

UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

ÉTUDE CHEZ LA DIATOMÉE *PHAEODACTYLUM TRICORNUTUM* :
LA STABILITÉ D'UN SYSTÈME D'EXPRESSION EXTRACHROMOSOMIQUE
CONTENANT UNE VOIE DE BIOSYNTHÈSE DE LA VANILLINE

THÈSE PRÉSENTÉE
COMME EXIGENCE PARTIELLE DU
DOCTORAT EN BIOLOGIE CELLULAIRE ET MOLÉCULAIRE

PAR
ANDREW DIAMOND

SEPTEMBRE 2022

Université du Québec à Trois-Rivières

Service de la bibliothèque

Avertissement

L'auteur de ce mémoire, de cette thèse ou de cet essai a autorisé l'Université du Québec à Trois-Rivières à diffuser, à des fins non lucratives, une copie de son mémoire, de sa thèse ou de son essai.

Cette diffusion n'entraîne pas une renonciation de la part de l'auteur à ses droits de propriété intellectuelle, incluant le droit d'auteur, sur ce mémoire, cette thèse ou cet essai. Notamment, la reproduction ou la publication de la totalité ou d'une partie importante de ce mémoire, de cette thèse et de son essai requiert son autorisation.

UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

DOCTORAT EN BIOLOGIE CELLULAIRE ET MOLÉCULAIRE (PH. D.)

Direction de recherche :

Isabel Desgagné-Penix

Directrice de recherche

Simon Barnabé

Codirecteur de recherche

Jury d'évaluation de la thèse :

Isabel Desgagné-Penix

Directrice de recherche

Simon Barnabé

Codirecteur de recherche

Hugo Germain

Président de jury

Tagnon Missihoun

Évaluateur interne

Dominique Michaud

Évaluateur externe

Thèse soutenue le 7 septembre 2022

À Jacob, Benjamin et Flavie Diamond

REMERCIEMENTS

Je souhaite remercier ma directrice de recherche, la professeure Isabel Desgagné-Penix de m'avoir partagé son enthousiasme envers la recherche et d'avoir été si inspirante tant du côté académique que du côté humain. C'est avec fierté que j'ai assisté au développement de votre laboratoire de recherche depuis sa création et le chemin parcouru est impressionnant. Je suis très heureux d'y avoir participé et d'avoir tant appris sous votre supervision. C'est avec une profonde gratitude que je remercie mon codirecteur, le professeur Simon Barnabé de m'avoir accueilli dans son laboratoire lors des mes études au baccalauréat. J'ai eu la chance de découvrir le monde passionnant des microalgues et de vouloir en faire le sujet de mes études graduées. Vous m'aurez démontré l'importance des collaborations et de leurs synergies pour le développement de projets de recherche passionnants.

Un remerciement particulier à Aparna Singh et Tarun Hotchandani pour votre soutien lors du début de nos études de cycles supérieurs. Un immense merci à Eddie Guillermo Sánchez Rueda, Nicolas Dufour, Aracely Maribel Diaz-Garza, Elisa Fantino et Félix Lavoie-Marchand d'avoir grandement participé à la progression des travaux de recherche de cette thèse. Merci aux professionnels de recherche Mélodie B. Plourde, Fatma Meddeb-Mouelhi, Natacha Mérindol et Sarah-Ève Gélinas pour votre soutien et le travail exceptionnel que vous réalisez pour le laboratoire. Je tiens à souligner le plaisir que j'ai eu de collaborer avec mes collègues de l'équipe des diatomées Fatima Awwad, Fadoua Dhaouadi, Aracely Maribel Diaz-Garza, Elisa Fantino, Anis Messaabi, Nicolas Sene et Nikunj Sharma. J'aimerais remercier tous mes autres collègues et anciens collègues, particulièrement Narimene Fradj, Kiran Jagdish Todkar, Shweta Kiran Todkar, Bulbul Ahmed, Saifur Rahman, Hur Madina, Manoj Koirala, Serge Basile Nouemssi, Bharat Bhusan Majhi, Claire Letanneur, Teura Barff, Geneviève Laperrière, Zainab Fakih, Ariane Garand, Daris Pazhukkunnel Simon, Alexandre Brisson, Ingrid Berenice Sanchez Carrillo, Manel Ghribi, Annabelle St-Pierre, Dorian Blondeau et Rémy Beauchemin.

Je tiens à remercier les professeurs du département de chimie, biochimie et physique et du département de biologie médicale pour l'accès leurs équipements. Un remerciement particulier au professeur Hugo Germain pour son soutien, ses discussions enrichissantes et sa collaboration envers l'équipe de laboratoire de ma directrice de recherche. Je remercie le Professeur Bogumil Karas et son équipe pour l'importante collaboration sur le projet de la vanilline. Je souhaite remercier Mme Catarina Leote F. Pio pour son soutien exceptionnel envers les étudiants de cycles supérieurs. Je tiens à exprimer toute ma reconnaissance et ma gratitude envers le conseil de recherches en sciences naturelles et en génie du Canada pour la bourse d'études supérieures du Canada accordée pour mes études doctorales.

Un remerciement spécial à ma famille pour leur encouragement envers mes études. Je remercie Catherine Béland pour son aide à la réalisation des figures lors de mes diverses présentations et publications. Merci à mes enfants Jacob, Benjamin et Flavie pour leur amour.

RÉSUMÉ

La vanilline est la molécule majoritairement responsable de l'arôme et de la saveur des extraits de vanille. Ce composé aromatique est utilisé principalement dans la préparation d'aliments et de parfums. Quoique la vanilline puisse être extraite des gousses de vanille, la chimie de synthèse est la méthode de production principale pour répondre à la forte demande des industries et des consommateurs. En alternative à ces deux types de production, la vanilline peut aussi être extraite de la biomasse lignocellulosique par des procédés d'oxydation de la lignine ou produite par des microorganismes. Ces derniers peuvent être des producteurs de vanilline intéressants, car ils peuvent faire la production de ce composé aromatique à partir d'une grande variété de substrats. Il est aussi possible de modifier génétiquement certains microorganismes pour leur procurer la capacité de produire de la vanilline *de novo* ou d'améliorer leur production initiale.

Les microorganismes photosynthétiques n'avaient pas encore été étudiés pour la production de vanilline. Ceux-ci pourraient avoir l'avantage de produire de la vanilline *de novo* à partir du dioxyde de carbone sans ajout externe de carbone organique. La diatomée *Phaeodactylum tricornutum* est un microorganisme photosynthétique modèle possédant un potentiel pour la production de vanilline. Le séquençage de son génome, l'annotation de celui-ci et le développement récent d'outils moléculaires ont permis de démontrer la capacité que possède ce microorganisme pour la production de molécules d'intérêts. Le récent développement d'un système de transformation par conjugaison bactérienne a permis de faciliter et d'augmenter le rendement de transformation de cette diatomée comparativement aux autres techniques telles que l'électroporation. De plus, la découverte d'un épisode comme premier système d'expression extrachromosomique offre l'avantage de propager une construction génique en éliminant les contraintes et les désavantages liés à l'intégration dans le génome des diatomées. L'intégration est imprévisible ce qui peut affecter négativement l'hôte transformé et cette méthode était la façon principale de propager des constructions géniques dans les diatomées avant la découverte de l'épisode chez *P. tricornutum*.

Pour les travaux de cette thèse, la diatomée *P. tricornutum* a été transformée par conjugaison bactérienne avec des épisodes contenant des gènes en lien avec la biosynthèse de la vanilline.

En premiers lieux, l'épisode transformé contenait les sept gènes de la voie de biosynthèse de la vanilline qui a récemment été élucidée chez *Vanilla planifolia*. La production des protéines hétérologues marquées avec une étiquette poly-histidine n'a pas été possible, et ce, bien que l'expression de certains gènes ait été confirmée. De plus, le séquençage des épisodes extraits de ces clones n'a pas permis de révéler des mutations pouvant expliquer complètement l'absence de production des protéines d'intérêts. Cependant, un des clones propageait un épisode réarrangé dont la séquence modifiée contenait des copies partielles de sa propre séquence.

En deuxième lieu, un second plasmide a été construit afin d'utiliser la protéine fluorescente verte (eGFP) comme protéine rapportrice pour étudier la production de protéines par l'épisome. Dans le but de réduire la quantité de protéines à produire, seule l'étape de conversion de l'acide férulique en vanilline a été étudiée. Cependant, une controverse a été mise de l'avant en 2017 concernant cette enzyme, c'est-à-dire la vanilline synthase (*VpVAN*). Cette enzyme a donc été remplacée par les enzymes feruloyl-CoA synthétase (fcs) et enoyl-CoA hydratase/aldolase (ech) dont les activités enzymatiques catalysent respectivement la conversion de l'acide férulique en feruloyl-CoA puis en vanilline. L'analyse des clones transformés avec cet épisome a démontré que des réarrangements de séquences étaient détectables. Pour les dix clones analysés, la fluorescence de la eGFP était détectable seulement dans les clones où les réarrangements modifiaient les gènes codant pour les protéines d'intérêts. Les résultats de ce chapitre permettent de mettre en lumière les limites de l'épisome comme système d'expression extrachromosomique chez la diatomée *P. tricornutum*.

Enfin, un article de type opinion est présenté dans cette thèse afin de démontrer qu'il n'y a pas seulement l'enzyme *VpVAN* qui doit être remise à l'étude, mais aussi la localisation intracellulaire de la voie de biosynthèse de la vanilline chez *V. planifolia*. Une étude récente avait démontré que le chloroplaste à lui seul permettait d'effectuer la biosynthèse de la vanilline. Cependant, la méthodologie utilisée dans cet article ne permet pas d'exclure l'implication d'enzymes cytosoliques et d'enzymes liés au réticulum endoplasmique afin de convertir la L-phénylalanine en vanilline.

Mots-clés : Vanilline, *Vanilla planifolia*, vanilline synthase, diatomée, *Phaeodactylum tricornutum*, système d'expression extrachromosomique, épisome, protéine fluorescente verte

TABLE DES MATIÈRES

REMERCIEMENTS	iv
RÉSUMÉ.....	vi
LISTE DES TABLEAUX.....	xii
LISTE DES FIGURES	xiv
LISTE DES ABRÉVIATIONS	xvii
CHAPITRE I	
INTRODUCTION.....	1
1.1 Production de vanille naturelle.....	1
1.2 Voie de biosynthèse de la vanilline chez la plante <i>Vanilla planifolia</i>	3
1.3 Intérêt commercial et industriel de la vanilline	8
1.4 Productions alternatives de la vanilline.....	9
1.4.1 Vanilline produite à partir de la lignine.....	10
1.4.2 Production de vanilline par des microorganismes de souche sauvage à partir de substrats autres que la lignine.....	11
1.4.3 Ingénierie métabolique des microorganismes pour la production de vanilline	11
1.5 Stratégies pour limiter la toxicité de la vanilline produite par les microorganismes.....	17
1.6 Les microorganismes photosynthétiques comme potentiels hôtes pour la production de vanilline par ingénierie métabolique	18
1.6.1 Les cyanobactéries.....	19
1.6.2 L'algue unicellulaire verte <i>Chlamydomonas reinhardtii</i>	19
1.6.3 La diatomée <i>Phaeodactylum tricornutum</i>	20
1.7 Objectifs.....	24
CHAPITRE II	
REPRODUCTION DE LA VOIE DE BIOSYNTHÈSE DE LA VANILLINE PROVENANT DE LA PLANTE <i>VANILLA PLANIFOLIA</i> VIA UN SYSTÈME D'EXPRESSION EXTRACHROMOSOMIQUE DANS LA DIATOMÉE <i>PHAEODACTYLUM TRICORNUTUM</i>	26
2.1 Introduction	26
2.2 Matériels et méthodes.....	29
2.2.1 Souches microbiennes et conditions de culture.....	29

2.2.2	Liste des produits chimiques.....	29
2.2.3	Préparation du milieu de culture L1 modifié pour <i>P. tricornutum</i>	30
2.2.4	Origine du plasmide p0521s-V et des clones de <i>P. tricornutum</i>	31
2.2.5	Préparation d'une courbe de calibration de la densité cellulaire de <i>P. tricornutum</i> en fonction de la densité optique.....	33
2.2.6	Prétraitement des culots cellulaires de <i>P. tricornutum</i> pour l'analyse des phénylpropanoïdes par chromatographie en phase liquide à haute performance couplée à un spectromètre de masse en tandem (HPLC-MS/MS).....	34
2.2.7	Extraction sur phase solide des échantillons prétraités de <i>P. tricornutum</i> pour analyse par HPLC-MS/MS	34
2.2.8	Analyse par HPLC-MS/MS des composés phénoliques libres extraits de <i>P. tricornutum</i>	35
2.2.9	Validation de la méthode pour l'analyse des composés phénoliques libres par HPLC-MS/MS	36
2.2.10	Détection des protéines avec une étiquette poly-histidine par immunobuvardage de type western	39
2.2.11	Extraction des ARN totaux et transcription inverse suivie d'une PCR (RT-PCR) en point final	40
2.2.12	Extraction des épisomes à partir des clones recombinants de <i>P. tricornutum</i>	41
2.2.13	Séquençage des épisomes extraits des clones de <i>P. tricornutum</i>	42
2.3	Résultats.....	43
2.3.1	Analyse de la production de métabolites spécialisés dans des clones de <i>P. tricornutum</i>	43
2.3.2	Analyse de la production des protéines d'intérêts par la méthode d'immunobuvardage de type western	58
2.3.3	Analyse qualitative de l'expression des cassettes d'expression par RT-PCR	59
2.3.4	Analyse des séquences des épisomes propagés par les clones V1-3 #2, Vc-1 #1 et Vp-7 #1	61
2.4	Discussion.....	63
2.5	Conclusion.....	69
CHAPITRE III		
INSTABILITY OF EXTRACHROMOSOMAL DNA TRANSFORMED INTO THE DIATOM <i>PHAEODACTYLUM TRICORNUTUM</i> FOR HETEROLOGOUS VANILLIN PRODUCTION		
3.1	Contribution des auteurs	70

3.2	Résumé de l'article (français).....	71
3.3	Article complet (anglais)	72
	Introduction	72
	Results and Discussion	74
	Fluorescence microscopy analysis of transconjugants demonstrated different fluorescent patterns in a single transformation event.....	74
	The sequences of the episomes recovered from <i>P. tricornutum</i> pDMi7 transconjugants showed rearrangements and mutations	77
	Fluorescent and non-fluorescent transconjugants tested positive for gene expression of the ech/fcs fusion protein	81
	Detection of heterologous proteins in <i>P. tricornutum</i> transconjugants depends on the percentage of producing cells in the total population and on the DNA sequence of the expression cassette.....	83
	The sequence of the expression cassette has an impact on the percentage of fluorescent transconjugant obtained by bacterial conjugation.....	91
	Conclusion	93
	Experimental procedures	95
	Microbial Strains and Growth Conditions	95
	<i>P. tricornutum</i> growth medium	95
	Plasmid Construction.....	96
	Transformation of <i>P. tricornutum</i> by bacterial conjugation from <i>E. coli</i> cells	97
	Fluorescence microscopy.....	98
	Episome DNA Isolation from <i>P. tricornutum</i> and episome rescue	99
	Whole Plasmid Sequencing	99
	Promoter region prediction	100
	Protein extraction.....	100
	Western blot.....	100
	Flow cytometry and fluorescence-activated cell sorting (FACS).....	101
	Total RNA extraction and RT-PCR.....	102
	References.....	103
	CHAPITRE IV	
	IS THERE A SPICE MISSING IN THE RECIPE? THE ELUCIDATED	
	INTRA-CELLULAR LOCALIZATION OF VANILLIN BIOSYNTHESIS	
	FROM L-PHENYLALANINE NEEDS MORE INVESTIGATIONS.....	107
4.1	Contribution des auteurs	107

4.2	Résumé de l'article (français).....	107
4.3	Article complet (anglais)	108
	Introduction	108
	Vanillin biosynthesis.....	110
	The chloroplast was reported to be sufficient to convert L-phenylalanine into vanillin.....	112
	Cytosolic and ER-anchored vanillin biosynthetic enzymes may co-purify with isolated chloroplasts.....	113
	Conclusion.....	115
	References.....	116
CHAPITRE V		
DISCUSSION, CONCLUSION ET PERSPECTIVES..... 120		
5.1	Discussion.....	120
5.1.1	Chapitre II – Reproduction de la voie de biosynthèse de la vanilline	120
5.1.2	Chapitre III – Instabilité des épisomes chez <i>P. tricornutum</i>	122
5.1.3	Chapitre IV – Réévaluation nécessaire pour la localisation intracellulaire de la voie de biosynthèse de la vanilline chez <i>V. planifolia</i>	126
5.2	Conclusion.....	127
5.3	Perspectives	128
5.3.1	Extraction des composés phénoliques	128
5.3.2	Évaluation du profil de l'activité des promoteurs.....	129
5.3.3	L'expression inductible des gènes d'intérêts	130
5.3.4	Gènes alternatifs pour la conversion de l'acide férulique en vanilline...	131
5.3.5	Étude sur la possibilité d'un mécanisme de répression des gènes.....	132
RÉFÉRENCES BIBLIOGRAPHIQUES..... 135		
ANNEXE A		
MUTATIONS ET VARIATIONS DANS LES SÉQUENCES DES PLASMIDES EXTRAITS DES CLONES V1-3 #2, VP-7 #1 ET VC-1 #1 PAR RAPPORT À LA SÉQUENCE DE RÉFÉRENCE DU PLASMIDE P0521S-V.. 153		
ANNEXE B		
FIGURES ET TABLEAUX SUPPLÉMENTAIRES DE L'ARTICLE PRÉSENTÉ AU CHAPITRE III..... 159		

LISTE DES TABLEAUX

Tableau		Page
2.1	Origine des gènes clonés dans le plasmide p0521s-V	32
2.2	Paramètres utilisés pour chaque paire d'amorces dans le but d'amplifier l'ADNc des cassettes d'expression.....	41
2.3	Transitions MRM observées chez les composés ciblés et le SI qui ont été sélectionnées pour la quantification par HPLC-MS/MS	46
2.4	Moyennes des taux des récupérations obtenues à la suite de l'extraction en phase solide avec les cartouches Strata-X-A	47
2.5	Moyenne des effets de matrice obtenus à la suite de l'extraction en phase solide avec les cartouches Strata-X-A	47
2.6	Moyenne des limites de détection (LDD) et de quantification (LDQ) pour les composés phénoliques analysés et le standard interne	49
2.7	Moyenne de la précision et de l'exactitude des contrôles de qualité à 0,5 mg/L.....	50
2.8	Concentrations des composés phénoliques ciblés dans les extraits des souches Vp-7 #1, Vc-1 #1, V1-3 #2 et p0521s qui ont été injectés au HPLC-MS/MS	53
3.1	Count of <i>P. tricornutum</i> fluorescent colonies using fluorescence microscopy 14 days after the bacterial conjugation with the episomes recovered from the <i>P. tricornutum</i> cells initially transformed with pDMi7..	93
A.1	Analyse des variations communes dans le séquençage des épisomes extraits des clones de <i>P. tricornutum</i> V1-3 #2, Vc-1 #1 et Vp-7 #1.....	153
A.2	Analyse des variations communes dans le séquençage des épisomes extraits des clones de <i>P. tricornutum</i> V1-3 #2 et Vc-1 #1.....	155
A.3	Analyse des variations dans le séquençage de l'épisome extrait du clone de <i>P. tricornutum</i> V1-3 #2.....	156
A.4	Analyse des variations dans le séquençage de l'épisome extrait du clone de <i>P. tricornutum</i> Vc-1 #1	157
A.5	Analyse des variations dans le séquençage de l'épisome extrait du clone de <i>P. tricornutum</i> Vp-7 #1	158

B.1	Count, and fluorescence detection of <i>P. tricornutum</i> colonies by fluorescence microscopy 14 days after conjugation for pControl	168
B.2	Count, and fluorescence detection of <i>P. tricornutum</i> colonies by fluorescence microscopy 14 days after conjugation for pDMi7 and the empty vector (pPtGE30)	168
B.3	Sequence analysis of the plasmid recovered from <i>P. tricornutum</i> transconjugants	169
B.4	List of oligonucleotides used in this study	175
B.5	List of plasmids used in this study	176
B.6	List of genetic parts used in this study	177

LISTE DES FIGURES

Figure		Page
1.1	Section transversale d'une gousse de vanille âgée de 6 mois.....	3
1.2	Voie de biosynthèse de la vanilline- β -D-glucoside dans les gousses de <i>V. planifolia</i>	5
1.3	Voie de biosynthèse proposée de la vanilline impliquant la molécule <i>p</i> -hydroxybenzaldéhyde	7
1.4	Voie de biosynthèse <i>de novo</i> de la vanilline dans la bactérie <i>E. coli</i>	14
1.5	Voie de biosynthèse de la vanilline utilisant le PET comme substrat de départ.....	16
1.6	Schéma des différentes morphologies cellulaires de <i>P. tricornutum</i>	22
2.1	Schéma représentant l'ajout d'un point d'entrée pour la voie de biosynthèse de la vanilline via l'enzyme TAL	27
2.2	Représentation graphique de l'épisome p0521s-V contenant les séquences codantes des enzymes de la voie de biosynthèse de la vanilline	45
2.3	Courbe de calibration de la vanilline et de ses intermédiaires.....	48
2.4	Courbe de calibration de la densité cellulaire en fonction de la densité optique à 730 nm pour les cultures de <i>P. tricornutum</i>	51
2.5	Courbe de croissance des clones p0521s (vecteur vide), Vp-7-#1, Vc-1 #1 et V1-3 #2 de <i>P. tricornutum</i>	52
2.6	Temps d'acquisition du standard de la vanilline à 1 ppm pour les trois ions fragmentés [M+H] ⁺ qui sont en plus grande abondance relative.....	54
2.7	Analyse de la présence de la vanilline dans la phase mobile.....	55
2.8	Analyse de la présence de la vanilline dans la solution de MeOH utilisée pour conditionner et laver les cartouches d'extraction en phase solide.....	56
2.9	Analyse de la présence de la vanilline dans la solution d'acide formique : MeOH (5:95, MeOH de grade LC-MS) qui est utilisée pour l'élution des composés d'intérêts lors de l'extraction en phase solide.....	57
2.10	Détection de protéines contenant une étiquette poly-histidines par immunobuvardage de type western	59

2.11	Détection de l'expression des cassettes DMi1, DMi2, DMi3 et DMi4 par migration sur gel d'agarose des produits de la RT-PCR provenant des clones p0521s, V1-3 #2, Vc-1 #1 et Vp-7 #1	60
2.12	Migration sur gel d'agarose 1 % de 500 ng d'ARNs totaux qui ont été utilisés en parallèle pour effectuer la RT-PCR	61
3.1	Transformation of <i>P. tricornutum</i> cells with a vanillin biosynthetic pathway	75
3.2	DNA digestion profiles and representation of the rearrangements of the recovered episomes from pDMi7 <i>P. tricornutum</i> transconjugants	78
3.3	Expression of the cassette of interest from pDMi7 <i>P. tricornutum</i> transconjugants	82
3.4	Enrichment of <i>P. tricornutum</i> GFP+ through fluorescence activated cell sorting	84
3.5	Stable and high level of GFP expression in DMi7-21-7 ^{GFP++} across subculturing.....	87
3.6	Comparison of GFP detected by western blot and by flow cytometry	88
4.1	The simplified biosynthetic pathway of vanillin and its potential intra-cellular localization	109
5.1	Production par <i>P. tricornutum</i> de la fluorescence eGFP sous le contrôle des promoteurs fcpA et GLNA	130
B.1	Observation of pPtGE33 colonies, that are producing eGFP and mCherry separated by a T2A self-cleaving peptide, with a fluorescent stereo microscope under (a) bright field, (b) GFP filter, and (c) RFP filter.....	159
B.2	Observation of different pControl colonies (a to e) with a fluorescent stereo microscope under bright field (left panel) and GFP filter (right panel).....	160
B.3	Observation of different pDMi7 colonies (a to f) with a fluorescent stereo microscope under bright field (left panel) and GFP filter (right panel).....	161
B.4	Analysis of putative promoter region of DMi7-21-3 clone product of plasmid rearrangements	162
B.5	Analysis of GFP fluorescence by flow cytometry from <i>P. tricornutum</i> cultures of DMi7-21-14, DMi7-21-15 and DMi7-21-16 that were detected as non-fluorescent (GFP-).....	163
B.6	Antibiotic resistant GFP- cells can be stably isolated from diatoms conjugated with a GFP and antibiotic resistance genes expressing episome following two cycles of fluorescence-activated cell sorting.....	164

B.7	Western blot anti-Myc tag on protein extracts from <i>P. tricornutum</i> transconjugants for the detection of the ech-fcs fusion protein production...	165
B.8	Colonies obtained 14 days after the bacterial conjugation with the episomes recovered from <i>P. tricornutum</i> initially transformed with pDMi7	166
B.9	Codon usage table for <i>P. tricornutum</i> extracted on the High-performance Integrated Virtual Environment-Codon Usage Tables (HIVE-CUTs) platform (Athey et al., 2017) using the refseq database on September 21st in 2018	167
B.10	Calibration curve for the optical density at 730 nm of <i>P. tricornutum</i> liquid culture and the cell count with a Neubauer improved cell counting chamber	167

LISTE DES ABRÉVIATIONS

%CV	Coefficient de variation
%GC	Contenu en nucléotides guanine et cytosine
%RE	Erreur relative
3DSD	3-déshydroshikimate déshydratase
4CL	4-hydroxycinnamoyl-CoA ligase
ADNc	ADN complémentaire
ARNm	ARN messenger
bsr	Blasticidine S désaminase
C3'H	Coumaroyl ester 3'-hydroxylase
C4H	Cinnamate 4-hydroxylase
CAR	Acide carboxylique réductase
CO ₂	Dioxyde de carbone
CoA	Coenzyme A
COMT	Acide caféique/acide 5-hydroxyfêrulique <i>O</i> -méthyltransférase
DCDDH	Acide 1,4-dicarboxylique déshydrogénase
DHBAI	3,4-dihydroxybenzaldehyde
DO _{730nm}	Densité optique à 730 nm
ech	Enoyl-CoA hydratase/aldolase
eGFP	Protéine fluorescente verte optimisée (de l'anglais « enhanced green fluorescent protein »)
eGFP+	qui émet de la fluorescence verte
eGFP-	dont l'émission de la fluorescence verte n'a pas été détectée
ETA	Éléments transposables autonomes

ER	Réticulum endoplasmique
FACS	Tri cellulaire induit par fluorescence
fbr-CM/PDH	Chorismate mutase/préphénate déhydrogénase
fbr-DAHPS	3-désoxy-D-arabinoheptulosonate-7-phosphate synthase
fcs	Trans-feruloyl-CoA synthétase
FMDV2A	Peptide 2A provenant du virus causant le syndrome pieds-mains-bouche
gltA	Gène codant pour la citrate synthase
HBS	p-hydroxybenzaldéhyde synthase
HCT	Hydroxycinnamoyltransférase
HPLC-MS/MS	Chromatographie en phase liquide à haute performance avec spectrométrie de masse en tandem
icdA	Isocitrate déshydrogénase
GUS	β -glucuronidase
LB	Milieu Luria-Bertani
LCC	Cutinase
LDD	Limite de détection
LDQ	Limite de quantification
m/z	Ratio de la masse sur la charge
MeOH	Méthanol
MRM	Multiple-reaction monitoring
NAT	Nourséothricine acétyltransférase
NCBI	National Center for Biotechnology Information
P2A	Peptide 2A teschovirus-1
PAL	Phénylalanine ammonia-lyase
PC	Acide protocatéchique
PCR	Réaction en chaîne de la polymérase

PES	Polyethersulfones
PEPS	Phosphoenolpyruvate synthétase
PET	Polytéréphtalate d'éthylène
PPDiox	Phénylpropanoïde 2,3-dioxygénase
PRA-PH/CH	Phosphoribosyl-ATP pyrophosphohydrolase/phosphoribosyl-AMP cyclohydrolase
RT-PCR	PCR à la suite d'une transcription inverse
S-COMT	Forme soluble de la catéchol- <i>O</i> -méthyltransférerase
SDS-PAGE	Gel de polyacrylamide 10 % en présence de laurylsulfate de sodium
Sh ble	Gène codant pour la résistance à la bléomycine
SI	Standard interne
S/B	Rapport signal sur bruit
TAL	Tyrosine ammonia-lyase
T2A	Peptide 2A du virus <i>Thosea asigna</i>
TKT	Transcétolase
TPADO	Téréphtalate 1,2-dioxygénase
UMPS	Uridine-5-monophosphate synthase
VA	Acide vanillique
vdh	Vanilline déshydrogénase
<i>Vp</i> VAN	Vanilline synthase de la plante <i>Vanilla planifolia</i>

CHAPITRE I

INTRODUCTION

1.1 Production de vanille naturelle

La vanille, sous forme d'extrait naturel ou artificiel, est une épice utilisée entre autres dans la préparation de plusieurs aliments. Le métabolite spécialisé majoritairement responsable de l'arôme et de la saveur de l'extrait de vanille naturelle est la vanilline. Ce composé aromatique fait partie d'un mélange de plus de 200 molécules donnant l'aspect riche et complexe de la vanille naturelle (Sinha, Sharma, et Sharma 2008).

La vanille naturelle est obtenue par la culture des gousses de certaines plantes de la famille des *Orchidaceae* du genre *Vanilla* (*V*). Parmi celles-ci, *V. tahitensis*, *V. pompona* et *V. planifolia* sont les seules espèces utilisées pour la culture commerciale (Kundu 2017). Pour sa qualité aromatique, 95 % de la production mondiale de vanille naturelle provient principalement de l'espèce *V. planifolia* (Gallage et Møller 2018). La qualité de la vanille produite varie selon plusieurs facteurs tels que l'espèce cultivée, la méthode de culture, le climat et la méthode de maturation des gousses après la récolte (Anuradha, Shyamala, et Naidu 2013; Brunschwig et al. 2016; Zhang et Mueller 2012).

La floraison des plants de vanilles se déroule sur une période de six à huit semaines où les fleurs vont par la suite s'ouvrir qu'une seule journée afin de permettre la pollinisation (Anuradha, Shyamala, et Naidu 2013). Le transfert du pollen par des pollinisateurs naturels est un phénomène rare qui se produit seulement dans quelques pays comme le Mexique, et ce pour seulement environ 1 % des fleurs (Hernández et Lubinsky 2010). La culture de la vanille nécessite donc une pollinisation manuelle à l'aide d'un bâton ayant le diamètre d'un cure-dent. Cette étape cruciale doit se faire entre l'ouverture de la fleur en matinée et sa fermeture l'après-midi (Anuradha, Shyamala, et Naidu 2013). Par la suite, entre six et huit semaines seront nécessaires pour former les gousses qui seront

récoltées au bout des sept à huit mois suivants lorsque l'extrémité de la gousse passera du vert au jaune.

Par la suite, l'étape de maturation des gousses récoltées permettant d'obtenir l'arôme unique de la vanille pourra prendre jusqu'à six mois supplémentaires selon la méthode utilisée (Anuradha, Shyamala, et Naidu 2013). Toutes les méthodes sont généralement composées des quatre mêmes étapes. Premièrement, les gousses sont exposées à la chaleur, par exemple en plongeant les gousses de deux à cinq minutes dans un bain d'eau à 65 °C. Cette étape a pour but d'arrêter le développement de la gousse afin d'empêcher la déhiscence, et ce, en induisant par la chaleur la sénescence du fruit (Odoux 2010). Deuxièmement, les gousses alternent durant 7 à 10 jours entre des périodes de séchage au soleil de quelques heures et des périodes d'étuvages dans des endroits clos. Cette deuxième étape a pour but de réduire le taux d'humidité des gousses à 60-70 % et ainsi de limiter la croissance de certains microorganismes néfastes à la qualité de la vanilline tels que les moisissures (Anuradha, Shyamala, et Naidu 2013). Troisièmement, les gousses sont séchées au soleil quelques heures et entreposées à l'ombre par la suite, et ce, à tous les jours pour une durée de deux à trois mois (Odoux 2010). Le taux d'humidité des gousses varie entre 25 et 38 % à la fin de cette troisième étape. Finalement, les gousses sont entreposées dans des boîtes en bois ou en métal pour une durée variant d'un mois à quelques mois (Odoux 2010). Le développement de l'arôme atteint son maximum durant cette étape finale.

Sans ce laborieux processus, la chaire des gousses ne contiendrait pas les composés aromatiques recherchés par les amateurs de vanille naturelle. La période de maturation permet ainsi de libérer les composés qui ont été accumulés sous leur forme glycosylée dans les gousses, telle que la vanilline- β -D-glucoside qui est alors converti en vanilline sous l'action de l'enzyme β -D-glucosidase (Odoux, Chauwin, et Brillouet 2003). Certains travaux ont montré que les microorganismes colonisant les gousses durant le processus de maturation sont aussi impliqués dans la production des composés aromatiques, par exemple via leur production de l'enzyme β -D-glucosidase (Röling et al. 2001; Chen et al. 2015; Escobar-Muciño et al. 2020; Xu et al. 2020).

Les plantes produisant de grandes quantités de vanilline doivent posséder la machinerie enzymatique pour glycosyler et ainsi accumuler cette molécule qui s'avère être toxique en forte concentration pour les êtres vivants (Fitzgerald, Stratford, et Narbad 2003; Boonchird et Flegel 1982). La vanilline- β -D-glucoside commence à s'accumuler dans la partie interne de la gousse à partir de trois mois après la pollinisation. Cette section de la gousse est décrite comme l'ensemble du mésocarpe interne, du « placental laminae », de l'endocarpe et des graines (Figure 1.1). Au moment de la récolte des gousses, la concentration de vanilline- β -D-glucoside est à son plus haut taux (Anuradha, Shyamala, et Naidu 2013).

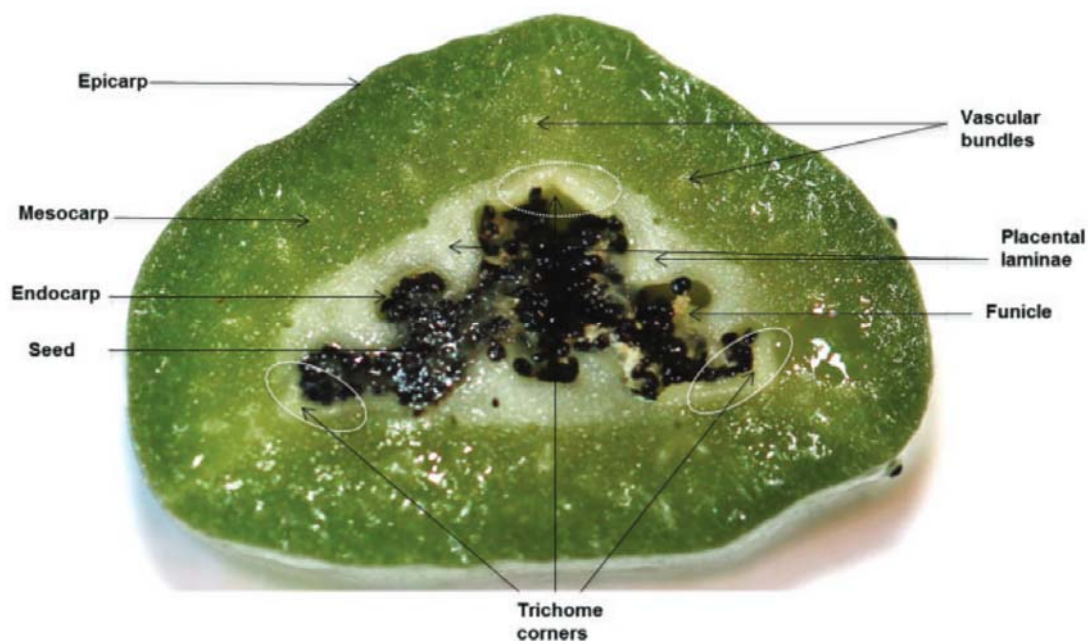


Figure 1.1 Section transversale d'une gousse de vanille âgée de 6 mois.
Source de l'image : Gallage et al. (2018).

1.2 Voie de biosynthèse de la vanilline chez la plante *Vanilla planifolia*

Au niveau moléculaire, la voie élucidée de la biosynthèse de la vanilline- β -D-glucoside (Figure 1.2) a été publiée en 2014 par Gallage et al. (2014). Au début de ce processus, l'enzyme phénylalanine ammonia-lyase (PAL) catalyse la désamination de l'acide aminé L-phénylalanine pour former l'acide trans-cinnamique. Par la suite, une réaction d'hydroxylation par la cinnamate 4-hydroxylase (C4H) convertie

cette molécule en acide *p*-coumarique. La formation du *p*-coumaroyl-CoA, par l'enzyme 4-hydroxycinnamoyl-CoA ligase (4CL), est obtenue par la réaction du groupe sulfhydryle du coenzyme A (CoA) avec le groupe carboxyle de l'acide *p*-coumarique. Ensuite, le groupe CoA est remplacé par une molécule shikimate ou quinate sous l'action de l'enzyme hydroxycinnamoyltransférase (HCT) pour produire la 4-coumaroyl shikimate/quinate. Une réaction d'hydroxylation est par la suite catalysée par la coumaroyl ester 3'-hydroxylase (C3'H) pour convertir la 4-coumaroyl shikimate/quinate en caffeoyl shikimate/quinate. À partir de cette molécule, l'enzyme HCT catalyse le remplacement du shikimate/quinate par le CoA pour former le métabolite caffeoyl-CoA. Le CoA est par la suite remplacé par un groupe hydroxyle via l'activité de l'enzyme 4CL. L'acide caféique ainsi formé est converti en acide férulique à la suite d'une méthylation catalysée par l'acide caféique/acide 5-hydroxyférulique *O*-méthyltransférase (COMT). La découverte de l'enzyme catalysant la conversion de l'acide férulique en vanilline permis à Gallage et al. (2014) d'élucider la voie de biosynthèse de la vanilline. Cette enzyme nommée vanilline synthase effectuerait l'étape de raccourcir la chaîne principale de l'acide férulique et celle de l'acide férulique glucoside de deux carbones. À la suite de la glycosylation de son groupement hydroxyle, la vanilline est ainsi accumulée sous la forme de vanilline- β -D-glucoside.

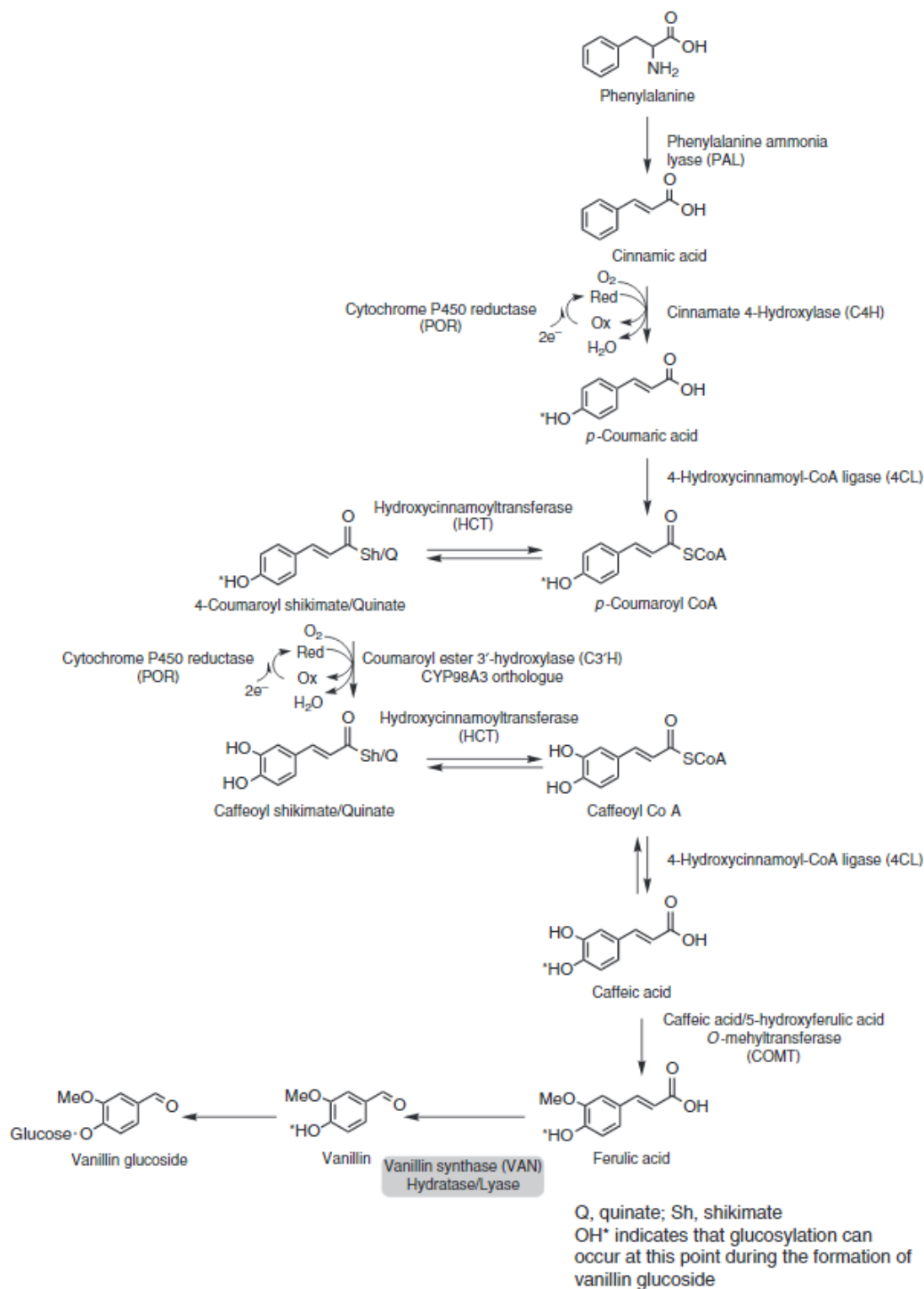


Figure 1.2 Voie de biosynthèse de la vanilline- β -D-glucoside dans les gousses de *V. planifolia*.

Source de l'image : Gallage et al. (2014).

La vanilline synthase de *Vanilla planifolia* (*VpVan*) est une enzyme de type hydratase/lyase (EC 4.1.2.41) qui partage des similarités avec la famille des protéinases à cystéine telles que la présence d'un peptide signal en extrémité N-terminal et un domaine propeptide. Ce domaine a généralement comme fonction d'inhiber l'activité de la protéinase, de réguler la structure de l'enzyme mature et de moduler la localisation de l'enzyme (Demidyuk et al. 2010). Gallage et al. (2014) ont démontré que la *VpVAN* est localisée dans les cellules de la partie interne des gousses de *V. planifolia*, c'est-à-dire au même site d'accumulation que la vanilline- β -D-glucoside.

Les travaux publiés par Gallage et al. (2014) ont par le fait même remis en question l'activité d'une enzyme qui était considérée auparavant comme candidate pour la voie de biosynthèse de la vanilline. Cette enzyme nommée *p*-hydroxybenzaldéhyde synthase (HBS) avait été partiellement purifiée par Podstolski et al. (2002) et elle avait démontré une activité de conversion de l'acide *p*-coumarique en *p*-hydroxybenzaldéhyde. Le mécanisme enzymatique de la HBS consistait à catalyser le raccourcissement de deux carbones au niveau de la chaîne principale de l'acide *p*-coumarique. Les résultats de cet article avaient ensuite été combinés à d'autres articles pour proposer une voie de biosynthèse potentielle de la vanilline (Figure 1.3) (Pak et al. 2004). Selon ce modèle, l'acide *p*-coumarique serait converti en *p*-hydroxybenzaldéhyde par l'enzyme HBS. Cette molécule serait ensuite hydroxylée en C3 pour former le 3,4-dihydroxybenzaldéhyde où par la suite le groupement 3-hydroxyle serait méthylé pour former la vanilline. La séquence de l'ADN codant pour la protéine HBS a par la suite été brevetée par Havkin-Frenkel, Podstolski, et Dixon (2003).

C'est en étudiant les enzymes proposées dans la littérature pour la biosynthèse de la vanilline- β -D-glucoside que Gallage et al. (2014) ont constaté que l'enzyme HBS ne catalysait pas la conversion de l'acide *p*-coumarique en *p*-hydroxybenzaldéhyde. Ils ont cependant détecté *in vitro*, avec cette enzyme, la conversion de l'acide férulique en vanilline et celle de l'acide férulique glucoside en vanilline- β -D-glucoside. L'enzyme nommée HBS fut alors renommée *VpVAN* (accession no. KP278240.1) par Gallage et al.

(2014). Ils ont aussi démontré que la molécule *p*-hydroxybenzaldéhyde n'était pas un précurseur potentiel de la vanilline- β -D-glucoside.

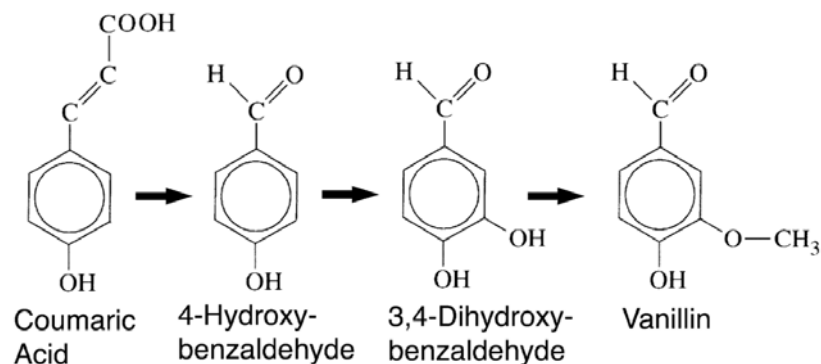


Figure 1.3 Voie de biosynthèse proposée de la vanilline impliquant la molécule *p*-hydroxybenzaldéhyde.
Source de l'image : Pak et al. (2004).

En ce qui concerne le site production intracellulaire de la vanilline chez *V. planifolia*, il aurait été détecté dans les chloroplastes par Gallage et al. (2018). Ils ont d'abord démontré par immunohistochimie que l'enzyme *VpVAN* était accumulée dans les chloroplastes des gousses de vanille. Ces organites ont par la suite été isolés à partir des gousses pour effectuer des essais de biosynthèse de la vanilline. En incubant les chloroplastes isolés avec le précurseur [¹⁴C]-phénylalanine, ils ont détecté une production de l'isotope [¹⁴C]-vanilline- β -D-glucoside. Gallage et al. (2018) ont donc conclu que les chloroplastes contenaient la machinerie enzymatique nécessaire pour convertir la L-phénylalanine en vanilline- β -D-glucoside.

Un groupe de chercheurs a tenté de reproduire les expériences de Gallage et al. (2014), mais ils n'ont pas réussi à confirmer l'activité enzymatique de *VpVAN* (Yang et al. 2017). Ces chercheurs ont mis de l'avant l'hypothèse que *VpVAN* pourrait faire partie d'un complexe protéique nécessaire pour la conversion de l'acide férulique en vanilline. D'un autre point de vue, deux publications indépendantes ont permis de démontrer qu'il est possible de produire de la vanilline en intégrant le gène *VpVAN* dans des cellules végétales. D'un côté, Chee et al. (2017) ont transformé ce gène dans des

cellules en culture de *Capsicum frutescens* et ils ont ainsi détecté une production de vanilline 190 fois plus élevée que dans les cellules de type sauvage. De l'autre côté, Arya et al. (2022) ont produit 26 fois plus de vanilline en intégrant le gène *VpVAN* dans le génome de cellules de riz. Ces deux publications représentent la preuve de concept pour produire de la vanilline par des cellules végétales génétiquement modifiées avec le gène codant pour l'enzyme *VpVAN*. Il n'y a cependant pas d'autres publications concernant l'utilisation de *VpVAN* dans d'autres types d'organismes.

1.3 Intérêt commercial et industriel de la vanilline

En plus de ses utilisations comme saveur, arôme et parfum, la vanilline a été étudiée entre autres pour ses propriétés antibactériennes et antifongiques, et ce, plus particulièrement dans le domaine alimentaire (Fitzgerald et al. 2004, 2005; Bai et al. 2021; Yang et al. 2021). Par exemple, ce composé phénolique posséderait un potentiel comme agent de conservation dans les préparations en poudre pour les nourrissons (Yemiş et al. 2011, 2012). Ces préparations commerciales peuvent être la source de contamination parfois létale chez les nouveau-nés par des bactéries du genre *Cronobacter*. La vanilline possède un effet bactéricide sur ces bactéries.

Quant à la production alimentaire, la vanilline a aussi été analysée pour sa capacité antibactérienne contre deux espèces du genre *Vibrio* qui sont problématiques dans les fermes aquacoles de poissons et de crevettes (Rossi et al. 2021). De plus, cette revue de littérature met de l'avant la problématique des souches microbiennes qui ont développé une résistance aux antibiotiques utilisés en aquaculture. Les métabolites spécialisés extraits de plantes, tels que la vanilline, ont le potentiel d'être une alternative aux antibiotiques. Il a aussi été démontré que le mélange de vanilline, de thymol et de certains acides organiques possède des propriétés antimicrobiennes. Cette préparation pourrait donc être un potentiel additif alimentaire permettant de réduire l'utilisation d'antibiotiques dans les fermes d'élevage de poulets ou de porcs (Rossi et al. 2020).

De plus, en inhibant la détection du quorum, la vanilline permettrait de réduire la formation de biofilms bactériens problématiques sur les membranes utilisées pour la production d'eau potable (Ponnusamy, Paul, et Kweon 2009; Kappachery et al. 2010). L'incorporation de vanilline dans les membranes de polyethersulfones (PES) permettrait de diminuer le développement des bactéries à leur surface, comparativement à une membrane composée seulement de PES (Sathish Kumar et al. 2017), ce qui pourrait augmenter leur durée de vie.

1.4 Productions alternatives de la vanilline

Pour produire 1 kg de vanilline dite naturelle, il faut extraire ce métabolite spécialisé d'environ 500 kg de gousses de vanille qui elles sont obtenues par la pollinisation d'approximativement 40 000 fleurs (Ramachandra Rao et Ravishankar 2000). La demande pour la vanilline en 2016 a été évaluée à 18 600 000 kg et une estimation calculée en 2017 prévoyait une hausse de 6,2 % de 2017 à 2025 (Martău, Călinoiu et Vodnar 2021). En 2018, la production annuelle de gousse de vanille était estimée à 7 500 000 kg (Martău, Călinoiu, et Vodnar 2021). Pour répondre à la demande mondiale, la vanilline doit être produite à partir d'autres sources que les plants de vanille. Ainsi, la vanilline peut aussi être produite par chimie de synthèse, par extraction de la lignine et par la culture de microorganismes de souches sauvages ou génétiquement modifiées.

La vanilline est produite par chimie de synthèse en utilisant principalement le gaïacol comme réactif de départ (Sillanpää et Ncibi 2017). L'avantage majeur de cette synthèse est sa faible production de coproduit permettant de limiter les coûts liés à la purification de la vanilline produite (Sillanpää et Ncibi 2017). Cependant, le gaïacol utilisé est principalement dérivé du pétrole et la vanilline ainsi produite n'est pas issue d'une méthode de production durable (Banerjee et Chattopadhyay 2018).

L'extraction de la vanilline à partir de la lignine permet une production en utilisant une biomasse renouvelable. Comme autre alternative à la synthèse chimique, la vanilline peut aussi être produite en utilisant les diverses techniques de biotechnologie dans un

objectif de production durable. L'utilisation des microorganismes pour la production de vanilline peut ainsi se faire en utilisant des substrats qui ne sont pas issus de l'industrie pétrochimique et même à partir de rejet à valoriser.

1.4.1 Vanilline produite à partir de la lignine

Les tissus végétaux sont composés principalement de trois polymères organiques, c'est-à-dire la cellulose, l'hémicellulose et la lignine. La lignine est un résidu obtenu à la suite de divers bioprocédés tels que la production des pâtes et papiers ainsi que de la production d'éthanol par fermentation des glucides provenant de la cellulose et de l'hémicellulose (Ragauskas et al. 2014). La vanilline peut être extraite de cette lignine par des procédés d'oxydation et d'électrolyse tels que présentés dans une revue de littérature récente (Costa, Vega-Aguilar, et Rodrigues 2021). Par exemple, Zirbes et al. (2020) ont développé un procédé d'électrolyse de la lignine Kraft à haute température, c'est-à-dire à 160 °C.

Les microorganismes ont l'avantage de produire de la vanilline à une température largement inférieure en utilisant entre autres comme matière première de la biomasse lignocellulosique renouvelable et des sous-produits de l'industrie agroalimentaire possédant un potentiel de revalorisation. Afin d'extraire de la biomasse lignocellulosique les molécules nécessaires à la biosynthèse de la vanilline, les microorganismes doivent être en mesure de dépolymériser la lignine. En développant des consortiums de microorganismes, les différentes étapes de ce processus peuvent être ainsi séparées entre les membres du consortium ou réalisées de façon synergique. Par exemple, des chercheurs ont développé un consortium capable de produire de la vanilline à partir de copeaux de bambou (Harshvardhan et al. 2017). De leur côté, le groupe de recherche de Sharma et al., a utilisé la combinaison d'un microorganisme et d'une enzyme pour produire de la vanilline en fermentant du son de blé (Sharma et al. 2021). Ils ont ainsi incubé 10 % p/v de son de blé avec une culture de la bactérie *Streptomyces* sp. ssr-198 et de l'enzyme feruloyl estérase produite par la bactérie *Enterococcus lactis* SR. Martău, Călinoiu et

Vodnar (2021) ont publié une revue de littérature recensant d'autres exemples de ce type de production de la vanilline.

1.4.2 Production de vanilline par des microorganismes de souche sauvage à partir de substrats autres que la lignine

Plusieurs microorganismes ont été étudiés pour leur capacité à produire de la vanilline en bioréacteur à partir de substrat spécifique. Par exemple, plusieurs bactéries et champignons peuvent métaboliser l'acide férulique, l'eugénol ou l'isoeugénol en vanilline (Banerjee et Chattopadhyay 2018). Par exemple, Paz et al. (2018) ont incubé la souche de *Bacillus aryabhatai* BA03 dans un milieu de culture contenant de l'acide férulique.

D'autres travaux ont été effectués avec des microorganismes photosynthétiques puisque ces derniers ont l'avantage de pouvoir croître sans apport externe de carbone organique. De plus, la capacité de ces organismes pour la bioséquestration du carbone permettrait de valoriser les rejets industriels de dioxyde de carbone (CO₂) (Seyed Hosseini, Shang, et Scott 2018). Par exemple, l'équipe de recherche de Onofrejevová et al. (2010) ont extrait de la vanilline à partir d'une souche sauvage de l'algue verte unicellulaire *Spongiocloris spongiosa* cultivée avec un mélange d'air contenant 2 % de CO₂.

1.4.3 Ingénierie métabolique des microorganismes pour la production de vanilline

Les techniques de biologie moléculaire permettent de modifier génétiquement une grande variété de microorganismes dans le but de leur faire produire de la vanilline, et ce, à partir de différents substrats. Une revue récente réalisée par Ma et al. (2022) recense divers cas où ces techniques ont été utilisées pour la production de composé aromatique.

1.4.3.1 Production de vanilline par la bactérie Escherichia coli

La bactérie *Escherichia coli* est un des microorganismes les plus étudiés pour le développement de souche dans un but de production de molécules d'intérêts (Pontrelli

et al. 2018). Cette bactérie est grandement utilisée pour les projets d'ingénierie métabolique, puisqu'elle peut, entre autres, être facilement modifiée génétiquement avec une vaste gamme d'outils moléculaires, elle possède une croissance rapide et son métabolisme a été grandement étudié (Pontrelli et al. 2018).

Cette bactérie n'a pas la capacité de produire de la vanilline, mais elle peut l'acquérir en étant transformée entre autres avec le gène codant pour l'isoeugenol monooxygénase (Yamada et al. 2008) ou une combinaison des gènes codants pour la trans-feruloyl-CoA synthétase (*fcs*) et l'enoyl-CoA hydratase/aldolase (*ech*) (Yoon et al. 2005a; Yoon et al. 2005b; Yoon et al. 2007; Barghini et al. 2007; Lee et al. 2009; Ni et al. 2015; Chakraborty, Gupta, et Kaur 2016). Notamment, Yamada et al. (2008) ont cloné dans *E. coli* le gène de l'isoeugenol monooxygénase provenant de la bactérie *Pseudomonas putida* IE27. Ils ont ainsi pu comparer la conversion de l'isoeugénole en vanilline par le microorganisme de souche sauvage *P. putida* IE27 et la souche d'*E. coli* génétiquement modifiée. Pour cette dernière, 81 % des molécules d'isoeugénol ont été converties en molécules de vanilline après 6 heures de culture en utilisant l'équivalent de 130 mg de biomasse sèche (Yamada et al. 2008). Alors que pour *P. putida* IE27, le taux de conversion a été de 71 % en 24 h de culture à partir de l'équivalent d'une biomasse sèche de 320 mg. Selon les conditions de culture utilisées, Yamada et al. (2008) ont démontré que la bactérie *E. coli* génétiquement modifiée pourrait être un meilleur système de production de vanilline comparativement au microorganisme de souche sauvage *P. putida* IE27 d'où provenait le gène d'intérêt.

Pour la production *de novo* de vanilline par la bactérie *E. coli*, Ni et al. (2015) ont partiellement reproduit la voie de biosynthèse de la vanilline de la plante *V. planifolia* (Figures 1.2 et 1.4). Comme substrat de départ à leur voie de biosynthèse artificielle, ils ont sélectionné la L-tyrosine au lieu de la L-phénylalanine. Dans cette voie, l'enzyme tyrosine ammonia-lyase (TAL) catalyse la conversion de la L-tyrosine en acide *p*-coumarique qui est ensuite converti en acide férulique par les enzymes C3H et COMT. Chez *V. planifolia*, l'enzyme *VpVAN* catalyse la conversion de l'acide férulique en vanilline. Cependant, Ni et al. (2015) n'ont pas réussi à produire cette enzyme dans

E. coli. En alternative, ils ont donc sélectionné les gènes *fcs* et *ech* provenant de la bactérie *Streptomyces* sp. V-1 et qui avaient été précédemment caractérisés par Yang et al. (2013). Pour obtenir une production *de novo* de la vanilline, Ni et al. (2015) ont dirigé le flux de carbone provenant du métabolisme primaire des glucides et des composés organiques simples, comme le glycérol et le xylose, vers la vanilline via la L-tyrosine (Figure 1.4). Pour ce faire, ils ont utilisé une souche d'*E. coli* développée par Nakagawa et al. (2011) qui a été génétiquement modifiée pour surproduire la L-tyrosine. Tout d'abord, l'inactivation du gène *tyrR* empêche la production d'un répresseur causant la rétro-inhibition de la biosynthèse des acides aminés aromatiques (Pittard, Camakaris et Yang 2005). De plus, la souche de Nakagawa et al. (2011) propageait un plasmide contenant les gènes codant pour la phosphoenolpyruvate synthétase (PEPS), la transcétolase (TKT) et des versions mutées des enzymes de la 3-désoxy-D-arabinoheptulosonate-7-phosphate synthase (fbr-DAHPS) et de la chorismate mutase/préphénate déhydrogénase (fbr-CM/PDH). La surproduction des enzymes PEPS et TKT permet d'augmenter le flux de carbone provenant de la glycolyse et de la voie des pentoses phosphates via le phosphoenolpyruvate et l'érythrose-4-phosphate respectivement. Quant aux enzymes fbr-DAHPS et fbr-CM/PDH, elles ont subi des mutations dans le but de les rendre résistantes à la rétro-inhibition (fbr) (Lütke-Eversloh et Stephanopoulos 2007).

