THÈSE DE DOCTORAT

NNT : 2024UPASL079

université PARIS-SACLAY

Optimisation de la fermentation de la biomasse végétale par *Lachnoclostridium phytofermentans*

Optimization of plant biomass fermentation by Lachnoclostridium phytofermentans

Thèse de doctorat de l'Université Paris-Saclay

École doctorale n° 577, Structure et Dynamique des Systèmes Vivants (SDSV)
Spécialité de doctorat : Biotechnologies
Graduate School : Sciences de la Vie et de la Santé
Référent : Université d'Évry Val d'Essonne

Thèse préparée dans l'unité de recherche
Génomique métabolique (Université Paris-Saclay, Univ Évry, CNRS, CEA)
Sous la direction de **Véronique DE BERARDINIS**, Directrice de recherche,
Et la co-direction de **Andrew TOLONEN**, Chargé de recherche.

Thèse soutenue à Paris-Saclay, le 28 Novembre 2024, par

Tom ZAPLANA

Composition du Jury

Membres du jury avec voix délibérative

Olga SOUTOURINA Professeure, Université Paris-Saclay	Présidente
Emmanuel GUEDON Directeur de recherche HDR, CNRS, Université de Lorraine	Rapporteur
Isabelle MEYNIAL-SALLES Professeure, INSA Toulouse	Rapportrice
Henri-Pierre FIEROBE Directeur de recherche, CNRS, Université Aix-Marseille	Examinateur
Nicolas LOPES FERREIRA Chargé de mission, Direction Générale – IFPEN	Examinateur
Olivier TENAILLON Directeur de recherche, INSERM, Institut Cochin, Université Paris-Cité	Examinateur

À ma mère

REMERCIEMENTS

Je tiens tout d'abord à exprimer ma gratitude envers les personnes avec qui j'ai eu l'opportunité de travailler durant ce doctorat. Je remercie Véronique De Bérardinis d'avoir rendu ce projet de thèse possible et Andrew Tolonen pour avoir cru en moi et pour avoir contribué à ma formation scientifique en m'ayant permis de me construire scientifiquement et professionnellement. Je suis également reconnaissant envers les membres de mon comité de suivi individuel, Henri-Pierre Fiérobe, François Wasels et Ioana Popescu pour leurs remarques constructives et leur disponibilité à chaque étape de ce projet. De plus, j'aimerais remercier les membres du jury d'avoir accepté de rapporter et d'examiner mon travail.

Je tiens à remercier tout particulièrement Ioana pour m'avoir permis de faire mes premiers pas dans un laboratoire et plus largement dans le domaine de la biologie synthétique et d'avoir fait naitre en moi le goût de l'ingénierie métabolique, que ce soit avec les cours à l'université ou durant iGEM ainsi que pour toutes les discussions stimulantes.

Dans l'équipe, j'aimais remercier, le support technique de Magali Boutard et Loïc Couturat ainsi que les différentes discussion et réflexion que j'ai pu avoir avec Solange Miele durant le second tiers de mon doctorat. J'aimerais aussi remercier les « nouveaux » (plus tant nouveaux) Drice Challal, Mathieu Cassarini, Naïda Eyogo et Soumaya Najah que j'ai pu former et interagir avec durant la fin de mon doctorat sans qui cette aventure scientifique n'aurait pas été la même. Plus largement j'aimerais remercier le reste des personnes du premier étage du Genoscope avec qui j'ai pu interagir.

Durant cette thèse j'ai pu réaliser l'activité de doctorant-conseil au sein du GIP-Genopole et je tiens à remercier très chaleureusement Véronique Le Boulch et Julien Picot avec qui j'ai pu travailler dans un projet d'ampleur pour lequel j'ai apprécié participer.

Sur le plan personnel, je veux absolument remercier mes amis d'iGEM, Cécile, Suzanne et Antoine et de l'IBGBI, Dana, Sabrina, Léa, Mehdi et Michael pour tous les moments de rigolade que j'ai pu passer avec vous.

Je tiens tout d'abord à exprimer ma profonde gratitude envers mon père, mon frère Quentin, et Sarah. Sans leur présence, leur soutien indéfectible et leurs encouragements constants, je n'aurais pas pu surmonter les moments difficiles de ces trois dernières années.

À Tara, mes mots ne pourront jamais être à la hauteur de ma reconnaissance. Faire nos thèses ensemble, dans deux domaines distincts de la biologie, a été une aventure aussi enrichissante qu'inoubliable, tant sur le plan scientifique que personnel. Depuis notre rencontre, tu as toujours été un pilier de soutien inestimable, me poussant à donner le meilleur de moi-même. Merci pour tout ce que tu fais, mais surtout pour tout ce que tu es.

Enfin, je voudrais rendre hommage à ma mère, dont la mémoire et les valeurs continuent de m'inspirer chaque jour. Elle reste présente dans mon cœur et dans tout ce que j'entreprends.

Un dernier remerciement spécial à Solange, Soumaya et Tara, qui ont pris le temps de relire cette thèse. Sans leurs commentaires et leurs corrections constructives, ce manuscrit serait bien différent.

TABLE DES MATIÈRES

REMERCIEMENTS *5*

TABLE DES MATIÈRES *7*

LISTE DES FIGURES *10*

LISTE DES TABLEAUX *11*

LISTE DES ABBRÉVIATIONS *12*

INTRODUCTION *14*

1 Caractéristiques des Clostridia **14**

1.1. Paroi cellulaire 14

1.2. Sporulation 16

1.3. Métabolisme 18

1.4. Phylogénie 20

1.5. Clostridia commensaux et pathogènes 21

1.6. Clostridia à intérêt industriels 23

2. Produits d'intérêts **41**

2.1. Applications dans le domaine de la santé 41

2.2. Biocarburants 42

3. Ingénierie des *Clostridia* **63**

3.1. Ingénierie non-rationnelle 63

3.2. Ingénierie rationnelle et biologie synthétique 68

OBJECTIFS DE LA THÈSE *87*

MATÉRIELS & MÉTHODES *89*

1. Souches bactériennes et conditions de culture **89**

1.1. Culture en bioréacteur 90

1.2. Évolution dirigée en GM3 anaérobie 91

1.3. Analyse de la croissance de L. phytofermentans 91

1.4. Quantification de la dégradation de la cellulose 92

2. Techniques de biologie moléculaire et génie génétique **93**

2.1. Plasmides utilisés dans cette étude 93

2.2. Méthode d'amplification et de clonage 95

2.3. Transformation bactérienne 96

3. Approche de biologie synthétique **96**

3.1. Adaptation et optimisation du biosenseur AlkS 96
3.2. Construction des voies de biosynthèses du butanol 97
3.3. Clonage de guides multiplexés dans pQdC12a 98

4. Tests fonctionnels et analyse d'expression génique 99

4.1. Quantification de la fluorescence et traitement des données 99
4.2. Mesure de la luminescence 100
4.3. Mesure de l'expression des ARNm par qRT-PCR 100
4.4. Répression métabolique par CRISPRi 101

5. Technique de chimie analytique 101

5.1. Quantification des produits de fermentation par HPLC 101
5.2. Analyse des acides gras cellulaires 102

6. Analyses génomiques et bio-informatiques 102

6.1. Séquençage génomique 102
6.2. Outils bio-informatiques 103

RÉSULTATS* *104

1. Évolution de *L. phytofermentans* pour la résistance au butanol 104

1.1. Processus d'évolution dirigé et sélection 105
1.2. Sélection des souches résistantes 107
1.3. Caractérisation génotypique par séquençage génomique 107
1.4. Caractérisation phénotypique de la souche évoluée 112
1.5. Discussion et Perspectives 127

2. Synthèse *de novo* du butanol chez *L. phytofermentans* 130

2.1. Construction d'une voie de combinatoire de biosynthèse du butanol 131
2.2. Quantification du butanol par chimie analytique et détection in vivo 136
2.3. Discussion et Perspectives 140

3. Régulation génique et prototypage métabolique par CRISPRi 143

3.1. Preuve de concept du système CRISPRi chez L. phytofermentans 144
3.2. Amélioration de la stratégie de clonage de guide pour le multiplexage 146
3.3. Design et stratégie expérimentale 149
3.4. Répression métabolique par CRISPRi 152
3.5. Discussion et Perspectives 157

CONCLUSION* *159

RÉFÉRENCES* *162

ANNEXES* *180

LISTE DES FIGURES

Figure 1 : Schéma de la paroi cellulaire des bactéries Gram positif. 16
Figure 2 : Cycle de vie d'un Clostridia 17
Figure 3 : Arbre phylogénétique de Lachnoclostridium phytofermentans et de quelques taxons de la Classe Clostridia basés sur des séquences d'ARNr 16S. 21
Figure 4 : Le métabolisme des Clostridia. 26
Figure 5 : Paradigme du cellulosome de Clostridium thermocellum. 29
Figure 6 : Optimisation des Designer Cellulosomes pour la dégradation de Galactomannan. 32
Figure 7 : Croissance, fermentation et morphologie cellulaire de L. phytofermentans sur différentes sources de carbone. 34
Figure 8 : Modèle de l'import et du métabolisme des carbohydrate chez L. phytofermentans. 36
Figure 9 : Métabolisme central de L. phytofermentans. 37
Figure 10 : Image de microscopie électronique à transmission de L. phytofermentans. 39
Figure 11 : Évolution de la production de bioéthanol par continent entre 2006 et 2019. 43
Figure 12 : Les générations de biocarburants. 49
Figure 13 : Les deux voies de production de biocarburants à partir de biomasse. 52
Figure 14 : Structure et composition de la lignocellulose de la paroi d'une cellule végétale. 54
Figure 15 : Transformation par voie enzymatique de la cellulose cristalline en glucose. 58
Figure 16 : Les 4 technologies de traitement de la biomasse lignocellulosique. 60
Figure 17 : Processus de production de biocarburant par l'entreprise Lanzatech. 62
Figure 18 : Dispositifs utilisés pour l'évolution adaptative en laboratoire (ALE). 65
Figure 19 : Schéma de culture dans un GM3. 67
Figure 20 : Le cycle DBTL 69
Figure 21 : Métabolisme de la fermentation ABE 72
Figure 22 : Co-culture pour la production de butanol à partir de cellulose par CBP. 77
Figure 23 : Maturation du pré-ARNg de la Cas9. 79
Figure 24 : Maturation du pré-ARNcr de la Cas12a. 82
Figure 25 : Fonctionnement du système CRISPR interference. 83
Figure 26 : Application du CRISPR Base Editing. 85
Figure 27 : Temps de génération de L. phytofermentans durant l'évolution en GM3. 106
Figure 28 : Alignement génomique fait avec le logiciel BreSeq. 109
Figure 29 : Visualisation des mutations dans les gènes Cphy_2809, 2380 et 0333. 110
Figure 30 : Courbe de croissance de L. phytofermentans WT et PHY24.0#6 en présence de butanol. 115
Figure 31 : Courbes de croissance de L. phytofermentans en présence d'alcools primairs. 117
Figure 32 : Métabolisme cellulolytique de L. phytofermentans WT et PHY24. 120
Figure 33 : Profils de fermentation de L. phytofermentans en bioréacteur. 122
Figure 34 : Abondance relative des acides gras cellulaires. 125
Figure 35 : Voie de synthèse optimisée pour la production de butanol. 133

Figure 36 : Stratégie de clonage de la voie de synthèse du butanol. 135
Figure 37 : Biosenseur original séparé sur deux plasmides. 139
Figure 38 : Adaptation du biosenseur sur un seul plasmide. 139
Figure 39 : Répression transcriptionnelle du Nanoluc chez L. phytofermentans par CRISPRi. 146
Figure 40 : Amélioration de la plateforme de clonage de guide et multiplexage du CRISPRi. 148
Figure 41 : Répression métabolique effectuer chez L. phytofermentans. 151
Figure 42 : Impact du CRISPRi sur les gènes du métabolisme central de L. phytofermentans. 154

LISTE DES TABLEAUX

Tableau 1 : Principaux métabolismes chez les Clostridia. ____ 19
Tableau 2 : Comparatif des châssis industriels utilisés. ____ 24
Tableau 3 : Comparaison des propriétés des bioalcools et de l'essence. ____ 46
Tableau 4 : Production de biomasse, production d'éthanol et superficie des terres nécessaires pour 100 Mds de Litres de carburants renouvelables aux USA. ____ 51
Tableau 5 : Comparaison récapitulative des caractéristiques de Cas9 et Cas12a ____ 81
Tableau 6 : Souches bactériennes utilisées dans cette thèse ____ 89
Tableau 7 : Milieux de cultures utilisé dans cette thèse ____ 90
Tableau 8 : Plasmides de départs utilisés dans cette thèse ____ 93
Tableau 9 : Plasmides construits dans cette thèse ____ 93
Tableau 10 : Abondance des différents acides gras selon leur état de saturation et leur taille. 126
Tableau 11 : Primers utilisé pour des PCR dans cette thèse ____ 180
Tableau 12 : Primers utilisé pour le clonage de guides CRISPR-Cas12a dans cette thèse ____ 181
Tableau 13 : Mutations intra-géniques de type SNP dans le génome de la souche PHY24.0#6 183

LISTE DES ABBRÉVIATIONS

rCDI : infections récurrentes à *C. difficile*

SCFA : *short-chain fatty acids*

CBM : *Carbohydrate-Binding Module*

ABC *transporters* : ATP synthase (ATP)-binding cassette transporters

BMC : *Bacterial Micro Compartments*

CBM : *Carbohydrate-Binding Module*

RON : *Research Octane Number*

SSF : *Simultaneous Saccharification and Fermentation*

SScF : *Simultaneous Saccharification and Co-Fermentation*

SHF : *Separate Hydrolysis and Fermentation*

CBP : *Consolidated BioProcessing*

GM3 : Génémat 3

DBTL : *Design-Build-Test-Lean*

ABE : Acétone-Butanol-Éthanol

CRISPR : *Clustered Regularly Interspaced Short Palindromic Repeats*

Cas : *CRISPR associated*

ARNg : ARN guide

ARNtracr : ARN CRISPR trans-activateur

ARNcr : ARN CRISPR

DR : *Direct Repeats*

ARNsg : ARN *single guide*

NHEJ : *Homologous End Joining*

HDR : *Homologous Directed Repair*

CRISPRi : *CRISPR-interference*

WT : *Wild Type*

HPLC : *High performance liquid chromatography*

SNP : *Single-nucleotide polymorphism*

ORF : Open Reading Frame

InDel : Insertion-délétions

PTS : *Phosphotransferase system*

PEP : phosphoénolpyruvate

EC50 : Concentration Efficace à 50%

MoClo : *Modular Cloning*

CDS : *Coding DNA Sequences*

RBS : *Ribosome Binding Site*

RFP : *Red Fluorescent Protein*

sfGFP *superfolder Green Fluorescent Protein*

PCR : *Polymerase Chain Reaction*

RT-qPCR : *Reverse transcriptase quantitative PCR*

aTc : anhydrotétracycline

INTRODUCTION

1 Caractéristiques des Clostridia

Les *Clostridia* représentent une *Classe* remarquable de bactéries qui occupent une place significative dans le domaine de la microbiologie. Caractérisées par leur nature anaérobie stricte et leur capacité à former des spores, les *Clostridia* sont largement répandues dans divers environnements, allant des sols aux intestins des humains et des animaux. Cette *Classe* bactérienne, appartenant au *phylum* des *Firmicutes* se distingue également par sa diversité, abritant à la fois des pathogènes redoutables pour l'homme, tels que *Clostridioides difficile* et *Clostridium tetani*, ainsi que des micro-organismes utilisé en industrie d'une grande importance, comme *Clostridium acetobutylicum* et *Clostridium autoethanogenum*. Cette polyvalence biologique fait des *Clostridia* un sujet d'étude fascinant pour comprendre les mécanismes de la pathogenèse, les interactions microbiennes et les applications biotechnologiques innovantes.

1.1. Paroi cellulaire

En 1884, Hans Christian Gram a élaboré une méthode de classification des bactéries basée sur leur paroi, les divisant en deux catégories : Gram-positif et Gram-négatif. Bien que cette observation ait permis de distinguer les types de parois, il est important de souligner que la coloration de Gram d'une bactérie ne correspond pas nécessairement à sa position dans un arbre phylogénétique. En effet, grâce à la phylogénie moderne et au séquençage des ARNr 16S des bactéries, une classification plus précise a été établie, révélant que des membres d'un même *phylum* peuvent présenter des caractéristiques de Gram positif ou négatif [1]. Ces deux types de parois se distinguent sur le plan biochimique et structural. Dans le contexte des procaryotes, la paroi cellulaire représente un élément indispensable à leur survie, qu'ils soient classés comme Gram positifs ou Gram négatifs. Elle forme une barrière entre le milieu intracellulaire et l'environnement extérieur, assurant ainsi la protection des composants cellulaires contre les contraintes physico-chimiques potentiellement néfastes. Toutefois, la fonction de la paroi n'est pas d'être totalement imperméable ; au contraire. Son rôle est de permettre les échanges nécessaires au bon

développement de la bactérie. Les nutriments externes sont importés, tandis que les molécules devant être expulsées de la cellule sont sécrétées, notamment dans le cadre de processus tels que la communication par *quorum sensing*, la production de produits de fermentation toxiques, le maintien de l'équilibre osmotique par le transport d'ions, ou encore la dégradation des polysaccharides par des enzymes.

Comme illustré dans la **Figure 1**, la paroi des bactéries Gram-positif se compose d'une unique membrane cytoplasmique, d'une épaisse couche de peptidoglycane, et, dans de nombreux cas chez les *Clostridia*, d'un *surface layer* [2]. Cette dernière est constituée de protéines monomériques s'auto-assemblant, jouant un rôle crucial dans la survie et la croissance cellulaire. Parmi ses fonctions, on peut citer la présentation d'enzymes, comme l'ancrage du cellulosome chez la majorité des *Clostridia* cellulolytiques.

La membrane plasmique est formée d'une bicouche lipidique composée de phospholipides, assurant le contrôle des échanges entre l'intérieur et l'extérieur de la cellule, avec une perméabilité sélective aux ions et aux petites molécules. La paroi des bactéries Gram-positif contient des acides lipotéichoïques et teichoïques exclusifs, jouant un rôle majeur dans sa robustesse. Les acides lipotéichoïques sont ancrés dans la membrane cytoplasmique via un domaine lipidique, tandis que les acides teichoïques sont liés de manière covalente aux couches de peptidoglycane [3].

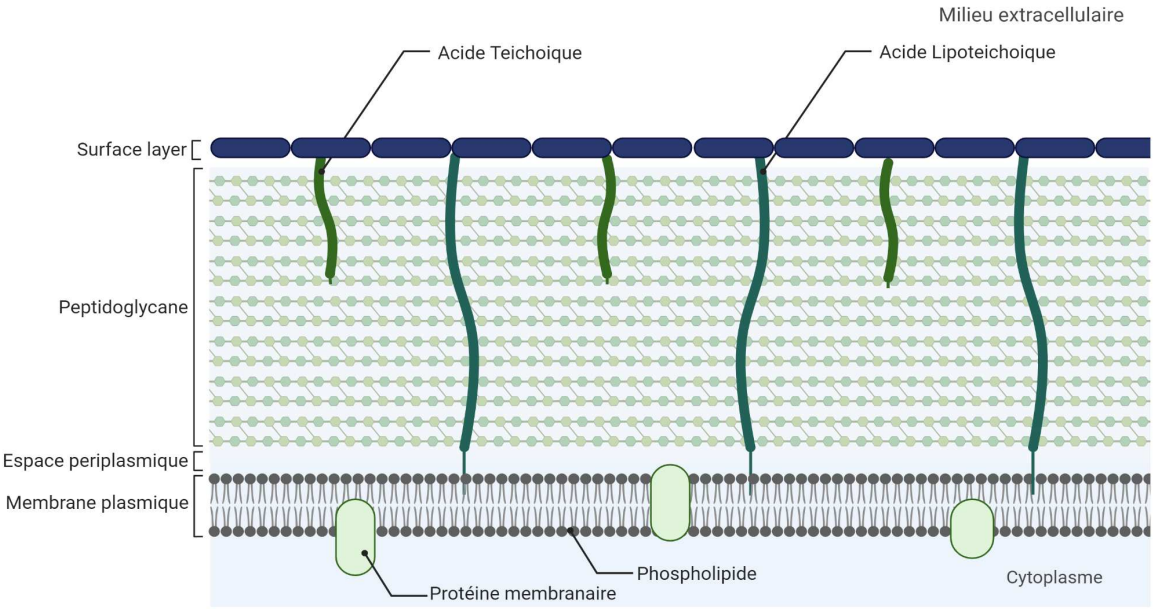


Figure 1 : Schéma de la paroi cellulaire des bactéries Gram positif.
La paroi est composée principalement d'une membrane plasmique, d'un peptidoglycane, d'acides teichoïques et lipoteichoïques, ainsi que d'une couche de proteins de structure appelé surface layer, celui-ci étant facultatif selon les espèces.

1.2. Sporulation

De manière similaire à la paroi cellulaire, la capacité des *Clostridia* à sporuler constitue un mécanisme essentiel qui accroît leur survie et leur adaptation à leur environnement. La sporulation, largement étudiée chez *Bacillus* mais également présente chez de nombreux Firmicutes, y compris les *Clostridia*, est principalement déclenchée par des conditions de privation de nutriments [4].

Le processus de sporulation implique plusieurs mécanismes moléculaires communs (**Figure 2**). Le gène maître *Spo0A*, transcrit par le facteur de transcription σH, puis traduit pour être phosphorylé (Spo0A~P) par des histidine kinases orphelines, variant selon les espèces de *Clostridia*. Spo0A~P initie la sporulation en activant une cascade de facteurs σ en aval (σF, σE, σG et σK), qui régulent et maintiennent l'état de sporulation. Ces mécanismes moléculaires entraînent des changements dans la composition de la paroi et du cytoplasme. Les *Clostridia* produisent de l'acide dipicolinique, qui, en se complexant avec les ions Ca2+, confère une thermorésistance. De plus, les *small acid-soluble proteins* (SASPs) offrent une protection contre la chaleur, les peroxydes et les radiations UV, tout en modifiant la conformation de l'ADN en une hélice A [4,5].

Lorsque les conditions environnementales sont favorables, une spore peut germer pour revenir à l'état de cellule végétative. Ce processus de germination, induit par des composés du milieu extérieur ou des facteurs physico-chimiques, implique la libération d'ions (H^+, K^+, Na^+), suivie de la libération d'acide dipicolinique et de $Ca2^+$. Le cortex de la spore est ensuite dégradé par des enzymes pour permettre l'absorption d'eau et le gonflement de la cellule. Les SASPs sont dégradées par des protéases de germination, libérant ainsi des acides aminés qui servent à la synthèse des protéines et au métabolisme [4].

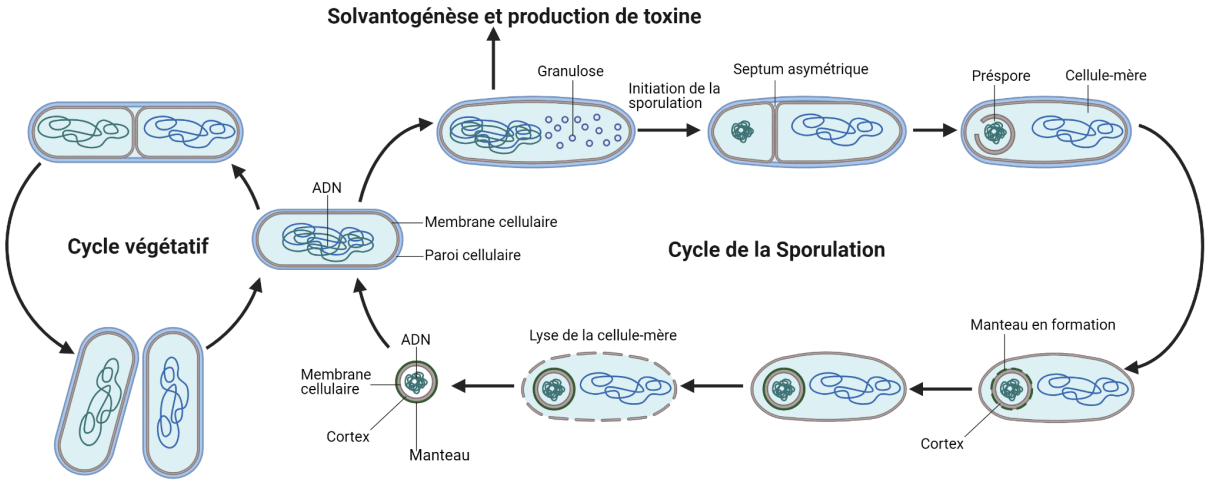


Figure 2 : Cycle de vie d'un Clostridia
En cas de conditions environnementales défavorables, les cellules quittent la phase végétative pour initier la morphogénèse vers la spore. La cellule s'allonge de plus en plus et des vésicules de granuloses s'accumulent, c'est à ce moment que les Clostridia pathogènes sécrètent leur toxine. Les cellules entament ensuite le processus de sporulation, et une division asymétrique se produit, donnant naissance aux compartiments de la cellule mère et de la pré-spore. La cellule mère englobe la pré-spore lors de la formation du cortex et de la paroi de la spore. Une fois le processus de maturation de la spore terminée, la cellule mère se lyse pour libérer la spore mature. Une fois des conditions favorables revenues, la spore donne naissance à une nouvelle cellule végétative. Adapté de [4],

1.3. Métabolisme

En raison de la variété des habitats dans lesquels ils prospèrent, tels que les tissus animaux, le sol et l'eau marine [6], les *Clostridia* exploitent une gamme tout aussi diversifiée de nutriments. Certaines espèces se spécialisent dans l'utilisation de types spécifiques de substrats ; par exemple, *Lachnoclostridium phytofermentans* se distingue par sa capacité à dégrader la biomasse végétale [7], tandis que d'autres, comme *Clostridium ljungdahlii*, sont capables de fixer le monoxyde et le dioxyde de carbone [8]. Les *Clostridia* se distinguent les uns des autres par leur métabolisme, qui peut reposer sur une grande variété de substrats, comme décrit dans le **Tableau 1**. Cette représentation ne couvre pas tous les métabolismes du genre *Clostridium* ; les métabolismes d'intérêt pour lesquels les *Clostridia* sont largement étudiés, tels que les métabolismes solvantogénique, acidogénique, *Wood-Ljungdahlii* et cellulolytique, seront plus amplement développés ultérieurement.

La diversité des métabolismes soutient l'idée générale selon laquelle les *Clostridia* sont des organismes hautement adaptés à leur environnement. Leur évolution est attestée par la diversité des espèces, et leur classification phylogénétique a été révisée à plusieurs reprises par le passé, comme nous le verrons dans la section suivante.

