thaliana* chloroplast, mitochondria and peroxisome proteomes. *Journal of Proteomics* **72** (3) : 367-378.

Tenaillon, M. I., J. D. Hollister, et B. S. Gaut. 2010. A triptych of the evolution of plant transposable elements. *Trends in Plant Science* **15** (8) : 471-478.

Thompson, J. D., D. G. Higgins, et T. J. Gibson. 1994. CLUSTAL W : improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22** (22) : 4673-4680.

Tosi, L. R. O., et S. M. Beverley. 2000. *Cis* and *trans* factors affecting *Mos1 mariner* evolution and transposition *in vitro*, and its potential for functional genomics. *Nucleic Acids Research* **28** (3) : 784-790.

Treangen, T. J., A.-L. Abraham, M. Touchon, et E. P. C. Rocha. 2009. Genesis, effects and fates of repeats in prokaryotic genomes. *FEMS Microbiology Reviews* **33** (3) : 539-571.

Turcotte, K., S. Srinivasan, et T. Bureau. 2001. Survey of transposable elements from rice genomic sequences. *The Plant Journal* **25** (2) : 169-179.

## U

Unsal, K., et G. T. Morgan. 1995. A novel group of families of short interspersed repetitive elements (SINEs) in *Xenopus* : evidence of a specific target site for DNA-mediated transposition of inverted-repeat SINEs. *Journal of Molecular Biology* **248** (4) : 812-823.

## V

- Van Berkel, J., F. Salamini, et C. Gebhardt. 1994. Transcripts accumulating during cold storage of potato (*Solanum tuberosum* L.) tubers are sequence related to stress-responsive genes. *Plant physiology* **104** (2) : 445-452.
- Van den Hoek, C., Mann, D.G., et H. M. Johns. 1995. *Algae : An introduction to phycology*. Nueva York, EUA, Cambridge.
- Van Gent, D. C., K. Mizuuchi, et M. Gellert. 1996. Similarities between initiation of V(D)J recombination and retroviral integration. *Science* **271** (5255) : 1592-1594.
- Vardi, A., K. Thamtrakoln, K. Bidle, et P. Falkowski. 2008. Diatom genomes come of age. *Genome Biology* **9** (12) : 245.
- Vardi, A., F. Formigini, R. Casotti, A. De Martino, F. Ribalet, A. Miralto, et C. Bowler. 2006. A stress surveillance system based on calcium and nitric oxide in marine diatoms. *PLoS Biology* **4** (3) : e60.
- Venner, S., C. Feschotte, et C. Biéumont. 2009. Dynamics of transposable elements : towards a community ecology of the genome. *Trends in Genetics* **25** (7) : 317-323.
- Vernhettes, S., M.-A. I. Grandbastien, et J. M. Casacuberta. 1997. *In vivo* characterization of transcriptional regulatory sequences involved in the defence-associated expression of the tobacco retrotransposon *Tnt1*. *Plant Molecular Biology* **35** (5) : 673-679.
- Vierling, E. 1991. The roles of heat shock proteins in plants. *Annual Review of Plant Physiology and Plant Molecular Biology* **42** : 579-620.
- Vitte, C., O. Panaud, et H. Quesneville. 2007. LTR retrotransposons in rice (*Oryza sativa*, L.) : recent burst amplifications followed by rapid DNA loss. *BMC Genomics* **8** : 218.
- Vitte, C., et O. Panaud. 2005. LTR retrotransposons and flowering plant genome size : emergence of the increase / decrease model. *Cytogenetic and Genome Research* **110** : 91-107.
- Vos, J. C., et R. H. Plasterk. 1994. *Tc1* transposase of *Caenorhabditis elegans* is an endonuclease with a bipartite DNA binding domain. *The EMBO journal* **13** (24) : 6125-6132.
- Vos, J. C., H. G. Van Luenen, et R. H. A. Plasterk. 1993. Characterization of the *Caenorhabditis elegans Tc1* transposase *in vivo* and *in vitro*. *Genes and development* **7** : 1244-1253.
- Vyas, D., et S. Kumar. 2005. Purification and partial characterization of a low temperature responsive Mn-SOD from tea (*Camellia sinensis* (L.) O. Kuntze). *Biochemical and Biophysical Research Communications* **329** (3) : 831-838.

## W

- Wagner, A. 2006. Periodic extinctions of transposable elements in bacterial lineages : evidence from intragenomic variation in multiple genomes. *Molecular Biology and Evolution* **23** (4) : 723-733.

- Wang, J., R. Staten, T. A. Miller, et Y. Park. 2005. Inactivated *mariner*-like elements (*MLE*) in pink bollworm, *Pectinophora gossypiella*. *Insect Molecular Biology* **14** (5) : 547-553.
- Ward, O. P., et A. Singh. 2005. Omega-3/6 fatty acids : Alternative sources of production. *Process Biochemistry* **40** (2) : 3627-3652.
- Warren, W. D., P. W. Atkinson, et D. A. O'Brochta. 1994. The *Hermes* transposable element from the house fly, *Musca domestica*, is a short inverted repeat-type element of the *hobo*, *Ac*, and *Tam3* (hAT) element family. *Genetics Research* **64** (2).
- Waters, E., B. Aebermann, et Z. Sanders-Reed. 2008. Comparative analysis of the small heat shock proteins in three angiosperm genomes identifies new subfamilies and reveals diverse evolutionary patterns. *Cell Stress and Chaperones* **13** (2) : 127-142.
- Waters, E., et I. Rioflorido. 2007. Evolutionary analysis of the small heat shock proteins in five complete algal genomes. *Journal of Molecular Evolution* **65** (2) : 162-174.
- Waters, E. R., et E. Vierling. 1999. The diversification of plant cytosolic small heat shock proteins preceded the divergence of mosses. *Molecular Biology and Evolution* **16** : 127-139
- Wessler, S. R. 1996. Plant retrotransposons : Turned on by stress. *Current Biology* **6** (8) : 959-961.
- Wessler, S. R., T. E. Bureau, et S. E. White. 1995. LTR-retrotransposons and MITEs : important players in the evolution of plant genomes. *Current Opinion in Genetics and Development* **5** (6) : 814-821.
- Wicker, T., F. Sabot, A. Hua-Van, J. Bennetzen, P. Capy, B. Chaloub, A. Flavell, P. Leroy, M. Morgante, O. Panaud, E. Paux, P. SanMiguel, et A. Schulman. 2007. A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics* **8** (12) : 973 - 982.
- William, K. W. L., et I. Morris. 1982. Temperature adaptation in *Phaeodactylum tricornutum* (Bohlin) : Photosynthetic rate compensation and capacity. *Journal of Experimental Marine Biology and Ecology* **58** (2-3) : 135-150.
- Wolfe-Simon, F., V. Starovoytov, J. R. Reinfelder, O. Schofield, et P. G. Falkowski. 2006. Localization and role of manganese superoxide dismutase in a marine diatom. *Plant physiology* **142** (4) : 1701-1709.
- Wolfe-Simon, F., D. Grzebyk, S. Schofield, et P. G. Falkowski. 2005. The role and evolution of superoxide dismutase in algae. *Journal of Phycology* **41** (3) : 453-465.
- Won, H., et S. S. Renner. 2003. Horizontal gene transfer from flowering plants to *Gnetum*. *Proceedings of the National Academy of Sciences* **100** (19) : 10824-10829.
- Wu, J.-T. 1999. A generic index of diatom assemblages as bioindicator of pollution in the Keelung River of Taiwan. *Hydrobiologia* **397** : 79-87.
- Xu, Z., X. Yan, S. Maurais, H. Fu, et D. G. O'Brien, J. Mottinger, and H. K. Dooner. 2004. *Jittery*, a *Mutator* Y-distant relative with a paradoxical mobile behavior : excision without reinsertion. *Plant Cell* **16** (5) : 1105-1114.

## Y

- Yamagushi, T., K. Nakayama, T. Hayashi, J. Yazaki, N. Kishimoto, S. Kikushi, et S. Koike. 2004. cDNA microarray analysis of rice anther genes under chilling stress at the microsporogenesis stage revealed two genes with DNA transposon *Castaway* in the 5'-flanking region. *Bioscience, Biotechnology, and Biochemistry* **68 (6)** : 1315-1323.
- Yang, G., D. H. Nagel, C. d. Feschotte, C. N. Hancock, et S. R. Wessler. 2009. Tuned for transposition : molecular determinants underlying the hyperactivity of a *Stowaway* MITE. *Science* **325 (5946)** : 1391-1394.
- Yang, G., C. F. Weil, et S. R. Wessler. 2006. A rice *Tc1/ mariner*-like element transposes in yeast. *Plant Cell* **18 (10)** : 2469-2478.
- Yeadon, P. J., et D. E. A. Catcheside. 1995. *Guest a 98* by inverted repeat transposable element in *Neurospora crassa*. *Molecular and General Genetics MGG* **247** : 105-109.
- Yildiz, M., et H. Terzi. 2008. Small heat shock protein responses in leaf tissues of wheat cultivars with different heat susceptibility. *Biologia* **63 (4)** : 521-525.
- Yongmanitchai, W., et O. P. Ward. 1991. Growth of and omega-3 fatty acid production by *Phaeodactylum tricornutum* under different culture conditions. *Applied and Environmental Microbiology* **57 (2)** : 419-425.

## Z

- Zelko, I. N., T. J. Mariani, et R. J. Folz. 2002. Superoxide dismutase multigene family : a comparison of the CuZn-SOD (SOD1), Mn-SOD (SOD2), and EC-SOD (SOD3) gene structures, evolution, and expression. *Free Radical Biology and Medicine* **33 (3)** : 337-349.
- Zerbib, D., P. Prentki, P. Gamas, E. Freund, D. J. Galas, et M. Chandler. 1990. Functional organization of the ends of *IS-1* : specific binding site for an *IS1*-encoded protein. *Molecular Microbiology* **4 (9)** : 1477-1486.
- Zhang, X., N. Jiang, C. Feschotte, et S. R. Wessler. 2004. *PIF*- and *Pong*-like transposable elements : distribution, evolution and relationship with *Tourist*-like miniature inverted-repeat transposable elements. *Genetics* **166 (2)** : 971-986.
- Zhang, L., A. Dawson, et D. J. Finnegan. 2001. DNA-binding activity and subunit interaction of the *mariner* transposase. *Nucleic Acids Research* **29 (17)** : 3566-3575.
- Zhaxybayeva, O., J. P. Gogarten, R. L. Charlebois, W. F. Doolittle, et R. T. Papke. 2006. Phylogenetic analyses of cyanobacterial genomes : quantification of horizontal gene transfer events. *Genome Research* **16 (9)** : 1099-1108.
- Zhou, M.-B., J.-J. Lu, H. Zhong, K.-X. Tang, et D.-Q. Tang. 2010. Distribution and polymorphism of *mariner*-like elements in the *Bambusoideae* subfamily. *Plant Systematics and Evolution* **289** : 1-11.

Zuker, C., J. Cappello, H. F. Lodish, P. George, et S. Chung. 1984. *Dictyostelium* transposable element *DIRS-1* has 350-base-pair inverted terminal repeats that contain a heat shock promoter. Proceedings of the National Academy of Sciences **81** (9) : 2660-2664.

# ANNEXES



## **Annexe 1 : Maintien des cultures de microalgues et axénie**

### Conservation des souches

Pour ralentir la croissance et conserver les espèces étudiées en limitant les repiquages, les souches sont cultivées sous faible éclaircment ( $20 \mu\text{mol photons m}^{-2} \text{s}^{-1}$ ) et photopériode réduite (4 h).

### Protocole d'axénie

Les cultures utilisées en physiologie et en biologie moléculaire sont des cultures axéniques, c'est-à-dire exemptes de bactéries, de champignons microscopiques. Le test d'axénie est réalisé par inoculation de la souche de microalgues dans deux milieux. Le premier milieu (FAG) est favorable à la croissance des mycètes et le deuxième est favorable au développement bactérien (FG). Pour cela, 0,5 mL de la culture de microalgues à tester sont introduits dans deux tubes à essais contenant respectivement 10 mL de milieu FAG et FG. Les tubes à essais sont placés à l'obscurité durant 5 à 6 jours à température ambiante pour limiter la croissance des microalgues. En l'absence de trouble visible qui correspondrait à un développement de microorganismes, les souches sont considérées comme axéniques. Dans le cas où une contamination est observée, l'utilisation d'antibiotiques et d'antifongiques est requise. Un mélange d'antibiotique / antifongique Sigma-Aldrich A5955 (10.000 unités pénicilline G, 10 mg streptomycine, 25  $\mu\text{g}$  amphotérycine B, par mL; Saint Quentin Fallavier, France) est ajouté dans la culture de microalgues à raison de 1,25 mL pour 125 mL d'EMA. Pour un retour à la phase exponentielle de croissance, la culture est placée pendant 5 jours dans des conditions de culture standard. Dans certains cas, l'obtention de l'axénie peut nécessiter l'ajout de plusieurs doses d'antibiotique / antifongique, en suivant la même procédure.

<u>Milieu FAG :</u>		<u>Milieu FG :</u>
Bacto-peptone :	3 g	4 g
Extrait de levure :	1 g	0,5 g
NH <sub>4</sub> SO <sub>4</sub> :	1 g	0 g
Glycérophosphate Na :	25 mg	25 mg
EDTA ferrique :	6 mg	6 mg
EMA 50 % qsp 1 L		EMA 75 % qsp 1L
pH 7,4 -7,8		pH 7,4 -7,8

## **Annexe 2 : Eau de Mer Artificielle (EMA)**

L'EMA utilisée comme milieu de culture est préparée d'après Harrison et al., (1980). Les solutions de sels sont préparées et autoclavées séparément. L'EMA est conservée à 4 °C.

		Concentration finale (M)
Solution I : sels anhydres	NaCl	$3,63 \cdot 10^{-1}$
	Na <sub>2</sub> SO <sub>4</sub>	$2,50 \cdot 10^{-2}$
	KCl	$8,04 \cdot 10^{-3}$
	NaHCO <sub>3</sub>	$2,07 \cdot 10^{-3}$
	KBr	$7,25 \cdot 10^{-4}$
	H <sub>3</sub> BO <sub>3</sub>	$3,72 \cdot 10^{-4}$
	NaF	$6,57 \cdot 10^{-5}$
Solution II : sels hydratés	MgCl <sub>2</sub> , 6H <sub>2</sub> O	$4,12 \cdot 10^{-2}$
	CaCl <sub>2</sub> , 2H <sub>2</sub> O	$9,40 \cdot 10^{-3}$
	SrCl <sub>2</sub> , 6H <sub>2</sub> O	$8,20 \cdot 10^{-5}$
Métaux	FeCl <sub>3</sub> , 6H <sub>2</sub> O	$6,56 \cdot 10^{-6}$
	Na <sub>2</sub> EDTA, 2H <sub>2</sub> O	$1,49 \cdot 10^{-5}$
	CoSO <sub>4</sub> , 7H <sub>2</sub> O	$5,69 \cdot 10^{-8}$
	Na <sub>2</sub> MoO <sub>4</sub> , 7H <sub>2</sub> O	$6,10 \cdot 10^{-9}$
	Na <sub>2</sub> SeO <sub>3</sub>	$1,00 \cdot 10^{-9}$
	NiCl <sub>2</sub> , 6H <sub>2</sub> O	$6,30 \cdot 10^{-9}$
	ZnSO <sub>4</sub> , 7H <sub>2</sub> O	$2,54 \cdot 10^{-7}$
	CuSO <sub>4</sub> , 5H <sub>2</sub> O	$3,92 \cdot 10^{-8}$
MnSO <sub>4</sub> , H <sub>2</sub> O	$2,42 \cdot 10^{-6}$	
Nutriments majeurs	FeNH <sub>4</sub> -citrate	$1,37 \cdot 10^{-6}$
	NaNO <sub>3</sub>	$5,49 \cdot 10^{-4}$
	NaH <sub>2</sub> PO <sub>4</sub> , 2H <sub>2</sub> O	$2,10 \cdot 10^{-5}$
	Na <sub>2</sub> SiO <sub>3</sub> , 9H <sub>2</sub> O	$1,05 \cdot 10^{-4}$
Vitamines	thiamine	$2,97 \cdot 10^{-7}$
	biotine	$4,09 \cdot 10^{-9}$
	B12	$1,47 \cdot 10^{-9}$
	acide nicotinique	$6,50 \cdot 10^{-10}$
	acide folique	$9,06 \cdot 10^{-9}$
	thymine	$9,52 \cdot 10^{-7}$
	ca-d-pantothenate	$8,39 \cdot 10^{-9}$
inositol	$1,11 \cdot 10^{-6}$	



36	C383B64	DTTChlPlnBOU	gi108861736	Shibataea chinensis clone Sc-1
37	C383B26	DTTChlPlnBOU	gi108861663	Phyllostachys pubescens clone PhK-3
38	C383B47	DTTChlPlnBOU	gi108861620	Hibanobambusa tranquillans clone Ht-1
39	C383B55	DTTChlPlnBOU	gi108861638	Oligostachyum sulcatum clone Os-1
40	C383B65	DTTChlPlnBOU	gi108861733	Schizostachyum funghomii clone Sf-1
41	C383B36	DTTChlPlnBOU	gi108861593	Bashania fargesii clone Bf-1
42	C383B31	DTTChlPlnBOU	gi108861689	Phyllostachys pubescens clone PhW-3
43	C383B60	DTTChlPlnBOU	gi108861650	Phyllostachys pubescens clone PhV-4
44	C383B62	DTTChlPlnBOU	gi108861723	Pseudosasa japonica clone Pj-1
45	C383B35	DTTChlPlnBOU	gi108861595	Brachystachyum densiflorum clone Bd-1
46	C383B43	DTTChlPlnBOU	gi108861612	Fargesia fungosa clone Ff-1
47	C383B29	DTTChlPlnBOU	gi108861642	Phyllostachys pubescens clone PhP-5
48	C383B30	DTTChlPlnBOU	gi108861714	Phyllostachys pubescens clone PhP-7
49	C383B56	DTTChlPlnBOU	gi108861659	Phyllostachys pubescens clone PhK-1
50	C383B28	DTTChlPlnBOU	Pseudostachyum polymorphum clone Pp-1	
51	C383B63	DTTChlPlnBOU	gi108861727	Sasa fortunei clone Sa-1
52	C383B67	DTTChlPln	AF432535	Lithachne humilis Lh339
53	C383B16	DTTChlPln	AF432519	Cynodon dactylon clone Cd1326
54	C380A1	DTTChlPlnPUB	AF435550	Setaria italica clone Si1967
55	C380B14	DTTChlPlnPUB	gi108861646	Phyllostachys pubescens clone PhV-2
56	C380B15	DTTChlPlnPUB	gi108861648	Phyllostachys pubescens clone PhV-3
57	C380B17	DTTChlPlnPUB	gi108861606	Dendrocalamopsis oldhamii clone Do-1
58	C380B19	DTTChlPlnPUB	gi108861729	Sasa sinica clone Ss-1
59	C380B16	DTTChlPlnPUB	gi108861610	Drepanostachyum luodianense clone Dl-1
60	C380B18	DTTChlPlnPUB	gi108861669	Phyllostachys pubescens clone PhL-3
61	C380B20	DTTChlPlnPUB	gi108861731	Sasa veitchii clone Sv-1
62	C380A2	DTTChlPlnPUB	AF435552	Sorghum bicolor (sorghum) clone Sb1316
63	C386A1	DTTChlPlnPUB	AF359544	Oryza nivara clone On1920
64	C386A2	DTTChlPlnPUB	AF435551	Sorghum bicolor (sorghum) clone Sb1315
65	C386A3	DTTChlPlnPUB	AF432589	Zea diploperennis clone Zd3417
66	D3789A1	DTTChlPln	AC147482	Medicago truncatula Chr7 mth-93e11
67	C375A1	DTTChlPlnIRI	AF435560	Iris x nelsonii clone In2816
68	C382A1	DTTChlPlnIRI	AF435556	Iris fulva clone If2802
69	C383A4	DTTChlPlnIRI	AF435554	Iris brevicaulis clone Ib615
70	C383A8	DTTChlPlnIRI	AF435561	Iris verna clone Iv2809
71	C383A9	DTTChlPlnIRI	AF435562	Iris verna clone Iv2811
72	C383B12	DTTChlPlnIRI	AF432558	Iris hexagona clone Ih2822
73	C383A5	DTTChlPlnIRI	AF435555	Iris brevicaulis clone Ib1815
74	C383A7	DTTChlPlnIRI	AF435559	Iris x nelsonii clone In2815
75	C383A6	DTTChlPlnIRI	AF435557	Iris fulva clone If2804
76	C739A1	DTTChlPlnIRI	AC006423	Arabidopsis thaliana Atmar4
77	D1282A2	DTTChlPlnIRI	AC007662	Arabidopsis thaliana chromosome 2 BAC F9A16
78	D1217A5	DTTChlPlnIRI	AF272705	Arabidopsis thaliana Atmar3
79	C794A1	DTTChlPlnIRI	AC006436	Arabidopsis thaliana Atmar7
80	C896A3	DTTChlPlnTHA	AL161498	Arabidopsis thaliana Atmar6
81	D1190A3	DTTChlPlnTHA	AF262043	Arabidopsis thaliana Atmar1
82	C382A2	DTTChlPlnLON	AF435563	Neomarica longifolia clone N12825
83	C382A3	DTTChlPlnLON	AF435564	Neomarica longifolia clone N12827
84	C382A4	DTTChlPlnLON	AF435565	Neomarica longifolia clone N12829
85	C382A6	DTTChlPlnLON	AF435571	Nicotiana tabacum clone Nt1026
86	D1681A1	DTTChlPlnLON	AY833551	Pisum sativum Psmar-2
87	D6244A1	DTTChlPlnLON	AC150442	Medicago truncatula Chr7 mth-2o12
88	C383B10	DTTChlPlnLON	AF435568	Glycine max clone Gm1021
89	D3491A1	DTTChlPlnLON	AF078934	Glycine max Soymar1
90	D3205A1	DTTChlPlnLON	AY833550	Pisum sativum Psmar-1B
91	D3208A1	DTTChlPlnLON	AY833549	Pisum sativum Psmar-1A
92	C381A1	DTTChlPlnMED	AF435566	Medicago sativa clone Ms193
93	C381A2	DTTChlPlnMED	AF435567	Medicago sativa clone Ms295
94	C382A5	DTTChlPlnMed	AF435569	Glycine max clone Gm1022
95	D3711A1	DTTChlPlnMed	AC137603	Medicago truncatula
96	C383B13	DTTChlPlnRAD	AF432513	Arundinaria simonii clone Ar1957
97	C383B49	DTTChlPlnRAD	gi108861626	Indosasa shibataeoides clone Is-1
98	C383B52	DTTChlPlnRAD	gi108861632	Monocladus saxatilis clone Ms-1
99	C383B22	DTTChlPlnRAD	gi108861655	Phyllostachys pubescens clone Ph-3
100	C383B48	DTTChlPlnRAD	gi108861624	Indocalamus latifolius clone Il-1
101	C383B44	DTTChlPlnRAD	gi108861618	Guadua angustifolia clone Ga-1
102	C383B18	DTTChlPlnRAD	AF432538	Pariana radiciflora clone Pr205
103	C383B23	DTTChlPlnRAD	gi108861657	Phyllostachys pubescens clone Ph-4
104	C383B19	DTTChlPlnRAD	AF432539	Pariana radiciflora clone Pr206
105	C383B38	DTTChlPlnRAD	gi108861604	Chusquea coronalis clone Ca-1
106	C451B51	DTTChlPlnRAD	AF432576	Zea mays clone Zm66
107	C452A7	DTTChlPlnRAD	AF432581	Zea mays subsp. mexicana clone Zx108
108	C383B14	DTTChlPlnSAT	AF432514	Avena sativa clone As1820
109	C383B17	DTTChlPlnSAT	AF432529	Hordeum vulgare clone Hv1828
110	C347A1	DTTChlPlnPis	AF435546	Oryza nivara clone Og1927
111	C994A1	DTTChlPlnPIS	AY833553	Pisum sativum Psmar-3B
112	D2205A1	DTTChlPlnPIS	AY833552	Pisum sativum Psmar-3A