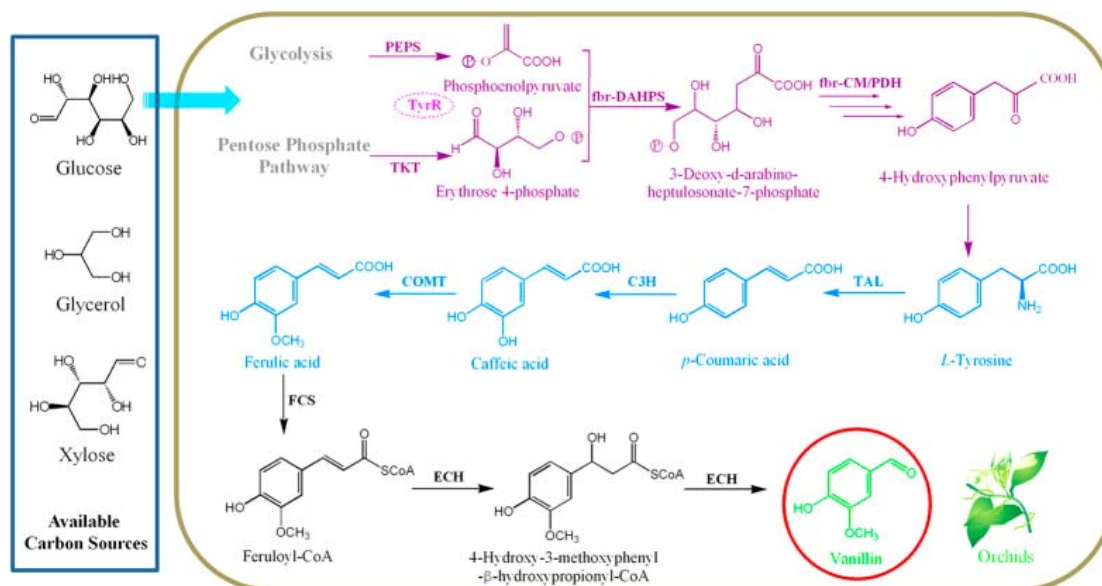


Figure 1.4 Voie de biosynthèse *de novo* de la vanilline dans la bactérie *E. coli*.

La portion en mauve de la voie représente le gène *TyrR* inactivé et les quatre enzymes permettant à la souche d'*E. coli* de surproduire l'acide aminé L-tyrosine. La partie en bleu représente la portion de la biosynthèse de l'acide férulique inspirée de *V. planifolia* par les enzymes *TAL*, *C3H* et *COMT*. Les réactions en noir illustrent la conversion de l'acide férulique en vanilline par les enzymes *FCS* et *ECH*. *TyrR*, répresseur de la biosynthèse des acides aminés aromatiques; *PEPS*, phosphoenolpyruvate synthétase; *TKT*, transcétolase; *fbr-DAHPS*, 3-désoxy-D-arabinoheptulosonate-7-phosphate synthase résistante à la rétro-inhibition; *fbr-CM/PDH*, chorismate mutase/préphénate déhydrogénase résistante à la rétro-inhibition; *TAL*, tyrosine ammonia-lyase; *C3H*, 4-coumarate 3-hydroxylase; *COMT*, Acide caféique/acide 5-hydroxyférulique *O*-méthyltransférase; *FCS*, trans-feruloyl-CoA synthétase; *ECH*, enoyl-CoA hydratase/aldolase.

Source de l'image : Ni et al. (2015).

D'un autre point de vue, il est aussi possible d'augmenter le rendement de production de la vanilline en améliorant cette fois-ci la production de coenzymes nécessaires à la voie de biosynthèse. À titre d'exemple, la conversion de l'acide férulique en vanilline par les enzymes *fcs* et *ech* nécessite une molécule de CoA qui sera par la suite convertie en une molécule d'acétyl-CoA. Si par la suite la CoA n'est pas régénérée efficacement à partir de l'acétyl-CoA, il en résultera une limitation dans le rendement de production de la vanilline. Pour limiter cet effet, l'ingénierie du cycle de Krebs chez *E. coli* permet de concentrer le flux métabolique vers la régénération de la CoA (Lee et al. 2009). Ils ont premièrement surexprimé le gène de ce cycle codant pour l'enzyme citrate synthase (*gltA*).

Par conséquent, la concentration de vanilline produite après 48h d'incubation en présence de 3,0 g/L d'acide férulique a été de 1,98 g/L comparativement à 0,91 g/L pour la souche sans cette modification. Deuxièmement, Lee et al. (2009) ont utilisé une souche d'*E. coli* dans laquelle le cycle de Krebs a été raccourci en supprimant le gène de l'isocitrate déshydrogénase (*icdA*). Cette mutation a eu comme effet d'accélérer la production de vanilline par *E. coli*. En combinant la surexpression du gène *gltA* avec cette souche mutante, ils ont détecté après 24h de culture une production de 1,87 g/L de vanilline en présence de 3,0 g/L d'acide férulique. En comparaison, en cultivant la souche d'*E. coli* surexprimant le gène *gltA* sans la seconde modification au cycle de Krebs, Lee et al. (2009) n'ont produit que 1,16 g/L de vanilline en 24 h, et ce, avec la même concentration d'acide férulique.

La modification génétique des microorganismes permet aussi de produire de la vanilline en utilisant comme substrat un rejet à valoriser. C'est le cas des travaux de Sadler et Wallace (2021) où un rejet de post-consommation à base de polytéréphtalate d'éthylène (PET) a été dégradé pour former de la vanilline. La voie de biosynthèse développée pour ce projet (Figure 1.5) impliquait l'enzymes cutinase (LCC) dont le gène a été isolé du métagénome d'un composte de branches et de feuilles (Sulaiman et al. 2012), la téréphtalate 1,2-dioxygénase (TPADO) de *Comamonas* sp., la CAR originaire de la bactérie *Nocardia iowensis* et la forme soluble de la catéchol-O-méthyltransférase (S-COMT) provenant de *Rattus norvegicus*. La TPADO est un hétéromère composé des sous-unités TphA1, TphA2, TphB2 et de l'acide 1,4-dicarboxylique déshydrogénase (DCDDH) (Sadler et Wallace 2021). La production de vanilline à partir du PET a été réalisée en deux étapes à l'aide de deux clones d'*E. coli*. Premièrement, l'enzyme LCC a été extraite du premier clone d'*E. coli*. Cette enzyme a été utilisée pour la dépolymérisation du PET *in vitro* à 72 °C dans le but de former des molécules d'acide téréphtalique. Deuxièmement, les acides téréphtaliques produits ont été incubés avec le second clone d'*E. coli* génétiquement modifié pour produire les enzymes TPADO, CAR et S-COMT. Il a été nécessaire de procéder en deux étapes, puisque l'activité optimale de la LCC est à 72 °C et la conversion de l'acide téréphtalique en vanilline par la souche d'*E. coli* est optimale à 22 °C.

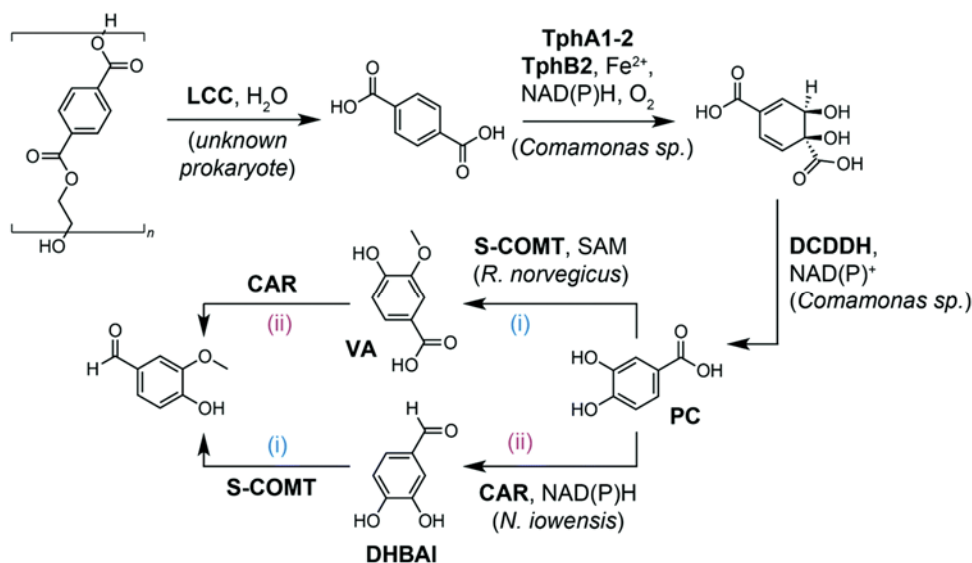


Figure 1.5 Voie de biosynthèse de la vanilline utilisant le PET comme substrat de départ.

L'enzyme LCC catalyse la dépolymérisation du PET en molécules d'acide téréphtalique. Par la suite, l'hétéromère TPADO via ses sous-unités TphA1, TphA2, TphB2 et DCDDH convertit l'acide téréphtalique en acide protocatéchique (PC). La conversion du PC en vanilline est ensuite possible de deux façons par les enzymes S-COMT et CAR. D'un côté, l'enzyme S-COMT convertit le PC en acide vanillique (VA) qui est par la suite transformé en vanilline par l'enzyme CAR. De l'autre côté, l'enzyme CAR catalyse la réduction du PC en 3,4-dihydroxybenzaldehyde (DHBAI) et cette molécule est convertie en vanilline à la suite de la méthylation catalysée par la S-COMT. PET, polytéréphtalate d'éthylène; LCC, cutinase; TPADO, téréphtalate 1,2-dioxygénase; PC, acide protocatéchique; S-COMT, forme soluble de la catéchol-O-méthyltransférase; CAR, acide carboxylique réductase.

Source de l'image : Sadler et Wallace (2021).

1.4.3.2 Production de vanilline par la levure *S. cerevisiae*

Du côté de la production *de novo* de la vanilline par la levure *S. cerevisiae*, la voie de biosynthèse conçue pour ce microorganisme a été reliée au métabolisme primaire de la voie du shikimate (Hansen et al. 2009; Strucko, Magdenoska, et Mortensen 2015; Strucko et al. 2017). La première réaction consiste à convertir l'acide 3-déshydroshikimique en acide protocatéchique (PC) via l'activité enzymatique de la 3-déshydroshikimate déshydratase (3DSD). Le gène codant pour cette enzyme a été amplifié à partir de l'ADN génomique de l'ascomycète *Podospora anserina*. La vanilline est par la suite formée par

la conversion du PC par la CAR et la O-méthyltransférase. Afin de limiter la toxicité de la vanilline, la molécule est accumulée par la levure sous la forme de la vanilline- β -D-glucoside. Cette dernière réaction est catalysée par la UDP-glycosyltransférase originaire de la plante *Arabidopsis thaliana*.

1.5 Stratégies pour limiter la toxicité de la vanilline produite par les microorganismes

Certains facteurs peuvent limiter la production de vanilline par des microorganismes, tels que la rétro-inhibition ou la toxicité. Une stratégie utilisée pour contrer ces situations est l'ajout de résine dans les cultures pour adsorber la vanilline et ainsi limiter sa concentration libre dans le milieu (Zhao et al. 2006; Hua et al. 2007). De cette façon, Zhao et al. (2006) ont utilisé la résine HD-8 pour produire de la vanilline par la bactérie *Bacillus fusiformis* CGMCC1347 en utilisant l'isoeugénol comme substrat. Quant à Hua et al. (2007), ils ont effectué leur production de vanilline en cultivant le microorganisme *Streptomyces* sp. V-1 dans un milieu contenant de l'acide férulique et la résine DM11. Parmi toutes les résines évaluées par Hua et al. (2007), la résine DM11 était celle qui adsorbait le plus de vanilline et à la fois le moins d'acide férulique. Cette caractéristique permettait de limiter la toxicité provenant de la vanilline tout en maximisant l'accès à l'acide férulique aux cellules de *Streptomyces* sp. V-1. L'utilisation des résines est donc un exemple de solution au problème de toxicité de la vanilline ne nécessitant pas de modifier génétiquement le microorganisme producteur.

Un autre facteur limitant l'accumulation de la vanilline est que cette molécule peut être dégradée par les microorganismes en un sous-produit possédant une toxicité inférieure. C'est le cas entre autres des microorganismes capables de produire une enzyme nommée vanilline déshydrogénase (*vdh*) qui catalyse l'oxydation de la vanilline en acide vanillique. Pour contrer cette perte de rendement, il est possible de supprimer le gène *vdh* codant pour cette enzyme. Cette stratégie a été utilisée avec l'actinobactérie *Amycolatopsis* sp. cultivée en présence d'acide férulique. La concentration maximale de vanilline accumulée dans le milieu de culture est passée de 2,9 mM pour la souche sauvage

à 6,8 mM pour le mutant (Fleige et al. 2013). L'inactivation d'un seul gène a ainsi permis d'augmenter le rendement de conversion de l'acide férulique en vanilline de 17,5 % à 41,1 %.

1.6 Les microorganismes photosynthétiques comme potentiels hôtes pour la production de vanilline par ingénierie métabolique

Les microorganismes photosynthétiques ont la capacité de croître en absence de source de carbone organique. De plus, leur capacité de séquestration du carbone leur permet de valoriser les rejets industriels de CO₂ tout en réduisant leurs impacts sur les gaz à effet de serre (Aslam et al. 2017; Mona et al. 2021). Le CO₂ capté par les microorganismes peut ainsi être converti en biomasse algale. Le potentiel de valorisation de cette biomasse provient des divers composés qui peuvent en être extraits. Par exemple, dû à la grande diversité de microalgue retrouvée dans la nature, une grande variété de pigments peuvent être extraits à partir de ces microorganismes (Zhuang et al. 2022). En ce sens, il est possible d'extraire entre 2 et 60 mg de fucoxanthine par gramme de biomasse de la diatomée *Phaeodactylum tricornutum* en fonction des conditions de cultures (Wang et al. 2021). La fucoxanthine est un composé d'intérêt entre autres pour ses propriétés anticancéreuses et anti-inflammatoires en plus de son potentiel effet anti-obésité (Xiao et al. 2020). D'un point de vue de la production de vanilline, les microalgues ne semblent pas être exploitées pour la production de ce composé aromatique. Cependant, de la vanilline peut être extraite d'une souche sauvage de la microalgue verte *S. spongiosa* (Onofrejevová et al. 2010). En revanche, il n'a pas encore été démontré qu'il est possible de modifier génétiquement des microalgues pour produire de la vanilline.

Dans le but de développer par ingénieries métaboliques des souches de microalgues productrices de vanilline, les organismes modèles peuvent être de bons candidats pour effectuer une preuve de concept. La littérature permet d'avoir accès à une grande quantité d'informations à leur sujet concernant les outils moléculaires, la métabolomique, la génomique et la protéomique. Les microorganismes photosynthétiques modèles sont la

cyanobactérie *Synechocystis* sp. PCC6803, l'algue unicellulaire verte *Chlamydomonas reinhardtii* et la diatomée *P. tricornutum* (Lin et al. 2019; Vavitsas et al. 2019).

1.6.1 Les cyanobactéries

Les cyanobactéries peuvent être transformées avec de l'ADN plasmidique qui est soit maintenu hors de l'ADN génomique soit intégré dans le génome par recombinaison homologue (Yadav, Rautela, et Kumar 2021). Il a été rapporté dans la littérature que l'instabilité génétique des cyanobactéries peut limiter le développement de souche (Jones 2014). Entre autres, le niveau de ploïdie de ces bactéries peut être instable passant d'un niveau variable de polyploïdie à la monoploïdie en fonction du stade et des conditions de la culture (Zerulla, Ludt, et Soppa 2016). Cette instabilité génétique pourrait ainsi nuire à la production d'un métabolite d'intérêt. Par exemple, le passage de la polyploïdie à la monoploïdie pourrait causer la perte de la construction génique d'intérêt si le clone cultivé est hétérozygote. Cependant, elles ont l'avantage d'avoir une croissance rapide ainsi qu'une grande variété d'outils moléculaires disponibles pour modifier génétiquement ces microorganismes (Vavitsas et al. 2019). Au moins 11 promoteurs inductibles ont été étudiés pour le contrôle de l'expression des gènes, dont ceux qui sont inductibles par exposition à la lumière verte (Abe et al. 2014), au rhamnose (Kelly et al. 2018) ou à certains métaux tels que le zinc, le cobalt et le cuivre (Peca et al. 2008). Par contre, les connaissances limitées sur le métabolisme des cyanobactéries sont un désavantage pour l'utilisation de ces microorganismes dans un but d'ingénierie du métabolisme. La création de modèles de prédiction lors de la conception de voie de biosynthèse artificielle peut ainsi être limitée pour ce microorganisme (Vavitsas et al. 2019).

1.6.2 L'algue unicellulaire verte *Chlamydomonas reinhardtii*

Du côté de l'algue unicellulaire verte *C. reinhardtii*, ses trois génomes, c'est-à-dire nucléaire, chloroplastique et mitochondrial, ont été séquencés (Merchant et al. 2007). De plus, les trois peuvent être modifiés génétiquement (Larosa et Remacle 2013; Siddiqui et al. 2020; Zhang, Wang et Wang 2021). Cette microalgue peut être transformée par

bombardement de microparticules recouvertes d'ADN, par électroporation, par agitation avec des billes de verres et par agroinfiltration (Scaife et al. 2015). Plusieurs outils moléculaires, éléments génétiques et banques de données sur le protéome, le transcriptome et le métabolome sont présentés dans diverses revues de littérature (Reijnders et al. 2014; Scaife et al. 2015; Hlavova, Turoczy, et Bisova 2015; Doron, Segal, et Shapira 2016). Des méthodes standardisées pour faciliter la construction de vecteurs ont aussi été développées (Lauersen, Kruse, et Mussnug 2015; Crozet et al. 2018). Pour limiter la taille du plasmide à transformer, il est possible de produire plus d'une protéine à partir d'un seul ARN messenger (ARNm). Pour ce faire, les gènes d'intérêts sont séparés par la séquence codante d'un peptide 2A d'origine virale (Dehghani et al. 2020).

Cependant, au niveau du génome nucléaire, l'ADN hétérologue est intégré aléatoirement par jonction d'extrémités non homologues. Ce phénomène limite le contrôle de la transformation et cause une variation d'expression entre les clones (Zhang, Wang et Wang 2021; Shahar et al. 2019). Il est alors possible que plusieurs copies s'intègrent dans le génome ou que l'intégration cause un effet de position qui peut être nuisible à l'expression du transgène ou à la croissance de la souche transformée (Vavitsas et al. 2019). Malgré tout le potentiel qu'offrent les connaissances disponibles sur *C. reinhardtii*, les projets d'ingénierie métabolique peuvent être limités par ce caractère aléatoire de la modification génétique.

1.6.3 La diatomée *Phaeodactylum tricornutum*

La diatomée *P. tricornutum* est un organisme photoautotrophe ou mixotrophe qui peut croître en présence de glucose, de glycérol, d'acétate ou de fructose (Cerón et al. 2006; Villanova et al. 2017; Villanova et Spetea 2021). Cette diatomée peut également croître en mode hétérotrophe lorsqu'elle est transformée avec un gène codant pour un transporteur de glucose, ce qui lui permet de croître à la noirceur lorsque du glucose est présent dans le milieu de culture (Zaslavskaja et al. 2001; Hamilton et al. 2016).

Contrairement aux autres diatomées, la présence de silice dans la paroi cellulaire de *P. tricornutum* n'est pas nécessaire pour sa croissance. (De Martino et al. 2011). L'absence d'un frustule rigide à base de silice entourant la cellule permettrait à *P. tricornutum* de modifier sa morphologie cellulaire et de s'adapter aux conditions environnementales (De Martino et al. 2011). En effet, les cellules de *P. tricornutum* peuvent être observées principalement sous quatre formes, soit la triradiée, la fusiforme, l'ovale ou la ronde (Figure 1.6) (De Martino et al. 2011). Les différentes conditions pour observer chacun des morphotypes ont été évaluées par De Martino et al. (2011). Par exemple, les cellules de formes ovales seront surreprésentées lorsque la température ou la salinité du milieu se retrouvent sous les conditions optimales, en absence d'agitation, sur un milieu de culture solide et lors de la phase stationnaire. La forme ovale est ainsi associée aux conditions de stress environnemental et la forme ronde augmente en proportion par la suite lorsque ces conditions sont maintenues. Les cellules ovales et rondes peuvent former des agrégats et se fixer aux parois de la verrerie utilisée en laboratoire. La forme ronde de son côté possède la capacité de former des biofilms où elle pourra y survivre pendant plusieurs mois. Les cultures de cellules rondes et ovales peuvent devenir majoritairement composées de cellules fusiformes lorsque les conditions de cultures redeviennent optimales. Par la suite, des cellules de la forme triradiée sont produites lorsque les conditions optimales sont maintenues. Par rapport aux trois autres formes, les cellules ovales ont la capacité unique de pouvoir se déplacer sur les surfaces solides, telles que les géloses ou les roches (Iwasa et Shimizu 1972). Cette propriété de *P. tricornutum* a pour conséquence de former deux types de colonies sur les géloses, des colonies rondes pour les cellules fusiformes et des colonies diffuses pour les cellules ovales. En plus des quatre formes précédemment décrites, les cellules de la souche CCMM 2004 peuvent être cruciformes dans des proportions qui peuvent varier entre 31,3 % et 55,3 % selon les conditions de culture (He, Han, et Yu 2014).

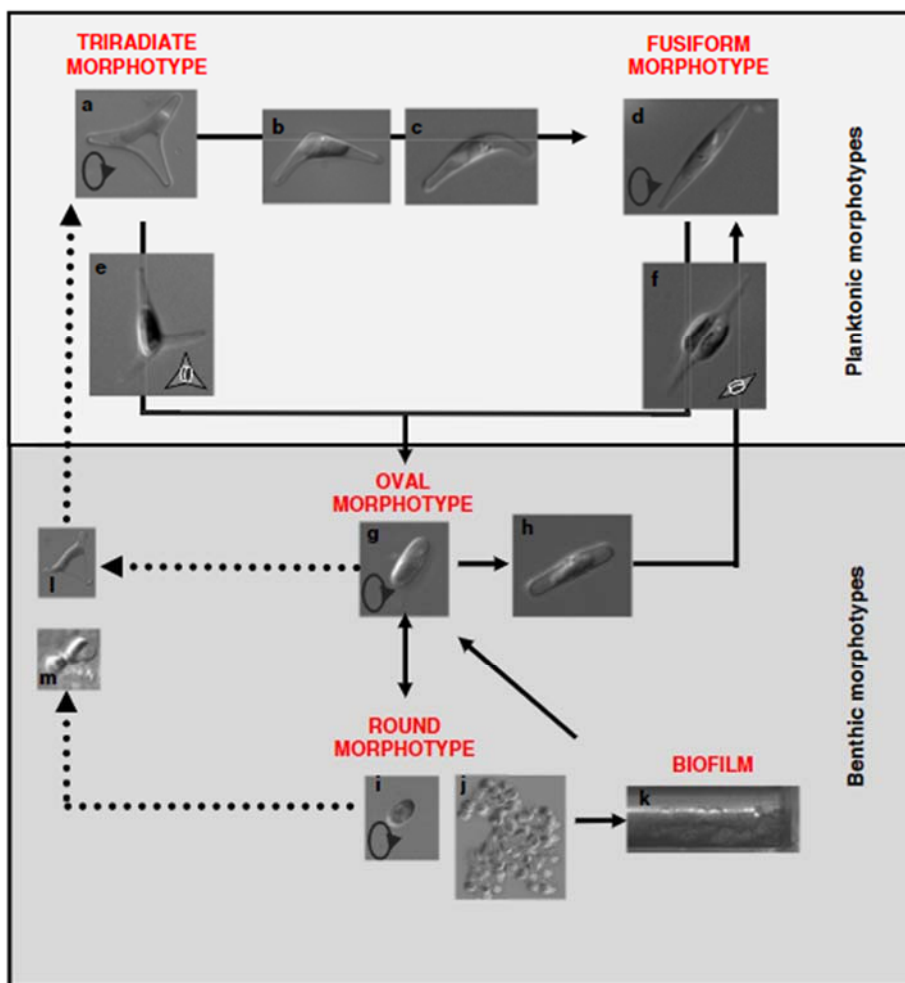


Figure 1.6 Schéma des différentes morphologies cellulaires de *P. tricornutum*. Les cellules triradiées (a) et fusiformes (d) sont associées au morphotype planctonique, c'est-à-dire libre dans l'eau tandis que les formes ovales (g) et rondes (i) sont considérées comme les morphotypes benthiques, c'est-à-dire vivant dans les fonds marins. Les cellules rondes et ovales forment des agrégats (j) et par la suite des biofilms (k). Les flèches indiquent vers quel type cellulaire se font les conversions morphologiques. Les photos b, c, e, f, h, l et m représentent des formes de transitions. Source de l'image : De Martino et al. (2011).

En ce qui a trait à l'ingénierie métabolique, la diatomée *P. tricornutum* peut être transformée par bombardement de microparticules (Miyagawa et al. 2009), par électroporation (Zhang et Hu, 2014) et par conjugaison bactérienne (Karas et al. 2015). Les outils moléculaires pour *P. tricornutum* ne sont pas aussi variés que ce qui est disponible pour *C. reinhardtii* et les cyanobactéries. Quelques promoteurs inductibles au niveau du génome nucléaire ont été étudiés pour la conception de plasmides.

Trois promoteurs nommés NR, DGAT1 et AMT sont inductibles de la même façon, c'est-à-dire en remplaçant l'ammonium dans le milieu de culture par du nitrate (Niu et al. 2012; Chu et al. 2016; Adler-Agnon (Shemesh) et al. 2018). Le promoteur AP1 du gène codant pour une phosphatase alcaline permet d'induire l'expression d'un gène hétérologue dans un milieu de culture contenant une forte concentration de phosphate (Lin et al. 2017). D'autres promoteurs ont été étudiés pour leur induction, par exemple lorsque la concentration en fer est limitée dans le milieu de culture (Yoshinaga et al. 2014). Quant au promoteur CA1, il permet l'expression d'un gène d'intérêt lorsque les cultures sont aérées avec l'air ambiant, mais il se retrouve réprimé lorsque la concentration de CO₂ est augmentée à 5 % (Harada et al. 2005).

En ce qui concerne les marqueurs de sélection, *P. tricornutum* peut acquérir la résistance à la zéocine, la blasticidine S et la nouséothricine en étant transformé respectivement avec les gènes de résistance à la bléomycine (*Sh ble*) (Zaslavskaia et al. 2000), le gène de la blasticidine S désaminase (*bsr*) (Buck et al. 2018) et celui de la nouséothricine acétyltransférase (*NAT*) (Karas et al. 2015). Des souches auxotrophes à l'uracile ou à l'histidine ont été développées en utilisant la technologie CRISPR/Cas9 afin d'inactiver respectivement du génome nucléaire les gènes *uridine-5-monophosphate synthase (UMPS)* et *phosphoribosyl-ATP pyrophosphohydrolase/phosphoribosyl-AMP cyclohydrolase (PRA-PH/CH)* (Slattery et al. 2020). Ces deux souches auxotrophes permettent d'obtenir des marqueurs de sélections chez *P. tricornutum* qui ne dépendent pas de l'utilisation des antibiotiques.

Comparativement à *C. reinhardtii*, un système d'expression extrachromosomique a été développé pour *P. tricornutum* permettant à un plasmide d'être propagé en tant qu'épisome (Karas et al. 2015). Le terme épisome chez les eucaryotes concerne une molécule d'ADN circulaire qui est répliqué par une cellule hôte sans s'intégrer dans son génome. L'épisome permet ainsi de propager de l'ADN hétérologue dans la diatomée en éliminant ainsi les problèmes d'effet de position liés à l'intégration. Afin de développer cet épisome, Karas et al. (2015) ont tout d'abord cloné des portions de chromosome de *P. tricornutum* dans le plasmide p0521s. Ils ont ainsi tenté d'évaluer si ces segments

d'ADN génomique pouvaient avoir des propriétés de centromère et ainsi permettre au plasmide de se répliquer dans la diatomée. En transformant par conjugaison bactérienne le plasmide p0521s contenant le gène de résistance *Sh ble*, des colonies résistantes à la zéocine ont été obtenues. Karas et al. (2015) ont ensuite retiré spécifiquement certaines portions du plasmide p0521s pour confirmer laquelle des séquences procurait la capacité d'épisome au plasmide. Ils ont ainsi démontré qu'ils n'obtenaient pas de clones à la suite de la conjugaison bactérienne en retirant la séquence CEN6-ARSH4-HIS3. Cette séquence provient du plasmide « vide » pCC1BAC-LC yeast qui a été utilisé comme base pour construire le plasmide p0521s. En analysant la séquence CEN6-ARSH4-HIS3, il a été démontré que la caractéristique qui permettait à cette séquence d'agir comme centromère dans *P. tricornutum* était son faible contenu en nucléotides guanine et cytosine (%GC) (Diner et al. 2016). Des essais ont permis de conclure que la majorité des séquences testées de plus de 500 pb avec un %GC de moins de 33 % permettaient à différents plasmides transformés dans *P. tricornutum* d'être propagés comme épisome (Diner et al. 2017).

La propagation d'ADN hétérologue via un épisome offre l'avantage d'éviter les effets de position, c'est-à-dire les intégrations dans les régions faiblement exprimées du génome et les inactivations aléatoires des gènes endogènes de *P. tricornutum*. Il est alors possible que les transformations soient prévisibles et reproductibles puisqu'il n'y aurait pas d'effets de position. Ces propriétés de l'épisome sont des avantages intéressants pour les projets d'ingénierie métabolique par rapport aux cyanobactéries et *C. reinhardtii*.

1.7 Objectifs

La voie de biosynthèse de la vanilline provenant de *V. planifolia* n'a jamais été complètement reproduite dans un microorganisme. L'épisome comme outil moléculaire est donc un bon candidat pour introduire cette voie de plusieurs gènes dans un microorganisme photosynthétique. L'objectif général de cette thèse est d'utiliser cet outil moléculaire récent dans le but d'évaluer le potentiel de biosynthèse *de novo* de la vanilline dans la diatomée *P. tricornutum*.

Les objectifs spécifiques de cette thèse sont divisés en deux axes; 1) évaluer si la voie de biosynthèse de la vanilline (Figure 1.2) peut être propagée via un épisode dans *P. tricornutum* en plus d'être fonctionnelle et 2) étudier la stabilité génétique de l'épisode ainsi que la production de protéine à partir de celui-ci.

Finalement, l'étude de la littérature réalisée au cours de cette thèse a permis de publier une publication de type opinion en lien avec la biosynthèse de la vanilline chez *V. planifolia* qui est présentée au chapitre IV. Elle contient des observations et commentaires par rapport aux résultats de Gallage et al. (2018) qui ont été utilisés pour conclure que les chloroplastes de *V. planifolia* contiennent l'ensemble de la machinerie enzymatique pour convertir la L-phénylalanine en vanilline. Les détails fournis par cette publication permettront d'apporter de nouvelles perspectives dans l'étude de la biosynthèse de la vanilline chez *V. planifolia*.

CHAPITRE II

REPRODUCTION DE LA VOIE DE BIOSYNTHÈSE DE LA VANILLINE PROVENANT DE LA PLANTE *VANILLA PLANIFOLIA* VIA UN SYSTÈME D'EXPRESSION EXTRACHROMOSOMIQUE DANS LA DIATOMÉE *PHAEODACTYLUM TRICORNUTUM*

2.1 Introduction

La découverte d'une séquence exogène permettant de maintenir potentiellement n'importe quel plasmide sous forme d'épisome chez la diatomée *P. tricornutum* (Karas et al. 2015) a mis de l'avant un nouveau potentiel d'ingénierie métabolique pour ce microorganisme. Ce nouvel outil moléculaire pourrait ainsi permettre d'insérer dans *P. tricornutum* des voies de biosynthèse impliquant plusieurs gènes hétérologues sans les problématiques liées à l'intégration dans le génome. Pour évaluer ce potentiel et dans le cadre de cette thèse de doctorat, une collaboration a été réalisée avec le professeur Bogumil Karas et son équipe de recherche. Il en a résulté une publication dans le journal ACS Synthetic Biology (Slattery et al. 2018) concernant l'étude de l'épisome chez *P. tricornutum*. Entre autres, les résultats présentés dans cet article ont démontré que plusieurs promoteurs endogènes à *P. tricornutum* (EF-1 α , 40SRPS8, H4-1B, γ -tubuline, RBCMT, FcpB, FcpC, FcpD et RBCMT) permettent d'exprimer le gène de résistance à la zéocine *Sh ble* contenu dans un épisome. Ces résultats ont permis de rendre disponibles de nouveaux promoteurs pour la conception de cassettes d'expression à insérer dans *P. tricornutum*.

La publication de Slattery et al. (2018) a aussi démontré que la diatomée *P. tricornutum* pouvait contenir et propager un épisome nommé p0521s-V contenant la voie de biosynthèse de la vanilline chez *V. planifolia* élucidée par Gallage et al. 2014 (Chapitre I – Figure 1.2). Les clones qui ont été analysés ont permis de démontrer qu'il était possible de propager l'épisome p0521s-V durant quatre mois sans qu'il n'y ait de réarrangements au cours de cette période. Cependant, des mutations ont été détectées au

niveau des épisomes extraits à partir des clones de *P. tricornutum* recombinants (Slattery et al. 2018). Ces mutations sont survenues soit durant le processus de clonage du plasmide, soit durant la transformation par conjugaison bactérienne ou lors du premier mois de culture des clones (Slattery et al. 2018). Les travaux présentés dans la publication de Slattery et al. (2018) ne concernaient pas la production de la vanilline par la diatomée contenant l'épisome p0521s-V. Les expériences en lien à ce sujet seront présentées dans ce chapitre.

Dans la voie de biosynthèse de la vanilline élucidée par Gallage et al. (2014) qui est illustrée à la figure 1.2 du chapitre I, l'acide aminé L-phénylalanine est le point de départ du métabolisme spécialisé de la vanilline. Afin de fournir une deuxième source de précurseurs à cette voie, l'enzyme TAL avait aussi été intégré dans l'épisome p0521s-V (Figure 2.1) étudié par Slattery et al. (2018).

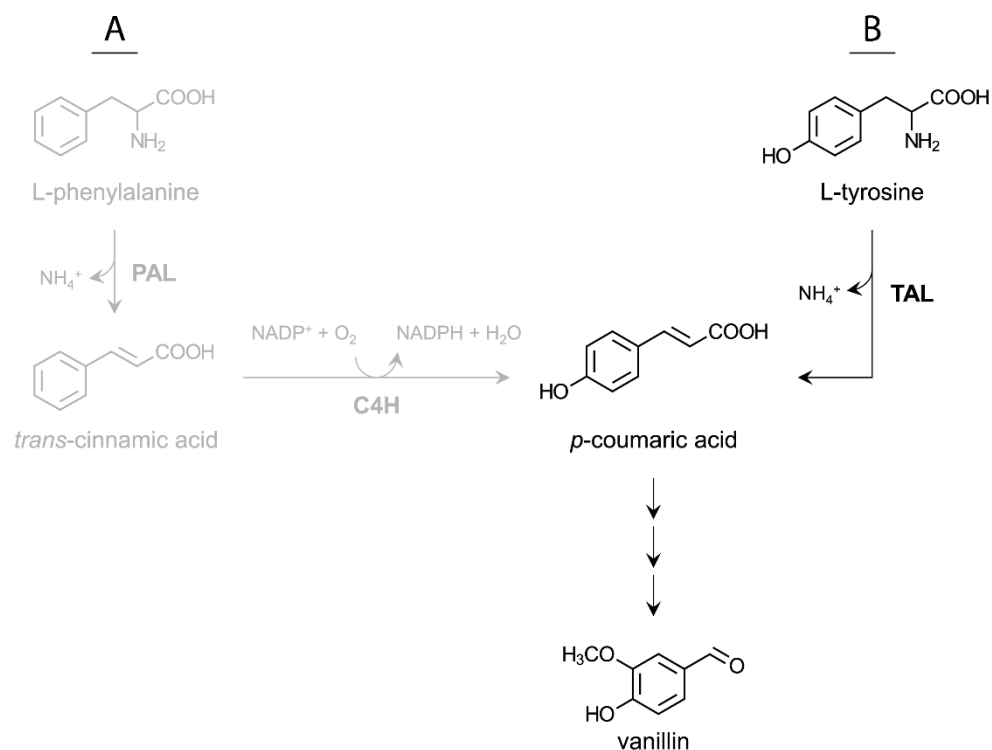


Figure 2.1 Schéma représentant l'ajout d'un point d'entrée pour la voie de biosynthèse de la vanilline via l'enzyme TAL.

(A) Biosynthèse de la vanilline à partir de la L-phénylalanine (Gallage et al. 2014). (B) Ajout d'un deuxième branchement pour la production de vanilline via la L-tyrosine. PAL, Phenylalanine ammonia-lyase; C4H, Cinnamate 4-hydroxylase; TAL, tyrosine ammonia-lyase.

Les nouveaux promoteurs étudiés par Slattery et al. (2018) ont été sélectionnés pour construire les cassettes d'expression de l'épisome p0521s-V. Seulement quatre promoteurs ont été utilisés pour les huit transgènes dans le but de limiter les différences entre les profils d'expression. Les gènes d'intérêts pour la voie de biosynthèse de la vanilline ont donc été clonés en paires dans l'épisome p0521s-V pour former quatre cassettes d'expression.

Dans le but de produire les huit protéines d'intérêts à partir de quatre cassettes d'expression, la séquence codante d'un peptide 2A a été insérée entre les gènes d'intérêts de chaque paire. Les peptides 2A sont des séquences peptidiques d'origine virale généralement de 18 à 22 acides aminés (Liu et al. 2017) qui permettent de produire deux protéines à partir d'un seul ARNm. La deuxième protéine est produite à la suite de l'inhibition de la formation du lien peptidique entre les deux derniers acides aminés, soit la glycine et la proline, de la séquence du peptide 2A. Ainsi, la protéine en cours de synthèse est libérée du ribosome et par la suite la traduction de l'ARNm se poursuit en produisant une deuxième protéine ayant en N-terminale la dernière proline du peptide 2A. Les différents peptides 2A possèdent une région conservée en C-terminale; D(V/I)EXNPG[↓]P où X représente n'importe qu'elle acide aminé et [↓] indique le site où le lien peptidique n'est pas formé (Wang et Marchisio 2021). Le mécanisme du peptide 2A causant la formation de deux protéines à partir d'un seul ARNm n'est pas encore complètement élucidé (Liu et al. 2017). Quelques modèles ont été proposés dans des revues de littérature. Par exemple, Brown et Ryan (2010) suggèrent entre autres que les peptides 2A causeraient un changement de conformation du ribosome qui permettrait une association du facteur de terminaison eRF1 au site catalytique de la liaison peptidique, et ce, sans la présence d'un codon de terminaison. Il s'en suivrait un relâchement de la protéine en cours de synthèse. Luke et Ryan (2018) ont quant à eux proposé que les peptides 2A induisent des changements dans la structure du ribosome causant une réorientation du peptidyl-ARNt. Le groupement ester du peptidyl-ARNt se retrouverait éloigné du site de formation du lien peptidique à la suite de ce changement d'orientation. Il y aurait ainsi inhibition de la formation du lien peptidique entre la glycine et la dernière proline du peptide 2A.

L'objectif de ce chapitre est d'évaluer le potentiel de l'épisome p0521s-V pour la production *de novo* de la vanilline par la diatomée *P. tricornutum*.

2.2 Matériels et méthodes

2.2.1 Souches microbiennes et conditions de culture

La bactérie *E. coli* NEB® 10-beta (n° cat. C3019H, New England BioLabs) a été cultivée dans le milieu Luria-Bertani (LB) supplémenté avec l'antibiotique chloramphénicol (15 mg/L) à 37 °C.

La diatomée *Phaeodactylum tricornutum* (n° cat. CCAP 1055/1, Culture Collection of Algae and Protozoa, Scotland, United Kingdom) nous a été fournie par le professeur Bogumil Karas (University Of Western Ontario). Cette diatomée a été cultivée dans un milieu de culture L1 modifié (section 2.2.3) sans silice à 18 °C sous un éclairage de lumière fluorescente blanche ayant une intensité de 75 $\mu\text{E}/\text{m}^2/\text{s}^1$ durant une photopériode de 16h/8h de lumière/noirceur. Les cultures liquides ont été agitées à 130 rpm.

2.2.2 Liste des produits chimiques

Les produits suivants ont été utilisés pour la préparation du milieu de culture L1 :

NaCl (n° cat. S271, Fisher Scientific), Na₂SO₄ (n° cat. S9627, Sigma-Aldrich), KCl (n° cat. P9541, Sigma-Aldrich), NaHCO₃ (n° cat. S-4772, Sigma-Aldrich), KBr (n° cat. 12612, Alfa Aesar), H₃BO₃ (n° cat. B-7660, Sigma-Aldrich), NaF (n° cat. S7920, Sigma-Aldrich), MgCl₂·6H₂O (n° cat. BP214, Fisher Scientific), CaCl₂·2H₂O (n° cat. C7902, Sigma-Aldrich), NaNO₃ (n° cat. S8170, Sigma-Aldrich), NaH₂PO₄·H₂O (n° cat. S369, Fisher Scientific), FeCl₃·6H₂O (n° cat. LC142901, LabChem), Na₂EDTA·2H₂O (n° cat. BP120, Fisher Scientific), CuSO₄·5H₂O (n° cat. C7631, Sigma-Aldrich), Na₂MoO₄·2H₂O (n° cat. S-6646, Sigma-Aldrich), ZnSO₄·7H₂O (n° cat. 4384-01, J.T. Baker), CoCl₂·6H₂O, MnCl₂·4H₂O (n° cat. M87, Fisher Scientific), H₂SeO₃ (n° cat. 12574, Alfa Aesar), NiSO₄·6H₂O, Na₃VO₄

(n° cat. P0758S, New England BioLabs), K_2CrO_4 , thiamine-HCl (n° cat. 47858, Sigma-Aldrich), biotine (n° cat. B0463, TCI America), cyanocobalamine (n° cat. V6629, Sigma-Aldrich).

2.2.3 Préparation du milieu de culture L1 modifié pour *P. tricornutum*

La composition du milieu de culture et sa préparation ont été modifiées et adaptées de l'article de Slattery et al. (2018). Un litre du milieu L1 modifié consiste à 917 mL d'eau distillée et stérilisée par autoclave, 50 mL d'une solution mère 20x de NaCl et Na_2SO_4 , 10 mL d'une solution mère 100x de sels anhydres, 20 mL d'une solution mère 50x de sels hydratés, 2 mL d'une solution mère de phosphate/sodium, 1 mL de la solution mère des métaux traces L1 et 0,5 mL de la solution mère de vitamines f/2. Pour préparer les géloses, 0,5 volume de milieu L1 modifié stérile a été mélangé à 0,5 volume d'une solution stérilisé par autoclave d'agar 2 % et versé dans des boîtes de Petri.

La solution mère 20x de NaCl et Na_2SO_4 consiste à 245 g/L de NaCl et 40,9 g/L de Na_2SO_4 stérilisés par autoclave. La solution mère 100x de sels anhydres consiste à 35 g/L de KCl, 10 g/L de $NaHCO_3$, 5 g/L de KBr, 1,5 g/L de H_3BO_3 et 0,15 g/L de NaF stérilisés par autoclave. La solution mère 50x de sels hydratés consiste à 277,5 g/L de $MgCl_2 \cdot 6H_2O$ et 38,5 g/L de $CaCl_2 \cdot 2H_2O$ stérilisés par autoclave. La solution mère de phosphate/sodium consiste à dissoudre 37,5 g $NaNO_3$ avec 2,5 g $NaH_2PO_4 \cdot H_2O$ dans 100 mL d'eau distillée et de stériliser la solution par filtration en utilisant un filtre avec des pores de 0,2 μm de diamètre. La solution mère des métaux traces L1 consiste à dissoudre et mélanger 3,15 g de $FeCl_3 \cdot 6H_2O$; 4,36 g de $Na_2EDTA \cdot 2H_2O$; 0,25 mL de $CuSO_4 \cdot 5H_2O$ (9,8 g/L dH₂O); 3,0 mL de $Na_2MoO_4 \cdot 2H_2O$ (6,3 g/L dH₂O); 1,0 mL de $ZnSO_4 \cdot 7H_2O$ (22,0 g/L dH₂O); 1,0 mL de $CoCl_2 \cdot 6H_2O$ (10,0 g/L dH₂O); 1,0 mL de $MnCl_2 \cdot 4H_2O$ (180,0 g/L dH₂O); 1,0 mL de H_2SeO_3 (1,3 g/L dH₂O); 1,0 mL de $NiSO_4 \cdot 6H_2O$ (2,7 g/L dH₂O); 0,1 mL de Na_3VO_4 (18,4 g/L dH₂O); et 1,0 mL de K_2CrO_4 (1,94 g/L dH₂O) dans 1L d'eau distillée puis de stériliser par filtration avec un filtre 0,2 μm de diamètre de pores. La solution mère de vitamines f/2 consiste à dissoudre 200 mg de thiamine-HCl avec 10 mL de biotine

(0,1 g/L dH₂O) et 1 mL de cyanocobalamine (1 g/L dH₂O) dans 1L dH₂O puis de stériliser par filtration avec un filtre 0,2 µm de diamètre de pores.

2.2.4 Origine du plasmide p0521s-V et des clones de *P. tricornutum*

Les clones analysés dans ce chapitre proviennent d'une collaboration avec le laboratoire du professeur Bogumil Karas (University Of Western Ontario). Nous avons sélectionné huit enzymes déjà caractérisées dans la littérature pour reproduire la voie de biosynthèse de la vanilline. L'épisome p0521s-V et les clones de *P. tricornutum* ont par la suite été fournis par l'équipe du professeur Bogumil Karas. La méthodologie détaillée de ce qui a été réalisé par son équipe peut être retrouvée dans l'article de Slattery et al. (2018).

Brièvement, voici ce qui a été réalisé au cours de cette collaboration pour produire l'épisome p0521s-V. Tout d'abord, les gènes codant pour les enzymes présentées dans le tableau 2.1 ont été assemblés en paires et ils ont été séparés par la séquence prolongée du peptide 2A provenant du virus causant le syndrome pieds-mains-bouche (FMDV2A) (Plucinak et al. 2015). La séquence de l'étiquette poly-histidine a été ajoutée à trois gènes (Figure 2.2B). Elle a été ajoutée en N-terminale aux gènes *ZmPAL* et *ZmTAL*. Pour le troisième gène, la séquence de l'étiquette a été ajoutée à l'extrémité 3' du gène *VpVAN*. Les paires de gènes séparées par la séquence du peptide FMDV2A ont été synthétisées, sans optimisation des codons, par la compagnie Bio Basic (Markham, Ontario, Canada). Les quatre gènes bicistroniques synthétisés ont ensuite été assemblés avec différents promoteurs et terminateurs pour former quatre cassettes d'expression nommées DMi1, DMi2, DMi3 et DMi4.

Tableau 2.1

Origine des gènes clonés dans le plasmide p0521s-V

Enzyme	Organisme d'origine	Numéro d'accès GeneBank	Références
Phenylalanine Ammonia-Lyase (<i>ZmPAL</i>)	<i>Zea mays</i>	BT062346.1	Zang et al. 2015
Cinnamate 4-hydroxylase (<i>AtC4H</i>)	<i>Arabidopsis thaliana</i>	U37235.1	Urban et al. 1997
Tyrosine ammonia-lyase (<i>ZmTAL</i>)	<i>Zea mays</i>	L77912.1	Rösler et al. 1997
Coumarate 3-hydroxylase (<i>AtC3H</i>)	<i>Arabidopsis thaliana</i>	NM_180006.2	Nair et al. 2002; Franke et al. 2002
Hydroxycinnamoyltransférase (<i>AtHCT</i>)	<i>Arabidopsis thaliana</i>	NM_124270.3	Hoffmann et al. 2003, 2004
4-coumarate:CoA ligase (<i>At4CL</i>)	<i>Arabidopsis thaliana</i>	BT005912.1	Block et al. 2014
Acide caféique/acide 5-hydroxyfêrulique O-méthyltransférase (<i>VpCOMT</i>)	<i>Vanilla planifolia</i>	AY555144.1	Pak et al. 2004
Vanilline synthase (<i>VpVAN</i>)	<i>Vanilla planifolia</i>	KP278240.1	Gallage et al. 2014

La cassette DMi1 est composée du gène de la *ZmPAL* à l'extrémité 5' et de la *AtC4H* à l'extrémité 3'. Cette cassette est sous le contrôle du promoteur et du terminateur H4-1B. Pour la cassette nommée DMi2, le gène de la *ZmTAL* se retrouve en 5' et le gène de la *AtC3H* se situe 3'. Le promoteur et le terminateur 40SRPS8 ont été clonés avec cette cassette. La cassette DMi3 est formée du gène de la *AtHCT* à l'extrémité 5' et du gène de la *At4CL* à l'extrémité 3'. Pour cette troisième cassette, les promoteur et terminateur EF-1 α ont été sélectionnés. La dernière cassette d'expression a été nommée DMi4 et les gènes *VpCOMT* et *VpVAN* ont été placés respectivement aux extrémités 5' et 3'. Pour cette dernière cassette, le gène bicistronique a été placé entre les promoteur et terminateur FcpD. Les quatre cassettes d'expressions ont été clonées ensemble en utilisant comme base le plasmide vide p0521s (n° cat. 62862, Addgene) (Karas et al. 2015) pour former le plasmide nommé p0521s-V. Pour ce faire, la technique d'assemblage par recombinaison

homologue a été utilisée dans la levure *Saccharomyce cerevisiae*. Pour la sélection des clones de *P. tricornutum* recombinants, les plasmides p0521s et p0521s-V contenaient le gène *Sh ble* procurant la résistance à la zeocine, un antibiotique appartenant à la famille des bléomycines.

Le plasmide p0521s-V a été transféré dans *P. tricornutum* en utilisant le système de conjugaison bactérienne développé par Karas et al. (2015). Pour l'analyse des épisomes contenus dans les clones, ils ont été extraits des clones de *P. tricornutum* et par la suite transformés dans *E. coli*. Cette étape a permis de produire via *E. coli* les quantités de l'ADN plasmidiques nécessaires pour leurs analyses. La présence des gènes dans les épisomes a été confirmée par PCR multiplex et par l'analyse du profil de digestion par des enzymes de restriction. Trois clones de *P. tricornutum* nommés V1-3 #2, Vc-1#1 et Vp-7 #1 ont alors été sélectionnés pour développer les protocoles présentés dans ce chapitre en lien avec la production de vanilline.

2.2.5 Préparation d'une courbe de calibration de la densité cellulaire de *P. tricornutum* en fonction de la densité optique

Une courbe de calibration a été produite pour calculer la densité cellulaire et faire le suivi la croissance des cultures de *P. tricornutum* via la densité optique. Pour ce faire, une culture de 7 jours d'un clone de *P. tricornutum* contenant le plasmide vide (clone p0521s) et une autre d'un clone contenant le plasmide recombinant (clone V1-3 #2) ont été utilisées. Des mesures de densité optique et de compte cellulaire à partir de plusieurs dilutions des cultures ont été effectuées. Pour chaque dilution, trois mesures de la densité optique à 730 nm (DO_{730nm}) ont été prises et trois décomptes cellulaires ont été réalisés sur un hématimètre avec cellules de Neubauer modifiée. Pour les décomptes cellulaires, 10 μ L des dilutions ont été chargés dans l'hématimètre. Les décomptes des cellules ont été faits sur 9 carrés de 0,2 mm par 0,2 mm pour les dilutions où la DO_{730nm} mesurée était supérieure à une valeur de 0,1. Lorsque la DO_{730nm} était de 0,1 ou moins, les décomptes ont été faits sur trois carrés de 1 mm par 1 mm. Les décomptes obtenus à partir des cultures des clones p0521s et V1-3 #2 ont été compilés ensemble pour former une courbe de calibration.

2.2.6 Prétraitement des culots cellulaires de *P. tricornutum* pour l'analyse des phénylpropanoïdes par chromatographie en phase liquide à haute performance couplée à un spectromètre de masse en tandem (HPLC-MS/MS)

Des cultures âgées de 7 jours des clones de *P. tricornutum* ont été utilisées pour l'extraction des métabolites d'intérêts. Pour cela, des cultures de 50 mL ont été inoculées pour obtenir une DO_{730nm} de départ d'environ 0,1. Au bout de 7 jours, la densité optique a été mesurée pour calculer la densité cellulaire. Les cultures de 50 mL ont été centrifugées à 4 °C pendant 10 minutes à 3 500 x g. Par la suite, les culots ont été lavés à deux reprises avec 5 mL de milieu L1. Les culots lavés ont été congelés dans l'azote liquide et conservés à -80 °C jusqu'au moment de l'extraction.

Pour l'extraction des métabolites, les culots cellulaires ont été remis en suspension avec 5 mL de méthanol:acétone:eau (7:7:6) (Ayoub, de Camargo, et Shahidi 2016). Un volume de 50 µL d'acide isovanillique 50 mg/L a été ajouté comme standard interne aux échantillons. L'extraction des métabolites a été effectuée en incubant les échantillons toute une nuit à -20 °C. Les échantillons ont ensuite été centrifugés pendant 10 minutes à 3 500 x g. Les surnageants ont été récoltés puis évaporés complètement en utilisant un concentrateur à vide Savant Speedvac (no cat. SPD1010-115, Thermo Scientific) sans chauffer et avec une pression réglée à 5,1 torr.

2.2.7 Extraction sur phase solide des échantillons prétraités de *P. tricornutum* pour analyse par HPLC-MS/MS

L'extraction des composés phénoliques sur phase solide a été réalisée en utilisant des cartouches Strata-X-A (n° cat. 8B-S123-FCH, Phenomenex) et un collecteur à vide (n° cat. AH0-6023, Phenomenex). Les cartouches ont d'abord été conditionnées avec 3 mL de méthanol (MeOH) et équilibrées avec 3 mL d'acétate d'ammonium 100 mM à pH 9,8. Un volume de 5 mL d'acétate d'ammonium 100 mM à pH 9,8 a été ajouté aux échantillons prétraités et secs (section 2.2.6). Les échantillons ont ensuite été élués et les cartouches ont été lavées en appliquant 3 mL d'acétate d'ammonium 100 mM à pH 9,8 puis à deux reprises avec 3 mL de MeOH grade LC-MS. Les cartouches ont été séchées pendant 5 minutes en augmentant la force de la pompe à vide près de la limite maximale

du collecteur à vide. Les composés d'intérêts ont par la suite été récoltés en éluant à trois reprises avec 1 mL d'acide formique:MeOH (5:95). Les échantillons récoltés ont été évaporés complètement sans chauffage avec un concentrateur à vide Savant Speedvac avec une pression réglée à 5,1 torr. Les échantillons secs ont été conservés à -20 °C à la noirceur jusqu'au moment de leur analyse au HPLC-MS/MS.

2.2.8 Analyse par HPLC-MS/MS des composés phénoliques libres extraits de *P. tricornutum*

Les extraits secs ont été reconstitués dans 250 µL d'une solution d'acide formique 0,1 % v/v et de MeOH grade LC/MS (80:20) afin de respecter les conditions initiales de la phase mobile utilisée au HPLC-MS/MS. Une dilution 1/10 des échantillons a été effectuée avec une solution d'acide formique 0,1 % v/v et de MeOH (80:20) avant leur injection dans le système HPLC-MS/MS. Une solution de blanc contenant seulement la phase mobile a été injectée tous les trois échantillons pour s'assurer qu'il ne se produisait pas de transfert d'échantillons provenant des injections précédentes.

Le chromatographe HPLC-MS/MS (Agilent, QC, Canada) utilisé a été équipé d'une colonne Kinetex EVO C18 (150 x 4,6 mm; 5 µm; 100 Å) (n° cat. 00F-4633-E0, Phenomenex). Le compartiment de la colonne a été réglé à une température de 30 °C.

Le programme d'élution du HPLC adapté de Rico et al. (2013) était un gradient utilisant différentes proportions d'une phase mobile aqueuse d'acide formique 0,1 % v/v (A) et d'une phase mobile organique constituée de MeOH (B). Le programme d'élution a été débuté en mode isocratique avec 20 % de la phase mobile B pendant 5 minutes (0-5 min). Un gradient a ensuite été appliqué pendant 25 minutes afin d'augmenter la proportion en solvant organique B à 60 % (5-30 min). Cette proportion de 60 % en phase mobile B a été maintenue pendant 5 minutes (30-35 min) avant de revenir graduellement aux conditions initiales en 3 minutes (35-38 min). Une période d'équilibration de la colonne de 2 minutes supplémentaires (38-40 min) a été ajoutée à la fin des analyses afin de s'assurer que le système serait prêt au lancement lors de la prochaine injection.

Un volume de 5 μL a été injecté et un débit de 0,4 mL/min a été appliqué tout au long de l'injection.

Les analytes ont été séparés, fragmentés et ionisés par le triple quadripôle du MS/MS. La fragmentation a été effectuée au niveau du deuxième quadripôle via une collision induite par l'application d'un champ électrique. Les paramètres utilisés dans la source Agilent Jet Stream du MS/MS pour l'analyse en mode positif étaient les suivants : un débit de 10 L/min et une température de 300 °C pour le gaz de focus (sheath gas) circulant autour du nébuliseur, une pression de 45 psi pour le nébuliseur, un débit de 11L/min et une température de 300 °C pour le gaz d'assèchement au cône, un voltage de 3500 V au capillaire et de 500 V à la pointe du nébuliseur. Les paramètres ont été inspirés d'une méthode publiée par de Souza et al. (2016).

2.2.9 Validation de la méthode pour l'analyse des composés phénoliques libres par HPLC-MS/MS

Les composés suivants ont été ciblés pour le développement de la méthode : acide *p*-coumarique, acide caféique, acide férulique, vanilline et acide isovanillique. L'acide isovanillique a été utilisé comme standard interne (SI). Chaque composé a été analysé selon leur transition « Multiple-reaction monitoring » (MRM), c'est-à-dire la transition du ratio de la masse sur la charge (m/z) de l'ion parent et le ratio m/z observé de l'ion fragmenté $[\text{M}+\text{H}]^+$. Les sections suivantes décrivent les paramètres de validation de la méthode HPLC-MS/MS.

2.2.9.1 L'efficacité d'extraction lors de l'utilisation de cartouches Strata-X-A

Les taux de récupérations des composés d'intérêts ont été comparés lors de leur extraction à partir de deux matrices différentes. Les matrices évaluées étaient la solution tampon d'acétate d'ammonium 100 mM à pH 9,8 et un culot cellulaire de la souche sauvage de *P. tricornutum*. Pour chaque matrice, deux solutions (nommées S1 et S2) ont été analysées avec des triplicatas. La solution S1 représente l'ajout des composés ciblés et du SI aux matrices avant leur extraction en phase solide avec les cartouches Strata-X-A.