Tableau 1 : Principaux métabolismes chez les Clostridia.
Adapté de [9]

Métabolismes	**Substrat(s)**	**Produits**	**Espèce type**
Cellulolytique	Sucres, polysaccharides	Éthanol, Acétate, H_2, CO_2, Lactate	*L. phytofermentans*
Wood-Ljungdahl	CO_2, CO	Éthanol, Acétate, H_2	*C. ljundahlii*
Propionate	Lactate	Propionate, Acétate, CO_2	*C. propionicum*
Propionate	Succinate	Butyrate, Acétate, CO_2, H_2	*C. mayombei*
Butyrate	Sucres	Butanol, Acétone	*C. butyricum*
Acétone/Butanol	Sucres	CO_2, H_2, Butyrate, Acétate	*C. acetobutylicum*
Alcolotrophique	Alcools et/ou Acides	Divers acides et solvants	*C. kluyveri*
Protéolytique	Protéines, peptides, acides aminés	Acétate, Propionate, Butyrate, CO_2, NH_4^+	*C. propionicum*
Réaction Stickland	Acides aminés	Acétate, CO_2, NH_4^+	*C. sticklandii*
Purinolytique	Purines	Acétate, formate, CO_2, NH_4^+	*C. purinilyticum*
Pyrimidinolytique	Pyrimidines	Acides, CO_2, NH_4^+	*C. oroticum*

1.4. Phylogénie

Dans cette exploration de la phylogénie des *Clostridia*, nous découvrirons que les *Clostridia* pathogènes et industriels sont étroitement liés sur le plan phylogénétique, et l'étude des uns peuvent nous éclairer sur les autres. Depuis la classification établie par *Collins et al. 1994*, qui a ordonné le genre *Clostridium* en 19 clusters, l'analyse phylogénétique basée sur le séquençage des ARNr 16S a révélé que seuls les organismes appartenant au cluster I devraient être classés dans le genre *Clostridium*. Cette constatation suggère que de nombreux organismes ont été incorrectement classés par le passé, soulignant la nécessité d'une reclassification[10].

Avec la publication de l'étude de *Yutin et al. 2013*, une nouvelle taxonomie a pu être établie en comparant un arbre phylogénétique basé sur un ensemble concaténé de 50 protéines ribosomales largement répandues avec les arbres phylogénétiques basés sur la sous-unité bêta de l'ARN polymérase (*RpoB*), l'ADN gyrase (*GyrB*) et l'ARN ribosomique 16S. Les résultats obtenus par cette méthode ont montré une concordance entre les trois arbres, ce qui a permis à Yutin et Galperin de reclasser de nombreuses espèces qui avaient été mal positionnées dans la phylogénie. Cette nouvelle taxonomie a remplacé l'ancienne dans les bases de données du NCBI en 2013, entraînant la séparation de certaines espèces de *Clostridium* entre différentes familles tout en maintenant leur appartenance à l'ordre *Eubacteriales* [11].

Depuis cette révision taxonomique, de nombreuses espèces ont vu leur désignation changer (*e.g Lachnoclostridium phytofermentans* anciennement *Clostridium phytofermentans* appartient maintenant à la famille *Lachnospiraceae* et non *Clostridiaceae*).

La **Figure 3** illustre un arbre phylogénétique des espèces reclassifiées en mentionnant les nouvelles familles phylogénétiques auxquelles les espèces ont été classées [12]. Cette reclassification des *Clostridia* se poursuit toujours, et tous les changements sont régulièrement mis à jour sur http://www.bacterio.net.

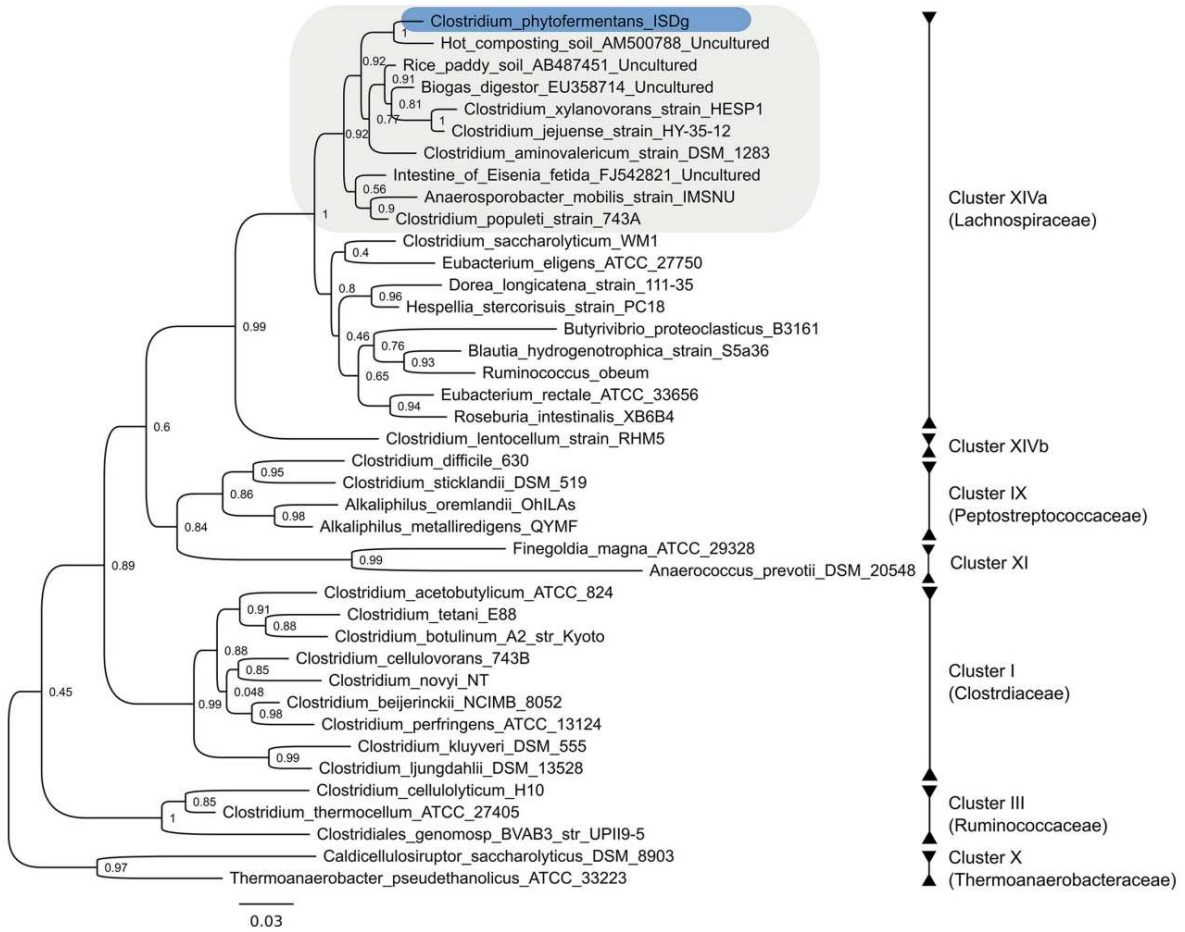


Figure 3 : Arbre phylogénétique de Lachnoclostridium phytofermentans et de quelques taxons de la Classe Clostridia basés sur des séquences d'ARNr 16S.
Les numéros des clusters correspondent au système de classification de Collins et al. 1994. Les valeurs bootstrap ont été déterminées pour 1000 réplicats. Tiré de [7]

1.5. Clostridia commensaux et pathogènes

Comme mentionné précédemment, il est acté que les *Clostridia* commensaux, pathogènes et ceux caractérisés comme ayant un intérêt industriel sont phylogénétiquement proches. Cette observation suggère que les outils génétiques et les méthodes développées pour l'étude d'un ensemble de *Clostridia* peuvent être appliqués à d'autres [13–17]. La *Classe* des *Clostridia* comprend de nombreuses espèces qui revêtent une importance cruciale pour la physiologie, la santé et le microbiote intestinal humain. Chez des patients sains, les *Clostridia* pathogènes sont généralement inoffensifs. Cependant, en cas de déséquilibre, tel que celui induit par la prise d'antibiotiques ou

d'autres traitements antimicrobiens, certains *Clostridia* opportunistes peuvent proliférer et prendre le dessus sur le reste du microbiote intestinal, entraînant des maladies.

Clostridioides difficile [12] (auparavant connu sous le nom de *Clostridium difficile*) est un représentant majeur des *Clostridia* pathogènes. Son infection est réputée pour provoquer des diarrhées sévères, colites, et est suggérée comme ayant un lien potentiel avec le développement de cancer du côlon. Selon une étude du *CDC Emerging Infections Program (EIP*) de 2015, une épidémie est apparue au début du XXIe siècle, infectant près d'un demi-million d'Américains (USA) et entraînant environ 29 000 décès en 2011 [18].

Typiquement, les infections à *C. difficile* touchent les patients vulnérables, tels que ceux immunodéprimés, âgés ou affaiblis. Étant principalement nosocomiales, elles surviennent souvent après des interventions chirurgicales ou des traitements antibiotiques. La prise de probiotiques après un traitement antibiotique est reconnue pour réduire le risque de développer des maladies causées par des souches antibiorésistantes [19]. Par exemple, des consortiums microbiens comme le SER-109 ont été développés pour lutter contre les infections récurrentes à *C. difficile* (rCDI). Ce consortium, composé de 77 genres bactériens, dont 36 % sont des *Lachnospiraceae*, a montré une efficacité significative dans une étude de phase 3 en réduisant la rCDI à la semaine 8. En 2023, la *Food and Drug Administration* américaine a donné son approbation au SER-109, commercialisé sous le nom de Vowst™, pour traiter les rCDI.

L'utilisation de probiotiques basés sur des *Clostridia* aurait de nombreux bénéfices. Ces bactéries dégradent un large panel de polysaccharides indigestes pour l'homme et produisent des métabolites aux nombreux bénéfices tels que les *short-chain fatty acids* (SCFAs) : acétate, propionate et butyrate. Les SCFA jouent un rôle essentiel dans la santé du tube digestif en nourrissant les cellules intestinales, en régulant l'inflammation et le microbiote, en influençant la motilité intestinale, en protégeant contre le cancer colorectal, en régulant l'appétit et en modifiant le pH intestinal. Ces effets contribuent à maintenir un système digestif sain et à prévenir diverses maladies gastro-intestinales [20].

1.6. Clostridia à intérêt industriels

Parallèlement, les *Clostridia* ont suscité un intérêt croissant dans le domaine industriel en tant que châssis pour la production de divers produits chimiques. Un châssis, dans ce contexte, désigne un organisme utilisé à des fins de production industrielle et pouvant être modifié génétiquement pour optimiser les rendements des produits d'intérêts. Poussés par l'essor des biotechnologies, les châssis disponibles se sont diversifiés dans les phénotypes qu'ils arborent. Actuellement, les châssis traditionnels des bioraffineries incluent principalement *Escherichia coli*, *Saccharomyces cerevisiae* et *Corynebacterium glutamicum* (**Tableau 2**).

Parmi ces hôtes, *E. coli* est couramment employé pour la production de biocarburants et d'autres alcools tels que le butanol et le propanol. Toutefois, sa moindre robustesse face aux variations de température, de salinité et de pH, ainsi que sa tolérance inférieure à l'alcool par rapport aux levures et aux *Clostridia*, limitent son utilisation en industrie.

Les levures, quant à elles, sont reconnues pour leur grande robustesse et leur tolérance élevée aux inhibiteurs et aux produits toxiques, mais elles ne possèdent pas la capacité de fermenter naturellement les pentoses. *C. glutamicum* est couramment employé dans la production d'acides aminés, d'éthanol et d'acides organiques. Bien qu'il présente une bonne résistance aux phénols et aux furanes que l'on trouve habituellement dans les hydrolysats lignocellulosiques, il n'est pas naturellement capable d'utiliser les pentoses [21].

Les *Clostridia*, pourtant moins connus du grand public, sont largement utilisés dans la production de nombreux produits chimiques. Parmi la grande diversité des *Clostridia*, quasiment tous les substrats connus peuvent être utilisés comme source de carbone : hexoses, pentoses, biomasse lignocellulosique ainsi que le gaz de synthèse. L'évolution de leur métabolisme solvantogénique leur a conféré une plus grande résistance aux alcools et solvants que les autres châssis. Leur culture en anaérobie qui semble initialement être un inconvénient, s'avère être un atout majeur à l'échelle industrielle. En effet, contrairement aux cultures aérobies qui nécessitent des systèmes complexes d'oxygénation du milieu de

culture, la culture en anaérobie ne nécessite qu'une agitation pour homogénéiser la culture offrant des gains énergétiques non négligeables.

Tableau 2 : Comparatif des châssis industriels utilisés.
Adapté de [22]

Organismes	***S. cerevisiae***	***E. coli***	***C. glutamicum***	***Clostridia***
Mode de culture	Aérobie Anaérobie	Aérobie, Anaérobie	Aérobie Anaérobie	Anaérobie
Application principale	Alcools, protéines eucaryotes	Molécule thérapeutiques, SCFAs, Biofuels	Acides aminés, alcools, acides organiques	SCFAs, Biofuels
Temps de génération en conditions optimales	90 min	20 min	3 - 5 heures	Une à plusieurs heures
Diversité de substrat	C6	C6, Glycérol, Acétate	C6, Glycérol, Acétate	C6, C5, Lignocellulose, Syngas
Valorisation de la biomasse cellulaire	★★★★★	★	★★	★★★★
Résistance aux conditions de culture industrielles	★★★★	★★	★★★★	★★★★
Facilité d'ingénierie génétique	★★★★★	★★★★★	★★★★	★★★

1.6.1. *Clostridia* solvantogènes

De nombreux *Clostridia* industriels sont étudiés en raison de leur phénotype solvantogène. Le phénotype de production de solvants et de SCFA est commun à tous les *Clostridia* industriels. Ce qui distingue ces espèces est la source de carbone utilisée par leur métabolisme : glycolytique (C5 et C6), acétogène (C1) ou lignocellulosique (biomasse) ainsi que les produits de fermentations **Figure 4** [23].

L'espèce la plus étudiée et représentative des *Clostridia* industriels est *Clostridium acetobutylicum*, elle est capable de fermentation ABE (acétone, butanol, éthanol) selon un ratio de 3:6:1. Son métabolisme se caractérise par sa subdivision en deux phases séquentielles. La première est la phase acidogène, qui se déroule lorsque *C. acetobutylicum* est en croissance exponentielle et se manifeste par une production rapide de SCFA (acétate et butyrate). La seconde est la phase solvantogène, qui survient lorsque la croissance cellulaire ralentit, permettant la production de solvants (acétone, butanol et éthanol : ABE), les SCFA étant partiellement réassimilés comme source de carbone durant cette phase [24].

Dans l'étude de *Nguyen et al. 2018*, l'équipe a modifié génétiquement *C. acetobutylicum* en délétant des gènes responsables de la production de SCFAs (*ptb*, *buk*), d'acétone (*crtAB*) et d'acide lactique (*ldhA*) pour rediriger les flux de carbones vers la production de butanol. De plus, des gènes optimisés ont été surexprimés pour augmenter sa production. En couplant cette approche d'ingénierie métabolique et une méthode de production et purification à petite échelle, cette équipe a pu passer de 10 g/L de butanol dans le fermenteur à 550 g/L de butanol en sortie de décanteur après distillation [25].

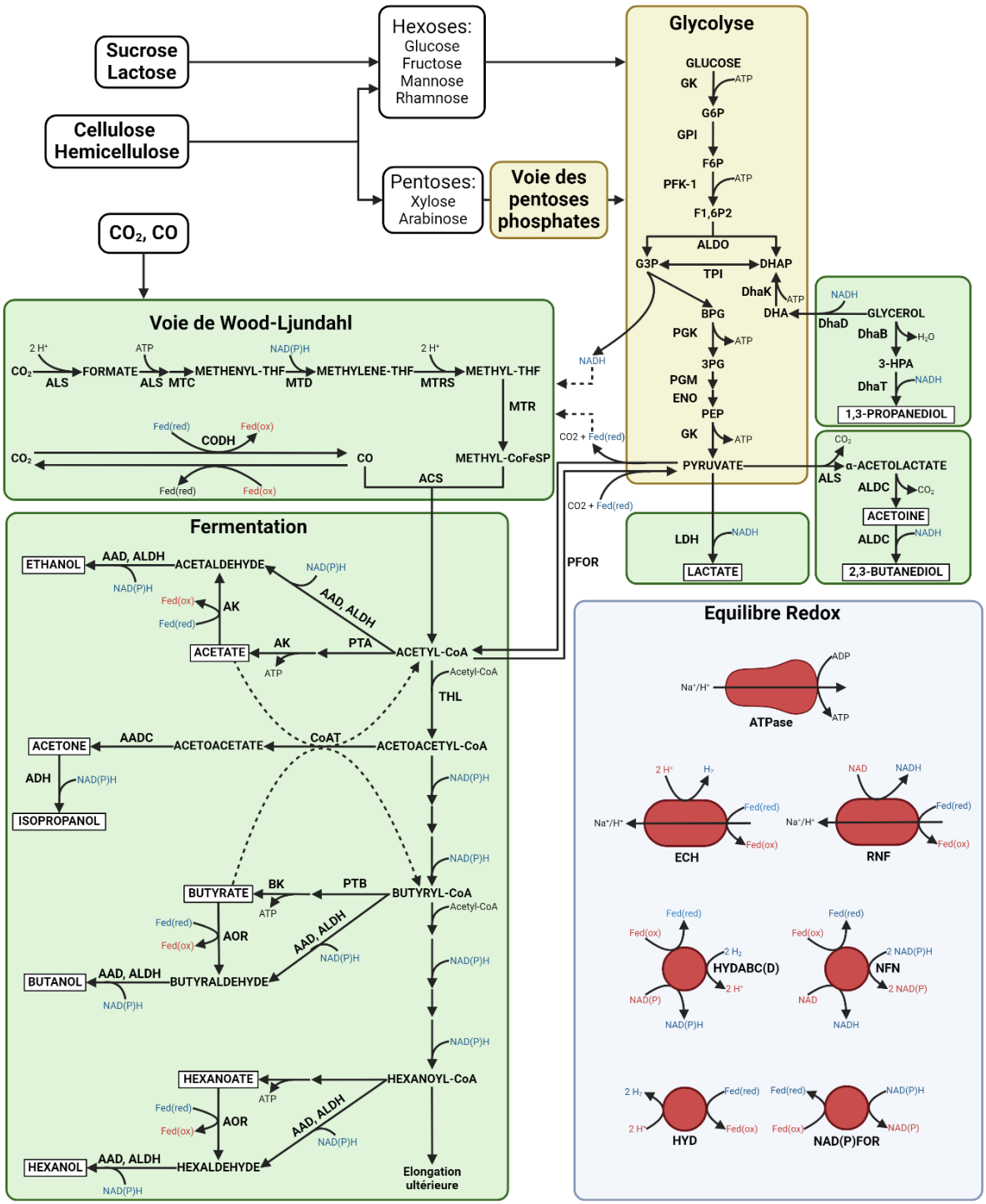


Figure 4 : Le métabolisme des Clostridia.

Les processus métaboliques oxydatifs (en jaune) génèrent des équivalents réduits (en bleu) ainsi que des métabolites intermédiaires. Les processus métaboliques réductifs (en vert) utilisent ces équivalents réduits (en bleu) pour la synthèse de produits (encadrés en noir). Les mécanismes d'équilibrage redox, permettant de maintenir l'équilibre des équivalents de réduction produits et consommés. Adapté de [23]

1.6.2. *Clostridia* acétogènes

Les *Clostridia* acétogènes se distinguent par leur capacité à produire de l'acétate à partir de deux molécules de dioxyde de carbone (CO_2). Ces organismes sont des autotrophes facultatifs, capables de croître par l'oxydation de molécules organiques telles que les sucres à 5 et 6 carbones ainsi que les alcools, ou bien par l'oxydation de molécules inorganiques comme le dihydrogène (H_2) ou le monoxyde de carbone (CO). La voie métabolique de *Wood-Ljungdahl* est centrale dans leur métabolisme, se composant d'une branche méthyle et d'une branche carbonyle et servant à la formation de l'intermédiaire central, l'acétyl-CoA[26]. Un modèle bien étudié pour ce groupe est *C. ljungdalhii*[27], bien que d'autres espèces, telles que l'acétogène *C. carboxidivorans* [28], produisent également des biocarburants (éthanol, le butanol et l'hexanol) ainsi que des SCFAs ('acétate, le butyrate et l'hexanoate) [29].

La voie de *Wood-Ljungdahl* est hautement économe en énergie. En croissance autotrophe, le CO_2, le CO et le CH_4 sont utilisés comme sources d'énergie et de carbone. En revanche, en croissance hétérotrophe, où les sucres en C5 et C6 servent de sources de carbone, la glycolyse libère du CO_2 et est couplée à la voie de *Wood-Ljungdahl*. En exploitant la catalyse du pyruvate en acétyl-CoA par la pyruvate ferrédoxine oxidoreductase (PFOR) pour assimiler le CO_2 produit, une nouvelle molécule d'acétyl-CoA est produite par la voie de *Wood-Ljungdahl*. En conséquence, les acétogènes ont la capacité de fixer l'entièreté du carbone des sucres. Pour une molécule de glucose, trois molécules d'acétyl-CoA (2 par Glycolyse + 1 en utilisant les 2 CO_2 du pyruvate) seront produites [30].

Les applications industrielles des acétogènes sont vastes et prometteuses. La fermentation des gaz de synthèse (syngas) permet de recycler les gaz résiduels d'un certain nombre d'industries telles que les industries métallurgiques et énergétiques utilisant des combustibles fossiles. Des entreprises telles que INEOS Bio, Synata Bio ou encore LanzaTech exploitent les *Clostridia* acétogènes autotrophes comme biocatalyseurs pour la production de biocarburants et d'autres molécules à valeur ajoutée. Par exemple, LanzaTech a annoncé produire de l'éthylène avec une neutralité carbone en modifiant leur

souche de *Clostridium autoethanogenum*, contribuant par conséquent à réduire l'émission de dioxyde de carbone dans l'atmosphère, une des principales sources d'émissions de l'industrie chimique [31].

1.6.3. *Clostridia* cellulolytique

La biomasse lignocellulosique représente une ressource abondante à la fois riche en carbone, économique et renouvelable. Elle présente une alternative prometteuse pour la transformation des économies locales en valorisant les déchets agricoles et municipaux pour produire des molécules industriellement intéressantes. Cette biomasse végétale peut se composer de cellulose, hémicellulose, lignine, pectine ou d'amidon suivant son origine. Une analyse détaillée de la composition de la biomasse végétale sera présentée dans le chapitre 2 de cette introduction.

Dans cette section, nous nous concentrerons sur les *Clostridia* cellulolytiques. Parmi les *Clostridia* cellulolytiques, *C. thermocellum* est particulièrement étudié en raison de son potentiel industriel élevé en tant que dégradeur de cellulose.

C. thermocellum, un thermophile capable de produire de l'éthanol directement à partir de la cellulose, a été isolé pour la première fois en 1926 par Viljoen et al. Des études ultérieures ont révélé que sa température de culture optimale est de 55°C, ce qui en fait un organisme intéressant pour la dégradation de la biomasse. Sa haute température de culture est un atout pour la production industrielle, car de nombreuses réactions biologiques sont exothermiques. Cette haute température optimale permet de réduire les besoins en refroidissement de grands volumes tout en réduisant grandement les risques de contamination [32]. Cependant, l'utilisation industrielle de *C. thermocellum* est limitée par sa sensibilité à des concentrations élevées d'éthanol, ne tolérant que jusqu'à 5 g/L dans son milieu de culture avant que sa croissance ne soit significativement inhibée. Cette sensibilité est étroitement liée à la composition et à la structure de sa paroi cellulaire, composée principalement de lipides ramifiés et linéaires à 16 carbones, ainsi que de plasmalogènes à 16 carbones. Cette composition membranaire confère à la membrane plasmique une fluidité élevée, qui peut être exacerbée par la présence d'éthanol. Lorsque la fluidité

devient excessive, la membrane perd son intégrité, affectant alors la croissance et la physiologie cellulaire [33].

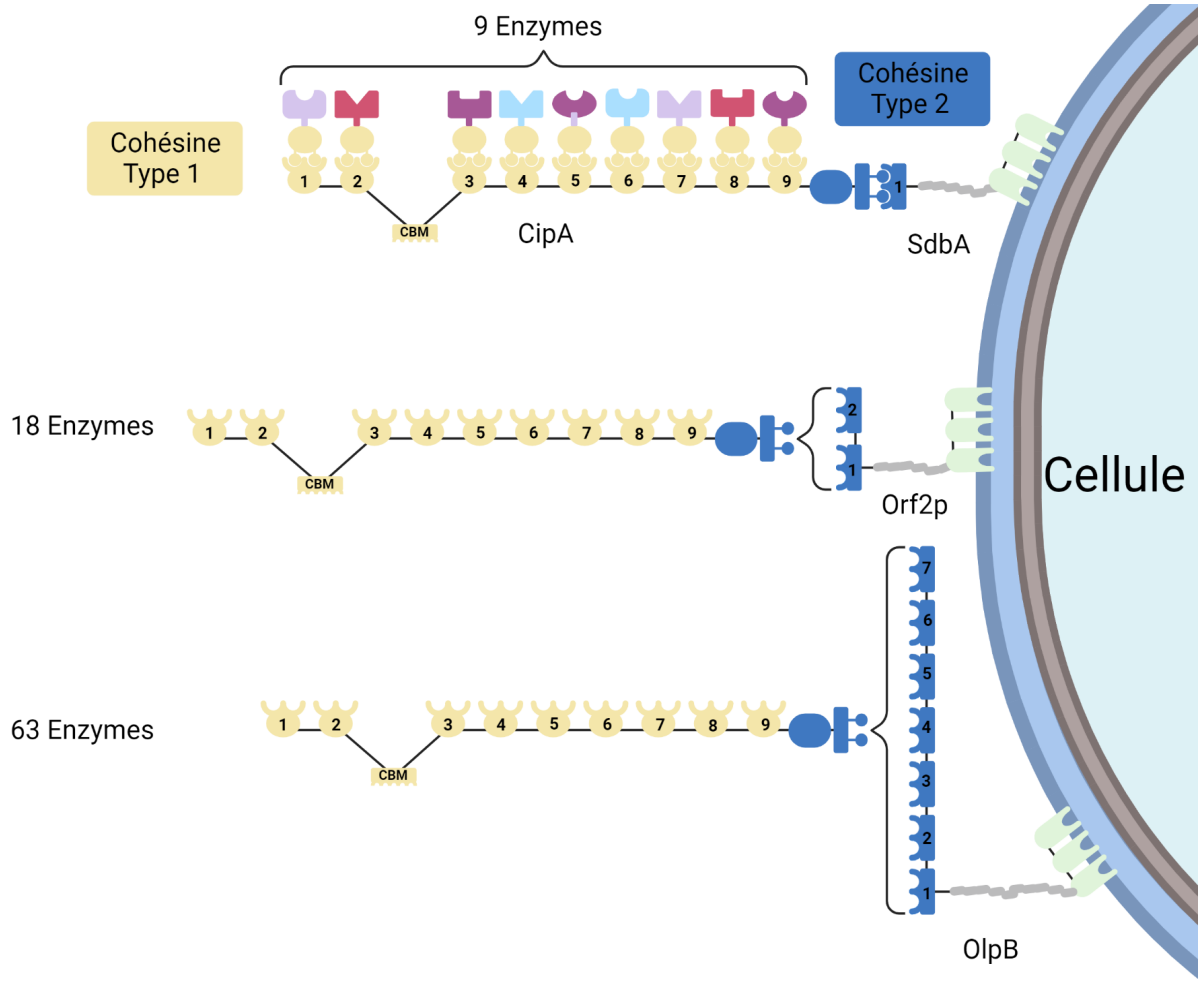


Figure 5 : Paradigme du cellulosome de Clostridium thermocellum.
Ce paradigme consiste en des enzymes individuelles contenant un domaine dockerine de type 1 intégré pouvant se lier aux domaines cohésines de type 1 de la scaffoldine primaire CipA. OlpB, Orf2p et SdbA sont des complexes protéiques qui s'ancrent dans la paroi cellulaire pour accueillir 1, 2 ou 7 cohésines de type 2. Le CBM permet à la cellule de s'accrocher à son substrat cellulosique. Adapté de [34]

Au début des années 1980, l'équipe d'Ed Bayer a fait une découverte majeure en identifiant le cellulosome chez *C. thermocellum* (**Figure 5**). Depuis,le cellulosome de cette bactérie est devenu le système le plus étudié et le mieux compris parmi tous les systèmes cellulolytiques. Il s'agit d'un complexe protéique comprenant des protéines structurelles telles que les dockerines, les cohésines, les scaffoldines et les enzymes impliquées dans la dégradation de la lignocellulose, telles que les hémicellulases, cellulases, pectinases, chitinases, glycosidases, xylanases, mannanases, xyloglucanases et estérases [35].