113	C379A3	DTTChlPlnVUL	AF432528	Hordeum vulgare clone Hv1827
114	C380A7	DTTChlPlnVUL	AF432515	Avena sativa clone As1821
115	C380B10	DTTChlPlnVUL	AF432590	Triticum aestivum clone Ta611
116	C380B11	DTTChlPlnVUL	AF432591	Triticum aestivum clone Ta613
117	C380A8	DTTChlPlnAES	AF432593	Triticum aestivum clone Ta3428
118	C380A9	DTTChlPlnAES	AF432594	Triticum aestivum clone Ta3432
119	C380B12	DTTChlPlnAES	AF432592	Triticum aestivum clone Ta614
120	C420A2	DTTChlPlnDAC	AF432520	Cynodon dactylon clone Cd1327
121	C460A4	DTTChlPlnDAC	AF432521	Cynodon dactylon clone Cd1329
122	C467A3	DTTChlPlnDAC	AF432522	Cynodon dactylon clone Cd1330
123	C435A2	DTTChlPlnTRI	AF432584	Zea mays subsp. mexicana clone Zx347
124	C457A4	DTTChlPlnTRI	AF435574	Tripsacum pilosum clone Tp1914
125	C474A2	DTTChlPlnTRI	AF435572	Tripsacum dactyloides clone Td137
126	C446A2	DTTChlPlnTRI	AF432579	Zea mays clone Zm1811
127	C461A3	DTTChlPlnTri	AF432583	Zea mays subsp. mexicana clone Zx1010
128	C455A8	DTTChlPlnZEA	AF432578	Zea mays clone Zm69
129	C463B16	DTTChlPlnZEA	AF432586	Zea diploperennis clone
130	C458A7	DTTChlPlnDUR	AF432518	Cynodon dactylon clone Cd1325
131	C466A4	DTTChlPlnDUR	AF432525	Ehrharta dura clone Ed1934
132	C468A3	DTTChlPlnDUR	AF432526	Ehrharta dura clone Ed203
133	C453A7	DTTChlPlnBIC	AF435549	Setaria italica (foxtail millet) clone Si1966
134	C458A6	DTTChlPlnBIC	AF435553	Sorghum bicolor clone Sb1317
135	C460A5	DTTChlPlnHUM	AF432536	Lithachne humilis clone Lh3310
136	C468A4	DTTChlPlnHUM	AF432537	Pariana radiceflora clone Pr1954
137	C497A7	DTTChlPlnHUM	AF432534	Lithachne humilis clone Lh337
138	C470A1	DTTChlPlnHUM	AF359542	Oryza sativa clone Os181
139	C470A2	DTTChlPlnHUM	AF359543	Oryza nivara clone On298
140	D5195A1	DTTChlPlnHUM	AP003294	Oryza sativa Japonica Group Osmar2b
141	D5565A1	DTTChlPlnHUM	AP000836	Oryza sativa Japonica Group
142	C475A2	DTTChlPlnHUM	AF435548	Orthoclada laxa clone Ol3311
143	C461A5	DTTChlPlnHum	AF435547	Pharus lappulaceus clone Ph100
144	C458A5	DTTChlPlnORY	AF359545	Oryza nivara clone Og1924
145	C459A5	DTTChlPlnORY	AF359540	Oryza sativa clone Os63
146	C376A3	DTTChlPlnRAV	AJ557159	Beta vulgaris Vulmar30.
147	C380B27	DTTChlPlnRAV	AJ557163	Beta corolliflora Cormar8.
148	C379A7	DTTChlPlnRAV	AJ557177	Beta nana Nanmar52.
149	C376A4	DTTChlPlnGAR	AJ557176	Beta nana Nanmar7.
150	C380B23	DTTChlPlnGAR	AJ557157	Beta vulgaris Vulmar33.
151	C380B28	DTTChlPlnGAR	AJ557165	Beta corolliflora Cormar16.
152	C381B10	DTTChlPlnGAR	AJ557154	Beta vulgaris Vulmar25.
153	C380B31	DTTChlPlnGAR	AJ557171	Beta corolliflora Cormar4.
154	C380B32	DTTChlPlnNAN	AJ557174	Beta nana Nanmar9.
155	C380B33	DTTChlPlnNAN	AJ557178	Beta nana Nanmar5
156	C382A7	DTTChlPlnCOR	AJ557160	Beta vulgaris Vulmar26.
157	C386A6	DTTChlPlnCOR	AJ557173	Beta nana Nanmar13.
158	C385A4	DTTChlPlnCOR	AJ557175	nana Nanmar11.
159	C386A5	DTTChlPlnCOR	AJ557164	Beta corolliflora Cormar9.
160	C385A3	DTTChlPlnCOR	AJ557172	Beta corolliflora Cormar22.
161	C384A9	DTTChlPlnCOR	AJ557167	Beta corolliflora Cormar71.
162	C383B15	DTTChlPlnAVE	AF432516	Avena sativa clone As1822
163	C474A3	DTTChlPlnAVE	AF432517	Avena sativa clone As3431
164	C485A4	DTTChlPlnAVE	AF432530	Hordeum vulgare Hv1828
165	C485A6	DTTChlPlnAVE	AF432532	Hordeum vulgare Hv2939
166	C485A5	DTTChlPlnAVE	AF432531	Hordeum vulgare Hv2936
167	C470A4	DTTChlPlnHOR	AF432527	Hordeum vulgare clone Hv1336
168	C470A5	DTTChlPlnHOR	AF432533	Hordeum vulgare Hv3423
169	C371A2	DTTChlPlnBET	AJ557187	Beta procumbens Procmar50.
170	C377A2	DTTChlPlnBET	AJ557185	Beta procumbens Procmar52.
171	C378B12	DTTChlPlnBET	AJ557189	Beta procumbens Procmar48.
172	C380B35	DTTChlPlnBET	AJ557183	Beta procumbens Procmar51.
173	C379A5	DTTChlPlnBET	AJ557155	Beta vulgaris Vulmar35.
174	C380B21	DTTChlPlnBET	AJ557153	Beta vulgaris Vulmar23.
175	C380B25	DTTChlPlnBET	AJ557161	Beta vulgaris Vulmar1.
176	C380B24	DTTChlPlnBET	AJ557158	Beta vulgaris Vulmar31.
177	C380B30	DTTChlPlnBET	AJ557169	Beta corolliflora Cormar41.
178	C380B26	DTTChlPlnBET	AJ557162	Beta vulgaris Vulmar109.
179	C383B68	DTTChlPlnBET	AJ557170	Beta corolliflora Cormar31.
180	D3909A1	DTTChlPlnBET	AJ556159	Beta vulgaris Vulmar1.
181	C379A8	DTTChlPlnBET	AJ557181	Beta nana Nanmar45.
182	C380B34	DTTChlPlnBET	AJ557180	Beta nana Nanmar4.
183	C380B38	DTTChlPlnBET	AJ557188	Beta procumbens Procmar49.
184	C380B41	DTTChlPlnBET	AJ557192	Beta procumbens Procmar45.
185	C379A6	DTTChlPlnBET	AJ557168	Beta corolliflora Cormar6.
186	C380B22	DTTChlPlnAVE	AJ557156	Beta vulgaris Vulmar24.
187	C380B29	DTTChlPlnPRO	AJ557166	Beta corolliflora Cormar75.
188	C380B40	DTTChlPlnPRO	AJ557191	Beta procumbens Procmar46.
189	C380B37	DTTChlPlnPRO	AJ557186	Beta procumbens Procmar44.

190 C380B39 DTTChlPlnPRO AJ557190 Beta procumbens Procmar47.  
191 C380B36 DTTChlPlnPRO AJ557184 Beta procumbens Procmar53.  
192 C371A1 DTTChlDia AF432523 Dianthus sp. Ds2833  
193 C384A1 DTTChlDiaNIT Nitzschia laevis Nilamar21  
194 C384A2 DTTChlDiaNIT Nitzschia laevis Nilamar22  
195 C379A2 DTTChlDiaWEI Thalassiosira weissflogii Thaweismar12  
196 C380A6 DTTChlDiaWEI Thalassiosira weissflogii Thaweismar16  
197 C381A9 DTTChlDiaWEI Thalassiosira weissflogii Thaweismar17  
198 C381A6 DTTChlDiaWEI Thalassiosira weissflogii Thaweismar11  
199 C381A7 DTTChlDiaWEI Thalassiosira weissflogii Thaweismar14  
200 C381A8 DTTChlDiaWEI Thalassiosira weissflogii Thaweismar15  
201 C390A1 DTTChlDiaTRI Amphora coffeaeformis Acomar31  
202 C390A2 DTTChlDiaTRI Entomoneis paludosa Entopalmar11  
203 C390A4 DTTChlDiaTRI Haslea crucigera Hcrumar11  
204 C390A9 DTTChlDiaTRI Phaeodactylum tricorutum Phatrimar14  
205 C390A3 DTTChlDiaTRI Entomoneis paludosa Entopalmar12  
206 C390A6 DTTChlDiaTRI Phaeodactylum tricorutum Phatrimar11  
207 C390A8 DTTChlDiaTRI Phaeodactylum tricorutum Phatrimar13  
208 C390A7 DTTChlDiaTRI Phaeodactylum tricorutum Phatrimar12  
209 C390A5 DTTChlDiaTRI Haslea crucigera Hcrumar13  
210 C390B10 DTTChlDiaTRI Phaeodactylum tricorutum Phatrimar21  
211 C392A3 DTTChlDiaTRI Haslea crucigera Hcrumar12  
212 C397A2 DTTChlDiaCOF Haslea crucigera Hcrumar31  
213 C401A2 DTTChlDiaCOF Amphora coffeaeformis Acomar21  
214 C378A9 DTTChlDiaLAE Nitzschia laevis Nilamar11  
215 C378B10 DTTChlDiaLAE Nitzschia laevis Nilamar12  
216 C384A3 DTTChlDiaLAE Nitzschia laevis Nilamar31  
217 C384A4 DTTChlDiaLAE Nitzschia laevis Nilamar41  
218 C384A5 DTTChlDiaLAE Nitzschia laevis Nilamar51  
219 C373A1 DTTChlDiaENT Entomoneis paludosa Entopalmar41  
220 C381A4 DTTChlDiaENT Entomoneis paludosa Entopalmar21  
221 C378A6 DTTChlDiaENT Amphora coffeaeformis Acomar41  
222 C378A7 DTTChlDiaENT Entomoneis paludosa Entopalmar31  
223 C378A8 DTTChlDiaENT Haslea crucigera Hcrumar21  
224 C375A2 DTTChlDiaACU Amphora acutiuscula Aacutimar14  
225 C375A3 DTTChlDiaACU Amphora acutiuscula Aacutimar15  
226 C375A4 DTTChlDiaACU Amphora acutiuscula Aacutimar16  
227 C378A4 DTTChlDiaACU Amphora acutiuscula Aacutimar12  
228 C379A1 DTTChlDiaACU Amphora acutiuscula Aacutimar11  
229 C378A5 DTTChlDiaACU Amphora acutiuscula Aacutimar13  
230 C394A1 DTTChlDiaAUR Amphora coffeaeformis Acomar13  
231 C396A5 DTTChlDiaAUR Odontalla aurita Odaumar14  
232 C396A2 DTTChlDiaAUR Amphora coffeaeformis Acomar11  
233 C396A3 DTTChlDiaAUR Odontalla aurita Odaumar11  
234 C397A3 DTTChlDiaAUR Odontalla aurita Odaumar12  
235 C397A1 DTTChlDiaAUR Amphora coffeaeformis Acomar12  
236 C396A4 DTTChlDiaAUR Odontalla aurita Odaumar13  
237 C372A2 DTTChlSTA Stauroneis constricta Staucmar11  
238 C372A4 DTTChlSTA Stauroneis constricta Staucmar13  
239 C372A5 DTTChlSTA Stauroneis constricta Staucmar14  
240 C372A3 DTTChlSTA Stauroneis constricta Staucmar12  
241 C378B11 DTTChlSTA Stauroneis constricta Staucmar23  
242 C381A5 DTTChlSTA Stauroneis constricta Staucmar21  
243 C387A1 DTTChlPAL Nitzschia palea Nipamar11  
244 C387A2 DTTChlPAL Nitzschia palea Nipamar12  
245 C387A3 DTTChlPAL Nitzschia palea Nipamar13  
246 C388A3 DTTChlPAL Nitzschia palea Nipamar21  
247 C587A1 DTT AB005236 Arabidopsis thaliana Atmar5  
248 C733A3 DTT AL161513 Arabidopsis thaliana Atmar9  
249 D1474A1 DTTTcuHBE X01747 Drosophila melanogaster HB2.  
250 D1894A2 DTTTcuHBE X01748/28-1921 Drosophila melanogaster HB1  
251 D5295A1 DTTG03 L03359 Moneuplotes crassus Tec1-1 ORF 1-3 genes  
252 D5328A1 DTTG03 L03360 Moneuplotes crassus Tec2-1 ORF 1-3 genes  
253 D5023A1 DTT AP003574 Oryza sativa Osmar4  
254 C613A2 DTTZat AAQB01010370 Drosophila willistoni  
255 D5670A1 DTTZat AAWU01037170 Culex quinquefasciatus strain  
256 D1733A1 DTTZat AAWT01010468 Schmidtea mediterranea S2F2  
257 D3489A1 DTTZat ABRM01000317 Hydra magnipapillata strain 105  
258 D3384A1 DTTZat ABRM01009058 Hydra magnipapillata strain 105  
259 D3907A1 DTTZat AAGE02018736 Aedes aegypti strain Liverpool cont1.18736  
260 D3483A1 DTTZat L00665 Caenorhabditis elegans Tc4  
261 D4305A1 DTTZat ABRM01020873 Hydra magnipapillata strain 105  
262 D1065A1 DTTZat L23169 Oxytricha fallax  
263 D2862A1 DTTZat AAWT01066039Schmidtea mediterranea S2F2  
264 D4142A1 DTTZat ABRM01000437 Hydra magnipapillata strain 105 ctg1101284840708  
265 D1196A1 DTTZat M67475 Anabaena sp. PCC 7120 IS895  
266 D1422A2 DTTTcuTSS L12206 Salmo salar Tss1-1

267	D5218A1	DTTZat	ABRM01025524 Hydra magnipapillata strain 105 ctg1101284786696
268	C595A2	DTT	AC005169 Arabidopsis thaliana Atmar8
269	D3660A1	DTTGam	AAAB01008960 Anopheles gambiae Parker
270	D1880A1	DTTGam	AAAB01008846 Anopheles gambiae Kiwi
271	D2983A1	DTTGam	AAAB01008831 Anopheles gambiae str. PEST Piper
272	D3197A1	DTTGam	AAAB01005441 Anopheles gambiae Pags
273	D5956A1	DTTGam	AAAB01008815 Anopheles gambiae Ozzie
274	D2453A1	DTTGam	AAAB01008849 Anopheles gambiae Whistler
275	D3050A1	DTTGam	AAAB01008879 Anopheles gambiae Mango
276	D1615A1	DTTGam	AAAB01006919 Anopheles gambiae Pi-Pi Truncated
277	D2936A1	DTTGam	X70186 Kluyveromyces lactis RAG3
278	D1682A1	DTTGam	L19880 Saccharomyces cerevisiae PDC2
279	D3680A1	DTTGam	AAAB01008960 Anopheles gambiae AgH1
280	D2533A1	DTTGam	AAAB01008815 Anopheles gambiae Julo
281	C437A5	DTTFot	U37228 Aspergillus niger var. awamori Vader
282	D3652A1	DTTGam	AAAB01008948 Anopheles gambiae lovejoy
283	D4726A1	DTTcuTCL	X79773 Cydia pomonella granulosis Virus TCL4.7
284	D1017A2	DTTGam	AAAB01008906 Anopheles gambiae Watteau
285	D3171A1	DTTGam	AC025715, Caenorhabditis elegans Tc5
286	D1175A3	DTTLem	AC006161 Arabidopsis thaliana Lem1
287	D1245A1	DTTLem	DQ888705 Arabidopsis thaliana RLd1 Lem1
288	D1301A6	DTTLem	DQ888708 Arabidopsis thaliana Coimbra-4 Lem1
289	D1301A7	DTTLem	DQ888709 Arabidopsis thaliana Coimbra-1 Lem1
290	D1301A9	DTTLem	EF526301 Arabidopsis thaliana Ge-0 Lem1
291	D1301A8	DTTLem	EF526300 Arabidopsis thaliana Wassilewskija Lem1
292	D1994A1	DTTFot	AF181821 Neurospora crassa Punt
293	C880A1	DTTcuQUE	L78797 Anopheles albimanus clone Que12
294	D1199A1	DTTcuQUE	L76255 Anopheles albimanus
295	D1680A1	DTTcuQUE	L76231 Anopheles albimanus Quetzal Que27
296	D1754A1	DTTcuQUE	L78796 Anopheles albimanus clone Que7
297	D2006A1	DTTcuQUE	L78798 Anopheles albimanus clone Que6
298	D2075A1	DTTcuQUE	L78795 Anopheles albimanus clone Que2
299	D1655A2	DTTcuUHU	X17356 Drosophila heteroneura Uhu.
300	D1656A1	DTTcuUHU	X63029 D.heteroneura Uhu transposon (lambda-Het4).
301	D1658A1	DTTcuUHU	X63028 D.heteroneura Uhu transposon (lambda-Het3).
302	D1558A1	DTTcuESS	U33469 Drosophila melanogaster element clone pSd.
303	D1736A1	DTTcuESS	U33463 Drosophila melanogaster element clone pS3
304	D1797A1	DTTcuESS	U33466 Drosophila melanogaster element clone pSa.
305	D1813A1	DTTcuESS	U33467 Drosophila melanogaster element clone pSb.
306	D1875A1	DTTcuESS	U33470 Drosophila melanogaster element clone pSsu.
307	D1828A1	DTTcuESS	U33461 Drosophila melanogaster element clone pS1.
308	D2298A1	DTTcuESS	U33462 Drosophila melanogaster element clone pS2.
309	D1785A1	DTTcuESS	U33464 Drosophila melanogaster element clone pS4.
310	D1590A2	DTTcuESS	U33468 Drosophila melanogaster element clone pSc.
311	D1806A1	DTTcuESS	U33465 Drosophila melanogaster element clone pS5.
312	D1730A1	DTTcuPAR	U26938 Drosophila virilis Paris
313	D2418A1	DTTcuPAR	U49973 Homo sapiens Tigger1
314	C989A1	DTTcuTC1	X54217 Caenorhabditis briggsae TCb1 (#5)
315	D1305A1	DTTcuTC1	X07827 Caenorhabditis briggsae TCb1 Barney
316	D1351A2	DTTcuTC1	X54218 Caenorhabditis briggsae TCb1 (#10)
317	D1607A2	DTTcuTC1	M64308 Caenorhabditis briggsae strain G16 Tcb2.
318	D1610A1	DTTcuTC1	X01005 Caenorhabditis elegans Tc1.
319	D1413A1	DTTcuTOP	AAAB01008944 Anopheles gambiae Topi 1
320	D1556A2	DTTcuTOP	AAAB01008849 Anopheles gambiae Topi 2
321	D3142A1	DTTcuTOP	U89799 Anopheles gambiae Topi I
322	D1446A1	DTTcuTOP	AAAB01008975 Anopheles gambiae Topi 3
323	D1515A1	DTTcuTOP	U89800 Anopheles gambiae Topi II
324	D2003A1	DTTcuTIA	AAAB01008987 Anopheles gambiae Tiang
325	D1444A1	DTTcuTSE	U89803 Anopheles gambiae Tsessebe III
326	D1983A1	DTTcuTSE	AAAB01008968 Anopheles gambiae Tsessebe
327	D2251A1	DTTcuTSE	U89801 Anopheles gambiae Tsessebe I
328	D2452A1	DTTcuTSE	U89802 Anopheles gambiae Tsessebe II
329	D1690A1	DTTcu	AAAB01008880 Anopheles gambiae Frisky
330	D1826A1	DTTFot	AF205929 Candida albicans transposon Cirt2
331	D1205A2	DTTcuTSS	L12201 Brachydanio rerio Tbr1
332	D1619A1	DTTcuTSS	L12207 Salmo salar Tss1-2
333	D1689A2	DTTcuTSS	L12209 Salmo gairdneri Tsg1
334	D1685A1	DTTcuTSS	L12208 Salmo salar Tss1-3
335	D1445A2	DTTcuPLA	AJ303069 Pleuronectes platessa Tc1 transposon, tn6 gene.
336	D1630A2	DTTcuPLA	AJ303068 Pleuronectes platessa tn5
337	D1653A1	DTTcuPLA	AJ303067 Salmo salar tn6
338	D1499A2	DTTcuTES	M93038 Eptatretus stoutii vasotocin gene. Tes1
339	D1777A1	DTTcuMIN	X61695 Drosophila hydei Minos
340	D1670A1	DTTcu	AAAB01008960 Anopheles gambiae Tango
341	D2121A1	DTTcu	X59837 Drosophila melanogaster Pogo
342	D1728A1	DTTcuBAR	X67681 Drosophila melanogaster Bari-1
343	D1038A2	DTTcuTC3	M77697 Caenorhabditis elegans cosmid B0303 Tc3