La solution S2 représente l'ajout des composés d'intérêts et du SI après l'extraction en phase solide avec les cartouches Strata-X-A. Le taux de récupération a été calculé avec le rapport des ratios d'aire sous la courbe des composés ciblés par rapport au SI (Équation 2.1). Pour chaque métabolite, les taux de récupération des deux matrices ont été comparés statistiquement avec l'analyse des variances de deux facteurs avec répétition d'expérience ($\alpha = 0.05$) en utilisant la macro complémentaire « utilitaire d'analyse » sur Microsoft Excel.

$$\text{Équation 2.1} \quad \text{Taux de récupération (\%)} = \frac{\text{Ratio d'aire pour } S_1}{\text{Ratio d'aire pour } S_2} \times 100$$

2.2.9.2 L'effet de matrice

Deux solutions (S3 et S4) ont été analysées avec des triplicatas lors de l'extraction en phase solide. La matrice de la solution S3 était un culot cellulaire de la souche sauvage de *P. tricornutum* et la solution S4 était une solution tampon d'acétate d'ammonium 100 mM à pH 9,8. Les protocoles de prétraitement et d'extraction sur phase solide ont été réalisés de la même manière pour les deux solutions. Pour les deux solutions, les composés ciblés et le SI ont été injectés à une concentration de 1 mg/L dans le système HPLC-MS/MS. L'effet de matrice a été calculé avec le rapport des ratios d'aire sous la courbe des composés ciblés par rapport au SI (Équation 2.2).

$$\text{Équation 2.2} \quad \text{Effet de matrice (\%)} = \frac{\text{Ratio d'aire pour } S_3}{\text{Ratio d'aire pour } S_4} \times 100$$

2.2.9.3 La linéarité

La linéarité de la courbe de calibration a été calculée en utilisant les concentrations de 0; 0,1; 0,3; 0,6; 1,0 et 2,0 mg/L. La courbe de calibration a été réalisée en ajoutant les composés d'intérêts et le SI aux culots cellulaires lors de leur remise en suspension pour effectuer le protocole de prétraitement (Section 2.2.6). Le SI a été ajouté afin d'obtenir une concentration calculée de 1 mg/L lors de l'injection dans le système HPLC-MS/MS. Les concentrations ont été testées avec des triplicatas à l'exception de la concentration de 1,0 mg/L qui a été effectuée avec des duplicatas.

2.2.9.4 Limite de détection (LDD) et limite de quantification (LDQ)

La LDD et la LDQ ont été calculées avec la plus petite concentration qui a été utilisée pour déterminer la courbe de calibration, c'est-à-dire 0,1 mg/L ($C = 0,1$ mg/L). Le rapport signal sur bruit (S/B) a été calculé pour chaque triplicata de cette concentration. La LDD (Équation 2.3) et la LDQ (Équation 2.4) ont été calculées en multipliant la concentration minimale de la courbe avec un facteur de 3 et 10 respectivement puis en calculant un ratio de la valeur obtenue par rapport au S/B (Peris-Vicente, Esteve-Romero, et Carda-Broch 2015). Les formules suivantes ont été utilisées pour calculer la LDD et la LDQ :

$$\text{Équation 2.3} \quad \text{LDD} = \frac{3 \times C}{\text{S/B}}$$

$$\text{Équation 2.4} \quad \text{LDQ} = \frac{10 \times C}{\text{S/B}}$$

2.2.9.5 La précision et l'exactitude

Les protocoles de prétraitement des culots cellulaires et de l'extraction sur phase solide ont été réalisés en ajoutant les composés ciblés et le SI à trois culots cellulaires de *P. tricornutum*. La concentration prévue de chaque composé était de 0,5 mg/L ($C_{\text{prévue}}$) au moment de l'injection dans le système HPLC-MS/MS. La concentration mesurée ($C_{\text{mesurée}}$) a été obtenue avec la courbe de calibration. La précision (Équation 2.5) est représentée en calculant le coefficient de variation (%CV) à partir de l'écart-type (s) et la moyenne (\bar{x}). L'exactitude (Équation 2.6) est exprimée par l'erreur relative (%RE). Les formules suivantes ont été utilisées pour calculer la précision et l'exactitude :

$$\text{Équation 2.5} \quad \%CV = \frac{s}{\bar{x}} \times 100$$

$$\text{Équation 2.6} \quad \%RE = \frac{C_{\text{mesurée}} - C_{\text{prévue}}}{C_{\text{prévue}}} \times 100$$

2.2.10 Détection des protéines avec une étiquette poly-histidine par immunobuvardage de type western

Pour l'extraction des protéines, des cultures de 20 mL de *P. tricornutum* ont été démarrées avec DO_{730nm} d'environ 0,1. Après 7 jours, les cultures ont été centrifugées à 1 500 xg pendant 20 minutes à 4 °C. Les culots ont été pesés et remis en suspension dans 1 mL de tampon d'extraction [50 mM Tris pH 7,4; 500 mM NaCl; 0,1 % Tween20, 1x cocktail d'inhibiteur de protéase] par 0,5 g de culot humide. La lyse cellulaire a été effectuée par sonication sur glace avec une amplitude de 35 % durant six périodes de 30 secondes qui ont été séparées par des pauses de 30 secondes. Les extraits ont été centrifugés à 20 000 xg durant 30 minutes à 4 °C. Les surnageants ont été conservés à -20 °C jusqu'à leur quantification avec le réactif DC™ (n° cat. 5000112, Bio-Rad) et la migration sur gel de polyacrylamide.

Pour chaque clone, 50 µg de protéines ont été séparés sur gel de polyacrylamide 10 % en présence de laurylsulfate de sodium (SDS-PAGE). Comme contrôle positif, 10 ng d'une protéine purifiée contenant une étiquette poly-histidine (n° cat. M0101, GenScript) a aussi été déposée sur gel. Le marqueur de poids moléculaire Precision Plus Protein Dual Color Standards (n° cat. 1610374, Bio-Rad) a aussi été migré sur gel. Les protéines ont été transférées en milieu humide sur une membrane PVDF Immun-Blot® (n° cat. 1620260, Bio-Rad) durant 2h30 à 100V constants. La membrane a été équilibrée 5 minutes dans une solution TBS (2,4 g/L Tris, 8,0 g/L NaCl, pH 7,6) puis elle a été lavée à trois reprises par des incubations avec agitation dans du TBST (TBS avec 0,1% de Tween® 20). La membrane a par la suite été incubée 1h à la température de la pièce dans du TBST supplémenté avec 5 % de lait en poudre écrémé. La membrane est ensuite incubée toute une nuit à 4 °C avec une solution de TBST contenant 3 % d'albumine de sérum bovin et un anticorps primaire anti-his (n° cat. MA121315, Invitrogen) dilué 1:1000. La membrane a de nouveau été lavée à trois reprises durant 10 minutes dans du TBST. Par la suite, la membrane a été incubée 1h à la température de la pièce dans une solution de TBST contenant 5 % de lait en poudre écrémé avec l'anticorps secondaire anti-souris conjugué à la peroxydase de raifort (n° cat. 170-5047, Bio-Rad) dilué 1:20 000. La membrane a été lavée avec du TBST à trois reprises durant 10 minutes. Finalement, la révélation a été

effectuée avec les substrats Clarity™ Western ECL (n° cat. 1705060S, Bio-Rad) par détection de chimiluminescence dans l'appareil d'imagerie ChemiDoc (n° cat. 12003153, Bio-Rad). La membrane a aussi été observée à la suite d'une coloration au rouge de Ponceau (5 % v/v d'acide acétique glacial et 0,1 % p/v du colorant rouge de Ponceau).

2.2.11 Extraction des ARN totaux et transcription inverse suivie d'une PCR (RT-PCR) en point final

Les ARN totaux ont été extraits à partir d'une culture de *P. tricornutum* contenant entre 1×10^7 et 1×10^8 cellules. La culture de *P. tricornutum* a été centrifugée et le culot cellulaire a été congelé dans l'azote liquide avant d'être conservé à -80 °C jusqu'au moment de l'extraction. Les ARN totaux ont été extraits et élués dans 50 µL d'eau exempts de ribonucléases en utilisant le RNeasy Plant Mini Kit (n° cat. 74904, QIAGEN). Pour éliminer la contamination d'ADN plasmidique, les échantillons d'ARN ont été par la suite incubés à 37 °C pendant 15 minutes en présence de 37 µL d'eau exempts de ribonucléases, 2,5 µL d'inhibiteur de RNase (n° cat. M0307S, New England BioLabs), 10 µL du tampon réactionnel 10x de la DNase I et 2,5 µL de DNase I 2000 U/mL (n° cat. M0303S, New England BioLabs). Les ARNs traités ont par la suite été purifiés en utilisant le Plant Total RNA Mini Kit (n° cat. RPD300, Geneaid). Afin de vérifier la qualité des ARNs extraits, 500 ng des ARNs totaux ont été déposés sur gel d'agarose 1 %.

La transcription inverse des ARNs totaux en ADN complémentaire (ADNc) a été réalisée en utilisant le High-Capacity cDNA Reverse Transcription Kit (n° cat. 4368814, Applied Biosystems). Pour chaque transcription inverse qui a été réalisée, un contrôle négatif a été fait en remplaçant la transcriptase inverse par de l'eau. La PCR a été réalisée avec la *Taq* polymérase et le tampon ThermoPol (n° cat. M0267, New England BioLabs) dans les conditions suivantes : dénaturation initiale (30 secondes à 95 °C); 30 cycles PCR de dénaturation (30 secondes), d'hybridation (40 secondes) et d'élongation (variable selon les cassettes d'expression) puis par une extension finale de 5 minutes à 68 °C. Les paramètres spécifiques pour l'amplification des quatre cassettes d'expression sont présentés dans le tableau 2.2. Les produits de la PCR ont été visualisés sur gel d'agarose 1 % en présence de SYBR Safe 1x (n° cat. S33102, Invitrogen).

Tableau 2.2

Paramètres utilisés pour chaque paire d'amorces dans le but d'amplifier l'ADNc des cassettes d'expression

Cassette	Séquence des amorces	Hybridation (°C)	Élongation (secondes)	Taille (paire de bases)
DMi1	Sens : GCTAACTACCCGCTGATG Anti-sens : TGAGCTAATCGACTTCTCTCAC	49	69	1154
DMi2	Sens : CGTGTTCTCCAAGATCACCAG Anti-sens : ATGCTTCCGTCGATCTGTTC	51	43	724
DMi3	Sens : TCTTGACTACCTGGAGATGCAG Anti-sens : GAGGTGTAGGGAAGGGAATTTG	51	30	505
DMi4.1	Sens : CATGTTGTTTCTGAAGCTCCAC Anti-sens : CTCCGATCCGTAGCTCTTC	50	42	705

2.2.12 Extraction des épisomes à partir des clones recombinants de *P. tricornutum*

Les épisomes ont été extraits en utilisant le Large Plasmid Mini Kit (n° cat. PDL300, Geneaid) et en adaptant les protocoles utilisés par Karas et al. (2015) et Slattery al. (2018). Pour chaque clone, un volume de 5 mL a été prélevé d'une culture de 7 jours et il a été centrifugé pendant 10 minutes à 3 500 x g. Les culots ont été remis en suspension avec 235 µL du tampon PDL1 (Geneaid), 5 µL d'une solution 100 mg/mL d'hémicellulase (n° cat. H2125, Sigma-Aldrich), 5 µL d'une solution 25 mg/mL de lysozyme (n° cat. AC611930010, Fisher Scientific) et 5 µL d'une solution 10 mg/mL de Zymolyase-20T (n° cat. NC0516655, Fisher Scientific). Les échantillons ont ensuite été incubés 30 minutes à 37 °C. Par la suite, 250 µL du tampon PDL2 (Geneaid) a été ajouté pour amorcer la lyse cellulaire. Les échantillons ont été mélangés par inversion puis incubés 2 minutes à la température de la pièce. La lyse a été neutralisée par l'ajout de 375 µL du tampon PDL3 (Geneaid). Les échantillons ont été de nouveau mélangés par inversion puis

incubés durant 2 minutes à la température de la pièce. Les échantillons ont été centrifugés dans une microcentrifugeuse à la vitesse maximale pendant 3 min. Les surnageants ont par la suite été transférés sur une colonne LP (Geneaid) et le protocole du Large Plasmid Mini Kit a été suivi pour les étapes de fixation, de lavage et d'éluion des épisomes selon les recommandations du fabricant.

2.2.13 Séquençage des épisomes extraits des clones de *P. tricornutum*

Un volume de 2 μL des épisomes extraits des clones *P. tricornutum* a été transformé par choc thermique dans des cellules compétentes d'*E. coli* NEB 10-beta selon le protocole du fabricant sauf pour l'étape d'étaler sur gélose les bactéries transformées. Pour cette étape, deux géloses LB avec chloramphénicol 15 mg/L ont été utilisées pour la sélection des bactéries transformées. Sur la première gélose, 100 μL des bactéries transformées ont été étalés. Le volume restant a été centrifugé à la température de la pièce pendant 5 minutes à 2 000 x g. Un volume de 700 μL a été retiré du surnageant et le liquide restant a servi pour remettre en suspension les cellules. À la suite de la remise en suspension, tout le volume a été étendu sur la deuxième gélose. Les géloses ont été incubées à 37 °C pour une nuit. Pour chaque transformation, une colonie isolée a été sélectionnée pour inoculer 5 mL de milieu LB liquide supplémenté avec chloramphénicol (15 mg/L). Des extractions d'ADN plasmidiques ont été réalisées à partir des cultures de 5 mL en respectant le protocole du Large Plasmid Mini Kit (n° cat. PDL300, Geneaid).

Les plasmides purifiés ont par la suite été dilués dans de l'eau à une concentration variant de 40 et 65 ng/ μL . Les plasmides ont été envoyés pour un séquençage complet au CCIB DNA Core (Massachusetts General Hospital, United States of America). Le CCIB DNA Core utilise la plateforme de séquençage à haut débit Illumina MiSeq. Finalement, les séquences fournies par le CCIB DNA Core ont été alignées avec la séquence de référence p0521s-V en utilisant le logiciel CLC Main Workbench 7. Les séquences des promoteurs et des terminateurs des cassettes d'expressions ont été alignées avec le programme megablast sur les bases de données « nucleotide collection (nr/nt) » et « Whole Genome Shotgun » (WGS) du National Center for Biotechnology Information

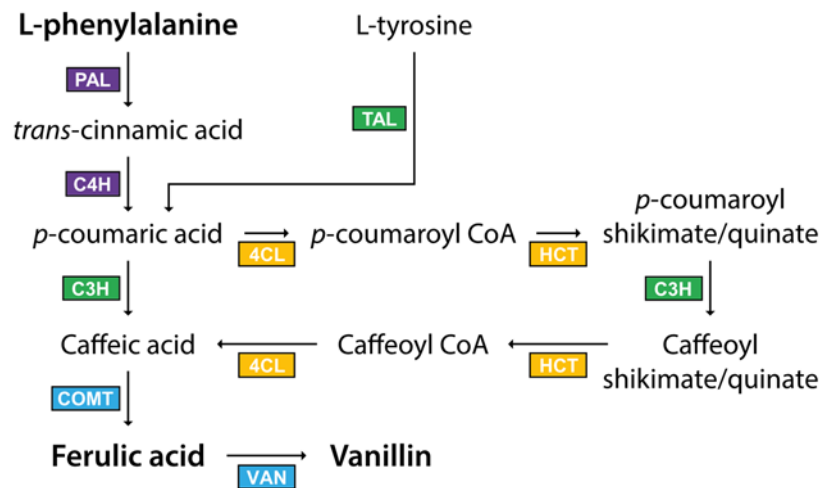
(NCBI). L'organisme *Phaeodactylum tricornutum* CCAP 1055/1 (taxid : 556484) a été sélectionné pour effectuer l'alignement.

2.3 Résultats

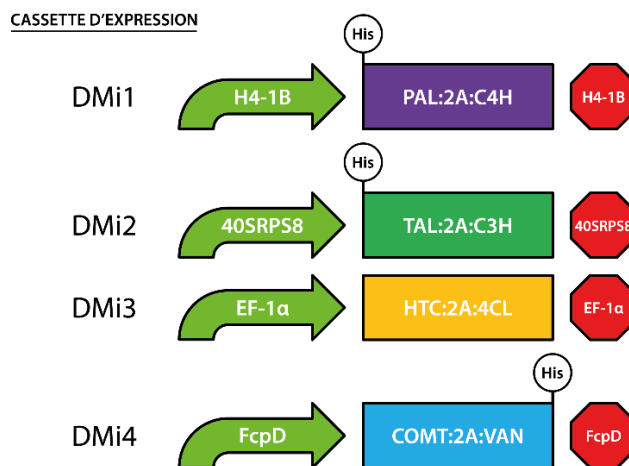
2.3.1 Analyse de la production de métabolites spécialisés dans des clones de *P. tricornutum*

Dans le but d'analyser la production de vanilline et de ses intermédiaires à partir d'un épisode dans la diatomée *P. tricornutum*, quatre clones ont été mis en culture. Le premier clone a été transformé avec le plasmide vide p0521s et il a été utilisé comme contrôle négatif lors des expériences. Ce plasmide vide a servi de base pour produire le plasmide p0521s-V (Slattery et al. 2018) contenant quatre cassettes d'expression codantes pour huit enzymes de la voie de biosynthèse de la vanilline (Figure 2.2A-C). Chaque paire de gènes a été séparée par la séquence du peptide FMDV2A (Figure 2.2B). Les trois autres clones provenaient d'une transformation par conjugaison bactérienne avec le plasmide p0521s-V et ils ont été nommés Vp-7-#1, Vc-1 #1 et V1-3 #2.

(A)



(B)



(C)

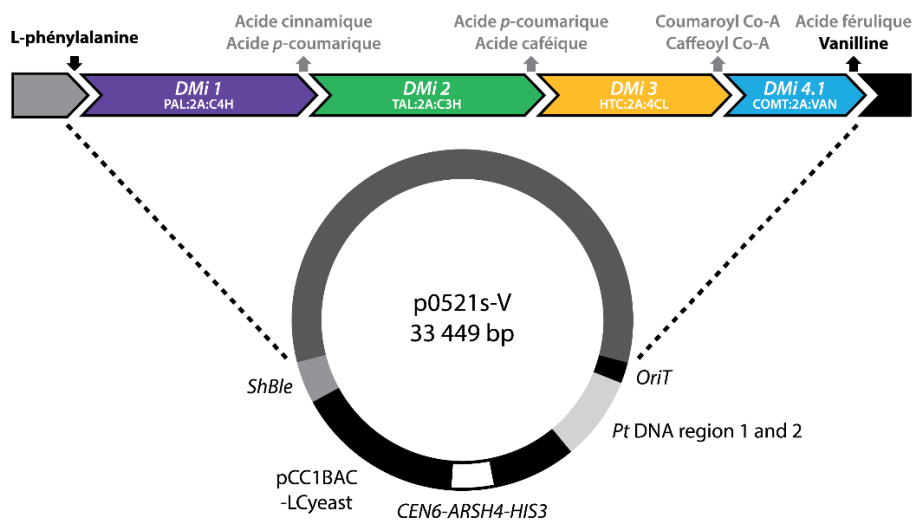


Figure 2.2 Représentation graphique de l'épisode p0521s-V contenant les séquences codantes des enzymes de la voie de biosynthèse de la vanilline.

(A) Schéma de la voie de biosynthèse de la vanilline introduite dans *P. tricornutum*. (B) Illustrations des quatre cassettes d'expression contenues dans le plasmide p0521s-V. (C) Carte du vecteur p0521s-V transformé dans la diatomée *P. tricornutum*. Les métabolites qui peuvent être produits à partir de chaque cassette d'expression sont indiqués en haut de la carte. L'image du plasmide p0521s-V est dessinée de façon à garder la taille relative des séquences. Le même code de couleur a été utilisé pour identifier les cassettes en (A), (B) et en (C). PAL, Phénylalanine ammonia-lyase; C4H, Cinnamate 4-hydroxylase; TAL, Tyrosine ammonia-lyase; C3H, Coumarate 3-hydroxylase; 4CL, 4 coumarate:CoA ligase; HCT, Hydroxycinnamoyltransférase; COMT, Acide caféique/acide 5-hydroxyfêrulique O-méthyltransférase; VAN, Vanilline synthase.

Dans le but d'évaluer si le plasmide p0521s-V permet aux clones *P. tricornutum* de produire de la vanilline, les composés phénoliques ont été extraits de culots cellulaires et analysés par HPLC-MS/MS. Pour ce faire, une méthode a été développée et validée pour la détection de la vanilline et de trois intermédiaires, c'est-à-dire l'acide *p*-coumarique, l'acide caféique et l'acide fêrulique. L'acide isovanillique a été choisi comme SI, car cette molécule a une structure semblable à la vanilline en plus de posséder un groupe carboxyle comme les intermédiaires ciblés. Malgré le fait que l'acide vanillique possède une structure encore plus similaire à la vanilline, cette molécule pas été sélectionné comme SI. Contrairement à l'acide isovanillique, il a été démontré que l'acide vanillique est détectable dans des extraits de *P. tricornutum* (Rico et al. 2013; Foo et al. 2017). Le SI ne doit pas être présent dans la matrice à extraire et doit posséder une structure chimique similaire aux analytes d'intérêts (Imre et al. 2019; Khodadadi et Pourfarzam 2020).

La méthode d'analyse a été validée selon le taux de récupération, l'effet de matrice, la linéarité, la limite de détection, la limite de quantification, la précision et l'exactitude. Le tableau 2.3 montre les transitions MRM qui ont été observées en plus grande proportion lors des analyses ainsi que les ions fragmentés $[M+H]^+$ sélectionnés pour la quantification et la qualification des métabolites.

Tableau 2.3

Transitions MRM observées chez les composés ciblés et le SI
qui ont été sélectionnées pour la quantification par HPLC-MS/MS

Métabolite	Transition MRM [M+H] ⁺ (<i>m/z</i>)	Ions [M+H] ⁺ (<i>m/z</i>)	
		Quantification	Qualification
Vanilline	153 → 125, 93 et 65	93	65
Acide férulique	195 → 177, 145, 117 et 83	177	117
Acide caféique	181 → 163, 145, 135, 117 et 89	163	117
Acide <i>p</i> -coumarique	165 → 147, 119, 91 et 65	147	119
Acide isovanillique (SI)	169 → 125, 93 et 65	93	125

Les taux de récupération présentés au tableau 2.4 démontrent qu'il n'y a pas de différence statistiquement significative entre le taux de récupération dans la solution tampon et celui dans des extraits de culot cellulaire ($p = 0,22$). Les résultats pour l'acide caféique indiquent que la méthode utilisée n'est pas adaptée pour ce métabolite, car de 70 à 74 % en moyenne de ce métabolite ont été perdus durant les processus d'extraction en phase solide.

Tableau 2.4

Moyennes des taux des récupérations obtenues à la suite de l'extraction en phase solide avec les cartouches Strata-X-A

Métabolite	Transition MRM [M+H] ⁺ (<i>m/z</i>)	Taux de récupération (%)	
		Solution tampon ^{*a} (<i>n</i> = 3)	Culot cellulaire ^a (<i>n</i> = 3)
Vanilline	153 → 93	87 (± 1)	92 (± 6)
Acide férulique	195 → 177	87 (± 1)	93 (± 4)
Acide caféique	181 → 163	30 (± 1)	26 (± 7)
Acide <i>p</i> -coumarique	165 → 147	96 (± 1)	95 (± 1)

* Acétate d'ammonium 100 mM à pH 9,8

^a Indique qu'il n'y a pas de différence statistique entre les colonnes ($p > 0,05$)

En comparant l'extraction en phase solide des standards initialement ajoutés à la solution tampon ou à un culot cellulaire, il a été constaté que l'effet de matrice du culot cellulaire est minime. En moyenne, la perte d'ions due à la matrice est de 1 % pour la vanilline, l'acide férulique et l'acide caféique tandis qu'elle est de 2 % pour l'acide *p*-coumarique (Tableau 2.5).

Tableau 2.5

Moyenne des effets de matrice obtenus à la suite de l'extraction en phase solide avec les cartouches Strata-X-A

Métabolite	Transition MRM [M+H] ⁺ (<i>m/z</i>)	Effet de matrice (%) (<i>n</i> = 3)
Vanilline	153 → 93	99 (± 3)
Acide férulique	195 → 177	99 (± 2)
Acide caféique	181 → 163	99 (± 3)
Acide <i>p</i> -coumarique	165 → 147	98 (± 2)

Une courbe de calibration a été réalisée en effectuant une extraction en phase solide des culots cellulaires contenant différentes concentrations des standards (Figure 2.3). À partir de cette courbe, le domaine de linéarité des standards a été déterminé. Le coefficient de détermination linéaire de 0,9394 pour l'acide caféique est le plus faible des quatre métabolites et il est probablement le reflet du faible taux de récupération. L'acide *p*-coumarique et l'acide férulique ont une excellente linéarité pour les cinq concentrations analysées comme le démontre leur coefficient détermination linéaire de 0,9995 et de 0,9991 respectivement. La droite de régression de la vanilline démontre une linéarité plus faible ($r^2 = 0,9667$) que l'acide *p*-coumarique et l'acide férulique, ce qui pourrait avoir un impact sur l'exactitude des valeurs calculées avec celle-ci.

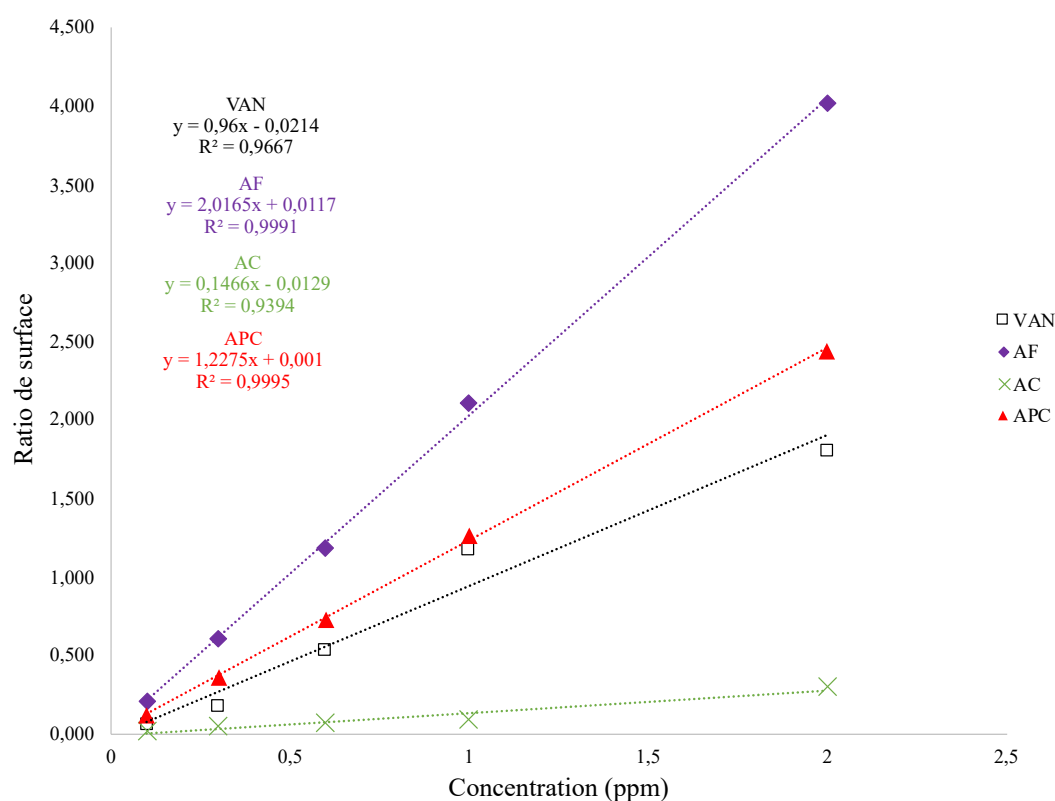


Figure 2.3 Courbe de calibration de la vanilline et de ses intermédiaires.

La vanilline et ses intermédiaires ont été ajoutés à des culots cellulaires pour valider la linéarité des standards de 0,1 à 2 ppm à la suite de l'extraction en phase solide. VAN, vanilline; AF, acide férulique; AC, acide caféique; APC, acide *p*-coumarique.

Le tableau 2.6 indique les LDD et LDQ calculés pour le SI, la vanilline et les intermédiaires ciblés par la méthode d'analyse. Les LDD varient entre 0,000149 ($\pm 0,000004$) et 0,005 ($\pm 0,001$) mg/L, tandis que les LDQ varient entre 0,0005 ($\pm 0,00001$) et 0,0170 ($\pm 0,004$) mg/L. La LDD et la LDQ de l'acide caféique sont les plus élevées avec des valeurs respectives de $0,005 \pm 0,001$ et $0,017 \pm 0,004$. Ce qui concorde avec la droite de régression présentée précédemment puisque la pente de la droite de régression et la sensibilité du HPLC-MS/MS sont proportionnelles.

Tableau 2.6

Moyenne des limites de détection (LDD) et de quantification (LDQ) pour les composés phénoliques analysés et le standard interne

Métabolite	Transition MRM [M+H] ⁺ (m/z)	LDD (mg/L) (n = 3)	LDQ (mg/L) (n = 3)
Vanilline	153 → 93	0,0005 ($\pm 0,0001$)	0,0016 ($\pm 0,0004$)
Acide férulique	195 → 177	0,00040 ($\pm 0,00004$)	0,0013 ($\pm 0,0001$)
Acide caféique	181 → 163	0,005 ($\pm 0,001$)	0,017 ($\pm 0,004$)
Acide <i>p</i> -coumarique	165 → 147	0,00060 ($\pm 0,00002$)	0,00200 ($\pm 0,00008$)
Acide isovanillique	169 → 93	0,000149 ($\pm 0,000004$)	0,00050 ($\pm 0,00001$)

Un contrôle de qualité a été réalisé pour confirmer la précision et l'exactitude de la méthode pour chacun des composés phénoliques sélectionnés (Tableau 2.7). Pour ce faire, les standards ont été mélangés avec un culot cellulaire de *P. tricornutum* puis ils ont été extraits sur phase solide. La concentration prévue de chaque métabolite lors de l'injection était de 0,5 mg/L. La concentration mesurée pour la vanilline est la plus faible avec 0,39 ($\pm 0,01$) mg/L indiquant une bonne précision avec un coefficient de variation de 3 %, mais une faible exactitude avec une erreur relative de 22 %. Pour l'acide caféique, la concentration mesurée de 0,7 ($\pm 0,1$) mg/L témoigne d'une faible précision avec un coefficient de variation de 17 % et une faible exactitude avec une erreur relative de 38. Les résultats de l'acide *p*-coumarique et de l'acide férulique démontrent une méthode

fiable pour ces deux composés avec des valeurs mesurées de 0,500 ($\pm 0,001$) et de 0,498 ($\pm 0,003$) mg/L respectivement.

Tableau 2.7

Moyenne de la précision et de l'exactitude des contrôles de qualité à 0,5 mg/L

Métabolite	Transition MRM [M+H] ⁺ (<i>m/z</i>)	Concentration prévue (mg/L)	Concentration mesurée (mg/L) (<i>n</i> = 3)	Précision (% CV) (<i>n</i> = 3)	Exactitude (% RE) (<i>n</i> = 3)
Vanilline	153 → 93	0,5	0,39 ($\pm 0,01$)	3	22
Acide fêrulique	195 → 177	0,5	0,498 ($\pm 0,003$)	1	< 1
Acide caféique	181 → 163	0,5	0,7 ($\pm 0,1$)	17	38
Acide <i>p</i> -coumarique	165 → 147	0,5	0,500 ($\pm 0,001$)	< 1	< 1

À la suite de la validation de la méthode, des analyses ont été réalisées sur des cultures des clones Vp-7-#1, Vc-1 #1 et V1-3 #2 et d'un clone transformé avec le plasmide vide p0521s. Une courbe de la densité cellulaire en fonction de la DO_{730nm} a été réalisée dans le but de faire le suivi des cultures en utilisant la densité optique (Figure 2.4).

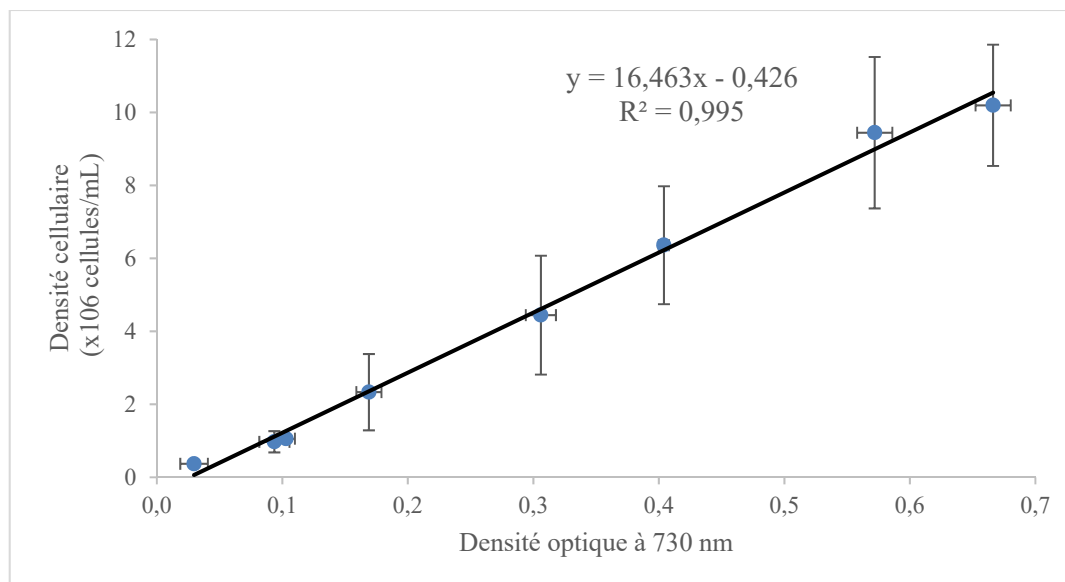


Figure 2.4 Courbe de calibration de la densité cellulaire en fonction de la densité optique à 730 nm pour les cultures de *P. tricornutum*.

La courbe de calibration a été réalisée en triplicata en combinant les données obtenues à partir des cultures d'un clone contenant le vecteur vide p0521s et du clone V1-3 #2 contenant le plasmide p0521s-V.

La DO_{730nm} a été utilisée pour faire le suivi des cultures sur une durée de 164 heures (Figure 2.5) et par la suite ces valeurs ont été converties en densité cellulaire en utilisant l'équation de la droite de régression à la figure 2.4. Les clones contenant l'épisme p0521s-V ont une croissance semblable au clone contenant l'épisme vide p0521s indiquant que l'épisme ne semble pas avoir d'impact sur la croissance cellulaire.

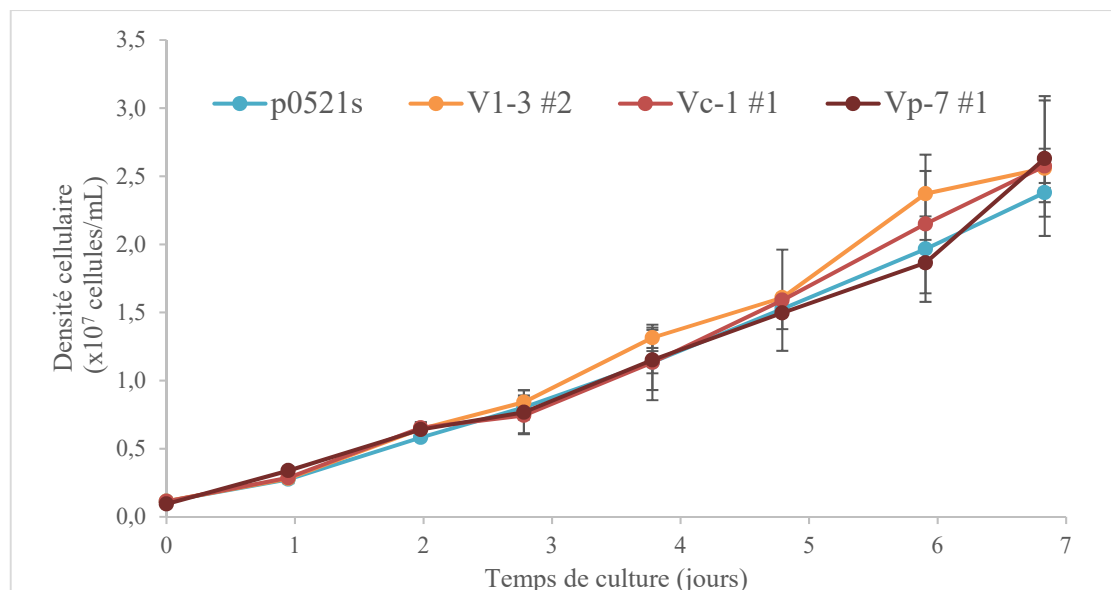


Figure 2.5 Courbe de croissance des clones p0521s (vecteur vide), Vp-7-#1, Vc-1 #1 et V1-3 #2 de *P. tricornutum*.

Les valeurs de densité cellulaire sont les moyennes de trois cultures pour chacun des clones.

Des analyses au HPLC-MS/MS ont été réalisées sur les extraits de culot cellulaire provenant des cultures de 50 mL des clones Vp-7-#1, Vc-1 #1, V1-3 #2 et p0521s. Parmi les quatre métabolites ciblés par la méthode d'analyse au HPLC-MS/MS, seule la vanilline a été détectée dans les quatre clones analysés. Cependant, les concentrations mesurées dans les extraits injectés au HPLC-MS/MS pour les quatre clones étaient très similaires, c'est-à-dire de 0,140 (\pm 0,001) à 0,141 (\pm 0,001) mg/L (Tableau 2.8). Ces échantillons avaient préalablement été dilués 10 fois après leur reconstitution dans 250 μ L avant d'être injectés dans le système HPLC-MS/MS.

Tableau 2.8

Concentrations des composés phénoliques ciblés dans les extraits des souches Vp-7 #1, Vc-1 #1, V1-3 #2 et p0521s qui ont été injectés au HPLC-MS/MS

Métabolite	Transition MRM [M+H] ⁺ (m/z)	Concentration dans les extraits (mg/L)			
		Vp-7 #1 (n = 3)	Vc-1 #1 (n = 3)	V1-3 #2 (n = 3)	p0521s (n = 3)
Vanilline	153 → 93	0,141 (± 0,001)	0,140 (± 0,001)	0,140 (± 0,001)	0,140 7 (± 0,0005)
Acide férulique	195 → 177	ND	ND	ND	ND
Acide caféique	181 → 163	ND	ND	ND	ND
Acide <i>p</i> -coumarique	165 → 147	ND	ND	ND	ND
Densité cellulaire des cultures (cellules/mL)		2,63 x 10 ⁷ (± 4,26 x 10 ⁶)	2,57 x 10 ⁷ (± 5,13 x 10 ⁶)	2,56 x 10 ⁷ (± 1,44 x 10 ⁶)	2,38 x 10 ⁷ (± 7,03 x 10 ⁵)

ND, non détecté

Puisqu'il n'y a pas de variation dans la concentration de vanilline détectée entre les quatre clones, un contrôle a été réalisé en effectuant le protocole d'extraction en phase solide sur un blanc de tampon d'acétate d'ammonium sans culot cellulaire, et ce, en triplicata. Le but de ce contrôle était de vérifier si la vanilline détectée provenait des extraits ou d'une contamination. Comme ce fut le cas pour les clones de *P. tricornutum*, seule la vanilline a été détectée, et ce à une concentration de 0,12 (± 0,01) mg/L. Puisque la contamination provient potentiellement de l'extraction en phase solide, les solvants utilisés pour ce protocole ont été analysés. En triplicata, un volume de 9 mL de la solution de MeOH a été évaporé à sec. Ce volume représente la quantité totale de MeOH utilisée aux étapes de conditionnement et de lavage de la cartouche. La même chose a été réalisée pour le 3 mL de la solution d'acide formique:MeOH (5:95, MeOH de grade LC-MS) qui est utilisé lors de l'étape d'élution du protocole. Avant leur injection dans le système HPLC-MS/MS, ces échantillons secs ont été reconstitués dans 250 µL d'une solution d'acide formique 0,1 % v/v et de MeOH (80:20). Comme contrôle, cette solution qui

représente la phase mobile a aussi été injectée dans le système HPLC-MS/MS. En comparant le temps d'acquisition du standard de vanilline qui est approximativement à 15,8 min (Figure 2.6), il n'y a pas eu de détection de vanilline dans la phase mobile (Figure 2.7). En ce qui concerne les deux solvants utilisés durant l'extraction en phase solide, la vanilline a été détectée pour les trois transitions MRM $[M+H]^+$ qui sont détectées en plus grande abondance relative dans les solutions de standards (Figures 2.8 et 2.9).

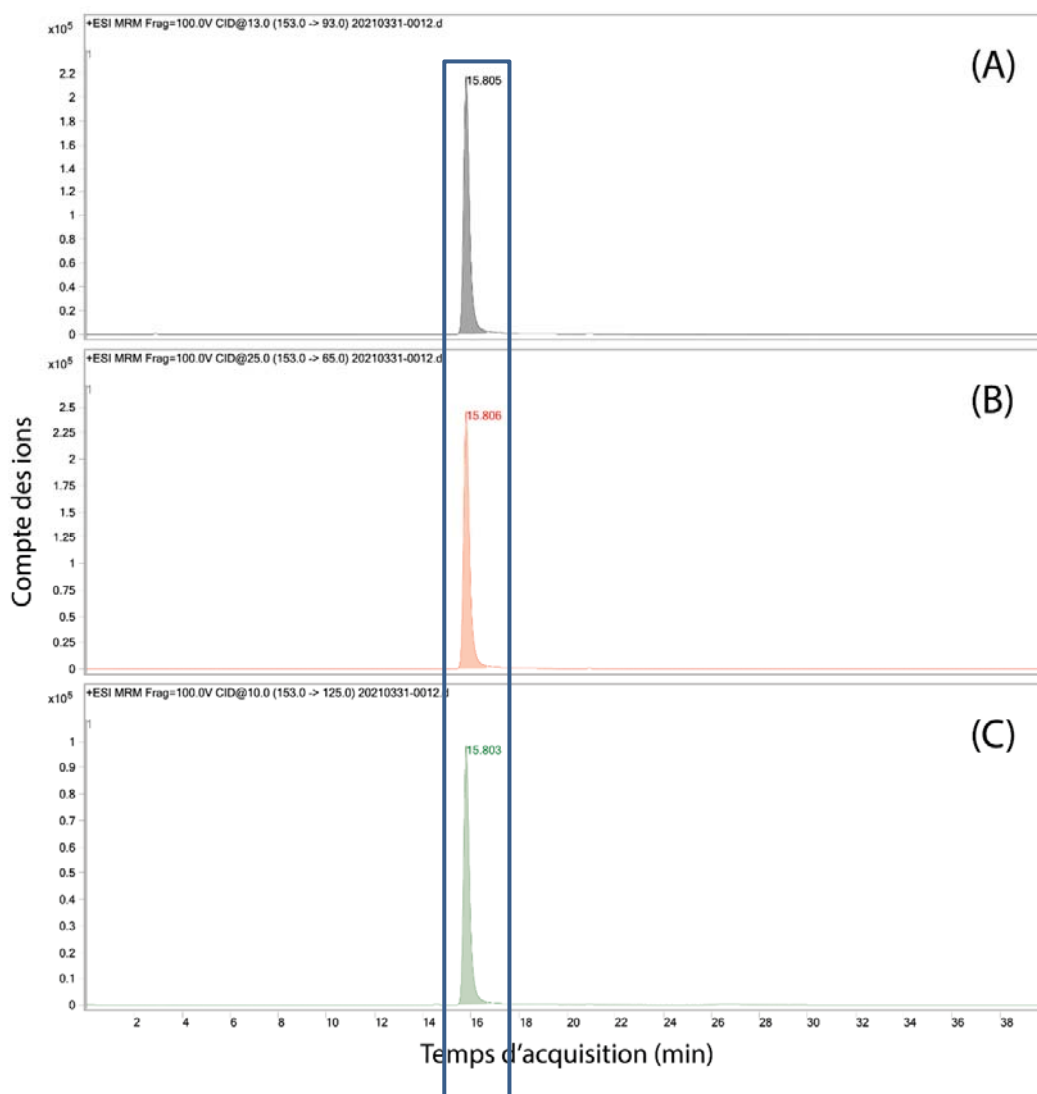


Figure 2.6 Temps d'acquisition du standard de la vanilline à 1 ppm pour les trois ions fragmentés $[M+H]^+$ qui sont en plus grande abondance relative.

(A) Transition MRM $[M+H]^+$ 153 \rightarrow 93. (B) Transition MRM $[M+H]^+$ 153 \rightarrow 65. (C) Transition $[M+H]^+$ MRM 153 \rightarrow 125. L'encadré bleu représente le temps d'acquisition du standard de la vanilline.

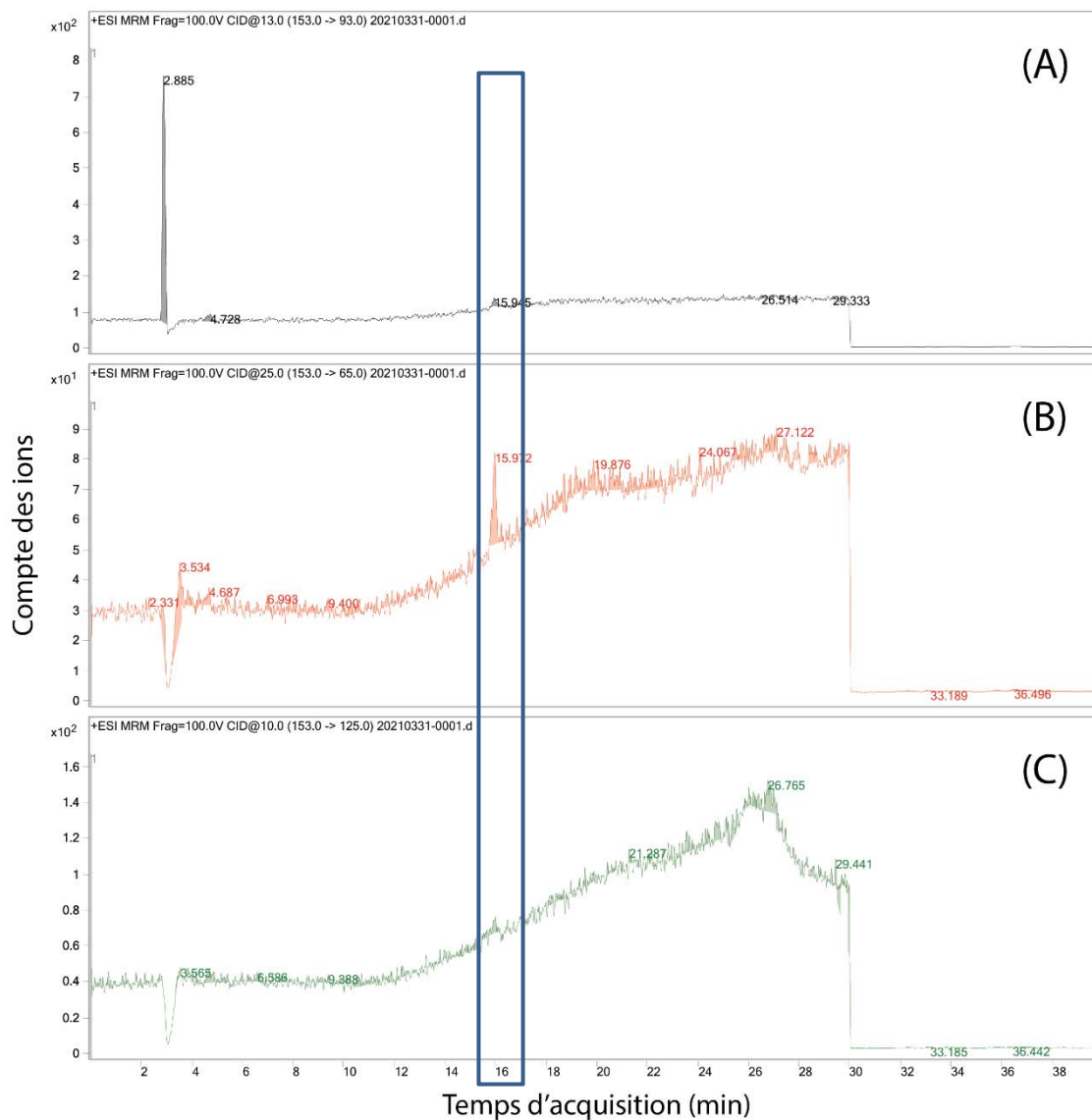


Figure 2.7 Analyse de la présence de la vanilline dans la phase mobile. (A) Transition MRM $[M+H]^+$ 153 \rightarrow 93. (B) Transition MRM $[M+H]^+$ 153 \rightarrow 65. (C) Transition $[M+H]^+$ MRM 153 \rightarrow 125. L'encadré bleu représente le temps d'acquisition du standard de la vanilline.

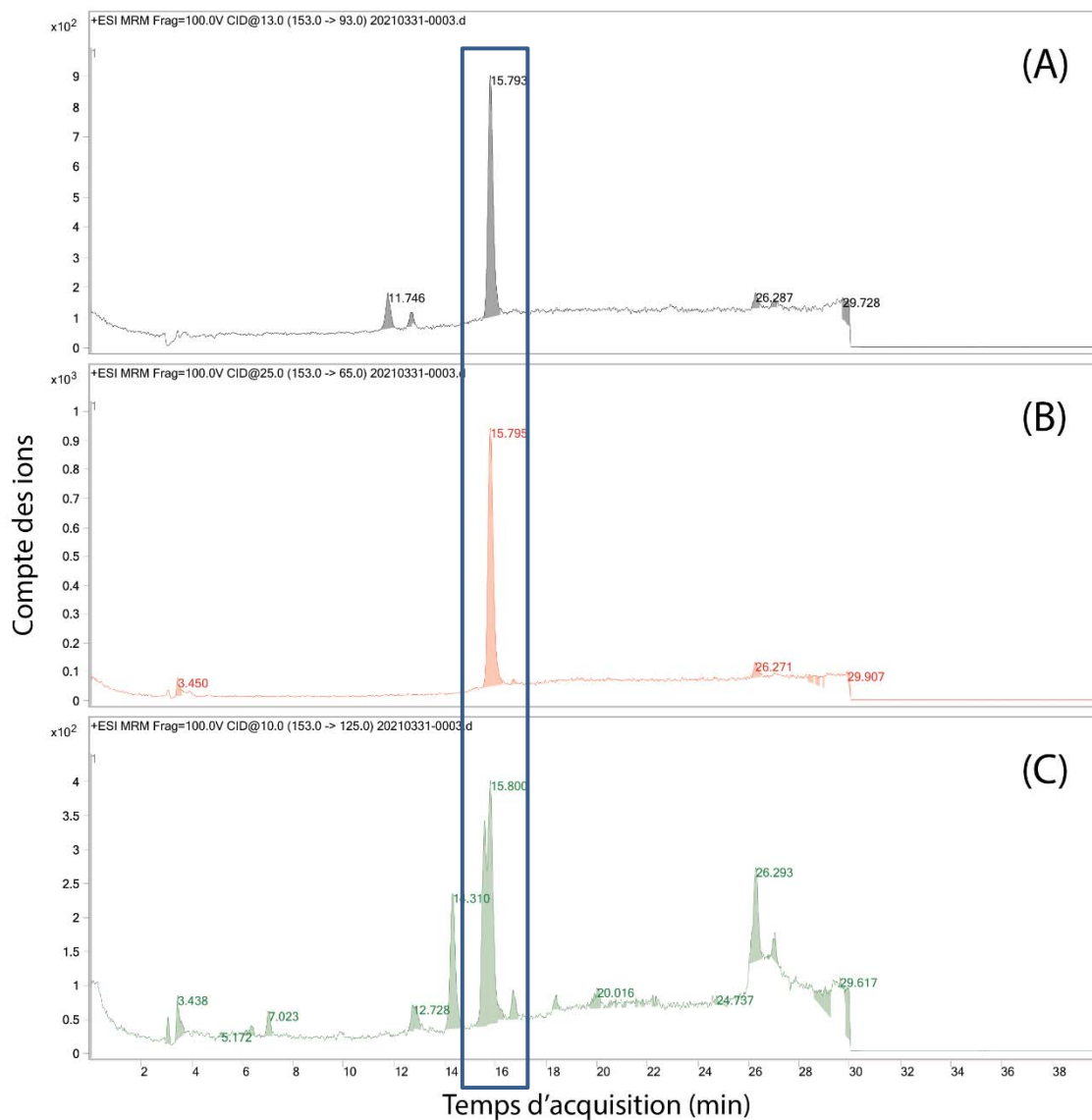


Figure 2.8 Analyse de la présence de la vanilline dans la solution de MeOH utilisée pour conditionner et laver les cartouches d'extraction en phase solide. (A) Transition MRM $[M+H]^+$ 153 \rightarrow 93. (B) Transition MRM $[M+H]^+$ 153 \rightarrow 65. (C) Transition $[M+H]^+$ MRM 153 \rightarrow 125. L'encadré bleu représente le temps d'acquisition du standard de la vanilline.

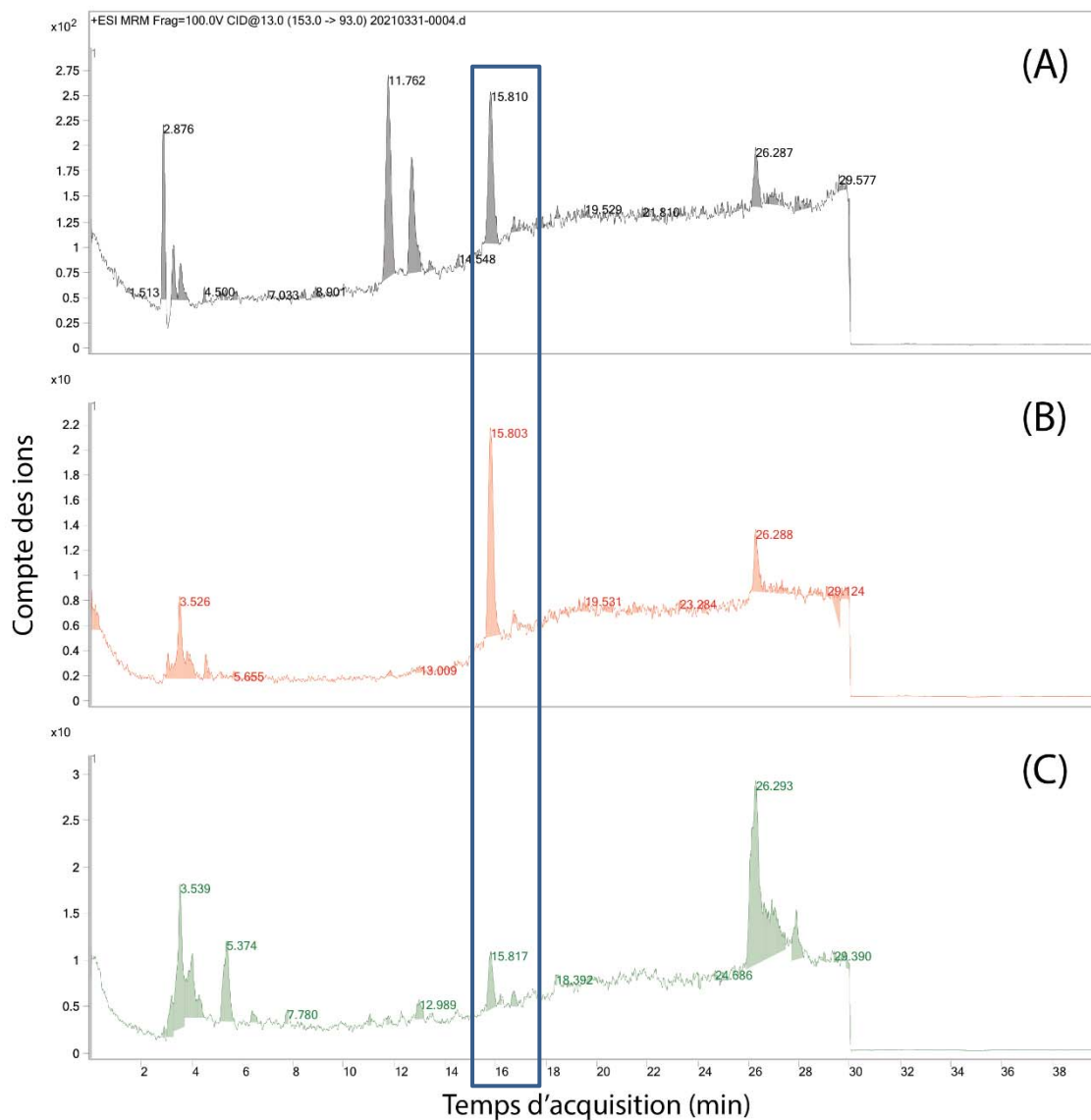


Figure 2.9 Analyse de la présence de la vanilline dans la solution d'acide formique : MeOH (5:95, MeOH de grade LC-MS) qui est utilisée pour l'élution des composés d'intérêts lors de l'extraction en phase solide. (A) Transition MRM $[M+H]^+$ 153 \rightarrow 93. (B) Transition MRM $[M+H]^+$ 153 \rightarrow 65. (C) Transition $[M+H]^+$ MRM 153 \rightarrow 125. L'encadré bleu représente le temps d'acquisition du standard de la vanilline.

Un changement a donc été effectué pour les sources de MeOH utilisé dans le prétraitement des culots cellulaires et dans l'extraction en phase solide. Cependant, il n'a pas été possible d'éliminer la contamination à la vanilline et le problème n'a pas été résolu. En ne sachant pas exactement d'où provenait cette contamination, il a été tenté de réduire

son impact en augmentant la quantité de cellules de *P. tricornutum* utilisée pour les extractions. Ceci aurait comme effet d'augmenter la quantité de composés phénoliques dans les extraits et d'obtenir un signal plus intense que celui provenant de la contamination de vanilline. Dans ce but, des extractions ont été réalisées à partir de cultures de 150 mL dont les densités cellulaires variaient entre $1,22 \times 10^7$ et $2,04 \times 10^7$ cellules/mL. Lors de l'étape d'éluion des extraits sur les cartouches Strata-X-A disponibles, ces dernières devenaient saturées en pigment lorsque la densité cellulaire des cultures était est au-dessus de 2×10^7 cellules/mL. Habituellement, la cartouche Strata-X-A retenait les pigments, mais pour ces échantillons une grande quantité de pigments était visible dans les éluats. Étant donné que l'efficacité d'extraction à l'approche du point de saturation est inconnue, les analyses par HPLC-MS/MS ont été arrêtées, car une nouvelle méthode devait être développée.

2.3.2 Analyse de la production des protéines d'intérêts par la méthode d'immunobuvardage de type western

La production de protéines hétérologues à partir des cassettes d'expression contenues dans le plasmide p0521s-V a été analysée. Pour ce faire, les protéines totales ont été extraites des clones Vp-7 #1, Vc-1 #1, V1-3 #2 et p0521s. La cassette DMi1 (*ZmPAL-2A-AtC4H*) contient une étiquette poly-histidines en N-terminale (Figure 2.2B) permettant de détecter une protéine de 81,4 si le peptide FMDV2A est clivé ou de 139,2 kDa dans le cas contraire. En ce qui concerne la cassette d'expression DMi2 (*ZmTAL-2A-AtC3H*), l'étiquette poly-histidines en N-terminale (Figure 2.2B) permet la détection d'une protéine clivée de 79,4 kDa ou d'une protéine non clivée de 137,6 kDa. La cassette DMi4 (*VpCOMT-2A-VpVAN*) contient une étiquette poly-histidines en position C-terminale (Figure 2.2B) qui permet la détection d'une protéine de 39,9 kDa pour la forme clivée et d'une protéine de 84,2 kDa pour la forme non clivée. La cassette DMi3 ne contenait pas d'étiquette poly-histidines. L'immunobuvardage de type western, selon les conditions expérimentales utilisées, n'a pas révélé de protéines d'intérêts avec les anticorps anti-his (Figure 2.10).