La pièce maîtresse du cellulosome chez *C. thermocellum* est une protéine macromoléculaire appelée CipA, qui a pour fonction d'être la scaffoldine qui soutient l'ensemble de la structure du cellulosome. Cette protéine comprend neuf domaines cohésines de type I qui se lient à des domaines dockerines de type I, lesquelles sont ensuite connectées aux domaines catalytiques des enzymes. CipA est ancré à la paroi bactérienne par une dockerine de type II et par des complexes protéiques d'ancrage tels que OlpB, Orf2p ou SdbA, qui possèdent des domaines cohésines de type 2 en nombre variable (1, 2 ou 7), permettant de fixer plusieurs CipA. De plus, CipA contient un module de liaison aux glucides (CBM) qui attache le cellulosome à son substrat glucidique, assurant la proximité des enzymes avec leur substrat et contribuant de cette manière à l'efficacité du cellulosome. Il existe de nombreuses variantes de cellulosomes, allant des plus simples aux plus complexes, avec 63 chez *C. thermocellum contre* 160 enzymes chez *Clostridium clariflavum*. *C. clariflavum* exprime des scaffoldines ne possédant pas de domaine dockerine, ce qui rend ses cellulosomes "*cell-free*" par opposition aux cellulosomes "*cell-bound*" ancrés à la paroi cellulaire [34].

Parallèlement à la découverte des systèmes cellulolytiques, le groupe d'Ed Bayer au Weizmann Institute a été pionnier dans l'ingénierie de cellulosomes modifiés, également appelés "*Designer Cellulosomes*". *Morais et al.* (2012) décrit la création de ces *Designer Cellulosomes*, où chaque protéine est considérée comme un module pouvant être assemblé à la scaffoldine en ajoutant les adaptateurs appropriés. En associant les domaines catalytiques des enzymes à une dockerine et à la cohésine correspondante (provenant du même organisme producteur du cellulosome) sur la scaffoldine, il est possible de contrôler

la composition enzymatique d'un cellulosome. Des améliorations ultérieures ont permis au cellulosome d'accepter une dockerine de type II, augmentant le nombre de cohésines et donc le nombre d'enzymes dans ce *Designer Cellulosomes* [36].

Enfin, Vanderstraeten et son équipe ont mis en place en 2022 une toolbox et un pipeline permettant une construction rapide et combinatoire pour comparer les rendements des *Designer Cellulosomes* et déterminer la meilleure combinaison d'enzymes pour un substrat donné (**Figure 6**) [37].

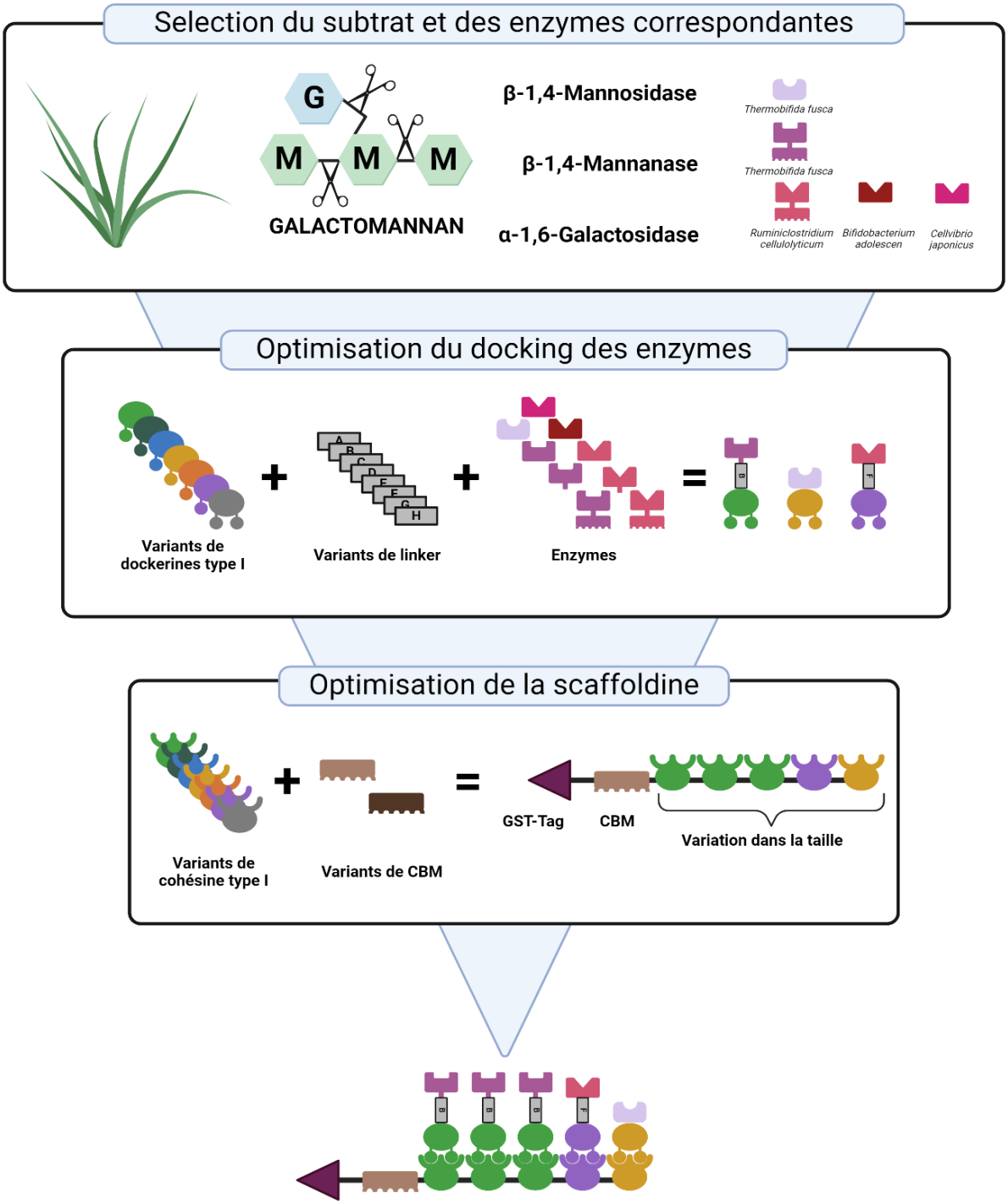


Figure 6 : Optimisation des Designer Cellulosomes pour la dégradation de Galactomannan.
Étape 1 : Sélection du substrat à dégrader et des enzymes nécessaires à sa dégradation en cherchant dans le répertoire des domaines catalytiques. Étape 2 : Optimisation du docking des enzymes. Des tests d'activité ont permis de sélectionner trois variantes optimisées de l'enzyme de docking à incorporer dans le complexe final. Étape 3 : Optimisation de la scaffoldine. L'analyse de l'activité des scaffoldines et des enzymes-dockerin sélectionnées on permit de construire un cellulosome de conception optimisée pour la dégradation de Galactomannan. Adapté de [37]

1.6.4. Lachnoclostridium phytofermentans

1.6.4.1. Généralités

Lachnoclostridium phytofermentans ISDg^{T} a été découvert pour la première fois par *Warnick et al.* 2002. C'est une bactérie anaérobie mésophile à Gram positif isolée du sol forestier près du réservoir de Quabbin dans l'État du Massachusetts (États-Unis). Cet organisme se développe en utilisant à la fois les composants insolubles et solubles de la biomasse végétale, en commençant par la dégradation des polysaccharides végétaux tels que la cellulose, l'hémicellulose, l'amidon et la pectine, suivie de la fermentation des pentoses et hexoses résultants majoritairement en éthanol, acétate, lactate et formate [38].

Selon la base de données CAZy, le génome de *L. phytofermentans* contient 171 enzymes à activité glucidique (CAZymes), le plus grand nombre parmi les *Clostridia* séquencés à ce jour. Contrairement à d'autres *Clostridia*, les CAZymes exprimées par *L. phytofermentans* ne sont pas organisées en cellulosome mais sont sécrétées sous forme libre dans l'environnement [39].

Les observations montrent des différences significatives dans la croissance et la production métabolique de *L. phytofermentans* en fonction du substrat carboné utilisé. La croissance est rapide en présence de glucose (**Figure 7A**), avec un temps de génération d'environ 4 heures, produisant environ 2,5 fois plus d'éthanol que d'acétate (**Figure 7D**). En revanche, en présence d'hémicellulose, la croissance est encore accélérée (**Figure 7B**) et la production d'éthanol et d'acétate est équivalente (**Figure 7E**). Enfin, l'utilisation de la cellulose comme substrat entraîne une croissance plus lente s'étalant sur plusieurs semaines (**Figure 7C**), avec une fermentation principalement axée sur la production d'éthanol (**Figure 7F**). La morphologie cellulaire de *L. phytofermentans* varie en fonction du substrat carboné. En présence de substrats solubles, les cellules sont généralement motiles avec un ou deux flagelles (**Figure 7G-H**), tandis qu'en présence de substrats insolubles, les cellules sont plus courtes, non flagellées et adhérentes à la biomasse (**Figure 7I**) [40]. Cette adhérence à la biomasse est une adaptation permettant d'augmenter la proximité des enzymes avec le substrat et de prévenir la compétition avec d'autres organismes. Alors que des bactéries cellulolytiques modèles telles que *Clostridium*

thermocellum et *Ruminiclostridium cellulolyticum* se fixent à leur substrat cellulosique grâce au *Carbohydrate-Binding Module* (CBM) de leur cellulosome, les mécanismes exacts d'adhérence au substrat chez *L. phytofermentans* restent à élucider, puisque cette fixation n'est pas assurée par le CBM du cellulosome.

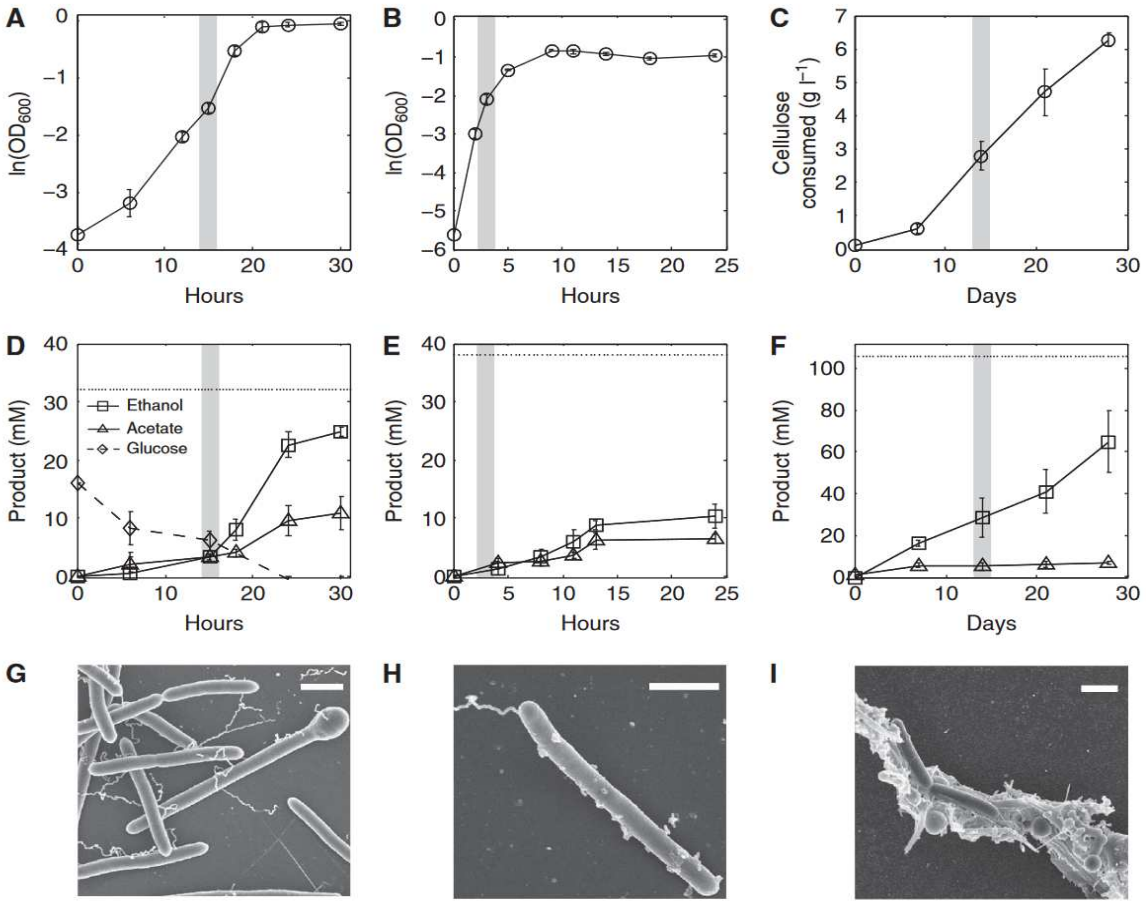


Figure 7 : Croissance, fermentation et morphologie cellulaire de L. phytofermentans sur différentes sources de carbone.

Les barres grises indiquent les moments des échantillons prélevés pour la spectrométrie de masse. La croissance sur glucose (A) et hémicellulose (B) a été quantifiée en OD_{600nm}. La croissance sur cellulose (C) a été mesurée en masse sèche de cellulose dans la culture. La production d'éthanol et d'acétate, principaux produits de fermentation, ainsi que la consommation de glucose dans le traitement au glucose, ont été mesurées par HPLC. Les lignes pointillées illustrent le rendement théorique maximal de l'éthanol. La microscopie électronique à balayage (MEB) montre des cellules se développant sur du glucose (G), de l'hémicellulose (H), et de la cellulose (I), avec une barre d'échelle blanche de 1 mm. Tiré de [40].

1.6.4.2. Métabolisme

En raison de sa grande variété de CAZymes, *L. phytofermentans* est capable de dégrader un large éventail de substrats glycosidiques, ensuite captés par la cellule via une diversité de transporteurs. Des analyses *in silico* ont prédit que le génome de *L. phytofermentans* contenait 572 gènes codant des transporteurs, dont 173 pour le transport de glucides. Parmi eux, les *ATP-Binding Cassettes* (ABC) représentent la majorité des transporteurs responsables de la translocation des glucides par hydrolyse de l'ATP, avec 158 gènes prédits dans le génome de *L. phytofermentans*. Les 15 gènes restants codent pour des transporteurs secondaires de type symport et antiport, qui assurent le transport des substrats avec l'implication de l'ATP (**Figure 8**). Une étude réalisée par *Cerisy et al.* 2019 a révélé que la diversité des CAZymes n'était pas le seul facteur déterminant de l'efficacité de la dégradation de la biomasse par *L. phytofermentans*. En effet, l'inactivation de gènes réalisées à l'aide d'un LI.LtrB *group II intron* (targetron) ciblant le domaine transmembranaire des gènes, ont mis en évidence l'importance des ABC-Transporteurs dans la captation des hexoses et pentoses [41].

En plus de son métabolisme des glucides particulièrement développé, le métabolisme énergétique de *L. phytofermentans* est conservateur (**Figure 9**). Il repose sur un duo de complexes protéiques et enzymatiques membranaires. Ce duo est constitué d'ATPases et d'oxidoréductases de ferrédoxine réduite. La ferrédoxine réduite générée lors de la conversion du pyruvate en acétyl-CoA par le pyruvate ferrédoxine oxidoréductase contribue à la production d'éthanol à la fois directement par la réduction du NAD+ et du NADP+, et indirectement, en participant à la conservation de l'énergie. Un complexe protéique hautement exprimé, appelé Rnf, joue un rôle crucial dans la production d'éthanol grâce à la ferrédoxine réduite. Rnf est une oxidoréductase NADH:ferrédoxine translocatrice de sodium, il crée un gradient d'ions sodium utilisé par la F1Fo-ATPase translocatrice de sodium pour produire de l'ATP à partir d'ADP. Ainsi, le complexe Rnf soutient la production d'éthanol en générant de l'ATP indirectement, réduisant la dépendance à la production d'acétate pour la génération d'ATP [7].

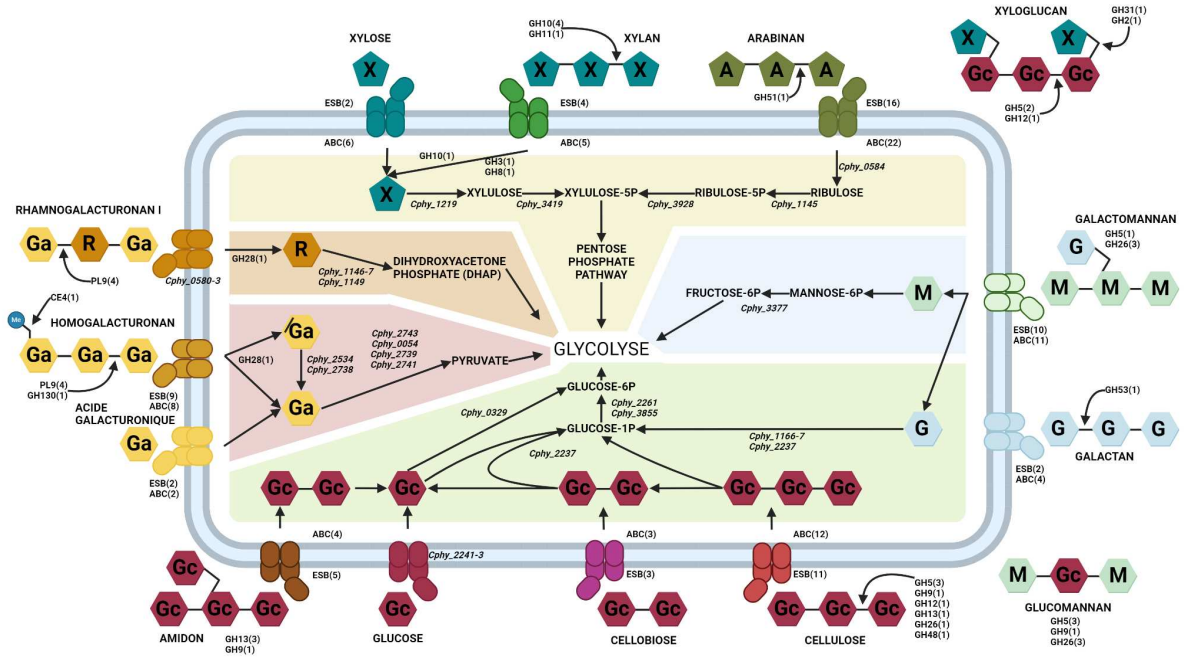


Figure 8 : Modèle de l'import et du métabolisme des carbohydrate chez L. phytofermentans.
Les CAZymes sont annotés par leur numéro de la base de données CAZy. Les enzymes métaboliques sont annotées par leur numéro NCBI. Les abréviations sont acide D-galacturonique (GA), L-rhamnose (R), D-mannose (M), D-glucose (Gc), D-galactose (G), D-xylose (X), L-arabinose (A), fructose (F), phosphate (P), voie du pentose phosphate (PPP), dihydroxyacétone-phosphate (DHAP), glycéraldéhyde-3-phosphate (G3P). Pour chaque substrat, le nombre de protéines de liaison aux solutés extra-cellulaires (ESB) et de transporteurs ABC (ABC) significativement régulés est indiqué. Les régions colorées montrent le métabolisme du glucose (vert), du mannose (bleu), du xylose et de l'arabinose (jaune), du rhamnose (orange) et de l'acide galacturonique (rouge). Adapté de [39]

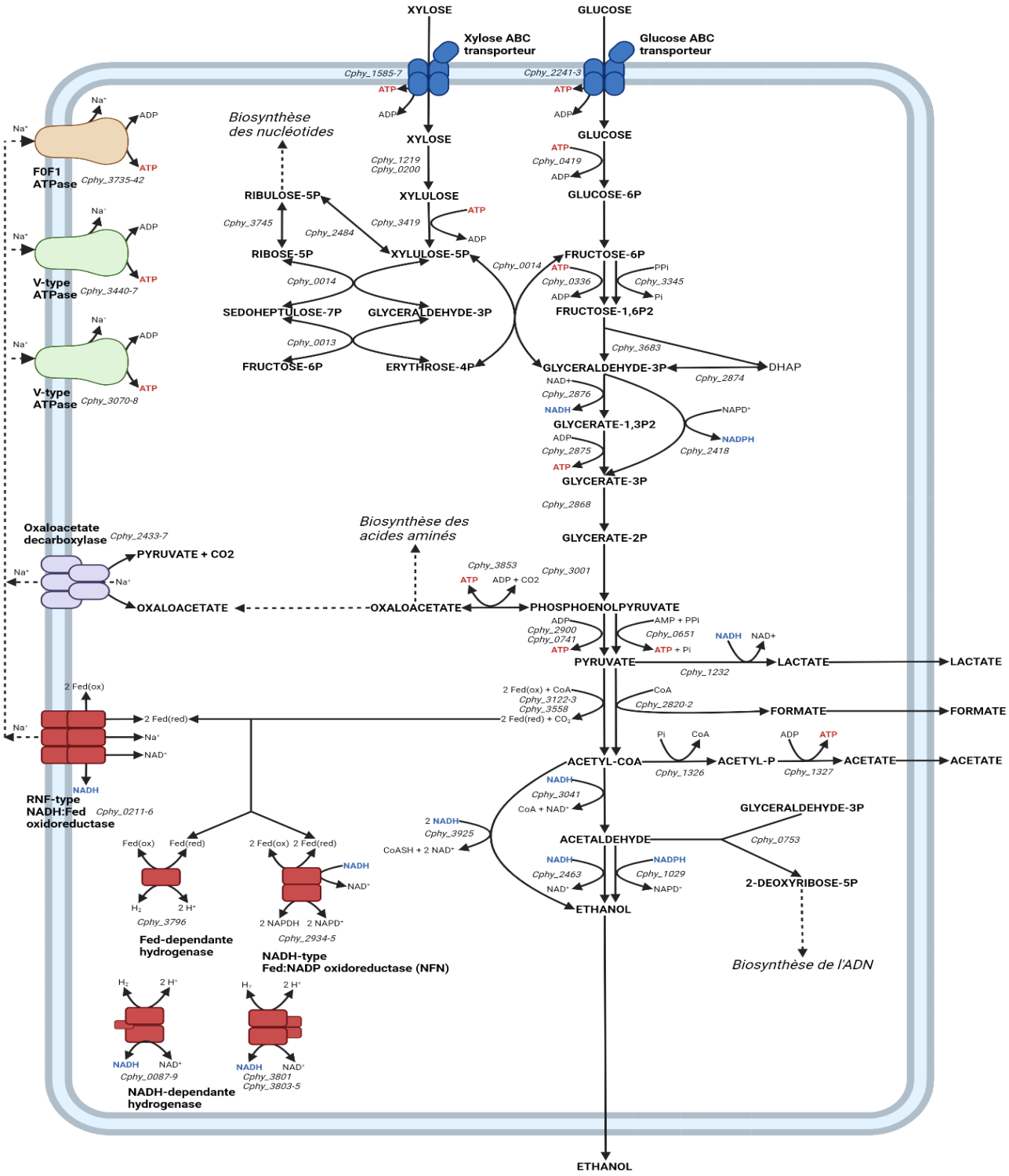


Figure 9 : Métabolisme central de L. phytofermentans.

Les co-facteurs réduits sont représentés en bleu, tandis que les molécules d'ATP sont en rouge. Ce schéma illustre les principales voies métaboliques impliquées dans la conversion des sucres en divers produits de fermentation, incluant l'éthanol, l'acétate et le lactate. Adapté de [7].

1.6.4.3. Autres études et brevets

En raison du nombre limité d'équipes travaillant sur *L. phytofermentans*, cet organisme demeure relativement méconnu : moins de 300 résultats mentionnent *L. phytofermentans* selon la base de données du NCBI. Malgré cela, plusieurs travaux de recherche ont mis en lumière des aspects intéressants de cet organisme et de ses enzymes.

Une découverte notable, mise en lumière par Petit et al. 2013 [42], a révélé la présence de *Bacterial Micro Compartments* (BMC) chez *Lachnoclostridium phytofermentans*. Ces BMC sont des organites bactériens polyédriques constitués exclusivement de protéines, renfermant diverses enzymes fonctionnelles [43]. L'étude a identifié deux *loci* codant ces BMC, dont l'un présente une forte homologie avec un locus de *Roseburia inulinivorans*, une bactérie commensale du microbiote intestinal, impliqué dans la conversion du fucose en propanol et propionate. La compartimentation au sein des BMC permet d'éviter la libération d'intermédiaires toxiques dans le cytoplasme, tels que les aldéhydes **Figure 10**.

Lorsque le rhamnose ou le fucose sont utilisés comme seule source de carbone, les études ont démontré que *L. phytofermentans* produit principalement de l'acétate, du propionate, de l'éthanol et du propanol [42]. La production de ces nouveaux produits de fermentation via l'utilisation des BMC suggère leur rôle essentiel dans l'optimisation des voies métaboliques et la protection cellulaire contre les composés toxiques. Par ailleurs, la compréhension des mécanismes de compartimentation des voies métaboliques pourrait être exploitée pour la synthèse d'autres composés d'intérêt, ouvrant ainsi de nouvelles perspectives en biotechnologie.