344	D1953A1	DTTTCuTC3	AF025458	Caenorhabditis elegans	cosmid C01B12 - Tc3
345	D1442A1	DTTTCuTC3	AF208675	edes aegypti	MsqTc3
346	D2547A1	DTTTCuTC3	AAAB01008811	Anopheles gambiae	Sunny
347	D4918A1	DTTG06	S80872	Aspergillus niger	Ant1
348	C762A1	DTTMat	AAAB01008975	Anopheles gambiae	An.gambiae.ITmD37D.1 - DD37D
349	D1281A6	DTTMat	AF098501	Caenorhabditis elegans	fosmid H28G03 Cemar6 CemaT1
350	D1281A7	DTTMat	AL021497	aenorhabditis elegans	YAC Y51A2D Cemar6 CemaT1
351	D1282A6	DTTMat	Z83129	Caenorhabditis elegans	Cosmid W06G6 Cemar6 CemaT1
352	D1282A8	DTTMat	AF047658	Caenorhabditis elegans	cosmid K03H6 Cemar6 CemaT1
353	D1282A9	DTTMat	CU457737	Caenorhabditis elegans	Cosmid C52D10 Cemar6 CemaT1
354	D1302A6	DTTMat	AC006622	Caenorhabditis elegans	Ce37Mar1
355	D1282A4	DTTMat	Z93391	Caenorhabditis elegans	Cosmid W04G5
356	D1282B11	DTTMat	Z81516	Caenorhabditis elegans	Cosmid F26H9 CemaT1 Cemar6
357	D1282A5	DTTMat	Z82258	Caenorhabditis elegans	Cosmid C27H2 Cemar6 CemaT1
358	D1300A3	DTTMat	U43131	Bombyx mori	clone Bmmarl.B7
359	D1307A2	DTTMat	U47917	Bombyx mori	Bmmarl
360	C771A2	DTTMat	Z81459	Caenorhabditis elegans	Cosmid C03H12 CemaT3
361	D1303A1	DTTMat	AC084515	Caenorhabditis briggsae	cosmid G17N15
362	D1303A2	DTTMat	AC084526	Caenorhabditis briggsae	cosmid G21D19
363	D1224A2	DTTMat	AF038612	Caenorhabditis elegans	cosmid F13B6 Cemar7 CemaT2
364	D1273A3	DTTMat	U46670	Caenorhabditis elegans	cosmid R09G11 Cemar7 CemaT2
365	D1296A5	DTTMat	Z81476	Caenorhabditis elegans	Cosmid C25F9 Cemar7
366	D1302A5	DTTMat	AC084524	Caenorhabditis briggsae	cosmid G20K13
367	C869A5	DTTMat	AF315725	Musca domestica	MdmaT1b
368	C961A3	DTTMat	AF315724	Musca domestica	MdmaT1a
369	D1151A3	DTTMat	AB054644	Sarcophaga peregrina	
370	C402A1	DTTMarMauFEL	L10471	Ctenocephalides felis	Cat.flea.10.7
371	D1188A1	DTTMarMauHYM	AJ312713	Bombus terrestris	Botmar1-2.
372	D1285A1	DTTMarMauHYM	AJ312712	Bombus terrestris	Botmar1-1
373	D1275A1	DTTMarMauHYM	AJ312714	Bombus terrestris	Botmar1-3.
374	C877A1	DTTMarMauDIP	X86158	Drosophila tsacasi	clone 27
375	D1277A1	DTTMarMauDIP	X86159	Zaprionus tuberculatus	clone 31
376	D1285A5	DTTMarMauDIP	AF373028	Musca domestica	MDMA1, HF, Mudmar1
377	D1286A1	DTTMarMauDIP	M14653	Drosophila mauritiana	peach Dromar
378	D1286A3	DTTMarMauDIP	AF355143	Blattella germanica	BGMA1, Bgmar1, GC
379	D1293A1	DTTMarMauDIP	X78906	Drosophila mauritiana	DMMELE1 Mos1
380	D1406A1	DTTMarMauDIP	X78907	Drosophila simulans	DSMEL2
381	D1288B11	DTTMarMauVUL	U24436	Mayetiola destructor	Desmar1, Madmar1
382	D1499A1	DTTMarMauVUL	U24436	Mayetiola destructor	Desmar1 OBSOLETE => D1288B11
383	D1284A2	DTTMarMauOCC	U15665	Metaseiulus occidentalis	Momar1
384	C425A2	DTTMarCAP	L10474	Forficula auricularia	ar.wig.5.10
385	C458A4	DTTMarCAP	L10473	Forficula auricularia	Ear.wig.5.1
386	C454B10	DTTMarCAP	U51181	Hydra littoralis	swiftwater hydra
387	C460A1	DTTMarCAP	U51183	Hydra vulgaris	
388	C496A1	DTTMarCAP	EF407577	Diasemopsis dubia	Dsmar 2.2
389	C498A1	DTTMarCAP	MC-Ddmar2.1	Diasemopsis dubia	Dsmar2.1
390	C494A1	DTTMarCAP	U51171	Hydra vulgaris	
391	C460A2	DTTMarCAP	U91351	Buenoa sp.	Backswimmer.5
392	C459A4	DTTMarCAP	U04467	Ceratititis capitata	Med fly 25
393	C441A4	DTTMarCAP	U04466	Anastrepha suspensa	Caribbean fruit fly 32
394	C459A2	DTTMarCAP	U51176	Girardia tigrina	
395	D1242B37	DTTMarELE	U10438	Caenorhabditis elegans	cosmid B0280 Cemar1
396	C445A4	DTTMarLIN	U51175	Girardia/Dugesia tigrina	4
397	C454A7	DTTMarLIN	U51174	Girardia/Dugesia tigrina	11
398	C454B29	DTTMarLIN	U91362	Nabis sp.	HMR-1997a Damsel.bug.4
399	C454A6	DTTMarLIN	U51168	Stylochus zebra	5
400	C454B12	DTTMarLIN	L10493	Ctenolepisma lineata	Silverfish.8.2
401	C937A3	DTTMarMelAUR	U19902	Apis mellifera	Ammar1 TnM1a
402	D1256A4	DTTMarMelNEO	U08093	Drosophila erecta	clone p12
403	D1276A5	DTTMarMelNEO	U08094	Drosophila erecta	clone p19
404	D1257A7	DTTMarMelNEO	U18308	Glossina palpalis	Gpmar1
405	D1253A1	DTTMarROS	U76904	Ceratititis capitata	Ccmar1.7
406	D1257A1	DTTMarROS	U76905	Ceratititis capitata	Ccmar1.13
407	D1268A3	DTTMarROS	U40493	Ceratititis capitata	Ccmar1.18
408	D1288A8	DTTMarROS	U76903	Ceratititis capitata	Ccmar1.4
409	D1255A1	DTTMarCecSAT	M63844	Hyalophora cecropia	HCMLE, Hcmar1
410	D1286B12	DTTMarCecSAT	AB006464	Attacus atlas	Aamar1
411	D1288A1	DTTMarCecTIG	X71979	Girardia tigrina	Gtmar1, Dtmar1
412	D1283A2	DTTMarCecINV	U68392	Heterorhabditis bacteriophora	Hbmar1.6
413	D1287A8	DTTMarCecMAM	U52077	Homo sapiens	Hsmar1
414	D1252A2	DTTMarCE2	L10990	Caenorhabditis elegans	C30A5, Cemar2e
415	D1253A3	DTTMarCE2	M98552	Caenorhabditis elegans	cosmid ZK370 Cemar2d
416	C457A1	DTTAtLCTE	L10492	Ctenolepisma lineata	Silverfish.8.1 partial cds.
417	C463A4	DTTAtLCTE	L10494	Ctenolepisma lineata	Silverfish.8.3 partial cds.
418	D1079A1	DTTMarWOO	AJ437558	Meloidogyne chitwoodi	Mcmar1-2.
419	D2000A1	DTTMarWOO	AJ437557	Meloidogyne chitwoodi	Mcmar1-1.
420	D1141A1	DTTAtLABY	AJ507224	Bythograea thermydron	Bytmar1.6.

421 D1278A4 DTTAtLABY AJ507220 *Bythograea thermydron* Bytmar1.2.  
422 D1299B12DTTAtLABY AJ507222 *Bythograea thermydron* Bytmar1.4.  
423 D1294A5 DTTAtLABY AJ507230 *Bythograea thermydron* Bytmar1.12.  
424 D1298A9 DTTAtLABY AJ507225 *Bythograea thermydron* Bytmar1.7.  
425 D1306A2 DTTAtLABY AJ507227 *Bythograea thermydron* Bytmar1.9.  
426 D1298B10DTTAtLABY AJ507226 *Bythograea thermydron* Bytmar1.8.  
427 D1299B13DTTAtLABY AJ507223 *Bythograea thermydron* Bytmar1.5.  
428 D1299B14DTTAtLABY AJ507228 *Bythograea thermydron* Bytmar1.10.  
429 D1299B15DTTAtLABY AJ507229 *Bythograea thermydron* Bytmar1.11.  
430 D1298B11DTTAtLABY AJ507231 *Bythograea thermydron* Bytmar1.13.  
431 D1282A1 DTTAtLABY AJ507221 *Bythograea thermydron* Bytmar1.3.  
432 D1302A3 DTTAtLABY AJ507219 *Bythograea thermydron* Bytmar1.1.  
433 D1170A1 DTTAtLIRR U11643 *Haematobia irritans* Hi4  
434 D1260A1 DTTAtLIRR U11641 *Haematobia irritans* Hi2  
435 D1288B12DTTAtLIRR U11642 *Haematobia irritans* Hi3 = Himar1  
436 D1293A3 DTTAtLIRR L06041 *Chrysoperla plorabunda* Cpmar1  
437 D1570A1 DTTAtLIRR U11645 *Haematobia irritans* Hi6  
438 D1587A1 DTTAtLIRR U11654 *Chrysoperla plorabunda* clone Cp8  
439 D1626A1 DTTAtLIRR U11651 *Chrysoperla plorabunda* clone Cp2  
440 D1689A1 DTTAtLIRR U11653 *Chrysoperla plorabunda* clone Cp4  
441 D1692A1 DTTAtLIRR U11644 *Haematobia irritans* Hi5  
442 D1713A1 DTTAtLIRR U11650 *Chrysoperla plorabunda* clone Cp1  
443 D1266A4 DTTAtLIRR U11649 *Mantispa pulchella* clone Mp1  
444 D1237A2 DTTAtLPOR AM906141 *Portunus tribuberculatus* PorTmar1.1  
445 D1237A3 DTTAtLPOR AM906142 *Portunus tribuberculatus* PorTmar1.2  
446 D1237A4 DTTAtLPOR AM906144 *Portunus tribuberculatus* PorTmar1.4  
447 D1238A1 DTTAtLPOR AM906145 *Portunus tribuberculatus* PorTmar1.5  
448 D1243A2 DTTAtLPOR AM906143 *Portunus tribuberculatus* PorTmar1.3  
449 D1334A3 DTTAtLPOR AM906138 *Portunus pelagicus* PorPmar1.3  
440 D1338A1 DTTAtLPOR AM906136 *Portunus pelagicus* PorPmar1.1  
451 D1329A1 DTTAtLCOA AM231069 *Pachygrapsus marmoratus* Pacmar1.1  
452 D1332A9 DTTAtLCOA AM231070 *Pachygrapsus marmoratus* Pacmar1.2  
453 D1241B17DTTAtLVER DQ174779 *Heliotis virescens* Hvmar1  
454 D1301A4 DTTAtLHOM U49974 *Homo sapiens* Hsmar2  
455 D2074A1 DTTAtLTCDC X59156 *Caenorhabditis elegans* Tc2.  
456 D1059A1 DTTMsq AF378002 *Anopheles gambiae* An.gambiae.ITmD37E1  
457 D1264A3 DTTMsq AF378001 *Aedes triseriatus* Ae.triseriatus.ITmD37E1  
458 D1225A2 DTTMsq AY038027 *Aedes atropalpus* ITm-D37E-3  
459 D1282B10DTTMsq AY038028 *Aedes atropalpus* ITm-D37E-4  
460 D1301A5 DTTMsq AF378000 *Ochlerotatus epactius* Oe.epactius.ITmD37E1  
461 D1293A4 DTTMsq AY038030 *Aedes atropalpus* ITm-D37E-6  
462 D1302A4 DTTMsq AF377999 *Aedes atropalpus* ITm-D37E-Ae-atropalpus1  
463 D1228A1 DTTMsq AY038029 *Aedes atropalpus* ITm-D37E-5  
464 D1385A1 DTTMsq AY038026 *Aedes atropalpus* ITm-D37E-2  
465 D1250B38DTTFot AF076632 *Fusarium oxysporum* Fot4  
466 D1836A1 DTTFot U74294 *Botryotinia fuckeliana* Flipper  
467 D1861A1 DTTFot Z33638 *Magnaporthe grisea* Pot2  
468 D1926A1 DTTFot AF434909 *Neocosmospora* sp.  
469 D1926A2 DTTFot AF443562 *Fusarium solani* var. minus Fot 1  
470 D1928A1 DTTFot X64799 *Fusarium oxysporum* Fot1-37  
471 D1926A3 DTTFot AF443563 *Nectria haematococca* Fot 1  
472 D2187A1 DTTFot AF076631 *Fusarium oxysporum* Fot3.  
473 D2198A1 DTTFot U78574 *Nectria haematococca* NHT1A  
474 D1656A2 DTTFot AJ608703 *Fusarium oxysporum* f. sp. lycopersici Fot5 ORF  
475 D2324A1 DTTFot U58946 *Aspergillus niger* var. awamori Tan1  
476 D1863A1 DTTFot AC000133 *Emericella nidulans*  
477 D1830A1 DTTFot U40479 *Cochliobolus carbonum* Fcc1  
478 D1860A3 DTTFot U60989 *Magnaporthe grisea* Pot3  
479 D1860A2 DTTFot AJ270953 *Podospira anserina* Pat  
480 D1122A2 DTTFot AF181822 *Neurospora crassa* dPunt  
481 D3717A1 DTTTPog X55039 *Homo sapiens* hCENP-B  
482 D1153A1 DTTTGam IS630 *Shigella sonnei* IS630  
483 D1281A3 DTTTcuIMP AF282722 *Fusarium oxysporum* sp. melonis Impala  
484 D1747A1 DTT L40593 *Phanerochaete chrysosporium* Pce1  
485 D3113A1 DTT AF134291 *Phanerochaete chrysosporium* strain BKM-F-1767 Pce4  
486 D3771A1 DTT AF134290 *Phanerochaete chrysosporium* strain BKM-F-1767 Pce3  
487 D4035A1 DTT AF134289 *Phanerochaete chrysosporium* strain BKM-F-1767 Pce2  
488 D1122A1 DTTTP01 *Rhizobium japonicum* ISRj1, RIATL  
489 D1146A2 DTTTP01 *Agrobacterium rhizogenes* strain A4 ISAr1 RIATL  
490 D3790A1 DTTTG02 X84950 *Cordyceps bassiana* (*Beauveria bassiana*) Hupfer1  
491 C948A1 DTTTPrk AY517905 *Salmonella enterica*

## **Annexe 4 : Etapes de la procédure d'agrégation de la méthode UPGM-VM**

Procédures de classifications.

Ici sont présentés les programmes qui sont implémentés pour la réalisation de la classification des éléments transposables.

### **Organisation des données :**

- 1) Acquisition des données : les données ont été téléchargées en fichiers de format Fasta contenant les séquences nucléotidiques. Un fichier correspond donc à une séquence soit à un élément transposable.
- 2) A chaque fichier de type Fasta est associé un fichier de format texte qui mémorise toutes les informations ou propriétés de l'élément par rapport aux autres éléments au sein de la classification.
- 3) Les programmes utilisent des bibliothèques prédéfinies contenant toutes les procédures nécessaires à la classification. Les programmes ne prennent en arguments que les noms des fichiers à traiter et le format du fichier de sortie souhaité. La grande majorité des programmes prennent en compte des fichiers listes pour automatiser la méthode.

### **Environnement de programmation :**

Tous les programmes sont utilisables sous environnements Windows XP et Vista.

Les programmes ont été écrits en langage ADA et Scilab. Ce dernier nécessite d'installer au préalable un environnement Scilab

### **Les Programmes :**

Les programmes décrits dans cette section ont permis de gérer les données, de réaliser le processus d'agrégation et de créer une visualisation de la formation des groupes.

**ET210\_Extract\_FAS\_NSI** : ce programme lit une séquence, convertit le nom de ce fichier en un nom au format NSI, en fonction de la longueur de la séquence, et crée un fichier texte associé avec le même nom.

**ET221\_Display\_NSI** : ce programme permet de vérifier la séquence du fichier Fasta et les informations la concernant dans le fichier texte.

- 1) Calcul des distances, alignement et processus d'agrégation : **ET 812\_calc\_agreg\_grp.**

Ce programme lit des fichiers Fasta, aligne les séquences qu'ils contiennent 2 à 2 et calcule les distances 2 à 2 pour enfin les grouper dans le processus d'agrégation.

- 2) Propriétés des groupes et séquences formant les groupes : **ET820\_liste\_grp.**

Ce programme crée un fichier qui contient toutes les informations sur le processus d'agrégation pas à pas. Il décrit chaque étape de la formation des groupes.

3) Réalisation de la figure synthétisant le processus d'agrégations. **ET830\_dessin\_grp** et **ET840\_tracer\_grp**.

Ces programmes lisent le fichier de description (produit par **ET812\_calc\_agreg\_grp**) du processus d'agrégation et créent un arbre de l'agrégation.