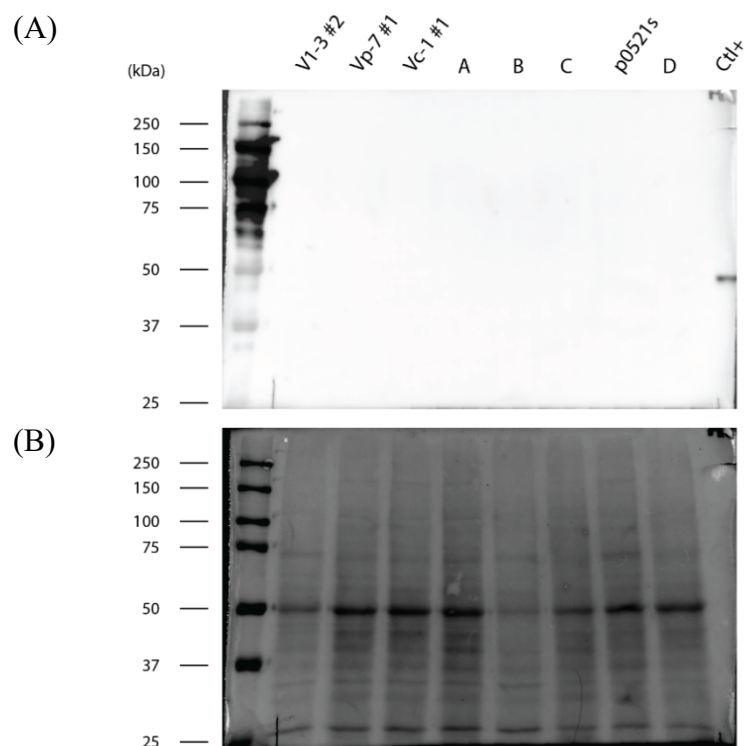


Figure 2.10 Détection de protéines contenant une étiquette poly-histidines par immunobuvardage de type western.

(A) Révélation de la membrane par détection de chimiluminescence.
 (B) Coloration de la membrane au rouge de Ponceau dans le but de révéler les protéines transférées sur gel.

2.3.3 Analyse qualitative de l'expression des cassettes d'expression par RT-PCR

Puisqu'il n'a pas été possible de détecter la production des protéines d'intérêts, nous avons procédé à l'analyse de l'expression des gènes d'intérêts dans le but de déterminer si les cassettes d'expressions étaient fonctionnelles. De manière qualitative, la présence ou l'absence d'expression des cassettes d'intérêts a été déterminé par migration des produits PCR sur gel d'agarose 1 % à la suite d'une RT-PCR (Figure 2.11). Pour chaque cassette, le site d'hybridation de l'amorce sens était localisé en amont de la séquence FMDV2A et le site pour l'amorce anti-sens était situé en aval de cette séquence. Ceci permettait d'assurer une amplification spécifique des ADNc contenant les pairs de gènes. En ce qui concerne le clone contenant le plasmide vide p0521s comme contrôle négatif, aucune amplification à partir des ADNc n'a été détectée. Cependant, une expression positive a été détectée pour les cassettes DMi2, DMi3 et DMi4 chez le

clone V1-3 #2. Pour les clones Vc-1 #1 et Vp-7 #1, seule l'ADNc de la cassette DMi3 a été amplifiée par PCR. La qualité des ARNs totaux extraits a été observée par migration sur gel d'agarose (Figure 2.12) en parallèle à la transcription inverse. Pour chaque échantillon, la bande de l'ARN 28S était plus intense que celle de l'ARN 18S. Une faible dégradation a été observée par la présence d'une faible trainée sous la bande correspondant à l'ARN 18S.

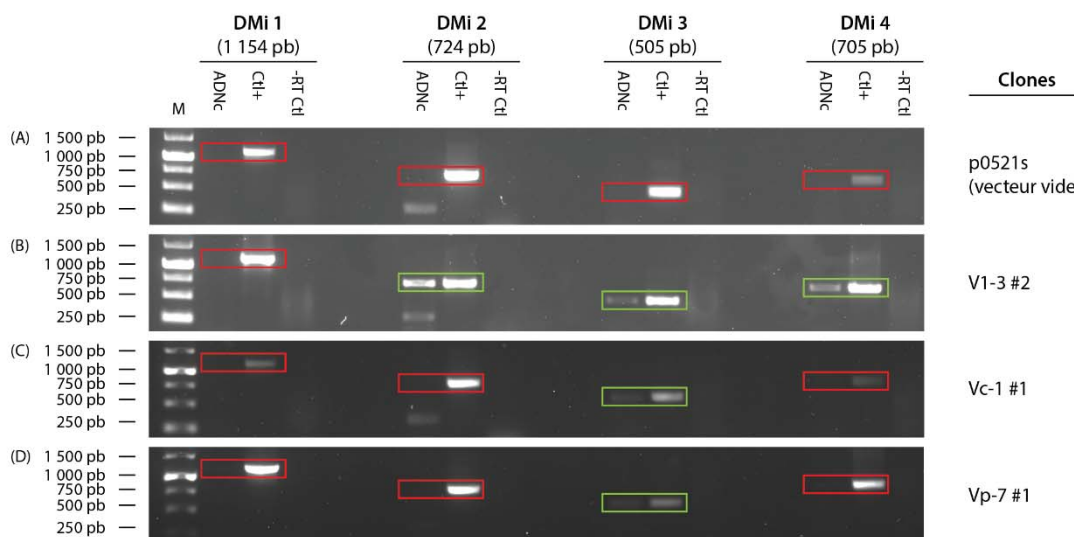


Figure 2.11 Détection de l'expression des cassettes DMi1, DMi2, DMi3 et DMi4 par migration sur gel d'agarose des produits de la RT-PCR provenant des clones p0521s, V1-3 #2, Vc-1 #1 et Vp-7 #1.

Résultat de la RT-PCR des 4 clones recombinants de *P. tricornutum* (A) p0521s, (B) V1-3 #2, (C) Vc-1 #1, (D) Vp-7 #1. Les plasmides contenant les gènes synthétisés par la compagnie Bio Basic ont été utilisés comme matrice d'ADN pour le contrôle positif (Ctl+). Les contrôles négatifs sont les échantillons de la rétrotranscription où la transcriptase inverse a été remplacée par de l'eau (-RT Ctl). ADNc, ADN complémentaire; M, marqueur de poids moléculaire.

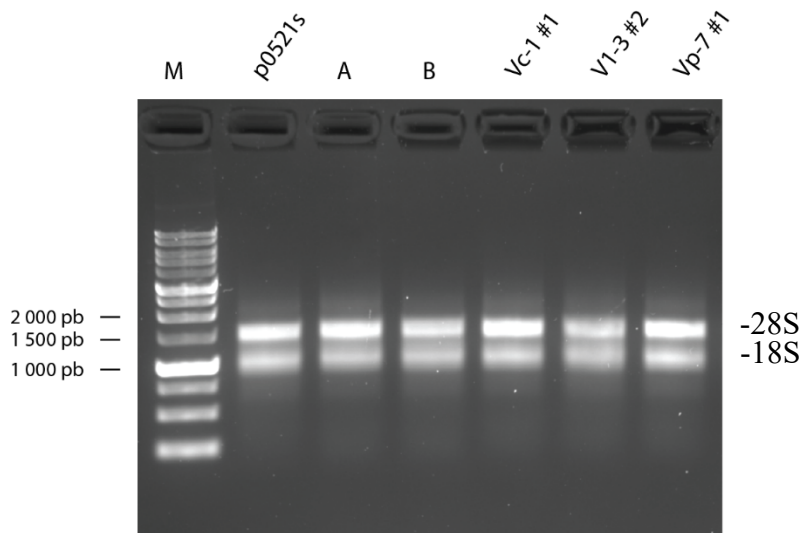


Figure 2.12 Migration sur gel d'agarose 1 % de 500 ng d'ARNs totaux qui ont été utilisés en parallèle pour effectuer la RT-PCR.
A et B représentent des échantillons qui ne sont pas reliés aux expériences présentées dans ce chapitre.

2.3.4 Analyse des séquences des épisomes propagés par les clones V1-3 #2, Vc-1 #1 et Vp-7 #1

Dans le but d'expliquer les résultats d'immunobuvardage de type western et de RT-PCR, les épisomes des clones V1-3 #2, Vc-1 #1 et Vp-7 #1 ont été extraits et séquencés. Un alignement de ces séquences a été fait en utilisant la séquence du plasmide p0521s-V comme référence et plusieurs variations ont été observées entre les séquences des clones et cette séquence (Annexe A). Plusieurs variations étaient communes aux trois clones (Annexe A – Tableau A.1) tandis que quelques modifications étaient spécifiques aux clones V1-3 #2 et Vc-1 #1 (Annexe A – Tableau A.2) puisque qu'elles étaient situées dans une région de l'épisome p0521s-V manquante à l'épisome extrait du clone Vp-7 #1.

Des mutations uniques à chacun des clones ont aussi été détectées (Annexe A – Tableaux A.3-A.5). Toutes ces mutations ont causé des variations entre 1 à 5 pb de longueur, à l'exception du réarrangement dans le plasmide provenant du clone Vp-7 #1 où une séquence de 4094 pb a été substituée par une séquence de 2414 pb (Annexe A –

Tableau A.5). Dans ce cas-ci, la région retirée du plasmide comprenait la séquence partielle du terminateur FcpD (Cassette DMi4), la cassette OriT, la séquence « *Pt DNA region 1* », la séquence « *Pt DNA region 2* » et une partie de la séquence oriV (Annexe A – Tableau A.5). Cette région a été remplacée par trois copies inversées partielles de la cassette DMi1 (Annexe A – Tableau A.5). La première copie inversée est une région de 1878 pb comprenant la séquence partielle du gène *ZmPAL*, la séquence du peptide FMDV2A, le gène *AtC4H* et une partie du terminateur H4-1B. La deuxième copie inversée est une portion de 52 pb du gène *ZmPAL*. La troisième copie inversée est une autre portion du gène *ZmPAL* d'une longueur de 484 pb. Le séquençage de l'épisme extrait du clone Vp-7 #1 démontre qu'à un stade inconnu, l'épisme peut être réarrangé et former ainsi un épisme chimère avec des duplicatas de sa propre séquence.

En analysant les potentielles mutations communes entre les clones (Annexe A – Tableaux A.1 et A.2) et leurs mutations spécifiques (Annexe A – Tableaux A.3-A.5), il n'est pas possible d'identifier clairement ce qui peut empêcher l'expression de certaines cassettes. Cependant, l'expression de la cassette DMi1 est la seule à ne pas avoir été détectée parmi les trois clones et ceux-ci ont en commun cinq substitutions d'une paire de bases au niveau du promoteur de cette cassette. Il est possible que ces mutations empêchent l'expression de la cassette DMi1. Pour ce qui est de l'absence d'expression de la cassette DMi2 dans les clones Vc-1 #1 et Vp-7 #1 (Chapitre II – Figure 2.11), il n'y a pas de mutation spécifique à ces deux clones dans le promoteur 40SRPS8 qui pourraient expliquer potentiellement ce phénomène. En ce qui concerne l'absence d'expression de la cassette DMi4, les plasmides extraits des clones Vc-1 #1 et Vp-7 #1 contiennent chacun une substitution différente d'une paire de bases au niveau du promoteur FcpD (Annexe A – Tableaux A.4 et A.5). L'absence d'expression pourrait donc provenir de ces deux mutations combinées aux mutations dans le promoteur FcpD de la cassette DMi4 qui sont communes aux trois clones (Annexe A – Tableau A.1).

Parmi toutes les mutations potentielles détectées dans les clones, seulement quatre ont un impact sur la séquence des protéines. Une substitution non synonyme (A755T) dans la cassette DMi4 au niveau du gène *VpVAN* (Annexe A – Tableau A.3) et un

changement du cadre de lecture du gène *4CL* (Annexe A – Tableau A.3) ont été séquencés dans le plasmide du clone V1-3 #2. Un changement dans un des cadres de lecture est aussi présent chez le clone Vc-1 #1, mais cette fois il est situé à partir de l'avant-dernier triplet du gène *AtC4H* causant une perte du codon-stop d'origine (Annexe A – Tableau A.4). La quatrième mutation pouvant affecter la séquence des protéines est une substitution non synonyme (D1093N) dans le gène *AtC4H* de la cassette DMi1 (Annexe A – Tableau A.5) qui a été observée dans la séquence du clone Vp-7 #1. Ces quatre mutations ne devraient pas empêcher la production des protéines, cependant elles vont modifier leur séquence et ceci pourrait avoir un impact sur l'activité des enzymes produites ou sur leur stabilité.

2.4 Discussion

À la suite de la collaboration avec l'équipe de professeur Bogumil Karas, nous avons publié un article démontrant que la diatomée *P. tricornutum* peut propager l'épisode p0521s-V d'une taille de plus de 33 kb et contenant la voie de biosynthèse de la vanilline (Slattery et al. 2018). Les expériences réalisées dans ce chapitre avaient pour objectif de déterminer si les cassettes d'expression de cet épisode étaient fonctionnelles et s'il était possible de produire de la vanilline à partir des clones de *P. tricornutum*.

En utilisant la méthode développée pour l'analyse des composés phénoliques au HPLC-MS/MS, les LDD de la vanilline, de l'acide férulique et de l'acide *p*-coumarique ont été calculées à $0,0005 \pm 0,0001$ mg/mL; $0,00040 \pm 0,00004$ mg/mL et $0,00060 (\pm 0,00002)$ mg/L respectivement. Puisqu'il n'y a pas eu de détection d'acide férulique et d'acide *p*-coumarique dans les extraits de culots cellulaires, il est possible que la quantité produite dans les cellules pour ces deux métabolites soit très faible ou que le rendement d'extraction soit insuffisant. Sans spécifier leurs conditions de culture et leur souche de *P. tricornutum*, Foo et al. (2017) ont détecté de faibles quantités de composés phénoliques dans des extractions à partir de biomasse sèche. Par exemple, ils ont extrait 41 ± 1 ng d'acide férulique par gramme de biomasse sèche.

Dans le but d'extraire les composés phénoliques et de les analyser par HPLC-MS/MS, il fallait déterminer quelle quantité de cellules à extraire était nécessaire pour effectuer le protocole. Comme point de départ, la courbe de croissance des cultures (Figure 2.5) a été utilisée pour calculer la quantité de cellules produites en sept jours dans une culture de 50 mL. Pour ce faire, les densités cellulaires au septième jour des clones p0521s, V1-3 #2, Vc-1 #1 et Vp-7 #1 variaient entre $2,38 \times 10^7 (\pm 7,03 \times 10^5)$ et $2,63 \times 10^7 (\pm 4,26 \times 10^6)$ cellules/mL (Tableau 2.8). La valeur arbitraire de $2,25 \times 10^7$ cellules/mL a donc été utilisée pour faire les calculs, car les quatre cultures avaient une densité cellulaire légèrement plus élevée. Une culture de 50 mL à $2,25 \times 10^7$ cellules/mL devrait contenir $1,13 \times 10^9$ cellules au total. Rico et al. (2013) ont détecté $37,0 \pm 0,9$ nmol d'acide férulique par 1×10^{10} cellules de *P. tricornutum*. Puisque notre quantité calculée de cellules est 8,8 fois inférieure à la quantité de cellules utilisée par Rico et al. (2013), il pourrait y avoir 4,2 nmol d'acide férulique dans $1,13 \times 10^9$ cellules. En utilisant la masse molaire de l'acide férulique qui est de 194,18 g/mol, il est possible de convertir la valeur de 4,2 nmol en 0,82 µg d'acide férulique. Cette masse d'acide férulique est alors ce qui pourrait être extrait à partir d'une culture de 50 mL à $2,25 \times 10^7$ cellules/mL si le rendement d'extraction est de 100 % et si nos cultures de *P. tricornutum* ont une production similaire aux cultures de Rico et al. (2013). Il est par la suite possible de calculer la concentration d'acide férulique qui serait injecté dans le HPLC-MS/MS selon la méthode développée à la section 2.2.7. Brièvement, suivant l'extraction en phase solide des métabolites, les extraits sont séchés et reconstitués dans 250 µL de phase mobile puis dilué de nouveau dix fois avec cette même phase mobile avant leur injection dans le HPLC-MS/MS. La concentration d'acide férulique au moment de l'injection serait donc potentiellement de 0,3 mg/L. Il est aussi possible d'injecter l'échantillon reconstitué sans faire de dilution pour augmenter par un facteur 10 la concentration injectée dans le HPLC-MS/MS. Selon la LDD et LDQ calculés lors de la validation de méthode (Tableau 2.6), il aurait été possible de détecter par HPLC-MS/MS l'acide férulique en utilisant des cultures de 50 mL âgés de sept jours. Rico et al. (2013) ont aussi rapporté qu'ils ont détecté $38,2 \pm 0,2$ nmol d'acide *p*-coumarique et 35 ± 1 nmol d'acide caféique dans des extraits de 10×10^{10} cellules.

Il faut tenir compte que Rico et al. (2013) n'ont pas utilisé les mêmes conditions expérimentales que celles utilisées dans nos expériences et que cela a un impact sur le métabolome des cellules de *P. tricornutum*. Plus précisément, ils n'ont pas spécifié la souche de *P. tricornutum* cultivée, ils ont maintenu leur culture à 24 °C sous illumination constante, ils n'ont pas spécifié si les cultures étaient agitées et ils ont utilisé un milieu de culture semi-synthétique produit avec de l'eau de mer stérilisée puis supplémentée en nutriment f/2 (Rico et al. 2013). Dans notre cas, la souche CCAP 1055/1 a été mise en culture à 18 °C sous agitation constante avec une photopériode de 16 h/8 h de lumière/noirceur dans le milieu L1 qui est un milieu de culture synthétique. Ces différences peuvent ainsi avoir un effet sur le métabolome de *P. tricornutum*. Par exemple, les longueurs d'onde de la lumière, la photopériode et la température ont un impact sur la production des lipides (Sirisuk et al. 2018; Sharma et al. 2020; Zhao, Han, et Cao 2020). De plus, les travaux de Rico et al. (2013) ont démontré que la composition du milieu de culture a un impact sur la quantité de composés phénoliques qui est extraite des cultures (Rico et al. 2013). Par exemple, en augmentant la concentration de Fe^{3+} de 900 nmol/L dans leur milieu semi-synthétique, ils ont détecté 60 ± 5 nmol par 10×10^{10} cellules d'acide férulique comparativement à $37,0 \pm 0,9$ nmol avec leur condition de contrôle. Tandis qu'en augmentant la concentration de Cu^{2+} de 790 nmol/L, la quantité d'acide férulique extraite est augmentée à 122 ± 10 nmol par 10×10^{10} cellules. Puisque le milieu de base utilisé par Rico et al. (2013) est de l'eau de mer stérilisée, il n'est pas possible de déterminer la concentration initiale en Fe^{3+} et Cu^{2+} de leur milieu contrôle avant l'ajout du supplément de métaux lourds. En comparaison, le milieu L1 que nous avons utilisé lors de nos expériences contient 11 654 nmol/L de Fe^{3+} et 9,8 nmol/L de Cu^{2+} . L'augmentation de la concentration en Cu^{2+} dans notre milieu de culture L1 pourrait permettre d'augmenter la concentration d'acide férulique extractible de la biomasse de *P. tricornutum*.

L'absence de détection des composés phénoliques ciblés pourrait donc être dû à un faible rendement des extractions, une trop faible biomasse extraite, une production des composés phénoliques plus faible que ce qui est décrit dans la littérature en fonction de nos conditions de culture ou une combinaison de ces facteurs. Puisqu'il n'y a pas de

système de criblage pour la production de vanilline dans notre système d'expression, un volume de 50 mL aurait permis de tester plusieurs clones ou plusieurs conditions de culture en étant moins limité par l'espace physique des cultures. Cependant, il est possible que ce ne soit pas suffisant pour détecter les métabolites d'intérêts.

Dans le but d'augmenter la quantité de biomasses à extraire, le protocole d'extraction devra être adapté afin de ne pas être affecté par la quantité de pigments qui est extraite conjointement avec les composés phénoliques. Le but est de ne pas saturer en pigments les cartouches d'extraction en phase solide. Dans la littérature, les analyses sur les composés phénoliques extraits de *P. tricornutum* ont été réalisées avec un gramme de biomasse lyophilisé (Rico et al. 2013; Foo et al. 2017). L'impact de la lyophilisation devrait être étudié au niveau du rendement de l'extraction des composés phénoliques comparativement à l'utilisation d'une biomasse humide telle qu'utilisée dans notre protocole d'extraction. Par exemple, la lyophilisation a significativement augmenté le rendement d'extraction des phénoliques totaux à partir des feuilles d'*Urtica dioica* (García et al. 2021). Roshanak, Rahimmalek et Goli (2016) ont rapporté la même conclusion pour l'extraction des feuilles de *Camellia sinensis* et de *Camellia assamica*. Cependant, dans le cas de *C. sinensis* et de *C. assamica*, les auteurs ont aussi démontré que la lyophilisation augmentait le rendement de la chlorophylle extraite conjointement avec les composés phénoliques.

Pour diminuer la proportion des pigments contaminants extraits, une étape pourrait être ajoutée au prétraitement de la biomasse pour retirer ces pigments. Par exemple, le noir de carbone graphitisé est capable d'adsorber les pigments tels que la chlorophylle et les caroténoïdes (Rodrigues et al. 2019; Senes et al. 2020; Nicácio et al. 2020). Cependant, il doit être limité sur le plan de la quantité utilisée, car il peut aussi interagir avec les composés phénoliques (Nicácio et al. 2020).

La validation de méthode du HPLC-MS/MS a démontré que l'acide caféique ne peut pas être analysé avec le protocole développé pour l'extraction en phase solide. En effectuant l'extraction sur cartouche Strata-X-A, il y a eu une perte moyenne de 70 %

lorsque l'acide caféique a été ajouté dans le tampon d'acétate d'ammonium 100 mM à pH 9,8 et de 74 % lorsque ce standard a été ajouté au culot cellulaire de *P. tricornutum*. Le pH 9,8 pourrait être trop élevé et causer la dégradation l'acide caféique durant le processus d'extraction. Cilliers et Singleton (1989) ont étudié la vitesse d'oxydation de l'acide caféique dissout dans l'eau à différent pH entre 4 et 8. Ils ont calculé qu'à une température de 20 °C, les constantes de vitesse de dégradation de l'acide caféique étaient de 360 à pH 6, de 1150 à pH 7 et de 2220 à pH 8. Ils n'ont pas testé à un pH basique plus élevé, mais si à pH 9,8 la constante de vitesse reste élevée ou augmente, il est possible que l'acide caféique ait été dégradé durant nos extractions en phase solide. Le pH de 9,8 du tampon a été choisi pour être au moins 2 unités de pH au-dessus du pKa le plus élevé des composés ciblés par la méthode, c'est-à-dire le pKa de 7,4 de la vanilline. Ainsi, près de 100 % des molécules devraient être ionisées à un pH de deux unités supérieures au pKa, ce qui devrait maximiser la rétention des composés sur les cartouches Strata-X-A. Pour analyser l'acide caféique, il faudrait alors changer la phase solide du système d'extraction pour faire l'extraction à un pH moins élevé ou il faudrait développer un deuxième protocole spécifique à l'acide caféique en diminuant le pH du tampon.

Le séquençage des épisomes extraits des clones Vp-7-#1, Vc-1 #1 et V1-3 #2 (Annexe A) a révélé plusieurs mutations. Cependant le clone Vp-7 #1 est le seul dont l'épisome a subi le réarrangement d'une séquence de plusieurs milliers de paires de bases (Annexe A – Tableaux A.3-A.5). Il a déjà été démontré que l'épisome peut être propagé de façon stable pendant au moins quatre mois dans *P. tricornutum* sans être réarrangé (Slattery et al. 2018), mais cela ne prouve pas que les réarrangements sont impossibles. Pour les trois clones, des variations de la séquence de certains promoteurs ont aussi été détectées (Annexe A – Tableaux A1 et A.3-A.5), mais ils n'expliquent pas les variations d'expression entre les clones démontrées par RT-PCR (Chapitre II – Figure 2.11). Pour vérifier l'impact de ces mutations, il faudrait cloner ces séquences altérées des promoteurs en amont d'un gène rapporteur tel que celui codant pour la eGFP. L'absence de production de fluorescence verte démontrerait que les promoteurs mutés ne sont pas fonctionnels.

De plus, il faut noter que les profils d'expression des promoteurs sélectionnés pour ce projet n'ont pas encore été caractérisés dans la littérature. L'utilisation d'un gène rapporteur permettrait de déterminer le niveau d'activité de ces promoteurs selon le stade et les conditions de culture de la diatomée. Ces informations permettraient d'améliorer les connaissances pour les futurs travaux d'ingénierie métabolique de *P. tricornutum*.

Un autre facteur qui peut affecter la production de protéines hétérologues est l'utilisation d'un peptide 2A. Ce peptide devait permettre de produire deux protéines à partir d'un seul ARNm, mais deux autres phénomènes peuvent se produire lors de la traduction des peptides 2A. D'un côté, il est possible que la traduction de l'ARNm s'arrête complètement au niveau du peptide 2A. Dans le cas de la cassette d'expression DMi4.1, l'étiquette poly-histidine est fusionnée en C-terminale à la protéine *VpVAN* qui est située en aval du peptide FMDV2A. Il ne serait donc pas possible de détecter cette protéine si la traduction s'arrête au niveau ce peptide 2A. De l'autre côté, l'ARNm peut être traduit au complet, et ce, sans interruption au niveau du peptide FDMV2. Les deux protéines se retrouveraient ainsi liées par une liaison covalente et il y aurait alors une augmentation de la taille de la protéine d'intérêt.

Le peptide FDMV2A a été utilisé pour ce projet dans le but former des gènes bicistroniques et ainsi réduire la quantité de promoteurs et terminateurs à cloner dans le plasmide p0521s-V. La séquence du peptide FMDV2A prolongé est d'une longueur de 39 acides aminés, comparativement à la majorité des peptides 2A qui ont une longueur moyenne de 18 à 22 acides aminés (Plucinak et al. 2015; Liu et al. 2017; Wang et Marchisio 2021). Selon Plucinak et al. (2015), le peptide FMDV2A prolongé a une efficacité de clivage près de 100 % chez *Chlamydomonas reinhardtii*. Ils ont aussi rapporté une meilleure efficacité de clivage avec le peptide FMDV2A prolongé comparativement au peptide 2A du virus *Thosea asigna*. Puisqu'il n'y avait pas d'informations disponibles dans la littérature au sujet des peptides 2A utilisés chez *P. tricornutum*, la séquence du FMDV2A a été sélectionnée en se basant sur les résultats provenant de la microalgue *C. reinhardtii*. L'étude des peptides 2A dans la diatomée

P. tricornutum permettrait de mieux prévoir leurs efficacités de clivage et leurs impacts sur la production de protéines hétérologues.

2.5 Conclusion

Les expériences réalisées dans ce chapitre sur les clones V1-3 #2, Vc-1 #1 et Vp-7 #1 contenant le plasmide p0521s-V n'ont pas permis de conclure si ceux-ci avaient la capacité de produire de la vanilline ou au moins les enzymes hétérologues impliquées dans la voie de biosynthèse de ce composé aromatique. En revanche, l'expression de certaines cassettes a pu être confirmée. Au cours de ces expériences, seulement trois clones ont été analysés, et ce, dans le but de développer et tester les protocoles d'analyses. Suite à la maîtrise des techniques, plusieurs clones auraient pu être par la suite analysés.

En ce qui concerne l'analyse des métabolites, un protocole devra être développé pour l'extraction des composés phénoliques en neutralisant ou diminuant la coextraction des pigments.

Tels que présentés dans ce chapitre, les résultats associés aux clones V1-3 #2, Vc-1 #1, Vp-7 #1 et p0521s démontrent qu'il y a un lien manquant entre les résultats de séquençages des épisomes et la variation des profils d'expression des cassettes entre les clones. L'étude des épisomes chez *P. tricornutum* via l'expression d'un gène rapporteur permettrait de mieux comprendre et de manipuler ce système d'expression extrachromosomique.

CHAPITRE III

INSTABILITY OF EXTRACHROMOSOMAL DNA TRANSFORMED INTO THE DIATOM *PHAEODACTYLUM TRICORNUTUM* FOR HETEROLOGOUS VANILLIN PRODUCTION

Andrew Diamond^{1*}, Aracely Maribel Diaz-Garza^{1*}, Jessica Li³, Samuel S. Slattery³,
Natacha Merindol¹, Elisa Fantino¹, Fatma Meddeb-Mouelhi^{1,2}, Bogumil J. Karas³,
Simon Barnabé^{1,2} and Isabel Desgagné-Penix^{1,2}

¹ Department of Chemistry, Biochemistry and Physics, Université du Québec à Trois-Rivières, Trois-Rivières, Québec, Canada

² Plant Biology Research Group, Université du Québec à Trois-Rivières, Trois-Rivières, QC, Canada

³ Department of Biochemistry, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada

* Les auteurs ont contribué également.

Le contenu de ce chapitre a fait l'objet d'un article qui a été soumis dans le journal *Plant Biotechnology Journal*.

3.1 Contribution des auteurs

AD, BJK, SB et IDP ont conçu le projet. Les expériences ont été menés par AD, AMDG, NM et EF. Les résultats ont été analysés par AD, AMDG, NM, EF, FMM, BJK, SB et IDP. L'article a été rédigé et corrigé par AD, AMDG, JL, SSS, NM, EF, FMM, BJK, SB et IDP.

3.2 Résumé de l'article (français)

La diatomée *Phaeodactylum tricornutum* possède un potentiel de plateforme d'ingénierie métabolique particulièrement depuis la découverte d'un épisome pouvant agir comme système d'expression extrachromosomique. Il y a cependant un manque de connaissance concernant l'efficacité et les limitations de ce type d'outils moléculaire. Dans ce but, nous avons étudié un épisome contenant une cassette d'expression composée de la protéine rapportrice eGFP lié par le peptide 2A du virus *Thosea asigna* (T2A) à une protéine de fusion des enzymes enoyl-CoA hydratase/aldolase (ech) et feruloyl-CoA synthétase (fcs) provenant de la bactérie *Streptomyces* sp. souche V-1. Moins de 10 % des clones contenant ce gène produisaient de la fluorescence verte détectable par microscopie à fluorescence. Suite à l'extraction des épisomes de dix clones, il a été démontré que les clones fluorescents contenaient un réarrangement au niveau de la protéine ech-fcs, tandis que ceux propageant une cassette d'expression sans mutation ne produisaient pas de fluorescence. En remplaçant la protéine ech-fcs par la protéine fluorescente mCherry, le taux de clones fluorescents par la eGFP est augmenté à plus de 80 %.

3.3 Article complet (anglais)

Introduction

Phaeodactylum tricornutum is a model organism for diatoms and a suitable platform for metabolic engineering. It is not only the best-characterized diatom so far known to accumulate high-value products but also is a viable organism for large-scale culture (Butler et al., 2020). *P. tricornutum* has been successfully used for recombinant protein production attributed to its high growth rates and high efficiency to express complex eukaryotic genes (Hempel and Maier, 2012; Taparia et al., 2019). In addition, the sequencing and annotation of the diatom's genome, done in 2008 (Bowler et al., 2008) and revised in 2021 (Filloramo et al., 2021) and in 2022 (Giguere et al., 2022) combined with the development of a variety of genetic tools, have enabled its use in biotechnology (Daboussi et al., 2014; Slattery et al., 2018). The genetic engineering of *P. tricornutum* using biolistic transformation led to the successful accumulation of docosahexaenoic acid (DHA) (Hamilton et al., 2014) and the production of betulin and its precursors (D'Adamo et al., 2019). *P. tricornutum* has also been engineered to overexpress a putative plastidial pyruvate transporter that led to an increase in biomass, lipid content, and growth (Seo et al., 2020).

The episomal expression method was recently added to the molecular tools for diatoms, allowing the design of consistent, complex and predictable platforms for protein production. This technique avoids the effects of random integration in the genome that could interrupt non-target genes by partial or multiple integrations of the expression cassette of interest (Diner et al., 2016; Karas et al., 2015). When compared to extrachromosomal episomal expression systems, random integration into the genome produced a higher, but variable, fluorescent signal detected from the reporter protein mVenus whereas the fluorescence intensity was more stable between the transconjugants harboring the episome (George et al., 2020). The variability of expression of heterologous genes following nuclear transformation can be due to the random integration impacting genes responsible for growth and development or to silencing due to the integration site. Although episomes have recently been successfully used for protein synthesis and

production of metabolites (Fabris et al., 2020; George et al., 2020), there is still a lack of knowledge on the success of episomal expression and the screening particularly when referring to rearrangement events and the production of heterologous proteins or metabolites.

In a previous study, Slattery et al. (2018) have demonstrated the possibility to introduce eight genes encoding enzymes of the vanillin biosynthesis pathway (Gallage et al., 2014) into *P. tricornutum* into a single episome. Despite the absence of mutations in the genes, enzyme production and activity in the selected transconjugants was not detected. Furthermore, the vanillin synthase (*VpVAN*) enzyme that we previously selected to convert ferulic acid into vanillin was reported to be problematic in a recent study that brought controversy about its catalytic activity (Yang et al., 2017). Thus, we aimed to engineer *P. tricornutum* for the production of vanillin using another strategy. As such, our constructs involved a combination of feruloyl-CoA synthetase (*fcs*) and enoyl-CoA hydratase/aldolase (*ech*) soluble enzymes from *Streptomyces* sp. strain V-1 (Figure 3.1a). These two enzymes have already been characterized for the conversion of ferulic acid to vanillin in *E. coli* (Ni et al., 2015; Yang et al., 2013). Specifically, we built an episomal expression cassette containing the enzymes *fcs* and *ech* with the enhanced green fluorescent protein (eGFP) as a reporter fluorescent protein. In this study, we focused on the production of heterologous proteins by *P. tricornutum* and the rearrangement events of the episomes. We demonstrated that the screening of *P. tricornutum* transconjugants by fluorescence detection using eGFP as a reporter protein can be misleading. On the one hand, no fluorescence could be detected using microscopy and flow cytometry in some zeocin resistant transconjugants, despite confirming a genetically intact whole plasmid by sequencing. On the other hand, in the zeocin resistant eGFP fluorescent transconjugants, the *fcs* gene was altered with deletions, frameshift mutations, or nonsynonymous substitution mutations. Thus, the screening based on the fluorescence of eGFP led us to the selection of transconjugants that did not contain an intact episome and an unaltered expression cassette. Previous studies proved the efficiency of the episome for the genetic engineering of *P. tricornutum* (Fabris et al., 2020; George et al., 2020; Moosburner et al., 2020; Russo et al., 2018; Sharma et al., 2018; Slattery et al., 2020). Nevertheless, our work

shows that toxicity of the heterologous gene product could lead to the episome rearrangements at an unknown step during the conjugation of diatoms.

Results and Discussion

Fluorescence microscopy analysis of transconjugants demonstrated different fluorescent patterns in a single transformation event

In a previous study, we constructed an episome encoding eight heterologous genes involved in vanillin biosynthesis that was stable and propagated in *P. tricornutum* over four months with no evidence of rearrangements (Slattery et al., 2018). However, we could not detect vanillin, *VpVAN* enzyme activity or accumulation in the transconjugants. Since there is a controversy surrounding the function of *VpVAN* in vanillin biosynthesis, we engineered a new construction for the production of vanillin based on a simpler pathway using two genes encoding enzymes already characterized for the conversion of ferulic acid to vanillin (Ni et al., 2015; Yang et al., 2013). Specifically, the *fcs* enzyme can convert the ferulic acid into feruloyl-CoA which will be converted into vanillin by the catalytic activity of *ech* (Figure 3.1a). The new episomal construction was built into the previously characterized plasmid pPtGE33 containing the *Thosea asigna* virus 2A self-cleaving peptide (T2A) linker cloned between eGFP and mCherry sequences and selection marker *Sh ble* as resistance cassette obtained from pPtGE30 plasmid (Slattery et al., 2018). There, the expression of the bi-cistronic gene construction of *eGFP-T2A-mCherry* is under the control of the 40SRPS8 promoter and the FcpA terminator which is known for increased heterologous gene expression (Slattery et al., 2018). Furthermore, the fluorescence from the reporter genes, eGFP and mCherry, will allow for high-throughput screening of transconjugants from liquid culture by flow cytometry or directly from plate using fluorescence microscopy (Annexe B – Figure B.1). Using this system, we constructed two episomes; pControl and pDMi7 for vanillin production (Figure 3.1b,c). In pDMi7, the expression cassette *eGFP-T2A-mCherry* of pControl was modified by replacing the *mCherry* sequence with the *ech* and *fcs* coding sequence linked by a flexible peptide sequence (GGGGS)₃ (Figure 3.1c).

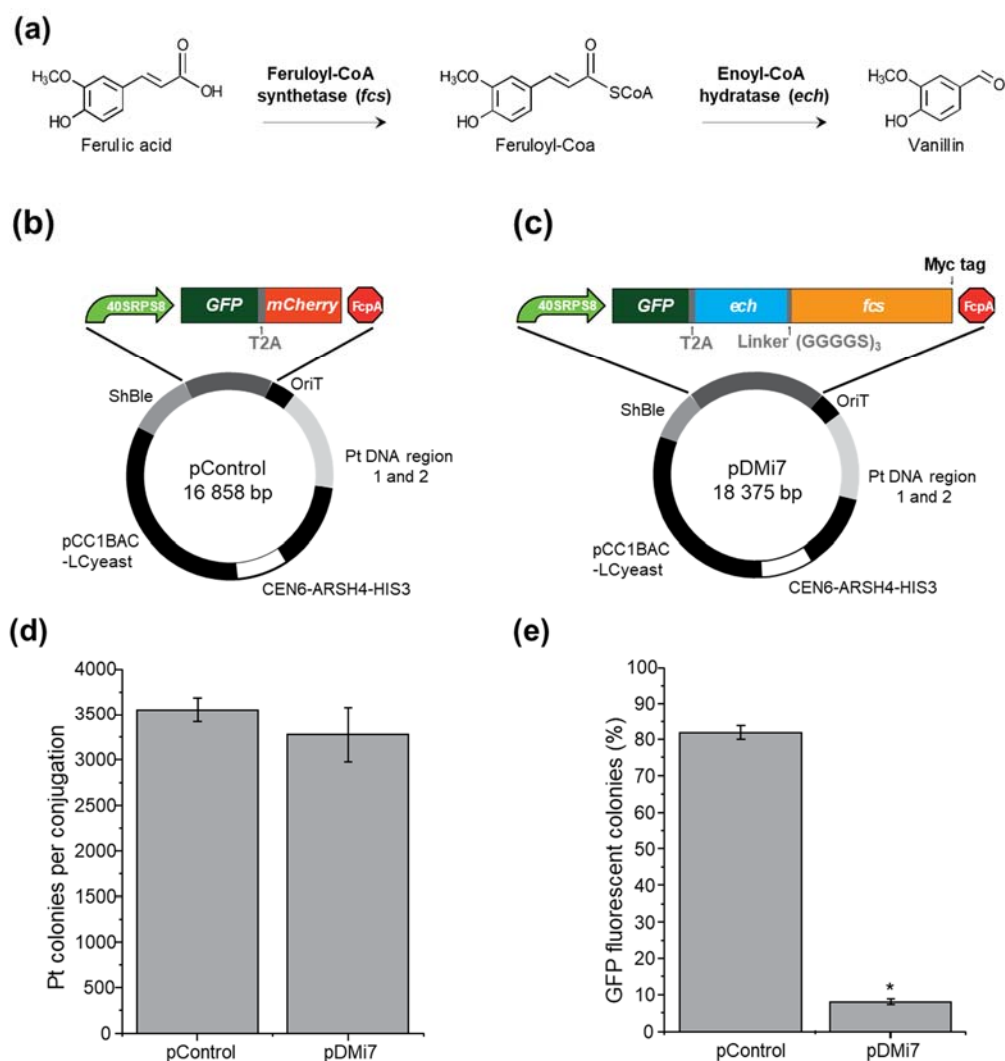


Figure 3.1 Transformation of *P. tricornutum* cells with a vanillin biosynthetic pathway.

(a) Biosynthetic pathway of two bacterial enzymes from *Streptomyces* sp. that convert ferulic acid into vanillin. (b) Plasmid scheme of pControl. The scheme is proportional to the relative size of the DNA sequences, with the exceptions of the size of the promoter and terminator. (c) Plasmid scheme of pDMi7 and the expression cassette containing the vanillin biosynthetic pathway. The scheme is proportional with the relative size of the DNA sequences, with the exceptions of the size of the promoter and terminator. (d) Total counts of colonies per bacterial conjugation done in parallel with an empty vector (pPtGE30) and pDMi7. (e) Percentage of GFP fluorescent colonies obtained from bacterial conjugations with pDMi7 and pControl observed by fluorescence microscopy. Data in (d) and (e) are means of three biological replicates except for the percentage of GFP fluorescent colonies from pDMi7 that was calculated from two biological replicates. The asterisk annotation indicates a significant difference from

pControl as determined by Welch's t-test ($p = 8.1E-06$). 40SRPS8, 40SRPS8 promoter; eGFP, enhanced green fluorescent protein; T2A, *Thosea asigna* virus 2A self-cleaving peptide; ech, enoyl-CoA hydratase/aldolase; fcs, feruloyl-CoA synthetase; FcpA, FcpA terminator.

First, we evaluated the transformation efficiency on zeocin resistant transconjugants 14 days after the transformation of *P. tricornutum* with pPtGE30 (empty vector) and pDMi7 by *E. coli* conjugation. The number of colonies obtained from pPtGE30 and pDMi7 conjugations was not significantly different ($p = 0.25$) (Figure 3.1d and Annexe B – Table B.2), indicating that the transformation was successful.

To evaluate the potential of screening directly on L1 agar plate media, the eGFP fluorescence production was observed using fluorescence microscopy on *P. tricornutum* zeocin resistant transconjugants. For pControl, the fluorescence was observed on $82.1 \pm 1.8\%$ and $76.0 \pm 1.0\%$ of the analysed colonies for the eGFP and mCherry fluorescence respectively (Figure 3.1d and Annexe B – Table B.1). The absence of eGFP (17.9%) and mCherry (24%) fluorescence in some of the pControl colonies analysed could be the result of *i*) an absence of fluorescent proteins production, *ii*) degradation of these proteins due to malfunction of the T2A self-cleaving peptide, or *iii*) mutations in the sequence of the expression cassette.

Unexpectedly, $8.2 \pm 0.8\%$ of pDMi7 colonies displayed eGFP fluorescence (Figure 3.1d) which is significantly different ($p < 0.00001$) as determined by Welch's t-test from the pControl results. This suggests a potential toxicity of pDMi7 genetic construction *ech* and *fcs* either from the coding sequences directly, the protein product or the activity of the transgene. A literature search did not yield any information on the toxicity of ech or fcs. Since the number of colonies obtained from pPtGE30 and pDMi7 conjugations was not significantly different (Figure 3.1e and Annexe B – Table B.2), it indicates that if pDMi7 was lethal for *P. tricornutum*, the diatom was able to counter the toxic properties from the plasmid. It is more likely that the *ech-fcs* protein fusion is toxic maybe due to its catalytic activity towards *P. tricornutum* metabolites.

It is interesting to note that the fluorescence of pDMi7 was not uniform in the transconjugant colonies. In some cases, we observed fluorescence as dots inside the colonies while in others, the fluorescence signal was homogeneous for only one part of the colony (Annexe B – Figures B.2 and B.3). Sometimes, the size or the shape of transformed *P. tricornutum* colonies could be interpreted as two colonies overlapping when only one of them was fluorescent. In other cases, it was more ambiguous to determine if two colonies had merged or if the heterogeneous fluorescence pattern was coming from a single colony. These observations were present for pControl transconjugants (Annexe B – Figures B.2 and B.3).

The sequences of the episomes recovered from *P. tricornutum* pDMi7 transconjugants showed rearrangements and mutations

We selected ten positive pDMi7 transconjugants displaying antibiotic resistance to investigate eGFP fluorescence from the reporter gene. Based on fluorescent microscopy and flow cytometry analysis, six out of the ten transconjugants were eGFP positives (GFP+) whereas four were not (GFP-). To further investigate this, we assessed the integrity of the pDMi7 with an episomes rescue experiment. For that, the recovered episomes from *P. tricornutum* cultures were transformed and replicated in *E. coli* from our six GFP+ transconjugants (namely DMi7-21-1, DMi7-21-3, DMi7-31-1, DMi7-31-3, DMi7-31-4, and DMi7-31-8) and four non-fluorescent (GFP-) ones (DMi7-21-14, DMi7-21-15, DMi7-21-16, and DMi7-31-10), extracted and digested (Figure 3.2a). The double digestion of the recovered episomes showed variable profiles compared to the control plasmid which is the initial pDMi7 following Gibson assembly. Only DMi7-21-14 exhibited an expected restriction enzyme digestion profile corresponding to the control plasmid (Figure 3.2a). This suggests that DNA rearrangements occurred in nine out of the ten episomes recovered from pDMi7 transconjugants whether they produced eGFP or not (Figure 3.2a).

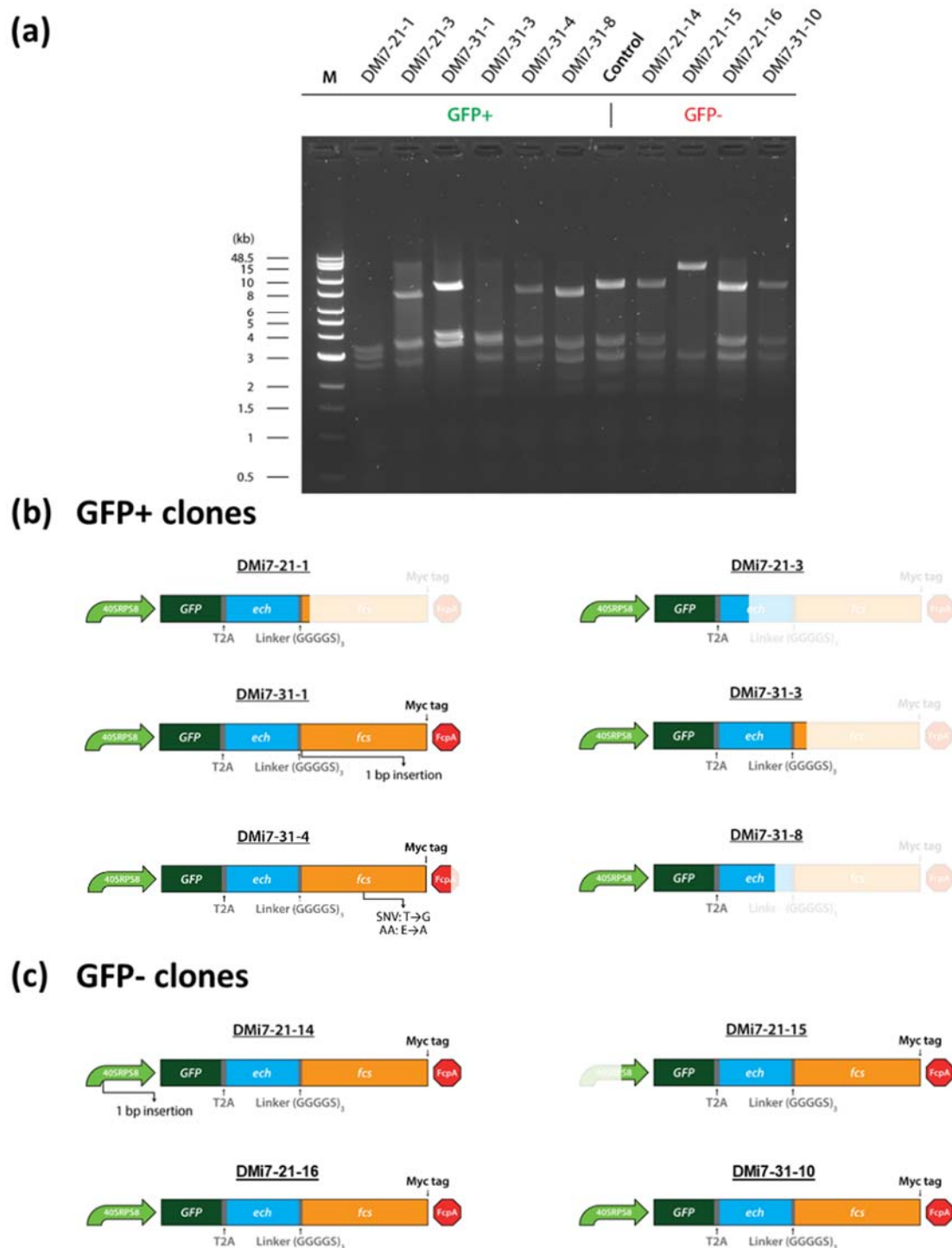


Figure 3.2 DNA digestion profiles and representation of the rearrangements of the recovered episomes from pDMi7 *P. tricornutum* transconjugants. (a) Double digestion by PstI and SpeI of the episomes recovered from *P. tricornutum* transconjugants showing rearrangement in their sequence. The control is the digestion of the plasmid pDMi7 used for the conjugation. (b and c) Schematic representation of the rearrangements in the expression

cassette from the episomes recovered from *P. tricornutum* transconjugants containing pDMi7. The GFP⁺ transconjugants are the one where GFP fluorescent cells have been detected by fluorescent microscopy and flow cytometry. Shaded parts represent deletions in the expression cassettes and is on scale. As a reference the schemes of DMi7-21-16 and DMi7-31-10 are identical to the designed expression cassette from pDMi7. SNV, single nucleotide variation (T to G); AA, amino acid (E to A). 40SRPS8, 40SRPS8 promoter; eGFP, enhanced green fluorescent protein; T2A, *Thosea asigna* virus 2A self-cleaving peptide; ech, enoyl-CoA hydratase/aldolase; fcs, feruloyl-CoA synthetase; FcpA, FcpA terminator.

To further analyze the profile of rearrangements, the whole episomes were sequenced using next-generation sequencing with the Illumina MiSeq platform. Interestingly, the episome sequence from the six GFP⁺ recovered showed the *fcs* coding sequence altered by mutations, or from complete or partial deletions (Figure 3.2b). The analyzed changes in the complete DNA sequences of the episomes are listed in table B.3 (Annexe B). The sequence of the expression cassette was truncated with total or partial deletion of the *fcs* gene and complete deletion of the FcpA terminator for DMi7-21-1, DMi7-21-3, DMi7-31-3, and DMi7-31-8, (Figure 3.2b). The total length of the deletions starting in the expression cassette for DMi7-21-1 and DMi7-31-3 were of 5,523 bp and 4,873 bp respectively. Regarding DMi7-31-8, the deletion of 2,552 bp was replaced by a partial and inverted duplicate from the *ech* sequence of 155 bp. In the case of DMi7-21-3 there was also a partial deletion of 5,004 bp starting in the *ech* gene. Moreover, in this transconjugant, the expression cassette was also present in two extra partial duplicates. The first one carries the partial sequence of *ech-fcs*. The second one consists of a partial *eGFP* sequence with the complete gene encoding the fusion protein (*ech-fcs*), however, lacking the promoter sequence of 40SRPS8. Regarding the clone DMi7-31-1, there was a change in the open reading frame of the *fcs* gene with a single insertion after the first twelve nucleotides as represent in figure 3.2. A nonsynonymous substitution was detected in the sequence of the *fcs* gene of the clone DMi7-31-4 causing a substitution of negatively charged amino acid glutamic acid (E) with a non-polar aliphatic alanine (A). Further work needs to be done to evaluate the impact of this mutation on the enzymatic activity of *fcs*. Altogether, the six GFP⁺ transconjugants demonstrated instability in their episomal sequences that was linked to the *fcs* sequence. This type of extrachromosomal DNA instability which influences the production of a heterologous

protein has never been reported from diatoms before. It is possible that the lack of data in the literature concerning this matter could be due to the fact that the episome in *P. tricornutum* has been discovered fairly recently (Karas et al., 2015). However, it is conceivable that this instability should not be specific to the *fcs* coding sequence used in this study. If so, it would then be possible that the episome instability could be problematic in some metabolic engineering experiments as it was reported for development of Cyanobacteria strains (Jones, 2014).

Intriguingly, no mutations were observed in the *eGFP-T2A-ech-fcs* coding sequence of the expression cassette in any of the non-fluorescent GFP- transconjugants (Figure 3.2c). Moreover, DMi7-21-16 and DMi7-31-10 had no modification in the 40SRPS8 promoter and in the FcpA terminator sequences. Regarding the two other eGFP negative transconjugants, the expression cassette from the clone DMi7-21-14 contained a single nucleotide insertion in the promoter sequence, and DMi7-21-15 was harbouring a deletion of the first half of the promoter sequence. Thus, the absence of eGFP fluorescence of those four transconjugants could not be explained by rearrangements and/or mutations in the expression cassette. Since the genes were codon optimized, the lack of fluorescence and the plasmid rearrangements of GFP+ transconjugants could not be explained by rare codon usage. Besides, the absence of fluorescence in the recombinant *P. tricornutum* transconjugants is not related to the DNA sequence, which suggests that regulation at other levels such as RNA silencing, protein degradation or low protein production below the sensitivity levels, could be responsible for this phenomenon.

Interestingly, a similar observation was reported recently. Defrel et al. transformed *P. tricornutum* by biolistic with the *uidA* gene, which encodes the β -glucuronidase (GUS) enzyme, linked by the 2A peptide to the *nourseothricin N-acetyl transferase* gene (NAT) (Defrel et al., 2021). No GUS activity was detected in two transconjugants with a colorimetric assay despite the absence of sequence modification of their transgene. With the development of a fluorometric assay, GUS activity was detected in both transconjugants and was 30 and 40 times lower compared to the clone with the strongest activity. The reason of the low activity from these two transconjugants was not elucidated.

In their case, the integration of the transgenes in poorly expressed genomic regions might be the cause of their low activity. This hypothesis cannot be applied to our work since the episomes are extrachromosomal expression systems. However, it is possible that the heterologous expression systems, either by random DNA integration or by episome, would be affected by defense mechanisms of *P. tricornutum* like RNA silencing. Regarding RNA silencing, De Riso et al. showed that the GUS reporter gene expression can be successfully silenced using constructs with sense and antisense guide RNAs in *P. tricornutum* (De Riso et al., 2009). Besides, they were able to modulate the expression of two endogenous genes encoding for phytochrome (Dph1) and cryptochrome/photolyase family 1 (CPF1), proving that RNA silencing can occur in *P. tricornutum* (De Riso et al., 2009). Further investigations could allow us to determine whether this mechanism could be the cause of the absence of fluorescent signals in transconjugants with no mutations.

Fluorescent and non-fluorescent transconjugants tested positive for gene expression of the ech/fcs fusion protein

To investigate the heterologous protein production from the pDMi7 transconjugants, we first tested the gene expression of the cassette by an endpoint RT-PCR, using primers that amplified a 209 bp long fragment that covered the end of *ech* and the beginning of the *fcs* genes. For this purpose, six transconjugants were analyzed including GFP+ and GFP- ones. To remove traces of episome that co-purified with the total RNA and could give false positives from the PCR analysis, the samples were treated with DNase I. Total RNA samples without reverse transcription were used as negative control for the PCR. All the transconjugants tested were positive indicating that they express the gene encoding the ech-fcs fusion protein (Figure 3.3). Based on the construction design, the eGFP should be produced from the same mRNA as the ech-fcs fusion protein, and then separated by the self-cleavage T2A linker.



Figure 3.3 Expression of the cassette of interest from pDMi7 *P. tricornutum* transconjugants.

Endpoint RT-PCR of *P. tricornutum* transconjugants that have been observed by fluorescence microscopy as fluorescent GFP+ (DMi7-21-1, -3, and -7) or non-fluorescent GFP- (DMi7-21-14, -15, -16). The plasmid pDMi7 has been used as a positive control for PCR (+Ctl) and the respective RNA samples (without reverse transcription) were used as negative controls (-Ctl RT).

The clone DMi7-21-3 gave a positive result even though the expression cassette is incomplete. In this case the 40SRPS8 promoter and the *eGFP* sequences are complete but the cassette does not have the reverse primer binding site located in the *fcs* gene (Figure 3.2b and Annexe B – Table B.3). In the episome recovered from this transconjugants, there were three partial copies of the cassette where only one of them contains the sequence that would give correct RT-PCR amplification. The first partial copy is presented in figure 3.2b, which is composed of 40SRPS8 promoter, *eGFP*, and each partial sequence lacking the binding site for the reverse primer (DMi7_R). The second partial copy of 2,400 bp is inserted after the reference position 13,888 bp. However, it is composed of *ech* partial sequence (3'-end) lacking the promoter, *eGFP*, and binding site for the forward primer (DMi7_F). Regarding the other copy, it is an insertion of 3,561 bp before the reference position 14,879. In this case, the 40SRPS8 promoter sequence and the first nucleotide of *eGFP* are lacking, but the *ech-fcs* sequence is complete and contains the RT-PCR primers binding sites. This could explain the cDNA amplification from DMi7-21-3. We hypothesized that the sequence upstream the duplicate, product of the episome rearrangements, could drive the expression of the *eGFP:c.1delA-T2A-ech-fcs*

cassette. For this, we analyzed the upstream sequence of the forward promoter binding site and found transcription factor binding sites predicted by PLANTCARE software (Lescot et al., 2002) and the presence of the transcription initiation (Inr) like consensus motif “TCAH₊₁W” (Annexe B – Figure B.4) which has been characterized before in this diatom (Kadono et al., 2015).

Detection of heterologous proteins in *P. tricornutum* transconjugants depends on the percentage of producing cells in the total population and on the DNA sequence of the expression cassette

To confirm the presence of eGFP fluorescence, transconjugants that were assessed by endpoint RT-PCR were also analysed by flow cytometry. We confirmed that fluorescent *P. tricornutum* transconjugants observed under fluorescence microscopy contained both eGFP⁺ and eGFP⁻ cells, with a majority of the cells that were non-fluorescent (Figure 3.4a-b, initial cultures). As such, the percentages of fluorescent cells obtained by flow cytometry were 33.8%, 5.65%, and 10.2% for DMI7-21-1, DMI7-21-3, and DMI7-21-7 respectively (Figure 3.4b; left panels). Regarding the non-fluorescent transconjugants (DMI7-21-14; DMI7-21-15 and DMI7-21-16), no eGFP fluorescent cells were detected by flow cytometry (Annexe B – Figure B.5; right panel).

two sequences of enrichment (initial cultures and results from first sorting). The total GFP⁺ population gate is shown as reference. Numbers indicate the mean fluorescence intensity of GFP in the total chlorophyll⁺ population. (b). Pseudocolor dot plots of control pPtGE30 and transconjugant DMi7-21-1, DMi7-21-3, and DMi7-21-7 cells at 530 nm (GFP) in the x axis, and autofluorescence at 448 nm in the y axis from initial culture used for initial sorting, cultures grown for 7 days after first sorting and used for second sorting (sGFP⁺) and cultures grown for 5 days after the first subculturing from the second sorting (GFP⁺⁺). Gates and frequencies of total GFP⁺, GFP⁺sorted (GFP sort and GFP⁺⁺ sort) and GFP-sorted populations were designed according to the autofluorescence of the negative control pPtGE30 at each day of experiment and are shown as reference. GFP, green fluorescent protein.