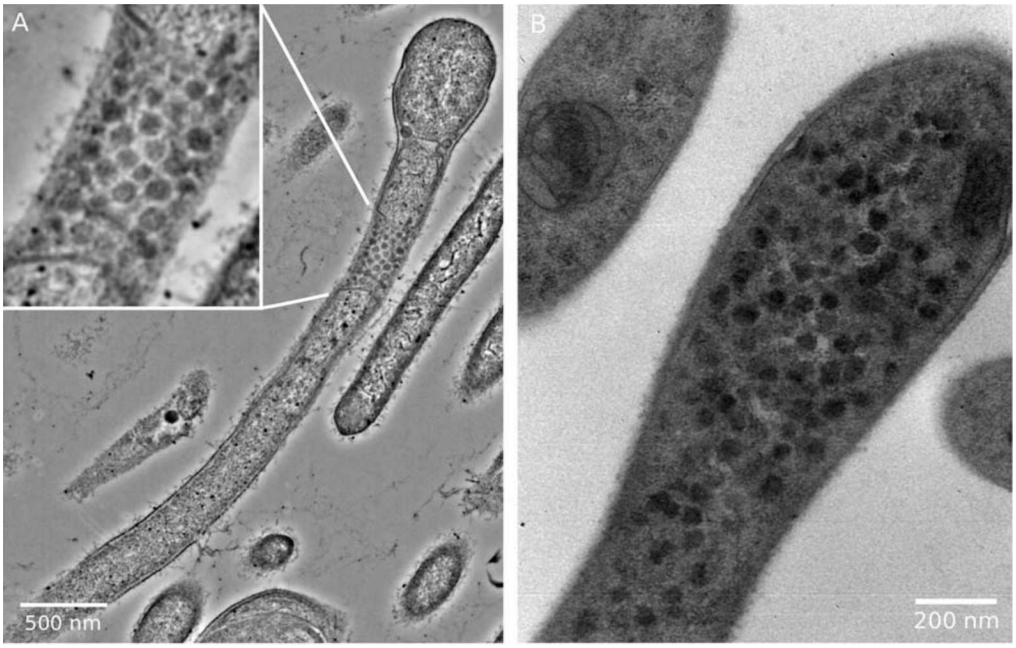


Figure 10 : Image de microscopie électronique à transmission de L. phytofermentans.
Cellules cultivées sur fucose (A) et rhamnose (B) montrant une cellule avec de multiples structures polyédriques. Tiré de [42]

L. phytofermentans offre un fort potentiel industriel, en particulier avec l'émergence de la co-culture telle que celle avec *S. cerevisiae*, permettant une production efficace d'éthanol à partir de sous-produits agro-industriels [44]. D'autre part, un *consortium* artificiel utilisant *E. coli* et *L. phytofermentans* favorise leur développement en synergie : *E. coli* consomme l'oxygène et les sous-produits inhibiteurs issus du catabolisme du cellobiose favorisant la croissance de *L. phytofermentans* qui, en retour, libère du glucose utilisable par *E. coli*, développant un cycle de croissance coopératif [45].

En plus de son potentiel en tant que châssis de bioproduction, *L. phytofermentans* suscite également un intérêt pour ses enzymes notamment par leur caractérisation dans des systèmes d'expression exogènes. Par exemple, la surexpression d'une xylose isomérase hautement efficace issue de *L. phytofermentans* , a conduit à une fermentation plus efficace du xylose chez *S. cerevisiae* [46]. Plus récemment, différentes voies d'assimilation du xylose, notamment via la xylose isomérase (XI) et la xylose réductase/xylitol déshydrogénase (XR/XDH), vers la voie des Pentoses Phosphates, ont été comparées pour la production d'éthanol. Cette étude a montré que l'expression seule de la xylose isomérase de *L. phytofermentans* permettait une assimilation de 84% du xylose. En complément, la co-

expression de XI et XR/XDH a permis d'obtenir un meilleur équilibre NAD^+/NADH, favorisant une meilleure détoxification des furanes produits lors de la consommation du xylose et augmentant ainsi le rendement d'éthanol de 38% [47].

Ces recherches soulignent l'intérêt de *L. phytofermentans* dans la production d'enzymes cellulolytiques, qui peuvent être exprimées plus facilement de manière exogène que les enzymes dotées de domaines dockerine. Les premiers brevets publiés par l'équipe à l'origine de la découverte de cet organisme, revendique son utilisation pour les applications de la fermentation de la biomasse végétale par *L. phytofermentans*. Ces brevets revendiquent l'utilisation de *L. phytofermentans* pour la fermentation d'une large gamme de biomasse végétale et pour la production d'éthanol et d'hydrogène [48]. Les différentes spécificités de *L. phytofermentans* ont permis d'éveiller son intérêt industriel pour la valorisation de la biomasse végétale, ces différents procédés et applications seront détaillés dans le prochain chapitre.

2. Produits d'intérêts

2.1. Applications dans le domaine de la santé

2.1.1. Protéines médicales

Les *Clostridia* sont souvent associés aux infections qu'ils provoquent, ce qui leur confère une mauvaise réputation. Pourtant, les facteurs de virulence responsables de ces maladies peuvent être valorisés dans le domaine médical. Par exemple, la toxine botulique produite par *Clostridium botulinum* peut être létale pour un adulte suite à une injection de seulement 2 ng ou une inhalation de 13 ng [49]. Cependant, administrée en microdose, elle a montré des bénéfices en médecine esthétique mais surtout dans le traitement de contractions spasmodiques telles que le blépharospasme, le spasme hémifacial, le strabisme et la dystonie cervicale [50]. Ces contractions se manifestent par des mouvements forts, répétés et incontrôlés des paupières ou par des contractions musculaires fortes, prolongées et involontaires. De plus, elle est également utilisée en prophylaxie contre les migraines chroniques chez les adultes [51]. La toxine botulique agit en inhibant l'action du complexe SNARE, responsable de la fusion entre les vésicules d'acétylcholine et la membrane présynaptique des cellules nerveuses. Ainsi, les récepteurs à l'acétylcholine ne sont pas stimulés et la contraction musculaire ne se produit pas. Ces mécanismes reflètent les symptômes de la paralysie flasque observée chez les patients atteints de botulisme.

Les collagénases sont également largement utilisées en thérapie. Principalement produites par *Clostridium histolyticum*, ces enzymes ont pour fonction de dépolymériser les fibres de collagène de type I et III en rompant les liaisons peptidiques et agissent comme facteurs de virulence par *C. histolyticum* pour provoquer la gangrène gazeuse [52]. Ces enzymes sont désormais utilisées dans le traitement de la contracture de Dupuytren et de la maladie de Peyronie [53]. Ces pathologies sont caractérisées par des nodules rigides sous la peau, limitant la mobilité. Les collagénases dégradent ces nodules/plaques de collagène, permettant ainsi de retrouver une partie de la mobilité initiale. Elles sont également utilisées pour aider à éliminer les tissus morts des blessures ainsi qu'en médecine esthétique pour

réduire la cellulite chez les femmes adultes en dégradant les fibres de collagène qui tendent à tirer la peau de manière irrégulière [54,55].

2.1.2. Protéines alimentaires

Un nombre croissant d'entreprises et de start-up, telles que SolarFood [56] ou encore Superbrewed Food [57], se concentre sur le développement et la production de nouvelles protéines alimentaires. À l'instar de SolarFood, dont le domaine phylogénétique du microorganisme utilisé n'est pas public, SuperBrewed Food utilise une ou plusieurs espèces de *Clostridia* pour produire des aliments tels que des barres énergétiques composées de protéines à haute valeur nutritionnelle. Dans un brevet publié en 2022 portant sur une fermentation de 20h, cette start-up garantit la production de protéines pure (20g/kg), d'acides aminés (7g/kg) mais également des SCFA, du butyrate et de l'acétate, décrits comme des éléments essentiels à l'équilibre intestinal [57].

2.2. Biocarburants

Les biocarburants sont des combustibles d'origine organique à partir de biomasse. Contrairement aux carburants conventionnels, ils sont produits dans un court laps de temps se comptant en jours contre des dizaines de millions d'années pour la formation de pétrole. Dans certains cas, la biomasse elle-même peut être utilisée comme biocarburant, par exemple les copeaux de bois. Généralement, le terme "biocarburant" fait référence à un carburant liquide ou gazeux utilisé pour alimenter des moteurs. Avant même l'utilisation généralisée du pétrole et autres dérivés, les premiers moteurs ont été développés avec pour source des biocarburants.

Les premiers moteurs ont été développés pour un fonctionnement à base de biocarburant par deux pionniers : Nikolaus Otto et Rudolf Diesel, qui ont créé respectivement un premier moteur à combustion interne fonctionnant à l'éthanol ainsi qu'un moteur éponyme alimenté par des huiles végétales telles que l'huile d'arachide. Encore aujourd'hui, les biocarburants sont en pleine extension. Comme le montre le graphique de la **Figure 11**, historiquement, l'Amérique du Sud est un pionnier dans la

production de biocarburants et est aujourd'hui le deuxième plus grand producteur de bioéthanol et de biodiesel après les États-Unis.

Les biocarburants se divisent en deux catégories : ceux à base d'alcool et ceux à base d'huile. A noter que, certains gaz sont également inclus dans la catégorie des biocarburants, tels que le H_2 et le CH_4

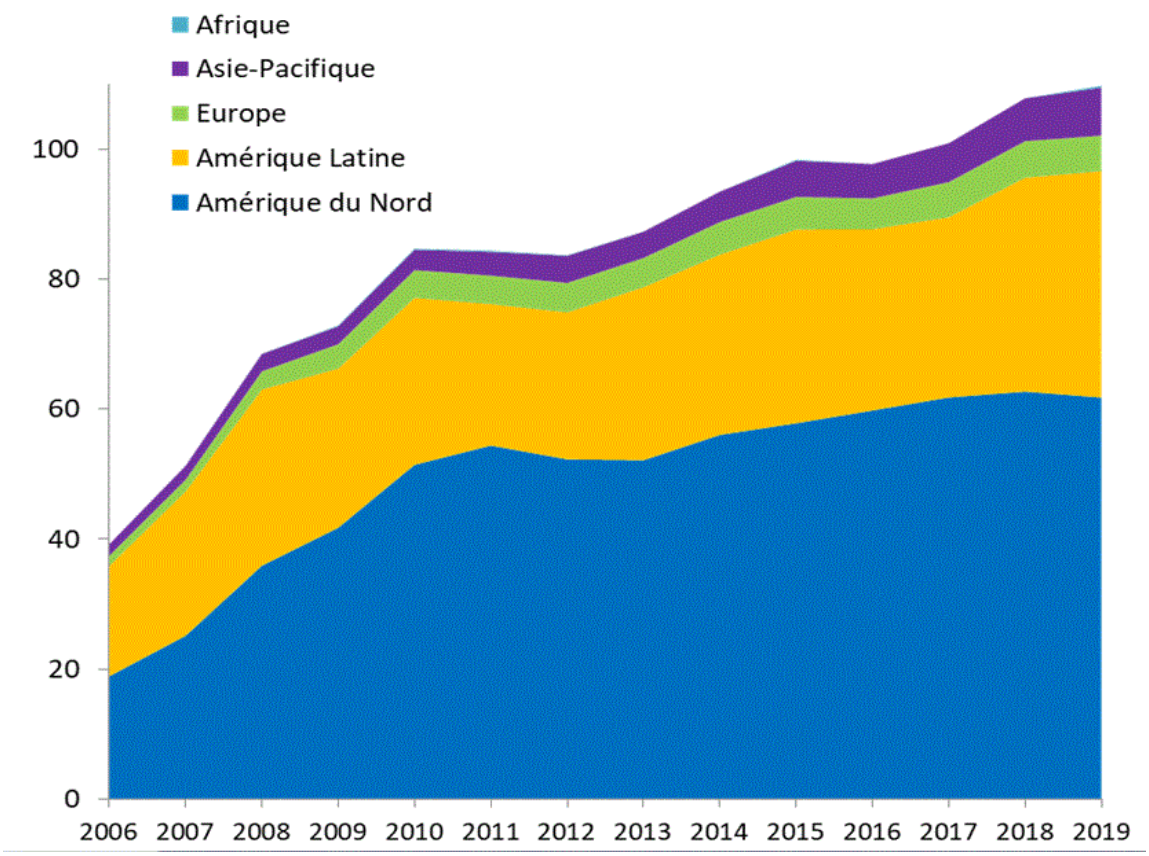


Figure 11 : Évolution de la production de bioéthanol par continent entre 2006 et 2019.
La production exprimée en milliards de litres est dominée par l'Amérique du Nord, suivie de l'Amérique Latine, de l'Europe, de l'Asie-Pacifique et de l'Afrique. Tiré de [58]

2.2.1. La filière huile

Un grand nombre d'espèces végétales sont connues et cultivées pour leurs propriétés oléifères, telles que le tournesol, le palmier à huile et le colza. Ces huiles sont extraites des graines par pressage à chaud ou à froid, et elles offrent deux possibilités d'utilisation en tant que biocarburant : elles peuvent être utilisées de manière pure et brute ou en association avec d'autres carburants dans les moteurs à allumage par compression (Diesel).

Cependant, l'une des limitations majeures est la viscosité des huiles végétales brutes qui nécessite une adaptation du moteur.

2.2.1.1. Bio-huile

La bio-huile est le liquide résultant de la transformation de la biomasse soit par pyrolyse rapide soit par liquéfaction hydrothermale. De couleur marron foncé et avec une odeur fumée, ce biocarburant est une mixture d'acides, d'aldéhydes, d'alcools, de cétones, d'esters, de phénols et d'oligomères. En raison de son mélange non défini, ce biocarburant n'est pas adapté à une utilisation sur route dans des moteurs de véhicules. Les propriétés indésirables de la bio-huile comprennent une forte teneur en eau, une viscosité élevée, une teneur élevée en oxygène et une corrosivité élevée [59].

Son coût de production élevée et la production d'un combustible de basse qualité rendent la commercialisation de la bio-huile difficile. Cependant, un projet pilote au Brésil vise à optimiser sa production à partir de la pyrolyse de résidus de canne à sucre, dans le but de l'utiliser pour le chauffage, la production d'électricité et de produits chimiques [60].

2.2.1.2. Biodiesel

Le biodiesel est produit à partir d'huiles végétales, de graisses jaunes, d'huiles de cuisson usagées ou de graisses animales. Ce carburant est obtenu par transestérification des triglycérides avec un alcool à courte chaîne, généralement du méthanol. Ce processus transforme les graisses et les huiles en biodiesel et en glycérine. La glycérine est un coproduit de la synthèse du biodiesel et peut être valorisée puisque couramment utilisée dans la fabrication de produits pharmaceutiques et cosmétiques.

Le biodiesel peut être mélangé et utilisé dans différentes concentrations. Les plus courantes sont le B5 (jusqu'à 5% de biodiesel) et le B20 (6% à 20% de biodiesel). Le B100 (biodiesel pur) est généralement utilisé comme base pour produire des mélanges moins concentrés et est rarement utilisé comme carburant de transport car moins rentable. Le B20 est un mélange courant car il offre un bon équilibre entre le coût, les émissions, les performances par temps froid, la compatibilité des matériaux et la capacité à agir comme solvant [61].

La principale problématique de la filière huile réside dans son besoin en terres arables pour cultiver des plantes oléagineuses. Dans la section suivante, nous examinerons les différentes générations de biocarburants. Produire des biocarburants à partir de plantes qui pourraient être utilisées pour l'alimentation humaine n'est pas durable, et des alternatives doivent être trouvées. Pour le biodiesel, par exemple, la culture de microalgues pourrait rendre cette filière à la fois plus rentable et plus durable. Cette alternative sera discutée en détail dans la partie consacrée aux générations de biocarburants.

2.2.2. La filière alcool

En parallèle à la production de gazole pour les moteurs Diesel, la filière de l'alcool représente la première source de biocarburant dans le monde. Selon l'Agence internationale de l'énergie, les États-Unis ont produit 57,5 milliards de litres d'éthanol et 14,7 milliards de litres de biodiesel en 2022. Tout comme pour la production d'huile végétale, il existe une grande diversité d'espèces végétales qui pourraient être utilisées comme source d'amidon ou de sucre plus simple [62]. Les plus cultivées sont le maïs, le blé, la canne à sucre et la betterave sucrière. De nouvelles recherches pourraient rendre la production de bioalcool plus pérenne, elles conduisent à la revalorisation de déchets au lieu de l'exploitation de ressources agricoles opposant l'alimentation de la population à la production de biocarburant.

Les bioalcools majoritairement produits sont l'éthanol, le méthanol et le butanol, chacun a des propriétés physico-chimiques en fonction du nombre de carbone qu'il possède, le **Tableau 3** récapitules les caractéristiques de chacun en tant que biocarburant.

Tableau 3 : Comparaison des propriétés des bioalcools et de l'essence.
Adapté de (75)

Biocarburants	Méthanol	Ethanol	Butanol	Essence
Caractéristiques techniques				
Valeur énergétique (MJ/L)	16,7	21,2	29,2	32,5
Ratio Air-Carburant	6.6	9	11-12	12-15
Research Octane Number (RON)	109	129	96	91-99
Production de CO_2 (kg/L)	1,7	2,2	2,9	4,7
Aspects pratiques				
Utilisation maximal seul	50%	85%	100%	100%
Moteur	Modifications nécessaire	Modifications mineures	Sans modification	Sans modification
Possibilité de stockage	Hautement corrosif, réservoirs spéciaux	Hygroscopique, réservoirs protégés	Comparable à l'essence, moins volatile	Facile à stocker, infrastructure existante

2.2.2.1. Biométhanol

Le biométhanol est encore largement méconnu en tant que carburant direct pour les véhicules, bien qu'il soit de plus en plus envisagé notamment pour lutter contre les émissions de particules fines dans l'atmosphère. Des projets d'envergure, tels que celui coordonné par TotalEnergies sur le site de Fraunhofer CBP, visent à produire du méthanol de manière renouvelable. Cette production implique la synthèse de méthanol en combinant du CO_2 industriel hautement concentré avec de l'hydrogène généré par électrolyse de l'eau, en utilisant l'énergie solaire ou éolienne [63].

Selon le *Methanol Institute*, le méthanol peut être utilisé comme additif à l'essence ou comme carburant autonome. Par exemple, le carburant A20 contient du méthanol, de l'éthanol et de l'essence [64]. De plus, une étude menée par *Wang et al.* 2019 a démontré qu'un mélange de 70% de méthanol dans de l'essence ayant un RON de 81,1 augmentait le rendement du

moteur de 30%. Il est même possible d'utiliser des mélanges de méthanol allant jusqu'à M50 sans qu'il y ait une détérioration significative (moins de 5%) de l'autonomie du véhicule [65].

2.2.2.2. Bioéthanol

Le bioalcool le plus produit est l'éthanol, obtenu par la fermentation d'amidon hydrolysé ou de sucres simples par la levure Saccharomyces cerevisiae. À ce jour, le bioéthanol est le bioalcool le plus largement produit. Au cours des deux dernières décennies, de nombreuses usines ont été construites pour produire du bioéthanol de manière plus économique et efficace. Ces enjeux seront discutés dans la section suivante, concernant les différentes générations de biocarburants [66].

Le bioéthanol peut remplacer partiellement l'essence conventionnelle jusqu'à 85%, ce qui constitue le carburant E85, composé de 85% d'éthanol et de 15% d'essence. En France, il est possible d'utiliser du E85 sans modification du moteur. Cependant, étant environ 30% moins calorique que le SP95, une modification de la fréquence d'injection du carburant en modifiant le calculateur du véhicule est nécessaire pour assurer un bon fonctionnement du moteur [67].

Au Brésil, des gammes de véhicules dits "Flexible-fuel" sont vendues pour utiliser toutes les concentrations d'éthanol comme carburant, allant de E20 à E100. Du fait de son RON élevé, l'éthanol résiste plus à l'auto-inflammation que l'essence. À des températures plus basses, il est plus *difficile* pour l'éthanol d'atteindre les conditions nécessaires pour s'enflammer spontanément dans le cylindre du moteur. Les problématiques de démarrage à froid (températures inférieures à 15°C) ont été résolues en ajoutant un deuxième réservoir plus petit près du moteur, en réchauffant ce réservoir cela facilite le démarrage du véhicule [68,69].

2.2.2.3. Biobutanol

Le biobutanol est l'un des biocarburants les plus prometteurs de la filière alcool. Tout comme le bioéthanol, il est produit par la fermentation de microorganismes. Cependant, la production à l'échelle industrielle du butanol est réalisée spécifiquement par des bactéries du genre *Clostridium*, telles que *Clostridium acetobutylicum*, *Clostridium beijerinckii*, *Clostridium saccharoperbutylacetonicum* ou *Clostridium saccharobutylicum*, qui fermentent des sucres simples ou des hydrolysats de biomasse d'amidon [70].

Pour démontrer les avantages de l'utilisation du butanol comme carburant, David Ramey a traversé les États-Unis en 2005 sur environ 10 000 miles à l'aide d'une voiture Ford Buick de série alimentée uniquement au butanol (B100), sans aucune modification du véhicule. En effet, les avantages de l'utilisation du butanol par rapport aux autres biocarburants sont nombreux et significatifs [71].

Le biobutanol peut être utilisé pur ou en mélange avec de l'essence. Contrairement à l'éthanol, aucune modification du moteur n'est nécessaire pour son utilisation. Du fait de sa faible pression de vapeur, il est plus sûr à manipuler que d'autres biocarburants. Il n'est pas hygroscopique, ce qui signifie qu'il peut être mélangé à l'essence dans les raffineries et y être stocké. Il est moins corrosif que l'éthanol, ce qui permet d'utiliser les infrastructures existantes telles que les réservoirs, les pipelines, les pompes et les stations-service. De plus, son contenu énergétique est plus élevé, ce qui augmente le rapport kilométrage/essence [72].

De manière générale, l'utilisation de bioalcools comme biocarburant permet de réduire les émissions de CO_2 dans l'atmosphère pour une même valeur énergétique par rapport à l'essence (**Tableau 3**). Il est souvent supposé que plus le bioalcool possède une chaîne carbonée longue, plus il est intéressant en tant que biocarburant. Cependant, le propanol n'a pas suscité beaucoup d'intérêt de la part des industries en raison de son coût de production élevée par rapport à l'énergie qu'il libère, soit environ 24 MJ/L. En effet, sa densité énergétique n'est pas plus élevée que celle de l'éthanol et est nettement inférieure à celle du butanol. Il pourrait cependant être envisageable de l'introduire dans des mélanges d'alcool, tels que le mélange propanol-butanol-éthanol [73].

2.2.3. Les générations de biocarburants

Avec la Ford T, conçue pour fonctionner au bioéthanol, Henry Ford avait pour ambition de faire produire le carburant de sa voiture par des agriculteurs. Cependant, avec la diminution croissante du prix du pétrole, l'essence a rapidement remplacé le bioéthanol. Dans cette partie, les différentes générations de biocarburants, avec leurs avancées et leurs

limitations, seront exposées. De manière générale, les biocarburants sont classifiés en quatre générations suivant la biomasse utilisée pour leur production (**Figure 12**) [74].

La biomasse est par définition un ensemble de matière organique, que ce soient les récoltes agricoles telles que les grains de maïs, ou les déchets issus de ces dernières, généralement non valorisés, comme les tiges ou les feuilles. Les biocarburants se distinguent par leur caractère renouvelable, offrant une alternative aux hydrocarbures fossiles (charbon, gaz et pétrole), dont les réserves sont estimées à s'épuiser dans respectivement 54, 63 et 112 ans selon EDF [75]. Il ne fait donc aucun doute que le domaine des biocarburants constitue un élément de plus en plus important dans le paysage énergétique et chimique.

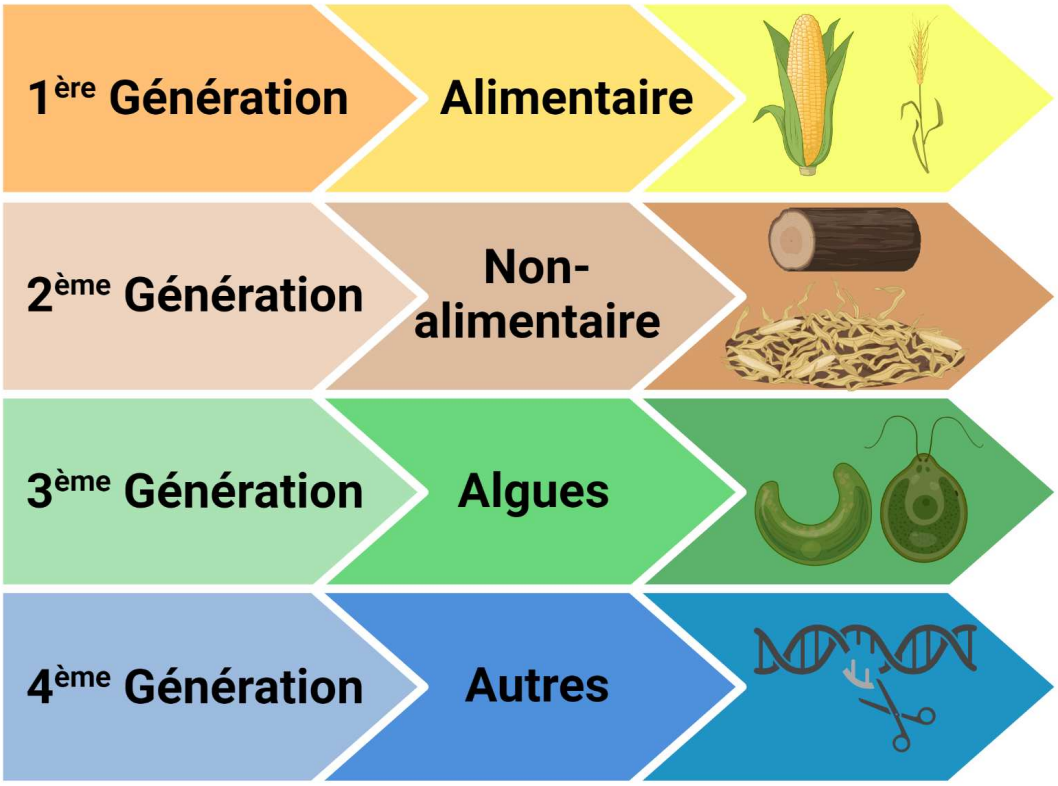


Figure 12 : Les générations de biocarburants.
Les générations de biocarburants sont classées en quatre catégories en fonction des matières premières utilisées. La 1ère génération est dérivée de ressources alimentaires comme le maïs, la 2ème de matières non-alimentaires comme la biomasse lignocellulosique, la 3ème des algues, et la 4ème explore des approches avancées utilisant des organismes génétiquement modifiés (OGM). Adapté de [74]

2.2.3.1. 1ère génération : Naissance des biocarburants

Les biocarburants de première génération sont issus de la fermentation ou de l'hydrolyse de matières premières provenant de cultures vivrières telles que la canne à sucre, la betterave sucrière, la mélasse ou encore des céréales comme le blé, l'orge et le maïs. Le maïs constitue la principale source de bioéthanol dans le monde, la majeure partie de cette production étant issue des États-Unis à hauteur de 42% de la production mondiale [76]. Pour la production d'éthanol à partir de maïs, les grains sont broyés et l'amidon qu'ils contiennent est liquéfié [77]. Le pH est ajusté pour permettre l'hydrolyse de l'amidon par l'alpha-amylase, ce qui forme des disaccharides. Ensuite, la saccharification est réalisée en ajoutant une deuxième enzyme, la glucoamylase, qui permet une seconde hydrolyse pour obtenir une solution de glucose [78]. Cette solution est ensuite fermentée par la levure *S. cerevisiae*. Une fois l'éthanol extrait, les coproduits peuvent être vendus et la biomasse cellulaire peut être utilisée comme source de protéines pour l'alimentation animale [79,80].