**Annexe 5 : Alignement des séquences de *MLE* de ~ 380 pb traduites *in silico* obtenues chez les 10 espèces de diatomées.**

```

*          20          *          40          *
Nipamar1.2 : IDEKWFELITKRVWKMYITPAELERGDVPHWTYVHKSHIHKVMFIVAVTRL : 50
Nipamar1.3 : IDEKWFELITKRVWKMYITPAE FERG DVPHWTYVHKSHIHKVMFIVAVTRM : 50
Nipamar2.1 : IDEKWFELITKRVWKMYITPAE FERG DVPHLKCVHKSQINKVMFIVAVTRM : 50
Nipamar1.1 : IDEKWFELITKRVWKMYITPAE LERGDVPHWTYVHKSHIHKVMFIVAVTRL : 50
Nipamar1.4 : IDEKWFELITKRVWKMYITPAE LERGDVPHWTYVHKSHIHKVMFIVAVTRL : 50
Thaweismar1.6 : IDEKWFHITMVNRKIYILHKNEP-EPTRKTR---HKKRHITKVMFIAAVARP : 46
Thaweismar1.5 : IDEKWFHITMVNRKIYILHKNEP-EPTRKTR---HKKRHITKVMFIAAVARP : 46
Thaweismar1.7 : IDEKWFHITMVNRKIYILHKNEP-EPTRKTR---HKKRHITKVMFIAAVARP : 46
Thaweismar1.1 : IDEKWFHITMVNRKIYILHKNEP-EPTRKTR---HKKRHITKVMFIAAVARP : 46
Thaweismar1.4 : IDEKWFHITMVNRKIYILHKNEP-EPTRKTR---HKKRHITKVMFIAAVARP : 46
Thaweismar1.2 : IDEKWFHITMVNRKIYILHKNEP-EPTRKTR---HKKRHITKVMFIAAVARP : 46
Odaumar1.1 : IDEKWFEMTQDNWRCCVAQSEEDANKAKAQ-AKHKSHIAKFMFICAAARP : 49
Odaumar1.2 : IDEKWFEMTQDNWRCCVAQSEEDANKAKAQ-AKHKSHIAKFMFICAAARP : 49
Accmar1.1 : IDEKWFEMLQDNWRCCVAQSEEDANKAKAQ-AKHKSHIAKFMFICAAARP : 49
Accmar1.2 : IDEKWFEMTQDNWRCCVAQSEEDANKAKAQ-AKHKSHIAKFMFICAAARP : 49
Odaumar1.3 : IDEKWFEMTQDNWRCCVAQSEEDANKAKAQ-AKHKSHIAKFMFICAAARP : 49
Odaumar1.4 : IDEKWFEMTQDNWLCQVTSERNANKVKAQQTAKHKSHIHKVMFICAVARS : 50
Accmar1.3 : IDEKWFEMTQDNWRCCVAQSEERNANKAKARQTKHKSHIIEVVFPCAVARP : 49
Nilamar1.1 : IDEKWFYLIKEGAKYYLAHDED-PPVFSVQ---NKNTVPKVMFICGLARP : 46
Nilamar1.2 : IDEKWFYLIKEGAKYYLAHDED-PPVFSVQ---NKNTVPKVMFICGLARP : 46
Nilamar4.1 : IDEKWFYLLKEGRRFYLADEE-RPYITQQ---SKQSI PKVMFICALSRP : 46
Nilamar5.1 : IDEKWFYLLKEGRRFYLADEE-RPYITQQ---SKQSI PKVMFICALSRP : 46
Nilamar3.1 : IDEKWFYLIKDGQRFYLAHDEK-PPYISQQ---HKQSI PKVMFICALARP : 46
Nilamar2.1 : IDEKWFYLLKEGQRYYLAPGEQ-LPHLTIQ---SKNYL PKVMFISAVARP : 46
Nilamar2.2 : IDEKWFYLLKEGQRYYLAPGEQ-LPHLTIQ---SKNYL PKVMFISAVARP : 46
Hacrumar1.3 : IDEKWFHMTKDGRRYILFGGEQ-PPDRRVQ---HKGFINPKVMFIAAVARP : 46
Entopalmar1.2 : IDEKWFHMTKDGRRYILFGGEQ-PPDRRVQ---HKGFINPKVMFIAAVARP : 46
Phatrimar1.1 : IDEKWFHMTKDGRRYILFGGEQ-PPDRRVQ---HKGFINPKVMFIAAVARP : 46
Phatrimar1.2 : IDEKWFHMTKDGRRYILFGGEQ-PPDRRVQ---HKGFINPKVMFIAAVARP : 46
Phatrimar1.3 : IDEKWFHMTKDGRRYILFRWRA-TSDRRVQ---HKGFINPKVMFIAAVARP : 46
Phatrimar1.4 : IDEKWFYMTKDGRRYILFAGEP-PPDRRVQ---HKGFINPKVMFIAAVARP : 46
Entopalmar1.1 : IDEKWFYMTKDGRRYILFAGEP-PPDRRVQ---HKGFINPKVMFIAAVARP : 46
Hacrumar1.1 : IDEKWFYLLTKDGRRYILFAGEP-PPDRRVQ---HKGFINPKVMFIAAVARP : 46
accmar3.1 : IDEKWFYLLTKDGRRYILFAGEP-PPDRRVQ---HKGFINPKVMFIAAVARP : 46
Phatrimar2.1 : IDEKWFYMTKDGRRYILFNGEK-PPDRRVQ---HKGFINPKVILIAAVARP : 46
Hacrumar1.2 : IDEKWFYMTKDGRRYILFNGEK-PPDRRVQ---HKGFINPKVMFIAAVARP : 46
Hacrumar2.1 : IDEKWFELTKLTRYVMGANEK-TPRRSVQ---HKGNVPKVMFISAIARP : 46
Accmar4.1 : IDEKWFELTKLTRYVMGANEK-TPRRSVQ---HKGNVPKVMFISAIARP : 46
Entopalmar3.1 : IDEKWFELTKLTRYVMCEQEK-TPRRSVQ---HKGNVPKVMFISAIARP : 46
Entopalmar2.1 : IDEKWFDFTKVTTRYIMAAGEK-LPNRTVQ---HKSHIDKVMFISAIARP : 46
Entopalmar4.1 : IDEKWFELTKVTTRYIMAPGEK-IPHRTVQ---HKSHVDKVMFIAATARP : 46
Aacutimar1.1 : IDEKWFHMTKDNQVFILSKNEE-PPKRTVQ---HKGYIGKVMFICAQARP : 46
Aacutimar1.2 : IDEKWFHMTKDNQVFILSKNEE-PPKRTVQ---HKGYIGKVMFICAQARP : 46
Aacutimar1.3 : IDEKWFHPSKDNQVFILSKNEH-PPKRTVQ---HKGYIGKVMFICAQARP : 46
Aacutimar1.5 : IDEKWFQMTKDNQVFILSKNKE-PPKRAVQ---HKCYIGKVMFICAQACP : 46
Aacutimar1.6 : IDEKWFHMTKDNQVFILSKNEE-PPKRTVQ---HKCYIGKVMFICAQARP : 46
Aacutimar1.4 : IDEKWFHMTKDNRVFILSNNEE-PPKRTVQ---HKCYIGKVTFLCAQARP : 46
Hacrumar3.1 : IDEKWFYLLTSNEKYILVSGEE-APQRKVQ---HKKHITKIMFIVAQARP : 46
Accmar2.1 : IDEKWFYLLSSNEKYILVSGEE-APQRKVQ---HKKHITKIMFIVAQARP : 46
Staucumar1.1 : IDEKWFYLLGRVCNHVWVGDPRLNPKIYGST---NTVQTIKVMFIVAQCKP : 47
Staucumar1.3 : IDEKWFYLLGRVCNHVWVGDPRLNPKIYGST---NTVQTIKVMFIVAQCKP : 47
Staucumar1.2 : IDEKWFYLLGRVCNHVWVGNPRNLPKIYGNT---NTVQTIKVMFIVAQCKP : 47
Staucumar1.4 : IDEKWFYLLGRVCNHIWVGGPRNLLKIYGST---NTVQTIKVMFIVAQCKL : 47
Staucumar2.3 : IDEKWFYLLSSVISNVCVGKPAKPFVVRGAT---DNTTMTNVTFLVAVAKP : 47
Staucumar2.1 : IDEKWFYLLSSVISNVCVGKPAKPFVVRGAT---DNTTMTNVTFLVAVAKP : 47

```

```

60          *          80          *          100
Nipamar1.2 : EYNKNN---EIKFDGHIQCWPIIQTVKAVQTSKNRSAGTDVTVQAVNITKN : 97
Nipamar1.3 : EYNKNN---EIKFDGHIQCWPIIQTVKAVQTSKNRSAGTDVTVQAVNITKN : 97
Nipamar2.1 : EYNKNN---EIKFDGHIQCWPIIQTVKAVQTSKNRSAGTDVTVQAVNITKN : 97
Nipamar1.1 : EYNKNN---EIKFDGHIQCWPIIQTVKAVQTSKNRSAGTDVTVQAVNITKN : 97
Nipamar1.4 : EYNKNN---EIKFDGHIQCWPIIQTVKAVQTSKNRSAGTDVTVQAVNITKN : 97
Thaweismar1.6 : RFDPHS---QVWFDGKLGWPIVHLVAAQRNSRNRPA GTLELQCRPI TKD : 93
Thaweismar1.5 : RFDPHS---QVWFDGKLGWPIVHLVAAQRNSRNRPA GTLELQCRPI TKD : 93
Thaweismar1.7 : RFDPHS---QVWFDGKLGWPIVHLVAAQRNSRNRPA GTLELQCRPI TKD : 93
Thaweismar1.1 : RFDPHS---QVWFDGKLGWPIVHLVAAQRNSRNRPA GTLELQCRPI TKD : 93
Thaweismar1.4 : RFDPHS---QVWFDGKLGWPIVHLVAAQRNSRNRPA GTLELQCRPI TKD : 93
Thaweismar1.2 : RFDPHS---QVWFDGKLGWPIVHLVAAQRNSRNRPA GTLELQCRPI TKD : 93
Odaumar1.1 : RFNDEG---MCAF DGKLGWPFVVEWEQARRSSHNRPA GAWELKMIRVTKE : 96
Odaumar1.2 : RFNDEG---MCAF DGKLGWPFVVEWEQARRSSHNRPA GAWELKMIRVTKE : 96
Accmar1.1 : RFNDEG---MCAF DGKLGWPFVVEWEQARRSSHNRPA GAWELKMIRVTKE : 96
Accmar1.2 : RFNDEG---MCAF DGKLGWPFVVEWEQARRSSHNRPA GAWELKMIRVTKE : 96
Odaumar1.3 : RFNDEG---MCAF DGKLGWPFVVEWEQARRSSHNRPA GTWELKTI GVTKE : 96
Odaumar1.4 : QFSDDG---MCTFDGKIGWPFVVEWEQARRSSHNC LAGTWELKMIRVTKE : 97
Accmar1.3 : QFN DNG---MCTFDGKIGWPFVVEWEQARRSSH NHPAGAWELKTI GVTKE : 96
Nilamar1.1 : RYVG----RQYFDGKIGIWP IGYMKPAERSSVNR PKGTPEWVSVKVDRI : 91
Nilamar1.2 : RYVG----RQYFDGKIGIWP IGYMKPAERSSVNR PKGTPEWVSVKVDRI : 91
Nilamar4.1 : RYVQET---RSMFDGKIGIWP IGFKEKPAARTSVNR PAGTLEFHSIKINRD : 93
Nilamar5.1 : RYVQET---RSMFDGKIGIWP IGFKEKPAARTSVNR PAGTLEFHSIKINRD : 93
Nilamar3.1 : RYIPEK---RCMFDGKIGIWP IGFKEKPAERTSRNR NAGTMEWHNVTVNRD : 93
Nilamar2.1 : RYLPHE---KKWFDGKIGWPI IGYEPAKRASKNR PAGTLEFVSEIVDRE : 93
Nilamar2.2 : RYLPHE---KKWFDGKIGWPI IGYEPAKRASKNR PAGTLEFVSEIVDRE : 93
Hacrumar1.3 : RFIETESG-GFHW DGKLG IWP IGDITKAIFASSRR PADADVWKNKNVNRD : 95
Entopalmar1.2 : RFIETESG-GFHW DGKLG IWP IGDITKAIFASSRR PADADVWKNKNVNRD : 95
Phatrimar1.1 : RFIETESG-GFHW DGKLG IWP IGDITKAIFASSRR PADADVWKNKNVNRD : 95
Phatrimar1.2 : RFIETESG-GFHW DGKLG IWP IGDITKAIFASSRR PADADVWKNKNVNRD : 95
Phatrimar1.3 : RFIETESG-GFHW DGKLG IWP IGDITKAIFASSRR PADADVWKNKNVNRD : 95
Phatrimar1.4 : RYIETENG-GYWWDGKLG IWP IGNITKAQRSSVNR PADADVWKNKTVNHQ : 95
Entopalmar1.1 : RYIETENG-GYWWDGKLG IWP IGNITKAQRSSVNR PADADVWKNKTVNHQ : 95
Hacrumar1.1 : RYIETENG-GYWWDGKLG IWP IGNITKAQRSSVNR PADADVWKNKTVNHQ : 95
accmar3.1 : RYIETENG-GYWWDGKLG IWP IGNITKAQRSSVNR PADADVWKNKTVNHQ : 95
Phatrimar2.1 : RTYETE HG-PEYWDGKIGIWP IGDMTKAQRASVNR PAEADVWKNKTVNRS : 95
Hacrumar1.2 : RTYETE HG-PEYWDGKIGIWP IGDMTKAQRASVNR PAEADVWKNKTVNRS : 95
Hacrumar2.1 : RIVN----GVIWDGKLG IFPVGEIEKAKNNSCR PANTPEWRNKTMCRD : 91
Accmar4.1 : RIVN----GVIWDGKLG IFPVGEIEKAKNNSCR PANTPEWRNKTMCRD : 91
Entopalmar3.1 : RIVN----GVIWDGKLG IFPVGEIEKAKNNSCR PANTPEWRNKTMCRD : 91
Entopalmar2.1 : RLID----GEMWDGKLG IFPVGYKEAERNSSCR PRGTMEWKNKSM DAA : 91
Entopalmar4.1 : RLID----GVMWDGKIG IFPVGEYEP AERNSSRR PKGTMEWKNKSMCAA : 91
Aacutimar1.1 : RMVNGRM----WDGKLG IWP FGHVGOAQRKSKYR PAGAPVWENESVDRD : 91
Aacutimar1.2 : RMVNGRM----WDGKLG IWP FGHVGOAQRKSKYR PAGAPVWENESVDRD : 91
Aacutimar1.3 : RMVNGRM----WDGKLG IWP FGHVGOAQRKSKYR PAGAPVWENESVDRD : 91
Aacutimar1.5 : RMVDGRM----WDGNIGIWP FGYVGEAQRSKHR PRGAPVWKNESIKQE : 91
Aacutimar1.6 : RMVDGQM----WDGKIGIWP FGYVGEAQRSKHR ARGAPVWKNESVKQE : 91
Aacutimar1.4 : -MVDGQM----WDEKIGIWP FGYVGEAQR-SKHR PRGAPVWNNESVKQE : 89
Hacrumar3.1 : RSLQSPSGSAVDWDGKIGIWP VGEYRPAERN SKHQKRG TICWHNESMNTA : 96
Accmar2.1 : RSLQSPSGSAVDWDGKIGIWP VGEYRPAERN SKHQKRG TICWHNESMNTA : 96
Staucmar1.1 : VLNERG---EVVFDGKIGCWPVVAEQVQCTRGQYAEQYIWHNRRMDR---A : 91
Staucmar1.3 : VLNERG---EVVFDGKIGCWPVVAEQVQCTRGQYAEQYIWHNRRMDR---A : 91
Staucmar1.2 : VLNERD---KVVFDNKLGCWPVVDQVQCTRGQYAG-CIWHNRKMDR---T : 90
Staucmar1.4 : VLNERG---KVVFDGKLG-WPVAEQVQCTRGQYAGQYIWHNRRKMDR---T : 90
Staucmar2.3 : CYNQGG---QT-FLAKIGYWPVVTVTAAQQTL-RMPVGTPIIKCKH--LN : 90
Staucmar2.1 : CYNQGG---QT-FLAKIGYWPVVTVTAAQQTL-RMPVGTPIIKCKH--LN : 90

```

```

*           120           *
Nipamar1.2 : ----YTVICDQISATGNQKKWPSWLKRN--TVIQQDNA- : 129
Nipamar1.3 : ----YTVICDQISATGNQKRWPSWLKRN--TVIQQDNA- : 129
Nipamar2.1 : ----YTVICYQISATGNQKKWPSWLKRN--TVIQQDN-- : 128
Nipamar1.1 : ----YTVYADQISATGNQKKWPSWLKRN--TVIQQDNA- : 129
Nipamar1.4 : ----YTVYADQISATGNQKKWPSWLKRN--TVIQQDNA- : 129
Thaweismar1.6 : VYSDFMCN-EVIPAIEQKWPQSYRNQP---IYIQQDN-- : 126
Thaweismar1.5 : VYSDFMCN-EVIPAIEQKWPQSYRNQP---IYIQQDNA- : 127
Thaweismar1.7 : VYSDFMCN-EVIPAIEQKWPQSYRNQP---IYIQQDNA- : 127
Thaweismar1.1 : VYSDFMCN-EVIPAIEQKWPQSYRNQP---IYIQQDNA- : 127
Thaweismar1.4 : VYSDFMCN-EVIPAIEQKWPQSYRNQP---IYIQQDNA- : 127
Thaweismar1.2 : VYSDFMCN-EVIPAIEQKWPQSYRNQQ---VYIQQDN-- : 126
Odaumar1.1 : VCRKHVTQ-KALESTCKKFPCKRSGQQV--ACIQQDNA- : 131
Odaumar1.2 : VCRKHVTQ-KALESTCKKFPCKRSGQQV--ACIQQDNA- : 131
Accmar1.1 : VCRKHVTQ-KALESTCKKFPCKRSGQQV--ACIQQDNA- : 131
Accmar1.2 : VCRKHVTQ-KALESTCKKFPCKRSGQQV--ACIQQDNA- : 131
Odaumar1.3 : VCRKCMIQ-TVLESNKQKFPCECSQQQV--VCIQQDNA- : 131
Odaumar1.4 : VHRKCMTQ-KVLESMCEKFP CGRLRQQA--VCIQQDNA- : 132
Accmar1.3 : VCRKHMVQ-KVLEPTCEKFPCKRSQQQV--AHIQQDNA- : 131
Nilamar1.1 : QYRNILMY-QLLEPAIEKWP TCLYTGRP--ILIQQDNA- : 126
Nilamar1.2 : QYRNILMY-QLLEPAIEKWP TCLYTGRP--ILIQQDNA- : 126
Nilamar4.1 : VYRDMMSR-LVLEPAILSKWPTCVYTGPP--ITIQQDNA- : 128
Nilamar5.1 : VYRDMMSR-LVLEPAILSKWPTCVYTGPP--ITIQQDNA- : 128
Nilamar3.1 : EYRHMMKY-KVLSAIIKWP TCVYTGPN--ITIQQDNA- : 128
Nilamar2.1 : KYRQLLIT-EVLEPAILRKWPHNNWEGKP--IYIQQDNA- : 128
Nilamar2.2 : KYRQLLIT-EVLEPAILRKWPHNIWEGKP--IYIQQDNA- : 128
Hacrumar1.3 : VYRDYLIQ-KLLEPAIVEKWPLADRRKRH--IRIQQDNA- : 130
Entopalmar1.2 : VYRDYLIQ-KLLEPAIVEKWPLADRRKRH--IRIQQDNA- : 130
Phatrimar1.1 : VYRDYLIQ-KLLEPAIVEKWPLADRRKRH--IRIQQDNA- : 130
Phatrimar1.2 : VYRDYLIQ-KLLEPAIVEKWPLADRRKRH--IRIQQDNA- : 130
Phatrimar1.3 : VYRDYLIQ-KLLEPAIVEKWPLADRRKRH--IRIQQDNA- : 130
Phatrimar1.4 : VYREYLIN-KLLEPAIEKWP AADRRKRH--IKIQQDNA- : 130
Entopalmar1.1 : VYREYLIN-KLLEPAIEKWP AADRRKRH--IKIQQDNA- : 130
Hacrumar1.1 : VYREYLIN-KLLEPAIEKWP AADRRKRH--IKIQQDNA- : 130
accmar3.1 : VYREYLIN-KLLEPAIEKWP AADRRKRH--IKIQQDNA- : 130
Phatrimar2.1 : VYRDYLIN-KLLEPAIIAKWPSADRRQRH--IKIQQDNA- : 130
Hacrumar1.2 : VYRDYLIN-KLLEPAIIAKWPSADRRQRH--IKIQQDNA- : 130
Hacrumar2.1 : RYRMLID-CILPAILEKWPLEDREGRT--IYIQQDNA- : 126
Accmar4.1 : RYRMLID-CILPAILEKWPLEDREGRT--IYIQQDNA- : 126
Entopalmar3.1 : RYRMLID-CILPAILEKWPLEDREGRT--IYIQQDNA- : 126
Entopalmar2.1 : RYRMMIDVRIE GILANWPLADREGRT--IYIQQDNA- : 127
Entopalmar4.1 : RYRMMID-CLIPAILANWPLEDREART--IYIQQD--- : 124
Aacutimar1.1 : TYRRYMLD-KVIPAILEKFP MRILRNGG--IKIQQDNA- : 126
Aacutimar1.2 : TYRRYMLD-KVIPAILEKFP MRILRNGG--IKIQQDNA- : 126
Aacutimar1.3 : TYRRYMLD-KVIPAILEKFP MRILRNGG--IKIQQDNA- : 126
Aacutimar1.5 : VYRRYMLD-KVIPAIIKFP MRCLKEG---VKIQQDNA- : 125
Aacutimar1.6 : VYRRYMLD-KVIPAIIKFP MRYLKEG---VKIQQDNA- : 125
Aacutimar1.4 : VYCQYMLD-KVIPAIIKFP MRYLKEG---VKIQQDNA- : 123
Hacrumar3.1 : RYKDMMIN-KVLEPAIKEKWP T-SLRRGHQQIRIQQDNA- : 132
Accmar2.1 : RYKDMMIN-KVLEPAIKEKWP T-SLRRGHQQIRIQQDNAI : 133
Staucmar1.1 : MYTSMIFN-KVIESIICRYPVT-NGAKLV---IQQDNA- : 124
Staucmar1.3 : MYTSMIFN-KVIESIICRYPVT-NGAKLV---IQQDNA- : 124
Staucmar1.2 : MYTSMIFN-QVIESIICQYPAT-NGAPLV---IQQDNA- : 123
Staucmar1.4 : MYTSMIFN-QVIELIIFQYPAT-NGAPLV---IQQDNA- : 123
Staucmar2.3 : NYTKKCCS-ILFCQQD-KFPAG-QMLPI---IQQDNA- : 121
Staucmar2.1 : NYTKKCCS-ILFCQQ-LKISCWPNVAPI---IQQDNA- : 122

```

## Annexe 6 : Les matrices d'identités

Les matrices d'identité qui ont permis de former les consensus sont présentées ci-dessous. L'identité est exprimée en pourcentage et les consensus constitués de plusieurs séquences sont indiqués en grisé.