Furthermore, we used cell sorting with flow cytometry to enrich the proportion of fluorescent cells in DMi7-21-1, DMi7-21-3, and DMi7-21-7 cultures (Figure 3.4). A first sorting was performed on the initial transconjugants DMi7-21-1, DMi7-21-3, and DMi7-21-7 (Figure 3.4b left panels). Both fluorescent (named sGFP⁺ cultures) and non-fluorescent cells (named sGFP⁻ cultures) were individually sorted. The sGFP⁺ and the sGFP⁻ cell cultures were grown for 7 days and then sorted for a second time (Figure 3.4b and Annexe B – Figure B.6; center panels). For this last round of sorting, only the fluorescent cells were sorted from the sGFP⁺ cultures and the resulting cultures from this enrichment were named GFP⁺⁺ (e.g., DMi7-21-1^{GFP++}) (Figure 3.4b; right panels). From the sGFP⁻ enrichment, only the non-fluorescent cells were sorted and the resulting cell cultures were named GFP⁻⁻ (Annexe B – Figure B.6; right panel). With the enrichment in eGFP⁺ cells, the percentage of fluorescent *P. tricornutum* cells increased from 2 to 9 fold as shown in figure 3.4 from 33.8% to 53.2% for DMi7-21-1^{GFP++}, from 5.65% to 48.8% for DMi7-21-3^{GFP++}, and from 10.2% to 96.4% for DMi7-21-7^{GFP++}.

Among the transconjugants tested, only DMi7-21-7^{GFP++} reached almost 100% of fluorescent cells. Even after three subculturing of DMi7-21-7^{GFP++}, the enriched proportion did not decrease and showed over 99% of fluorescent cells (Figure 3.5). The stability of the fluorescent cells population in the enriched DMi7-21-7^{GFP++} culture and the failure to recover the episomes from the initial culture cells of DMi7-21-7 could suggest that the DNA episome was integrated into the diatom genome.

Intriguingly, the GFP- sorted cells such as DMI7-21-1^{GFP--} and DMI7-21-3^{GFP--} come from clones with an intact eGFP sequence (Figure 3.2b). This suggests that protein production such as GFP can be stably turned off in *P. tricornutum* episome (Annexe B – Figure B.6). Although, DMI7-21-1^{GFP++} and DMI7-21-3^{GFP++} display similar eGFP frequency and mean fluorescence intensity (MFI), a distinct subpopulation of eGFP+ cells was only observed in DMI7-21-1^{GFP++}, compared to a continuous range of eGFP intensity (from negative to positive cells) was observed in DMI7-21-3^{GFP++} (Figure 3.4a; right panel). Based on this observation, it is possible that some colonies of pDMI7 are originating from a mix or heterogenous transconjugants as shown in (Annexe B – Figures B.2 and B.3) with a subpopulation of cells producing eGFP and another not producing eGFP. In all, the results demonstrated that using cell sorting by flow cytometry to enrich a cell population can increase the production of a heterologous protein from a culture of *P. tricornutum* transconjugants without optimization of the media or the culture conditions. This can also be achieved by the droplet-based microfluidic techniques that can be used to accelerate the screening of microalgal transconjugants and to increase the cell population producing the eGFP (Yu et al., 2021).

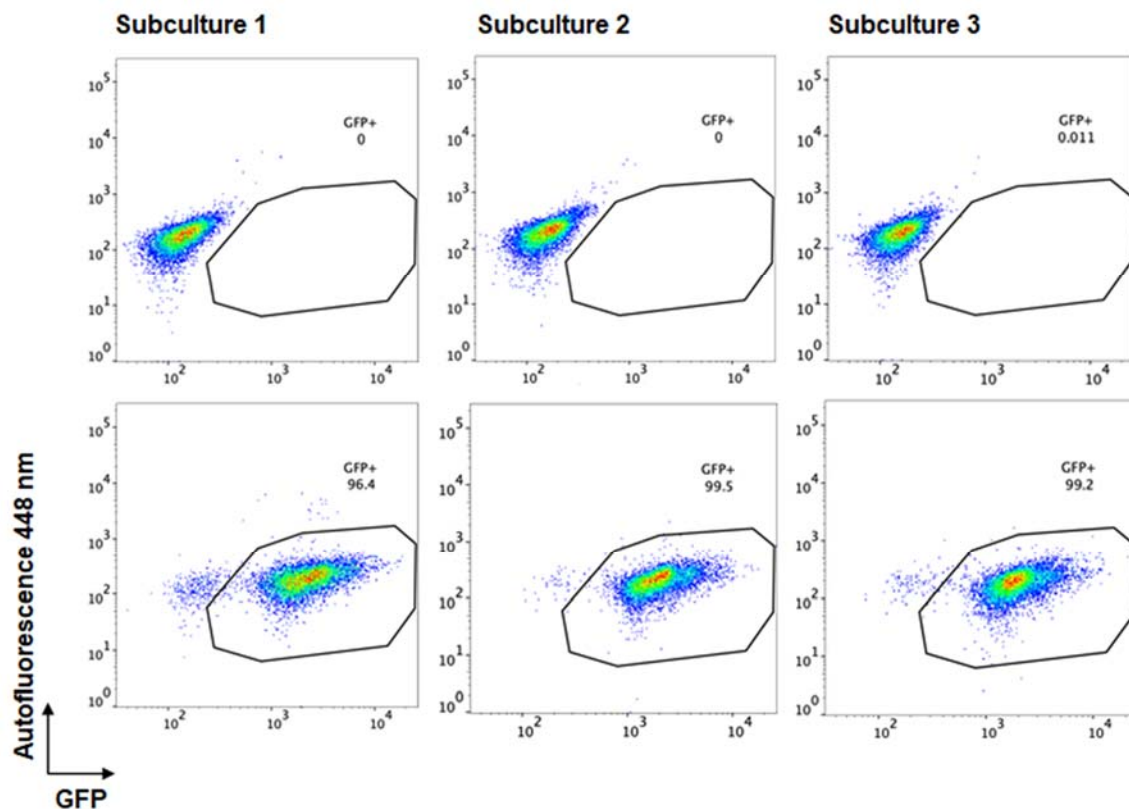


Figure 3.5 Stable and high level of GFP expression in DMi7-21-7^{GFP++} across subculturing.

Pseudocolor dot plots of pPtGE30 (upper lane) and DMi7-21-7^{GFP++} (lower lane) cultures with 530 nm (GFP) in the x axis and autofluorescence at 448 nm in the y axis of the first three subculturing after the second sorting. Gates and frequencies of total GFP were designed according to the negative control pPtGE30 autofluorescence and are shown as reference. GFP, green fluorescent protein.

Next, we assessed the production of eGFP using Western blot analysis (Figure 3.6a). Proteins were extracted from *P. tricornutum* transconjugants including empty vector controls, and initial unsorted (DMi7-21-1, DMi7-21-3, and DMi7-21-7), and sorted (GFP⁺⁺ and GFP⁻⁻) cell cultures. As expected, the negative control (empty vector pPtGE30) showed no GFP specific bands (Figure 3.6a). Positive control (pPtGE33) showed three bands (28, 36, and 58 kDa) where the lower (28 kDa) and the upper (58 kDa) bands corresponded to the cleaved and uncleaved form of the protein eGFP-T2A-mCherry that have a theoretical molecular weight of 29 and 56 kDa respectively (Figure 3.6a). The results are consistent with previously reported ones (Slattery et al., 2018). The 36 kDa

band appears to be a byproduct of the 2A. Indeed, green and red fluorescent proteins linked by a 2A peptide have been used to study the 2A cleavage efficiency. In some cases, it was reported that byproducts were detected in immunoblotting experiments with anti-GFP antibodies (Kim et al., 2011; Minskaia and Ryan, 2013; Souza-Moreira et al., 2018). It raised the question of whether the assigned bands reported in these investigations as cleaved and uncleaved proteins and the byproducts detected by Western blot could be eGFP cleaved and/or uncleaved variants with post-translational modifications.

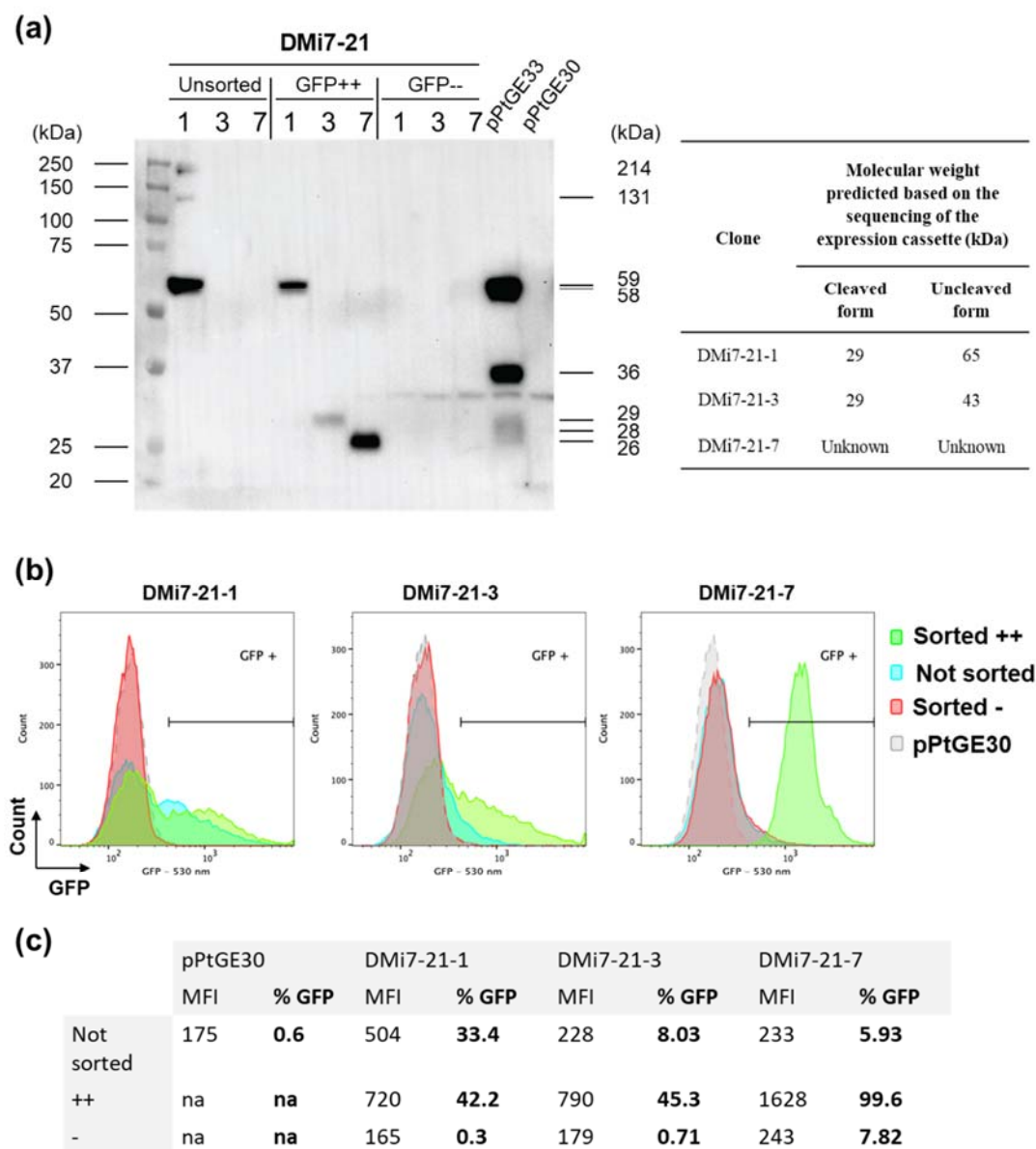


Figure 3.6 Comparison of GFP detected by western blot and by flow cytometry.

(a) Western blot anti-GFP of unsorted, GFP+ sorted (GFP++), GFP- sorted (GFP--) cultures of DMi7-21-1, DMi7-21-3 and DMi7-21-7 compared to pPtGE30 (empty vector) and pPtGE33 (positive control). The size of the proteins from the ladder are identified at the left of the blot. The sizes measured for the detected proteins are indicated at the right of the blot. The table at the right of the blot contains the expected size of the proteins based on their sequencing result. (b) Histogram plots from the same culture of GFP++ sorted (green), GFP-- sorted (red) and unsorted (blue) cultures of DMi7-21-1, DMi7-21-3 and DMi7-21-7 compared to pPtGE30 (grey) autofluorescence. The GFP+ population gate is shown. (c) Table of mean fluorescence intensity (MFI) at 530 nm (GFP) and frequency of GFP+ population (% GFP in bold) in sorted and unsorted DMi7-21-1, -3 and -7 cultures, as compared to pPtGE30 (autofluorescence). Sorted ++, GFP positive population that was sorted twice; Sorted --, GFP negative population that was sorted twice; na, not applicable.

Regarding the unsorted cultures, eGFP was detected only in DMi7-21-1 (Figure 3.6a) and this transconjugant exhibited 33.4% of eGFP+ cells (Figure 3.6b-c). The analysis of the DMi7-21-1 episome sequence revealed the deletion of the initial stop codon and a partial loss of the *fcs* gene sequence (Figure 3.2b and Annexe B – Table B.3). Based on the protein size of 59 kDa detected in the Western blot, it suggests that the eGFP was not cleaved by the self-cleaving T2A peptide. From the sequence of DMi7-21-1 episome, the expected protein sizes were 29 kDa and 65 kDa for the cleaved and uncleaved eGFP, respectively. If the *eGFP-T2A-ech-fcs* cassette would have been unaltered, the cleaved (29 kDa) and uncleaved protein (115 kDa) could have been expected. However, it appears that a deletion in the DNA led to an unexpected 59 kDa eGFP protein. Regarding the unsorted initial cultures of DMi7-21-3 and DMi7-21-7 transconjugants, no band was detected, and only 8.0% and 5.9% of cells were fluorescent, respectively (Figure 3.6a-c). For the group of transconjugants enriched in non-fluorescent cells (GFP-), no proteins were detected. The percentages of eGFP fluorescent cells for the cultures of DMi7-21-1^{GFP--} and DMi7-21-3^{GFP--} are close to 0%. In the case of the culture of DMi7-21-7^{GFP--}, 7.82% of the cells were eGFP fluorescent which is close to the level of the unsorted culture (Figure 3.6c).

As for enriched sorted GFP++ cells, different sizes of eGFP was detected in each of the three transconjugants (Figure 3.6a). For example, the previously observed 59 kDa

band was detected in *P. tricornutum* DMi7-21-1^{GFP++} but not in the other enriched transconjugants. DMi7-21-3^{GFP++} contained a band at 29 kDa which matches the theoretical molecular weight of eGFP associated with the cleaved form of the T2A peptide. Thus, DMi7-21-1^{GFP++} seems then to produce a protein with an uncleaved T2A peptide, while DMi7-21-3^{GFP++} is producing a cleaved protein. In HeLa cells and in the TnT Quick Coupled Transcription/Translation System it was demonstrated that the sequence upstream of the 2A peptide influences its cleavage efficiency (Minskaia and Ryan, 2013). This investigation demonstrated that an addition of a spacer composed of three amino acids (Gly-Ser-Gly) between the N-terminal protein and the 2A peptide can increase the efficiency of the cleavage. In the expectation that this spacer would also increase the cleavage efficiency of the 2a peptides in *P. tricornutum*, we added this spacer upstream of the T2A peptide in all the episomes designed in this study. However, based on the sequencing results of the episome DNA from these two transconjugants, there is no difference in the sequence upstream the T2A region (Figure 3.2 and Annexe B – Table B.3). In this context to explain the difference between the DMi7-21-1 and DMi7-21-3 related to the heterologous proteins produced, it could then be possible that the cleavage of the T2A peptide was influenced by the whole mRNA sequence and/or its structure. This could be explained by the fact that for both, when the T2A would be translated, the only difference that could have an impact on the cleavage would be the mRNA sequence downstream the T2A. In the case of the enriched culture of DMi7-21-7^{GFP++}, composed of almost 100% of eGFP fluorescent cells, the molecular weight of the detected band is 26 kDa which is lower than the band obtained from DMi7-21-3^{GFP++} (Figure 3.6a). In the impossibility to recover the episome from this clone and to sequence it, we cannot analyse the sequence of this protein, but it suggests a truncated form of eGFP.

Based on the Western blot results (Figure 3.6a), there is knowledge that needs to be acquired regarding the cleavage efficiency of the T2A peptide in *P. tricornutum*. Multiple 2A peptides have been compared for their cleavage efficiency in *Drosophila* (Daniels et al., 2014), in CHO cells (Chng et al., 2015), in yeast (Souza-Moreira et al., 2018), in the silkworm *Bombyx mori* (Wang et al., 2015), in the zebrafish embryos (Kim et al., 2011), in human cell lines (Kim et al., 2011; Liu et al., 2017), in adult mice

(Kim et al., 2011), in mouse cell lines (Liu et al., 2017), and recently in *P. tricornutum* (Defrel et al., 2021). In all, the 2A peptides with the best cleavage efficiency were the T2A, the porcine teschovirus-1 2A (P2A), and the ERBV-1 peptide. In the case of *P. tricornutum*, Defrel et al. compared the activity of the GUS enzyme linked to NAT by the P2A or T2A peptides (Defrel et al., 2021). A higher β -glucuronidase activity was observed when the P2A was used between the two enzymes and when the GUS enzyme was in C-terminal position. Based on a relative comparison with their reference strain named Gus5, the relative β -glucuronidase intensity for the transconjugants with GUS in C-terminal was higher than 50% for 10 of 24 and for 23 of 31 clones, respectively for T2A and P2A groups. When *GUS* gene was upstream of the 2A peptide sequence, the transconjugants with a relative β -glucuronidase intensity that was higher than 50% were 2 of 18 for the group with the T2A peptide and 5 of 20 for the ones with the P2A peptide. The authors proposed that when the GUS enzyme is cloned in N-terminal of the construct, the residual 20 amino acids resulting from the cleavage of the 2A peptide could negatively affect its β -glucuronidase activity. The cleavage efficiency of the 2A peptides was not measured by Defrel et al. and it was then not correlated with the β -glucuronidase activity. More studies are required for better understanding 2A peptides for *P. tricornutum* engineering. For its potential as a molecular tool, it will be necessary to investigate the correlation between the sequence of the 2A peptides, their cleavage efficiency, the position effect of the proteins in the expression cassette, and the activities of the proteins of interest.

Next, we analyzed the production of the ech-fcs fusion protein by Western blot. The fusion protein was tagged with a Myc tag in C-terminal, but no tagged proteins were detected on the blot (Annexe B – Figure B.7).

The sequence of the expression cassette has an impact on the percentage of fluorescent transconjugant obtained by bacterial conjugation

Based on the sequencing results from the rescued episomes and flow cytometry, we observed three contexts in which the reporter protein eGFP was detected in the DMi7 transconjugants. First of all, there was a partial or complete deletion of the *fcs* sequence

in DMi7-21-1, DMi7-21-3, DMi7-31-3, and DMi7-31-8. Second, there was a change in the open reading frame (ORF) affecting the *fcs* sequence in DMi7-31-1. Finally, there was a nonsynonymous substitution in the *fcs* sequence of DMi7-31-4 that could have an impact on the structure of the protein. In all these scenarios, the *eGFP-T2A-ech-fcs* protein could not be produced unaltered (Figure 3.2 and Annexe B – Table B.3).

There still could be a possibility that those modifications in the *fcs* sequence and the rearrangement in the episomes could have removed toxic elements or parts that inhibited *P. tricornutum* translation of the unaltered *eGFP-T2A-ech-fcs* gene sequence. The modifications could also be consequences of a random phenomenon. To determine what could be the impact of rearrangements, episomal DNA was extracted and recovered from *P. tricornutum* and then transformed into *E. coli* for bacterial conjugation into wild type strain of *P. tricornutum*. As results, no colonies were obtained from the conjugation of the episomes recovered from DMi7-21-1, DMi7-21-16, DMi7-31-3, DMi7-31-4, and DMi7-31-8 transconjugants. The lack of transconjugants could be explained by the complete or partial deletion of the OriT sequence (Annexe B – Table B.3) necessary to activate the transfer of the episome from *E. coli* to *P. tricornutum* (Strand et al., 2014). The episomes rescued from non-fluorescent DMi7-21-14, DMi7-21-15, and DMi7-31-10 transconjugants were successfully transformed in *P. tricornutum* and led to frequencies of 13.1%, 12.6%, and 15.0% of fluorescent colonies, respectively, which are similar to the results obtained with the conjugation of pDMi7 (Figure 3.1d and Table 3.1). In these rescued episomes, the coding sequence of the expression cassette did not contain any mutation or rearrangement (Figure 3.2c and Annexe B – Table B.3). From this point of view, there is no difference between transforming *P. tricornutum* with those three episomes or with the original episome pDMi7. The eGFP fluorescence of the transconjugants from these rescued episomes is then possibly originating from mutated or rearranged episomes, like the transconjugants from the conjugation with the episome pDMi7. Different results were obtained with the episomes rescued from the two remaining fluorescent transconjugants. Episome originating from the fluorescent clone DMi7-31-1 contained a frameshift in the ORF of the *fcs* sequence (Figure 3.2 and Annexe B – Table B.3). Once it was reintroduced into *P. tricornutum*, 84.1% of the colonies were

detected as fluorescent using fluorescence microscopy (Table 3.1). This percentage of fluorescent colonies is close to what was obtained from the conjugation with the episome pControl where the *ech-fcs* sequence was replaced by *mCherry* (Figure 3.1b,d). Regarding the episome from the clone DMi7-21-3, the quantity of colonies was lower than what was obtained with the others rescued episomes (Annexe B – Figure B.8) with 46% of fluorescent colonies (Table 3.1). Based on the sequencing of the episome recovered from the clone DMi7-21-3, the bacterial conjugation should not be possible. In this episome, there are only two partial duplicates of the OriT cassette where the first partial copy of the OriT cassette contains the first 131 bp and the second one has the first 158 bp on a total length of 771 bp.

Table 3.1 Count of *P. tricornutum* fluorescent colonies using fluorescence microscopy 14 days after the bacterial conjugation with the episomes recovered from the *P. tricornutum* cells initially transformed with pDMi7.

pDMi7 subclones of	GFP detected in the culture used for episome rescue	GFP fluorescent colonies of the pDMi 7 subclones (%)
DMi7-21-3 Replicate #1	GFP +	46.4 (13/28)
DMi7-31-1		84.1 (334/397)
DMi7-21-14	GFP -	13.1 (47/358)
DMi7-21-15		12.6 (42/333)
DMi7-31-10		15.0 (53/354)

GFP + means that GFP fluorescence was initially detected by flow cytometry in the cultures used for episome rescue; GFP - means that GFP fluorescence was not initially detected by flow cytometry in the cultures used for episome rescue.

Conclusion

To engineer *P. tricornutum* for the production of vanillin, we designed a bicistronic expression cassette containing a T2A peptide into an episomal system. We were not able to produce vanillin due to rearrangement of the episomal cassette in *P. tricornutum*. Based on Western blot results and the sequencing of the episomes, it appears that the

sequence downstream of the T2A peptide could have an impact on its cleavage efficiency. Future investigations focusing on 2A peptides in *P. tricornutum* should elucidate the links between the selected 2A peptides, their cleavage efficiency, the genes order in the expression cassette, and their impacts on the proteins produced downstream or upstream the 2A peptide.

The results presented here introduce the possibilities of mechanisms used by *P. tricornutum* to prevent the production of a heterologous protein of interest from extrachromosomal expression systems. It's also possible that the GFP fluorescence and the protein of interest ech-fcs could not be detected. This could be due to a low production or the degradation of the heterologous proteins in the clones harboring an unaltered expression cassette in pDMi7. Indeed, the bacterial conjugation of *P. tricornutum* with an episome harboring the expression cassette for the polyprotein eGFP-T2A-mCherry resulted in 82.1% of colonies producing the eGFP fluorescence. A significant reduction to 8.1% of fluorescent colonies was observed by replacing the *mCherry* sequence by the coding sequences of the enzymes ech and fcs from *Streptomyces* sp. strain V-1.

We herein demonstrated that a screening that is not specific to the production of the proteins or the metabolites of interest can led to the selection of transconjugants that contain a rearranged or mutated episome at the level of the expression cassette. A screening could have been based on the sequence of the episome. However, it would most likely have selected transconjugants where the reporter protein eGFP and the proteins of interest ech and fcs are not detected.

We successfully enriched cultures of *P. tricornutum* by fluorescence-activated cell sorting by flow cytometry. Following the enrichment, it was possible to increase the protein production of a clone without optimizing the culture conditions. Further investigation will be necessary to understand the dynamic of the cell population of the *P. tricornutum* transconjugants. The observations by fluorescence microscopy showed that some colonies are almost completely fluorescent, and some others are only partially

fluorescent. We could not confirm the reason for these differences in fluorescence patterns and what could be the impact on the cultures made from these colonies.

It is the first time that rearrangement of episomes in *P. tricornutum* is investigated. We demonstrated that the episome can be unstable at a molecular level in *P. tricornutum*. Nevertheless, recent publications by other groups presented successful uses of this extrachromosomal expression system. This article should therefore be interpreted as a warning about the limits of the episome in *P. tricornutum* and the importance of the construction strategy. It can be concluded that the screening strategy used in this paper led to the selection of transconjugants that did not contain the designed episome.

Experimental procedures

Microbial Strains and Growth Conditions

E. coli (NEB® 10-beta, New England Biolabs, Canada) was grown in Luria-Bertani (LB) Broth supplemented with appropriate antibiotics (chloramphenicol (15 mg L⁻¹)). *E. coli* (Epi300, Epicenter) was grown in LB Broth supplemented with appropriate antibiotics (gentamicin (20 mg L⁻¹) or chloramphenicol (15 mg L⁻¹) and gentamicin (20 mg L⁻¹)). *P. tricornutum* (CCAP 1055/1, Culture Collection of Algae and Protozoa) was grown in modified L1 medium without silica at 18 °C under cool white fluorescent lights (75 μE m⁻² s⁻¹) and a photoperiod of 16 h light:8 h dark with an agitation of 130 rpm for liquid cultures.

***P. tricornutum* growth medium**

L1 media preparation was adapted from Slattery et al. (2018). It consisted of 917 mL of autoclaved distilled water, 50 mL of 20x stock of NaCl and Na₂SO₄, 10 mL of 100x stock of anhydrous salt, 20 mL of 50x stock solution of hydrous salt, 2 mL of sodium phosphate (NP) stock, 1 mL L1 trace metals stock, 0.5 mL f/2 vitamin solution.

The 20x stock of NaCl and Na₂SO₄ sterilized by autoclave consisted of 245 g L⁻¹ NaCl and 40.9 g L⁻¹ Na₂SO₄. The 100x stock of anhydrous salt sterilized by autoclave consisted of 35 g L⁻¹ KCl, 10 g L⁻¹ NaHCO₃, 5 g L⁻¹ KBr, 1.5 g L⁻¹ H₃BO₃, and 0.15 g L⁻¹ NaF. The 50x stock of hydrous salt sterilized by autoclave consisted of 277.5 g L⁻¹ MgCl₂·6H₂O and 38.5 g L⁻¹ CaCl₂·2H₂O. The sodium phosphate (NP) stock was made in 100 mL H₂O and consisted of 37.5 g NaNO₃ and 2.5 g NaH₂PO₄·H₂O and was sterilized through a 0.2 µm filter. The L1 trace metal stock solution was made by mixing 3.15 g FeCl₃·6H₂O, 4.36 g Na₂EDTA·2H₂O, 0.25 mL (9.8 g L⁻¹ dH₂O) CuSO₄·5H₂O, 3.0 mL (6.3 g L⁻¹ dH₂O) Na₂MoO₄·2H₂O, 1.0 mL (22.0 g L⁻¹ dH₂O) ZnSO₄·7H₂O, 1.0 mL (10.0 g L⁻¹ dH₂O) CoCl₂·6H₂O, 1.0 mL (180.0 g L⁻¹ dH₂O) MnCl₂·4H₂O, 1.0 mL (1.3 g L⁻¹ dH₂O) H₂SeO₃, 1.0 mL (2.7 g L⁻¹ dH₂O) NiSO₄·6H₂O, 0.1 mL (100 mM, Cat. P0758S, NEB) Na₃VO₄, and 1.0 mL (1.94 g L⁻¹ dH₂O) K₂CrO₄ in 1 L H₂O and was sterilized through a 0.2-µm filter. The F/2 vitamin stock solution was made by mixing 200 mg thiamine-HCl, 10 mL of a 0.1 g L⁻¹ biotin stock, and 1 mL of a cyanocobalamin 1 g L⁻¹ stock in 1L H₂O and was sterilized through a 0.2-µm filter.

For agar plates, equal parts of sterilized liquid L1 medium and autoclaved 2% agar were combined and poured into Petri dishes.

Plasmid Construction

All plasmid constructs were done by Gibson assembly using the NEBuilder® HiFi DNA Assembly Bundle for Large Fragments (New England Biolabs, Canada). Amplicons used to do the assemblies were amplified by PCR with PrimeSTAR GXL DNA Polymerase (Takara Bio, Japan) following the manufacturer's protocol. Episome pDMi7 was made by replacing the URA3 element in pPtGE30 with an expression cassette containing 40SRPS8 promoter and FcpA terminator driving *eGFP* linked with a T2A self-cleaving peptide linker to the enzyme *ech* fused with the enzyme *fcs* by a flexible linker (GGGGS)₃ (Figure 3.1a). Expression vector pControl was derived from pDMi7 by inserting *mCherry* ORF next to the T2A peptide instead of *ech*, flexible linker (GGGGS)₃, and *fcs* genes.

The *ech*, and *fcs* containing a Myc tag in 3' genes were codon optimized and were synthesized by Bio Basic (Markham, Ontario, Canada) using the codon usage list extracted from the High-performance Integrated Virtual Environment-Codon Usage Tables (HIVE-CUTs) platform (Athey et al., 2017) with the refseq database on September 21st in 2018 (Annexe B – Figure B.9). The *eGFP* and *T2A* DNA were amplified from pPtGE33, and the 40SRPS8 promoter was amplified from pPtGE19 (Slattery et al., 2018).

All forward and reverse primers used are listed in Annexe B – Table B.4. Plasmids list used and constructed for this study is available in Annexe B – Table B.5. The list of all DNA sequences used in this study is available in Annexe B – Table B.6.

Transformation of *P. tricornutum* by bacterial conjugation from *E. coli* cells

Conjugation protocol was adapted from Karas et al. and Slattery et al. (Karas et al., 2015; Slattery et al., 2018).

Preparation of P. tricornutum cells. A liquid culture of *P. tricornutum* of 4 to 8 days old was used as the starter culture. The cell concentration of 250 μ L culture was adjusted to 1.0×10^8 cells mL^{-1} based on the optical density of a 1/5 diluted sample and calibration curve (Annexe B – Figure B.10). To adjust the concentration, the cells were centrifuged at 3,500 x g at room temperature. The 250 μ L of concentrated culture is plated on $1/2 \times \text{L1}$ 1% agar plates and grown for 4 days. The cells were then scrapped three times with 400 μ L of L1 media and the concentration was adjusted to 5.0×10^8 cells mL^{-1} .

Preparation of E. coli cells. The *E. coli* cells used for bacterial conjugation contain the conjugative plasmid pTA-Mob (Strand et al., 2014) and the cargo plasmid with the expression cassette of interest. A 50 mL of LB was inoculated with 1 mL of an overnight 5 mL culture of *E. coli* and incubated at 37 °C to A_{600} of 0.8–1.0, centrifuged for 10 min at 3,000 x g and resuspended in 500 μ L of SOC media.

Bacterial conjugation. 200 μL of *P. tricornutum* and 200 μL of *E. coli* cells were mixed to initiate the conjugation. The cell mixture was plated on $1/2 \times \text{L1}$ 5% LB 1% agar plates, incubated for 2h at 30 °C in the dark, and then moved to grow for two days at 18 °C in the light. After 2 days, the cells were scrapped twice with 650 μL of L1 media. The scrapped cells were plated with 200 μL three times and the remaining volume ($< 200 \mu\text{L}$) on the fourth plates of $1/2 \times \text{L1}$ 1% agar plates supplemented with Zeocin 50 $\mu\text{g mL}^{-1}$ (or nourseothricin 200 $\mu\text{g mL}^{-1}$ for pPtGE33 transconjugants only). Colonies appeared after 10–14 days of incubation at 18 °C with a photoperiod of 16 h light:8 h dark.

Counts of *P. tricornutum* (Pt) colonies per conjugation were compared statistically by Welch's t-test ($\alpha= 0.05$) in Microsoft Excel with the Data AnalysisToolPak.

The *P. tricornutum* transconjugants were named as follows in our study: (Name of the episome used for conjugation) – (Round of conjugation) (Number of the *E. coli* colony obtained from the Gibson assembly of pDMi7) – (Number of *P. tricornutum* colony after the bacterial conjugation)→As an example: DMi7-21-1

Fluorescence microscopy

Colonies were analysed 14 days after conjugation under a Fluorescent Stereo Microscope Leica M165 FC with GFP filter (Leica cat # 10 447 408) for eGFP fluorescence (emission of 525-550 nm) and RFP filter (Leica cat # 10 450 224) for mCherry fluorescence detection (emission of 605-670 nm). Colonies were observed with a magnification of 80 to 120x.

The percentage of eGFP fluorescent colonies were compared statistically by Welch's t-test ($\alpha= 0.05$) in Microsoft Excel with the Data AnalysisToolPak.

Episome DNA Isolation from *P. tricornutum* and episome rescue

The recovery of episomes from *P. tricornutum* was adapted from Karas et al. (2015), Slattery et al. (2018), and the manufacturer's protocol of the Large Plasmid Mini Kit (Geneaid Biotech Ltd., Taiwan). Briefly, 5 mL of a 7 days old culture of *P. tricornutum* was centrifuged for 10 min at 3,500 x g. The pellet was resuspended in 235 μ L of PDL1 buffer (Geneaid) supplemented with 5 μ L of hemicellulase (100 mg/mL), 5 μ L of lysozyme (25 mg/mL) and 5 μ L of 20T zymolyase solution (10 mg/mL). The mixed solution was incubated for 30 min at 37 °C. To initiate the lysis, 250 μ L of PDL2 buffer was added to the solution and mixed by inversions 5 to 10 times and incubated 2 min at room temperature. The lysis was neutralized by the addition of 375 μ L of PDL3 buffer and mixed by inversions 5 to 10 times and incubated 2 min at room temperature. The mixture was centrifuged for 3 min on a microcentrifuge at maximal speed at room temperature. The manufacturer's protocol from the Large Plasmid Mini Kit (Geneaid) was then followed for the steps of DNA binding, washing, and elution.

To complete the episome rescue, 2 μ L of the miniprep were transformed by heat shock in NEB 10-beta chemically competent *E. coli* cells following the manufacturer's protocol up to the spreading of transformed cells on LB plates. At this point, 100 μ L of the cell mixture was spread on a LB plate with chloramphenicol. The remaining volume was centrifuged 2,000 x g at room temperature for 5 min. The supernatant was partially removed (700 μ L) and the remaining volume is used to resuspend the pellet. The total volume of the cell mixture is then plated on a LB plate with chloramphenicol. After an overnight incubation at 37 °C, an isolated colony was used to inoculate 5 mL of LB culture to proceed for a miniprep following the manufacturer's protocol from the Large Plasmid Mini Kit (Geneaid).

Whole Plasmid Sequencing

Episome pDMi7 was sequenced following the Gibson assembly and served as the reference sequence for further alignment made with CLC Main Workbench 7.7 (QIAGEN, Germany) with the "very accurate" alignment parameter. pDMi7 episomes

from Gibson assembly and episome rescue were completely sequenced by CCIB DNA Core (Massachusetts General Hospital, United States of America) through their Next-Generation sequencing Illumina MiSeq platform.

Promoter region prediction

To analyze the sequence of a putative promoter in DMi7-21-3 clone, a region of 1,000 bp before the first ATG of the longest predicted open reading frame from the sequence result of episome rearrangements was analysed using PlantCARE (Lescot et al., 2002) software for predicting transcription factor binding sites. Besides, to determine if it had the potential consensus transcription initiation sequence from *P. tricornutum* we searched for “TCAH₊₁W” in the selected region (Kadono et al., 2015).

Protein extraction

One-week-old cultures were centrifuged at 1,500 x g for 20 minutes at 4 °C. Pellets were weighed and resuspended with a ratio 1.3 g FW/mL in an extraction buffer (50 mM Tris pH 7.4, 500 mM NaCl, 0.1% Tween20, 1x protease inhibitor cocktail). Sonication was performed 6 times at 35% amplitude, 30 sec on, 30 sec off for 3 min total. Protein extracts were centrifuged at 20,000 x g for 30 min at 4 °C. Supernatant containing the total soluble protein fractions were kept at -20 °C to be used for Western blot. The proteins were quantified with the RC DC™ Protein Assay Kit I (Bio-Rad cat # 5000121).

Western blot

For protein detection, 50 µg of total proteins were loaded in 10% SDS-PAGE. The proteins were then transferred to the 0.2 µm PVDF membrane and transfer settings were; 100V constant and 400 mA for 2 h. Primary antibodies were incubated overnight at 4 °C. Primary antibody for eGFP was purchased from Cedarlane (Ontario L7L 5R2 Canada, cat. #CLH106AP) and for Myc Tag from ThermoFisher Scientific (Illinois 61101

USA, cat. #MA1-21316). Both were used at a 1:1,000 dilution in Tris-buffered saline, 0.1% Tween 20 (TBST) solution with 3% BSA. After three washes with TBST solution, the blots were incubated for 1 hour in a 1:20,000 dilution of secondary antibody in TBST solution with 5% milk. The secondary antibody Immun-Star Goat Anti-Mouse (GAM)-HRP Conjugate was purchased from Bio-Rad (Ontario L5T 1C9 Canada, cat.#1705047). *P. tricornutum* clone containing pPtGE33 was used as a positive control for eGFP expression. A quantity of 10 ng of Multiple Tag from GenScript (cat. #M0101) was used as a positive control for Myc Tag detection. After three washes of the membrane with TBST solution, protein detection was realized by using Clarity Max Western ECL Substrate-Luminol solution from Bio-Rad (cat #1705062S). Chemiluminescence detection and Ponceau S stained (Glacial Acetic Acid 5% v/v, Ponceau Red dye 0.1% m/v) of the blots were visualized using ChemiDoc Imaging System with Image Lab Touch Software (Bio-Rad cat # 12003153) and Image Lab™ Software (Bio-Rad cat # 1709690). The molecular weight of the protein corresponding to the detected band was calculated with Image Lab™ Software and the method point-to-point (semi-log).

Flow cytometry and fluorescence-activated cell sorting (FACS)

The BD FACSMelody (BD Biosciences, La Jolla, CA, USA) equipped with blue (488 nm), red (640 nm) and violet (405 nm) lasers was used to sort *P. tricornutum* transformed transconjugants according to eGFP production. Prior to the first sort, selected transconjugants were grown in L1 liquid medium supplemented with corresponding antibiotic and grown for 7 days. *P. tricornutum* cultures were washed in L1 medium, filtered on a 100 µm Nylon Net filter (Merck Millipore, Ireland) and diluted to an $OD_{730} = 0.1$ in L1 media prior to sorting.

Events were acquired at a fixed flow rate of 1 and at least 10,000 events were analysed. Cells were gated according to FSC-A (forward scatter area) and SSC-A (side scatter area) parameters and doublons were excluded according to FSC-H (height) vs. FSC-W (width) and SSC-H vs SSC-W plots. Chloroplast autofluorescence was measured on the PerCP channel (700/54 nm). Cells with high levels of PerCP fluorescence

were further gated whereas cells with high non-specific autofluorescence were excluded based on their emission in the 448/45 nm channel. eGFP was further analysed on the 527/32 nm band-pass filter channel. Sorting parameters were set on purity parameter. Sorted cells were collected in 1.5 mL Eppendorf containing 500 μ L of L1 media without antibiotics. Sorted cells were centrifuged 10 min at 3,500 x g and 90–95% of the supernatant was removed and replaced by 500 μ L of L1 media supplemented with zeocin 50 μ g/mL and chloramphenicol 25 μ g/mL.

For the second round of sorting, 1st round-sorted cultures were incubated for 7 days and diluted in 1 mL to an OD₇₃₀ of 0.1. The cultures were then grown for another 7 days before being sorted a second time, according to the same procedure. Figures and statistics were analysed using BD FlowJo version 10 software (BD Biosciences, La Jolla, CA, USA, 2020). At least 10 000 events were acquired for each sample.

Total RNA extraction and RT-PCR

Total RNA from *P. tricornutum* from flash frozen biomass (10 mL of a 7 days old culture) was extracted using the RNeasy Plant Mini Kit according to the manufacturer's protocol (QIAGEN, Germany). Followed by DNase I (NEB, Canada) treatment at 37 °C for 15 minutes according to the manufacturer's protocol to remove episome contamination and purification using RB columns from Plant Total RNA Mini Kit (GENEAID, Taiwan). The RNA quality was confirmed by migration at 100V on a 1% agarose gel for 35 min. RT-PCR was carried out using High-Capacity cDNA Reverse Transcription Kit according to the manufacturer's protocol (Applied Biosystems, USA). For PCR reaction, all the transconjugants were tested using a forward primer binding at *ech* (DMI7_F) and a reverse primer that binds at the beginning of *fcs* (DMI7_R). The reaction conditions were: initial denaturation at 95 °C for 30s; 30 cycles of 95 °C (30s), 50 °C (40s) and 68 °C (16s), with a final extension at 68 °C for 5 min. PCR products were visualized in 1% agarose gel and the length of 259 bp was expected for the positive amplification. The forward and reverse primers sequence used for PCR are listed in Annexe B – Table B.4.

References

- Athey, J., Alexaki, A., Osipova, E., Rostovtsev, A., Santana-Quintero, L.V., Katneni, U., Simonyan, V. and Kimchi-Sarfaty, C. (2017) A new and updated resource for codon usage tables. *BMC Bioinformatics* **18**, 391.
- Bowler, C., Allen, A.E., Badger, J.H., Grimwood, J., Jabbari, K., Kuo, A., Maheswari, U., Martens, C., Maumus, F., Otilar, R.P., Rayko, E., Salamov, A., et al. (2008) The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **456**, 239-244.
- Butler, T., Kapoore, R.V. and Vaidyanathan, S. (2020) *Phaeodactylum tricornutum*: A Diatom Cell Factory. *Trends Biotechnol* **38**, 606-622.
- Chng, J., Wang, T., Nian, R., Lau, A., Hoi, K.M., Ho, S.C., Gagnon, P., Bi, X. and Yang, Y. (2015) Cleavage efficient 2A peptides for high level monoclonal antibody expression in CHO cells. *MAbs* **7**, 403-412.
- Daboussi, F., Leduc, S., Maréchal, A., Dubois, G., Guyot, V., Perez-Michaut, C., Amato, A., Falciatore, A., Juillerat, A., Beurdeley, M., Voytas, D.F., Cavarec, L. and Duchateau, P. (2014) Genome engineering empowers the diatom *Phaeodactylum tricornutum* for biotechnology. *Nature communications* **5**, 3831.
- D'Adamo, S., Schiano di Visconte, G., Lowe, G., Szaub-Newton, J., Beacham, T., Landels, A., Allen, M.J., Spicer, A. and Matthijs, M. (2019) Engineering the unicellular alga *Phaeodactylum tricornutum* for high-value plant triterpenoid production. *Plant biotechnology journal* **17**, 75-87.
- Daniels, R.W., Rossano, A.J., Macleod, G.T. and Ganetzky, B. (2014) Expression of multiple transgenes from a single construct using viral 2A peptides in *Drosophila*. *PloS one* **9**, e100637.
- De Riso, V., Raniello, R., Maumus, F., Rogato, A., Bowler, C. and Falciatore, A. (2009) Gene silencing in the marine diatom *Phaeodactylum tricornutum*. *Nucleic Acids Res* **37**, e96.
- Defrel, G., Marsaud, N., Rifa, E., Martins, F. and Daboussi, F. (2021) Identification of Loci Enabling Stable and High-Level Heterologous Gene Expression. *Front Bioeng Biotechnol* **9**, 734902.
- Diner, R.E., Bielinski, V.A., Dupont, C.L., Allen, A.E. and Weyman, P.D. (2016) Refinement of the Diatom Episome Maintenance Sequence and Improvement of Conjugation-Based DNA Delivery Methods. *Front Bioeng Biotechnol* **4**, 65.

- Fabris, M., George, J., Kuzhiumparambil, U., Lawson, C.A., Jaramillo-Madrid, A.C., Abbriano, R.M., Vickers, C.E. and Ralph, P. (2020) Extrachromosomal Genetic Engineering of the Marine Diatom *Phaeodactylum tricornutum* Enables the Heterologous Production of Monoterpenoids. *ACS synthetic biology* **9**, 598-612.
- Filloramo, G.V., Curtis, B.A., Blanche, E. and Archibald, J.M. (2021) Re-examination of two diatom reference genomes using long-read sequencing. *BMC genomics* **22**, 379.
- Gallage, N.J., Hansen, E.H., Kannangara, R., Olsen, C.E., Motawia, M.S., Jørgensen, K., Holme, I., Hebelstrup, K., Grisoni, M. and Møller, B.L. (2014) Vanillin formation from ferulic acid in *Vanilla planifolia* is catalysed by a single enzyme. *Nature communications* **5**, 4037.
- George, J., Kahlke, T., Abbriano, R.M., Kuzhiumparambil, U., Ralph, P.J. and Fabris, M. (2020) Metabolic Engineering Strategies in Diatoms Reveal Unique Phenotypes and Genetic Configurations With Implications for Algal Genetics and Synthetic Biology. *Front Bioeng Biotechnol* **8**, 513.
- Giguere, D.J., Bahcheli, A.T., Slattery, S.S., Patel, R.R., Browne, T.S., Flatley, M., Karas, B.J., Edgell, D.R. and Gloor, G.B. (2022) Telomere-to-telomere genome assembly of *Phaeodactylum tricornutum*. *PeerJ* **10**, e13607.
- Hamilton, M.L., Haslam, R.P., Napier, J.A. and Sayanova, O. (2014) Metabolic engineering of *Phaeodactylum tricornutum* for the enhanced accumulation of omega-3 long chain polyunsaturated fatty acids. *Metab Eng* **22**, 3-9.
- Hempel, F. and Maier, U.G. (2012) An engineered diatom acting like a plasma cell secreting human IgG antibodies with high efficiency. *Microbial cell factories* **11**, 126.
- Jones, P.R. (2014) Genetic instability in cyanobacteria - an elephant in the room? *Front Bioeng Biotechnol* **2**, 12.
- Kadono, T., Miyagawa-Yamaguchi, A., Kira, N., Tomaru, Y., Okami, T., Yoshimatsu, T., Hou, L., Ohama, T., Fukunaga, K., Okauchi, M., Yamaguchi, H., Ohnishi, K., Falcatore, A. and Adachi, M. (2015) Characterization of marine diatom-infecting virus promoters in the model diatom *Phaeodactylum tricornutum*. *Scientific reports* **5**, 18708.
- Karas, B.J., Diner, R.E., Lefebvre, S.C., McQuaid, J., Phillips, A.P., Noddings, C.M., Brunson, J.K., Valas, R.E., Deerinck, T.J., Jablanovic, J., Gillard, J.T., Beerli, K., Ellisman, M.H., Glass, J.I., Hutchison, C.A., 3rd, Smith, H.O., Venter, J.C., Allen, A.E., Dupont, C.L. and Weyman, P.D. (2015) Designer diatom episomes delivered by bacterial conjugation. *Nature communications* **6**, 6925.

- Kim, J.H., Lee, S.R., Li, L.H., Park, H.J., Park, J.H., Lee, K.Y., Kim, M.K., Shin, B.A. and Choi, S.Y. (2011) High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLoS one* **6**, e18556.
- Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., Rouzé, P. and Rombauts, S. (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res* **30**, 325-327.
- Liu, Z., Chen, O., Wall, J.B.J., Zheng, M., Zhou, Y., Wang, L., Vaseghi, H.R., Qian, L. and Liu, J. (2017) Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. *Scientific reports* **7**, 2193.
- Minskaia, E. and Ryan, M.D. (2013) Protein coexpression using FMDV 2A: effect of "linker" residues. *Biomed Res Int* **2013**, 291730.
- Moosburner, M.A., Gholami, P., McCarthy, J.K., Tan, M., Bielinski, V.A. and Allen, A.E. (2020) Multiplexed Knockouts in the Model Diatom *Phaeodactylum* by Episomal Delivery of a Selectable Cas9. *Frontiers in microbiology* **11**, 5.
- Ni, J., Tao, F., Du, H. and Xu, P. (2015) Mimicking a natural pathway for de novo biosynthesis: natural vanillin production from accessible carbon sources. *Scientific reports* **5**, 13670.
- Russo, M.T., Aiese Cigliano, R., Sanseverino, W. and Ferrante, M.I. (2018) Assessment of genomic changes in a CRISPR/Cas9 *Phaeodactylum tricornutum* mutant through whole genome resequencing. *PeerJ* **6**, e5507.
- Seo, S., Kim, J., Lee, J.W., Nam, O., Chang, K.S. and Jin, E. (2020) Enhanced pyruvate metabolism in plastids by overexpression of putative plastidial pyruvate transporter in *Phaeodactylum tricornutum*. *Biotechnology for biofuels* **13**, 120.
- Sharma, A.K., Nymark, M., Sparstad, T., Bones, A.M. and Winge, P. (2018) Transgene-free genome editing in marine algae by bacterial conjugation - comparison with biolistic CRISPR/Cas9 transformation. *Scientific reports* **8**, 14401.
- Slattery, S.S., Diamond, A., Wang, H., Therrien, J.A., Lant, J.T., Jazey, T., Lee, K., Klassen, Z., Desgagné-Penix, I., Karas, B.J. and Edgell, D.R. (2018) An Expanded Plasmid-Based Genetic Toolbox Enables Cas9 Genome Editing and Stable Maintenance of Synthetic Pathways in *Phaeodactylum tricornutum*. *ACS synthetic biology* **7**, 328-338.
- Slattery, S.S., Wang, H., Giguere, D.J., Kocsis, C., Urquhart, B.L., Karas, B.J. and Edgell, D.R. (2020) Plasmid-based complementation of large deletions in *Phaeodactylum tricornutum* biosynthetic genes generated by Cas9 editing. *Scientific reports* **10**, 13879.

- Souza-Moreira, T.M., Navarrete, C., Chen, X., Zanelli, C.F., Valentini, S.R., Furlan, M., Nielsen, J. and Krivoruchko, A. (2018) Screening of 2A peptides for polycistronic gene expression in yeast. *FEMS Yeast Res* **18**.
- Strand, T.A., Lale, R., Degnes, K.F., Lando, M. and Valla, S. (2014) A new and improved host-independent plasmid system for RK2-based conjugal transfer. *PloS one* **9**, e90372.
- Taparia, Y., Zarka, A., Leu, S., Zarivach, R., Boussiba, S. and Khozin-Goldberg, I. (2019) A novel endogenous selection marker for the diatom *Phaeodactylum tricornutum* based on a unique mutation in phytoene desaturase 1. *Scientific reports* **9**, 8217.
- Wang, Y., Wang, F., Wang, R., Zhao, P. and Xia, Q. (2015) 2A self-cleaving peptide-based multi-gene expression system in the silkworm *Bombyx mori*. *Scientific reports* **5**, 16273.
- Yang, H., Barros-Rios, J., Kourteva, G., Rao, X., Chen, F., Shen, H., Liu, C., Podstolski, A., Belanger, F., Havkin-Frenkel, D. and Dixon, R.A. (2017) A re-evaluation of the final step of vanillin biosynthesis in the orchid *Vanilla planifolia*. *Phytochemistry* **139**, 33-46.
- Yang, W., Tang, H., Ni, J., Wu, Q., Hua, D., Tao, F. and Xu, P. (2013) Characterization of two *Streptomyces* enzymes that convert ferulic acid to vanillin. *PloS one* **8**, e67339.
- Yu, Z., Geisler, K., Leontidou, T., Young, R.E.B., Vonlanthen, S.E., Purton, S., Abell, C. and Smith, A.G. (2021) Droplet-based microfluidic screening and sorting of microalgal populations for strain engineering applications. *Algal Res* **56**, None.

CHAPITRE IV

IS THERE A SPICE MISSING IN THE RECIPE? THE ELUCIDATED INTRACELLULAR LOCALIZATION OF VANILLIN BIOSYNTHESIS FROM L-PHENYLALANINE NEEDS MORE INVESTIGATIONS

Andrew Diamond¹, Simon Barnabé¹, Isabel Desgagné-Penix^{1,2}

¹ Department of Chemistry, Biochemistry and Physics, Université du Québec à Trois-Rivières, Trois-Rivières, Québec, Canada

² Plant Biology Research Group, Université du Québec à Trois-Rivières, Trois-Rivières, QC, Canada

L'article de type « Viewpoint » a été accepté pour publication dans le journal *Plant Biology*.

4.1 Contribution des auteurs

AD a développé l'idée du thème présenté dans cet article et a rédigé l'article. SB et IDP ont corrigé et amélioré l'article.

4.2 Résumé de l'article (français)

Le chloroplaste a été désigné comme le site de la biosynthèse de la vanilline chez la plante *Vanilla planifolia* à partir de l'acide aminé L-phénylalanine. Cette conclusion avait été tirée en incubant des chloroplastes isolés avec cet acide aminé. Ainsi, la production de vanilline détectée menait à l'interprétation que le chloroplaste était suffisant pour convertir la L-phénylalanine en vanilline. Cependant, en effectuant une revue de littérature sur le lien entre le chloroplaste et le réticulum endoplasmique (RE), il a été possible de

suggérer que les chloroplastes isolés n'étaient probablement pas purs. Puisque la liaison physique entre le RE et le chloroplaste peut résister à certaine lyse cellulaire, il est probable que les chloroplastes aient été co-isolés avec des fragments de RE. De plus, des études ont démontré que certaines enzymes accumulées dans le cytoplasme pouvaient former des liaisons non covalentes avec des enzymes ancrées dans les membranes du RE. Ces liaisons permettaient à ces enzymes cytoplasmiques d'être isolées avec les chloroplastes et les fragments du RE tout en conservant leur activité enzymatique. L'article présenté dans le chapitre IV démontre que la localisation intracellulaire de la conversion de la L-phénylalanine en vanilline chez *V. planifolia* doit être réévaluée en étudiant l'apport des enzymes cytoplasmiques et du RE.

4.3 Article complet (anglais)

Introduction

Vanilla extract is a complex mix of over 200 aromatic compounds extracted from cured vanilla pods, from which, vanillin (4-hydroxy-3-methoxybenzaldehyde) is the major and the most popular flavor compound in the world (Sinha et al., 2008). The main natural resource comes from the pods which are obtained by cultivating vanilla orchids, mainly *Vanilla x tahitensis*, *V. pompona* and *V. planifolia*, the last one being the most cultivated for commercial production due to its aromatic qualities (Gallage and Møller, 2018). Considerable effort has been made to completely elucidate the vanillin biosynthesis pathway in this species with some disagreement reported. However, it is widely accepted that L-phenylalanine and hydroxycinnamic acids (cinnamic, *p*-coumaric, caffeic and ferulic acids) are precursors of vanillin biosynthesis (Figure 4.1a) (Negishi et al., 2009; Gallage et al., 2018). However, there is still debate around the identity of the final enzyme and the intracellular localization of the metabolic pathway. Given its commercial importance, vanillin biosynthesis needs to be resolved to allow for a promising toolkit with which synthetic biologists can use for alternative production.

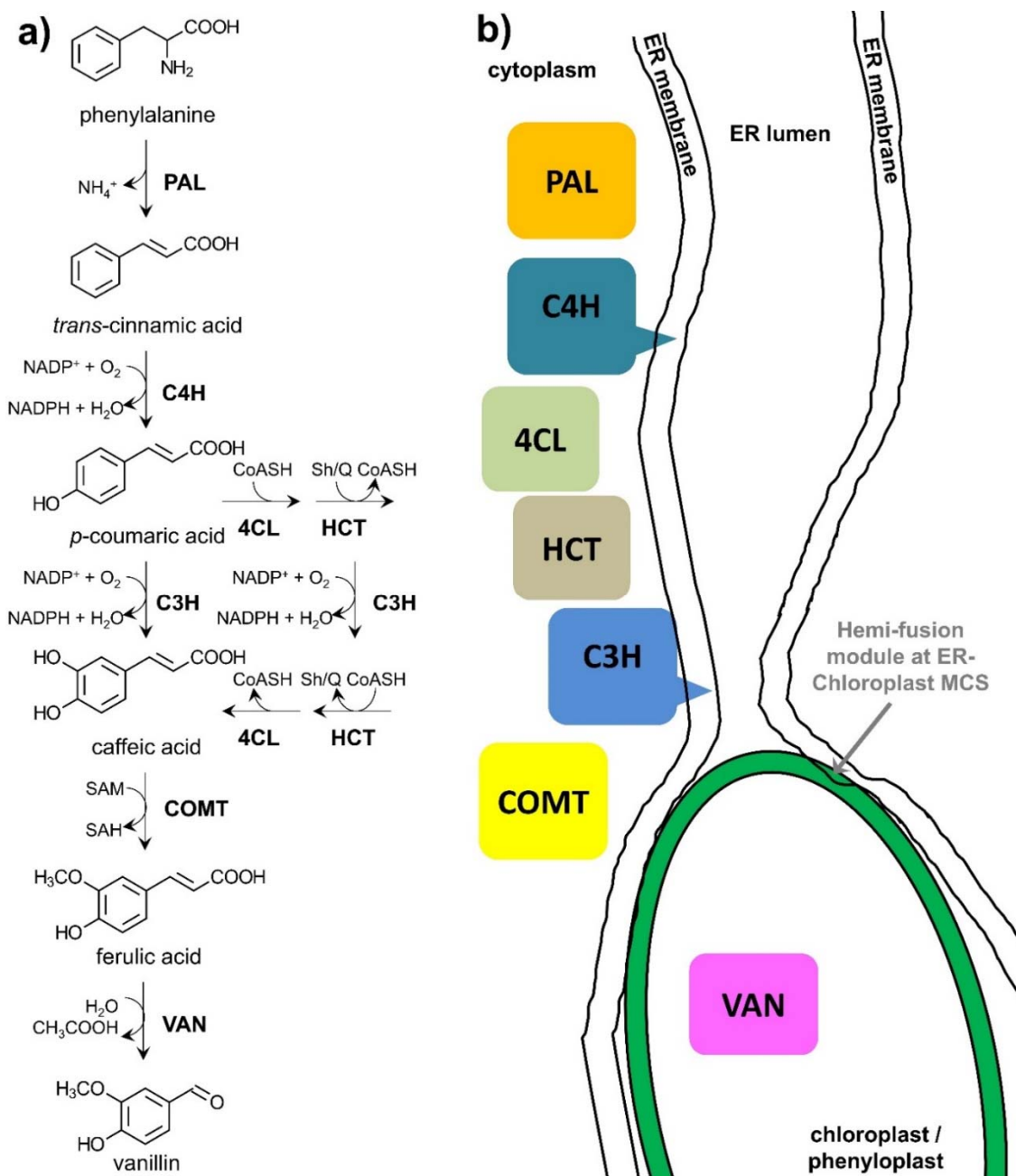


Figure 4.1 The simplified biosynthetic pathway of vanillin and its potential intra-cellular localization.

(a) The enzymes proposed by Gallage and colleagues, 2014; 2018 to convert the vanillin from the *L*-phenylalanine. (b) A schematic representation of the proposed intra-cellular localization of the enzymes involved in the vanillin biosynthesis. The diagram showing the continuity of chloroplast and ER membranes was reported using freeze-fracturing techniques, transorganellar complementation and regulation of enzyme activities (McLean et al., 1988; Mehrshahi et al., 2013). The black and green lines overlapping depict the two outer leaflets merging on each side. Abbreviations are in the text.