Malheureusement la production d'éthanol à partir de maïs n'est pas rentable. Par exemple, en 2012, le prix du maïs sur le marché était proche de 338$/tonne, aboutissant entre 400 et 450 L d'éthanol selon les procédés de production. Avec un prix de vente de l'éthanol à 0,68$/L, cela ne permettait pas d'amortir le prix d'achat du maïs par la seule vente de l'éthanol. Encore aujourd'hui, les coproduits sont donc vendus comme nourriture pour le bétail pour équilibrer les coûts [81]. De plus, cette génération de biocarburants entre en directe concurrence avec l'alimentation humaine et n'offre donc pas de solution viable à long terme.

2.2.3.2. 2ème génération : Dégradation de la lignocellulose

La seconde génération de biocarburants englobe les processus de valorisation de la lignocellulose. Celle-ci peut provenir de déchets agricoles et municipaux lignocellulosiques valorisables. La principale différence avec la première génération réside dans la matière première utilisée pour produire les biocarburants, la deuxième étant exclusivement constituée de déchets issus de l'activité humaine ou de la biomasse végétale provenant de cultures non vivrières tels que le *Miscanthus* x *giganteus* et le *Panicum virgatum* [82]. Le cycle de vie des biocarburants de deuxième génération montre une

augmentation du gain énergétique net par rapport aux biocarburants de première génération (**Tableau 4**).

Tableau 4 : Production de biomasse, production d'éthanol et superficie des terres nécessaires pour 100 Mds de Litres de carburants renouvelables aux USA.
Adapté de [83]

Matières premières	Grain de maïs	Canne de maïs	Total maïs	*Panicum*	*Miscanthus*
Biomasse exploitable (tonnes/acre)	11,12	8,15	19,27	11,37	32,62
Éthanol (gal/acre)	4235	2806	7071	3938	11206
Million d'hectares nécessaires pour 100 Mds de litres d'éthanol	4,77	7,4	2,75	5,18	1,62
% des terres cultivées récoltées en 2006 aux États-Unis	24,4	37,2	14,8	26,5	9,3

La biomasse lignocellulosique utilisée pour la production de biocarburants de deuxième génération est classée en trois catégories [84]:

- **Homogène** : comprend les copeaux de bois (coût approximatif de 100 à 120 $/tonne).
- **Quasihomogène** : comprend les résidus forestiers et agricoles (coût approximatif de 60 à 80 $/tonne).
- **Non homogène** : comprend les déchets municipaux (coût approximatif de 0 à 60 $/tonne).

Ces différentes biomasses peuvent être traitées par des procédés thermochimiques ou biochimiques (**Figure 13**).

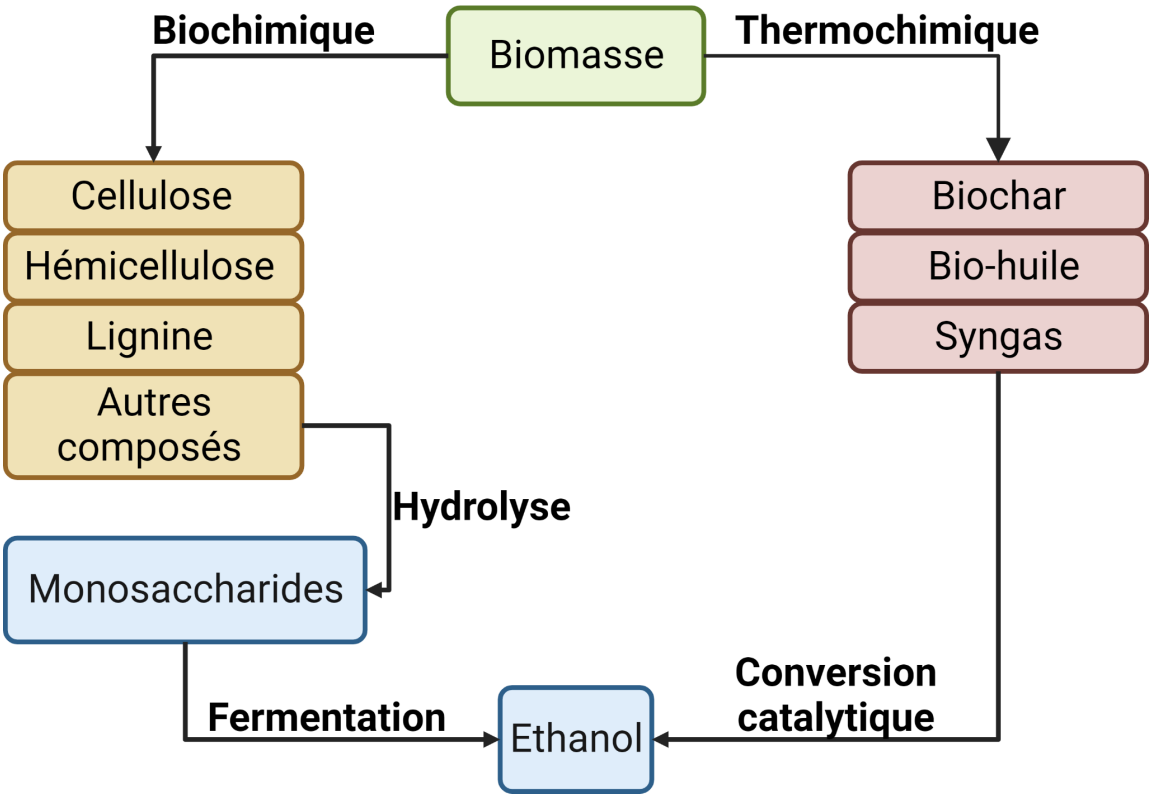


Figure 13 : Les deux voies de production de biocarburants à partir de biomasse.
La voie biochimique implique l'hydrolyse de la cellulose, hémicellulose et lignine en monosaccharides, suivie de la fermentation pour produire de l'éthanol. La voie thermochimique convertit la biomasse en biochar, bio-huile et syngas puis en éthanol via des processus catalytiques. Adapté de [81]

Dans cette partie, nous décrirons les différentes approches pour valoriser la biomasse lignocellulosique par des voies thermochimiques et biochimiques. Il est d'abord essentiel de définir précisément la composition de la biomasse afin d'exploiter pleinement son potentiel et d'adapter les prétraitements chimiques, biologiques et biochimiques en conséquence.

Composition de la biomasse

La lignocellulose, ensemble de polymères les plus abondants sur Terre, se compose principalement de trois polymères, dont deux d'origine glucidique, la cellulose et l'hémicellulose, ainsi qu'un polyphénolique, la lignine. La lignine confère une rigidité et une résistance aux parois cellulaires en se liant aux autres polymères glucidiques [84]. La cellulose, quant à elle, constitue le principal polysaccharide structurel de la paroi cellulaire primaire des plantes, composée de chaînes linéaires d'unités de D-glucose liées β(1→4) [85]. En

moyenne, la biomasse sèche se compose de 40% de cellulose, 25% d'hémicellulose, 25% de lignine, et 10% de pectine, protéines et cendres. Ces proportions peuvent toutefois varier considérablement d'une espèce de plante à une autre en fonction de leurs besoins. Les arbres présentent une concentration plus élevée de cellulose leur conférant une grande robustesse tandis que les céréales affichent un taux d'hémicellulose supérieur conférant une souplesse aux graminées [84].

La cellulose est constituée d'une chaîne de disaccharides appelée cellobiose, un dimère de glucose (**Figure 14**). Ces chaînes de cellulose s'assemblent par des interactions de Van der Waals et des liaisons hydrogènes pour former des microfibrilles, qui sont ensuite regroupées en fibres de cellulose. Dans la biomasse, la cellulose se présente à la fois sous forme cristalline et amorphe. La cellulose cristalline, la forme ordonnée, représente la majeure partie de la cellulose, tandis qu'une petite fraction, sous forme non organisée, constitue la cellulose amorphe, plus sensible à la dégradation enzymatique [86].

L'hémicellulose constitue le second polymère le plus abondant de la lignocellulose. Contrairement à la cellulose, elle n'est pas chimiquement homogène, ce qui engendre une grande variété de monosaccharides qui la composent. Ces monosaccharides comprennent des pentoses (xylose et arabinose), des hexoses (glucose, galactose, rhamnose et mannose), ainsi que des acides uroniques (acides 4-O-méthyl glucuronique, D-glucuronique et D-galacturonique) [87]. Comme la lignocellulose, la composition de l'hémicellulose varie selon les espèces végétales mais les polymères qui la composent sont également variables. Par exemple, les polymères de pentoses (pentosanes) les plus courants dans la paille et l'herbe sont le xylane, l'arabinoxylane, le glucuronoxylane, l'arabino-glucuronoxylane, l'arabino-4-O-méthyl-glucuro-noxylane et le galacto-arabino-glucuronoxylane. Les hexosanes majoritaires dans l'hémicellulose sont les glucanes et les mannanes.

Lors de la fermentation de l'hémicellulose, il est intéressant de noter que du furfural est produit, et agirait comme un inhibiteur de la croissance des micro-organismes. C'est pourquoi l'étape de prétraitement peut être indispensable dans certains cas pour éviter une perte de rendement due à des produits de dégradation toxiques [88]

Contrairement à la cellulose et à l'hémicellulose, la lignine est un hétéropolymère aromatique résultant de l'assemblage d'unités monomériques : le p-hydroxyphényl (H), le guaiacyl (G) et le syringyl (S). Ces trois monomères phénylpropanoïdes dérivent des monolignols : l'alcool p-coumarylique, l'alcool conifériylique et l'alcool sinapylique. L'abondance de chacune de ces unités monomériques varie selon l'espèce végétale et le tissu de la plante. Par exemple, la lignine se compose de trois monomères chez les graminées. Tandis que les bois tendres (gymnospermes ou conifères) contiennent une proportion proche de (G) et (S), la lignine des bois durs (angiospermes ou feuillus) est composée quasiment exclusivement de (G) [84]. Présente dans la paroi primaire des cellules végétales, la lignine confère rigidité, imperméabilité et résistance aux infections microbiennes et au stress oxydatif. Elle agit comme une colle, se liant de manière covalente à la cellulose et à l'hémicellulose pour rendre la biomasse insoluble dans l'eau [88].

En fonction de la catégorie de biomasse utilisée (homogène, quasi-homogène et non-homogène), les approches employées pour la transformer seront différentes. De fait, une biomasse non-homogène sera transformée par voie thermochimique, alors qu'une biomasse homogène pourra être utilisée par voie biochimique.

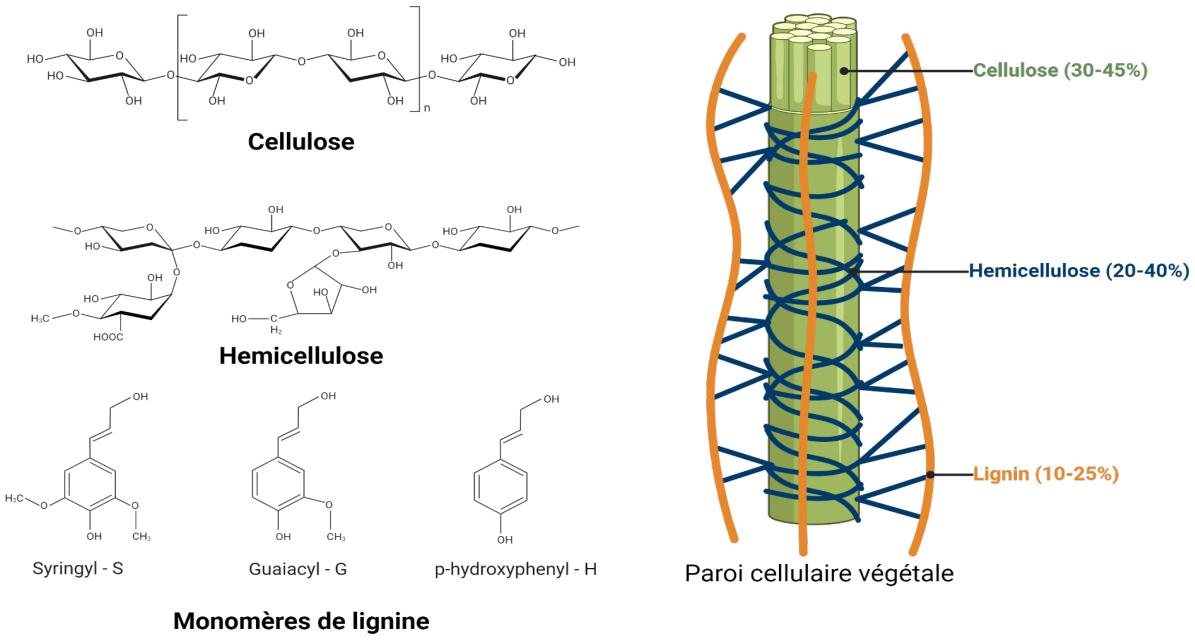


Figure 14 : Structure et composition de la lignocellulose de la paroi d'une cellule végétale.
Les trois polymères constitutifs sont : la cellulose, l'hémicellulose et la lignine, composé de syringyl, guaiacyl et p-hydroxyphényl

Transformation thermochimique

Les approches thermochimiques regroupent les différents processus au cours desquels la biomasse est soumise à des températures élevées. Selon la température utilisée, les produits obtenus diffèrent et peuvent être classés en trois fractions distinctes : solide (biochar), liquide (huile de pyrolyse) et gazeuse (syngas). À des températures basses (250 - 300°C) et en l'absence d'oxygène, la biomasse subit une torréfaction, produisant du biochar, un solide couramment utilisé pour améliorer la qualité des sols agricoles. À des températures plus élevées (550 - 750°C) et toujours en l'absence d'oxygène, le processus de pyrolyse génère de la bio-huile, comme décrit précédemment. À des températures très élevées (750 - 1200°C) et avec une faible quantité d'oxygène, la gazéification produit principalement du syngas (CO, CO_2, H_2), avec des coproduits tels que le biochar et la bio-huile. Un avantage notable de ces procédés thermochimiques est leur quasi-autosuffisance énergétique, car l'oxydation partielle ou totale de la biomasse est une réaction exothermique qui contribue au maintien de la température nécessaire à la transformation de la lignocellulose. Cette méthode permet généralement de produire de l'éthanol à moindre coût : environ 360 litres d'éthanol peuvent être obtenus à partir d'une tonne de biomasse, avec un coût de production avoisinant les 0,3$/L. Toutefois avec ce procédé, l'utilisation de biomasse homogène est peu recommandée en raison de son coût élevé ; il est préférable de réserver ce type de biomasse riche pour les processus de transformation biochimique [81,86]

Transformation biochimique

L'approche de transformation biochimique est considérée comme plus durable que l'approche thermochimique et pourrait révolutionner la manière dont nous traitons nos déchets lignocellulosiques. Ce procédé repose sur trois grandes étapes : la digestion ou prétraitement de la biomasse, l'hydrolyse puis la fermentation. L'objectif idéal est de valoriser cette biomasse en produisant des molécules à valeur ajoutée moyenne à élever. Cependant, cela nécessite une expertise approfondie en ingénierie biologique ainsi qu'une optimisation des processus de production à l'échelle industrielle. Cette section abordera les technologies et les processus actuels de valorisation de la biomasse végétale.

1. Pré-traitements de la biomasse

Pour valoriser la biomasse lignocellulosique, il est nécessaire de la rendre exploitable, cela passe par une ou plusieurs étapes de prétraitements qui peuvent être combinées selon la source de la biomasse. A l'heure actuelle, les prétraitements sont divisés en 4 catégories :

Les techniques dites **physiques** sont utilisées pour libérer les fibres de celluloses pour les rendre exploitables par la suite. Ces techniques sont : l'irradiation aux micro-ondes, l'homogénéisation à haute pression et le broyage. Le coût énergétique de ces approches les rend non-viables économiquement à grande échelle.

Les pré-traitements **physico-chimiques** incluant principalement l'explosion de vapeur d'eau. L'explosion de vapeur est un traitement combiné de chaleur et d'eau où les matières végétales sont chauffées dans de l'eau sous pression, de sorte que la température soit supérieure à 100°C et la pression bien supérieure à la pression atmosphérique. La haute température de l'eau brise les fibres de la biomasse, rendant accessible la cellulose. Les désavantages sont la libération de composés toxiques dérivant de la lignine et de l'hémicellulose.

Les pré-traitements **chimiques** tels que des acides, des bases, des solvants biologiques et organiques sont utilisés pour séparer la lignine de la biomasse lignocellulosique avec un faible taux de libération d'inhibiteurs de croissance. Ces techniques utilisent de grande quantité de réactifs onéreux et nécessitent la mise en place de matériel résistant à la corrosion.

Les pré-traitements **biologiques** tirent profit de la capacité de certains microorganismes à être cellulolytiques (généralement des *Fungi*) pour dégrader partiellement la lignocellulose. Par exemple, *Trichoderma reesei*, un champignon réputé pour sa capacité à produire des quantités importantes d'enzymes dégradant la biomasse végétale [89]. Ce type de prétraitement est le plus lent mais le plus prometteur car il est le moins coûteux en énergie et avec l'essor de la biologie de synthèse, de plus en plus d'organismes cellulolytiques comme *L. phytofermentans* émergent comme étant des châssis cellulaires

intéressants. Une fois la cellulose accessible, les enzymes cellulolytiques sont purifiées pour être ajoutées à la biomasse pour effectuer la saccharification [86].

2. Traitement de la biomasse

Après un ou plusieurs pré-traitements de la biomasse, qui dépendent du type de biomasse utilisée, plusieurs technologies sont couramment utilisées ou en développement pour valoriser la biomasse et produire des biocarburants. Parmi celles-ci, on trouve :

***Separate Hydrolysis and Fermentation* (SHF)**. La technologie d'hydrolyse et de fermentation séparées (SHF) est une méthode de production de biocarburant qui divise les étapes d'hydrolyse et de fermentation en deux phases distinctes. Bien que cette approche offre des avantages et des inconvénients, elle reste largement utilisée et étudiée. Parmi ses avantages, la SHF permet d'optimiser indépendamment les étapes d'hydrolyse et de fermentation, ce qui peut améliorer le rendement final en biocarburant. De plus, elle offre la flexibilité d'utiliser différents micro-organismes pour chaque étape, ce qui peut être bénéfique si certains micro-organismes sont mieux adaptés à l'hydrolyse ou à la fermentation. Cependant, la technologie SHF comporte également des inconvénients. La nécessité de mettre en place deux systèmes de fermentation/hydrolyse distincts peut être coûteuse et complexe. De plus, les transferts de matière et d'énergie entre les deux phases peuvent entraîner une perte de rendement en biocarburant, ainsi qu'un risque de contamination si plusieurs organismes sont utilisés. En résumé, bien que la technologie SHF offre des possibilités d'optimisation des processus, elle peut également présenter des défis logistiques et des risques associés à la séparation des étapes d'hydrolyse et de fermentation.

***Simultaneous Saccharification and Fermentation* (SSF)** et ***Simultaneous Saccharification and co-Fermentation* (SScF).** La saccharification et fermentation simultanées (SSF) ainsi que la saccharification et co-fermentation simultanées (SScF) sont deux méthodes qui fusionnent la saccharification enzymatique et la fermentation dans un même bioréacteur. L'avantage principal de la SSF est sa capacité à réaliser l'ensemble du processus de production de biocarburant en une seule étape, ce qui peut être plus efficace

en termes de rendement, de temps et de coûts [90]. Cependant, un inconvénient majeur de la SSF est que les pentoses issus de l'hémicellulose ne sont pas entièrement utilisés comme source de carbone, entraînant ainsi une sous-utilisation de la biomasse. La SSCF représente une amélioration de la SSF car elle intègre des voies de dégradation des xyloses, permettant ainsi l'utilisation des pentoses comme source de carbone. Cela conduit à une meilleure valorisation de la biomasse en exploitant l'hémicellulose comme substrat de production de biocarburants. Cependant, ces approches présentent également des inconvénients, tels qu'une diminution de l'efficacité de l'hydrolyse de la biomasse en raison de l'incorporation des cellulases et hémicellulases dans le milieu de fermentation. Cela crée un environnement moins favorable à l'hydrolyse de la biomasse. Néanmoins, cette perte d'efficacité peut être compensée par la réduction des coûts des procédés, permettant ainsi de produire du biocarburant à moindre coût par rapport à la méthode SHF [91].

Pour récapituler, pour les processus de SHF, SSF et SScF, l'hydrolyse de la biomasse est le plus souvent réalisée par des enzymes préalablement produites, principalement par l'organisme producteur de cellulase *Trichoderma reesei* [92]. Le processus d'hydrolyse de la cellulose est illustré dans la **Figure 15.** Une endoglucanase est ajoutée pour décomposer les chaînes de cellulose cristalline et les rendre amorphes. Ensuite, une exoglucanase libère le cellobiose des liaisons β-(1-4), et enfin une glucosidase est ajoutée pour convertir le cellobiose en glucose, démarrant ainsi le processus de fermentation.

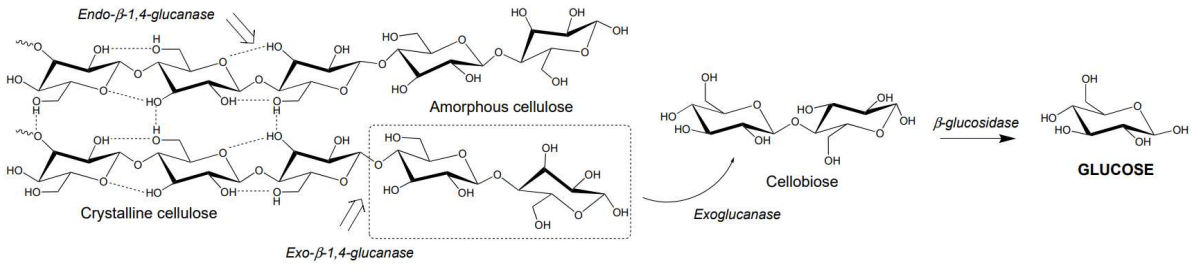


Figure 15 : Transformation par voie enzymatique de la cellulose cristalline en glucose.
Les endo- et exo-β-1,4-glucanases dépolymérisent la cellulose, libérant ainsi le cellobiose. Ensuite, la β-glucosidase intervient en clivant la liaison β 1,4, transformant ainsi le cellobiose en 2 molécules de glucose. Tiré de [86].

***Consolidated BioProcessing* (CBP).** Le *Consolidated BioProcessing* (CBP) est une approche de production de biocarburant qui intègre les étapes de production des cellulases et hémicellulases, de saccharification et de fermentation en une seule étape (**Figure 16**). Cette méthode simplifie le processus de production de biocarburant en utilisant des microorganismes capables de réaliser toutes les étapes nécessaires dans un même bioréacteur à partir de la biomasse lignocellulosique [86]. Dans la nature, ces microorganismes se trouvent notamment parmi les bactéries de l'ordre *Clostridial* ou dans le tube digestif de certains animaux comme les bovins, les ovins et les caprins. Le CBP a le potentiel d'être 25% moins coûteux que les approches SHF et SS(c)F [93]. Certains microorganismes sont modifiés par génie génétique pour améliorer leur capacité de dégradation de la biomasse, de production de biocarburant et de tolérance à celui-ci. Par exemple, *Clostridium thermocellum* a été génétiquement modifié pour être plus résistant et produire davantage d'éthanol. Une approche *bottom-up* consisterait à modifier des organismes déjà efficaces dans la production d'éthanol et à leur conférer la capacité de dégrader la biomasse lignocellulosique [94]. Depuis sa découverte dans des échantillons de sol en 2002, *L. phytofermentans*, autrefois appelé *Q Microb* par la société Qteros, est utilisé comme organisme modèle dans les CBP. *L. phytofermentans* a été identifié comme un excellent candidat pour la fermentation d'une variété de fibres végétales, produisant de grandes quantités d'éthanol avec peu de coproduits et montrant une haute tolérance à l'éthanol ce qui en fait un organisme idéal pour des applications industrielles à grande échelle [95].

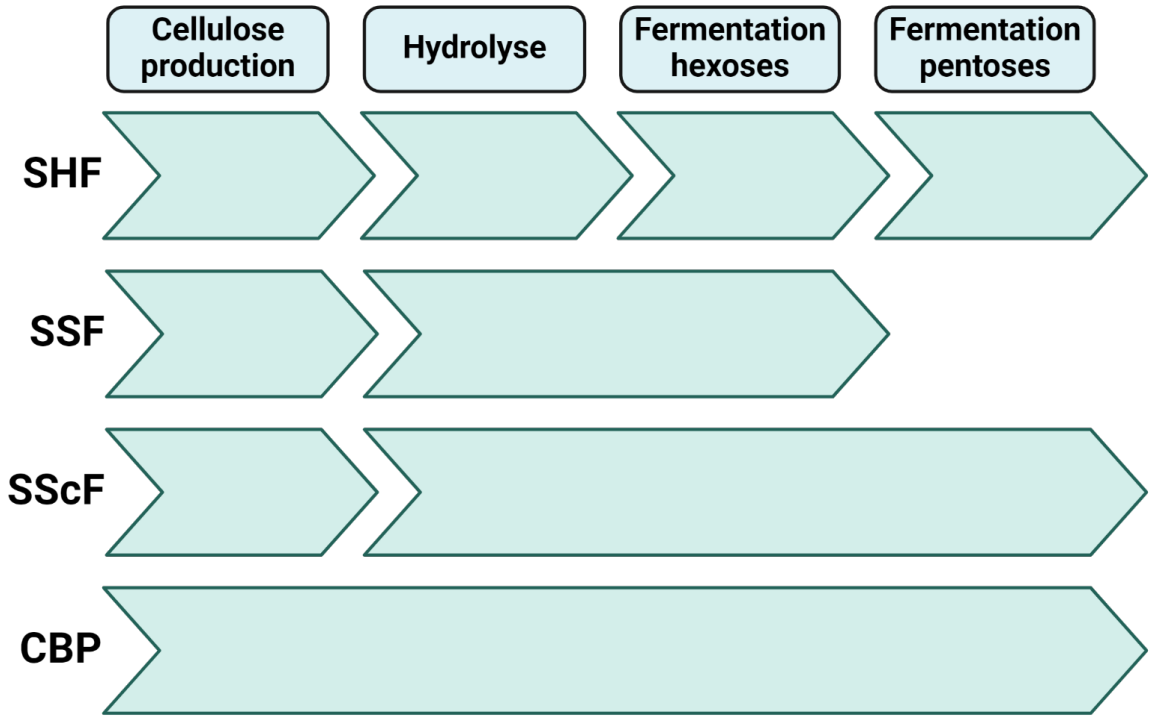


Figure 16 : Les 4 technologies de traitement de la biomasse lignocellulosique.
Abréviations SHF : Hydrolyse et Fermentation séparées, SSF : Saccharification et Fermentation Simultanée, SScF : Saccharification et co-Fermentation Simultanée, CBP : Bio-Processus Consolidé. Adapté de [93]

2.2.3.3. 3ème génération : Des lipides provenant des algues

Peu de temps après la Seconde Guerre mondiale, la culture des algues était principalement étudiée pour l'alimentation en raison de leur teneur élevée en lipides, plutôt que pour la production de biocarburants [96]. L'intérêt pour les macro-algues et les microalgues a augmenté dans les années 1970. En 1978, le programme de recherche *Aquatic Species Program*, dirigé par le Département de l'énergie des États-Unis, visait à développer un nouveau carburant à base d'huile d'algue pour concurrencer les carburants fossiles dérivés du pétrole [97]. Bien que la production à grande échelle dans des bassins extérieurs ait été démontrée comme une preuve de concept, le projet a été abandonné en 1996 en raison de la compétitivité des prix du pétrole à l'époque [97,98].