### A. coffeaeformis

consensus	Acomarcons1			Acomarcons2	Acomarcons3	Acomarcons4
	Acomar1.1	Acomar1.2	Acomar1.3	Acomar2.1	Acomar3.1	Acomar4.1
Acomar1.1	100	96,5	85,9	50,9	54,3	50,3
Acomar1.2		100	86,4	51,1	56,9	51,6
Acomar1.3			100	52,6	56,9	50,5
Acomar2.1				100	56,9	51,1
Acomar3.1					100	58,7
Acomar4.1						100

### E. paludosa

consensus	Entopalmarcons1		Entopalmarcons2	Entopalmarcons3	Entopalmarcons4
	Entopalmar1.1	Entopalmar1.2	Entopalmar2.1	Entopalmar3.1	Entopalmar4.1
Entopalmar1.1	100	77,4	55,4	62,3	52,1
Entopalmar1.2		100	54,9	60,5	51,8
Entopalmar2.1			100	66,9	73,2
Entopalmar3.1				100	67,7
Entopalmar4.1					100

### H. crucigera

consensus	Hacrumarcons1			Hacrumarcons2	Hacrumarcons3
	Hacrumar1.1	Hacrumar1.2	Hacrumar1.3	Hacrumar2.1	Hacrumar3.1
Hacrumar1.1	100	75,8	76,2	62,1	60,7
Hacrumar1.2		100	74,2	57,4	59,9
Hacrumar1.3			100	58,2	63,0
Hacrumar2.1				100	53,7
Hacrumar3.1					100

### N. laevis

consensus	Nilamarcons1		Nilamarcons2		Nilamarcons3	Nilamarcons4	Nilamarcons5
	Nilamar1.1	Nilamar1.2	Nilamar2.1	Nilamar2.2	Nilamar3.1	Nilamar4.1	Nilamar5.1
Nilamar1.1	100	95,8	59,6	59,1	66,7	66,9	66,4
Nilamar1.2		100	59,6	60,2	66,9	66,9	66,4
Nilamar2.1			100	98,4	58,9	58,9	58,9
Nilamar2.2				100	59,6	59,1	59,1
Nilamar3.1					100	70,3	69,5
Nilamar4.1						100	99,0
Nilamar5.1							100

### A. acutiuscula

consensus	Aacutimarcons1					
	Aacutimar1.1	Aacutimar1.2	Aacutimar1.3	Aacutimar1.4	Aacutimar1.5	Aacutimar1.6
Aacutimar1.1	100	98,9	96,8	75,7	77,6	79,4
Aacutimar1.2		100	97,6	76,7	78,0	79,9
Aacutimar1.3			100	74,9	76,5	77,8
Aacutimar1.4				100	92,3	90,9
Aacutimar1.5					100	91,2
Aacutimar1.6						100

*O. aurita*

<i>consensus</i>	<i>Odaumarcons1</i>			
	<i>Odaumar1.1</i>	<i>Odaumar1.2</i>	<i>Odaumar1.3</i>	<i>Odaumar1.4</i>
<i>Odaumar1.1</i>	100	98.5	93.2	85.6
<i>Odaumar1.2</i>		100	93.2	85.1
<i>Odaumar1.3</i>			100	86.1
<i>Odaumar1.4</i>				100

*T. weissflogii*

<i>consensus</i>	<i>Thaweismarcons1</i>					
	<i>Thaweismar1.1</i>	<i>Thaweismar1.2</i>	<i>Thaweismar1.4</i>	<i>Thaweismar1.5</i>	<i>Thaweismar1.6</i>	<i>Thaweismar1.7</i>
<i>Thaweismar1.1</i>	100	97,2	99,2	99,2	75,5	99,2
<i>Thaweismar1.2</i>		100	98,2	98,2	75,9	97,9
<i>Thaweismar1.4</i>			100	99,7	75,2	99,2
<i>Thaweismar1.5</i>				100	75,2	99,2
<i>Thaweismar1.6</i>					100	75,8
<i>Thaweismar1.7</i>						100

*N. palea*

<i>consensus</i>	<i>Nipamarcons1</i>				<i>Nipamarcons2</i>
	<i>Nipamar1.1</i>	<i>Nipamar1.2</i>	<i>Nipamar1.3</i>	<i>Nipamar1.4</i>	<i>Nipamar2.1</i>
<i>Nipamar1.1</i>	100	98,4	97,2	100	51,0
<i>Nipamar1.2</i>		100	98,2	98,4	51,0
<i>Nipamar1.3</i>			100	97,2	51,0
<i>Nipamar1.4</i>				100	51,0
<i>Nipamar2.1</i>					100

*P. tricornutum*

<i>consensus</i>	<i>Phatrimarcons1</i>				<i>Phatrimarcons2</i>
	<i>Phatrimar1.1</i>	<i>Phatrimar1.2</i>	<i>Phatrimar1.3</i>	<i>Phatrimar1.4</i>	<i>Phatrimar2.1</i>
<i>Phatrimar1.1</i>	100	99,2	99,7	77,2	74,1
<i>Phatrimar1.2</i>		100	99,0	76,9	74,1
<i>Phatrimar1.3</i>			100	77,2	74,1
<i>Phatrimar1.4</i>				100	74,2
<i>Phatrimar2.1</i>					100

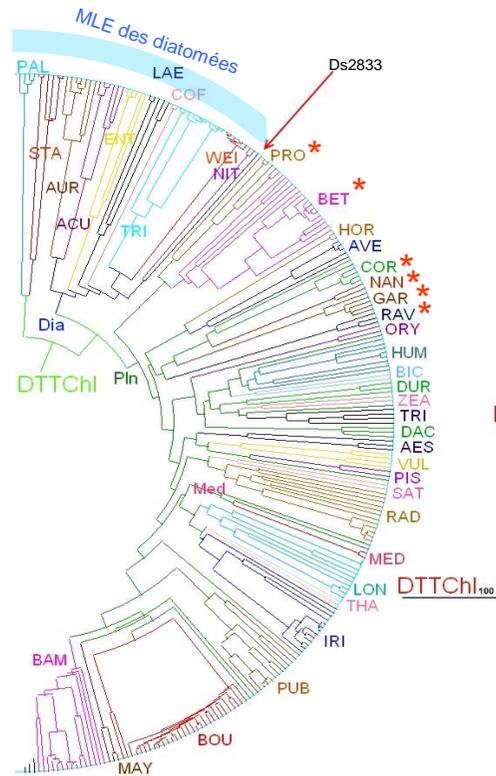
*S. constricta*

<i>consensus</i>	<i>Staucmarcons1</i>				<i>Staucmarcons2</i>		
	<i>Staucmar1.1</i>	<i>Staucmar1.2</i>	<i>Staucmar1.3</i>	<i>Staucmar1.4</i>	<i>Staucmar2.1</i>	<i>Staucmar2.2</i>	<i>Staucmar2.3</i>
<i>Staucamar1.1</i>	100	90,6	99,5	90,9	58,0	58,0	57,7
<i>Staucamar1.2</i>		100	90,6	90,1	58,5	58,5	58,5
<i>Staucamar1.3</i>			100	91,1	58,5	58,5	58,2
<i>Staucamar1.4</i>				100	57,2	57,2	57,1
<i>Staucamar2.1</i>					100	100	98,7
<i>Staucamar2.2</i>						100	98,7
<i>Staucamar2.3</i>							100

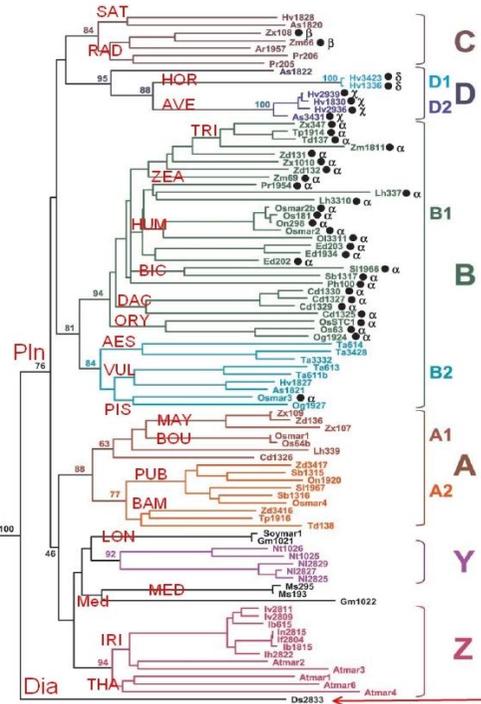
### **Annexe 7 : Méta-analyse des *MLE* de végétaux**

Les séquences de la sous-famille Plantis utilisées pour l'analyse DTT491 sont des séquences qui ont été traitées préalablement dans la littérature et dont la position phylogénétique a déjà été étudiée par Feschotte et Wessler en 2002 (pour les *MLE* issus d'Angiospermes) et Jacobs et al. 2004 (pour les *MLE* issus du genre *Beta*). Lors de la classification DTT491, les séquences issues de ces analyses ont été classées au sein de la superfamille Tc1/ *mariner*. La comparaison de la classification des *MLE* issus de végétaux par la méthode UPGM-VM par rapport aux positions phylogénétiques déterminées par l'étude de Feschotte et Wessler (2002) et le travail de Jacobs et al. (2004) est présentée dans la figure ci-après.

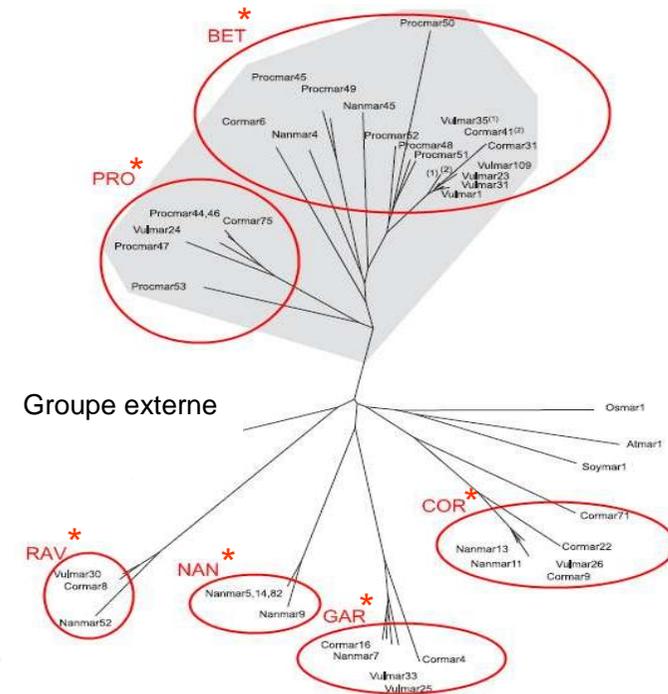
### Classification DTT491



### Phylogénie des MLE de plantes (Feschotte et Wessler, 2002)



### Phylogénie des MLE du genre Beta (Jacobs et al., 2004)



### Comparaison de la classification agrégative DTT491 par rapport aux arbres phylogénétiques des MLE de plantes terrestres

La représentation partielle de la classification agrégative DTT491 est, ici, comparée à l'arbre phylogénétique établi par de Feschotte et Wessler (2002) sur les MLE issus de différentes espèces d'Angiospermes et à l'arbre phylogénétique établi par Jacobs et al. (2004) sur les MLE dans le genre *Beta*.

Les noms des tribus, qui ont été obtenues avec la classification agrégative (à la périphérie de la demi-rosace) sont représentés (en rouge) sur l'arbre phylogénétique au niveau des nœuds correspondants aux groupes identifiés par Feschotte et Wessler (2002). L'élément Ds2833 inclus dans la sous-famille Diatomis y est indiqué (flèche rouge).

Les groupes correspondants entre la classification et l'analyse des MLE du genre *Beta* sont indiqués par un \* et entourés en rouge sur l'arbre de Jacobs et al. (2004).

Toutes les tribus obtenues à partir de la classification montrent une très bonne correspondance avec les groupes de MLE identifiés au préalable avec des méthodes phylogénétiques classiques, validant ainsi, l'utilisation de cette nouvelle méthode de classification des ET.

## **Characterization of *mariner-like* transposons in marine diatoms**

D. Hermann<sup>1,3</sup>, A. Caruso<sup>1</sup>, Q.-T. Bui<sup>2</sup>, B. Jacquette<sup>1</sup>, A. Morant-Manceau<sup>1</sup> and N. Casse<sup>1,3</sup>

<sup>1</sup> Laboratoire Mer, Molécules, Santé, EA 2160 Université du Maine, Avenue Olivier Messiaen, 72085 Le Mans cedex 9, France

<sup>2</sup> Laboratoire de Biologie Cellulaire, Institut Jean-Pierre Bourgin, INRA, 78026 Versailles, France

<sup>3</sup>Corresponding authors : [Dorothee.Hermann.Etu@univ-lemans.fr](mailto:Dorothee.Hermann.Etu@univ-lemans.fr) and [Nathalie.Casse@univ-lemans.fr](mailto:Nathalie.Casse@univ-lemans.fr)

*Mariner-like* elements (*MLEs*) are ubiquitous DNA transposons. They have been reported in a wide range of animals, and in recent years have also been detected in terrestrial plants. In contrast, there has been little investigation of *MLEs*, in the marine environment. So far, they have been investigated and found in crustaceans, but there has been no investigation in marine microalgae. In this paper, we report the first evidence of *MLE* fragments in marine photosynthetic organisms, the diatoms (Bacillariophyceae). Using degenerated primers previously employed to detect *MLEs* in various flowering plants, sequences corresponding to *mariner* elements (that belong to the Tc1-*mariner* superfamily) were obtained from 10 marine diatom species including the model species *Phaeodactylum tricornutum*. Analysis of these sequences reveals homologies with plant *MLEs*, but phylogenetic trees indicate that diatom sequences constitute a distinct cluster that could be grouped with the organisms of the *ITm D39D* family, which are found in terrestrial plants. Wide divergences between sequences from the same species reflect invasion mechanisms alternating with diversification events. These findings suggest that *mariner* subfamily elements could be widespread in marine diatoms. The search for complete *mariner-like* elements that are able to transpose is under way, and if discovered they could provide new biotechnology tools for diatom transformation.

**Key words** : transposable elements, *mariner-like* element, diatoms, microalgae, Bacillariophyceae

## 1. Introduction

Transposable elements are DNA sequences that are able to move and replicate within genomes. They are almost ubiquitous in living organisms, and display extremely diverse transposon families (for a review see Wicker et al., 2007). The activity of transposable elements and their wide distribution mean that they constitute the biggest and most rapidly evolving component of the eukaryotic genome (Hua Van et al., 2005). Transposable elements are commonly divided in two main classes on the basis of their transposition mechanism (Finnegan, 1992). Class I elements, also known as retrotransposons, use a reverse transcriptase to transpose via an RNA intermediate; the copy then being pasted elsewhere in the genome. Class II elements, usually known as transposons, transpose directly from DNA to DNA by a cut-and-paste mechanism. Class II elements have been extensively studied, with particular emphasis on the *IS630/Tc1-mariner* family, which is the most highly diversified. The length of *mariner* elements ranges from 1.5 kilobases in metazoans to 5 kb in plants. They are characterized by short terminal inverted repeats (TIRs), measuring from 30 to 50 bp, which flank a gene encoding a transposase. This gene is constituted of a single open reading frame (ORF) encoding an approximately 330-amino acid transposase. *Mariner* transposases share a DDD (aspartate residues) core feature that catalyses the transposition. This triad is crucial, since any mutation results in the loss of transposase activity (Lohe et al., 1997). The transposase recognizes these ITRs, and then excises and inserts the transposon into a target sequence within the genome. Many elements in the *mariner* subfamily are defective (Robertson, 1993; Hartl et al., 1997; Feschotte et al., 2002) as a result of mutations in the transposase gene.

There has been considerable investigation of *mariner* elements in metazoans (Robertson 1993, 1997; Jacobson et al. 1986; Auge-Gouillou et al. 1995). Recent studies have shown that *mariner-like* elements (*MLEs*) are widespread in plant genomes (Feschotte et al. 2002, 2005), and it is assumed that transposable elements are responsible for the increase in plant DNA content (Bennetzen et al. 2002, Bennetzen 2005). Partial *mariner-like* sequences have been identified in numerous flowering plants (Feschotte & Wessler, 2002), but only four complete and potentially active elements have been reported : *Soymar1* isolated from soybean (Jarvick & Lark, 1998), *Osmar1* from the rice bacterial artificial chromosome (Tarchini et al., 2000), *Vulmar1* from sugar beet (Jacobs et al., 2004) and *Psmar1* from the garden pea (Macas et al., 2005).

Diatoms (Bacillariophyceae) are ubiquitous brown unicellular microalgae. With 100,000 species reported in all aquatic environments ranging from marine to freshwater areas and pelagic to benthic habitats, they are the most diversified group of photosynthetic eukaryotes (Round et al., 1990). They have a huge

ecological impact since they are responsible for almost 40% of the global productivity of the planet (Raven & Waite, 2004). The diatom group can be divided in two subgroups on the basis of the morphology of their siliceous cell wall (Round et al., 1990). The centric group is composed of diatoms which exhibit a radial symmetry and are thought to be the most ancestral, since their presence is recorded in the earliest fossils from the Cretaceous period (Gersonde & Harwood, 1990). The pennate diatoms all exhibit bilateral symmetry and potential mobility. The phylogenetic analyses performed, have not yet identified the ancestry of the diatoms (Wiebe et al., 1996; Williams, 2007).

Nevertheless, information is emerging as a result of the sequencing of the whole genomes of two diatom species : *Phaeodactylum tricorutum* (pennate) and *Thalassiosira pseudonana* (centric) (Scala & Bowler, 2001; Scala et al., 2002; Armbrust et al., 2004; Lopez et al., 2005; Monstant et al., 2005; Bowler et al., 2008). These diatoms share various features that make them perfect models for the study of genetics and evolution : little is known about diatom genome structure even though they have small genomes, some of them are quite easy to grow under monoclonal and axenic conditions, have a short generation time, and the cells can easily be transformed genetically with weakly silicified diatom species.

A comparative study between of metazoans or bacteria and diatoms reveals differential expression and diversity rates in the repetitive DNA fraction. This component of genomes is thought to consist predominantly of class I *Ty/copia* LTR transposable elements in the model diatoms (Maumus et al., 2009). Retrotransposon activity is assumed to be the major driving force behind the creation of specific genes (Bowler et al., 2008). To date, only a few studies have investigated the occurrence of *MLEs* in marine organisms (crustaceans) (Casse et al., 2002, 2006; Halaima-Toumi et al., 2004; Bui et al., 2007, 2008). We therefore investigated *mariner-like* transposons in 13 diverse marine diatom species, including the two diatom model species *P. tricorutum* and *T. pseudonana*. PCR approach was done using *MLE5A* and *MLE3A*, the degenerated primers used by Feschotte and Wessler (2002) to identify *MLEs* in terrestrial plant genomes. Partial *MLE* sequences were used to generate phylogenetic trees in order to compare the positions of diatom *MLEs* to those of plants.

## **2. Material and methods**

### *2.1. Marine diatom species : culture conditions and DNA extraction*

Thirteen species were selected from the extensive and highly diversified group of diatoms (Bacillariophyceae). This huge group consists of at least 25, 000 genera, some of which are monospecific (Round, 1990). We sampled ten pennate and three centric diatoms (Table1). Our choice was made on the basis of

the diversity of morphological and ecological traits. For example, some species have a thin, silicified walls (the *Cylindrotheca* and *Phaeodactylum* genera) others are asymmetric or twisted diatoms (the *Amphora* and *Entomoneis* genera). Most of the diatoms sampled were epipelagic, then lived in or on sediment, but some were benthic (*E. paludosa* and *P. tricorutum*). Two species belonging to the *Amphora* and *Nitzschia* genera were collected : one from a temperate area (French atlantic coast) and the other from warm seawater (Vietnamese coast), as shown in Table 1. Although diatoms are single-cell, some species may form colonies (*Thalassiosira*, *Odontella* genera). The two model diatoms, the genomes of which had been sequenced, *Phaeodactylum tricorutum* and *Thalassiosira pseudonana*, were included in these samples.

Thirteen diatom strains were grown in batch cultures in artificial seawater (Harrison et al., 1980) at a temperature of  $16 \pm 1^\circ\text{C}$  for the species from European coast, and of  $23 \pm 1^\circ\text{C}$  for the species from Vietnamese coast (*Amphora acutiuscula* and *Nitzschia palea*).

The light/dark period was 14 h/10 h, with  $100 \mu\text{mol photons m}^{-2} \text{s}^{-1}$ . These cultures were maintained under axenic conditions. Genomic DNA extractions were performed according to Doyle & Doyle (1990) using hexadecyltrimethyl-ammonium bromide (CTAB).