Vanillin biosynthesis

Several possible phenylpropanoid metabolic pathways (CoA-dependent or not, β -oxidative or not) leading to vanillin have been proposed. Pulse-chase experiments with ^{14}C -labeled compounds showed that the biosynthetic pathway for vanillin is cinnamic acid \rightarrow 4-coumaric acid \rightarrow caffeic acid \rightarrow ferulic acid \rightarrow vanillin in mature vanilla pods (Negishi et al., 2009). In *V. planifolia*, vanillin biosynthesis occurs in plastids or ‘phenyloplasts’ which are differentiated plastids. It was reported that all the enzymes required for the conversion of [^{14}C]-phenylalanine to [^{14}C]-vanillin glucoside were confined within that organelle (Gallage et al. 2018).

Vanillin biosynthesis starts from the primary metabolism via the amino acid L-phenylalanine (Figure 4.1a). It is first converted into cinnamic acid by the phenylalanine ammonia lyase (PAL) and then cinnamate 4-hydroxylase (C4H), a cytochrome P450-dependent monooxygenase, catalyzes a hydroxylation to form *p*-coumaric acid. Next, coumarate 3-hydroxylase (C3H) [also known as *p*-coumaroyl (shikimate/quinic) 3-hydroxylase] hydroxylates free *p*-coumaric acid or *p*-coumaroyl residue at C3 to produce caffeic acid or caffeoyl residue, respectively (Figure 4.1a). Lately, a non-membrane bound C3H enzyme [also known as bifunctional cytosolic ascorbate peroxidase] catalyzing the direct 3-hydroxylation of *p*-coumaric acid to caffeic acid was characterized from *Brachypodium distachyon* and *A. thaliana* (Barros et al., 2019) but as of yet, not reported in *V. planifolia*. The CoA-dependent non- β -oxidative pathway would pass through the enzymatic activities of 4-hydroxycinnamoyl-CoA ligase (4CL), hydroxycinnamoyltransferase (HCT), and C3H to yield caffeic acid, which is methylated by the caffeic acid *O*-methyl transferase (COMT) to produce ferulic acid (Figure 4.1a).

A CoA-independent non- β -oxidative pathway has been proposed where hydroxycinnamic acids intermediates undergo lateral chain shortening via reverse aldol reaction, releasing acetate and benzaldehyde. In *V. planifolia*, a 4-hydroxybenzaldehyde synthase (HBS) was purified and shown to convert *p*-coumaric acid to 4-hydroxybenzaldehyde and acetate (Podstolski et al., 2002).

Gallage et al. (2014) identified the last reaction involving the conversion of ferulic acid to vanillin. The enzyme named vanillin synthase (*VpVAN*), belong to the hydratase/lyase type (Gallage et al., 2014). Even if a complete biosynthetic pathway has been proposed for vanillin, there is still disagreement around the catalytic activity of *VpVAN*. Indeed, Yang and colleagues (2017) investigated the enzymatic activity of *VpVAN* in heterologous organisms (*Escherichia coli*, *Saccharomyces cerevisiae*, *Medicago truncatula* hairy roots, and *Arabidopsis thaliana* seedlings) and repeated the *in vitro* transcription/translation experiment that was used to first characterize *VpVAN*, but were unable to confirm the activity of the enzyme (Yang et al., 2017). The differences in the results might be explained by the lack of codon optimizations in the systems used and the use of lower sensitivity analytical equipment, as discussed by Gallage et al. (2018). As proposed by Dixon, *VpVAN* may be part of a protein complex in which another enzyme could have assisted vanillin biosynthesis (Dixon, 2018). This would support the hypothesis that *VpVAN* alone may not be sufficient to catalyze the conversion of ferulic acid into vanillin (Yang et al., 2017). In light of those conflicting results, the debate around *VpVAN* activity is not over. In addition, a second enzyme named phenylpropanoid 2,3-dioxygenase (PPDiox) was reported to catalyze the cleavage of the ferulic acid side chain to form vanillin in *V. planifolia* (Negishi and Negishi, 2017). Remarkably, it appears that proteins from two different protein families can catalyze the same reaction.

Independent studies using heterologous systems support the ability of *VpVAN* to convert ferulic acid into vanillin (Havkin-Frenkel et al., 2006; Havkin-Frenkel and Podstolski, 2007; Chee et al., 2017; Arya et al., 2022). For example, Chee and colleagues (2017) expressed *VpVAN* in callus cultures of *Capsicum frutescens* [where vanillin is naturally produced as a precursor of capsaicin biosynthesis] and reported an increase in vanillin accumulation (Chee et al., 2017). Another example involves metabolically engineered rice cell cultures with *VpVAN* as a plant-based alternative to microbial vanillin production systems (Arya et al., 2022). In this study, rice calli derived from embryonic rice cells were engineered with a codon-optimized *VpVAN* gene using *Agrobacterium*-mediated transformation. High-performance liquid chromatography identified the biosynthesis of vanillin in transgenic calli lines (Arya et al., 2022).

Although both studies did not directly investigate the catalytic activity of *Vp*VAN, they serve as proof-of-concept to confirm its involvement in vanillin production.

The chloroplast was reported to be sufficient to convert L-phenylalanine into vanillin

In vanilla pods, vanillin is stored as vanillin glucoside in phenyloplasts (Brillouet et al., 2014) and using immunofluorescence, the enzyme *Vp*VAN was also located in the chloroplast (Gallage et al., 2018) thus making the conversion site of vanillin the same as its storage site. Gallage and colleagues investigated whether chloroplasts harbor the whole vanillin biosynthesis pathway using a radioactive labelling technique (Gallage et al., 2018). Intact chloroplasts from *V. planifolia* pods were isolated using Percoll gradients, then incubated with the radiolabeled precursor [¹⁴C]-phenylalanine resulting in [¹⁴C]-vanillin glucoside production detected by thin-layer chromatography. They concluded that all enzymes necessary to convert L-phenylalanine into vanillin glucoside were present in the chloroplast. In this investigation, the intactness and purity of the isolated chloroplasts were determined by light microscopy; however, this method does not allow the detection of co-isolated fragments from other organelles.

It should be noted that few studies have reported on the presence of enzymes of the phenylpropanoid pathway in the chloroplasts, including the PAL and C4H (Saunders and McClure, 1975), as well as unidentified enzymes that can convert *p*-coumaric acid into caffeic acid (Satô, 1966). It was also reported that the major isoform of PAL extracted from spinach leaves was found outside of the chloroplast, but two other isoforms were present inside the chloroplast (Nishizawa et al., 1979). Nevertheless, it was proposed that the interaction between the endoplasmic reticulum (ER) and the chloroplast may be necessary for complete biosynthetic machinery (Møller and Laursen, 2021). From the review of various publications, we suggest herein that this hypothesis should be seriously considered to clarify the intra-cellular localization of the conversion of L-phenylalanine into vanillin (Figure 4.1b).

Cytosolic and ER-anchored vanillin biosynthetic enzymes may co-purify with isolated chloroplasts

Even though some enzymes are considered to be located in different cellular compartments, they can sometime interact nonetheless. As such, it was demonstrated that the cytosolic enzyme PAL1, an isoform from *Nicotiana tabacum* involved in lignin biosynthesis, was able to associate with C4H, a membrane enzyme anchored in the ER with its active site located in the cytoplasm, to form a metabolon (*i.e.* multi-protein complex) (Achnine et al., 2004; Bassard et al., 2012a; Bassard et al., 2012b). The formation of this protein complex promotes a synergy between those two enzymes. This association was studied in microsomes, vesicle-like particles formed during the breakdown of cells and mostly made from ER fragments. PAL was reported as partially associated to the microsomes (Rasmussen and Dixon, 1999) It was also demonstrated that in microsome extracts, *p*-coumaric acid was produced from L-phenylalanine more efficiently than from cinnamic acid (Czichi and Kindl, 1975; 1977; Rasmussen and Dixon, 1999). This implies that in cell lysates, PAL could associate with the ER through its interaction with C4H. This association could provide the catalytic activity to convert L-phenylalanine to *p*-coumaric acid through the first steps of the phenylpropanoid pathway.

Gallage and colleagues suggested the chloroplast as the unique site of vanillin biosynthesis from L-phenylalanine (Gallage et al., 2018). However, this conclusion neglects the proximity and near continuity between ER and plastids, which have been demonstrated through electronic microscopy and transorganellar complementation investigations (McLean et al., 1988; Whatley et al., 1991; Kaneko and Keegstra, 1996; Mehrshahi et al., 2013). The literature regarding the membrane contact sites (MCSs) between those two organelles has been recently reviewed (Block and Jouhet, 2015; Wang and Dehesh, 2018; Liu and Li, 2019). Indeed, MCS could be important for several metabolic processes. The ER-plastid MCSs have been proposed to allow the transfer of intermediate metabolites involved in the biosynthesis of specialized metabolites such as vitamins and lipids (Mehrshahi et al., 2013; Negi et al., 2018; Michaud and Jouhet, 2019). As an example, the disruption of tocopherol biosynthesis through mutations of three

plastid-localized enzymes, was complemented via the retargeting of those enzymes to the ER (Mehrshahi et al., 2013) suggesting that the enzymes localized in the ER could have access to the nonpolar metabolites pool from the plastids. Another example from Negi et al. (2018) involves the lipid metabolism of stomatal guard cells. It was demonstrated that compared to mesophyll cells, the low biosynthesis of lipids from the prokaryotic pathway, which is restricted to the plastids, was compensated by the eukaryotic pathway involving intermediates from the ER. The possible metabolism of lipids from the ER-plastid MCSs was reviewed by Michaud and Jouhet (2019). The interaction between the ER and the plastids during biosynthesis of vanillin or other phenylpropanoids has not been reported yet. This could be the topic for further studies on the vanillin biosynthesis.

In addition, the strong attraction between the ER and the chloroplasts was demonstrated through optical fragmentation of *A. thaliana* protoplasts (Andersson et al., 2007a; b). Protoplasts were fragmented with a nitrogen laser, and a chloroplast was captured with optical tweezers. The displacement of the captured chloroplast caused stretching of the ER strands that accompanied the moving plastid (Andersson et al., 2007b). Moreover, confocal microscopy observation of isolated chloroplasts showed a colocalization between the fluorescence from GFP coming from the ER lumen, and the chloroplasts (Andersson et al., 2007a; b). Andersson and colleagues have also detected activity from two ER-associated enzymes, NADH-dependent cytochrome c reductase and phosphatidylcholine synthase, from isolated chloroplasts (Andersson et al., 2007b), highlighting the possibility that ER fragments, potentially attached through MCS, could be co-isolated with chloroplasts.

Additional investigations based on omics sciences could identify candidate enzymes of the vanillin biosynthetic pathway and along with their subcellular localization. Multiple strategies based on subcellular proteomics have been reviewed recently (Christopher et al., 2021) and should be useful for this topic. A starting point could be to identify the candidate enzymes involved in the vanillin biosynthesis that are found in the isolated chloroplasts then to confirm their localization by immunofluorescence in intact cells. If all the necessary enzymes are detected into the plastids, it would reinforce the

conclusion from Gallage et al. (2018). If the enzymes proposed by Gallage et al. (2018) are located inside the chloroplast, a method could be developed to inactivate the enzymes found outside the isolated chloroplast before incubating them with the radiolabeled precursors. This way, the precursors will not be consumed by enzymes co-purified with the intact chloroplasts. The impact of those contaminating proteins will be the main challenge that will be faced by the researchers working on this topic.

Conclusion

Vanillin is the most widely used flavor compound in the world and is employed extensively in food, beverage, perfume, and pharmaceutical industries. Several biosynthetic pathways for vanillin have been proposed (Havkin-Frenkel and Belanger, 2007; Dixon, 2010; Kundu, 2017; Khoyratty et al., 2018), but a complete CoA-independent non- β -oxidative pathway involving *VpVAN* for the last step was reported (Gallage et al., 2014). Although, there is intense debate about *VpVAN* activity, recent independent studies using heterologous systems support the ability of *VpVAN* to convert ferulic acid into vanillin. Gallage and colleagues isolated chloroplasts to describe their biosynthetic capacity to form vanillin (Gallage et al., 2018). In this study, the method used to isolate chloroplasts did not prevent the co-isolation of fragments from other organelles that could contain vanillin biosynthetic enzymes. The chloroplastic localization of *VpVAN* has been demonstrated, but the localization of the other enzymes of the vanillin biosynthetic pathway has not (Gallage et al., 2018). We propose that cytosolic and ER-anchored enzymes could have been co-isolated with the chloroplasts studied by Gallage and colleagues. Furthermore, it's possible that those extra-chloroplastic enzymes could be implicated or even necessary to complete the conversion of L-phenylalanine into vanillin (Figure 4.1b). Based on the different investigations presented herein, we suggest that the localization of the vanillin biosynthetic machinery needs re-evaluation.

References

- Achnine, L., Blancaflor, E.B., Rasmussen, S. and Dixon, R.A. (2004) Colocalization of L-phenylalanine ammonia-lyase and cinnamate 4-hydroxylase for metabolic channeling in phenylpropanoid biosynthesis. *The Plant Cell* **16**, 3098-3109.
- Andersson, M.X., Goksör, M. and Sandelius, A.S. (2007a) Membrane contact sites: physical attachment between chloroplasts and endoplasmic reticulum revealed by optical manipulation. *Plant signaling & behavior* **2**, 185-187.
- Andersson, M.X., Goksör, M. and Sandelius, A.S. (2007b) Optical manipulation reveals strong attracting forces at membrane contact sites between endoplasmic reticulum and chloroplasts. *The Journal of biological chemistry* **282**, 1170-1174.
- Arya, S.S., Mahto, B.K., Sengar, M.S., Rookes, J.E., Cahill, D.M. and Lenka, S.K. (2022) Metabolic Engineering of Rice Cells with Vanillin Synthase Gene (VpVAN) to Produce Vanillin. *Molecular biotechnology* **64**, 861-872.
- Barros, J., Escamilla-Trevino, L., Song, L., Rao, X., Serrani-Yarce, J.C., Palacios, M.D., Engle, N., Choudhury, F.K., Tschaplinski, T.J., Venables, B.J., Mittler, R. and Dixon, R.A. (2019) 4-Coumarate 3-hydroxylase in the lignin biosynthesis pathway is a cytosolic ascorbate peroxidase. *Nature communications* **10**, 1994.
- Bassard, J.E., Mutterer, J., Duval, F. and Werck-Reichhart, D. (2012a) A novel method for monitoring the localization of cytochromes P450 and other endoplasmic reticulum membrane associated proteins: a tool for investigating the formation of metabolons. *The FEBS journal* **279**, 1576-1583.
- Bassard, J.E., Richert, L., Geerinck, J., Renault, H., Duval, F., Ullmann, P., Schmitt, M., Meyer, E., Mutterer, J., Boerjan, W., De Jaeger, G., Mely, Y., Goossens, A. and Werck-Reichhart, D. (2012b) Protein-protein and protein-membrane associations in the lignin pathway. *The Plant Cell* **24**, 4465-4482.
- Block, M.A. and Jouhet, J. (2015) Lipid trafficking at endoplasmic reticulum-chloroplast membrane contact sites. *Curr Opin Cell Biol* **35**, 21-29.
- Brillouet, J.M., Verdeil, J.L., Odoux, E., Lartaud, M., Grisoni, M. and Conéjéro, G. (2014) Phenol homeostasis is ensured in vanilla fruit by storage under solid form in a new chloroplast-derived organelle, the phenyloplast. *Journal of experimental botany* **65**, 2427-2435.
- Chee, M.J., Lycett, G.W., Khoo, T.J. and Chin, C.F. (2017) Bioengineering of the Plant Culture of *Capsicum frutescens* with Vanillin Synthase Gene for the Production of Vanillin. *Molecular biotechnology* **59**, 1-8.
- Christopher, J.A., Stadler, C., Martin, C.E., Morgenstern, M., Pan, Y., Betsinger, C.N., et al. (2021). Subcellular proteomics. *Nat Rev Methods Primers* **1**.

- Czichi, U. and Kindl, H. (1975) Formation of p-coumaric acid and o-coumaric acid from L-phenylalanine by microsomal membrane fractions from potato: Evidence of membrane-bound enzyme complexes. *Planta* **125**, 115-125.
- Czichi, U. and Kindl, H. (1977) Phenylalanine ammonia lyase and cinnamic acid hydroxylases as assembled consecutive enzymes on microsomal membranes of cucumber cotyledons: Cooperation and subcellular distribution. *Planta* **134**, 133-143.
- Dixon, R.A. (2010). "Vanillin Biosynthesis – Not as Simple as it Seems?" in *Handbook of Vanilla Science and Technology*. Wiley-Blackwell), 292-298.
- Dixon, R.A. (2018). "Vanillin Biosynthesis – Not as Simple as it Seems?" in *Handbook of Vanilla Science and Technology, Second Edition* (eds D. Havkin-Frenkel and F.C. Belanger). Wiley-Blackwell, 435-445.
- Gallage, N.J., Hansen, E.H., Kannangara, R., Olsen, C.E., Motawia, M.S., Jørgensen, K., Holme, I., Hebelstrup, K., Grisoni, M. and Møller, B.L. (2014) Vanillin formation from ferulic acid in *Vanilla planifolia* is catalysed by a single enzyme. *Nature communications* **5**, 4037.
- Gallage, N.J., Jørgensen, K., Janfelt, C., Nielsen, A.J.Z., Naake, T., Dunski, E., Dalsten, L., Grisoni, M. and Møller, B.L. (2018) The Intracellular Localization of the Vanillin Biosynthetic Machinery in Pods of *Vanilla planifolia*. *Plant & cell physiology* **59**, 304-318.
- Gallage, N. J., and Møller, B. L. (2018). "Vanilla: The Most Popular Flavour" in *Biotechnology of Natural Products*, eds. W. Schwab, B. M. Lange, and M. Wüst (Cham: Springer International Publishing), 3-24. doi:10.1007/978-3-319-67903-7_1.
- Havkin-Frenkel, D., and Podstolski, A. (2007). *Vanillin production. USA patent, US2007/0231864 A1*. USA patent application.
- Havkin-Frenkel, D., Zylstra, G., Frenkel, C., and Belanger, F. (2006). *Production of vanillin in microbial cells. USA patent, US20060172402 A1*. USA patent application.
- Kaneko, Y. and Keegstra, K. (1996) Plastid biogenesis in embryonic pea leaf cells during early germination. *Protoplasma* **195**, 59-67.
- Khoyratty, S., Kodja, H. and Verpoorte, R. (2018) Vanilla flavor production methods: A review. *Industrial Crops and Products* **125**, 433-442.
- Kundu, A. (2017) Vanillin biosynthetic pathways in plants. *Planta* **245**, 1069-1078.
- Liu, L. and Li, J. (2019) Communications Between the Endoplasmic Reticulum and Other Organelles During Abiotic Stress Response in Plants. *Front Plant Sci* **10**, 749.

- McLean, B., Whatley, J.M. and Juniper, B.E. (1988) Continuity of chloroplast and endoplasmic reticulum membranes in *Chara* and *Equisetum*. *New Phytologist* **109**, 59-65.
- Møller, B. L., and Laursen, T. (2021). "Chapter Eight - Metabolons and bio-condensates: The essence of plant plasticity and the key elements in development of green production systems," in *Advances in Botanical Research*, eds. R. Verpoorte, G.-J. Witkamp, and Y. H. Choi (Academic Press), 185-223. doi:10.1016/bs.abr.2020.09.006.
- Negi, J., Munemasa, S., Song, B., Tadakuma, R., Fujita, M., Azoulay-Shemer, T., et al. (2018). Eukaryotic lipid metabolic pathway is essential for functional chloroplasts and CO₂ and light responses in *Arabidopsis* guard cells. *Proc Natl Acad Sci USA* **115**(36), 9038-9043.
- Negishi, O. and Negishi, Y. (2017) Phenylpropanoid 2,3-dioxygenase involved in the cleavage of the ferulic acid side chain to form vanillin and glyoxylic acid in *Vanilla planifolia*. *Biosci Biotechnol Biochem* **81**, 1732-1740.
- Negishi, O., Sugiura, K. and Negishi, Y. (2009) Biosynthesis of vanillin via ferulic acid in *Vanilla planifolia*. *Journal of agricultural and food chemistry* **57**, 9956-9961.
- Nishizawa, A.N., Wolosiuk, R.A. and Buchanan, B.B. (1979) Chloroplast phenylalanine ammonia-lyase from spinach leaves: Evidence for light-mediated regulation via the ferredoxin/Thioredoxin system. *Planta* **145**, 7-12.
- Podstolski, A., Havkin-Frenkel, D., Malinowski, J., Blount, J.W., Kourteva, G. and Dixon, R.A. (2002) Unusual 4-hydroxybenzaldehyde synthase activity from tissue cultures of the vanilla orchid *Vanilla planifolia*. *Phytochemistry* **61**, 611-620.
- Rasmussen, S. and Dixon, R.A. (1999) Transgene-mediated and elicitor-induced perturbation of metabolic channeling at the entry point into the phenylpropanoid pathway. *The Plant Cell* **11**, 1537-1552.
- Satô, M. (1966) Metabolism of phenolic substances by the chloroplasts-II.: Conversion by the isolated chloroplasts of p-coumaric acid to caffeic acid. *Phytochemistry* **5**, 385-389.
- Saunders, J.A. and McClure, J.W. (1975) Phytochrome controlled phenylalanine ammonia lyase in *Hordeum vulgare* plastids. *Phytochemistry* **14**, 1285-1289.
- Sinha, A.K., Sharma, U.K. and Sharma, N. (2008) A comprehensive review on vanilla flavor: extraction, isolation and quantification of vanillin and others constituents. *International journal of food sciences and nutrition* **59**, 299-326.
- Wang, J.Z. and Dehesh, K. (2018) ER: the Silk Road of interorganellar communication. *Current opinion in plant biology* **45**, 171-177.

- Whatley, J.M., McLean, B. and Juniper, B.E. (1991) Continuity of chloroplast and endoplasmic reticulum membranes in *Phaseolus vulgaris*. *New Phytologist* **117**, 209-217.
- Yang, H., Barros-Rios, J., Kourteva, G., Rao, X., Chen, F., Shen, H., Liu, C., Podstolski, A., Belanger, F., Havkin-Frenkel, D. and Dixon, R.A. (2017) A re-evaluation of the final step of vanillin biosynthesis in the orchid *Vanilla planifolia*. *Phytochemistry* **139**, 33-46.

CHAPITRE V

DISCUSSION, CONCLUSION ET PERSPECTIVES

5.1 Discussion

5.1.1 Chapitre II – Reproduction de la voie de biosynthèse de la vanilline

Les méthodes d'extraction et d'analyse au HPLC-MS/MS réalisées au chapitre II n'ont pas permis d'analyser la production de la vanilline ou de ses intermédiaires puisqu'aucun signal significatif n'a été détecté pour les composés d'intérêts dans les échantillons. La méthode d'extraction devra donc se faire à partir d'une plus grande biomasse afin d'augmenter la quantité de composés d'intérêts présents dans les extraits à analyser au HPLC-MS/MS. Il est aussi possible que le problème de la méthode provienne d'un faible rendement d'extraction des composés d'intérêts hors des cellules. De plus, il faudra choisir une nouvelle méthode d'extraction en phase solide pour l'analyse de l'acide caféique, le pH basique utilisé dans la méthode d'extraction a possiblement causé l'oxydation de la molécule puisqu'un faible taux de récupération a été obtenu (Chapitre II – Tableau 2.4). Si les pigments sont encore problématiques en saturant les colonnes d'extraction en phase solide, il faudra ajouter une étape pour les éliminer avant d'effectuer l'extraction en phase solide ou utiliser une méthode qui n'extrait pas ces molécules.

En ce qui concerne la détection des protéines par immunobuvardage de type western présentée au chapitre II, trois possibilités peuvent expliquer l'absence de détection des protéines avec une étiquette poly-histidine. Premièrement, les protéines peuvent être présentes dans les extraits, mais sous le seuil de détection de l'immunobuvardage de type western anti-his. Dans ce cas-ci, il faudrait par exemple concentrer les protéines en utilisant l'étiquette poly-histidine avant d'effectuer la méthode de détection. Deuxièmement, les protéines peuvent ne pas avoir été produites par les clones au moment où la biomasse cellulaire a été récoltée. Remplacer les protéines contenant l'étiquette

poly-histidine par une protéine fluorescente permettrait au moins de confirmer que la production des protéines est possible au temps choisi pour la récolte. Troisièmement, il est possible que les protéines ne s'accumulent pas dans les cellules, par exemple si elles sont dégradées après leur production.

En ce qui concerne l'hypothèse concernant l'absence de production des protéines d'intérêts au moment de la récolte des cellules, il est important de noter que les promoteurs clonés dans p0521s-V (Chapitre II – Figure 2.2B) n'ont été que partiellement caractérisés. Pour cette étude, Slattery et al. (2018) ont sélectionné les promoteurs étudiés puisqu'ils avaient le potentiel d'être hautement actifs en fonction des gènes qu'ils contrôlaient chez *P. tricornutum*. Ces promoteurs ont par la suite été clonés individuellement en amont du gène de résistance *Sh ble* dans le plasmide p0521s. Pour chaque promoteur, Slattery et al. (2018) ont aussi évalué l'impact du terminateur sur l'expression du gène *Sh ble* en clonant le promoteur soit en combinaison à son terminateur natif, soit avec le terminateur FcpA. Ces plasmides ont par la suite été transformés dans *P. tricornutum* par conjugaison bactérienne. Pour chaque construction, le nombre de colonies de *P. tricornutum* résistantes à la zéocine a été compté et le niveau d'expression relatif du gène *Sh ble* a été mesuré par RT-PCR quantitatif (qRT-PCR) à partir des cultures liquides en phase de croissance exponentielle. La caractérisation des promoteurs est donc partielle, car ces informations permettent d'estimer le niveau d'expression qu'à la suite de la transformation ou lors de la phase de croissance exponentielle des cultures liquides, et ce, selon leurs conditions de culture. De plus, l'utilisation du gène *Sh ble* peut créer une pression au niveau de l'expression, c'est-à-dire qu'afin de survivre la cellule doit exprimer ce gène. Il est donc possible que l'activité du promoteur diminue en remplaçant le gène *Sh Ble* par un autre gène d'intérêt. Parmi les quatre cassettes d'expression du plasmide p0521s, il n'y a pas eu d'expression détectée pour la cassette DMi1 dans les trois clones analysés (Chapitre II Figure 2.11). Les plasmides ont été séquencés et des mutations ont été détectées au niveau du promoteur H4-1B de cette cassette. En tenant compte des considérations précédentes, il est aussi possible que l'absence d'expression de la cassette DMi1 dans les trois clones soit due à une inactivité du promoteur au stade la culture choisie pour faire l'analyse RT-PCR.

La stabilité d'un épisome propagé pendant quatre mois de culture avait été précédemment démontrée par Slattery et al. (2018), mais les réarrangements dans les épisomes n'avaient pas encore été étudiés. Quant à George et al. (2020), ils ont étudié la possibilité d'intégration de l'épisome dans le génome de *P. tricornutum*. Pour ce faire, ils ont séquencé le génome nucléaire ainsi que les épisomes extraits de divers clones de *P. tricornutum* qui ont été transformés par conjugaison bactérienne. De leur côté, ils n'ont pas détecté d'intégration d'ADN hétérologues dans le génome nucléaire de *P. tricornutum* ou de réarrangement au niveau des épisomes. En ce qui concerne les résultats présentés dans le chapitre II de cette thèse, le séquençage du plasmide extrait du clone Vp-7 #1 (Annexe A – Tableau A.5), démontre qu'un épisome réarrangé peut être propagé par *P. tricornutum*. Dans ce cas-ci le réarrangement consiste à une délétion de 4 094 pb et de trois insertions qui sont des copies partielles de la cassette d'expression DMi1. Des réarrangements avaient été rapportés par Karas et al. (2015) lorsqu'ils ont développé la méthode de conjugaison bactérienne pour la transformation de *P. tricornutum*. Cependant, dans leur cas les réarrangements étaient des chimères entre le plasmide conjugatif p0521s et le plasmide RP4 contenant le matériel génétique nécessaire pour produire les protéines impliquées dans la conjugaison bactérienne. Ils ont réglé ce problème en changeant le plasmide RP4 par un dérivé, c'est-à-dire le plasmide pTA-Mob (Strand et al. 2014) dont la séquence *oriT* avait été retirée. Le séquençage du plasmide extrait du clone Vp-7 #1 (Annexe A – Tableau A.5) démontre que le plasmide conjugatif p0521s-V (Figure 2.2C) peut former avec lui-même des plasmides chimères.

5.1.2 Chapitre III – Instabilité des épisomes chez *P. tricornutum*

À la suite des résultats présentés au chapitre II, la stabilité de l'épisome chez *P. tricornutum* a été étudiée. À ce jour, le manque d'informations à ce sujet ne permet pas de connaître les limites de l'épisome comme outil moléculaire. Tel que mentionné dans le chapitre III, il y a déjà quelques publications qui démontrent que ce système d'expression extrachromosomique peut être fonctionnel (Russo et al. 2018; Sharma et al. 2018; Fabris et al. 2020; George et al. 2020; Moosburner et al. 2020; Slattery et al. 2020). Cependant, il y a un manque de résultats et d'observations concernant ce qui advient aux

épisomes propagés par les clones de *P. tricornutum* lorsque la production des protéines hétérologues ou des métabolites d'intérêt n'est pas stable. Ce type d'informations permettrait à la communauté scientifique d'avoir des exemples à tenir en compte lors de l'analyse des clones de *P. tricornutum* propageant un épisode. Par exemple, Jones (2014) a mis de l'avant les problèmes d'instabilités génétiques observés dans plusieurs cas de modifications génétiques des cyanobactéries. Dans un des cas présentés par l'auteur, une souche de *Synechococcus* sp. PCC 7002 a été transformée pour produire du mannitol, mais certains clones ne pouvaient produire que temporairement ce métabolite (Jacobsen et Frigaard 2014). Lorsque la production de mannitol n'était plus observée, les clones retrouvaient un taux de croissance semblable à la souche sauvage. Ils ont alors analysé un des clones ayant perdu sa capacité de production de mannitol. Le séquençage du gène d'intérêt a révélé la délétion d'un nucléotide causant un changement dans le cadre de lecture du gène et empêchant ainsi la production d'une enzyme fonctionnelle. Dans un autre cas répertorié par Jones (2014), plusieurs gènes ont été introduits par Kusakabe et al. (2013) dans le génome de la souche PCC 7942 de *Synechococcus elongatus* pour produire de l'isopropanol. À la suite de la transformation de la cyanobactérie, Kusakabe et al. (2013) ont vérifié la présence des gènes d'intérêts dans quatre clones en utilisant la technique de la colonie-PCR. Dans trois clones, les fragments amplifiés d'un des gènes étaient plus longs que ce qui était attendu. Dans le quatrième clone, le fragment amplifié par la colonie-PCR était de la bonne taille, mais le séquençage de ce fragment révéla une variation d'un nucléotide causant une substitution non synonyme (G66D). Le changement d'une glycine pour un acide aspartique a eu comme impact de diminuer l'activité catalytique de l'enzyme. Comme mentionné par Jones (2014), ces exemples d'instabilités génétiques ne sont pas des généralités retrouvées dans tous les projets d'ingénierie métabolique des cyanobactéries. Il faut cependant tenir compte que les instabilités génétiques puissent être des obstacles à la réussite de certains projets si ces phénomènes ne sont pas connus ou s'ils sont ignorés lors de l'interprétation des résultats. C'est l'idée que Jones (2014) a mis de l'avant en consacrant une publication pour mettre en lumière les instabilités génétiques des cyanobactéries: « Chances are it will not be reported, nor will the nature of the “instability” be investigated. That is often the fate of negative results; few people are willing to spend resources in chasing up loose ends, which are

unlikely to enhance or result in a publication. Instead, they will try it again or try something different, hoping it will work the next time. » (Jones, 2014).

Dans le même ordre d'idée que Jones (2014), Poust et al. (2014) ont même suggéré que des banques de données publiques devraient être créées afin de rendre disponibles les résultats non publiés. Dans une publication sur l'étude de la production d'une polycétide synthase par ingénierie génétique de microorganisme, Poust et al. (2014) ont indiqué que la majorité des publications à ce sujet ne comprenaient pas les résultats considérés comme des échecs. Ce manque d'informations cruciales limite le développement des connaissances. Ainsi, des algorithmes pourraient extraire des informations à partir de ces compilations de données dans le but d'optimiser la conception de projet de recherche en lien avec la production de polycétide synthase. Des banques de données de ce genre devraient être créées afin de rassembler les données en lien avec les épisomes transformés dans les diatomées. Ces informations pourraient servir d'avertissement au moment de la conception des projets d'ingénierie métabolique concernant les potentielles complications en lien avec les épisomes. De plus, ces banques de données pourraient servir de références afin de proposer des explications concernant les résultats négatifs.

Ces types d'observations en lien avec l'instabilité des épisomes n'avaient pas encore été rapportés chez *P. tricornutum*. Les résultats de séquençage des clones analysés au chapitre III (Figure 3.2 et Annexe B – Table B.3) ont démontré que des instabilités génétiques pouvaient aussi être observées au niveau des épisomes propagés par *P. tricornutum*. En transformant la diatomée par conjugaison bactérienne avec le plasmide pDMi7, deux groupes de clones ont été obtenus. Il y avait ceux dont la fluorescence de la protéine rapportrice eGFP a pu être détectée et ceux dont ce ne fut pas le cas. Pour les six clones fluorescents (eGFP+) analysés, soit DMi7-21-1, DMi7-21-3, DMi7-31-1, DMi7-31-3, DMi7-31-4 et DMi7-31-8, les mutations ont eu un impact sur la production de l'enzyme fcs. L'identification de ce qui est à l'origine de ces mutations permettrait de limiter ces phénomènes et d'améliorer le processus de modifications génétiques de *P. tricornutum*. En ce qui concerne les quatre clones étudiés dont la fluorescence n'a pas été observée par microscopie à fluorescence ou par cytométrie en flux (eGFP-),

c'est-à-dire DMi7-21-14, DMi7-21-15, DMi7-21-15 et DMi7-31-10, la séquence du gène *eGFP-T2A-ech-fcs* n'a pas été modifiée. Les mutations ne peuvent donc pas expliquer l'impossibilité de détecter la fluorescence dans ces clones. Il est possible que la cassette d'expression soit régulée au niveau post-transcriptionnel empêchant ainsi la production de la protéine d'intérêt. Des investigations à ce sujet permettraient de confirmer si cela est possible et par la suite d'identifier et de neutraliser ces mécanismes de régulations ou du moins de limiter leur effet.

Par exemple, l'algue unicellulaire verte *C. reinhardtii* réprime fréquemment l'expression des transgènes qui ne sont pas soumis à une pression de sélection (Schroda 2019). Pour augmenter l'expression des gènes hétérologues, des souches de *C. reinhardtii* ont été soumises à des expériences de mutagenèse par exposition à des rayonnements ultraviolets (Neupert, Karcher, et Bock 2009; Kurniasih et al. 2016). Ces expériences ont permis d'augmenter l'expression des gènes hétérologues. De leur côté, Neupert, Karcher et Bock (2009) ont identifié deux clones, nommés UVM4 et UVM11, dont la résistance à l'émétine avait été augmentée par rapport aux clones qui n'avaient pas été exposés à l'agent mutagène. L'analyse des mutations produites dans les clones UVM4 et UVM11 a permis d'identifier qu'un gène prétendu de la famille Sir2 des histones désacétylases serait à l'origine de l'augmentation de l'expression des gènes hétérologues (Neupert et al. 2020). Kurniasih et al. (2016) ont quant à eux sélectionné des clones dont les mutations aléatoires causées par les rayonnements ultraviolets permettaient de survivre dans de plus grandes concentrations de zéocine.

Au chapitre II, l'immunobuvardage de type western n'avait pas permis de confirmer la production des protéines hétérologues marquées avec une étiquette poly-histidine par les clones de *P. tricornutum* V1-3 #2, Vc-1 #1 et Vp-7 #1 (Chapitre II – Figure 2.10). Au chapitre III, la détection de la eGFP par cette même technique a été réalisée sur des cultures des clones DMi7-21-1, DMi7-21-3 et DMi7-21-7 produisant de la fluorescence (Chapitre III – Figure 3.6). Pour chacun des clones, des cultures enrichies en cellules eGFP⁺ ou eGFP⁻ par un tri cellulaire induit par fluorescence (FACS) ont aussi été analysées. Malgré la présence de cellules eGFP⁺ dans les cultures initiales des clones

DMi7-21-3 et DMi7-21-7, la détection de la eGFP n'a été obtenue que dans ces cultures enrichies en cellules eGFP⁺. L'analyse par cytométrie en flux a démontré que les cultures initiales de DMi7-21-3 et DMi7-21-7 émettaient de la fluorescence eGFP, mais que le pourcentage de cellules eGFP⁺ étaient inférieur à 10 % (Chapitre III – Figure 3.6b-c). Cependant, le criblage des clones DMi7-21-3 et DMi7-21-7 par la technique d'immunobuvardage de type western n'aurait pas permis de confirmer que ces clones peuvent produire la eGFP. Donc, il est possible d'appliquer cette observation aux cultures des clones V1-3 #2, Vc-1 #1 et Vp-7 #1 analysés au chapitre II. La technique d'immunobuvardage de type western n'était peut-être pas assez sensible pour détecter la production des protéines marquées avec une étiquette poly-histidine. Puisque ces clones ne produisent pas de protéines fluorescentes, il n'est pas possible d'enrichir par FACS ces cultures en cellules produisant les protéines d'intérêts. Dans ce cas-ci, il faudrait purifier et concentrer les protéines marquées avec une étiquette poly-histidine à partir des extraits de protéines totales avant d'effectuer l'immunobuvardage de type western anti-his.

5.1.3 Chapitre IV – Réévaluation nécessaire pour la localisation intracellulaire de la voie de biosynthèse de la vanilline chez *V. planifolia*

Gallage et al. (2018) ont conclu que le chloroplaste à lui seul était suffisant pour convertir la L-phénylalanine en vanilline chez *V. planifolia*. L'analyse des conclusions de Gallage et al. (2018) ainsi que de la littérature au sujet du lien entre le chloroplaste et le réticulum endoplasmique permet d'émettre l'hypothèse que le chloroplaste n'est pas suffisant pour effectuer cette biosynthèse. Il est probable que des enzymes soient localisées dans le cytoplasme et que d'autres soient ancrées dans le réticulum endoplasmique. Cette hypothèse présentée au chapitre IV met de l'avant l'idée qu'il faudrait réévaluer la localisation intracellulaire de la machinerie enzymatique permettant de faire la biosynthèse de la vanilline à partir de la L-phénylalanine. Puisque les travaux de cette thèse ne permettaient pas d'étudier cette hypothèse, la publication a été réalisée dans le but de la partager avec la communauté scientifique. L'investigation de cette hypothèse pourrait permettre de faire progresser l'état des connaissances sur la biosynthèse de la vanilline chez *V. planifolia*.

Les résultats rapportés par Gallage et al. (2018) démontrent cependant que l'enzyme *VpVAN* est accumulée dans le chloroplaste. Si une nouvelle tentative devait être réalisée pour produire cette enzyme par *P. tricornutum*, il faudrait s'assurer que la protéine *VpVAN* soit accumulée dans le chloroplaste. Pour ce faire, il faudrait modifier le peptide signal de *VpVAN* afin de s'assurer qu'il contienne la séquence consensus nécessaire pour transporter la protéine dans le chloroplaste de *P. tricornutum* (Kilian et Kroth 2005; Gruber et al. 2007). Il est aussi possible de transporter *VpVAN* dans le chloroplaste en concevant une protéine de fusion avec une protéine connue pour s'accumuler dans le chloroplaste (Liu et al. 2016).

5.2 Conclusion

L'épisome est un outil moléculaire pour *P. tricornutum* dont l'utilité a déjà été démontrée par quelques groupes de recherche. Cependant, les exemples démontrant les limitations de l'épisome telles que les instabilités en fonction des cassettes d'intérêts ne sont pas disponibles dans la littérature au meilleur de mes connaissances. En ce qui concerne la détection de la vanilline produite par les clones de *P. tricornutum*, elle n'a pas été possible avec les diverses méthodes développées pour la réalisation de cette thèse. L'écriture de celle-ci s'est donc concentrée sur les réarrangements de l'épisome et leurs impacts. Cette thèse permettra de rendre disponibles à la communauté scientifique les résultats concernant les possibles limitations de l'épisome comme outil moléculaire chez *P. tricornutum* dans le but de mieux planifier son utilisation.

Si le projet de production de vanilline par *P. tricornutum* devait se poursuivre, la priorité serait évidemment de quantifier la vanilline produite par les clones propageant les plasmides p0521s-V et pDMi7. Il est possible que la quantification de la vanilline démontre que ces deux plasmides ne permettent pas de produire de la vanilline. Si tel est le cas, il faudrait envisager des stratégies pour limiter les réarrangements et les potentielles régulations au niveau post-transcriptionnel observées chez les différents clones. Les enzymes ech et fcs du plasmide pDMi7 avaient été sélectionnés pour remplacer l'enzyme *VpVAN* dans la voie de biosynthèse de la vanilline. Cette décision était basée

sur la littérature qui suggérait que cette enzyme ne possède peut-être pas l'activité enzymatique de conversion de l'acide férulique en vanilline. Les enzymes ech et fcs pourraient être aussi remplacées par d'autres enzymes pour poursuivre ce projet de recherche.

5.3 Perspectives

5.3.1 Extraction des composés phénoliques

Puisqu'il n'y a pas eu de système de criblage développer pour la production de vanilline dans ce projet de thèse, la méthode de macération avait été sélectionnée pour extraire les composés phénoliques de *P. tricornutum*. Cette technique permettait de traiter plusieurs échantillons en même temps sans équipement spécialisé. De plus, Rico et al. (2013) ont utilisé la macération pour extraire les composés phénoliques de *P. tricornutum*. Cependant, en utilisant la méthode présentée au chapitre II ainsi que des variantes qui n'ont pas été publiées dans cette thèse, il n'a pas été possible de détecter les mêmes phénylpropanoïdes que Rico et al. (2013) avaient détecté dans leurs extraits. Il est possible que le rendement des extractions par macération fût trop faible. La prochaine étape pour ce projet de thèse serait donc d'évaluer d'autres alternatives pour l'extraction des composés phénoliques. Des revues de littérature sont disponibles à propos des diverses méthodes d'extraction telles que l'extraction assistée par micro-ondes, par ultrason ou par un fluide supercritique (Alara, Abdurahman, et Ukaegbu 2021; Chanioti, Katsouli, et Tzia 2021). D'autres solvants d'extractions peuvent aussi être évalués, tels que divers solvants organiques ou les solvants eutectiques profonds. Ces derniers sont des mélanges de deux ou plusieurs composés qui ensemble ont une température de fusion plus faible que chacun des composés séparés (Fernández et al. 2018). Certains solvants eutectiques profonds sont nommés naturel. C'est le cas par exemple des mélanges aqueux d'un sel tel que chlorure de choline et d'un glucide comme le glucose (Choi et al. 2011). Ces solvants naturels eutectiques profonds peuvent stabiliser des composés qui se dégradent à la lumière s'ils sont conservés dans des solvants organiques tels que le méthanol (Jeliński, Przybyłek, et Cysewski 2019; Fan et al. 2021). Fan et al. (2021) ont aussi démontré que

la lutéine pouvait être extraite de l'algue unicellulaire verte *Scenedesmus* sp. en utilisant un mélange de fenchol et de thymol. La lutéine extraite de cette façon avait une plus grande résistance à la chaleur et une meilleure stabilité d'entreposage dans ce mélange qu'avec le méthanol. De plus, les solvants naturels eutectiques profonds peuvent avoir un meilleur rendement d'extraction que les solvants organiques. Par exemple, comparativement à une solution d'éthanol 70 %, le mélange de chlorure de choline et d'acide citrique permettait d'extraire une plus grande quantité de composés phénoliques à partir du grignon d'olive, un sous-produit de l'extraction de l'huile d'olive (Chanioti et Tzia 2018).

5.3.2 Évaluation du profil de l'activité des promoteurs

Comme mentionné précédemment, il y a peu d'informations disponible sur les promoteurs qui ont été sélectionnés pour construire les plasmides p0521s-V et pDMi7. L'utilisation d'une protéine fluorescente telle que la GFP permettrait de quantifier la fluorescence produite au cours du temps pour chaque promoteur selon le stade de la culture ou des conditions de culture à tester. Ce genre d'information permettrait d'améliorer le processus de conception des cassettes d'expression. Par exemple, une étude a été réalisée pour comparer l'activité de différents promoteurs transformés dans *P. tricornutum* tels que le promoteur *fcpA* et le promoteur de la glutamine synthétase (GLNA) (Erdene-Ochir et al. 2016). Pour ce faire, Erdene-Ochir et al. (2016) ont quantifié dans le temps la fluorescence relative de la GFP dont le gène était sous le contrôle de ces promoteurs. Ils ont ainsi mesuré que sous le contrôle du promoteur *fcpA*, l'intensité de la fluorescence émise par la GFP était maximale au 6^e jour de croissance (Figure 5.1A). Au même moment, l'intensité de la GFP était environ deux à trois fois moins élevée lorsque son expression se retrouvait sous le contrôle du promoteur de la GLNA (Figure 5.1B). Cependant, au 11^e jour de culture, la fluorescence relative mesurée était 4 à 5 fois plus élevée pour le promoteur de la GLNA comparativement au promoteur *fcpA* (Figure 5.1). Pour le projet de production de vanilline, ce genre de connaissance pourrait être combiné à l'analyse du métabolome de *P. tricornutum* dans le temps. En sachant à

quel moment la production d'acide férulique par la diatomée est à son maximum, il serait possible de choisir un promoteur dont l'activité est très forte au même moment.

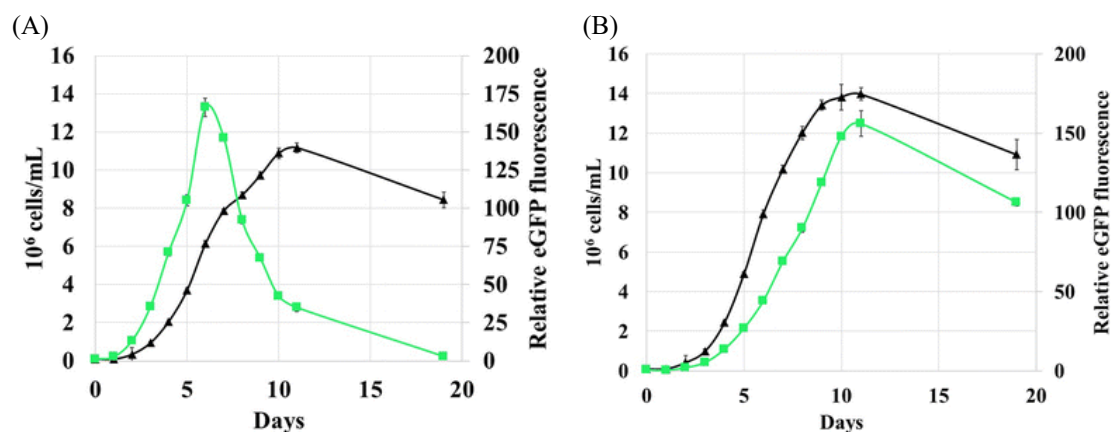


Figure 5.1 Production par *P. tricornutum* de la fluorescence eGFP sous le contrôle des promoteurs *fcpA* et GLNA.

La courbe de croissance des cultures est représentée en noir et la fluorescence relative de la eGFP est illustrée en vert. Les clones ont été transformés avec le gène de la eGFP sous le le contrôle des promoteurs (A) *fcpA* ou (B) de la GLNA (les 501 nucléotides en amont du gène codant pour la GLNA).

Source de l'image : Erdene-Ochir et al. (2016).

5.3.3 L'expression inductible des gènes d'intérêts

Il est possible que l'instabilité des épisomes ou la régulation potentielle au niveau post-transcriptionnel soient induites en réponse à la production de la protéine eGFP-T2A-ech-fcs du plasmide pDMi7. Pour limiter ces processus et permettre la production de cette protéine, il faudrait évaluer l'utilisation d'un promoteur inductible. Comme mentionné au chapitre I, quelques promoteurs inductibles sont utilisés chez *P. tricornutum*, mais certains peuvent avoir une expression basale (Niu et al. 2012; Chu et al. 2016; Adler-Agnon (Shemesh) et al. 2018). Il faudrait évaluer si cette faible expression peut être suffisante pour induire les processus d'instabilités génétiques ou de répression des gènes. Si c'est le cas, il faut s'assurer d'utiliser des promoteurs complètement inactifs en absence d'inducteur. Il sera peut-être nécessaire de développer de nouveaux promoteurs inductibles si ceux disponibles dans la littérature ne permettent pas le contrôle nécessaire de l'expression.

Une autre option pourrait être de modifier génétiquement une souche de *P. tricornutum* pour lui permettre de croître en hétérotrophie à la noirceur. Ainsi il serait possible d'utiliser un promoteur qui est déjà caractérisé comme étant inductible par la lumière, tel que le promoteur FcpB, et qui, selon les résultats publiés par Siaut et al. (2007), a le potentiel d'être complètement inactif à la noirceur. Pour ce faire, il faudrait tout d'abord transformer *P. tricornutum* avec le gène d'un transporteur de glucose. Cette diatomée pourrait ainsi croître à la noirceur dans un milieu contenant du glucose (Zaslavskaja et al. 2000; Hamilton et al. 2016). La diatomée obtenue pourrait par la suite être transformée avec le gène *eGFP-T2A-ech-fcs*, c'est-à-dire le gène d'intérêt utilisé dans le plasmide pDMi7, sous le contrôle du promoteur FcpB. L'induction de l'expression se ferait simplement en transférant la culture à la lumière. Cette méthode aurait cependant comme inconvénient de nécessiter un apport externe en carbone organique. La production *de novo* de la vanilline sans ajout de carbone organique au milieu de culture était cependant un des avantages pour l'utilisation des microorganismes photosynthétiques.

5.3.4 Gènes alternatifs pour la conversion de l'acide férulique en vanilline

La première voie de biosynthèse de la vanilline étudiée au chapitre II comprenait l'enzyme *VpVAN* qui devait catalyser la conversion de l'acide férulique en vanilline. L'article de type opinion (Chapitre IV) inséré dans cette thèse relate entre autres la controverse liée à l'activité enzymatique de *VpVAN*. Pour l'instant, deux groupes de recherche ont pu insérer le gène codant pour *VpVAN* dans des cellules de plantes et ainsi détecter une augmentation de la production de vanilline (Chee et al. 2017; Arya et al. 2022). Des efforts supplémentaires en recherche pourraient être nécessaires pour caractériser cette enzyme dans le but de la rendre fonctionnelle dans d'autres types d'organismes.

Tel que présenté au chapitre III, il a été tenté pour les travaux de cette thèse de remplacer l'enzyme *VpVAN* par une combinaison des enzymes *fcs* et *ech*. Dans ce cas-ci, la conversion de l'acide férulique en feruloyl-CoA par l'enzyme *fcs* nécessite de l'ATP et de la CoA (Yang et al. 2013). Il est possible que la consommation de ces deux métabolites

cause une charge importante sur le métabolisme de *P. tricornutum*. Le débalancement du métabolisme peut avoir un impact négatif sur le taux de croissance et de mutation du microorganisme transformé (Wu et al. 2016). La modification génétique de la diatomée dans le but d'optimiser la régénération de la CoA pourrait être une piste à étudier pour améliorer les biosynthèses nécessitant ce coenzyme (Lee et al. 2009; Su et al. 2022).

Le remplacement de l'enzyme *VpVAN* ou de la combinaison des enzymes *fcs* et *ech* pourrait être une des alternatives à tenter pour la production de vanilline chez *P. tricornutum*. Entre autres, la production de vanilline peut aussi se faire à partir de l'acide protocatéchique via l'activité des enzymes S-COMT et CAR (Chapitre I – Figure 1.5). Cette voie de biosynthèse peut aussi être problématique pour le métabolisme de *P. tricornutum*, car l'enzyme CAR nécessite comme coenzyme l'ATP et le NADPH. Quant à elle, l'enzyme S-COMT catalyse la réaction de méthylation en consommant une molécule de SAM. Il est donc possible que la charge métabolique de cette voie de biosynthèse ait aussi un impact négatif sur la production de vanilline et sur la stabilité génétique. Comme autre possibilité, une enzyme isolée de la plante *V. planifolia* par Negishi et Negishi (2017), c'est-à-dire la PPDiox, peut catalyser la conversion de l'acide férulique en vanilline. Cependant, des publications indépendantes concernant l'étude de cette enzyme ne sont pas encore disponibles. Il est donc possible que cette enzyme nécessite d'autres investigations avant de pouvoir permettre la production de vanilline dans des projets d'ingénieries métaboliques chez les diatomées.

5.3.5 Étude sur la possibilité d'un mécanisme de répression des gènes

Les clones eGFP- analysés au chapitre III, soit DMi7-21-14, DMi7-21-15 et DMi7-21-16, ne produisent pas de fluorescence verte (Annexe B – Figure B.5), mais ils expriment tout de même le gène d'intérêt codant pour la protéine eGFP-T2A-*ech-fcs* (Chapitre III – Figure 3.3). Il est alors possible pour ces clones qu'un mécanisme de répression des gènes inhibe la production de la protéine d'intérêt. En déterminant si c'est bien le cas et en étudiant ce mécanisme, il pourrait être possible

de développer des techniques pour obtenir des clones de *P. tricornutum* produisant la protéine d'intérêt.

Une revue de la littérature concernant la répression des gènes chez *P. tricornutum* démontre qu'il est possible de réduire artificiellement l'expression de gènes ciblés. Par exemple, Kaur et Spillane (2015) ont effectué une preuve de concept en diminuant la quantité de caroténoïdes accumulés par cette diatomée. Pour ce faire, ils ont utilisé des micro-ARNs artificiels ciblant le gène de la phytoène synthase qui est impliqué au niveau de la biosynthèse de ces pigments. Quant à Zhu et al. (2016), ils ont diminué l'activité de la UGPase en transformant *P. tricornutum* avec des séquences anti-sens et des répétitions inversées du gène codant pour cette enzyme. De leur côté, Pan et al. (2017) ont étudié chez la diatomée le métabolisme du propionyl-CoA en induisant une répression par des séquences anti-sens de deux gènes codant respectivement pour les sous-unités α et β de la propionyl-CoA carboxylase. La stratégie des ARN interférant a aussi été utilisée chez *P. tricornutum* lors de l'étude du métabolisme des stéroïdes (Fabris et al. 2014), la biosynthèse des diatoxanthines (Lavaud et al. 2012), la production *de novo* des pyrimidines (Sakaguchi et al. 2011), la fixation du carbone en C4 lors de la photosynthèse (Haimovich-Dayana et al. 2013), la régulation de la division cellulaire (Huysman et al. 2013) et la régulation de la réponse à la lumière (Bailleul et al. 2010; Juhas et al. 2014; Schellenberger Costa et al. 2013).

Même si l'expression des gènes peut être artificiellement réduite, ce mécanisme endogène de régulation n'est pas complètement élucidé. L'analyse des petits ARN interférents (pARNi) chez *P. tricornutum* a démontré une corrélation négative entre l'expression des éléments transposables autonomes (ETA) et leur couverture par les pARNi (Rogato et al. 2014). De plus, Rogato et al. (2014) ont étudié le profil de méthylation des ETA ciblés par les pARNi et ils ont ainsi observé que les ETA faiblement exprimés avaient un haut taux de méthylation. Rogato et al. (2014) ont donc proposés que le profil d'expression des ETA pourrait être lié à leur niveau de méthylation. Du même point de vue, De Riso et al. (2009) ont démontré, en utilisant des séquences anti-sens et des répétitions de régions inversées du gène d'intérêt, que la répression des transgènes

était associée à la méthylation *de novo* des cytosines. La corrélation entre le niveau d'expression d'un gène et son taux de méthylation a aussi été couverte par Veluchamy et al. (2013). Il serait donc possible de retrouver chez *P. tricornutum* un mécanisme de répression des transgènes situés sur l'épisome qui serait contrôlé par la méthylation de l'ADN et les pARNi.

L'état des connaissances actuelles ne permet pas de déterminer si des histones peuvent s'associer aux épisomes chez *P. tricornutum*. Si c'est le cas, le profil épigénétique des histones est un autre aspect de recherche qui devrait être investigué en lien avec la répression des gènes. En ce sens, Geis et al. (2022) ont étudié les ADNs viraux non intégrés en début d'infection des cellules humaines par le virus de l'immunodéficience humaine (VIH). Ils ont ainsi observé une corrélation négative entre le niveau d'expression des gènes viraux extrachromosomiques et le taux de méthylation des histones H3.3.

La possibilité d'un mécanisme de défense par *P. tricornutum* induisant la répression des transgènes extrachromosomiques devrait être étudiée pour optimiser les futurs travaux d'ingénierie métabolique.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Abe, K., Miyake, K., Nakamura, M., Kojima, K., Ferri, S., Ikebukuro, K. and Sode, K. (2014) Engineering of a green-light inducible gene expression system in *Synechocystis* sp. PCC6803. *Microb Biotechnol* **7**, 177-183.
- Adler-Agnon, Z., Leu, S., Zarka, A., Boussiba, S. and Khozin-Goldberg, I. (2018) Novel promoters for constitutive and inducible expression of transgenes in the diatom *Phaeodactylum tricornutum* under varied nitrate availability. *Journal of Applied Phycology* **30**, 2763-2772.
- Anuradha, K., Shyamala, B.N. and Naidu, M.M. (2013) Vanilla--its science of cultivation, curing, chemistry, and nutraceutical properties. *Critical reviews in food science and nutrition* **53**, 1250-1276.
- Arya, S.S., Mahto, B.K., Sengar, M.S., Rookes, J.E., Cahill, D.M. and Lenka, S.K. (2022) Metabolic Engineering of Rice Cells with Vanillin Synthase Gene (VpVAN) to Produce Vanillin. *Molecular biotechnology* **64**, 861-872.
- Aslam, A., Thomas-Hall, S.R., Mughal, T.A. and Schenk, P.M. (2017) Selection and adaptation of microalgae to growth in 100% unfiltered coal-fired flue gas. *Bioresource technology* **233**, 271-283.
- Ayoub, M., de Camargo, A.C. and Shahidi, F. (2016) Antioxidants and bioactivities of free, esterified and insoluble-bound phenolics from berry seed meals. *Food Chem* **197**, 221-232.
- Bai, H., Zhou, D., Zhang, X., Cao, Y., Xiao, X., Zhang, Y. and Yu, Y. (2021) The responses of *Salmonella enterica* serovar Typhimurium to vanillin in apple juice through global transcriptomics. *International journal of food microbiology* **347**, 109189.
- Bailleul, B., Rogato, A., de Martino, A., Coesel, S., Cardol, P., Bowler, C., Falciatore, A. and Finazzi, G. (2010) An atypical member of the light-harvesting complex stress-related protein family modulates diatom responses to light. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 18214-18219.
- Banerjee, G. and Chattopadhyay, P. (2019) Vanillin biotechnology: the perspectives and future. *Journal of the science of food and agriculture* **99**, 499-506.
- Barghini, P., Di Gioia, D., Fava, F. and Ruzzi, M. (2007) Vanillin production using metabolically engineered *Escherichia coli* under non-growing conditions. *Microbial cell factories* **6**, 13.