Les algues sont considérées comme ayant le potentiel de révolutionner le marché de l'énergie. Elles présentent de nombreux avantages, notamment le fait qu'elles n'ont besoin que de CO_2, de nutriments, d'eau et de soleil pour se développer. Certaines cultures d'algues peuvent même se développer dans les eaux usées, ce qui réduit les coûts en

éliminant le besoin d'eau douce. De plus, les biocarburants issus des algues pourraient produire jusqu'à 30 fois plus d'énergie que les cultures vivrières telles que le soja, en raison de leur haute productivité en biomasse et de leur capacité à stocker de grandes quantités de lipides [99]

En pratique, les algues nécessitent d'être cultivées en présence de lumière, soit en extérieur dans des bassins, soit dans des photo bioréacteurs. Les micro algues subissent une phase de privation en nutriments pour favoriser l'accumulation des lipides dans les cellules sous forme de triglycérides [100,101] Ensuite, les cellules sont lysées et des solvants ou des techniques de séparation par gravité sont utilisés pour extraire les acides gras des sucres et protéines. Le biodiesel est ensuite produit par une réaction de Trans estérification des triglycérides, avec l'ajout de méthanol pour former du glycérol et des esters méthyliques, également appelés biodiesel [102]

Malgré ces avantages, de nombreux défis doivent être relevés pour amener la production de biodiesel à partir d'algues à l'échelle commerciale et rendre son coût compétitif par rapport aux combustibles fossiles [66]

2.2.3.4. 4ème génération : Intégration de la biologie synthétique

Les biocarburants de quatrième génération proviennent d'organismes génétiquement modifiés à l'aide d'outils de biologie synthétique pour adapter leur métabolisme afin de produire efficacement le biocarburant souhaité en utilisant les sources de carbone et d'énergie les plus abordables possibles. De nombreux projets d'ingénierie métabolique sont appliqués aux micro algues photosynthétiques pour augmenter leur rendement de production et leur capacité de stockage de lipides [103–105]

D'autres approches sont proposées par des entreprises telles que Lanzatech, qui revalorise le CO_2 atmosphérique et utilisent des méthodes de gazéification de déchets municipaux et la capture directe à la sortie des cheminées d'évacuation d'usines (**Figure 17**). Lanzatech est l'entreprise la plus avancée technologiquement en termes d'ingénierie de souche pour la production de biocarburants en utilisant un *Clostridia* acétogène [31,106].

Bien que la quatrième génération soit encore à un stade précoce, elle représente actuellement le meilleur candidat pour l'avenir des biocarburants. Les trois premières générations ont montré leur efficacité par le passé, mais la biologie synthétique semble être ce qui pourrait dépasser les limites techniques et économiques des générations précédentes. Cependant, il convient de nuancer ce discours, car des projets comme celui exposé dans cette thèse sont considérés comme de deuxième génération, bien que des techniques avancées de biologie synthétique soient utilisées, permettant d'obtenir des rendements de production de biocarburants impossibles à atteindre sans ces approches.

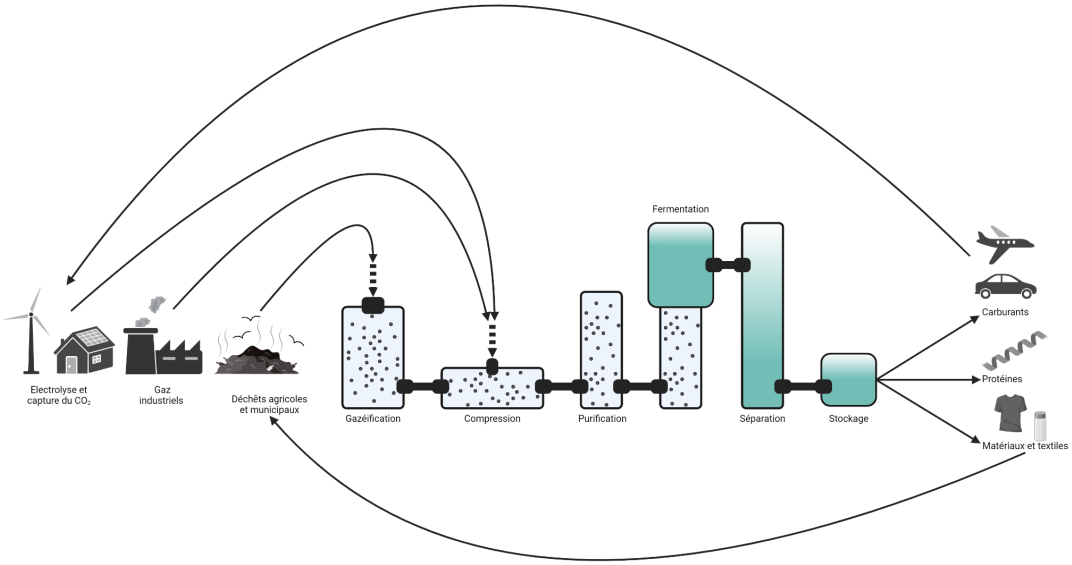


Figure 17 : Processus de production de biocarburant par l'entreprise Lanzatech.
Les gaz de synthèse sont générés par la gazéification des déchets agricoles et municipaux, la compression des gaz industriels ainsi que par l'électrolyse du CO_2 atmosphérique. Ces gaz sont ensuite comprimés, purifiés et injectés dans les cuves de fermentation pour produire les molécules d'intérêt. Enfin, les produits sont séparés, purifiés et stockés. Adapté de [31].

3. Ingénierie des *Clostridia*

L'ingénierie biologique se divise en deux approches distinctes : les approches non rationnelles, qui exploitent des paramètres inconnus pour cibler un phénotype spécifique comme l'évolution adaptative en laboratoire (ALE) et les approches rationnelles, qui requièrent une compréhension approfondie des mécanismes biologiques à manipuler. Ces deux approches se complètent mutuellement et contribuent à la mise en place de plateformes de production industrielle.

3.1. Ingénierie non-rationnelle

3.1.1. Évolution Adaptative en Laboratoire (ALE)

L'évolution est un processus permettant aux organismes vivants de s'adapter aux conditions environnementales en favorisant les phénotypes les plus adaptés. Ce processus d'adaptation implique des mutations et des recombinaisons. Pour accélérer l'évolution de microorganismes et les orienter vers un phénotype d'intérêt, il est possible de générer en laboratoire un grand nombre de générations afin d'accumuler les mutations, puis de sélectionner les individus les mieux adaptés. En milieu naturel, la plupart des mutations semblent neutres ou délétères tandis que les bénéfiques sont plus rares. En revanche, dans une expérience de ALE, grâce à la pression de sélection appliquée, les organismes porteurs de mutations délétères sont rapidement éliminés, forçant l'émergence de mutations bénéfiques au fil des générations [107].

L'ingénierie non-rationnelle en biologie présente des avantages et des inconvénients. D'une part, cette approche permet d'explorer un large éventail de solutions potentielles, offrant ainsi la possibilité de découvrir des phénotypes intéressants ou des solutions totalement innovantes. L'exemple le plus parlant est l'expérience qu'a mené Richard Lenski. Son équipe a réalisé la culture de *E. coli* durant 50 000 générations dans un milieu minimal, le résultat le plus inattendu est l'évolution pour l'utilisation de citrate, un agent chélateur du milieu. Cette évolution a permis aux bactéries ayant l'ensemble de ces mutations de croître plus rapidement, de cette façon de devenir majoritaire dans la culture [108].

Cependant, le manque de contrôle peut conduire à des résultats imprévisibles et difficiles à reproduire. Dès lors, cette approche peut être chronophage et coûteuse en temps et en ressources, nécessitant souvent des expériences répétées et des analyses approfondies. Enfin, il existe un risque que les résultats obtenus ne soient pas optimaux ou ne répondent pas pleinement aux besoins du système, en raison du manque de capacité à guider activement le processus vers des solutions spécifiques [107]

3.1.1.1. Stratégies Traditionnelles d'ALE

Différentes méthodes conventionnelles sont employées pour une expérience *Adaptive laboratory evolution* (ALE) (**Figure 18**). Ces techniques reposent sur le maintien prolongé d'une culture de microorganismes, en les diluant à intervalles réguliers ou lorsque leur croissance atteint un seuil prédéterminé.

La première consiste à réaliser des dilutions successives de cultures. Cette méthode ne nécessite aucun équipement particulier mais nécessite une personne quotidiennement et présente un risque de contamination à chaque dilution. Des dispositifs de culture continus tels que les chemostats et les turbidostats permettent de maintenir une culture en phase exponentielle, accélérant ainsi la vitesse des générations par rapport aux dilutions en série. Un chemostat est un bioréacteur où les microorganismes sont cultivés dans un environnement contrôlé. Du milieu frais est ajouté au récipient à un rythme constant, tandis que la culture est retirée au même rythme que le milieu y est ajouté, maintenant ainsi le volume de culture constant. Un turbidostat, quant à lui, contrôle le taux de dilution selon un seuil prédéfini de densité cellulaire proportionnel à la turbidité. Contrairement au chemostat, le flux de milieu frais n'est donc pas continu. Cependant, les chemostats et turbidostats sont coûteux, bien que leur capacité à contrôler la culture et les nutriments dans les expériences d'ALE soit indéniable.

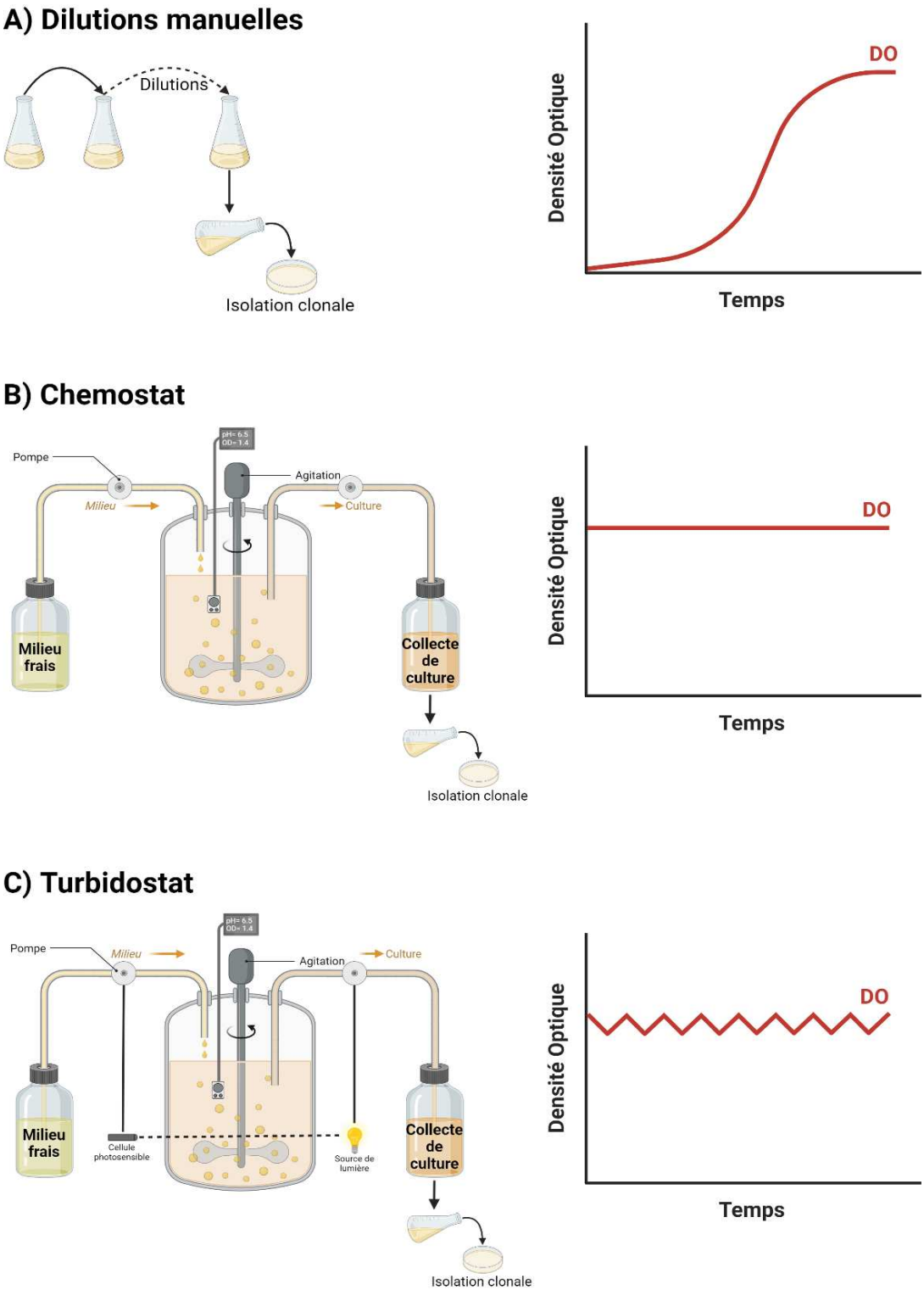


Figure 18 : Dispositifs utilisés pour l'évolution adaptative en laboratoire (ALE).
(A) Les cultures sont réalisées dans des erlenmeyers où les conditions de croissance sont contrôlées et sont diluées à fréquence fixe. Au cours de chaque cycle d'adaptation, le facteur de stress est maintenu constant ou augmente progressivement. (B) La culture continue en chemostat se déroule dans un bioréacteur où les conditions de croissance sont stabilisées par un flux continu de milieux frais et un prélèvement continu de la culture en cours, maintenant un volume stable. (C) La culture continue en turbidostat se réalise dans un bioréacteur où les conditions de croissance restent quasiment constantes. L'équilibre est régulé par la mesure de la densité optique (densité cellulaire) dans le réacteur, cela permet une évolution de la culture en s'adaptant à la vitesse de croissance des organismes à faire évoluer. Adapté de [107].

3.1.1.2. Génémate 3

Le Génémate 3 (GM3) est un dispositif unique capable de fonctionner à la fois comme un chemostat ou un turbidostat, tout en permettant l'utilisation de deux conditions de culture différentes : l'une optimale et l'autre stressante, dirigeant ainsi l'évolution vers le phénotype désiré. Initialement développé en 1999 à l'Institut Pasteur par Rupert Mützel et Philippe Marlière, ce dispositif servait à la sélection de la prolifération accélérée des cellules en suspension [109]. La version que nous utilisons au CEA-Genoscope a été grandement améliorée, permettant de cultiver des microorganismes dans deux milieux distincts : un milieu "relaxant", offrant tous les nutriments nécessaires à une croissance optimale, et un milieu "stressant", conçu pour orienter l'évolution vers des résistances spécifiques, notamment en augmentant progressivement l'exposition à une molécule d'intérêt (**Figure 19**).

Le GM3 se compose de deux chambres de volume identique, équipées d'un système fluidique pour les dilutions de cultures et l'ajout de milieu frais. Pour induire l'évolution, le programme "medium swap" est utilisé, permettant d'alterner entre les milieux relaxant (R) et stressant (S) à des intervalles définis, à la manière d'un chemostat. Un second programme, fonctionnant comme un turbidostat, réduit le temps de génération des souches soumises aux conditions stressantes, augmentant ainsi leur valeur sélective (fitness), tout en maintenant leur phénotype évolué [109].

Dans une étude menée dans notre laboratoire, le GM3 a été employé pour accroître la tolérance de *Lachnoclostridium phytofermentans* aux produits de dégradation de la lignine. Les intermédiaires issus de la synthèse des monolignols, tels que le férulate, présentent une toxicité qui constitue un obstacle majeur pour la croissance microbienne. Au fil des semaines, les concentrations de férulate dans le milieu ont été progressivement augmentées. Le milieu stressant (S) était ajouté de manière cyclique dès que la densité cellulaire atteignait un seuil critique, mesuré par turbidimétrie. Ce processus entraînait une réduction de la densité cellulaire en raison de la toxicité du férulate, suivi d'un ajout de milieu relaxant (R) pour permettre l'amplification des cellules résistantes. Ce cycle était répété jusqu'à l'obtention du phénotype souhaité.

L'efficacité du GM3 dans ce contexte a été démontrée par Cerisy et al. 2017, qui ont utilisé le dispositif en conditions anaérobies pour l'évolution de *L. phytofermentans*. En seulement 93 jours d'évolution, avec une augmentation progressive des concentrations en férulate, une souche isolée a montré une capacité à croître presque normalement en présence de 6 g/L de férulate, mettant en évidence la puissance du GM3 comme outil d'évolution pour l'acquisition de nouveaux phénotypes [110].

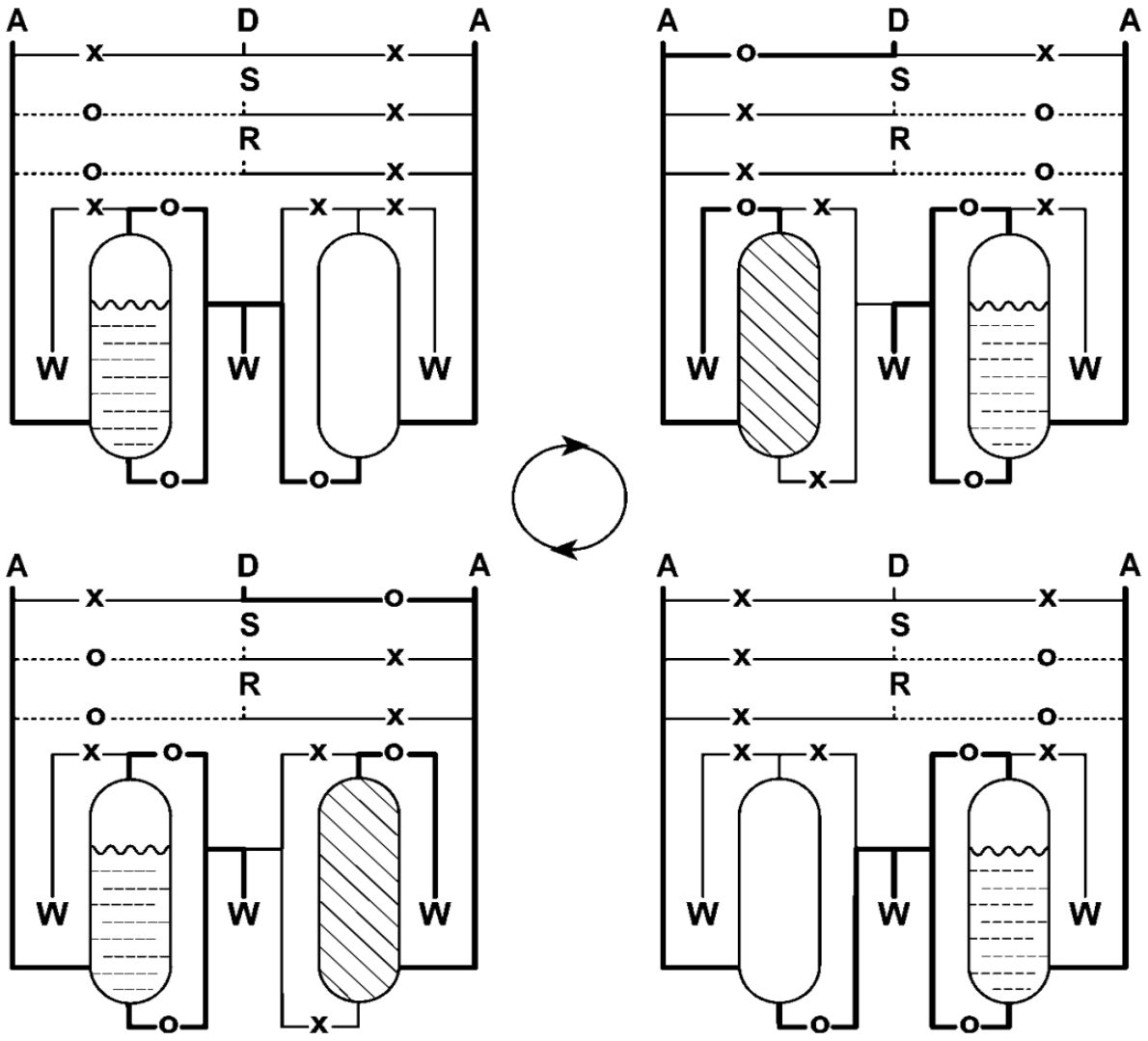


Figure 19 : Schéma de culture dans un GM3.
Les fluides s'écoulent depuis des réservoirs pressurisés contenant de l'air (A), de l'eau de rinçage (non représentée), de l'hydroxyde de sodium (D), un milieu nutritif relaxant (R) et un milieu nutritif stressant contenant la molécule pour laquelle une tolérance est voulue (S), puis lors des étapes de dilution, rinçage, et stérilisation les déchets sont envoyés vers une bouteille poubelle (W). X et O représentent respectivement les états fermés et ouverts des vannes. Les flux intermittents et constants sont représentés respectivement par des pointillés et en gras. Les tirets dans les chambres indiquent la population bactérienne et les rayures obliques, l'hydroxyde de sodium. 1) La culture est dans la chambre de gauche et vidange après rinçage de la chambre de droite. 2) Transfert de la culture de la chambre gauche vers la droite et stérilisation de la chambre de gauche. 3) La culture est dans la chambre de droite et vidange après rinçage de la chambre de gauche. 4) Transfert de la culture de la chambre de droite vers la gauche et stérilisation de la chambre de droite. Tiré de [109]

3.2. Ingénierie rationnelle et biologie synthétique

L'ingénierie rationnelle en biologie est une approche méthodique qui vise à concevoir et modifier des systèmes biologiques en s'appuyant sur une compréhension approfondie de leurs mécanismes et fonctions. La biologie synthétique, discipline émergente à l'intersection de la biologie, de la chimie, de l'informatique et des sciences de l'ingénieur, en est un parfait exemple. Elle consiste à appliquer des principes d'ingénierie pour modifier génétiquement des organismes, avec pour objectif de créer des systèmes biologiques complexes dotés de nouvelles fonctions. Cette démarche repose sur la construction de nouveaux composants biologiques, notamment des segments d'ADN fonctionnels, dans le but de mieux comprendre les mécanismes du vivant et de développer des applications innovantes en biotechnologie.

En combinant la rigueur du design rationnel avec les outils de bio-informatique, la biologie synthétique accélère les découvertes et les innovations biotechnologiques. Pour concevoir de nouveaux systèmes biologiques, il est crucial d'adopter un cadre d'ingénierie standardisé, garantissant rapidité et reproductibilité. L'un des piliers de cette approche est la méthodologie itérative DBTL (*Design-Build-Test-Learn*), qui se compose des étapes de conception, de construction, d'analyse et d'apprentissage. Chaque cycle repose sur les enseignements tirés des précédents, permettant une optimisation continue du processus [111].

Le cycle DBTL se décompose en quatre étapes (**Figure 20**) :

Design : Cette première phase consiste à concevoir le système biologique souhaité et le protocole expérimental associé. Les chercheurs utilisent des outils bio-informatiques et des connaissances préalables pour concevoir des circuits génétiques ou des voies métaboliques qui coderont les fonctions désirées. Les implications éthiques et les mesures de sécurité sont également prises en compte lors de cette phase.

Build : La construction physique des séquences d'ADN implique souvent la synthèse de l'ADN et leur clonage pour multiplier ses séquences, ainsi que son intégration génomique ou épisomale dans l'organisme hôte, également appelé châssis.

Test : Une fois le système biologique construit et intégré dans le châssis, celui-ci est soumis à des tests expérimentaux. Des quantifications par chimie analytique des rendements de production de molécules d'intérêt ou l'expression de gènes rapporteurs révèlent les activités cellulaires d'intérêt. Les données collectées sont utilisées pour évaluer les performances du système synthétique au sein de l'organisme.

Learn : Pendant cette phase d'apprentissage, les données recueillies lors des expériences sont analysées et comparées aux résultats escomptés. Les chercheurs identifient les succès, les limites et les axes d'amélioration. Ces connaissances sont ensuite utilisées pour affiner le design et apporter les ajustements nécessaires au système biologique pour l'itération suivante.