## 2.2. PCR amplifications, cloning and sequencing

The PCR primers used to search for *mariner* transposase in marine diatoms were those previously defined by Feschotte & Wessler (2002) in flowering plants. These degenerated primers are derived from conserved amino acid regions encoding the plant *mariner* transposase IDEKWF (*MLE5A* : 5'-ATHGATGARAARTGGTTC-3') and IQQDNA (*MLE3A* : 5'GCATTRTCYTGTYGDAT-3'). PCR was performed on each species in 50  $\mu\text{l}$  of reaction medium with about 100 ng of genomic DNA, 0, 2  $\mu\text{M}$  of the primer, 150  $\mu\text{M}$  of each dNTP, and 0,5 U of *Taq* polymerase (Promega, Madison, WI). The PCR program consisted of an initial denaturing step of 5 min at  $94^\circ\text{C}$ , followed by 40 cycles of  $94^\circ\text{C}$  for 1 min,  $47^\circ\text{C}$  for 1 min,  $72^\circ\text{C}$  for 30 sec, and ending with  $72^\circ\text{C}$  for 10 min. PCR products were visualized on a 1% agar gel with 0.5% ethidium bromide solution added before electrophoresis. Fragments of the expected size (~ 380 bp) were eluted from the gel using Wizard SV Gel<sup>®</sup> and PCR Clean Up System<sup>®</sup> (Promega, Madison WI). The DNA fragments were then cloned in pGEM-T-easy Vector System I<sup>®</sup> plasmid (Promega, Madison WI), and sequenced by the Cogenics-Genome express company (England, <http://www.cogenics.com>)

## 2.3. Diatom sequence analyses

Nucleotide sequences were conceptually translated into amino acids using Infobiogen web services (<http://bioinfo.hku.hk/services/menuserv.html>), and tested for homology with plant transposases using TBLASTX at the NCBI site (<http://blast.ncbi.nlm.nih.gov>). All sequences have been added to the EMBL database, and their accession numbers are provided in Table 1.

Consensuses were established for sequences that displayed more than 75% nucleotide identity. When sequences displayed less than 75% identity, two consensuses were considered, and designated cons1 and cons2, respectively. The names assigned to diatom *MLEs* are derived from Robertson & Asplund (1996), thus *Acomarcons1* means *Amphora coffeaeformis mariner consensus 1*. When a species exhibited more than two different consensus sequences, they were designated consX (X = 3, 4 or 5). Protein sequences were aligned with BioEdit Sequence Alignment Editor Copyright© (1997-2007 Tom Hall) using the CLUSTAL method with the default parameters (Thompson et al., 1994). When necessary, frameshifts were introduced to maintain the alignment of the reading frame.

#### 2.4 Plant *MLE* dataset

In order to compare diatom *MLEs* and plant *MLEs*, alignments were made with 41 *MLE* sequences from the angiosperm taxon. The sequences were chosen from the lineages of plant *MLEs* determined by the phylogenetic analysis carried out by Feschotte & Wessler (2002). 380-bp bamboo *MLEs* found in Genbank, and the two recently-established complete sequences *Vulmar1* from *Beta vulgaris* (Jacobs et al., 2004) and *Psmar1* from *pisum sativum* (Macas et al., 2005) were also added to this dataset. Only the 380-bp from the catalytic domain of the complete elements was selected for the alignment and phylogenetic analyses.

#### 2.5 Phylogenetic analyses

Three different methods were used to generate phylogenetic trees : neighbor-joining (NJ), maximum parsimony (MP) with PAUP 4.0b08 (Swofford, 2002), and the Bayesian inference (BI) using Mr Bayes v3.0b4 (Ronquist & Huelsenbeck, 2003).

NJ pairwise distance analyses were carried out using the MEAN model, and for the MP analysis, we performed a heuristic search using the tree-bisection-reconnection (TBR) branch swapping option with 10 addition replicates. The node robustness was calculated by 1000 bootstrapping replicates, and 131 out of characters were included on 132 characters, for each method.

The Bayesian inference method was performed with the amino acid model determined by ProtTest (Abascal et al., 2005), which indicated the WAG+G as the best model according to the Akaike Information Criterion. This model was implemented in Mr Bayes 3.1.2 (Ronquist & Huelsenbeck, 2003), using three heated chains and a single cold chain in all Markov Chain Monte Carlo (MCMC) analyses. We conducted two independent MCMC runs with 1 million generations, and each run was initiated with random trees. We sampled trees and parameters every 100 generations. Stationarity was assessed by examining the average standard deviation of split frequencies approaching zero reflecting the fact that the two tree samples were becoming increasingly similar. We also examined the Potential Scale Reduction Factor (PSRF) approaching one in order to check the runs' convergence. The first 25% of the sampled trees were discarded as burnin.

In order to determine the relationship classification of the diatom *MLEs* within the plant *MLEs* family, trees were rooted using several animal elements belonging to sister families of plant *MLEs* in the *Tc1-mariner* superfamily (Shao & Tu, 2001). *Tc1-like* elements : *Tc3* from *Caenorhabditis elegans* and animal *MLE* elements : *Pacmar* from *Pachygrapsus marmoratus*, *Mos1* from *Drosophila mauritiana* and the human element *Hsmar1* were tested as outgroups. All the trees obtained from the different outgroups using three phylogenetic methods (NJ, MP, BI) displayed the same topology. The tree built with MEGA4 software (Tamura et al., 2007) from Bayesian analysis with the outgroup cluster *Mos1* and *Hsmar1* was presented and completed with bootstrap values from the NJ and MP methods. All bootstrap analyses were performed with 1000 replications.

### 3. Results

#### 3.1. PCR amplification and characterization of mariner-like transposase fragments

Using the degenerated primers *MLE5A* and *MLE3A*, approximately 380-bp fragments in the 13 marine diatom species were amplified (Fig 1). In ten species, we successfully amplified various fragments between 372 and 397-bp in length (Table 1). A minimum of five fragments of the expected size (about 380 bp), and at least three amplified fragments of other sizes were cloned and sequenced for each species. After comparison with databases, only the 380-bp fragments were found to show homology with plant *mariner*-like transposases. A fragment was identified as a transposase *MLE* if BLASTX score versus plant *MLEs* was over 80. For each species in which an *MLE* was detected, at least four 380-bp *mariner* partial sequences were obtained. For the ten species in which *MLEs* were detected, we successfully characterized 55 approximately 380-bp sequences. At least one open reading frame (ORF) for the gene coding the transposase was detected per diatom species (Table 1). In this study, 25 *MLE* consensus sequences were established. The ten species included three species,

*A. coffeaeformis*, *N. laevis*, *E. paludosa*, which exhibited more than two different consensus sequences, and only one species, *A. acutiuscula* that exhibited only one consensus. Each consensus sequence was conceptually translated into amino acid sequences and proteins, and the sequence lengths ranged between 124 and 132 amino acids. Despite their differences, the diatom transposases all exhibited approximately four conserved protein patterns (Fig 2). The KVMFL and the A(x)ARPR patterns (at positions 40 to 44 and 46 to 51, in the amino acid sequences respectively) were present in 20 out of the 25 diatom *MLE* sequences. In addition, the (F/W)DGKLGWIP (position 63 to 71) and KWP (position 108 to 110) patterns were observed in 17 and 19 of the 25 consensus sequences respectively. These four motifs were not restricted to the consensus, as the 55 *MLE* amino acid sequences translated *in silico* also shared these patterns.

### 3.2. Phylogenetic relationships of marine diatom *MLE* sequences

The 25 diatom consensus *MLE* sequences were assigned to 7 clusters on the basis of the percentage nucleic homology calculated by pairwise sequence comparisons (Table 2). A cluster was established when the sequences exhibited an identity of more than 75%. The phylogenetic relationships between diatom *MLE* sequences are represented in the consensus tree (on 22 parsimonious trees) constructed using the MP method with PAUP (tree lengths = 764, CI, excluding uninformative characters = 0,729; RI=0,749) (Fig 3). All the diatom sequences are clearly separated from the root corresponding to the two terrestrial *mariner* plant elements *Osmar1* and *Soymar1*, and the phylogenetic distinction is also supported by the node value (100). Clusters 2, 4, 5 and 6 are supported by high bootstrap values. The node bootstrap values for clusters 1, 3 and 7 are less than 50%, indicating that these clusters are not clearly defined. Clusters 1, 2, 3 and 4 contain sequences from a combination of different diatom species, whereas clusters 5, 6 and 7 all contain consensus sequences isolated from a single species. Cluster 1 contains 2 consensus sequences, *Odaumarcons1* and *Acomarcons1*, which share 96% homology (Table 2), and are derived from two distinct species *O. aurita* and *A. coffeaeformis*. Cluster 2 includes consensus sequences from four diatom species : *P. tricornutum* (*Phatrimarcons1*, 2), *H. crucigera* (*Hacrumarcons1*), *E. paludosa* (*Entopalmarcons1*) and *A. coffeaeformis* (*Acomarcons3*). Pairwise comparison indicates that they share between 80 and 99% of homology (Table 2). Cluster 3 consists of just 2 consensus sequences, *Hcrumarcons3* and *Acomarcons2*, which exhibit 98% homology. Cluster 4 consists mainly of three consensus sequences from *E. paludosa* (*Entopalmarcons2*, 3 and 4) associated with *Hcrumarcons2* and *Acomarcons4*. Pairwise comparison indicates homology ranging from 62% to 99% (Table 2). The degree of homology between the consensus sequences that constitute each of the mixed clusters (1, 2, 3, and 4) is less than 50% (Table 2). The consensus

isolated from one species are either assigned to distinct clusters 5, 6, 7 or separated (*Aacumarcons1*, *Thaweissmarcons1*). As mentioned above, clusters 5, 6 and 7 all consist exclusively of sequences isolated from a single diatom species, *N. laevis*, *N. palea* and *S. constricta*, respectively. The homology within these groups is between 57% and 97% in cluster 5, and 92% in clusters 6. Cluster 7 contains only two sequences, *Staucmarcons1* and 2, isolated from *S. constricta*. These two consensus sequences exhibit a very little homology (24%), and their homology with the other sequences is less than 50%. Three consensus sequences, *Aacumarcons1*, *Nilamarcons2* and *Thaweissmarcons1*, isolated from *A. acutiuscula*, *N. laevis* and *T. weissflogii* respectively, do not cluster with any other sequences.

### 3.3. Phylogenetic relationships between the MLE sequences of marine diatoms and those of terrestrial plants

The phylogenetic relationships between *MLE* fragments of marine diatoms and of terrestrial plants were determined using three classic methods NJ, MP, and Bayesian inference (Fig 4). Terrestrial plant sequences were extracted from the NCBI database (accession numbers are shown in Table 3). Somewhat different consensus tree topologies resulted from the two methods NJ and MP (consensus tree index from 30 trees, CI = 0,597, RI = 0,741). The topology resulting from the Bayesian inference (BI) method was similar to those of NJ and MP, and only the BI consensus tree is presented (Fig 4). This tree indicated four clusters. The marine Diatom *MLE* sequences were assigned to three groups, *MLE-Di1*, *MLE-Di2* and *MLE-Di3*, which were distinct from the terrestrial *MLE* sequences cluster in the last clade whatever the method was used. These groups were correlated to the clusters established from Table 2. The *MLE-Di1* group includes consensus sequences from *S. constricta* and *N. palea*, respectively, the clusters 6 and 7 (Fig 3), and also the consensus sequences from *T. weissflogii*. This group include two clusters and one of the isolated sequences previous obtained in *MLE* diatom phylogeny (Fig 3) and is supported by a high node value.

The *MLE-Di2* group corresponds to clusters 2, 3, 4 and 5 in Figure 3. This group included 18 of the 25 diatom consensus sequences, and was supported by node values of more than 80 by all methods. The isolated sequences *Aacumarcons1* and *Nilamarcons2* in the diatom phylogeny were included in this group. The *MLE-Di3* group contains sequences *Odaumarcon1* and *Acomarcons1* sequences that are very different from those in the other diatom sequences. This is consistent with their positions in the diatom phylogeny shown in Figure 3, in which they are assigned to cluster 1.

The last group (indicated by an arrow, Fig 4), includes all the *MLE* sequences from terrestrial plants and is defined as the plant *MLE* clade. This group was supported by node values of more than 93. This finding is

consistent with the plant *MLE* phylogenetic analysis results of Feschotte & Wessler (2002) who identified plant *MLE* lineages A, B, C, Y and Z in a monophyletic group. These lineages were clustered in subgroups, which less clearly supported the node values (Figure 4).

The diatom sequences are assigned to three different sister groups, and are not assigned to the same clade as the terrestrial plant *MLE* fragments. This was observed with all three phylogenetic methods used.

#### 4. Discussion

*Mariner*-like elements have been identified in terrestrial plants and four complete elements, *Osmar1*, *Soymar1*, *Vulmar1* and *Psmar1*, have been characterized in rice (Tarchini et al., 2000), soybean (Jarvick & Lark, 1998), sugar beet (Jacobs et al., 2004), and the pea (Macas et al., 2005) respectively. Moreover, using the degenerated primers *MLE5A* and *MLE3A*, numerous partial *MLE* sequences have been identified by Feschotte and Wessler (2002) in a wide variety of terrestrial plant species. Some partial *MLE* sequences have also been identified in 33 bamboo species (Feschotte & Wessler, 2002; unpublished data Tang, 2006), suggesting that *MLEs* are widespread in terrestrial plants.

The partial *MLE* sequences obtained in this study are the first to have been identified in marine diatoms. We were able to characterize *MLE* transposase fragments in 10 of the 13 diatom species tested using degenerated primers. These primers make it possible to amplify the coding region for the transposase located between the two first aspartate residues (D) of the DDD catalytic domain. This signature motif is common in the *mariner*-like family, and it is highly conserved (Shao & Tu, 2001). This aspartate core is essential for catalyzing the transposition involving the excision and subsequent re-integration of the transposon (Capy et al., 1997; Lohe et al., 1997).

From the 10 diatom species tested, we obtained 55 nucleic sequences and 25 amino acid consensus sequences corresponding to *MLEs*. As about 75% of the diatoms studied contained *mariner*-like encoding sequences in their genomes, it seems likely that *MLEs* are widespread in marine diatom genomes. However, the failure to detect *MLEs* in other diatom species, such as *T. pseudonana*, does not necessarily mean that these genomes contain no *MLE*, but could be due to the use of inappropriate degenerated primers. In order to test this hypothesis, more specific primers will have to be designed on the basis of diatom *MLE* sequences.

Pairwise comparisons and the alignment of diatom transposases reveal a high level of heterogeneity among the sequences. Nevertheless the presence of some conserved amino acid residues between *MLEs* of marine diatoms and plants confirm that diatom transposases all belong to the Tc1-*mariner* superfamily. In fact,

partial *MLE* sequences amplified in marine diatoms using *MLE5A* and 3A displayed the typical features (KVMFL, A(x)ARPR, (F/W)DGKLG IWP, and the KWP) encountered in most terrestrial plant *MLEs* (fig 5). In the full-length *mariner* elements *Osmar1*, 2 and *Soymar1*, which are included in the DD39D *mariner* group (Brillet et al., 2007), these four patterns show slight differences (one or two amino acids may be missing). In contrast, these patterns are not found in the *vulmar1* sequence. The NCBI BLAST comparisons were done on the matching region including the first 2 Ds of the presumed DD39D transposase catalytic domain. Taken together, these observations suggest that the marine diatom *MLE* transposases could well belong to the DD39D group.

A very high degree of pairwise identity was found between the amino acid sequences of *N. laevis* and *E. paludosa* (80% identity between *Entopalmarcons2* and *Entopalmarcons4*, ~ 65% identity between *Entopalmarcons3* and *Entopalmarcons2* and 4, and 70 to 90% identity between *Nilamarcons3*, 4 and 5). This high degree of conservation could reflect either the amplification of a single *MLE* subfamily, which has thus become prevalent in the host genomes, or could be the result of a recent invasion of these genomes by several very similar *MLE* transposons. In contrast, the sequences can differ widely in a given species (for example, in *A. coffeaeformis*, the consensus sometimes shares a higher percentage of identity with the consensus of other species, Table 2), and it is possible that the degenerated primers allow the amplification of the more mutated fragments of *MLEs* in the diatoms to occur.

In most of the species studied, each PCR product cloned and sequenced came from a different PCR reaction, meaning that each sequence could correspond to a different gene encoding transposase, which implies that almost all the sequences obtained were unique. Similar results have also been reported in terrestrial plants (Feschotte & Wessler, 2002), in the genus *Beta* (Jacobs et al., 2004), and for the *Ty-copia*-like retrotransposons (Kumar & Bennetzen, 1999). Indeed no complete gene-encoding transposase sequences has so far been identified, and if these results were attributable to a recent invasion, we would expect to find some active elements still present in the host genomes.

Moreover, in some cases, the divergences are greater between sequences isolated from the same species (*A. coffeaeformis* and *H. crucigera*) than between sequences from different species, which is not consistent with the hypothesis of a recent genome invasion of the host genome. Similar results have been reported in coastal crabs by Bui et al. (2008). This type of *MLE* distribution has often been reported, and is the result of equilibrium between two kinds of regulatory mechanism : i) *MLE* diatom sequences could result from a pre-existing transposase lineage followed by diversification events and vertical inheritance (Lohe et al., 1995; Le Rouzic & Capy, 2006) in which excessive sequence divergences are the result of stochastic losses. An example of this is

seen in *mos1* in *Drosophila melanogaster* (Capy et al., 1992; Hartl et al., 1997), in which different subfamilies of *mariner*-like elements co-exist in the same species (Feschotte & Wessler, 2002; Jacobson et al., 2004, Carr, 2007). The other mechanism that could lead to a distribution of this sort would be ii) differences in diatom sequences as a result of horizontal transfer events (Lohe et al., 1995; Robertson & Lampe, 1995; Robertson, 1997; Diao et al., 2006; Bui et al., 2008; Bowler et al., 2008). This assumption is reinforced by the ecosystem in which the diatoms live : permanent contact with viruses and bacteria in seawater could have the effect of promoting gene transfer. Evidence of horizontal transfer between marine organisms has already been observed for *MLEs* (Casse et al., 2006; Bui et al. 2008; Fortune et al., 2008). The horizontal transfer phenomenon is also posited to have occurred during the evolutionary history of the model diatom *P. tricornutum*, and to explain the heterogeneity of the sequences observed. A large number of predicted genes are thought to have been derived from this kind of transfer between bacteria and diatoms (Bowler et al., 2008). It is assumed that horizontal transfer is more common in diatoms than in other eukaryotes, and it could have constituted the major driving force during diatom evolution (Bowler et al., 2008). Moreover, horizontal transfer has already been reported to occur in the *mariner* subfamily (Robertson & Lampe, 1995; Robertson, 1997; Casse et al., 2006, Bui et al., 2008). It is not possible to confirm either hypothesis without knowing the complete sequences of the transposases and the phylogeny of the host, and it is more likely that both these mechanisms have contributed to the regulation and the evolution of diatom genomes. Genetic studies of diatoms are beginning, and the annotation of the fully-sequenced genomes of *P. tricornutum* and *T. pseudonana* will provide a useful framework for further studies of the molecular evolution of diatom. More complete sequences of *MLE* transposases are currently under investigation.

Amplified fragments obtained in this study were similar to those belonging to the terrestrial plant *MLE* group also known as *ITm* D39D (Shao & Tu, 2001; Brillet et al., 2007). The relationships between diatom *MLE* sequences and terrestrial plant *MLE* transposases investigated by constructing phylogenetic trees using three different methods clearly show that diatom *MLE* sequences do not cluster in the group of terrestrial plant *MLEs* referenced in databases. Diatom *MLEs* are distributed in three sister groups, designed *MLE-Di1*, 2 and 3, which are supported by high bootstrap robustness with all the three methods. The BI analyses revealed the best resolved tree. The lengths of the branches in the group *MLE-Di1* suggested that the *MLE* consensus sequences from *N. palea*, *S. constricta* have diverged and evolved faster than the other diatom *MLEs* resulting in the clustering of these sequences. This is consistent with the diatom *MLEs* alignment (Fig 2) and phylogeny (Fig 3), in which these sequences appear to be very different and isolated, suggesting that they have probably more undergone

greater mutation than the other diatom *MLEs*. This result does not reflect the putative ancestral state of the diatom host, since *Staucmarcons1*, *Staucmarcons2*, *Nipamarcons1* and *Nipamarcons2* were isolated from pennate diatom whereas the *Thaweismarcons1* branch is shorter although isolated from a centric diatom *T. weissflogii*, which is thought to belong to the ancestral group of diatoms (Gersonde & Harwood, 1990 ; Medlin et al., 1993).

Phylogenetic trees indicate that plant *MLEs* form a monophyletic group as described by Feschotte & Wessler, (2002). The four main lineages of plant *MLEs* were detected in our analysis.

If more data from some species that display greater divergences (*N. palea*, *S. constricta*, *T. weissflogii*) were to be included, these sequences could be brought closer to those of other Diatom sequences revealing the monophyly of this marine diatom group. The identification of complete or additional sequences in diatoms could allow us to assign *Acomarcons1* and *Odaumarcons1* to the *MLE-Di2* diatom cluster, because different *MLE* subfamilies have probably been amplified. In contrast, even if new sequences are added, it seems unlikely that the diatom sequence clusters would form a single monophyletic group with the *MLE* sequences of plants, the conserved features of which are domains with functions that so far remain unknown.