- Block, A., Widhalm, J.R., Fatihi, A., Cahoon, R.E., Wamboldt, Y., Elowsky, C., Mackenzie, S.A., Cahoon, E.B., Chapple, C., Dudareva, N. and Basset, G.J. (2014) The Origin and Biosynthesis of the Benzenoid Moiety of Ubiquinone (Coenzyme Q) in Arabidopsis. *The Plant Cell* **26**, 1938-1948.
- Boonchird, C. and Flegel, T.W. (1982) In vitro antifungal activity of eugenol and vanillin against *Candida albicans* and *Cryptococcus neoformans*. *Canadian journal of microbiology* **28**, 1235-1241.
- Brown, Jeremy D., and Martin D. Ryan. 2010. "Ribosome 'Skipping': 'Stop-Carry On' or 'StopGo' Translation." In *Recoding: Expansion of Decoding Rules Enriches Gene Expression*, edited by John F. Atkins and Raymond F. Gesteland, 101-21. New York, NY: Springer New York.
- Brunschwig, C., Rochard, S., Pierrat, A., Rouger, A., Senger-Emonnot, P., George, G. and Raharivelomanana, P. (2016) Volatile composition and sensory properties of *Vanilla × tahitensis* bring new insights for vanilla quality control. *Journal of the science of food and agriculture* **96**, 848-858.
- Buck, J.M., Río Bártulos, C., Gruber, A. and Kroth, P.G. (2018) Blastocidin-S deaminase, a new selection marker for genetic transformation of the diatom *Phaeodactylum tricornutum*. *PeerJ* **6**, e5884.
- Cerón García MC, García Camacho F, Sánchez Mirón A, Fernández Sevilla JM, Chisti Y, Molina Grima E. 2006. "Mixotrophic Production of Marine Microalga *Phaeodactylum Tricornutum* on Various Carbon Sources." *Journal of Microbiology and Biotechnology* **16** (5): 689-94.
- Chakraborty, D., Gupta, G. and Kaur, B. (2016) Metabolic engineering of *E. coli* top 10 for production of vanillin through FA catabolic pathway and bioprocess optimization using RSM. *Protein Expr Purif* **128**, 123-133.
- Chanioti, S. and Tzia, C. (2018) Extraction of phenolic compounds from olive pomace by using natural deep eutectic solvents and innovative extraction techniques. *Innovative Food Science & Emerging Technologies* **48**, 228-239.
- Chee, M.J., Lycett, G.W., Khoo, T.J. and Chin, C.F. (2017) Bioengineering of the Plant Culture of *Capsicum frutescens* with Vanillin Synthase Gene for the Production of Vanillin. *Molecular biotechnology* **59**, 1-8.
- Chen, Y., Gu, F., Li, J., He, S., Xu, F. and Fang, Y. (2015) Involvement of Colonizing *Bacillus* Isolates in Glucovanillin Hydrolysis during the Curing of *Vanilla planifolia* Andrews. *Applied and environmental microbiology* **81**, 4947-4954.
- Chng, J., Wang, T., Nian, R., Lau, A., Hoi, K.M., Ho, S.C., Gagnon, P., Bi, X. and Yang, Y. (2015) Cleavage efficient 2A peptides for high level monoclonal antibody expression in CHO cells. *MAbs* **7**, 403-412.

- Choi, Y.H., van Spronsen, J., Dai, Y., Verberne, M., Hollmann, F., Arends, I.W., Witkamp, G.J. and Verpoorte, R. (2011) Are natural deep eutectic solvents the missing link in understanding cellular metabolism and physiology? *Plant physiology* **156**, 1701-1705.
- Chu, L., Ewe, D., Río Bártulos, C., Kroth, P.G. and Gruber, A. (2016) Rapid induction of GFP expression by the nitrate reductase promoter in the diatom *Phaeodactylum tricornutum*. *PeerJ* **4**, e2344.
- Cilliers, J.J.L. and Singleton, V.L. (1989) Nonenzymic autoxidative phenolic browning reactions in a caffeic acid model system. *Journal of agricultural and food chemistry* **37**, 890-896.
- Costa, C.A.E., Vega-Aguilar, C.A. and Rodrigues, A.E. (2021) Added-Value Chemicals from Lignin Oxidation. *Molecules* **26**.
- Crozet, P., Navarro, F.J., Willmund, F., Mehrshahi, P., Bakowski, K., Lauersen, K.J., Pérez-Pérez, M.E., Auroy, P., Gorchs Rovira, A., Sauret-Gueto, S., Niemeyer, J., Spaniol, B., Theis, J., Trösch, R., Westrich, L.D., Vavitsas, K., Baier, T., Hübner, W., de Carpentier, F., Cassarini, M., Danon, A., Henri, J., Marchand, C.H., de Mia, M., Sarkissian, K., Baulcombe, D.C., Peltier, G., Crespo, J.L., Kruse, O., Jensen, P.E., Schroda, M., Smith, A.G. and Lemaire, S.D. (2018) Birth of a Photosynthetic Chassis: A MoClo Toolkit Enabling Synthetic Biology in the Microalga *Chlamydomonas reinhardtii*. *ACS synthetic biology* **7**, 2074-2086.
- Daniels, R.W., Rossano, A.J., Macleod, G.T. and Ganetzky, B. (2014) Expression of multiple transgenes from a single construct using viral 2A peptides in *Drosophila*. *PLoS one* **9**, e100637.
- De Martino, A., Bartual, A., Willis, A., Meichenin, A., Villazán, B., Maheswari, U. and Bowler, C. (2011) Physiological and molecular evidence that environmental changes elicit morphological interconversion in the model diatom *Phaeodactylum tricornutum*. *Protist* **162**, 462-481.
- De Riso, V., Raniello, R., Maumus, F., Rogato, A., Bowler, C. and Falciatore, A. (2009) Gene silencing in the marine diatom *Phaeodactylum tricornutum*. *Nucleic Acids Res* **37**, e96.
- de Souza, M.P., Bataglioni, G.A., da Silva, F.M.A., de Almeida, R.A., Paz, W.H.P., Nobre, T.A., Marinho, J.V.N., Salvador, M.J., Fidelis, C.H.V., Acho, L.D.R., de Souza, A.D.L., Nunomura, R.C.S., Eberlin, M.N., Lima, E.S. and Koolen, H.H.F. (2016) Phenolic and aroma compositions of pitomba fruit (*Talisia esculenta* Radlk.) assessed by LC-MS/MS and HS-SPME/GC-MS. *Food Research International* **83**, 87-94.

- Defrel, G., Marsaud, N., Rifa, E., Martins, F. and Daboussi, F. (2021) Identification of Loci Enabling Stable and High-Level Heterologous Gene Expression. *Front Bioeng Biotechnol* **9**, 734902.
- Dehghani, J., Adibkia, K., Movafeghi, A., Maleki-Kakelar, H., Saeedi, N. and Omidi, Y. (2020) Towards a new avenue for producing therapeutic proteins: Microalgae as a tempting green biofactory. *Biotechnol Adv* **40**, 107499.
- Demidyuk, I.V., Shubin, A.V., Gasanov, E.V. and Kostrov, S.V. (2010) Propeptides as modulators of functional activity of proteases. *Biomolecular concepts* **1**, 305-322.
- Diner, R.E., Bielinski, V.A., Dupont, C.L., Allen, A.E. and Weyman, P.D. (2016) Refinement of the Diatom Episome Maintenance Sequence and Improvement of Conjugation-Based DNA Delivery Methods. *Front Bioeng Biotechnol* **4**, 65.
- Diner, R.E., Noddings, C.M., Lian, N.C., Kang, A.K., McQuaid, J.B., Jablanovic, J., Espinoza, J.L., Nguyen, N.A., Anzelmann, M.A., Jr., Jansson, J., Bielinski, V.A., Karas, B.J., Dupont, C.L., Allen, A.E. and Weyman, P.D. (2017) Diatom centromeres suggest a mechanism for nuclear DNA acquisition. *Proceedings of the National Academy of Sciences of the United States of America* **114**, E6015-E6024.
- Doron, L., Segal, N. and Shapira, M. (2016) Transgene Expression in Microalgae-From Tools to Applications. *Front Plant Sci* **7**, 505.
- Erdene-Ochir, E., Shin, B.-K., Huda, M.N., Kim, D.H., Lee, E.H., Song, D.-G., Kim, Y.-M., Kim, S.M. and Pan, C.-H. (2016) Cloning of a novel endogenous promoter for foreign gene expression in *Phaeodactylum tricorutum*. *Applied Biological Chemistry* **59**, 861-867.
- Escobar-Muciño, E., Luna-Guevara, M.L., Ramos-Cassellis, M.E., Amador-Espejo, G.G., Castañeda-Lucio, M. and Arenas-Hernández, M.M.P. (2020) Evaluation of process involved in the production of aromatic compounds in Gram-negative bacteria isolated from vanilla (*Vanilla planifolia* ex. Andrews) beans. *J Appl Microbiol* **128**, 1086-1098.
- Fabris, M., Matthijs, M., Carbonelle, S., Moses, T., Pollier, J., Dasseville, R., Baart, G.J.E., Vyverman, W. and Goossens, A. (2014) Tracking the sterol biosynthesis pathway of the diatom *Phaeodactylum tricorutum*. *New Phytol* **204**, 521-535.
- Fabris, M., George, J., Kuzhiumparambil, U., Lawson, C.A., Jaramillo-Madrid, A.C., Abbriano, R.M., Vickers, C.E. and Ralph, P. (2020) Extrachromosomal Genetic Engineering of the Marine Diatom *Phaeodactylum tricorutum* Enables the Heterologous Production of Monoterpenoids. *ACS synthetic biology* **9**, 598-612.
- Fan, C., Liu, Y., Shan, Y. and Cao, X. (2021) A priori design of new natural deep eutectic solvent for lutein recovery from microalgae. *Food Chem* **376**, 131930.

- Fernández, M., Boiteux, J., Espino, M., Gomez, F.J.V. and Silva, M.F. (2018) Natural deep eutectic solvents-mediated extractions: The way forward for sustainable analytical developments. *Anal Chim Acta* **1038**, 1-10.
- Fitzgerald, D.J., Stratford, M., Gasson, M.J., Ueckert, J., Bos, A. and Narbad, A. (2004) Mode of antimicrobial action of vanillin against *Escherichia coli*, *Lactobacillus plantarum* and *Listeria innocua*. *J Appl Microbiol* **97**, 104-113.
- Fitzgerald, D.J., Stratford, M. and Narbad, A. (2003) Analysis of the inhibition of food spoilage yeasts by vanillin. *International journal of food microbiology* **86**, 113-122.
- Fitzgerald, D.J., Stratford, M., Gasson, M.J. and Narbad, A. (2005) Structure-function analysis of the vanillin molecule and its antifungal properties. *Journal of agricultural and food chemistry* **53**, 1769-1775.
- Fleige, C., Hansen, G., Kroll, J. and Steinbüchel, A. (2013) Investigation of the *Amycolatopsis* sp. strain ATCC 39116 vanillin dehydrogenase and its impact on the biotechnical production of vanillin. *Applied and environmental microbiology* **79**, 81-90.
- Foo, S.C., Yusoff, F.M., Ismail, M., Basri, M., Yau, S.K., Khong, N.M.H., Chan, K.W. and Ebrahimi, M. (2017) Antioxidant capacities of fucoxanthin-producing algae as influenced by their carotenoid and phenolic contents. *Journal of biotechnology* **241**, 175-183.
- Franke, R., Humphreys, J.M., Hemm, M.R., Denault, J.W., Ruegger, M.O., Cusumano, J.C. and Chapple, C. (2002) The *Arabidopsis* REF8 gene encodes the 3-hydroxylase of phenylpropanoid metabolism. *The Plant journal: for cell and molecular biology* **30**, 33-45.
- Gallage, N.J., Hansen, E.H., Kannangara, R., Olsen, C.E., Motawia, M.S., Jørgensen, K., Holme, I., Hebelstrup, K., Grisoni, M. and Møller, B.L. (2014) Vanillin formation from ferulic acid in *Vanilla planifolia* is catalysed by a single enzyme. *Nature communications* **5**, 4037.
- Gallage, N.J., Jørgensen, K., Janfelt, C., Nielsen, A.J.Z., Naake, T., Dunski, E., Dalsten, L., Grisoni, M. and Møller, B.L. (2018) The Intracellular Localization of the Vanillin Biosynthetic Machinery in Pods of *Vanilla planifolia*. *Plant & cell physiology* **59**, 304-318.
- Gallage, N. J., and Møller, B. L. (2018). "Vanilla: The Most Popular Flavour" in *Biotechnology of Natural Products*, eds. W. Schwab, B. M. Lange, and M. Wüst (Cham: Springer International Publishing), 3-24. doi:10.1007/978-3-319-67903-7_1.

- Garcia, L.M., Ceccanti, C., Negro, C., De Bellis, L., Incrocci, L., Pardossi, A. and Guidi, L. (2021) Effect of Drying Methods on Phenolic Compounds and Antioxidant Activity of *Urtica dioica* L. Leaves. *Horticulturae* **7**, 10.
- Geis, F.K., Sabo, Y., Chen, X., Li, Y., Lu, C. and Goff, S.P. (2022) CHAF1A/B mediate silencing of unintegrated HIV-1 DNAs early in infection. *Proceedings of the National Academy of Sciences of the United States of America* **119**.
- George, J., Kahlke, T., Abbriano, R.M., Kuzhiumparambil, U., Ralph, P.J. and Fabris, M. (2020) Metabolic Engineering Strategies in Diatoms Reveal Unique Phenotypes and Genetic Configurations With Implications for Algal Genetics and Synthetic Biology. *Front Bioeng Biotechnol* **8**, 513.
- Gruber, A., Vugrinec, S., Hempel, F., Gould, S.B., Maier, U.G. and Kroth, P.G. (2007) Protein targeting into complex diatom plastids: functional characterisation of a specific targeting motif. *Plant Mol Biol* **64**, 519-530.
- Haimovich-Dayan, M., Garfinkel, N., Ewe, D., Marcus, Y., Gruber, A., Wagner, H., Kroth, P.G. and Kaplan, A. (2013) The role of C4 metabolism in the marine diatom *Phaeodactylum tricornutum*. *New Phytol* **197**, 177-185.
- Hamilton, M.L., Powers, S., Napier, J.A. and Sayanova, O. (2016) Heterotrophic Production of Omega-3 Long-Chain Polyunsaturated Fatty Acids by Trophically Converted Marine Diatom *Phaeodactylum tricornutum*. *Marine drugs* **14**.
- Hansen, E.H., Møller, B.L., Kock, G.R., Büchner, C.M., Kristensen, C., Jensen, O.R., Okkels, F.T., Olsen, C.E., Motawia, M.S. and Hansen, J. (2009) De novo biosynthesis of vanillin in fission yeast (*Schizosaccharomyces pombe*) and baker's yeast (*Saccharomyces cerevisiae*). *Applied and environmental microbiology* **75**, 2765-2774.
- Harada, H., Nakatsuma, D., Ishida, M. and Matsuda, Y. (2005) Regulation of the expression of intracellular beta-carbonic anhydrase in response to CO₂ and light in the marine diatom *Phaeodactylum tricornutum*. *Plant physiology* **139**, 1041-1050.
- Harshvardhan, K., Suri, M., Goswami, A. and Goswami, T. (2017) Biological approach for the production of vanillin from lignocellulosic biomass (*Bambusa tulda*). *Journal of Cleaner Production* **149**, 485-490.
- Havkin-Frenkel, Daphna, Andrzej Podstolski, and Richard Dixon. 2003. Vanillin biosynthetic pathway enzyme from *Vanilla planifolia*. USPTO 20030070188A1. *US Patent*, issued April 10, 2003.
- He, L., Han, X. and Yu, Z. (2014) A rare *Phaeodactylum tricornutum* cruciform morphotype: culture conditions, transformation and unique fatty acid characteristics. *PloS one* **9**, e93922.

- Hernández, Juan Hernández, and Pesach Lubinsky. 2010. "Cultivation Systems." In *Vanilla*, edited by Eric Odoux and Michel Grisoni, 75-96. Medicinal and Aromatic Plants — Industrial Profiles. CRC Press.
- Hlavova, M., Turoczy, Z. and Bisova, K. (2015) Improving microalgae for biotechnology-From genetics to synthetic biology. *Biotechnol Adv* **33**, 1194-1203.
- Hoffmann, L., Besseau, S., Geoffroy, P., Ritzenthaler, C., Meyer, D., Lapierre, C., Pollet, B. and Legrand, M. (2004) Silencing of hydroxycinnamoyl-coenzyme A shikimate/quinate hydroxycinnamoyltransferase affects phenylpropanoid biosynthesis. *The Plant Cell* **16**, 1446-1465.
- Hoffmann, L., Maury, S., Martz, F., Geoffroy, P. and Legrand, M. (2003) Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism. *The Journal of biological chemistry* **278**, 95-103.
- Hua, D., Ma, C., Song, L., Lin, S., Zhang, Z., Deng, Z. and Xu, P. (2007) Enhanced vanillin production from ferulic acid using adsorbent resin. *Applied microbiology and biotechnology* **74**, 783-790.
- Huysman, M.J., Fortunato, A.E., Matthijs, M., Costa, B.S., Vanderhaeghen, R., Van den Daele, H., Sachse, M., Inzé, D., Bowler, C., Kroth, P.G., Wilhelm, C., Falciatore, A., Vyverman, W. and De Veylder, L. (2013) AUREOCHROME1-mediated induction of the diatom-specific cyclin dsCYC2 controls the onset of cell division in diatoms (*Phaeodactylum tricornutum*). *The Plant Cell* **25**, 215-228.
- Imre, S., Tero-Vescan, A., Dogaru, M.T., Kelemen, L., Muntean, D.L., Curticapean, A., Szegedi, N. and Vari, C.E. (2019) With or Without Internal Standard in HPLC Bioanalysis. A Case Study. *Journal of chromatographic science* **57**, 243-248.
- Iwasa, K. and Shimizu, A. (1972) Motility of the diatom, *Phaeodactylum tricornutum*. *Exp Cell Res* **74**, 552-558.
- Jacobsen, J.H. and Frigaard, N.U. (2014) Engineering of photosynthetic mannitol biosynthesis from CO₂ in a cyanobacterium. *Metab Eng* **21**, 60-70.
- Jeliński, T., Przybyłek, M. and Cysewski, P. (2019) Natural Deep Eutectic Solvents as Agents for Improving Solubility, Stability and Delivery of Curcumin. *Pharm Res* **36**, 116.
- Jones, P.R. (2014) Genetic instability in cyanobacteria - an elephant in the room? *Front Bioeng Biotechnol* **2**, 12.
- Juhas, M., von Zadow, A., Spexard, M., Schmidt, M., Kottke, T. and Büchel, C. (2014) A novel cryptochrome in the diatom *Phaeodactylum tricornutum* influences the regulation of light-harvesting protein levels. *The FEBS journal* **281**, 2299-2311.

- Kappachery, S., Paul, D., Yoon, J. and Kweon, J.H. (2010) Vanillin, a potential agent to prevent biofouling of reverse osmosis membrane. *Biofouling* **26**, 667-672.
- Karas, B.J., Diner, R.E., Lefebvre, S.C., McQuaid, J., Phillips, A.P., Noddings, C.M., Brunson, J.K., Valas, R.E., Deerinck, T.J., Jablanovic, J., Gillard, J.T., Beerli, K., Ellisman, M.H., Glass, J.I., Hutchison, C.A., 3rd, Smith, H.O., Venter, J.C., Allen, A.E., Dupont, C.L. and Weyman, P.D. (2015) Designer diatom episomes delivered by bacterial conjugation. *Nature communications* **6**, 6925.
- Kaur, S. and Spillane, C. (2015) Reduction in carotenoid levels in the marine diatom *Phaeodactylum tricornutum* by artificial microRNAs targeted against the endogenous phytoene synthase gene. *Marine biotechnology* **17**, 1-7.
- Kelly, C.L., Taylor, G.M., Hitchcock, A., Torres-Méndez, A. and Heap, J.T. (2018) A Rhamnose-Inducible System for Precise and Temporal Control of Gene Expression in Cyanobacteria. *ACS synthetic biology* **7**, 1056-1066.
- Khodadadi, M. and Pourfarzam, M. (2020) A review of strategies for untargeted urinary metabolomic analysis using gas chromatography-mass spectrometry. *Metabolomics* **16**, 66.
- Kilian, O. and Kroth, P.G. (2005) Identification and characterization of a new conserved motif within the presequence of proteins targeted into complex diatom plastids. *The Plant journal : for cell and molecular biology* **41**, 175-183.
- Kim, J.H., Lee, S.R., Li, L.H., Park, H.J., Park, J.H., Lee, K.Y., Kim, M.K., Shin, B.A. and Choi, S.Y. (2011) High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PloS one* **6**, e18556.
- Kundu, A. (2017) Vanillin biosynthetic pathways in plants. *Planta* **245**, 1069-1078.
- Kurniasih, S.D., Yamasaki, T., Kong, F., Okada, S., Widyaningrum, D. and Ohama, T. (2016) UV-mediated *Chlamydomonas* mutants with enhanced nuclear transgene expression by disruption of DNA methylation-dependent and independent silencing systems. *Plant Mol Biol* **92**, 629-641.
- Kusakabe, T., Tatsuke, T., Tsuruno, K., Hirokawa, Y., Atsumi, S., Liao, J.C. and Hanai, T. (2013) Engineering a synthetic pathway in cyanobacteria for isopropanol production directly from carbon dioxide and light. *Metab Eng* **20**, 101-108.
- Larosa, V. and Remacle, C. (2013) Transformation of the mitochondrial genome. *Int J Dev Biol* **57**, 659-665.
- Lauersen, K.J., Kruse, O. and Mussgnug, J.H. (2015) Targeted expression of nuclear transgenes in *Chlamydomonas reinhardtii* with a versatile, modular vector toolkit. *Applied microbiology and biotechnology* **99**, 3491-3503.

- Lavaud, J., Materna, A.C., Sturm, S., Vugrinec, S. and Kroth, P.G. (2012) Silencing of the violaxanthin de-epoxidase gene in the diatom *Phaeodactylum tricornutum* reduces diatoxanthin synthesis and non-photochemical quenching. *PLoS one* **7**, e36806.
- Lee, E.G., Yoon, S.H., Das, A., Lee, S.H., Li, C., Kim, J.Y., Choi, M.S., Oh, D.K. and Kim, S.W. (2009) Directing vanillin production from ferulic acid by increased acetyl-CoA consumption in recombinant *Escherichia coli*. *Biotechnology and bioengineering* **102**, 200-208.
- Lin, H.-Y., Yen, S.-C., Kuo, P.-C., Chung, C.-Y., Yeh, K.-L., Huang, C.-H., Chang, J. and Lin, H.-J. (2017) Alkaline phosphatase promoter as an efficient driving element for exogenic recombinant in the marine diatom *Phaeodactylum tricornutum*. *Algal Research* **23**, 58-65.
- Lin, W.R., Tan, S.I., Hsiang, C.C., Sung, P.K. and Ng, I.S. (2019) Challenges and opportunity of recent genome editing and multi-omics in cyanobacteria and microalgae for biorefinery. *Bioresour. Technol.* **291**, 121932.
- Liu, X., Hempel, F., Stork, S., Bolte, K., Moog, D., Heimerl, T., Maier, U.G. and Zauner, S. (2016) Addressing various compartments of the diatom model organism *Phaeodactylum tricornutum* via sub-cellular marker proteins. *Algal Research* **20**, 249-257.
- Liu, Z., Chen, O., Wall, J.B.J., Zheng, M., Zhou, Y., Wang, L., Vaseghi, H.R., Qian, L. and Liu, J. (2017) Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. *Scientific reports* **7**, 2193.
- Luke, G.A. and Ryan, M.D. (2018) "Therapeutic applications of the 'NPGP' family of viral 2As". *Rev Med Virol* **28**, e2001.
- Lütke-Eversloh, T. and Stephanopoulos, G. (2007) L-tyrosine production by deregulated strains of *Escherichia coli*. *Applied microbiology and biotechnology* **75**, 103-110.
- Ma, Q., Liu, L., Zhao, S., Huang, Z., Li, C., Jiang, S., Li, Q. and Gu, P. (2022) Biosynthesis of vanillin by different microorganisms: a review. *World J Microbiol Biotechnol* **38**, 40.
- Martău, G.A., Călinoiu, L.-F. and Vodnar, D.C. (2021) Bio-vanillin: Towards a sustainable industrial production. *Trends in Food Science & Technology* **109**, 579-592.
- Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, S.J., Witman, G.B., Terry, A., et al. (2007) The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* **318**, 245-250.
- Minskaia, E. and Ryan, M.D. (2013) Protein coexpression using FMDV 2A: effect of "linker" residues. *Biomed Res Int* **2013**, 291730.

- Miyagawa, A., Okami, T., Kira, N., Yamaguchi, H., Ohnishi, K. and Adachi, M. (2009) Research note: High efficiency transformation of the diatom *Phaeodactylum tricornutum* with a promoter from the diatom *Cylindrotheca fusiformis*. *Phycological Research* **57**, 142-146.
- Mona, S., Malyan, S.K., Saini, N., Deepak, B., Pugazhendhi, A. and Kumar, S.S. (2021) Towards sustainable agriculture with carbon sequestration, and greenhouse gas mitigation using algal biochar. *Chemosphere* **275**, 129856.
- Moosburner, M.A., Gholami, P., McCarthy, J.K., Tan, M., Bielinski, V.A. and Allen, A.E. (2020) Multiplexed Knockouts in the Model Diatom *Phaeodactylum* by Episomal Delivery of a Selectable Cas9. *Frontiers in microbiology* **11**, 5.
- Nair, R.B., Xia, Q., Kartha, C.J., Kurylo, E., Hirji, R.N., Datla, R. and Selvaraj, G. (2002) Arabidopsis CYP98A3 mediating aromatic 3-hydroxylation. Developmental regulation of the gene, and expression in yeast. *Plant physiology* **130**, 210-220.
- Nakagawa, A., Minami, H., Kim, J.S., Koyanagi, T., Katayama, T., Sato, F. and Kumagai, H. (2011) A bacterial platform for fermentative production of plant alkaloids. *Nature communications* **2**, 326.
- Negishi, O. and Negishi, Y. (2017) Phenylpropanoid 2,3-dioxygenase involved in the cleavage of the ferulic acid side chain to form vanillin and glyoxylic acid in *Vanilla planifolia*. *Biosci Biotechnol Biochem* **81**, 1732-1740.
- Neupert, J., Gallaher, S.D., Lu, Y., Strenkert, D., Segal, N., Barahimipour, R., Fitz-Gibbon, S.T., Schroda, M., Merchant, S.S. and Bock, R. (2020) An epigenetic gene silencing pathway selectively acting on transgenic DNA in the green alga *Chlamydomonas*. *Nature communications* **11**, 6269.
- Neupert, J., Karcher, D. and Bock, R. (2009) Generation of *Chlamydomonas* strains that efficiently express nuclear transgenes. *The Plant journal: for cell and molecular biology* **57**, 1140-1150.
- Ni, J., Tao, F., Du, H. and Xu, P. (2015) Mimicking a natural pathway for de novo biosynthesis: natural vanillin production from accessible carbon sources. *Scientific reports* **5**, 13670.
- Nicácio, A.E., Rodrigues, C.A., Jardim, I.C.S.F., Visentainer, J.V. and Maldaner, L. (2020) Modified QuEChERS method for phenolic compounds determination in mustard greens (*Brassica juncea*) using UHPLC-MS/MS. *Arabian Journal of Chemistry* **13**, 4681-4690.
- Niu, Y.F., Yang, Z.K., Zhang, M.H., Zhu, C.C., Yang, W.D., Liu, J.S. and Li, H.Y. (2012) Transformation of diatom *Phaeodactylum tricornutum* by electroporation and establishment of inducible selection marker. *Biotechniques* **52**.

- Odoux, Eric. 2010. "Vanilla Curing." In *Vanilla*, edited by Eric Odoux and Michel Grisoni, 173-88. Medicinal and Aromatic Plants — Industrial Profiles. CRC Press.
- Odoux, E., Chauwin, A. and Brillouet, J.M. (2003) Purification and characterization of vanilla bean (*Vanilla planifolia* Andrews) beta-D-glucosidase. *Journal of agricultural and food chemistry* **51**, 3168-3173.
- Onofrejšová, L., Vasícková, J., Klejšdus, B., Stratil, P., Misurcová, L., Krácmár, S., Kopecký, J. and Vacek, J. (2010) Bioactive phenols in algae: the application of pressurized-liquid and solid-phase extraction techniques. *Journal of pharmaceutical and biomedical analysis* **51**, 464-470.
- Pak, F.E., Gropper, S., Dai, W.D., Havkin-Frenkel, D. and Belanger, F.C. (2004) Characterization of a multifunctional methyltransferase from the orchid *Vanilla planifolia*. *Plant Cell Rep* **22**, 959-966.
- Pan, Y., Yang, J., Gong, Y., Li, X. and Hu, H. (2017) 3-Hydroxyisobutyryl-CoA hydrolase involved in isoleucine catabolism regulates triacylglycerol accumulation in *Phaeodactylum tricornutum*. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **372**.
- Paz, A., Outeiriño, D., Pinheiro de Souza Oliveira, R. and Domínguez, J.M. (2018) Fed-batch production of vanillin by *Bacillus aryabhatai* BA03. *N Biotechnol* **40**, 186-191.
- Peca, L., Kós, P.B., Máté, Z., Farsang, A. and Vass, I. (2008) Construction of bioluminescent cyanobacterial reporter strains for detection of nickel, cobalt and zinc. *FEMS Microbiol Lett* **289**, 258-264.
- Peris-Vicente, Juan, Josep Esteve-Romero, and Samuel Carda-Broch. 2015. "Validation of Analytical Methods Based on Chromatographic Techniques: An Overview." In *Analytical Separation Science*, 1757-1808. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA. <https://doi.org/10.1002/9783527678129.assep064>.
- Pittard, J., Camakaris, H. and Yang, J. (2005) The TyrR regulon. *Mol Microbiol* **55**, 16-26.
- Plucinak, T.M., Horken, K.M., Jiang, W., Fostvedt, J., Nguyen, S.T. and Weeks, D.P. (2015) Improved and versatile viral 2A platforms for dependable and inducible high-level expression of dicistronic nuclear genes in *Chlamydomonas reinhardtii*. *The Plant journal : for cell and molecular biology* **82**, 717-729.
- Podstolski, A., Havkin-Frenkel, D., Malinowski, J., Blount, J.W., Kourteva, G. and Dixon, R.A. (2002) Unusual 4-hydroxybenzaldehyde synthase activity from tissue cultures of the vanilla orchid *Vanilla planifolia*. *Phytochemistry* **61**, 611-620.

- Ponnusamy, K., Paul, D. and Kweon, J.H. (2009) Inhibition of Quorum Sensing Mechanism and *Aeromonas hydrophila* Biofilm Formation by Vanillin. *Environmental Engineering Science* **26**, 1359-1363.
- Pontrelli, S., Chiu, T.Y., Lan, E.I., Chen, F.Y., Chang, P. and Liao, J.C. (2018) *Escherichia coli* as a host for metabolic engineering. *Metab Eng* **50**, 16-46.
- Poust, S., Hagen, A., Katz, L. and Keasling, J.D. (2014) Narrowing the gap between the promise and reality of polyketide synthases as a synthetic biology platform. *Current opinion in biotechnology* **30**, 32-39.
- Sathish Kumar R., Arthanareeswaran, G., Lukka Thuyavan, Y. and Ismail, A.F. (2017) Enhancement of permeability and antibiofouling properties of polyethersulfone (PES) membrane through incorporation of quorum sensing inhibition (QSI) compound. *Journal of the Taiwan Institute of Chemical Engineers* **72**, 200-212.
- Ragauskas, A.J., Beckham, G.T., Biddy, M.J., Chandra, R., Chen, F., Davis, M.F., Davison, B.H., Dixon, R.A., Gilna, P., Keller, M., Langan, P., Naskar, A.K., Saddler, J.N., Tschaplinski, T.J., Tuskan, G.A. and Wyman, C.E. (2014) Lignin valorization: improving lignin processing in the biorefinery. *Science* **344**, 1246843.
- Ramachandra Rao, S. and Ravishankar, G.A. (2000) Vanilla flavour: production by conventional and biotechnological routes. *Journal of the science of food and agriculture* **80**, 289-304.
- Reijnders, M.J., van Heck, R.G., Lam, C.M., Scaife, M.A., dos Santos, V.A., Smith, A.G. and Schaap, P.J. (2014) Green genes: bioinformatics and systems-biology innovations drive algal biotechnology. *Trends Biotechnol* **32**, 617-626.
- Rico, M., López, A., Santana-Casiano, J.M., González, A.G. and González-Dávila, M. (2013) Variability of the phenolic profile in the diatom *Phaeodactylum tricornutum* growing under copper and iron stress. *Limnology and Oceanography* **58**, 144-152.
- Rodrigues, C.A., Nicacio, A.E., Visentainer, J.V., Maldaner, L. and Jardim, I.C.S.F. (2019) Determination of Phenolic Compounds in Red Sweet Pepper (*Capsicum annum* L.) Using a Modified Modified QuEChERS QuEChERS Method Method and UHPLC-MS/MS UHPLC-MS/MS Analysis Analysis and Its Relation to Antioxidant Activity. *Journal of the Brazilian Chemical Society* **30**, 1229-1240.
- Rogato, A., Richard, H., Sarazin, A., Voss, B., Cheminant Navarro, S., Champeimont, R., Navarro, L., Carbone, A., Hess, W.R. and Falciatore, A. (2014) The diversity of small non-coding RNAs in the diatom *Phaeodactylum tricornutum*. *BMC genomics* **15**, 698.

- Röling, W.F., Kerler, J., Braster, M., Apriyantono, A., Stam, H. and van Verseveld, H.W. (2001) Microorganisms with a taste for vanilla: microbial ecology of traditional Indonesian vanilla curing. *Applied and environmental microbiology* **67**, 1995-2003.
- Roshanak, S., Rahimmalek, M. and Goli, S.A. (2016) Evaluation of seven different drying treatments in respect to total flavonoid, phenolic, vitamin C content, chlorophyll, antioxidant activity and color of green tea (*Camellia sinensis* or *C. assamica*) leaves. *J Food Sci Technol* **53**, 721-729.
- Rösler, J., Krekel, F., Amrhein, N. and Schmid, J. (1997) Maize phenylalanine ammonia-lyase has tyrosine ammonia-lyase activity. *Plant physiology* **113**, 175-179.
- Rossi, B., Toschi, A., Piva, A. and Grilli, E. (2020) Single components of botanicals and nature-identical compounds as a non-antibiotic strategy to ameliorate health status and improve performance in poultry and pigs. *Nutr Res Rev* **33**, 218-234.
- Rossi, B., Esteban, M.A., García-Beltran, J.M., Giovagnoni, G., Cuesta, A., Piva, A. and Grilli, E. (2021) Antimicrobial Power of Organic Acids and Nature-Identical Compounds against Two *Vibrio* spp.: An In Vitro Study. *Microorganisms* **9**.
- Russo, M.T., Aiese Cigliano, R., Sanseverino, W. and Ferrante, M.I. (2018) Assessment of genomic changes in a CRISPR/Cas9 *Phaeodactylum tricorutum* mutant through whole genome resequencing. *PeerJ* **6**, e5507.
- Sadler, J.C. and Wallace, S. (2021) Microbial synthesis of vanillin from waste poly(ethylene terephthalate). *Green Chem* **23**, 4665-4672.
- Sakaguchi, T., Nakajima, K. and Matsuda, Y. (2011) Identification of the UMP synthase gene by establishment of uracil auxotrophic mutants and the phenotypic complementation system in the marine diatom *Phaeodactylum tricorutum*. *Plant physiology* **156**, 78-89.
- Scaife, M.A., Nguyen, G., Rico, J., Lambert, D., Helliwell, K.E. and Smith, A.G. (2015) Establishing *Chlamydomonas reinhardtii* as an industrial biotechnology host. *The Plant journal : for cell and molecular biology* **82**, 532-546.
- Schellenberger Costa, B., Sachse, M., Jungandreas, A., Bartulos, C.R., Gruber, A., Jakob, T., Kroth, P.G. and Wilhelm, C. (2013) Aureochrome 1a is involved in the photoacclimation of the diatom *Phaeodactylum tricorutum*. *PLoS one* **8**, e74451.
- Schroda, M. (2019) Good News for Nuclear Transgene Expression in *Chlamydomonas*. *Cells* **8**.
- Senes, C.E.R., Nicácio, A.E., Rodrigues, C.A., Manin, L.P., Maldaner, L. and Visentainer, J.V. (2020) Evaluation of Dispersive Solid-Phase Extraction (d-SPE) as a Clean-up Step for Phenolic Compound Determination of *Myrciaria cauliflora* Peel. *Food Analytical Methods* **13**, 155-165.

- Seyed Hosseini, N., Shang, H. and Scott, J.A. (2018) Biosequestration of industrial off-gas CO₂ for enhanced lipid productivity in open microalgae cultivation systems. *Renewable and Sustainable Energy Reviews* **92**, 458-469.
- Shahar, N., Landman, S., Weiner, I., Elman, T., Dafni, E., Feldman, Y., Tuller, T. and Yacoby, I. (2019) The Integration of Multiple Nuclear-Encoded Transgenes in the Green Alga *Chlamydomonas reinhardtii* Results in Higher Transcription Levels. *Front Plant Sci* **10**, 1784.
- Sharma, A.K., Nymark, M., Sparstad, T., Bones, A.M. and Winge, P. (2018) Transgene-free genome editing in marine algae by bacterial conjugation - comparison with biolistic CRISPR/Cas9 transformation. *Scientific reports* **8**, 14401.
- Sharma, A., Singh, J., Sharma, P., Tomar, G.S., Singh, S., Grover, M. and Nain, L. (2021) One-pot microbial bioconversion of wheat bran ferulic acid to biovanillin. *3 Biotech* **11**, 462.
- Sharma, N., Fleurent, G., Awwad, F., Cheng, M., Meddeb-Mouelhi, F., Budge, S.M., Germain, H. and Desgagné-Penix, I. (2020) Red Light Variation an Effective Alternative to Regulate Biomass and Lipid Profiles in *Phaeodactylum tricornutum*. *Applied Sciences* **10**, 2531.
- Siaut, M., Heijde, M., Mangogna, M., Montsant, A., Coesel, S., Allen, A., Manfredonia, A., Falciatore, A. and Bowler, C. (2007) Molecular toolbox for studying diatom biology in *Phaeodactylum tricornutum*. *Gene* **406**, 23-35.
- Siddiqui, A., Wei, Z., Boehm, M. and Ahmad, N. (2020) Engineering microalgae through chloroplast transformation to produce high-value industrial products. *Biotechnol Appl Biochem* **67**, 30-40.
- Sillanpää, Mika, and Chaker Ncibi. 2017. "Biochemicals." In *A Sustainable Bioeconomy: The Green Industrial Revolution*, edited by Mika Sillanpää and Chaker Ncibi, 141-83. Cham: Springer International Publishing.
- Sinha, A.K., Sharma, U.K. and Sharma, N. (2008) A comprehensive review on vanilla flavor: extraction, isolation and quantification of vanillin and others constituents. *International journal of food sciences and nutrition* **59**, 299-326.
- Sirisuk, P., Ra, C.H., Jeong, G.T. and Kim, S.K. (2018) Effects of wavelength mixing ratio and photoperiod on microalgal biomass and lipid production in a two-phase culture system using LED illumination. *Bioresource technology* **253**, 175-181.
- Slattery, S.S., Diamond, A., Wang, H., Therrien, J.A., Lant, J.T., Jazey, T., Lee, K., Klassen, Z., Desgagné-Penix, I., Karas, B.J. and Edgell, D.R. (2018) An Expanded Plasmid-Based Genetic Toolbox Enables Cas9 Genome Editing and Stable Maintenance of Synthetic Pathways in *Phaeodactylum tricornutum*. *ACS synthetic biology* **7**, 328-338.

- Slattery, S.S., Wang, H., Giguere, D.J., Kocsis, C., Urquhart, B.L., Karas, B.J. and Edgell, D.R. (2020) Plasmid-based complementation of large deletions in *Phaeodactylum tricornutum* biosynthetic genes generated by Cas9 editing. *Scientific reports* **10**, 13879.
- Souza-Moreira, T.M., Navarrete, C., Chen, X., Zanelli, C.F., Valentini, S.R., Furlan, M., Nielsen, J. and Krivoruchko, A. (2018) Screening of 2A peptides for polycistronic gene expression in yeast. *FEMS Yeast Res* **18**.
- Strand, T.A., Lale, R., Degnes, K.F., Lando, M. and Valla, S. (2014) A new and improved host-independent plasmid system for RK2-based conjugal transfer. *PloS one* **9**, e90372.
- Strucko, T., Buron, L.D., Jarczynska, Z.D., Nødvig, C.S., Mølgaard, L., Halkier, B.A. and Mortensen, U.H. (2017) CASCADE, a platform for controlled gene amplification for high, tunable and selection-free gene expression in yeast. *Scientific reports* **7**, 41431.
- Strucko, T., Magdenoska, O. and Mortensen, U.H. (2015) Benchmarking two commonly used *Saccharomyces cerevisiae* strains for heterologous vanillin- β -glucoside production. *Metab Eng Commun* **2**, 99-108.
- Su, B., Lai, P., Yang, F., Li, A., Deng, M.R. and Zhu, H. (2022) Engineering a Balanced Acetyl Coenzyme A Metabolism in *Saccharomyces cerevisiae* for Lycopene Production through Rational and Evolutionary Engineering. *Journal of agricultural and food chemistry* **70**, 4019-4029.
- Sulaiman, S., Yamato, S., Kanaya, E., Kim, J.J., Koga, Y., Takano, K. and Kanaya, S. (2012) Isolation of a novel cutinase homolog with polyethylene terephthalate-degrading activity from leaf-branch compost by using a metagenomic approach. *Applied and environmental microbiology* **78**, 1556-1562.
- Urban, P., Mignotte, C., Kazmaier, M., Delorme, F. and Pompon, D. (1997) Cloning, yeast expression, and characterization of the coupling of two distantly related *Arabidopsis thaliana* NADPH-cytochrome P450 reductases with P450 CYP73A5. *The Journal of biological chemistry* **272**, 19176-19186.
- Vavitsas, K., Crozet, P., Vinde, M.H., Davies, F., Lemaire, S.D. and Vickers, C.E. (2019) The Synthetic Biology Toolkit for Photosynthetic Microorganisms. *Plant physiology* **181**, 14-27.
- Veluchamy, A., Lin, X., Maumus, F., Rivarola, M., Bhavsar, J., Creasy, T., O'Brien, K., Sengamalay, N.A., Tallon, L.J., Smith, A.D., Rayko, E., Ahmed, I., Le Crom, S., Farrant, G.K., Sgro, J.Y., Olson, S.A., Bondurant, S.S., Allen, A.E., Rabinowicz, P.D., Sussman, M.R., Bowler, C. and Tirichine, L. (2013) Insights into the role of DNA methylation in diatoms by genome-wide profiling in *Phaeodactylum tricornutum*. *Nature communications* **4**, 2091.

- Villanova, V., Fortunato, A.E., Singh, D., Bo, D.D., Conte, M., Obata, T., Jouhet, J., Fernie, A.R., Marechal, E., Falciatore, A., Pagliardini, J., Le Monnier, A., Poolman, M., Curien, G., Petroutsos, D. and Finazzi, G. (2017) Investigating mixotrophic metabolism in the model diatom *Phaeodactylum tricorutum*. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **372**.
- Villanova, V. and Spetea, C. (2021) Mixotrophy in diatoms: Molecular mechanism and industrial potential. *Physiologia plantarum* **173**, 603-611.
- Wang, S., Wu, S., Yang, G., Pan, K., Wang, L. and Hu, Z. (2021) A review on the progress, challenges and prospects in commercializing microalgal fucoxanthin. *Biotechnol Adv* **53**, 107865.
- Wang, X. and Marchisio, M.A. (2021) Synthetic polycistronic sequences in eukaryotes. *Synth Syst Biotechnol* **6**, 254-261.
- Wang, Y., Wang, F., Wang, R., Zhao, P. and Xia, Q. (2015) 2A self-cleaving peptide-based multi-gene expression system in the silkworm *Bombyx mori*. *Scientific reports* **5**, 16273.
- Wild, J., Hradecna, Z. and Szybalski, W. (2002) Conditionally amplifiable BACs: switching from single-copy to high-copy vectors and genomic clones. *Genome Res* **12**, 1434-1444.
- Wu, G., Yan, Q., Jones, J.A., Tang, Y.J., Fong, S.S. and Koffas, M.A.G. (2016) Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications. *Trends Biotechnol* **34**, 652-664.
- Xiao, H., Zhao, J., Fang, C., Cao, Q., Xing, M., Li, X., Hou, J., Ji, A. and Song, S. (2020) Advances in Studies on the Pharmacological Activities of Fucoxanthin. *Marine drugs* **18**.
- Xu, F., Chen, Y., Cai, Y., Gu, F. and An, K. (2020) Distinct Roles for Bacterial and Fungal Communities During the Curing of Vanilla. *Frontiers in microbiology* **11**, 552388.
- Yadav, I., Rautela, A. and Kumar, S. (2021) Approaches in the photosynthetic production of sustainable fuels by cyanobacteria using tools of synthetic biology. *World J Microbiol Biotechnol* **37**, 201.
- Yamada, M., Okada, Y., Yoshida, T. and Nagasawa, T. (2008) Vanillin production using *Escherichia coli* cells over-expressing isoeugenol monooxygenase of *Pseudomonas putida*. *Biotechnology letters* **30**, 665-670.
- Yang, H., Barros-Rios, J., Kourteva, G., Rao, X., Chen, F., Shen, H., Liu, C., Podstolski, A., Belanger, F., Havkin-Frenkel, D. and Dixon, R.A. (2017) A re-evaluation of the final step of vanillin biosynthesis in the orchid *Vanilla planifolia*. *Phytochemistry* **139**, 33-46.

- Yang, J., Chen, Y.Z., Yu-Xuan, W., Tao, L., Zhang, Y.D., Wang, S.R., Zhang, G.C. and Zhang, J. (2021) Inhibitory effects and mechanisms of vanillin on gray mold and black rot of cherry tomatoes. *Pestic Biochem Physiol* **175**, 104859.
- Yang, W., Tang, H., Ni, J., Wu, Q., Hua, D., Tao, F. and Xu, P. (2013) Characterization of two *Streptomyces* enzymes that convert ferulic acid to vanillin. *PLoS one* **8**, e67339.
- Yemiş, G.P., Pagotto, F., Bach, S. and Delaquis, P. (2011) Effect of vanillin, ethyl vanillin, and vanillic acid on the growth and heat resistance of *Cronobacter* species. *J Food Prot* **74**, 2062-2069.
- Yemiş, G.P., Pagotto, F., Bach, S. and Delaquis, P. (2012) Thermal tolerance and survival of *Cronobacter sakazakii* in powdered infant formula supplemented with vanillin, ethyl vanillin, and vanillic acid. *J Food Sci* **77**, M523-527.
- Yoon, S.H., Lee, E.G., Das, A., Lee, S.H., Li, C., Ryu, H.K., Choi, M.S., Seo, W.T. and Kim, S.W. (2007) Enhanced vanillin production from recombinant *E. coli* using NTG mutagenesis and adsorbent resin. *Biotechnol Prog* **23**, 1143-1148.
- Yoon, S.H., Li, C., Kim, J.E., Lee, S.H., Yoon, J.Y., Choi, M.S., Seo, W.T., Yang, J.K., Kim, J.Y. and Kim, S.W. (2005a) Production of vanillin by metabolically engineered *Escherichia coli*. *Biotechnology letters* **27**, 1829-1832.
- Yoon, S.-H., Li, C., Lee, Y.-M., Lee, S.-H., Kim, S.-H., Choi, M.-S., Seo, W.-T., Yang, J.-K., Kim, J.-Y. and Kim, S.-W. (2005) Production of vanillin from ferulic acid using recombinant strains of *Escherichia coli*. *Biotechnology and Bioengineering* **10**, 378-384.
- Yoshinaga, R., Niwa-Kubota, M., Matsui, H. and Matsuda, Y. (2014) Characterization of iron-responsive promoters in the marine diatom *Phaeodactylum tricornutum*. *Mar Genomics* **16**, 55-62.
- Zang, Y., Jiang, T., Cong, Y., Zheng, Z. and Ouyang, J. (2015) Molecular Characterization of a Recombinant *Zea mays* Phenylalanine Ammonia-Lyase (ZmPAL2) and Its Application in trans-Cinnamic Acid Production from L-Phenylalanine. *Appl Biochem Biotechnol* **176**, 924-937.
- Zaslavskaja, L.A., Lippmeier, J.C., Shih, C., Ehrhardt, D., Grossman, A.R. and Apt, K.E. (2001) Trophic conversion of an obligate photoautotrophic organism through metabolic engineering. *Science* **292**, 2073-2075.
- Zaslavskaja, L.A., Lippmeier, J.C., Kroth, P.G., Grossman, A.R. and Apt, K.E. (2000) Transformation of the diatom *Phaeodactylum tricornutum* (Bacillariophyceae) with a variety of selectable marker and reporter genes. *Journal of phycology* **36**, 379-386.

- Zerulla, K., Ludt, K. and Soppa, J. (2016) The ploidy level of *Synechocystis* sp. PCC 6803 is highly variable and is influenced by growth phase and by chemical and physical external parameters. *Microbiology (Reading)* **162**, 730-739.
- Zhang, C. and Hu, H. (2014) High-efficiency nuclear transformation of the diatom *Phaeodactylum tricornutum* by electroporation. *Mar Genomics* **16**, 63-66.
- Zhang, M.P., Wang, M. and Wang, C. (2021) Nuclear transformation of *Chlamydomonas reinhardtii*: A review. *Biochimie* **181**, 1-11.
- Zhang, S. and Mueller, C. (2012) Comparative analysis of volatiles in traditionally cured Bourbon and Ugandan vanilla bean (*Vanilla planifolia*) extracts. *Journal of agricultural and food chemistry* **60**, 10433-10444.
- Zhao, L.-Q., Sun, Z.-H., Zheng, P. and He, J.-Y. (2006) Biotransformation of isoeugenol to vanillin by *Bacillus fusiformis* CGMCC1347 with the addition of resin HD-8. *Process Biochemistry* **41**, 1673-1676.
- Zhao, T., Han, X. and Cao, H. (2020) Effect of Temperature on Biological Macromolecules of Three Microalgae and Application of FT-IR for Evaluating Microalgal Lipid Characterization. *ACS Omega* **5**, 33262-33268.
- Zhu, B.H., Shi, H.P., Yang, G.P., Lv, N.N., Yang, M. and Pan, K.H. (2016) Silencing UDP-glucose pyrophosphorylase gene in *Phaeodactylum tricornutum* affects carbon allocation. *N Biotechnol* **33**, 237-244.
- Zhuang, D., He, N., Khoo, K.S., Ng, E.P., Chew, K.W. and Ling, T.C. (2022) Application progress of bioactive compounds in microalgae on pharmaceutical and cosmetics. *Chemosphere* **291**, 132932.

ANNEXE A

MUTATIONS ET VARIATIONS DANS LES SÉQUENCES DES PLASMIDES EXTRAITS DES CLONES V1-3 #2, VP-7 #1 ET VC-1 #1 PAR RAPPORT À LA SÉQUENCE DE RÉFÉRENCE DU PLASMIDE P0521S-V

Note : La séquence du plasmide p0521s-V a été utilisée pour faire l'alignement avec les séquences des plasmides extraits des clones de *P. tricornutum* V1-3 #2, Vc-1 #1 et Vp-7 #1. Le premier nucléotide du gène *Sh ble* est désigné comme position de référence #1 dans la séquence de p0521s-V.

Tableau A.1

Analyse des variations communes dans le séquençage des épisomes extraits des clones de *P. tricornutum* V1-3 #2, Vc-1 #1 et Vp-7 #1

Position de référence	Description des mutations en comparaison avec le plasmide vide p0521s	Régions	Effet
795	Substitution 1 pb (T→C)	Promoteur H4-1B (Cassette DMi1)	Inconnu
816	Substitution 1 pb (A→T)		Inconnu
818	Substitution 1 pb (T→C)		Inconnu
923	Substitution 1 pb (T→C)		Inconnu
1 082	Substitution 1 pb (C→T)		Inconnu
6 071	Substitution 1 pb (A→C)	Promoteur 40SRPS8 (Cassette DMi2)	Inconnu
6 095	Substitution 1 pb (A→G)		Inconnu
10 825-10840	Indel (CCCCCCCCCTTTCAA→GGGGGGGGGGGGGGGG)	URA3	Inconnu
10 842 (+1)	Insertion 1 pb (T)		Inconnu
11 988-11 997	Indel (GGGGGGGGGG→CCCCCCCCCCCCCCCC)		Inconnu
16 313-16 318	Substitution 4 pb (T CGTT →CGTGGA)	Terminateur EF-1α (Cassette DMi3)	Inconnu
16 324	Substitution 1 pb (T→G)		Inconnu
16 333	Substitution 1 pb (A→C)		Inconnu

Position de référence	Description des mutations en comparaison avec le plasmide vide p0521s	Régions	Effet
16 335	Substitution 1 pb (C→A)		Inconnu
16 346	Substitution 1 pb (C→T)		Inconnu
16 365	Substitution 1 pb (C→A)		Inconnu
16 369-16 370	Substitution 2 pb (GC→TG)		Inconnu
16 374	Délétion 1 pb (G)		Inconnu
16 377	Substitution 1 pb (A→T)		Inconnu
16 406	Substitution 1 pb (T→A)		Inconnu
16 450	Substitution 1 pb (T→C)		Inconnu
16 506	Substitution 1 pb (G→C)		Inconnu
16 667	Substitution 1 pb (G→A)		Inconnu
16 692-16693	Substitution 2 pb (AG→TA)		Inconnu
16 737-16 745	Délétion 9 pb (CGTCATCGT)		Inconnu
16 813	Substitution 1 pb (T→G)	Promoteur FcpD (Cassette DMi4)	Inconnu
16 841	Substitution 2 pb (A→C)		Inconnu
16 847	Substitution 1 pb (G→C)		Inconnu
17 183	Substitution 1 pb (T→G)		Inconnu

Tableau A.2

Analyse des variations communes dans le séquençage des épisomes extraits des clones de *P. tricornutum* V1-3 #2 et Vc-1 #1

Position de référence	Description des mutations en comparaison avec le plasmide vide p0521s	Régions	Effet
19 731-19 732	Substitution 2 pb (CA → TG)	Terminateur FcpD (Cassette DMi4)	Inconnu
19 742	Substitution 1 pb (G → A)		Inconnu
19 767	Substitution 1 pb (T → C)		Inconnu
19 775	Substitution 1 pb (C → T)		Inconnu
19 799	Substitution 1 pb (C → G)		Inconnu
19 824	Substitution 1 pb (C → A)		Inconnu

Tableau A.3

Analyse des variations dans le séquençage de l'épisode extrait du clone
de *P. tricornutum* V1-3 #2

Position de référence	Description des mutations en comparaison avec le plasmide vide p0521s	Régions	Effet
1 675	Substitution 1 pb (G→A)	<i>ZmPAL</i> (Cassette DMi1)	Mutation silencieuse
5 118	Substitution 1 pb (A→C)	Terminateur H4-1B (Cassette DMi1)	Inconnu
5 123	Substitution 1 pb (T→C)		Inconnu
5 210	Substitution 1 pb (T→C)		Inconnu
5 261-5 265	Indel (TATAA→GTATG)		Inconnu
5 281	Substitution 1 pb (T→G)		Inconnu
5 286	Substitution 1 pb (G→T)		Inconnu
5 427	Substitution 1 pb (G→A)		Inconnu
5 626	Substitution 1 pb (A→C)	Promoteur 40SRPS8 (Cassette DMi2)	Inconnu
15 251 (+1)	Insertion 1 pb (G)	<i>4CL</i> (Cassette DMi3)	Changement du cadre de lecture du gène <i>4CL</i> après la 812 ^e pb. La longueur totale de <i>4CL</i> est de 1701 pb.
19 503	Substitution 1 pb (G→A)	<i>VpVAN</i> (Cassette DMi4)	Substitution non synonyme (A755T)
19 542 (+1)	Insertion 1 pb (T)	Terminateur FcpD (Cassette DMi4)	Inconnu
19 787-19 791	Substitution 4 pb (TTTCG→CCTTT)		Inconnu
22 883	Délétion 1 pb (A)	Pt DNA region 2	Inconnu
22 886-22 887	Substitution 2 pb (AG→GA)		Inconnu
22 913 (+1)	Insertion 1 pb (A)		Inconnu

Tableau A.4

Analyse des variations dans le séquençage de l'épisome extrait du clone
de *P. tricornutum* Vc-1 #1

Position de référence	Description des mutations en comparaison avec le plasmide vide p0521s	Régions	Effet
5 072-5 073	Délétion 2 pb (AA)	<i>AtC4H</i> (Cassette DMi1)	Changement dans le cadre de lecture à partir de l'avant-dernier triplet du gène. Perte du codon stop d'origine.
9 969	Délétion 1 pb (A)	Terminateur 40SRPS8 (Cassette DMi2)	Inconnu
11 975	Insertion 1 pb (A)	URA3	Inconnu
12 233-12 234	Délétion 2 pb (CC)		Inconnu
16 631	Délétion 1 pb (T)	Terminateur EF-1 α (Cassette DMi3)	Inconnu
17 235	Substitution 1 pb (G→A)	Promoteur FcpD (Cassette DMi4)	Inconnu
19 787	Substitution 1 pb (T→G)	Terminateur FcpD (Cassette DMi4)	Inconnu
19 791	Substitution 1 pb (G→A)		Inconnu
22 884	Substitution 1 pb (A→G)	Pt DNA region 2	Inconnu
22 887	Substitution 1 pb (G→A)		Inconnu
32 911-32 914	Délétion 4 pb (GACA)	Promoteur FcpC	Inconnu

Tableau A.5

Analyse des variations dans le séquençage de l'épisome extrait du clone
de *P. tricornutum* Vp-7 #1

Position de référence	Description des mutations en comparaison avec le plasmide vide p0521s	Régions	Effet
4571	Substitution 1 pb (G→A)	<i>AtC4H</i> (Cassette DMi1)	Substitution non synonyme (D1093N)
6 199-6 201	Délétion 3 pb (CAC)	<i>ZmTAL</i> (Cassette DMi2)	Délétion d'un codon complet (7H) qui réduit d'une histidine la séquence de l'étiquette de poly-histidines. La détection de cette étiquette par les anticorps anti-his pourrait être diminuée.
11 975	Insertion 1 pb (A)	URA3	Inconnu
16 593	Substitution 1 pb (G→A)	Terminateur EF-1 α (Cassette DMi3)	Inconnu
17 192	Délétion 1 pb (A)	Promoteur FcpD (Cassette DMi4)	Inconnu
19 590-23 683	Région réarrangée. Délétion de cette séquence et insertion des séquences suivantes dans l'ordre : <ul style="list-style-type: none"> • Copies inversées des séquences suivantes; le gène <i>ZmPAL</i> partiel, la séquence du peptide FMDV2A, le gène <i>AtC4H</i> et le terminateur H4-1B partiel (Position de référence 3 408-5 285). La copie du gène <i>AtC4H</i> contient une substitution 1 pb (G→A) qui causerait une substitution non synonyme (D1093N) si cette séquence était traduite en respectant le cadre de lecture de la cassette DMi1. • Copie partielle et inversée de <i>ZmPAL</i> (Position de référence 3 356-3 407) • Copie partielle et inversée de <i>ZmPAL</i> (Position de référence 2 924-3 407) 	Terminateur FcpD (Cassette DMi4), cassette OriT, <i>Pt</i> DNA region 1, <i>Pt</i> DNA region 2 et oriV	La perte partielle du terminateur FcpD peut affecter l'expression de la cassette DMi4. La perte de la cassette OriT empêche de transformer par conjugaison bactérienne ce plasmide. Les séquences nommées <i>Pt</i> DNA region 1 et <i>Pt</i> DNA region 2 ne sont pas essentiels, mais elles augmenteraient le nombre de clones obtenus par conjugaison bactérienne (Karas et al. 2015). La séquence OriV est une origine de réplication inductible pour <i>E. coli</i> et sa perte partielle pourrait empêcher d'augmenter par induction le nombre de copies du plasmide (Wild, Hradecna, et Szybalski 2002).
23 688	Délétion 1 pb (T)	oriV	Inconnu

ANNEXE B

FIGURES ET TABLEAUX SUPPLÉMENTAIRES DE L'ARTICLE PRÉSENTÉ AU CHAPITRE III

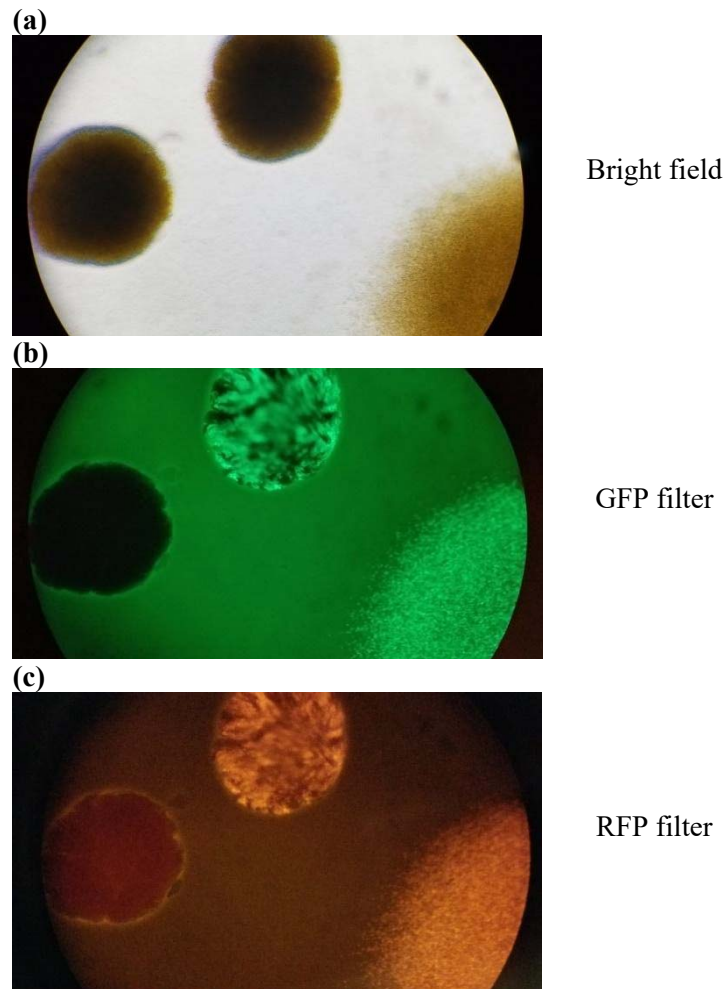


Figure B.1 Observation of pPtGE33 colonies, that are producing eGFP and mCherry separated by a T2A self-cleaving peptide, with a fluorescent stereo microscope under (a) bright field, (b) GFP filter, and (c) RFP filter.