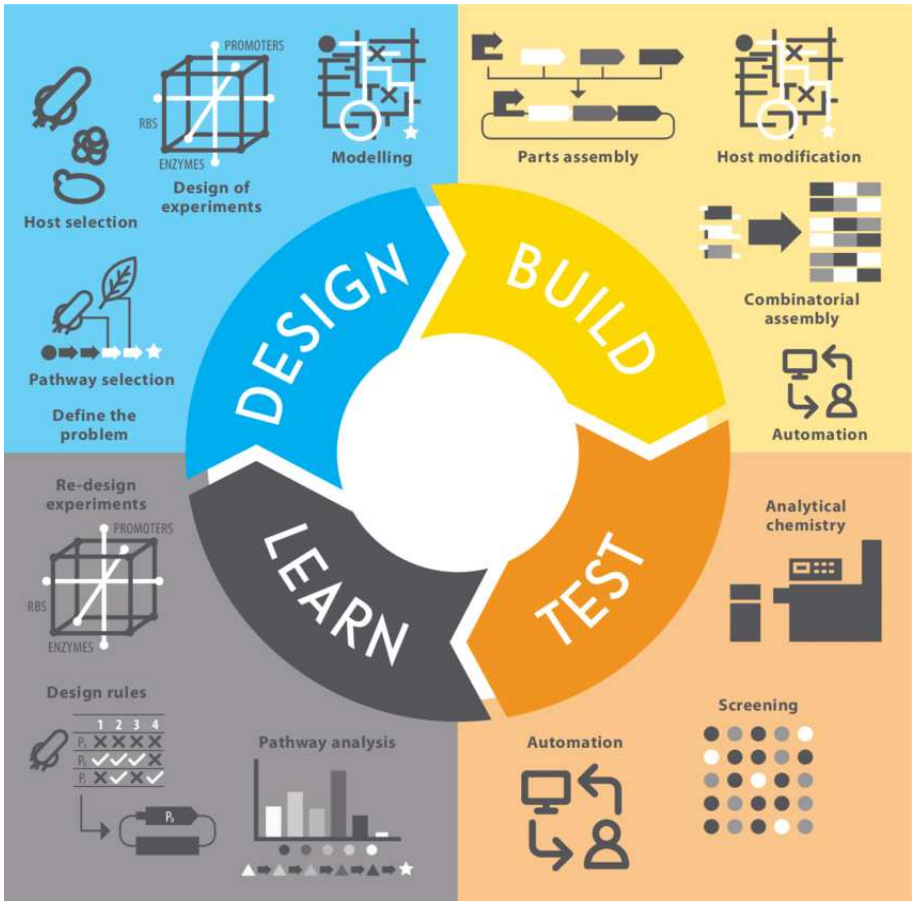


Figure 20 : Le cycle DBTL
Un processus itératif clé en biologie synthétique. Les quatre étapes essentielles du cycle DBTL sont Design, Build, Test et Learn, elles conduisent à l'optimisation des systèmes biologiques conçus et à une meilleure compréhension des mécanismes biologiques sous-jacents Tiré de [112]

3.2.1. Ingénierie du métabolisme

L'ingénierie métabolique vise à modifier le métabolisme d'un organisme, de façon permanente ou temporaire, pour améliorer ses performances métaboliques. Deux approches principales sont privilégiées dans ce domaine pour optimiser le métabolisme. La première approche consiste à introduire de nouvelles voies métaboliques pour générer un nouveau produit d'intérêt ou augmenter le rendement d'un produit déjà existant chez l'organisme hôte. Cette stratégie n'exclut pas la seconde approche, qui implique la redirection des flux métaboliques existants au sein de l'organisme. En effet, ces deux méthodes peuvent être combinées pour maximiser l'efficacité et la productivité de l'organisme modifié. Cela peut être réalisé de manière permanente par des délétions ou des insertions génomiques inactivant des gènes ciblés, ou de manière transitoire en utilisant des systèmes de répression. L'objectif principal est de combiner ces deux approches pour construire des souches optimisées à des fins industrielles. Dans le cadre de ce projet, je vais détailler comment l'ingénierie du métabolisme peut être appliquée pour la production de notre molécule d'intérêt : le butanol.

3.2.1.1. Bioproduction du butanol

La production de butanol a été initiée pour la première fois en 1861 par Louis Pasteur, qui cultiva une culture mixte de micro-organismes en conditions anaérobies [113]. Cette culture, probablement composée d'organismes de la classe des *Clostridia*, était capable de réaliser la fermentation Acétone-Butanol-Ethanol (ABE) dans un ratio de 3:6:1. Plus tard, le biochimiste Chaim Weizmann fut le premier à isoler *Clostridium acetobutylicum* et à le cultiver à des fins industrielles, tout en développant les procédés d'extraction associés. Sa méthode de production d'acétone joua un rôle crucial dans la fabrication d'explosifs à base de cordite pour l'industrie de guerre britannique durant la Première Guerre mondiale [24].

Initialement considéré comme un sous-produit indésirable de la fermentation ABE, le butanol a vu son potentiel reconnu en tant que molécule plateforme dans l'industrie chimique. Cependant, l'émergence d'un nouveau procédé chimique de production de butanol, basé sur la synthèse de l'oxyde de propylène, provoqua le déclin de la fermentation ABE, devenue non compétitive face aux procédés pétrochimiques [114]. Les coûts du butanol synthétique furent indexés au marché du propylène, rendant sa

production sensible aux fluctuations des prix du pétrole brut. Ces dernières années, l'instabilité des prix du pétrole a conduit à la réouverture de certaines usines de production de butanol en Chine et au Brésil.

Fermentation ABE chez les *Clostridia* solvantogènes

Le châssis microbien le plus étudié pour la production native de butanol est *Clostridium acetobutylicum*, reconnu pour son métabolisme en deux phases distinctes. Durant la phase acidogène, caractérisée par une croissance exponentielle, cette bactérie produit rapidement des acides gras à chaîne courte (SCFA), tels que l'acétate et le butyrate. Ensuite, lors de la phase solvantogénique, la croissance cellulaire ralentit, induisant la production de solvants comme l'acétone, le butanol et l'éthanol (ABE), tandis que les SCFA sont partiellement réassimilés comme source de carbone (**Figure 21**). Cependant, dans les fermentations en batch, le rendement maximal de butanol reste modeste, rarement au-delà de 13 g/L, en raison de sa forte toxicité pour les cellules [115].

Bien que *C. acetobutylicum* soit historiquement utilisé pour la production de butanol, plusieurs défis subsistent, tels que la vulnérabilité aux inhibiteurs présents dans les matières lignocellulosiques, une faible productivité liée à de longs temps de fermentation, et des coûts élevés dus à la production d'autres solvants. Des limitations techniques, comme l'instabilité phénotypique, la faible tolérance au butanol et le manque d'outils d'ingénierie génomique fiables, compliquent également le processus [116].

Le développement de systèmes de manipulation génétique, comme les plasmides navettes (pIM1, pMTL500E et pLK1-MCS), capables de se répliquer dans plusieurs organismes, a permis d'introduire et de maintenir du matériel génétique étranger dans les souches d'intérêt [117]. Cependant, pour faciliter la transformation génétique, il est essentiel de surmonter les systèmes endogènes de restriction. Pour ce faire, des systèmes de méthylation in vivo ont été développés pour *C. acetobutylicum ATCC 824* et *C. saccharobutylicum NCP 262*.

L'ingénierie métabolique permet également d'améliorer la production de butanol. Dans une étude menée par *Nguyen et al. 2018,* deux stratégies ont été appliquées. La première consistait à inactiver les voies métaboliques concurrentes à la synthèse de butanol

en supprimant les gènes responsables de la production de butyrate, d'acétone et de lactate (*ptb-buk, ctfAB, ldhA*). La seconde approche visait à surexprimer les gènes impliqués dans la synthèse de butanol (*atoB, hbd1, crt, bcd-etfAB, adhE2*). En combinant ces stratégies avec une méthode de purification à l'échelle de laboratoire, le rendement en butanol a été considérablement augmenté, passant de 10 g/L dans le fermenteur à 550 g/L après distillation [25].

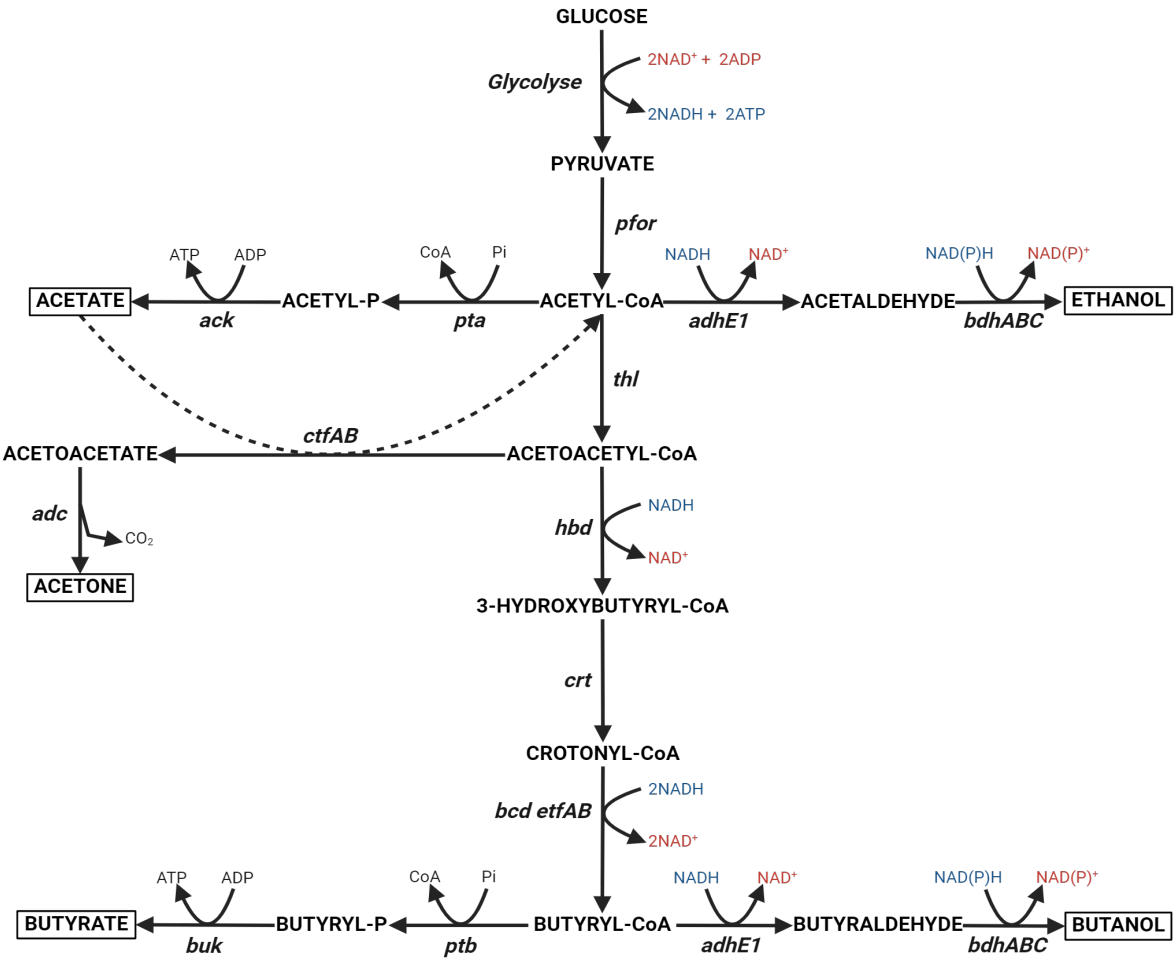


Figure 21 : Métabolisme de la fermentation ABE
Ce schéma illustre les voies métaboliques principales menant à la production d'acétone, de butanol et de butyrate à partir du glucose, via des intermédiaires clés tels que l'acétyl-CoA et le butyryl-CoA.

Production recombinante de butanol par des *Clostridia* cellulolytiques

Bien que *C. acetobutylicum* permet une production naturelle de butanol, les *Clostridia* cellulolytiques offrent un avantage distinct en raison de leur capacité à utiliser une grande diversité de substrats. Pour rendre la production de butanol rentable, plusieurs approches sont envisageables : soit par l'amélioration du rendement de production

(comme observé dans la section précédente en modifiant les souches de *C. acetobutylicum*) grâce à l'optimisation génétique et des conditions de culture spécifiques, soit par l'utilisation de sources de carbone moins coûteuses que l'amidon ou que les sucres simples.

Le CBP requiert des micro-organismes capables de dégrader efficacement la lignocellulose tout en produisant des métabolites d'intérêt, mais peu sont naturellement équipés pour cette double fonction. Actuellement, trois stratégies principales sont utilisées pour atteindre cet objectif : l'ingénierie de la voie du butanol chez les *Clostridia* cellulolytiques capables de dégrader la biomasse, la modification des *Clostridia* solvantogènes pour sécréter des cellulases libres ou des cellulosomes, et la culture mixte de *Clostridia* cellulolytiques et de *Clostridia* solvantogènes, producteurs de butanol. Chacune de ces approches offre des avantages uniques et contribue à la recherche sur la production de butanol par CBP.

Production de butanol à partir de cellulose par des *Clostridia* cellulolytiques par CBP

La première approche, c'est-à-dire l'ingénierie de la voie du butanol chez les *Clostridia* cellulolytiques, offre une alternative pour la production durable de biobutanol. En parallèle, le butanol a également été produit à partir de cellulose cristalline en introduisant une voie dépendante du CoA dans C. cellulolyticum [118]. La thiolase (*atoB*), la 3-hydroxybutyryl-CoA déshydrogénase (*hbd*), la crotonase (*crt*), la butyryl-CoA déshydrogénase (*bcd*) et la butyraldéhyde/ butanol déshydrogénase (adhE2) ont été assemblées sous le contrôle du promoteur de la thiolase Pthl de *Clostridium acetobutylicum*. Le transformant final a produit 0,12 g/l de butanol à partir de la cellulose. Cependant, le titre de butanol obtenu à partir de cellulose cristalline est encore bien inférieur à celui obtenu à partir de sucres simples. Le titre en butanol peut être encore amélioré en éliminant les sous-produits tels que l'acétate et le lactate. Toutefois, en raison de la complexité des voies de synthèse à implémenter, cinq étapes pour l'isobutanol et six étapes pour le butanol - l'ingénierie des *Clostridia* pour la production de butanol représente un défi de taille.

C. cellulovorans est un châssis *Clostridia*l cellulolytique prometteur pour la production de butanol, car il peut se développer sur la lignocellulose et accumuler une grande quantité de butyrate, le précurseur du butanol chez les souches qui le produisent [119,120]. Du fait de sa production native du butyrate, *Yang et al. (2015)* ont pu surexprimer le gène de l'aldéhyde/alcool déshydrogénase (*adhE2*) de *C. acetobutylicum* chez C. *cellulovorans* , permettant à la souche recombinante de produire 1,42 g/L de butanol en 252 h [121]. Avec le développement d'outils génétiques tels que TargeTron [122] et le système CRISPR/Cas {Citation} une voie acétone-butanol-éthanol (ABE) dépendante du CoA a été introduite chez *C. cellulovorans* pour rediriger le flux de carbone du butyrate vers le butanol, améliorant encore la production de butanol.

Dans ce projet de thèse, cette approche a été explorée en transformant *L. phytofermentans* pour lui faire produire du butanol à partir de cellulose, les stratégies et méthodes liées à cette approche seront développées dans le chapitre résultats.

Surexpression de cellulosome et cellulase chez des *Clostridia* producteurs de butanol

La souche ATCC 824 de *C. acetobutylicum* a été précédemment utilisée pour la production hétérologue et la sécrétion de composants cellulosomales individuels ainsi que de mini-cellulosomes de *C. thermocellum* et *R. cellulolyticum* [123]. Dans l'étude de *Mingardon et al. (2011)*, six gènes codant les cellulases cellulosomales les plus caractérisées de *R. cellulolyticum* ont été clonés, produits et sécrétés avec succès chez *C. acetobutylicum* [124]. Dans ces recherches, toutes les sous-unités cellulosomales, les cellulases et les protéines d'échafaudage, ont été exprimées à partir de plasmides multi-copies, leur expression étant induite par un promoteur affaibli ou un promoteur contenant l'opérateur lac. En effet, le clonage des enzymes cellulolytiques en aval d'un promoteur fort s'est avéré toxique pour les cellules d'*Escherichia coli*. Malgré cela, dans tous ces systèmes, les niveaux de protéines sécrétées, en particulier les grandes cellulases de la famille 9 et de la famille 48, étaient très faibles, et les souches recombinantes étaient incapables de se développer sur de la cellulose cristalline.

Dans une étude plus récente menée par *Chanal et al. 2011*, divers modules d'échafaudage, tels que le module de liaison aux hydrates de carbone de la famille 3a combiné aux modules X2 de *R. cellulolyticum* ou *C. acetobutylicum*, ont été identifiés comme pouvant déclencher la sécrétion de cellulases plus importantes lorsqu'ils sont introduits dans la cellulase en tant que domaine de transport. Cependant, aucune autre preuve de la solubilisation de la cellulose n'a été apportée. L'utilisation de plasmides à copies multiples pour l'expression des gènes de cellulase soulève des inquiétudes concernant la stabilité du phénotype, en particulier si l'expression des cellulases exerce un effet toxique sur la cellule, comme décrit par *Mingardon et al*. Ces problèmes ont été illustrés par la description faite par les auteurs de la génération d'une souche hétérologue de *C. acetobutylicum* capable d'exprimer Cel48F de *R. cellulolyticum*. De plus, étant donné qu'aucun plasmide clostridien ne présente une stabilité ségrégative complète en l'absence de sélection antibiotique, leur utilisation dans un processus CBP à l'échelle industrielle serait impraticable [125].

En conclusion, l'introduction du phénotype de dégradation de la cellulose chez des *Clostridia* solvantogènes est une entreprise complexe. En plus des difficultés liées au clonage des enzymes chez *E. coli*, la stratégie doit prendre en compte le niveau d'expression et d'activité des enzymes introduites pour qu'elles puissent répondre aux besoins en glucose de la cellule hôte.

Production de butanol par des consortiums microbiens en CBP

La fabrication de butanol à partir de cellulose par des consortiums microbiens basés sur le CBP représente la méthode la plus évidente pour atteindre des rendements optimaux en butanol à partir de biomasse lignocellulosique. En effet, cette approche ne repose pas sur l'ingénierie de microorganismes donc ne nécessite pas d'outils moléculaire pour la modification de ceux-ci. La charge métabolique généré par l'implémentation des phénotypes recombinants que nous avons pu voir dans les sections précédentes peuvent être un des facteurs limitant permettant la production de butanol par CBP.

Pour relever ces défis, l'utilisation de consortiums microbiens synthétiques est étudiée afin de compartimenter les fonctions métaboliques natives chez différents hôtes (**Figure 22**). Cette division du travail permet aux consortiums microbiens artificiels de réaliser des tâches complexes de manière plus efficace [44,126]. Une étude intéressante met en lumière la symbiose entre *C. cellulovorans* et *C. beijerinckii*, où *C. cellulovorans* dégrade la lignocellulose pour fournir des sucres fermentables ainsi que du butyrate à *C. beijerinckii*, qui à son tour réassimile et utilise le butyrate pour la production de butanol. Cette coopération démontre un potentiel prometteur dans la production de butanol à partir de la biomasse végétale. Cette co-culture a permis de produire 8.30 g/L de butanol à partir de 68.6g/L rafles de maïs extraits par traitement alcalin au bout de 80h [126].

Cependant, malgré les avantages potentiels des consortiums microbiens, leur utilisation pose d'autres défis. Les systèmes de co-culture composés d'espèces de *Clostridium* cellulolytiques et solvantogènes ont montré des rendements plus élevés en butanol de 2,7 à 11,5 g/L, mais une efficacité de dégradation de la biomasse nettement inférieure (de 0,12 à 0,2 g/g de cellulose) en raison de la production de co-produits. Ajoutons que la fermentation en co-culture est *difficile* à contrôler et à transposer à plus grande échelle en raison des conditions de croissance différentes pour chaque espèce impliquée. Il est donc indispensable d'optimiser les conditions de culture pour chaque couple de souches utilisé [120].

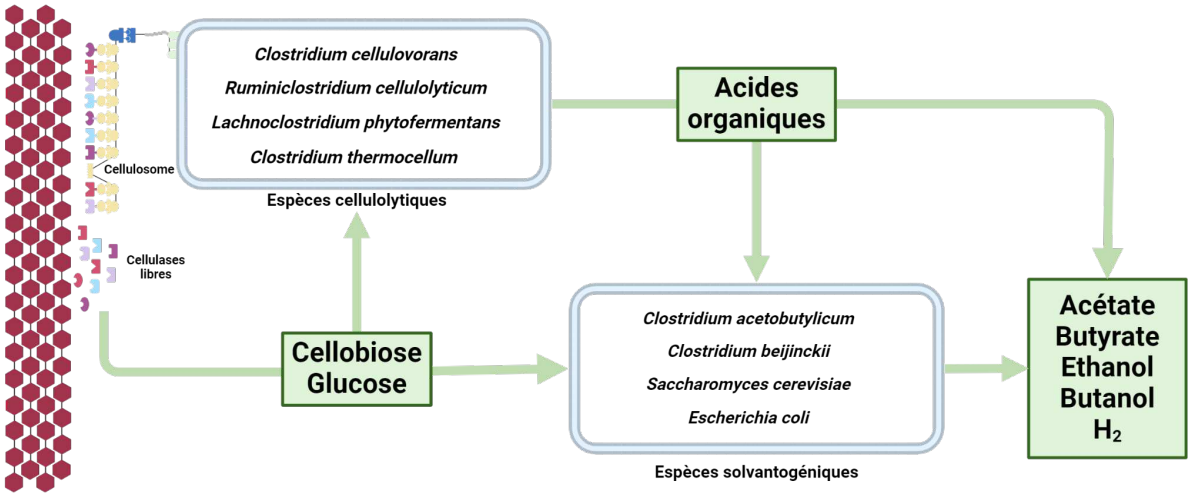


Figure 22 : Co-culture pour la production de butanol à partir de cellulose par CBP.
Les espèces cellulolytiques dégradent la cellulose en cellobiose et glucose, qui sont ensuite convertis en acides organiques par ces mêmes espèces ou des espèces solvantogéniques, menant à la production d'acétate, de butyrate, d'éthanol, de butanol et d'hydrogène.

En résumé, bien que les consortiums microbiens synthétiques offrent une approche plus pratique et réalisable pour la production de butanol, chaque méthode présentée dans cette partie comporte ses avantages et inconvénients. Des avancées futures en biologie synthétique seront nécessaires pour favoriser le développement des biocarburants de seconde génération.Compte tenu des progrès restant pour rentabiliser la production de butanol par la voie biologique chez les *Clostridia*, d'autres alternatives ont été explorées, comme l'utilisation d'autres châssis hétérologues pour la production de butanol, notamment *Saccharomyces cerevisiae* [127], *Synechococcus elongatus* [128], *Bacillus subtilis* [129] et *Escherichia coli* [130].

3.2.1.2. CRISPR/Cas : un outil génétique puissant et versatile

Dans le cas de l'implémentation de phénotypes complexes tels que des voies métaboliques ou des complexes enzymatiques, la nécessité de technique de modification génétique efficace est primordiale. En 2007, avant l'utilisation d'un système CRISPR/Cas chez les *Clostridia*, le TargeTron était utilisé pour effectuer des modifications génétiques via l'insertion site-spécifique d'introns mobiles du groupe II [122]. Malgré de nombreux projets d'ingénierie métabolique et de caractérisation fonctionnelle de gène par inactivation comme l'élucidation des mécanismes d'imports des sucres par *L. phytofermentans* [41]. L'apparition de *off-targets* et la faible fiabilité de l'inactivation des gènes ciblés restreignent l'utilisation du TargeTron [131]. Dans le même objectif, l'échange d'allèles via simple ou

double *crossover* est une technique largement utilisée pour la modification génomique de précision. Cette méthode repose sur le principe de la recombinaison homologue. Cependant, une des problématiques majeures est l'inefficacité de la recombinaison homologue couplée à la faible efficacité de transformation chez les *Clostridia* [132].

Depuis la découverte du système CRISPR-Cas et son utilisation comme outils de génie génétique par les co-prix Nobel Emmanuelle Charpentier et Jennifer Doudna dans les années 2010, l'application de ce système pour la réalisation de modifications génomiques autant chez les eucaryotes que chez les procaryotes dont les *Clostridia s'est largement développé* [133–135]. Malheureusement, là encore, le manque d'efficacité des mécanismes *Non-Homologous End Joining* (NHEJ), et *Homologous Directed Repair* (HDR) ainsi que la faible efficacité de transformation en font un système *difficile* à mettre en place chez ces organismes [136]. Afin d'outrepasser ces limitations, d'autres applications ont vu le jour.

CRISPR-Cas9

Originellement, le système CRISPR utilisé pour effectuer des modifications génomiques provient de la bactérie *Streptococcus pyogenes* (spCas9) [137]. Plus particulièrement, le système CRISPR-Cas9 appartient au type II. Dans ce type, son ARN guide (ARNg) se compose de deux molécules d'ARN : un ARN CRISPR (ARNcr) et un ARN CRISPR trans-activateur (ARNtracr). Une fois appareillés par homologie de base de séquences répétées, ces ARN forment une seule molécule guide. L'ARNcr a pour fonction de reconnaître de manière spécifique la région d'ADN à cibler, tandis que l'ARNtracr active les domaines endonucléases de la Cas9 [137,138]. L'ARNcr, contient des régions appelées "spacers" flanquées de séquences répétées (Direct Repeats - DR). Ces spacers vont permettre la spécificité du système CRISPR-Cas.

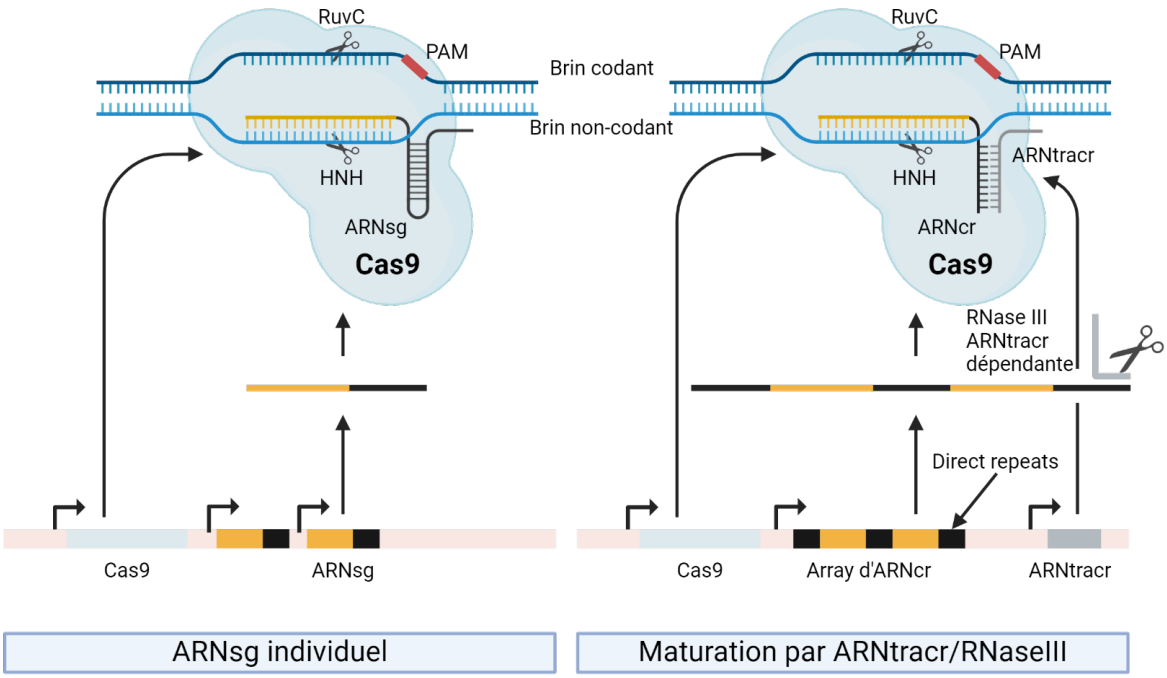


Figure 23 : Maturation du pré-ARNg de la Cas9.
Dans le cas d'un ARNsg, celui-ci est transcrit puis se complex directement avec la Cas9 pour cibler la région d'intérêt, lorsqu'un système à double ARN est utilisé, un ARNtracr est transcrit ainsi qu'un matrice d'ARNcr contenant des DR. Par action de la RNase III dépendante de l'ARNtracr, la matrice sera maturée et chaque ARNcr contiendra un spacer individuel.