This divergence between marine diatom *MLEs* and plant *MLEs* is not surprising, because of differences between the evolution of marine diatoms and that of terrestrial plants. Three main factors must be taken into account : the environment (marine *versus* terrestrial), cellular organization (unicellular versus multicellular), and the origin of diatoms that derived from a secondary endosymbiosis (Falkowski et al., 2004) and have been less studied than green algae or human parasites (Baldauf, 2003; Monstant et al., 2005). Furthermore, the comparison of molecular gene diversification in diatoms and in metazoans has highlighted rapid rates of diversification in gene expansion, differences in the equilibrium between the gain and loss of introns, and in the mobilization of transposable elements in the diatom genome of *P. tricornutum* (Bowler et al., 2008).

Our study reveals for the first time the presence of *MLE* transposons in marine diatoms. Most of the diatoms tested contained fragments of *MLE* transposons, probably belonging to several different subfamilies. The NCBI blast correspondences between the *MLE* sequences of diatoms and terrestrial plants were located on the catalytic region, between the first two aspartic amino acid of the DD39D motif of the plant transposon catalytic triad. This finding clearly shows that the marine diatom *MLE* transposases belong to the Tc1-*mariner* superfamily (Wicker et al., 2007), and suggests they could belong to the DD39D group. Although diatom *MLE* sequences are similar to those of terrestrial plants, it has not yet been confirmed that they belong to the DD39D group. The phylogenetic relationships between marine diatom *MLEs* and plant *MLEs* show well defined three

sister diatom *MLE* groups and the monophyly of plant *MLE* group. This finding suggests that the *MLE* were present in the basis of the plant lineages then, identification of *MLEs* in others lineages as macroalgae, bryophytes, etc, will be relevant to explore the evolutionary relationships between *MLEs* in vegetal lineages.

## References

Abascal, F., Zardoya, R., Posada, D., 2005. ProtTest : Selection of best-fit models of protein evolution. *Bioinformatics* 21, 9, 2104-2105.

Armbrust, E. V., Berges J. A., Bowler C., Green B. R., Martinez D., Putnam N. H., Zhou S., Allen A. E., Apt K. E., Bechner M., Brzezinski, M. A., Chaal, B. K., Chivittii, A., Davis, A. K., Demarest, M. S., Detter, J. C., Glavina, T., Goodstein, D., Hadi, M. Z., Hellsten, U., Hildebrand, M., Jenkins, B. D., Jurka, J., Kapitonov, V. V., Kröger, N., Lau, W. W., Lane, T. W., Larimer, F. W., Lippmeier, J. C., Lucas, S., Medina, M., Monstant, A., Obornik, M., Parker, M. S., Palenik, B., Pazour, G. J., Richardson, P. M., Ryneerson, T. A., Saito, M. A., Schwartz, D. C., Thamatrakoln, K., Valentin, K., Vardi, A., Wilkerson, F. P., Rokhsar, D. S., 2004. The genome of the diatom *Thalassiosira pseudonana* : ecology, evolution, and metabolism. *Science* 306, 79-86.

Augé-Gouillou, C., Bigot, Y., Pollet, N., Hamelin, M. H., Meunier-Rotival, M., Periquet, G., 1995. Human and other mammalian genomes contain transposons of the mariner family. *FEBS Lett.* 368, 3, 541-546.

Baldauf, S. L., 2003. The deep Roots of eukaryotes. *Science* 300, 1703-1706.

Bennetzen, J. L., Ma, J., Devos, K. M., 2002. Mechanisms of recent genome size variation in flowering plants. *Ann. Bot.* 95, 127-132.

Bennetzen, J. L., 2005. Transposable elements, gene creation and genome rearrangement in flowering plants. *Curr. Opin. Genet. Dev.* 15, 621-627.

Bowler, C., Allen, A., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., Maheswari, U., Martens, C., Maumus, F., Otiillard, R. P., Rayko, E., Salamov, A., Vandepoele, K., Beszteri, B., Gruber, A., Heijde, M., Katinka, M., Mock, T., Valentin, K., Verret, F., Berges, J. A., Brownlee, C., Cadoret, JP., Chiovitti, A., Choi, C. J., Coesel, S., De Martino, A., Detter, J. C., Durkin, C., Falciatore, A., Fournet, J., Haruta, M., Huysman, M. J. J., Jenkins, B. D., Jiroutova, K., Jorgensen, R. E., Joubert, Y., Kaplan, A., Kröger, N., Kroth, P. G., La Roche, J., Lindquist, E., Lommer, M., Martin-Jézéquel, V., Lopez, P., Lucas, S., Mangogna, M., McGinnis, K., Medlin, L. K., Monstant, A., Oudot-Le Secq, M. P., Napoli, C., Obornik, M., Schnitzler Parker, M., Petit, J. L., Porcel, B. M., Poulsen, N., Robison, M., Rychlewski, L., Ryneerson, T. A., Schmutz, J., Shapiro, H., Siaux, M., Stanley,

M., Sussman, M. R., Taylor, A. R., Vardi, A., Dassow, P. V., Vyverman, W., Willis, A., Wyrwicz, L. C., Rokhsar D. S., Weissenbach, J., Armbrust, E. V., Green, B. R., Van De Peer, Y., Grigoriev, I. V., 2008. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 406, 239-243.

Brillet, B., Bigot, Y., Augé-Gouillou, C., 2007. Assembly of the Tc1 and *mariner* transposition initiation complexes depends on the origins of their transposase DNA binding domains. *Genetica* 130, 105-120.

Bui, Q. T., Delaurière, L., Casse, N., Nicolas, V., Laulier, M., Chénais, B., 2007. Molecular characterization and phylogenetic position of a new *mariner*-like element in the coastal crab, *Pachygrapsus marmoratus*. *Gene* 396, 2, 248-256.

Bui, Q. T., Casse, N., Leignel, V., Nicolas, V., Chénais, B., 2008. Widespread occurrence of *mariner* transposons in coastal crabs. *Mol. Phylogenet. Evol.* 47, 1181-1189.

Capy, P., Koga, A., David, J. R., Hartl, D. L., 1992. Sequence analysis of active *mariner* elements in natural populations of *Drosophila simulans*. *Genetics* 130, 499-506.

Capy, P., Langin, T., Bigot, Y., Brunet, F., Daboussi, M. J., Periquet, G., David, J. R., Hartl, D. L., 1994. Horizontal transmission versus ancient origin : *mariner* in the witness box. *Genetica* 93 (1-3), 161-170.

Capy, P., Langin, T., Higuët, D., Maurer, P., Bazin, C., 1997. Do the integrase of LTR-retrotransposons and class II element transposase have a common ancestor? *Genetica* 100 (1-3), 63-72.

Casse, N., Pradier, E., Demattei, M. V., Bigot, Y., Laulier, M., 2002. *Mariner* transposons are genetics parasites occurring in the genome of hydrothermal invertebrates. *Cah. Biol. Mar.* 43, 325-328.

Casse, N., Bui, Q. T., Nicolas, V., Renault, V., Bygot, Y., Laulier, M., 2006. Species sympatry and horizontal transfers of *mariner* transposons in marine crustacean genomes. *Mol. Phylogenet. Evol.* 40, 609-619.

Carr, M., 2008. Multiple subfamilies of *mariner* transposable elements are present in stalk-eyed flies (Diptera : diopsidae). *Genetica.* 132, 2, 113-122

Crooks, G. F., Hon, G., Chandonia, J. M., Brenner, S. E., 2004. Weblogo : a sequence logo generator. *Genome Res.* 14, 1188-1190.

Diao, X., Freeling, M., Lisch, D., 2006. Horizontal transfer of a plant transposon. *PLoS Biol.* 4, 1, 0119-0128.

Doyle, J. J., Doyle, J. L., 1990. Isolation of plant DNA from fresh tissue. *Focus* 12, 13-15.

Falkowski, P. G., Katz, M. E., Quigg, A., Raven, J. A., Schofield, O., Taylor, F. J. R., 2004. The evolution of modern eukaryotic phytoplankton. *Science* 305, 354-360.

Feschotte, C., Wessler, S. R., 2002. *Mariner*-like transposases are widespread and diverse in flowering plants. *Proc. Natl. Acad. Sci.* 99, 280-285.

Feschotte, C., Jiang, N., Wessler, S. R., 2002. Plant transposable elements : where genetics meet genomics. *Genetics* 3, 320-341.

Feschotte, C., Osterlund, M. T., Peeler, R., Wessler, S. R., 2005. DNA-binding specificity of rice *mariner*-like transposases and interaction with Stowaway *MITEs*. *Nucleic Acids Res.* 33, 2153-2165.

Finnegan, D. J., 1992. Transposable elements. *Curr. Opin. Genet. and Dev.* 2, 821-867.

Fortune, P. M., Roulin, A., Panaud, O., 2008. Horizontal transfer of transposable elements in plants. *Commu. Int. Biol.* 1,1, 1-4.

Gersonde, R., Harwood, D. M., 1990. Lower Cretaceous Diatoms from ODP Leg 113 Site 693 (Weddell sea). Part I. Vegetative cells. In *Proceedings of the Ocean Drilling Program. 113* (Baker, P. f., Kennett, J. P., et al., editors). 365-402.

Halaima-Toumi, N., Casse, N., Demattei, M. V., Renault, S., Pradier, E., Bigot, Y., Laulier, M., 2004. The GC-rich transposon *Bythmar1* from the deep-sea hydrothermal crab, *Bythogrea thermydron*, may encode three transposase isoforms from a single ORF. *Mol. Biol. Evol.* 59, 6, 747-760.

Hall, T. A., 1999. BioEdit : a user-friendly biological sequences alignment editor and analysis program for Windows 95/98/NT. *Nucleic. Acids Symp. Ser.* 41, 95-98.

Harrison, P. J., Waters, R. E., Taylor, F. J. R., 1980. A broad spectrum artificial seawater medium for coastal and open ocean phytoplankton. *J. Phycol.* 16, 28-35.

Hartl, D. L., Lozovskaya, E. H., Nurminsky, D. L., Lohe, A. R., 1997. What restricts the activity of *mariner*-like transposable elements? *Trends Genet.* 13, 5, 197-201.

Hua Van, A., Le Rouzic, A., Maisonhaute, C., Capy, P., 2005. Abundance, distribution and dynamics of retrotransposable elements and transposons : similarities and differences. *Cytogenet. Genome Res.* 110, 426-440.

Jacobs, G., Dechyeva, D., Menzel, G., Dombrowski, C., Schmidt, T., 2004. Molecular characterisation of *vulmar1*, a complete *mariner* transposon of sugar beet and diversity of *mariner*- and *En/Spm*-like sequences in the genus *Beta*. *Genome* 47, 1192-1201.

Jacobson, J. W., Medhora, M. M., Hartl, D., L., 1986. Molecular structure of a somatically unstable transposable element in *Drosophila*. *Proc. Natl. Acad. Sci.* 83, 8684-8688.

Jarvick, T., Lark, K. G., 1998. Characterization of *Soymar1*, a *mariner* element in soybean. *Genetics* 149, 1569-1574.

- Kumar, A., Bennetzen, J. L., 1999. Plant retrotransposons. *Annu. Rev. Genet.* 33, 479-532.
- Le Rouzic, A., Capy, P., 2006. Population genetics models of competition between transposable elements subfamilies. *Genetics* 174, 785-793.
- Lohe, A. R., Moriyama, E. N., Lidholm, D. A., Hartl, D., 1995. Horizontal transmission, vertical inactivation, and stochastic loss of *mariner*-like transposable elements. *Mol. Biol. Evol.* 12, 1, 62-72.
- Lohe, A. R., De Aguiar, D., Harl, D., 1997. Mutation in the *mariner* transposase : the DD(35)E consensus sequence is non functional. *Proc. Natl. Acad. Sci. USA* 94, 1293-1297.
- Lopez, P., Descles, J., Allen, A. E., Bowler, C., 2005. Prospects in diatom research. *Curr. Opin. Biotechnol.* 16, 180-186.
- Macas, J., Klobitzkova, A., Neumann, P., 2005 Characterization of Stowaway MITEs in Pea (*Pisum sativum* L.) and identification of their potential master elements. *Genome* 48, 831-839.
- Maumus, F., Allen, A. E., Mhiri, C., Hu, H., Jabbari, K., Vardi, A., Grandbastien, M.-A., Bowler, C., 2009. Potential impact of stress activated retrotransposons on genome evolution in marine diatom. *BMC Genomics* 10 : 624.
- Medlin, L. K., Williams, D. M., Sims, P. A., 1993. The evolution of the diatoms (Bacillariophyta). I. Origin of the group and assessment of the monophyly of its major subdivisions. *Eur. J. Phycol.* 28, 261-275.
- Montsant, A., Jabbari, K., Maheswari, U., Bowler C., 2005. Comparative genomics of the pennate Diatom *Phaeodactylum tricornutum*. *Plant Physiol.* 137, 500-513.
- Raven, J. A., Waite, A. M., 2004. The evolution of silicification in diatoms : inescapable sinking and sinking as escape? *New Phytol.* 162, 45-61.
- Robertson, H. M., 1993. The *mariner* transposable element is widespread in insects. *Nature* 362, 241-245.
- Robertson, H. M., Lampe, D. J., 1995. Recent horizontal transfert of a *mariner* transposable element among and between Diptera and Neuroptera. *Mol. Biol. Evol.* 12, 5, 125-139.
- Robertson, H. M., 1997. Multiple *mariner* transposons in flatworm and hydras are related to those of insects. *J. Hered.* 88, 195-201.
- Robertson, H. M., Asplund, M. L., 1996. *BmmarI* : a basal lineage of *mariner* family of the transposable element silkworm moth, *Bombyx mori*. *Insect. Biochem. Mol. Biol.* 26(8-9), 945-954.
- Ronquist, F., Huelsenbeck, J. P., 2003. Mr Bayes 3 : Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572-1574.

Round, F. E., Crawford, R. M., Mann, D. G., 1990. The diatoms. Cambridge : Cambridge university press, 747p.

Scala, S., Carels, N., Falciatore, A., Chiusano, L. C., Bowler, C., 2002. Genome properties of the diatom *Phaeodactylum tricornutum*. Plant Physiol. 129, 993-1002.

Scala, S., Bowler, C., 2001. Molecular insights into the novel aspects of diatom biology. Cell. Mol. Life Sci. 58, 1666-1673.

Shao, H., Tu, Z., 2001. Expanding the diversity of the IS630- Tc1-mariner Superfamily : discovery of a unique DD37E transposon and reclassification of DD37D and DD39D transposons. Genetics 159, 1103-1115.

Swofford, D., 2002. Paup (Phylogenetic Analysing Using Parsimony). Sinauer Associates, Sunderland Massachusetts, USA.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4 : Molecular Evolutionary Genetics Analysis (MEGA) software versions 4.0. Mol. Biol. Evol. 24, 1596-1599.

Tarchini, R., Biddle, P., Wineland, R., Tingey, S., Rafalski, A., 2000. The complete sequence of 340 kb of DNA around the rice Adh1-Adh2 region reveals interrupted colinearity with maize chromosome 4. Plant Cell. 12, 381-392.

Thompson, J.D., Higgins, D.G., and Gibson, T.J., 1994. CLUSTAL W : improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673-4680.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., San Miguel, P., Schulman, A., H., 2007. An unified classification system for eukaryotic transposable elements. Nat. Rev. Genet. 8, 973-982.

Wiebe, H. C., Kooistra, F., Medlin, L. K., 1996. Evolution of the diatom (Bacillariophyta). IV.A Reconstruction of their age from Small Subunit rRNA coding regions and the fossil record. Mol Phylogenet. Evol 6, 3, 391-407.

Williams, D., 2007. Diatom phylogeny : Fossils, molecules and the extinction of evidence. C. R. Palevol. 6, 505-514.

### **Figure 1. *MLE* transposase fragments amplified by PCR**

PCR were performed using the degenerated primers *MLE5A* and *MLE3A* (Feschotte & Wessler, 2002) and allowed to amplify approximately 380-bp fragments of *MLE* transposase genes. Lane M : molecular weight marker 100-bp DNA ladder, Lane 1 : *Amphora acutiuscula*, Lane 2 : *Amphora coffeaeformis*, Lane 3 : *Entomoneis paludosa*, Lane 4 : *Haslea crucigera*, Lane 5 : *Nitzschia laevis*, Lane 6 : *Nitzschia palea*, Lane 7 : *Odontella aurita*, Lane 8 : *Phaeodactylum tricornutum*, Lane 9 : *Stauroneis constricta.*, Lane 10 : *Thalassiosira weissflogii* , Lane C : negative control (DNA-free water).

### **Figure 2. Diatom *MLE* sequence alignment**

Diatom protein sequences were obtained by conceptual translation from the nucleic consensus sequences. The alignment was performed with BioEdit, and visualized using GENEDOC programs. Black indicates 100% identity, dark gray 80% identity, and pale gray an identity of between 50 and 80%. The motifs used to design the primers are underlined. The main diatom amino acid patterns are boxed.

### **Figure 3. Phylogenetic relationships between diatom *MLE* transposase fragments**

A tree was generated using the amino acid consensus sequences of the ten diatom species processing *MLE* transposases. This tree was generated using a MP method. The MP bootstrap values supporting this analysis are indicated in regular font, except when the values were < 50%. The time used for the MP method was 0.06 sec. Diatom sequence accession numbers are indicated in Table 1. The circles indicate sequences assigned to the same cluster. Partial sequences from the two complete plant mariner elements, *Osmar1* (*Oryza sativa mariner* element 1, Tarchini et al., 2000) and *Soymar1* (*Soybean mariner* element 1, Jarvick & Larck, 1998), were used for comparison.

### **Figure 4. Phylogenetic relationships between *MLE* transposase fragments of marine diatoms versus those of terrestrial plants**

The tree was generated by comparing the *MLE* consensus sequences of diatoms (Table 1 and Fig 2) to those of terrestrial plants from the NCBI database. The diatom *MLE* sequence clusters are represented as bold branches and all the diatom *MLE* sequences are highlighted in gray. Three lineages were designated as *MLE-Di1*, 2 and 3. The clade containing terrestrial plant sequences is indicated by an arrow and all the plant *MLEs* are hatched, the previously detected lineages (Feschotte & Wessler, 2002) are indicated by brackets. All accession numbers for

plant *MLEs* are given near to the species. The terrestrial plant sequences consisted of bamboo sequences and plant sequences from Feschotte and Wessler 2002 (indicated by the code used in the paper, e.g : *Hordeum vulgare* Hv1828). We also used the sequences extracted from three complete *mariner* elements *Osmar1*, *Soymar1*, *Vulmar1* and *Psmar1* : these sequences are indicated in bold. To root this tree, we used a cluster constitute by animal *MLE* sequences, *Mos1* (*Drosophila melanogaster mariner* element) and *Hsmar1* (*Homo sapiens mariner* element). This tree was generated by Bayesian inference method with Mr Bayes and MEGA4 software (Tamura et al., 2007). The BI posterior probability values are underlined for the Bayesian method and completed with the bootstrap values of the main nodes supporting this analysis from NJ and MP methods indicated in bold and in italics respectively. Values under 50% are not indicated.

### **Figure 5. Conserved motifs in *MLE* sequences of diatoms and plants.**

The figures were constructed using the Weblogo website ([http : //weblogo.berkeley.edu/logo.cgi](http://weblogo.berkeley.edu/logo.cgi)). The biggest letters correspond to the motifs most frequently conserved in sequences. The *MLE* primers are boxed, and the relatively highly-conserved motifs are indicated in brackets. They are referred to by numbers : 1 for KVMFL, 2 for A(x)ARPR, 3 for (F/W)DGKLG IWP, and 4 for the KWP motif.

### **Table 1. Diatom species and origins, and detection of the presence of *mariner* transposase**

The species from Vietnam are identified by an \*. The two groups of diatom are distinguished by superscript symbols, <sup>p</sup> for pennate and <sup>c</sup> for centric diatom. For each species, at least five fragments corresponding to *MLE* transposases were sequenced. Most of these fragments exhibit an open reading frame (ORF). Sequences are referenced in NCBI with accession numbers. The consensus performed from these sequences are shown in Table 2.

### **Table 2. Pairwise divergence matrix of the diatom amino acid consensus sequences**

The homology is expressed as a percentage. The sequences were aligned using CLUSTAL W software, and the matrix was performed with GENEDOC using default parameters. The *MLE* consensus shown correspond to the following species : *Acomarcons*, *Amphora coffeaeformis*; *Odaumarcons*, *Odontella aurita*; *Entopalmarcons*, *Entomoneis paludosa*; *Hcrumarcons*, *Haslea crucigera*; *Phatrimarcons*, *Phaeodactylum tricornutum*; *Aacumarcons*, *Amphora acutiuscula*; *Nilamarcons*, *Nitzschia laevis*; *Nipamarcons*, *Nitzschia palea*; *Staucumarcons*, *Stauroneis constricta*; *Thaweismarcons*, *Thalassiosira weissflogii*. The clusters indicated in Table 2 are also shown in Figure 3.

### **Acknowledgements**

We thank the reviewers for their stimulating remarks, which have helped us to improve our findings. We thank V. Leignel for providing us helpful discussions. The English text was revised by Monika Ghosh. The studies were funded by PhD grants from the French regions Pays de Loire and Centre.

Figure 1.