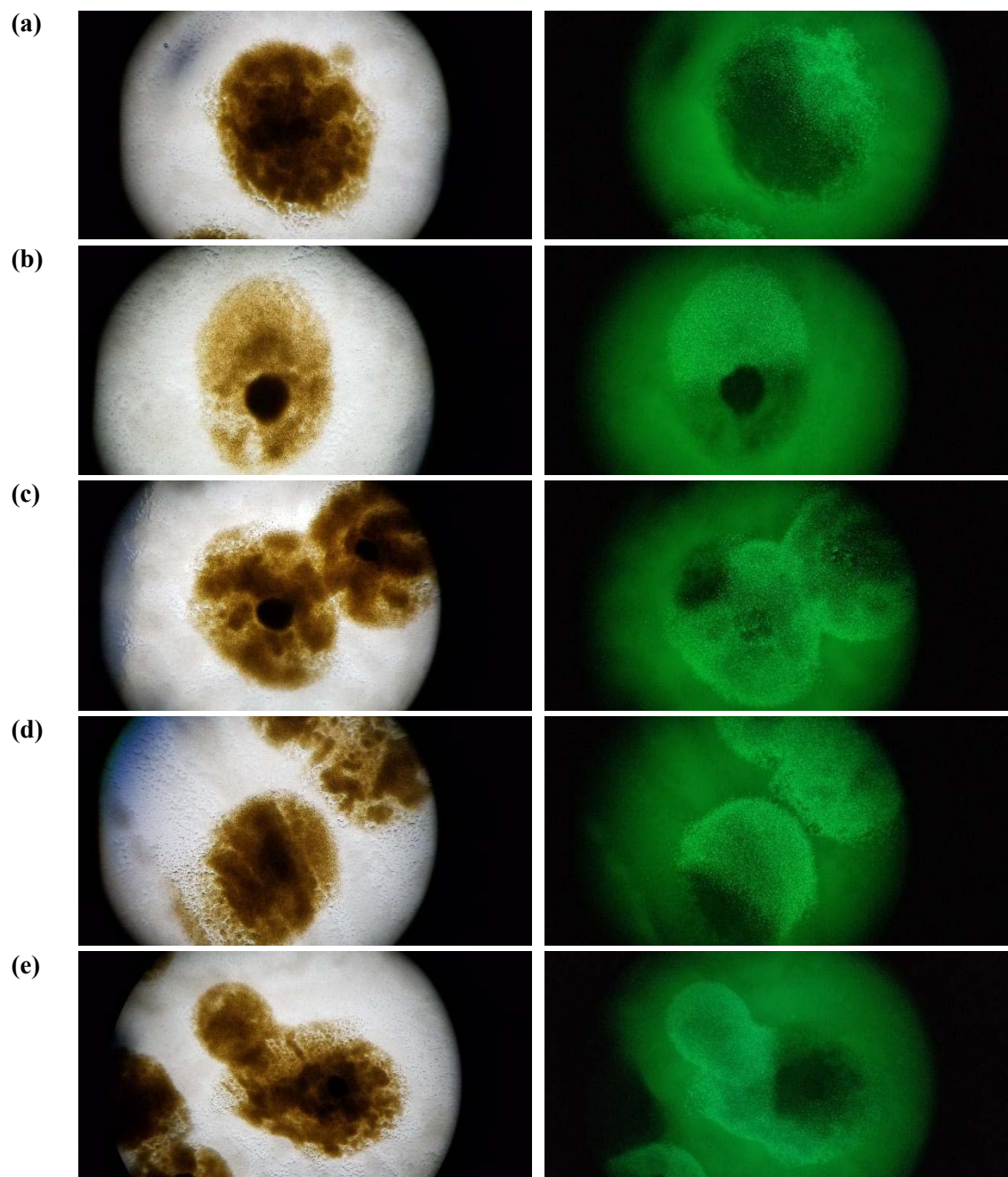


Figure B.2 Observation of different pControl colonies (a to e) with a fluorescent stereo microscope under bright field (left panel) and GFP filter (right panel).

These transconjugants are expressing GFP and mCherry separated by a T2A self-cleaving peptide.

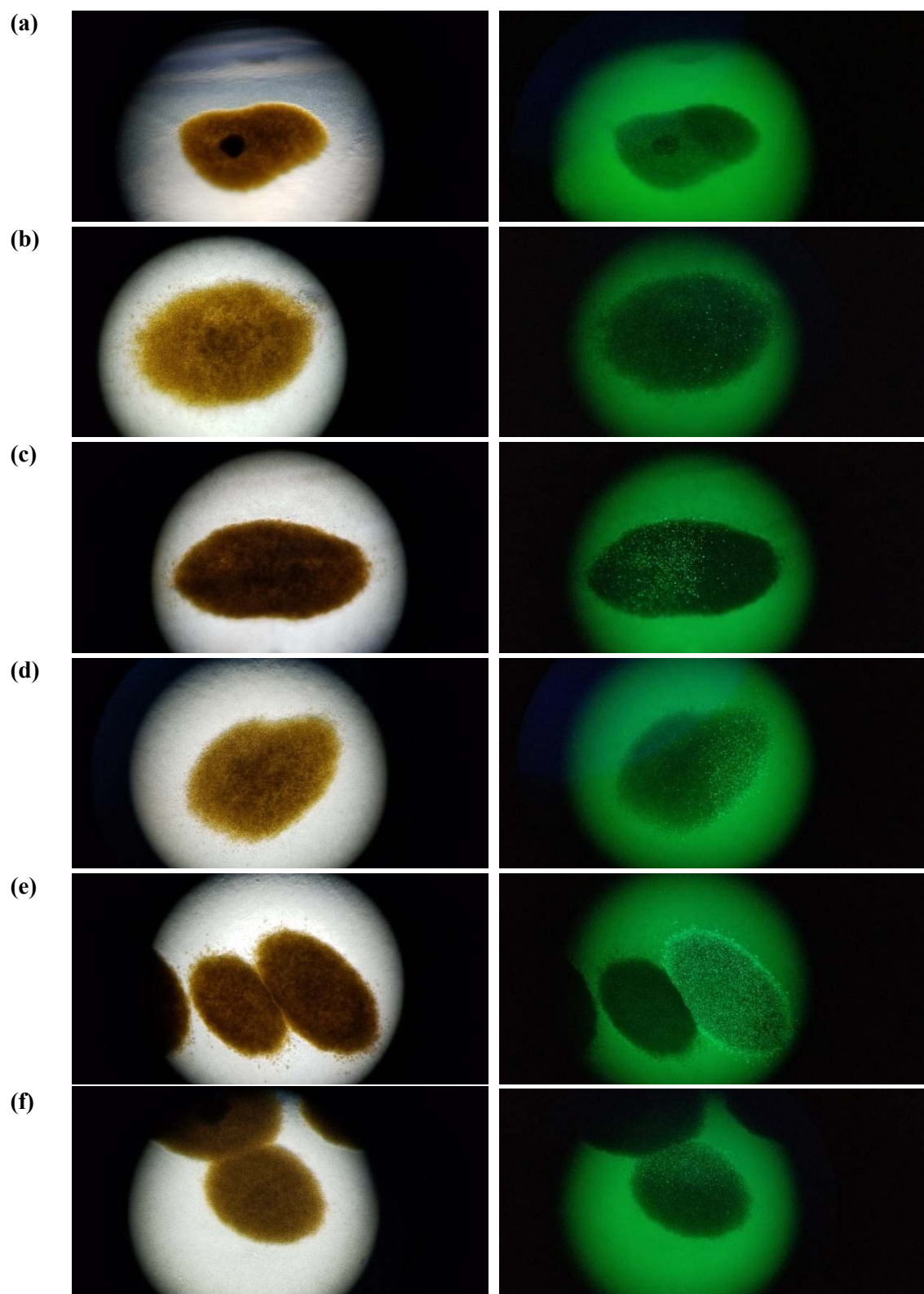


Figure B.3 Observation of different pDMi7 colonies (a to f) with a fluorescent stereo microscope under bright field (left panel) and GFP filter (right panel). These transconjugants are producing eGFP separated by a T2A self-cleaving peptide from the ech and fcs proteins fused with a flexible linker.

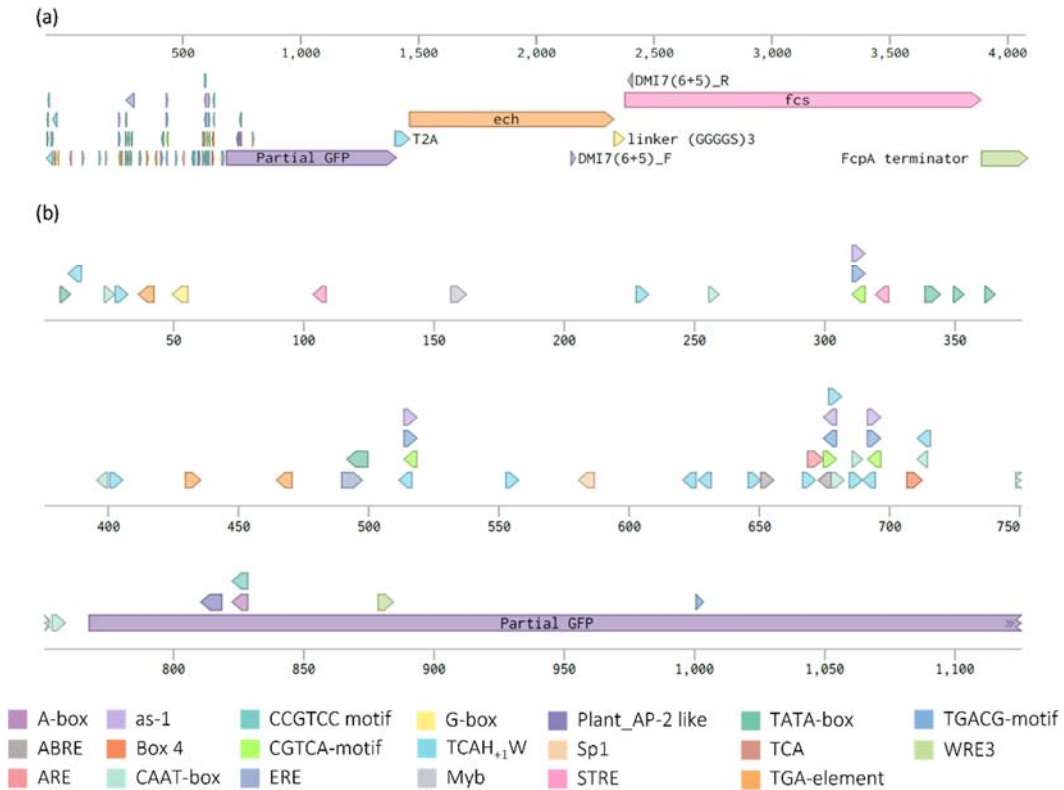


Figure B.4 Analysis of putative promoter region of DMI7-21-3 clone product of plasmid rearrangements.

(a) Complete scheme of the second copy of DMI7 construction without 40SRS8 promoter sequence. (b) Zoom on the first 1,150 bp of the scheme in A. Transcription factor binding sites found using PlantCARE and transcription initiation potential conserve motif (TCAH₁W) of *P. tricornutum*. Schemes were done using Benchling.

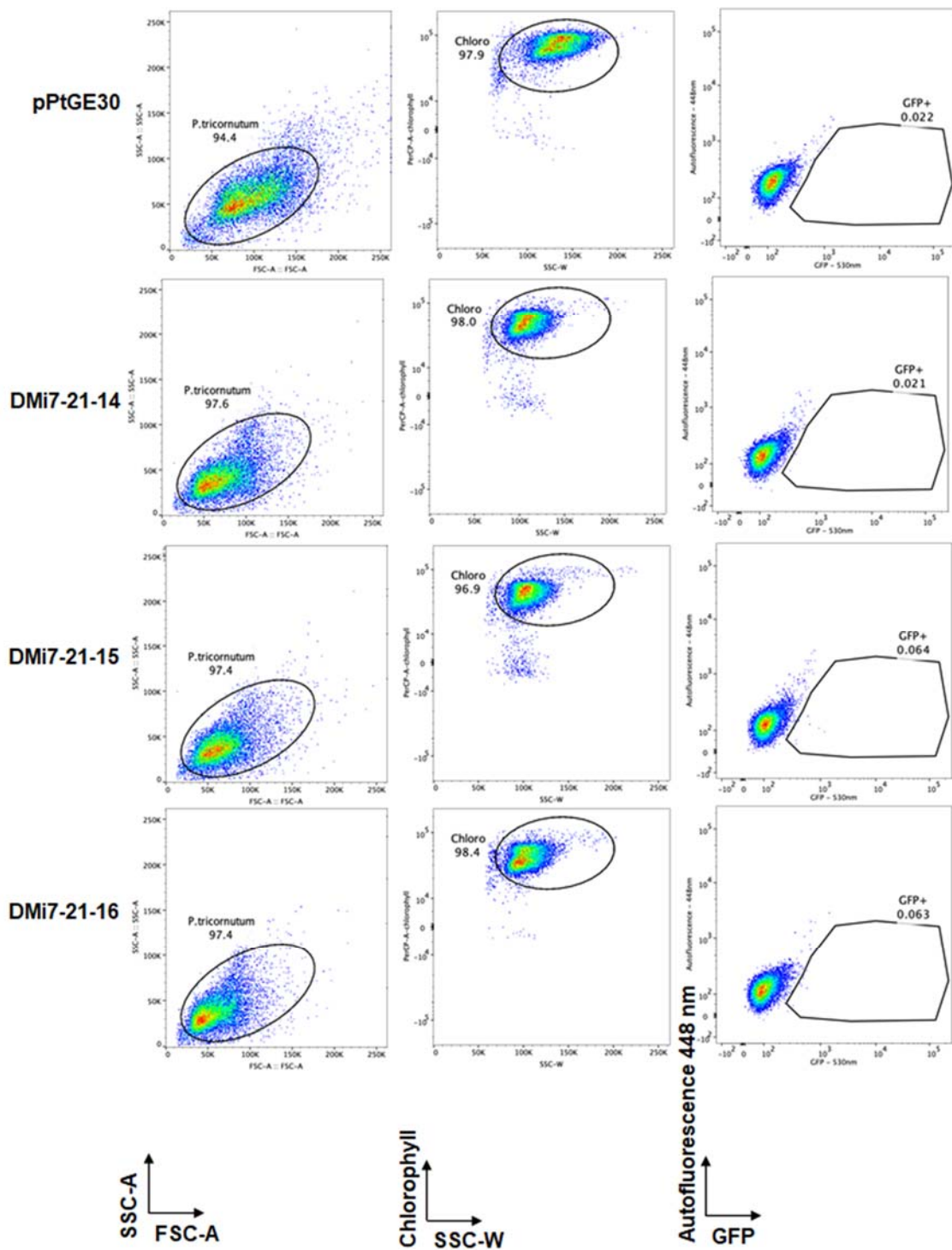


Figure B.5 Analysis of GFP fluorescence by flow cytometry from *P. tricornutum* cultures of DMI7-21-14, DMI7-21-15 and DMI7-21-16 that were detected as non-fluorescent (GFP-).

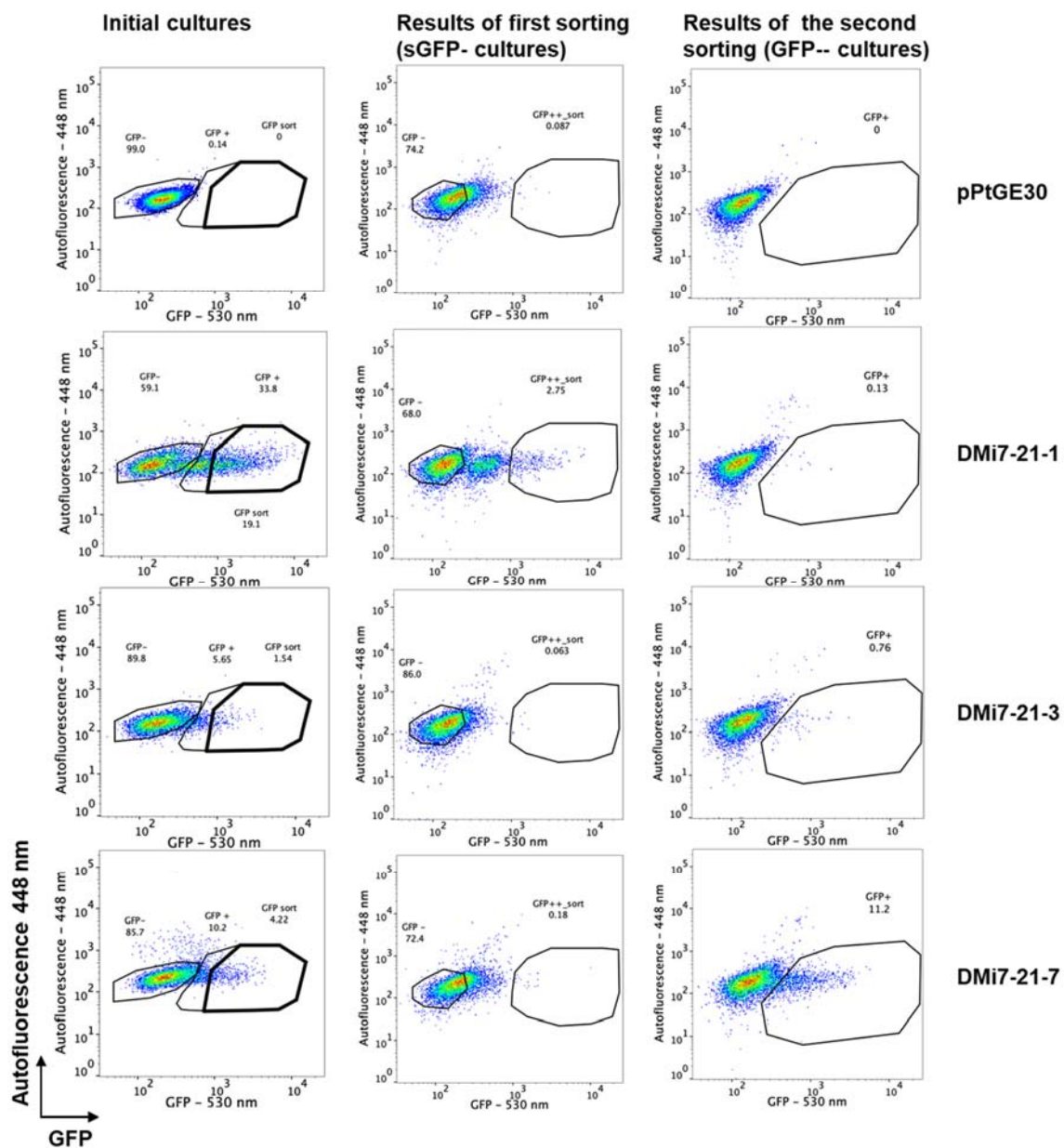
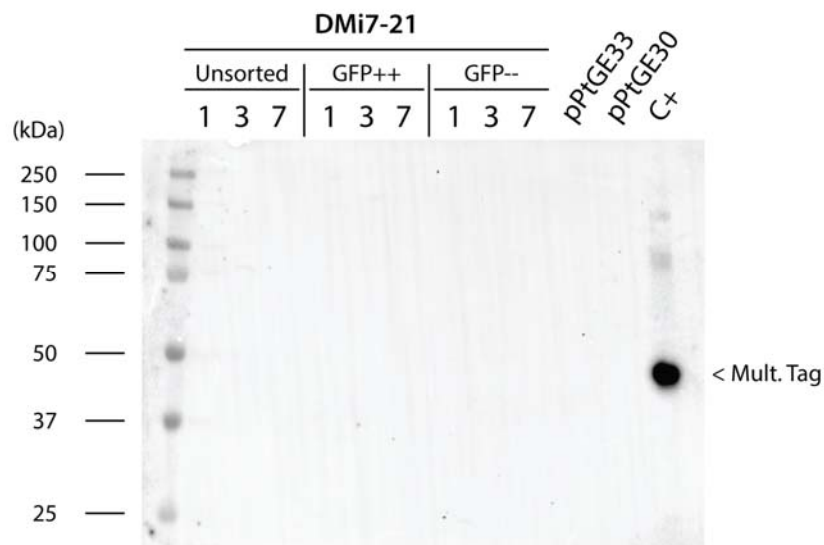


Figure B.6 Antibiotic resistant GFP- cells can be stably isolated from diatoms conjugated with a GFP and antibiotic resistance genes expressing episome following two cycles of fluorescence-activated cell sorting. Pseudocolor dot plots of pPtGE30 and DMi7-21-1, DMi7-21-3 and DMi7-21-7 cultures with 530 nm (GFP) in the x axis and autofluorescence at 448 nm in the y axis across days following two sequences of enrichment (initial cultures and results from first sorting). Gates and frequencies of total GFP+, sGFP+ (GFP sort and GFP++ sort) and sGFP- population were designed according to the autofluorescence of the negative control pPtGE30 at each day of experiment and are shown as reference.

(a)



(b)

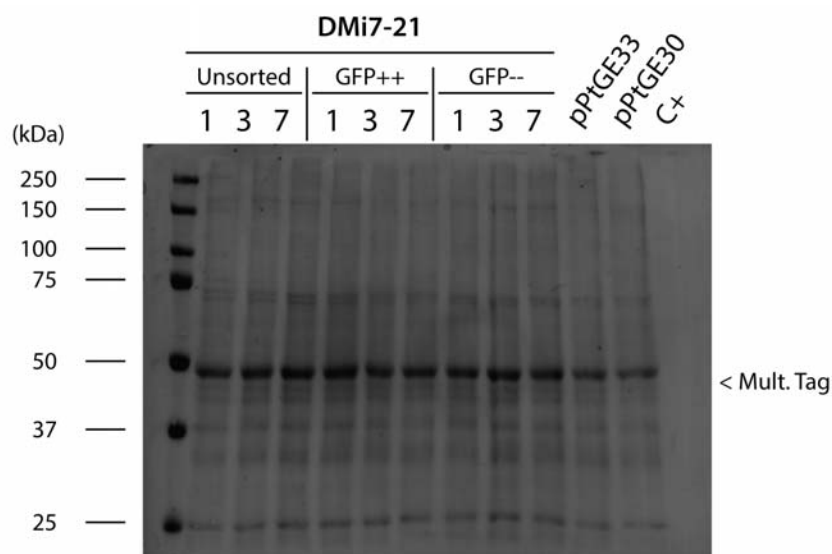


Figure B.7 Western blot anti-Myc tag on protein extracts from *P. tricornutum* transconjugants for the detection of the ech-fcs fusion protein production.

(a) Western blot analysis using anti-myc antibodies on proteins extracted from unsorted, GFP+ sorted (GFP++), GFP- sorted (GFP--) cultures of DMi7-21-1, DMi7-21- 3 and DMi7-21-7 compared to pPtGE30 (negative control as empty vector), pPtGE33 (negative control as a strain expressing heterologous proteins without myc tag) and a Multiple Tag purified protein (C+; positive control for protein with a myc tag). (b) Visualization of the quantity of proteins transferred on the membrane by Ponceau S staining.

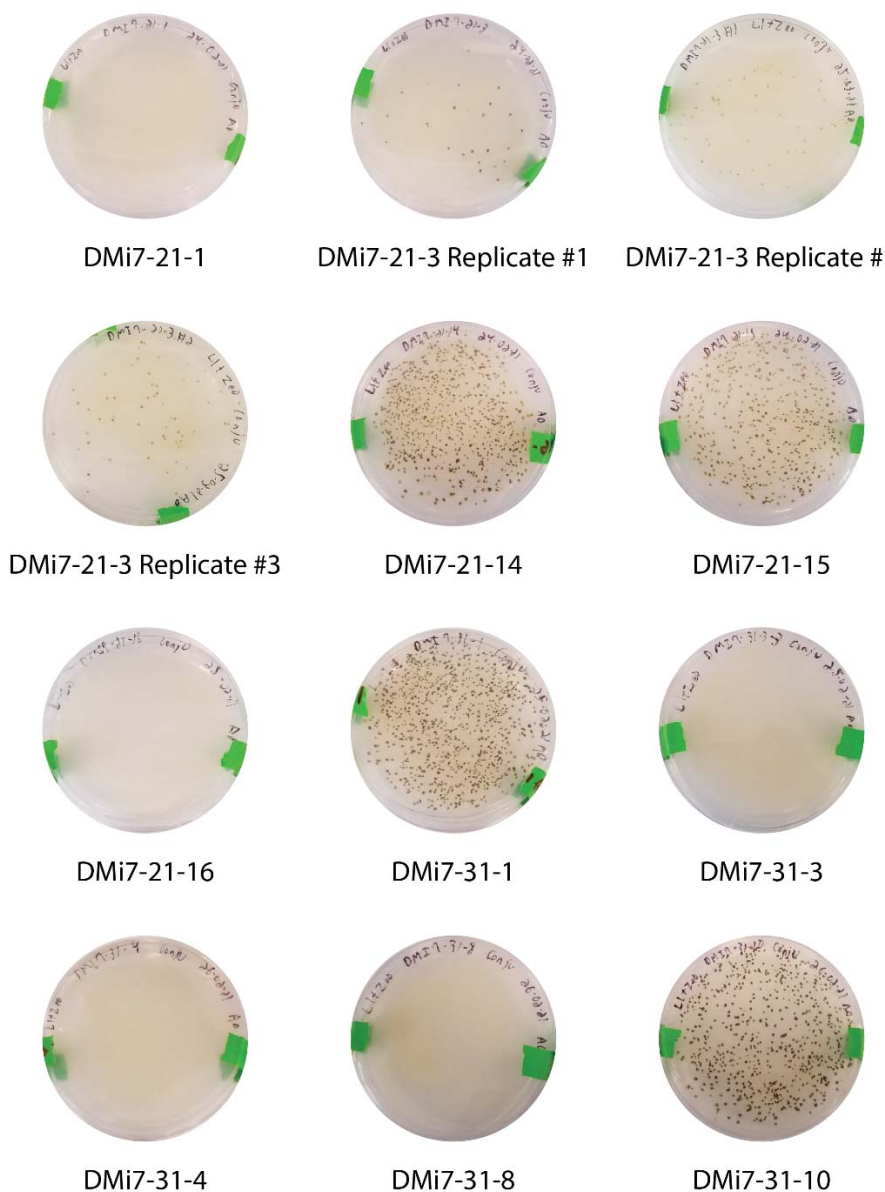


Figure B.8 Colonies obtained 14 days after the bacterial conjugation with the episomes recovered from *P. tricornutum* initially transformed with pDMi7.

For each plate, 200 μ L has been spread from the mixture of *E. coli* and *P. tricornutum* cells scraped before the transfer on L1 plates with zeocin. Absence of colony is observed for bacterial conjugation with rescued plasmid from DMi7-21-1, DMi7-21-16, DMi7-31-3, DMi7-31-4, and DMi7-31-8 clones. The total volumes of the cells mixtures were 630 μ L for DMi7-21-1, 690 μ L for DMi7-21-3 replicate #1, 650 μ L for DMi7-21-3 replicate #2, 790 μ L for DMi7-21-3 replicate #3, 670 μ L for DMi7-21-14, 680 μ L for DMi7-21-15, 690 μ L for DMi7-21-16, 690 μ L for DMi7-31-1, 580 μ L for DMi7-31-3, 680 μ L for DMi7-31-4, 730 μ L for DMi7-31-8, and 720 μ L for DMi7-31-10.

Table contains 10408 CDSs (4843509 codons)

TTT 22.47 (108824)	TCT 13.01 (63006)	TAT 9.15 (44297)	TGT 7.66 (37104)
TTC 14.88 (72069)	TCC 17.30 (83795)	TAC 16.31 (78975)	TGC 8.05 (38976)
TTA 7.64 (37019)	TCA 10.69 (51786)	TAA 0.75 (3616)	TGA 0.52 (2515)
TTG 27.08 (131152)	TCG 17.95 (86934)	TAG 0.62 (3021)	TGG 13.44 (65120)
CTT 15.22 (73737)	CCT 11.42 (55317)	CAT 10.60 (51352)	CGT 12.80 (62003)
CTC 18.24 (88340)	CCC 13.70 (66351)	CAC 14.02 (67914)	CGC 13.91 (67389)
CTA 9.16 (44366)	CCA 12.55 (60789)	CAA 23.04 (111606)	CGA 14.03 (67955)
CTG 16.89 (81788)	CCG 14.18 (68687)	CAG 17.78 (86094)	CGG 10.41 (50432)
ATT 22.92 (111005)	ACT 12.88 (62365)	AAT 17.81 (86259)	AGT 12.40 (60068)
ATC 15.10 (73139)	ACC 17.17 (83179)	AAC 19.86 (96173)	AGC 12.30 (59551)
ATA 6.94 (33618)	ACA 13.71 (66423)	AAA 22.81 (110468)	AGA 6.05 (29316)
ATG 22.50 (108978)	ACG 17.40 (84267)	AAG 23.59 (114235)	AGG 4.22 (20452)
GTT 17.69 (85671)	GCT 20.61 (99808)	GAT 29.37 (142264)	GGT 17.72 (85811)
GTC 21.03 (101877)	GCC 27.72 (134265)	GAC 29.60 (143351)	GGC 16.42 (79532)
GTA 11.55 (55922)	GCA 17.79 (86153)	GAA 38.82 (188001)	GGA 19.85 (96138)
GTG 18.00 (87181)	GCG 18.98 (91951)	GAG 21.90 (106056)	GGG 9.85 (47703)

Figure B.9 Codon usage table for *P. tricornutum* extracted on the High-performance Integrated Virtual Environment-Codon Usage Tables (HIVE-CUTs) platform (Athey et al., 2017) using the refseq database on September 21st in 2018.

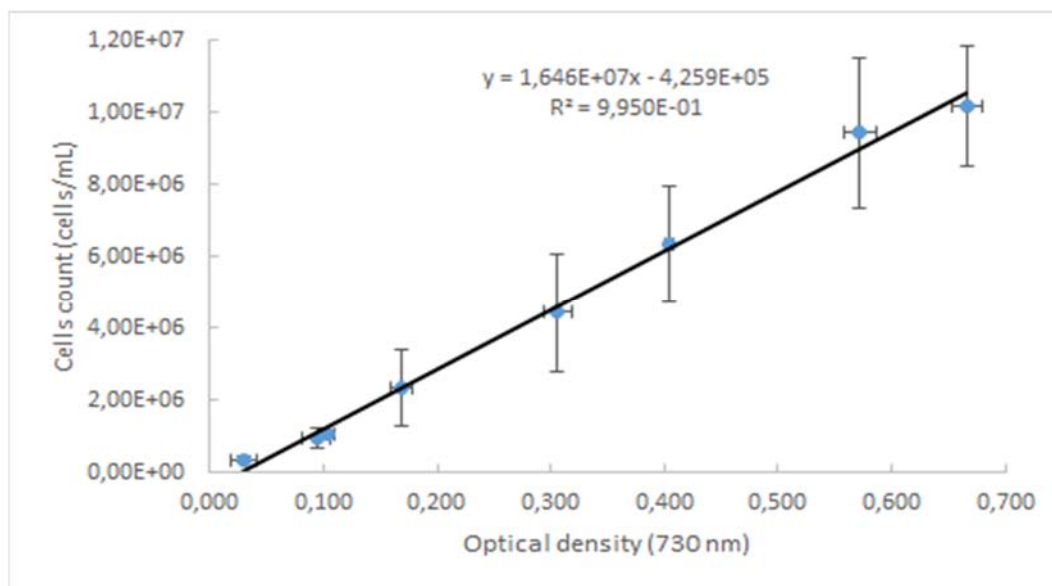


Figure B.10 Calibration curve for the optical density at 730 nm of *P. tricornutum* liquid culture and the cell count with a Neubauer improved cell counting chamber.

Table B.1 Count, and fluorescence detection of *P. tricorntutum* colonies by fluorescence microscopy 14 days after conjugation for pControl.

Conjugation		Total count of colonies	Average of colonies per conjugation	GFP fluorescent colonies (%)	Average of GFP fluorescent colonies (%)	mCherry fluorescent colonies (%)	Average of mCherry fluorescent colonies (%)
pControl	#1	2025	2131±153	81.1 (185/228)	82.1±1.8	77.2 (176/228)	76.0±1.0
	#2	2061		84.1 (122/145)		75.2 (109/145)	
	#3	2307		81.1 (167/206)		75.7 (156/206)	

Table B.2 Count, and fluorescence detection of *P. tricorntutum* colonies by fluorescence microscopy 14 days after conjugation for pDMi7 and the empty vector (pPtGE30).

Conjugation		Total count	Average	Percentage of eGFP fluorescent colonies	Average of eGFP fluorescent colonies
pDMi7	#1	3046	3281±305	8.7 (74/849)	8.2±0.8
	#2	3626		7.7 (71/928)	
	#3	3170		NA	
pPtGE30 (Empty vector)	#1	3474	3555±142	0.0 (0/100)	0.0±0.0
	#2	3719		0.0 (0/100)	
	#3	3471		0.0 (0/100)	

NA: Not available

Table B.3 Sequence analysis of the plasmid recovered from *P. tricornutum* transconjugants.

Note: The first nucleotide of the Sh ble gene is set as the reference position 1. The reference sequence is the sequencing results of pDMi7 following the Gibson assembly.

Clone	Reference position	Description	Features affected	Effect
DMi7-21-1 (GFP+)	3,255-8,777	Deletion of 5,523 bp.	<i>fcs</i> , FcpA terminator, OriT cassette, Pt region 1 and 2, and OriV	The <i>fcs</i> protein cannot be produced. Bacterial conjugation with this plasmid is impossible due to the loss of OriT cassette. Pt region 1 and 2 are not essential, but they increase the number of colonies obtained from bacterial conjugation due to an unknown phenomenon (Karas et al., 2015). The partial deletion of OriV, an inducible origin of replication, can inhibit the increase of copy number obtained by induction (Wild, Hradecna, and Szybalski 2002).
DMi7-21-3 (GFP+)	2,603-7,606	Deletion of 5,004 bp.	<i>ech</i> , linker (GGGGS) ₃ , <i>fcs</i> , FcpA terminator, OriT cassette and Pt region 1	The <i>ech-fcs</i> polyprotein cannot be produced from this cassette. The bacterial conjugation of this plasmid will not be possible from this sequence.
	13,885	Transition of 1 bp (T->G).	sopB	Unknown
	13,888(+1)	Insertion of a 2,400 bp fragment that is an inverted duplicate of the partial sequence of <i>ech-fcs</i> gene, the complete FcpA terminator and the partial OriT cassette (reference positions 2,604-5,003).	sopB	Unknown.
	13,889-14,878	Deletion of 990 bp.	sopB, <i>sopc</i> and <i>cos</i>	Unknown.

Clone	Reference position	Description	Features affected	Effect
	14,879(-1)	Insertion of a 3,561 bp fragment that is an inverted duplicate of the partial sequence of <i>GFP-ech-fcs</i> gene, the complete FcpA terminator and the partial OriT cassette (reference positions 1,471-5,031).	sopB	The partial duplicate of the OriT cassette contains only 23 bp on 110 bp of the OriT sequence and nothing from the traJ sequence. The bacterial conjugation of this plasmid would probably not be possible from this sequence.
	17,753(+1)	Insertion of 1 bp (C).	FcpC promoter	Unknown
DMi7-21-14 (GFP-)	1,021(+1)	Insertion of 1 bp (A).	FcpC promoter	Unknown
	8,262-8,263	Deletion of 2 bp.	OriV	Unknown
DMi7-21-15 (GFP-)	390-1,236	Deletion of 847 bp.	FcpC promoter and 40SRPS8 promoter	The expression of <i>Sh ble</i> could be affected by the loss of its terminator. The expression of the cassette <i>eGFP-T2A-ech-fcs</i> could also be affected by the partial loss of its promoter.

Clone	Reference position	Description	Features affected	Effect
	5,011-8,054	Rearrangement. The sequence of reference that was 3,044 bp long has been replaced by a sequence of 4,670 bp. This rearranged sequence contains a duplicate of the partial <i>eGFP-T2A-ech-fcs</i> sequence, FcpA terminator, OriT cassette, and Pt region 1 sequences (reference positions 1,741-6,198). Compared to the reference sequence, this copy contains a SNV (T→C) at the original position 2,003 that would cause a nonsynonymous substitution (L179P) if the ORF of the <i>eGFP-T2A-ech-fcs</i> cassette would have been translated. The remaining part of the rearranged sequence is a partial and inverted duplicate of the <i>sopC</i> and <i>cos</i> sequences (reference positions 14,555-14,766).	OriT cassette, Pt region 1 and region 2	Based on the sequencing, the cassette <i>eGFP-T2A-ech-fcs</i> is unaltered. The partial duplicate of this cassette should not be expressed due to its upstream sequence that is the FcpA terminator.
DMi7-21-16 (GFP-)	5,012-5,391	Deletion of 380 bp.	OriT cassette	The bacterial conjugation of this plasmid will not be possible.
	14,173-14,285	Deletion of 113 bp	<i>sopC</i>	Unknown
	18,227-18,234	Deletion of 8 bp	FcpC promoter	Unknown

Clone	Reference position	Description	Features affected	Effect
DMi7-31-1 (GFP+)	3,171(+1)	Insertion of 1 bp (A) after the first 12 bp of the gene <i>fcs</i> .	<i>fcs</i>	There is a frameshift that changes the ORF of the <i>fcs</i> gene. The protein <i>fcs</i> cannot be produced.
	7,538	Transition of 1 bp (C→A).	Pt region 1	Unknown
	10,022-10,146	Deletion of 125 bp.	<i>incC</i>	Unknown
	10,147-13,295	Inversion of the 3,779 bp sequence.	<i>incC</i> , CEN6-ARSH4-HIS3 cassette, <i>sopA</i> and <i>sopB</i>	Unknown
	13,926-14,147	Deletion of 222 bp.	<i>sopB</i> and <i>sopC</i>	Unknown
DMi7-31-3 (GFP+)	3,3365	Transition of 1 bp (A->T).	<i>fcs</i>	The SNV would have causes a nonsynonymous substitution (E69V) in the <i>fcs</i> protein
	3,366-8,238	Deletion of 4,873 bp.	<i>fcs</i> , FcpA terminator, OriT cassette, Pt region 1 and 2, and OriV	The <i>fcs</i> protein cannot be produced. Bacterial conjugation with this plasmid is impossible.
DMi7-31-4 (GFP+)	3,722	Transition of 1 bp (A->C).	<i>fcs</i>	The SNV causes a nonsynonymous substitution (E188A) in the <i>fcs</i> protein
	4,807-5,438	Deletion of 632 bp.	FcpA terminator and OriT cassette	The bacterial conjugation of this plasmid will not be possible.
	13,897-14,452	Deletion of the 466 bp sequence and insertion of a partial duplication of the <i>GFP</i> sequence (reference positions 1,971-2,148)	<i>sopB</i> and <i>sopC</i>	Unknown

Clone	Reference position	Description	Features affected	Effect
	14,453-14,921	Inversion of the 469 bp sequence.	sopC and cos	Unknown
	14,922-15,190	Deletion of the 269 bp sequence and insertion of a partial duplication of the <i>lacZ</i> cassette (reference positions 15,645-15,800)	cos	Unknown
DMi7-31-8 (GFP+)	2,918-4,011	Deletion of 1,094 bp.	<i>ech</i> , linker (GGGGS) ₃ and <i>fcs</i>	The unaltered <i>ech</i> and <i>fcs</i> proteins cannot be produced.
	4,012-4,166	Inversion of the 155 bp sequence.	<i>fcs</i>	The <i>fcs</i> protein cannot be produced.
	4,167-5,469	Deletion of 1,303 bp.	<i>fcs</i> , FcpA terminator and OriT cassette	The unaltered <i>fcs</i> protein cannot be produced. The bacterial conjugation of this plasmid will not be possible.
	8,439(+1)	Insertion of a 1,450 bp sequence that contains a duplicate of partial <i>oriV</i> (reference positions 8,440-8,518) followed by a duplicate of partial Pt region 1, complete Pt region 2 and partial <i>oriV</i> sequence (reference positions 7,343-8,713).	<i>oriV</i>	The alteration of <i>OriV</i> can inhibit the increase of copy number of the plasmid by induction.
	14,453(+1)	Insertion of 1 bp (G)	sopC	Unknown
DMi7-31-10 (GFP-)	13,660-15,623	Deletion of 1,303 bp.	sopB, sopC, cos, loxP and <i>lacZ</i> cassette	Unknown

Clone	Reference position	Description	Features affected	Effect
Mutation found in pDMi7 and pControl	839-840	Due to an error in the design of the primer D634F used in the Gibson assembly, there is a loss of the following 80 bp that should have been between reference positions 839 and 840: CCAAGTGCCTAGCGTGCAC ATTGCATGGTTAGACCTCCT TGACGACTGTGAG	FcpC terminator	Unknown

Table B.4 List of oligonucleotides used in this study.

Name	Sequence (5' to 3') (Priming sequence for PCR)	Description	Use
Prom+GFP+T2A_F	CGTTGATCTTGCACCGAAGGAATCAGAGAATAGAATACC <u>ATGGTGAGCAAGGGCGAGGAG</u>	Forward for <i>eGFP</i> with 40SRPS8 promoter homology	Gibson assembly of pDMi7
Prom+GFP+T2A_R	TTCCGTTCCGACTCGTCCGTTTCCGACGGCGGTCGACAT <u>GGGTCCGGGGTTTTCTTCGAC</u>	Reverse for <i>eGFP</i> and <i>T2A</i> with <i>ech</i> homology	Gibson assembly of pDMi7
T2A+ech+linker_F	CTCCTCACCTGCGGAGACGTGCAAGAAAACCCCGACCC <u>ATGTCGACCGCCGTCGGAAAC</u>	Forward for <i>ech</i> with T2A homology	Gibson assembly of pDMi7
T2A+ech+linker_R	CTCCACCTCCGGATCCACCTCCTCCGGAACCTCCACCTCC <u>TTTTCCGGATCAAAGGCCGACA</u>	Reverse for <i>ech</i> and linker (GGGS) ₃ homology	Gibson assembly of pDMi7
Linker+fcs+term_F	TGGAGGTTCCGGAGGAGGTGGATCCGGAGGTGGAGGTTCC <u>ATGCGTAACCAAGGCTTAGG</u>	Forward for <i>fcs</i> with (GGGS) ₃ homology	Gibson assembly of pDMi7
Linker+fcs+term_R	GAAAGTGTCCAGCCAAAGTCGAGGTAGTTGTTGCGG <u>TTACAAATCTTCTTCGGAAATCAATT TTTGTTCAAC</u>	Reverse for <i>fcs</i> with FcpA terminator homology	Gibson assembly of pDMi7
D634F*	TTGCATGGTTAGACCTCCTTGACGACTGTGAGCCTACATCCTTCTGCAACAAGCTGCAAT <u>CCCT GCGATAGACCTTTTCC</u> *	Forward for 40SRPS8 promoter with FcpC terminator homology	Gibson assembly of pDMi7
Prom+GFP_R	GCACCACCCCGGTGAACAGCTCCTCGCCCTTGCTCACCAT <u>GGTATTCTATTCTCTGATTC</u>	Reverse for 40SRPS8 promoter with <i>eGFP</i> homology	Gibson assembly of pDMi7
fcs+term_F	ATTTGGTGAACAAAAATTGATTTCGAAGAAGATTGTA <u>ACCGCAACAACCTACCTCGACT</u>	Forward for FcpA terminator with <i>fcs</i> homology	Gibson assembly of pDMi7
D630R	<u>TTTGCAAACCAAGTTCGACAACCTGCGTACGGCCTGTTGCAAAGATCTACCACCGCTCTGG</u>	Reverse primer in partial HIS3 ORF. Used to amplify half of the backbone.	Gibson assembly of pDMi7 and pControl
D631F	<u>GGCCCGTGGTGGAGTAAAAAGGTTTGGATCAGGATTGCGCCTTTGGATGAGGCACTTT</u>	Forward primer in partial HIS3 ORF. Used to amplify half of the backbone.	Gibson assembly of pDMi7 and pControl
D634R	ATACTTCTACCCCTTTGTTTCTTGGACTGCGTGAGTTTGAAAAAGGTCTATCGCAGGG <u>ATTGCAAGCTTGTGACAGCTCGTCCATG</u>	Reverse for FcpC terminator with 40SRPS8 promoter homology	Gibson assembly of pDMi7
Ctl_FcpAterm_F	CATGGACGAGCTGTACAAGTAA <u>CCGCAACAACCTACCTCGACTTTGGCTGGGACAC</u>	Forward for FcpA terminator with mCherry homology	Gibson assembly of pControl
DMI7_Ctl_R	GAAAGTGTCCAGCCAAAGTCGAGGTAGTTGTTGCGG <u>TTACTTGTACAGCTCGTCCATG</u>	Reverse for mCherry with FcpA terminator homology	Gibson assembly of pControl
mCherry7_F	<u>GATCCGGAGAAGGACGTGGATCCCTCCTCACCTGCGGAGACGTGCAAGAAAACC</u>	Forward for mCherry with T2A sequence homology	Gibson assembly of pControl
GFP7_R	<u>GGTTTTCTTCGACGTCCTCCGACGGTGAGGAGGGATCCACGTCCTTCTCCGGATC</u>	Reverse for GFP with T2A sequence homology	Gibson assembly of pControl

*Note: As mentioned in the table S3, the D634F primer had a mistake in the design that caused the loss of 80 nucleotides close to the end of the FcpC terminator from *Sh Ble* cassette. The sequence of D634F should have been instead: AATCAAGCTAGTATTGCAGCTTGTGTCAGAAGCCTACATCCTTCTGCAACAAGCTGCAATCCCTGCGATAGACCTTTTCC

Table B.5 List of plasmids used in this study.

Plasmid	Description	Reference
pPtGE19	pPtPuc3 with 40SRPS8 Promoter and Terminator driving <i>Sh ble</i>	(Slattery et al., 2018)
pPtGE30	<i>P. tricornutum</i> expression vector, FcpD promoter and FcpA terminator driving <i>Sh ble</i>	(Slattery et al., 2018)
pPtGE33	<i>P. tricornutum</i> expression vector, FcpD promoter and FcpA terminator driving eGFP and mCherry linked by a T2A self cleaving peptide linker	(Slattery et al., 2018)
pDMi7	<i>P. tricornutum</i> expression vector, 40SRPS8 promoter and FcpA terminator driving eGFP linked with a T2A self-cleaving peptide linker to the enzyme ech fused with the enzyme fcs by a flexible linker (GGGGS) ₃ .	This study
pControl	<i>P. tricornutum</i> expression vector, 40SRPS8 promoter and FcpA terminator driving eGFP and mCherry linked by a T2A self cleaving peptide linker	This study
pTA-Mob	Conjugative plasmid that contains the genes to produce the proteins that will establish the bacterial conjugation for the transformation. These proteins will be essential for the formation of the physical link for the transfer of the cargo plasmid of interest from <i>E. coli</i> to <i>P. tricornutum</i> cells.	(Strand et al., 2014)

Table B.6 List of genetic parts used in this study.

Name	Sequence	Description
GFP	ATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCTGGTCGAGCTGGA CGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCT ACGGCAAGCTGACCCTGAAGTTCATCTGCACCACGGCAAGCTGCCCGTGCCTGGCCCA CCCTCGTGACCACCTGACCTACGGCGTGAGTGTTCAGCCGCTACCCCGACCACATGA AGCAGCAGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCT TCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACC CTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGG GCACAAGCTGGAGTACAACACAAGACGGCCACAACGTCTATATCATGGCCGACAAGCAGA AGAACGGCATCAAGGTGAACCTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAG CTCGCCGACCACTACCAGCAGAACACCCCATCGGCGACGGCCCGTGTGCTGCCCGAC AACCCTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCA CATGGTCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTA CAAGGAATTC	Enhanced green fluorescent protein
T2A	GGATCCGGAGAAGGACGTGGATCCCTCCTCACCTGCGGAGACGTGAAGAAAACCCCGG ACCC	Self-cleaving <i>Thosea asigna</i> virus 2A peptide from pPtGE33 (Slattery et al., 2018)
mCherry	GGACCGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTT CAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCG AGGGCCGCCCCACGAGGGGACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCC CTGCCCTTCGCTGGACATCTGTCCCTCAGTTCATGTACGGCTCAAGGCCATCGTGA AGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCTTCCCCGAGGGCTTCAAGTGGG AGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTCC AGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCTCCGACGGCC CCGTAATGCAGAAGAAAACCATGGGCTGGGAGGCCCTCCTCCGAGCGGATGTACCCCGAG GACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTA CGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGACGCTGCCCGGCCCT ACAACGTCAACATCAAGTTGGACATCACCTCCACAACGAGGACTACACCATCGTGGAAC AGTACGAACGCGCCGAGGGCCGCCACTCCACCGCGGCATGGACGAGCTGTACAAG	mCherry fluorescent protein
ech	ATGTCGACCGCCGTCGGAACGGACGAGTCCGAACGGAACCGTGGGGAGAAAACGGTCTT GGTTGAATTTGATGAAGGAATTGCCTGGGTGATGTTGAACCGACCGGATAAAACGAAAACGC CATGAACCCAACGTTGAACGATGAAAATGGTCCGAGTTTTGGATCACTTGAAGGAGATGA TCGATGTCGTGTTTTGGTTTTGACGGGTGCCGGAGAATCCTTTTCCGCCGGTATGGATTTG AAAGAATACTTTTCGTGAAGTTGATGCCACGGGATCCACCGCCGTCAAAATTAAGTGCGA CGCGCCTCCGCTGAATGGCAGTGGAAAACGATTGGCCAAGTGGAGTAAACCCACGATTGCT ATGGTGAACGGATGGTGTGTTTTGGAGGAGCCTTACCCCATGGTTCGCTTGTGATTTGGCTT TTGCTGATGAAGATGCCCCGATTTGGATTGTCGGAAGTAACTGGGGAATTCCCCAGGAG GAGTGGTTTTCGCGAGCTTTGGCCGCCACGGTCCCGCAACGTGATGCCTTGTACTACATTAT GACCGGAGAACCCTTTGATGGACGACGTGCCGCCGAAATGCGATTGGTCAACGAAGCCTT GCCCCGCTGATCGCTTGCGCGAACGAACCCGAGAAGTGCCTTGAATTTGGCCTCGATGAA CCAAGTTGTTTTGCACGCCGCCAAAACGGGATACAAGATTGCCAAGAAATGCCATGGG AACAAGCCGAAGATTACTTGTACGCCAAAATTGGATCAATCGCAATTTGCCGATAAAGCTG GTGCCCGAGCCAAAGGTTTGACGCAATTTTTGGATCAAAAATCGTACCGACCGGGTTTGT CGGCCTTTGATCCGGAAAAA	Codon optimized Enoyl-CoA hydratase ORF (Sequence of origin: GenBank accession # KC847406.1)
fcs	ATGCGTAACCAAGGCTTAGGTAGTTGGCCCGTGGAGCGCGCACGAATGAGTCCACATGCC ACGGCCGTTAGGCACGGAGGCACCGCATTGACCTACGCTGAACTAAGTAGGCGAGTAGC CCGTCTAGCTAACGGGCTCCGGGCGGGCGGCGTAAGACCGGGTGACCGTGTAGCGTACTT AGGTCTAATCACCCAGCATATCTTGAGACATTGTTTCGCTTGTGGTCAAGCCGGAGCAGT GTTTGTTCGGTAAATTTTACTAGTGGTTCCTGAACTTGACCACGCACTAGCTGATAGT GGCGCCTCTGTGCTGATCCACACTCCTGAGCAGCTGAAACAGTAGCGGCCCTAGCAGCC GGACGTTTATTGAGAGTACCGGCCGGTGAACCTGACGCAGCTGATGATGAACCCCGGAC CTGCCGGTAGGACTAGATGACGTGTGCCTATTGATGTACACCTCAGGTAGCACAGGTAGA CCTAAGGGAGCGATGCTGACGCACGGTAACTTAACATGGAATTGCGTAAACGTATTAGTT GAAACTGATCTAGCCTCTGATGAGCGTGCCTTAGTAGCGGCACCTTTTTCCACGGCGCC GCCTAGGAATGGTGTGTTTTGCCACCTTGTGAAAGGAGGAACGGTCAATTTGCACAGC	Codon optimized Feruloyl-CoA synthetase ORF (Sequence of origin: GenBank accession # KC847405.1)

Name	Sequence	Description
	GCCTTTGATCCCGGAGCCGTCTTGAGTGCCGTGGAACAAGAACGTGTGACGTTGGTGT GGTGTCCCAACCATGTACCAAGCCATTGCCGCTCACCCCGCTGGCGGTCCGCCGATCTC TCGTCCTTGCGAACGTTGTTGTGTGGCGGAGCCCCAGTGCCCGCGATTGGCGTCGCGTT ACTTGGATCGTGGATTGGCTTTTGTCCAAGGATACGGTATGACGGAAGCCGCGCTGGAG TGTTGGTCTTGGACCGCGCCACGTCGCCGAAAAAATTGGTTCTGCCGGCGTCCCGTCTT TTACCGATGTCCGCTTGGCCGACCTCAGGAGAACCGTTCCACCCGGTGAAAAAGG TGAAATTGTCGTTTCGGGTCCCAATGTCATGAAAGGATACTGGGGACGACCAGAAGCCAC GGCAGAAGTTCTCCGTGATGGATGGTTTCACTCGGGTGACGTCGCCACCGTTGATGGAGA CGGATACTTTCACGTCGTCGATCGATTGAAGGATATGATTATTTCCGGAGGAGAAAAAT CTACCCAGCCGAAGTTGAAAAACGAATTGTACGGCTACCCCGGAGTGGAAGCCTGTGCCGT CATTGGAGTCCAGATCCTCGTTGGGGAGAAGTCGGAAGCCGTCGTCGTGCCAGCTGA TGGATCGCAATTGATGGAGACGAATTGTTGGCATGGTTGCGAACCCGATTGGCTGGTTA CAAAGTTCCCAAATCCGTAGAAATTCACAGATCGTTTGCCACGACGGGTTCCGGAAAAAT TTTGAAAGGTGAAGTCCGACGACGATTGGT	
40SRPS8 Promoter	CCCTGCGATAGACCTTTTCCAAACTCACGCAGTCCAAGAAAACAAAGGGTGAGAAGTA TACGCACCTTTCGGTTTCGGCATAATTCTTAAACTCTTGTGGTCACTTTCTTGTGAAGAAG CTAGGGGCACTCGTTTTCCCTCAGAGCCTGCAAACACAAAATTCTGCAGTCAATTGTCC CAACACTCGCAAGCCGTATGCGCAAGCAACGATGCGCAGAAGGCCGTGGATGGATGGC GACTCGCGATATGGCTTCTTGGGTGCGCCAGTGTGGTATGTCCGGCGTATGTCAATACGCG AATTCGGACGACTGGCATCTCTAGGAGGAGGATTCCTTCTTTATGACATGTTTATTTAT ATACATTGATGCTTTCGACAGTCGGAAGTAATAAATGAATTTATTTCAAGACTACCTAT ACTCCTTTGACTTGTTCGACTAATCTTACCGCTTACTAAAATCTCGAAATCACGTTGACC TCTCGCACGAAAATTTTGTGCTGGACGCTACGCACTCGGCCAAATTCTTCTCGGTCTC GTCGTCGCAATTGTCGTTGCGTTGATCTTGCACCGAAGGAATCAGAGAATAGAATACC	40SRPS8 Promoter
FcpA Terminator	CCGCAACAACCTACCTCGACTTTGGCTGGGACACTTTCAGTGAGGACAAGAAGCTTCAGAA GCGTGCTATCGAACTC AACCAGGACGTGCGGCACAAATGGGCATCCTTGCTCTCATGGTGCACGAACAGTTGGG AGTCTCTATCCTTCTT AAAAAATTTAATTTTCATTAGTTGCAGTCACTCCGCTTTGGTTT	Fucoxanthin-chlorophyll a/c binding protein A Promoter