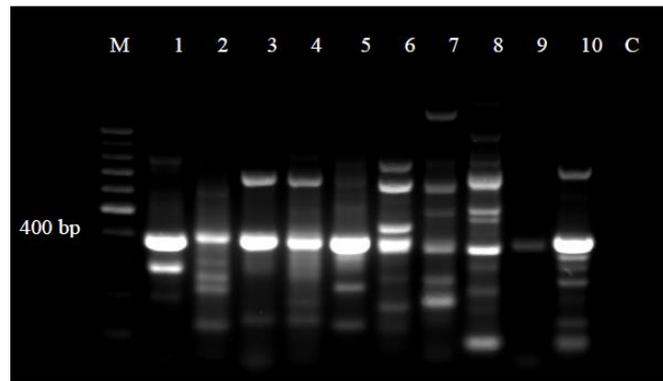


Figure 2

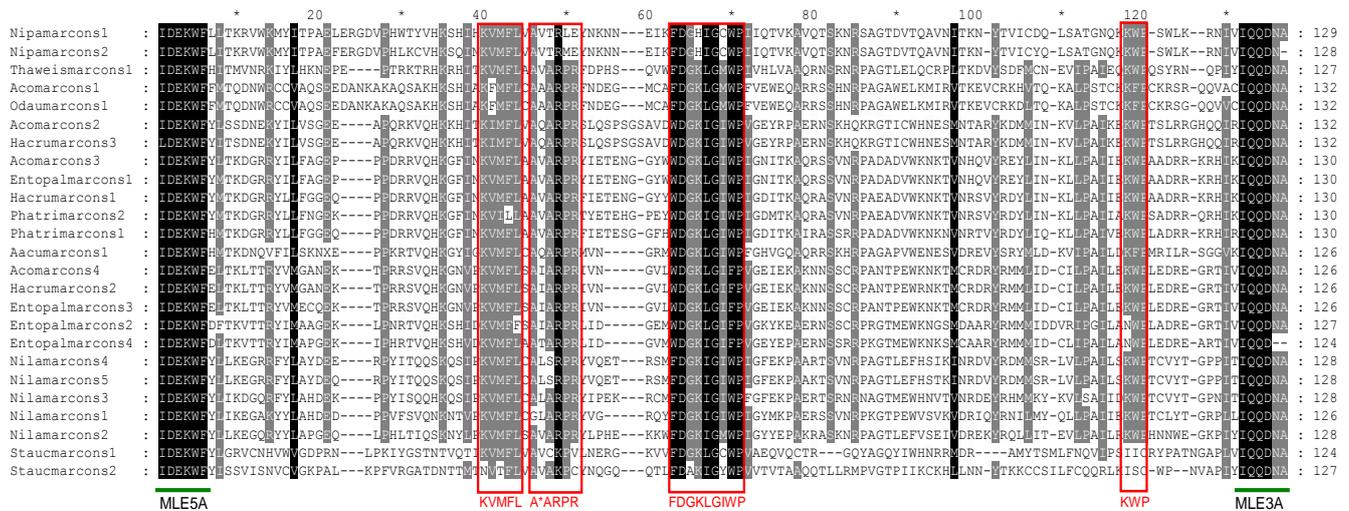


Figure 3

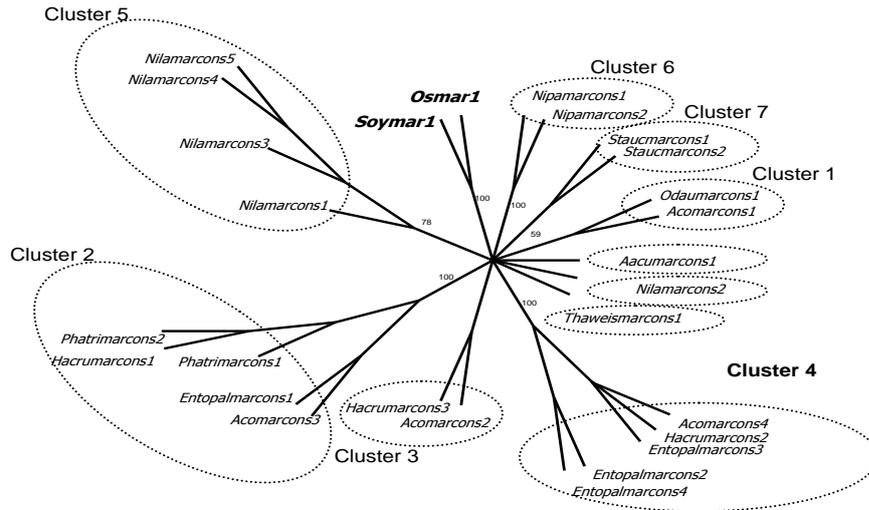


Figure 4.

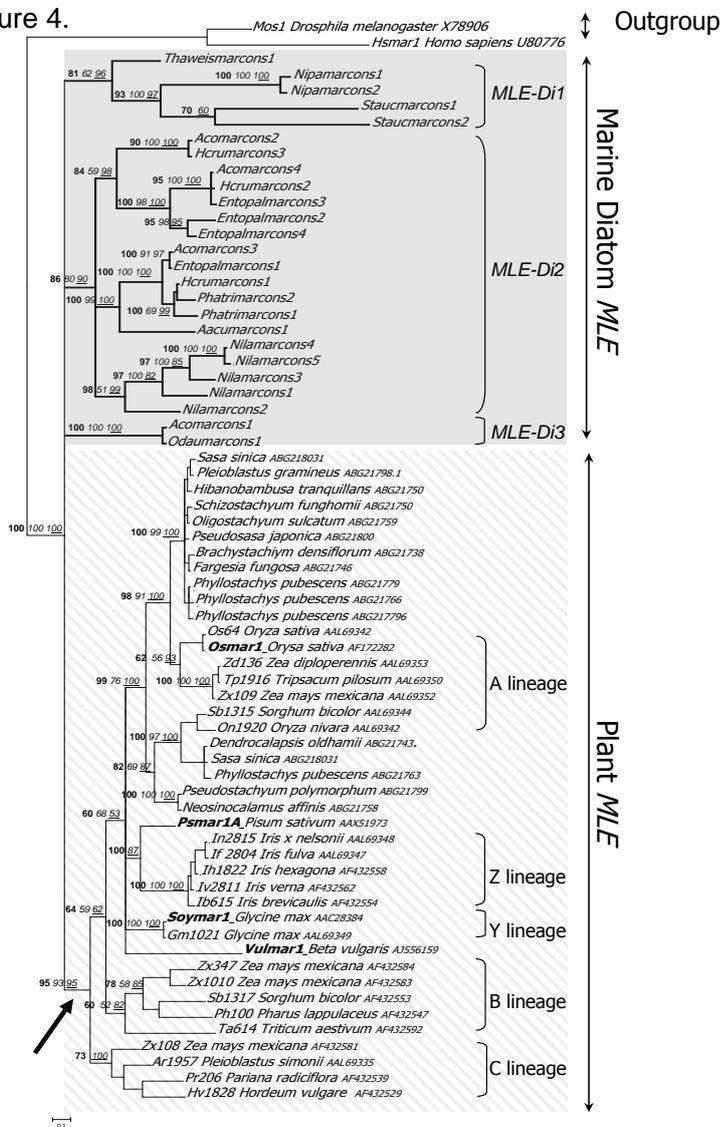


Figure 5.

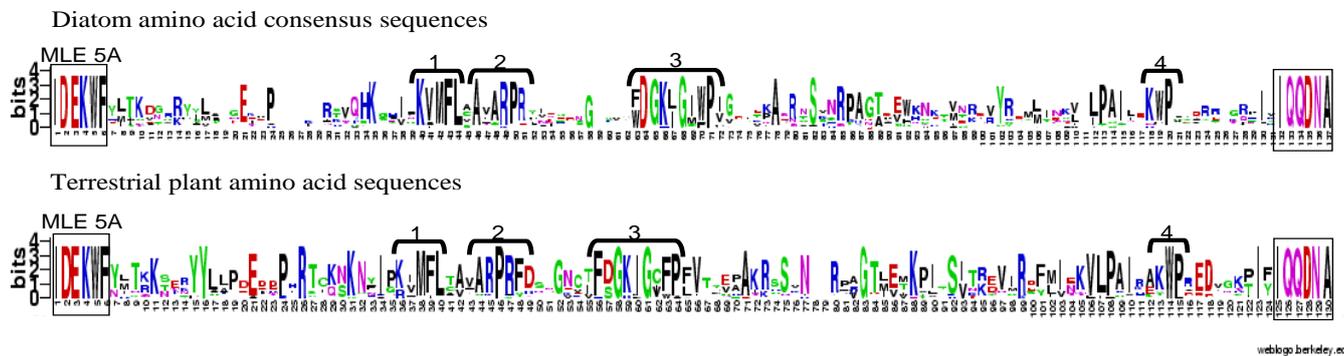


Table1

Species	Number of 380-bp <i>MLE</i> fragments	Number of ORFs in initial nucleotide sequences	Number of <i>MLE</i> consensus sequences	Accession numbers
1) <i>Amphora acutiuscula</i> * <sup>P</sup>	6	5	1	FN546185 - FN546190
2) <i>Amphora coffeaeformis</i> <sup>P</sup>	6	5	4	FN546191 - FN546196
3) <i>Cylindrotheca fusiformis</i> <sup>P</sup>	0	0	0	.....
4) <i>Cylindrotheca closterium</i> <sup>P</sup>	0	0	0	.....
5) <i>Entomoneis paludosa</i> <sup>P</sup>	5	5	4	FN546197 - FN546201
6) <i>Haslea crucigera</i> <sup>P</sup>	5	5	3	FN546202 - FN546206
7) <i>Nitzschia laevis</i> <sup>P</sup>	6	6	5	FN546207 - FN546212
8) <i>Nitzschia palea</i> * <sup>P</sup>	5	5	2	FN546213 - FN546217
9) <i>Odontella aurita</i> <sup>C</sup>	4	1	1	FN546218 - FN546221
10) <i>Phaeodactylum tricornutum</i> <sup>P</sup>	5	5	2	FN546222 - FN546226
11) <i>Stauroneis constricta</i> <sup>P</sup>	7	2	2	FN546227 - FN546233
12) <i>Thalassiosira pseudonana</i> <sup>C</sup>	0	0	0	.....
13) <i>Thalassiosira weissflogii</i> <sup>C</sup>	6	6	1	FN546234 - FN546239
<b>total</b>	<b>55</b>	<b>45</b>	<b>25</b>	

**Table 2. Pairwise divergence matrix of the diatom amino acid consensus sequences**

	Cluster 1		Cluster 2				Cluster 3				Cluster 4				Cluster 5				Cluster 6		Cluster 7				
	<i>Acomarcons1</i>	<i>Odaumarcons1</i>	<i>Acomarcons3</i>	<i>Entopalarcons1</i>	<i>Hacumarcons1</i>	<i>Phatrimarcons1</i>	<i>Phatrimarcons2</i>	<i>Aacumarcons1</i>	<i>Acomarcons2</i>	<i>Hacumarcons3</i>	<i>Acomarcons4</i>	<i>Hacumarcons2</i>	<i>Entopalarcons3</i>	<i>Entopalarcons2</i>	<i>Entopalarcons4</i>	<i>Nilamarcons4</i>	<i>Nilamarcons5</i>	<i>Nilamarcons3</i>	<i>Nilamarcons1</i>	<i>Nilamarcons2</i>	<i>Thaweismarcons1</i>	<i>Nipamarcons1</i>	<i>Nipamarcons2</i>	<i>Staucumarcons1</i>	<i>Staucumarcons2</i>
<i>Acomarcons1</i>		96	34	35	35	35	32	36	27	28	29	28	29	28	30	32	30	35	29	31	36	28	27	23	18
<i>Odaumarcons1</i>			34	35	35	35	32	36	26	27	30	29	30	28	30	32	30	35	30	32	36	27	26	24	18
<i>Acomarcons3</i>				99	89	81	80	49	45	45	51	50	50	45	47	43	42	47	46	45	41	28	28	26	20
<i>Entopalarcons1</i>					90	82	81	50	44	45	50	50	45	46	42	41	46	45	44	41	27	28	25	20	
<i>Hacumarcons1</i>						88	88	50	44	45	50	50	43	45	43	43	46	45	45	41	26	27	26	19	
<i>Phatrimarcons1</i>							78	46	42	43	47	46	47	42	45	42	41	45	44	44	38	26	27	22	19
<i>Phatrimarcons2</i>								49	43	44	51	50	51	45	48	40	40	43	41	44	39	26	27	22	18
<i>Aacumarcons1</i>									42	43	41	41	40	42	40	40	38	43	37	39	41	23	22	24	17
<i>Acomarcons2</i>										98	41	41	41	45	48	40	39	43	39	39	39	24	23	28	21
<i>Hacumarcons3</i>											41	41	41	46	48	40	38	42	38	38	40	24	24	27	21
<i>Acomarcons4</i>												99	97	63	68	40	40	43	43	43	35	22	20	25	20
<i>Hacumarcons2</i>													96	62	67	40	40	43	43	43	35	22	20	25	19
<i>Entopalarcons3</i>														62	68	40	40	43	43	34	22	20	24	19	
<i>Entopalarcons2</i>															80	39	38	42	37	41	35	23	20	26	17
<i>Entopalarcons4</i>																40	39	43	41	44	37	26	23	26	21
<i>Nilamarcons4</i>																	97	72	58	56	41	28	27	23	21
<i>Nilamarcons5</i>																		71	57	56	40	28	27	23	21
<i>Nilamarcons1</i>																			60	50	38	28	28	25	19
<i>Nilamarcons3</i>																				54	33	28	28	24	21
<i>Nilamarcons2</i>																					41	29	28	27	19
<i>Thaweismarcons1</i>																						32	33	26	26
<i>Nipamarcons1</i>																							92	19	19
<i>Nipamarcons2</i>																								19	19
<i>Staucumarcons1</i>																									24
<i>Staucumarcons2</i>																									

# COMMUNICATIONS ET PUBLICATIONS

---

## Publication

---

**2011 Hermann D.**, Egue F., Caruso A., Jacquette B., Morant-Manceau A., Casse N.  
“Thermal stress expression of *Mariner*-like transposable element in *Phaeodactylum tricornutum*”  
*European Journal of Phycology*, en préparation.

### ***Publication soumise dans une revue scientifique avec comité de lecture***

**2010 Hermann D.**, Caruso A., Bui Q-T, Jacquette B., Morant-Manceau A., Casse N.  
“Characterization of *Mariner*-like transposons in marine diatoms”  
*Molecular Phylogenetics and Evolution*,  
Re-soumise après révision, chez l'éditeur.

---

## Communications orales

---

**2010 Dorothee Hermann**, Aurore Caruso, Annick Morant-Manceau, Gerard Tremblin, Boris Jacquette, Yohan Graglia, Sophie Hiard et Nathalie Casse.

«Expression d'éléments génétiques transposables chez trois diatomées marines induite par un stress» 29<sup>ième</sup>  
**Congrès International** de l'Association des Diatomistes de Langue Française, Quebec, Canada, 7-10 septembre

**2010** Gérard Tremblin, **Dorothee Hermann**, Iles Makhloufi, Boris Jacquette, Aurore Caruso, Lionel Ulmann, Virginie Mimouni, Annick Manceau

«Mise en évidence et estimation des stress chez des diatomées marines par fluorimétrie modulée» 29<sup>ième</sup>  
**Congrès International** de l'Association des Diatomistes de Langue Française, Quebec, Canada, 7-10 septembre

**2009 Hermann D.**, Caruso A., Bui Q-T, Jacquette B., Morant-Manceau A., Casse N.

“Eléments Transposables de type *mariner* chez les diatomées marines”  
16<sup>ème</sup> **Congrès National** des Eléments Transposables (CNET), Le Mans, France, 1-3 juillet

**2009 Hermann D.**, Casse N., Caruso A., Jacquette B., Bui Q.T., Morant-Manceau A.

“Eléments transposables génétiques de type *mariner* chez les microalgues marines”  
Colloque de la Société Française de Phycologie « ALGIMAC : Les Algues : leur importance actuelle. Biodiversité, environnement, exploitation et santé », Concarneau 9-10 décembre.

Acte publié dans **Cryptogamie Algologie**, vol 30 (novembre 2009) (pp. 311-341)

---

## Communications affichées

---

**2010 Hermann D.**, Caruso A., Morant-Manceau A., Tremblin G., Jacquette B., Graglia Y., Hiard S., Casse N.

“Expression of the *mariner*-like elements in the marine diatom *Phaeodactylum tricornutum* under thermic stress”  
**International Congress** of mobile DNA, Montréal, Canada, 25-29 Avril.

**2009 Hermann D.**, Casse N., Morant-Manceau A., Tremblin G., Jacquette B., Hiard S., Guo N., Graglia Y., Caruso A., “Le stress, une stratégie pour détecter les éléments transposables de type *mariner* chez des microalgues marines” 16<sup>ème</sup> **Congrès National** des Eléments Transposables (CNET), Le Mans, France, 1-3 juillet

**2008 Hermann D.**, Caruso A., Casse N., Hardivillier Y., Pradier E., Jacquette Boris, Sallenave-Namont C., Robiou du Pont T., Chenais B., Morant-Manceau A. “Tc1/*mariner*-like elements in marine diatoms and fungi”

**International Congress** on Transposable Elements, Saint-Malo, France, 20-23 avril

## CARACTERISATION D'ELEMENTS TRANSPOSABLES DE TYPE MARINER CHEZ LES DIATOMEES MARINES (Bacillariophycées)

**Résumé :** Les éléments transposables (ET) sont des séquences d'ADN capables de se déplacer dans les génomes qui les hébergent. Les ET de type *mariner* (*mariner*-like element : *MLE*) ont été caractérisés chez les animaux et chez les plantes à fleurs terrestres mais pas chez les microalgues marines. Dans le présent travail, les *MLE* ont été recherchés chez les diatomées marines (Bacillariophycées) qui sont des microalgues possédant une enveloppe externe siliceuse. Elles ont colonisé tous les milieux aquatiques et constituent une part majeure du phytoplancton; de ce fait, elles jouent un rôle écologique clé dans le milieu marin.

La caractérisation des *MLE* des diatomées a été réalisée au moyen d'approches moléculaire et bio-informatique. La présence de *MLE* dans le génome de 10 espèces de diatomées a été mise en évidence grâce à l'amplification de fragments d'environ 380 pb. Ces fragments correspondent à une partie de la séquence conservée codant l'enzyme responsable de la transposition des *MLE* : la transposase. L'analyse des séquences obtenues, par des méthodes phylogénétiques, ainsi que la classification des *MLE* de diatomées mettent en évidence leur appartenance au groupe des *MLE* végétaux de la superfamille *Tc1-mariner*. Néanmoins, les séquences *MLE* de diatomées divergent considérablement par rapport aux *MLE* des plantes terrestres. Afin de déterminer si les diatomées ont la capacité de produire la transposase, l'expression des *MLE* a été recherchée chez des diatomées soumises à des stress thermiques de courte durée (5 h). Nos travaux montrent que les *MLE* sont exprimés chez les trois espèces testées incluant la diatomée modèle *Phaeodactylum tricornerutum* dont le génome a été séquencé récemment. L'expression des *MLE* des diatomées est variable selon les conditions thermiques et selon les espèces.

L'ensemble de nos résultats suggère que les *MLE* sont ubiquistes dans les génomes de diatomées et qu'ils sont présents de manière ancestrale dans la lignée végétale. Les *MLE* des diatomées forment trois nouvelles sousfamilles de la superfamille *Tc1-mariner*, la plus répandue des superfamilles d'ET. De plus, leur expression suggère qu'il existe des *MLE* capables de se déplacer dans le génome des diatomées. Si la transposition était vérifiée, ils pourraient alors être développés comme outils de mutagenèse.

**Mots-clés :** éléments transposables, *mariner*-like element, diatomées marines, stress, Bacillariophycée.

## CHARACTERIZATION OF MARINER-LIKE ELEMENTS IN MARINE DIATOMS (Bacillariophyceae)

**Abstract:** Transposable elements (TE) are sequences able to move between two loci in the host genomes. Mariner-like element (*MLE*) are well characterized in animal and land plant genomes but not in marine microalgae. In this work, we have looked for *MLE* in marine diatoms (Bacillariophyceae) that are microalgae having a siliceous wall. They have colonized all aquatic environments and are a major component of the phytoplankton, so they play a major role in the ecology of marine environments.

To characterize the diatom *MLE*, molecular and bio-informatics approaches were used. The presence of *MLE* was detected in ten diatom species which exhibited sequence fragments of about 380 pb. These fragments were identified as a section of the conserved sequence which encodes the enzyme responsible for *MLE* transposition, the transposase. Phylogeny and classification analysis of these fragments revealed that diatom *MLE* belong to the *Tc1-mariner* superfamily, and more precisely to the vegetal *MLE* group. Nevertheless, diatom *MLE* sequences diverged from the flowering plant *MLE*. In order to determine if *MLE* transposase is produced in diatoms, *MLE* expression was looked for in diatoms under thermal stresses. Our results showed that diatom *MLE* were expressed in the three species tested, including the model diatom *Phaeodactylum tricornerutum* which was completely sequenced recently. Diatom *MLE* expression was dependent on the thermal conditions and on the species.

Our results suggest that *MLE* are widespread in diatom genomes and that they have an ancestral presence in the green lineage. Diatom *MLE* cluster in three subfamilies in the huge ET *Tc1-mariner* superfamily. Finally, the diatom *MLE* expression detected could reflect the existence of active *MLE* in diatom genomes. If this hypothesis were verified, it could lead to the development of mutagenesis tools.

**Key-words:** transposable element, *mariner*-like element, marine diatoms, stress, Bacillariophyceae.