Une fois associé, la Cas9 et l'ARNg vont induire une cassure double brin de l'ADN, généralement à 3 pb en amont d'un protospacer adjacent motif (PAM) 5'-NGG-3' (propre à spCas9) [139]. Pour remplacer le complexe de guidage à double ARN, un ARN guide chimérique (ARNsg) a été désigné pour combiner les propriétés de l'ARNtracr et de l'ARNcr, facilitant ainsi les applications du système CRISPR-Cas9 [140]. Cependant si le but est de cibler plusieurs régions d'ADN en même temps, aussi appelé multiplexage, l'utilisation du complexe ARNtracr-ARNcr peut être plus judicieux (**Figure 23**). D'autres alternatives à l'ARNtracr ont été explorées en intégrant des ribozymes capables d'auto-clivage ou encore des ARNt dans le pré-ARNcr [141].

La protéine Cas9 possède deux domaines nucléase responsables de son activité endonucléasique, le HNH et le RuvC. À la suite de l'appariement des bases entre l'ARNg et la séquence cible, le domaine HNH clive le brin complémentaire tandis que le domaine RuvC clive le brin non complémentaire entraînant une cassure double brin d'ADN. Cependant, si l'un de ces domaines est inactivé (notamment par la mutation D10A pour HNH ou la mutation H840A pour RuvC), cela résulte en un mutant nucléase qui introduit

une encoche sur un seul brin, lui conférant le nom de Cas9 nickase (Cas9n) [142]. Bien que la recombinaison homologue soit moins efficace, cela permet d'outrepasser la létalité de la cassure double brin, permettant la coexistence d'un mélange de population cellulaire ayant ou non intégré l'ADN d'intérêt. Cette approche nécessite donc une étape de criblage en plus pour identifier la souche au génome modifié [136]

CRISPR-Cas12a

Peu après la découverte du système CRISPR-Cas9, le système de type V CRISPR-Cas12a (anciennement CRISPR-Cpf1) a été caractérisé. Ce dernier, composé d'une endonucléase utilisant un seul ARNg, est rapidement devenu une alternative populaire à CRISPR-Cas9. Contrairement à Cas9, Cas12a reconnaît une PAM riche en T (5'-TTTV-3' où V est n'importe quelle base sauf T) localisée à l'extrémité 5' des 23 paires de bases du protospacer, ce qui diffère de l'ARNg de Cas9 où la PAM se trouve à l'extrémité 3'. De plus, Cas12a offre un avantage majeur par rapport à Cas9: elle ne dépend ni de la RNase III ni de l'ARNtracr puisque son activité RNase intrinsèque lui permet d'assurer la maturation des pré-ARNcr résultant en un ARNg plus compact et facilitant le ciblage avec plusieurs ARNg simultanément en utilisant le même plasmide (**Figure 24**) (**Tableau 5**) [137,143,144]. Un aspect intéressant de Cas12a est qu'elle possède un seul domaine endonucléasique de type RuvC, lui permettant de cliver le brin d'ADN contenant la séquence PAM à une position éloignée de l'extrémité 3' du PAM de 19 nucléotides, tandis que l'autre brin est coupé à 23 nucléotides du PAM, créant ainsi une extrémité 5' dépassant de 4 bases, rendant plus rare la religation du génome par rapport à une cassure double brin cohésive. De nombreuses études ont exploré le potentiel d'orthologues de Cas12a provenant d'une variété de bactéries pour l'édition génomique chez les cellules humaines [144,145]. En raison de ses caractéristiques uniques, un intérêt croissant s'est développé récemment sur les applications de Cas12a chez les procaryotes.

Chez les *Clostridia*, quelques études ont rapporté l'utilisation de CRISPR-Cas12a pour la répression et la suppression de gènes [146,147]. Sa capacité à reconnaître les PAM riches en T en fait un candidat idéal pour les organismes à forte teneur en AT, tels que les *Clostridia*, notamment *L. phytofermentans* dont le génome contient 65 % d'AT. De plus,

plusieurs études ont suggéré que Cas12a entraîné moins d'erreurs de ciblage et une toxicité réduite par rapport à Cas9 [7]. Pour une utilisation chez notre organisme d'intérêt *Lachnoclostridium phytofermentans*, l'utilisation du variant issu de *Lachnospiraceae bacterium*, LbCas12a semble prometteur.

Tableau 5 : Comparaison récapitulative des caractéristiques de Cas9 et Cas12a
Adapté de[143,148]

	Cas9	**Cas12a**
Taille de la protéine	1000-1600aa	1300aa
ARN CRISPR	ARNsg ou ARNcr + ARNtracr	ARNcr
Type de clivage	Extrémitées franches	Extrémitées cohésives
Nucléases	Deux domaines nucléases HNH et RuvC	Un seul domaine nucléase RuvC-Nuc
PAM	5'-NGG-3'	5'-TTTV-3'
Maturation du pré-ARNcr	Besoin de la RNase III et de l'ARNtracr	Activité RNase intrinsèque
Multiplexage	Besoin de l'ARNtracr	Aucune modification nécessaire

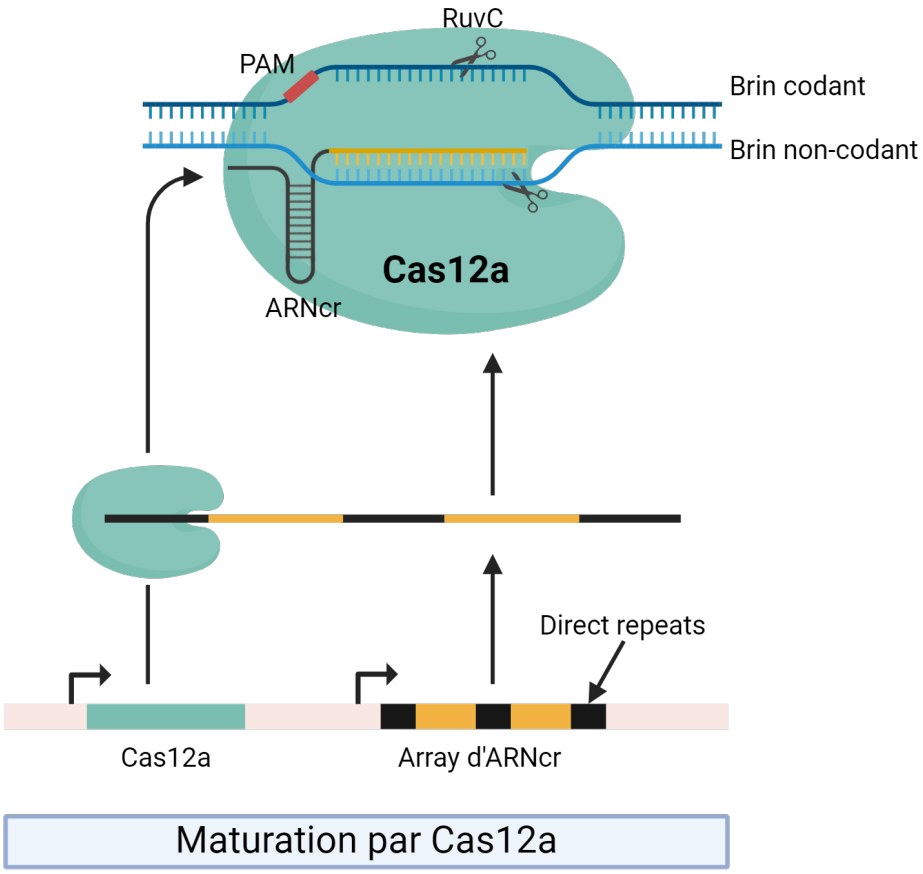


Figure 24 : Maturation du pré-ARNcr de la Cas12a.
De la même manière que dans le cas de la Cas9 avec un système à double ARN, la matrice d'ARNcr devra être maturer au niveau des DR pour générer des ARNcr fonctionnels. Ici, l'activité RNase de la Cas12a réalise la maturation de l'ARNcr.

CRISPR-Interférence

L'utilisation des systèmes CRISPR-Cas s'étend au-delà de l'édition génomique, ils peuvent être appliqués pour altérer l'expression génique. En effet, au lieu d'effectuer une délétion de gène qui impliquerait une perte totale de fonction, le CRISPR-interférence (CRISPRi) implique une Cas dites "morte" ayant ses domaines DNases inactivés par une mutation ciblée (deadCas9/12a ou dCas9/12a). Pour spCas9, les mutations D10A et H840A inactivent respectivement les domaines RuvC et HNH. Pour Cas12a, selon l'orthologue utilisé, différentes mutations dans le domaine RuvC ont été rapportées [142,148]. Ces protéines, rendues catalytiquement inactives se lient à des séquences d'ADN spécifiques définies par l'ARN guide (ARNg), empêchant ainsi l'initiation ou l'élongation de la transcription [149] (**Figure 25**). Contrairement à l'édition du génome, CRISPRi ne nécessite pas de matrice de réparation, ce qui réduit considérablement la taille des plasmides utilisés. De plus, cette

méthode permet de cibler des gènes essentiels tout en évitant les effets létaux d'une élimination totale. Des recherches approfondies ont été menées chez les *Clostridia*, mettant en évidence l'efficacité du CRISPRi dans la répression de l'expression de gènes cibles, notamment chez *C. acetobutylicum* et *C. beijerinckii* l pour lesquels la répression atteint jusqu'à 99 % [150]. L'efficacité de la répression dépend de divers facteurs, en particulier la séquence cible de l'ARNg, la localisation du ciblage par rapport au site de transcription et l'utilisation d'un seul ou de plusieurs ARNg.

Bien que CRISPRi offre des perspectives prometteuses, des défis subsistent tels que la possibilité de réprimer l'expression d'un gène contenu dans un opéron sans impacter l'expression des gènes en aval. Néanmoins, cette méthode représente un outil puissant pour étudier la fonctionnalité des gènes et offre de nouvelles possibilités dans la manipulation précise des réseaux génétiques et métaboliques [150,151].

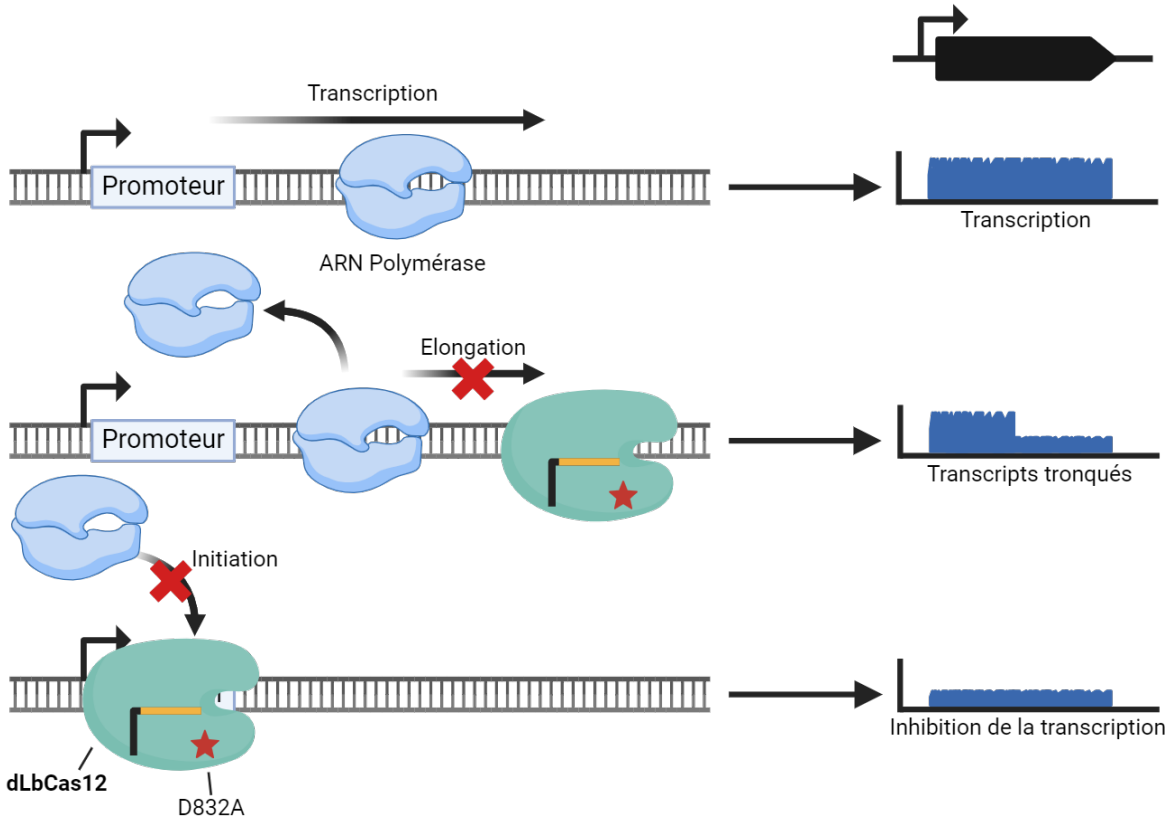


Figure 25 : Fonctionnement du système CRISPR interference.
En conditions contrôle l'ARN polymérase se fixe au niveau du promoteur pour transcrire le gène, avec la méthode CRISPRi, le gène peut être ciblé au niveau du CDS pour former des transcrits interrompus qui ne seront pas traduits ou ciblé directement le promoteur contenant le TSS pour empêcher de manière stérique l'interaction entre l'ARN polymérase et son promoteur. Pour la LbCas12a, la mutation générant un mutant inactif est D832A.

CRISPR-Base Editing

Révolutionnant l'édition génomique, une nouvelle technique basée sur CRISPR-Cas a émergé, combinant une désaminase avec le système CRISPR-Cas pour une édition nucléotidique ciblée. Cette avancée permet de modifier un seul nucléotide sans le besoin de clivage double brin d'ADN ou de HDR. Les premiers éditeurs de base (BE) ont été développés en fusionnant dCas9 avec une cytidine désaminase, initiant ainsi la conversion de C-G en T-A. Malgré des résultats montrant une efficacité d'édition faible, ces études ont suscité un intérêt dans le domaine des thérapies géniques [152]. Certaines modifications ont été apportées au système, notamment l'ajout de l'inhibiteur de l'ADN l'uracile glycosylase (UGI) et l'utilisation de la nickase Cas9 (D10A) qui ont permis d'augmenter l'efficacité de l'édition tout en minimisant les réversions C-G indésirables. En parallèle, les éditeurs de base adénine (ABE) ont été développés, permettant la conversion de A-T en G-C dans le génome. Des progrès significatifs ont été réalisés pour améliorer la spécificité et l'efficacité de ces éditeurs, élargissant ainsi leur champ d'application.

Les applications des éditeurs de base sont vastes, allant de l'ingénierie métabolique à la biologie synthétique. Par exemple, l'inactivation de gènes peut être réalisée sans cassure double brins ni recombinaison homologue, grâce à des méthodes telles que CRISPR-STOP ou i-Silence (**Figure 26**) [153,154]. Dans l'approche CRISPR-STOP, les éditeurs de cytidine ciblent les codons spécifiques CAG, CGA ou CAA dans le brin codant ou ACC dans le brin non codant et les convertir en codons stop TAG, TGA ou TAA. Si la mutation se produit avant les domaines fonctionnels de la protéine ciblée, elle conduit à une protéine tronquée donc inactive dû au codon stop prématuré [154]. De la même manière, i-Silence met en œuvre un ABE utilisé pour effectuer la conversion du codon Start ATG en GTG dans le cas du brin codant ou TAC en TGC en ciblant le brin non codant [153]

Un exemple révélateur de l'utilisation de ces outils est la mise en place d'un système d'édition de base multiplexé chez *Lactococcus lactis*, permettant l'inactivation simultanée de plusieurs gènes pour activer une voie métabolique spécifique. Cette approche offre une méthode efficace pour reprogrammer les voies métaboliques dans les microorganismes [155].

L'utilisation de ces outils chez les *Clostridia* a été limitée mais prometteuse, démontrant son potentiel pour l'étude de ces organismes et leurs applications biotechnologiques [156,157].

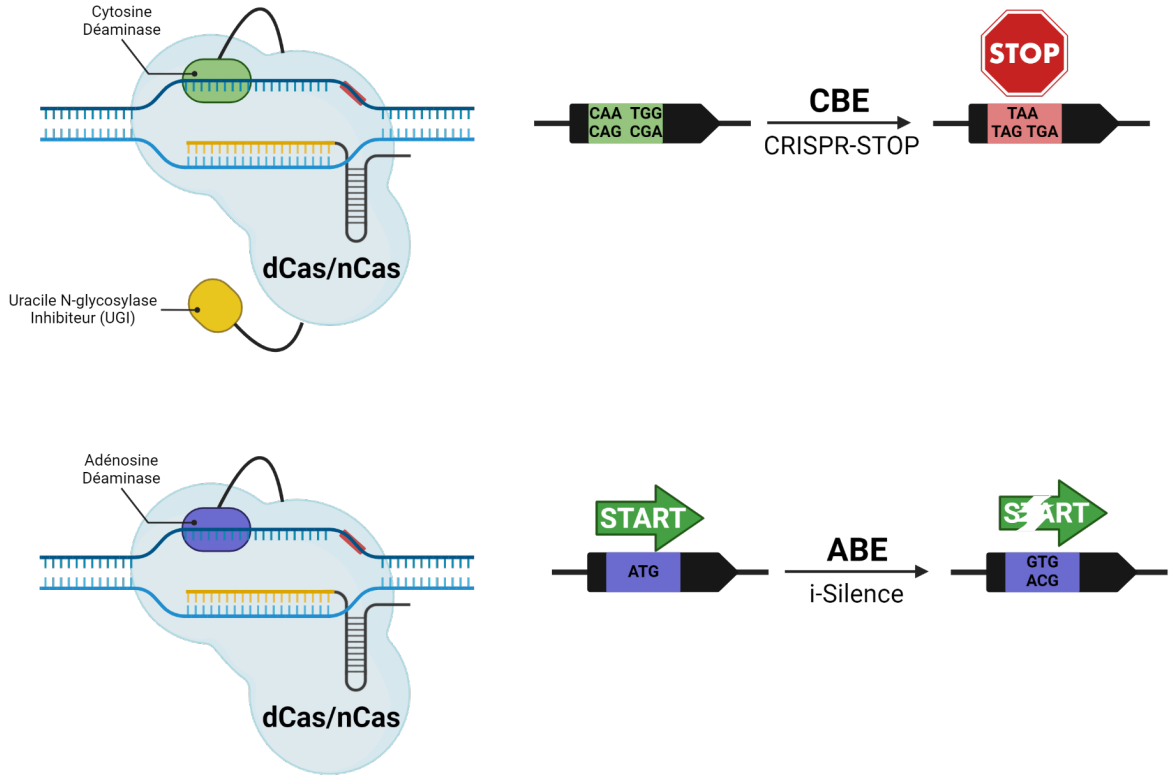


Figure 26 : Application du CRISPR Base Editing.
CRISPR-STOP utilise un dCas ou nCas fusionné à une Cytosine désaminase ainsi que l'inhibiteur de l'uracile N-Glycosylase (UGI) empêchant la réparation de la mutation U inséré dans le génome. i-Silence utilise une Adenosine deaminase pour éditer les codon start ATG

OBJECTIFS DE LA THÈSE

Les objectifs de cette thèse s'inscrivent dans une démarche d'optimisation de la conversion de la biomasse végétale en produits d'intérêt, en utilisant les *Clostridia*, une classe de bactéries anaérobies, avec un focus particulier sur *Lachnoclostridium phytofermentans*. Ce projet de recherche ambitionne de pousser les frontières de l'ingénierie métabolique de *L. phytofermentans* à travers plusieurs axes d'étude.

Dans un premier temps, l'objectif est de faire évoluer *L. phytofermentans* pour accroître sa résistance aux fortes concentrations de butanol, un biocarburant prometteur capable de remplacer les carburants fossiles grâce à ses propriétés physico-chimiques avantageuses. Cette évolution sera réalisée dans un automate de cultures continues (GM3). Une fois une souche résistante isolée, son phénotype sera caractérisé et comparé à celui de la souche sauvage WT. Des hypothèses actuelles suggèrent que cette résistance pourrait être attribuée à des modifications structurelles de la paroi cellulaire et de la membrane plasmique, lesquelles seront investiguées. Parallèlement, une analyse approfondie des mutations acquises lors de l'évolution dirigée en présence de butanol sera menée, en recourant au séquençage génomique.

Dans un deuxième temps, l'accent sera mis sur la conception et l'introduction d'une voie de synthèse du butanol chez *L. phytofermentans*, en s'appuyant sur les principes de la biologie synthétique. Cette étape exploite les concepts de standardisation, modularité et abstraction, ainsi que le processus itératif du cycle DBTL pour favoriser un prototypage rapide et efficace. La détection du butanol produit sera quantifiée par HPLC et validée à l'aide d'un biosenseur spécifique.

Enfin, l'implémentation d'un système de CRISPR interférence basé sur dLbCas12a sera réalisée chez *L. phytofermentans*. Ce système permettra, dans un premier temps, de valider la fonctionnalité de la répression génique en ciblant un gène rapporteur. Par la suite, la répression ciblée de gènes clés du métabolisme central sera mise en œuvre pour étudier son impact. Les effets de ces répressions seront évalués par qRT-PCR et HPLC, afin de comprendre comment moduler le métabolisme pour optimiser la production de biocarburants.

L'ensemble de ces objectifs vise non seulement à approfondir notre compréhension du métabolisme de *L. phytofermentans*, mais aussi à exploiter pleinement son potentiel pour la production durable de biocarburants, ouvrant ainsi la voie à des solutions énergétiques innovantes et respectueuses de l'environnement.

MATÉRIELS & MÉTHODES

1. Souches bactériennes et conditions de culture

Les souches bactériennes utilisées dans cette étude sont répertoriées dans le **Tableau 6** et les milieux utilisés dans le **Tableau 7**. Les souches d'*E. coli* ont été manipulées selon les protocoles standards (*Green et Sambrook. 2018*) [158]. Si nécessaire, les antibiotiques ont été ajoutés à la concentration suivante 25 µg mL^{-1} de chloramphénicol ou 50 µg mL^{-1} de kanamycine.

Tableau 6 : Souches bactériennes utilisées dans cette thèse

Souches	Génotypes	Sources
Escherichia coli MDS42	Délétions multiples, *ΔrecA*, M15-	[159]
Escherichia coli DB3.1	*gyrA*462, provient de la souche HB101	[160]
Lachnoclostridium phytofermentans ISDg^T	WT	[38]
Lachnoclostridium phytofermentans PHY24.0#6	Issue de l'évolution adaptative en GM3	Cette étude

Les souches de *L. phytofermentans* ont été cultivées dans le milieu GS2. La source de carbone soluble (glucose ou cellobiose), a été ajoutée à une concentration finale de 0,5 %. La cellulose, elle a été pesée et ajoutée pour être à environ 15g/L. Le milieu solide GS2-Agar est supplémenté avec 15g/L d'agar. A partir d'un stock glycérol de *L. phytofermentans*, une préculture a été réalisée (dilution au 1/10ème dans le milieu GS2). Cette préculture a été cultivée à 30°Cdans une atmosphère anaérobie en utilisant des jarres scellées (BD Jars, 260626) et des sachets réducteurs (BD GasPaks, 260678). Les cultures durent entre 24 à 48 heures selon les souches inoculées. Au bout de cette durée, les cultures ont été de nouveau diluées avec un facteur de 1/10ème dans un nouveau milieu GS2. Après 24 heures d'incubation cette culture en fin de phase exponentielle est utilisée pour les différentes expériences.

Tableau 7 : Milieux de cultures utilisé dans cette thèse

Milieux	Compositions	Sources
Lysogeny Broth (LB)	- Tryptone : 10 g/L - Extrait de levure : 5 g/L - NaCl : 10 g/L	161
S.O.C	- Tryptone : 20 g/L - Extrait de levure : 5 g/L - NaCl : 0,5 g/L - KCl : 2,5 mM - $MgCl_2$: 10 mM - $MgSO_4$: 10 mM - Glucose : 20 mM	162
GS2	Base : - KH_2PO_4 : 1,5 g/L - K_2HPO_4 : 2,9 g/L - Urée : 2,1 g/L - MOPS : 2 g/L - Citrate de sodium dihydraté : 3 g/L - Extrait de levure : 6 g/L - L-Cystéine HCl : 2 g/L - 6N KOH : 3,5 mL/L - Resazurine 0,1% : 1 mL/L Sels 10X : - $MgCl_2 \cdot 6H_2O$: 10 g/L - $CaCl_2 \cdot H_2O$: 1,5 g/L - $FeSO_4 \cdot 7H_2O$: 0,0125 g/L Source de carbone : - Glucose / Cellobiose : 5 g/L - Cellulose : 15 g/L	38

1.1. Culture en bioréacteur

Les cultures en jarres ont été réalisées dans des tubes falcon avec un volume maximal de 10 mL de milieu GS2. Des bioréacteurs (Sartorius BiostatB-DCU 5L) ont été utilisés pour mener des expériences de culture en conditions de pH contrôlé, avec un volume plus important que dans des tubes de quelques millilitres. La cuve du bioréacteur, ayant une capacité maximale de 5 L, a été utilisée avec un volume de travail de 2 L de milieu. Le milieu GS2, utilisé dans les bioréacteurs, contenait comme source de carbone du glucose à une concentration de 0,5 %. Ce milieu a été inoculé avec une culture saturée de *L. phytofermentans* de 50 mL, diluée au 1/40ème. Les conditions de culture dans les bioréacteurs ont été rigoureusement contrôlées : le pH a été maintenu à 7 par l'ajout d'une

solution de NaOH ou de KOH, tandis que la température a été maintenue constante à 30 °C à l'aide d'un cryostat. Une agitation douce (100 RPM) a été appliquée pour assurer une bonne répartition du NaOH ou du KOH dans le milieu de culture. Des prélèvements régulier (toutes les 2 heures) ont été réalisés pour analyser les cinétiques de fermentation au cours du temps.

1.2. Évolution dirigée en GM3 anaérobie

Des clones de *L. phytofermentans* résistants au butanol ont été sélectionnés à l'aide d'un automate GM3, un dispositif de culture continue à deux chambres, conçu pour maintenir des conditions anaérobies grâce à un flux continu de gaz N_2 à 100 % dans les cultures. Une culture de 50 mL dans du milieu GS2 a été maintenue à 30°C, avec des mesures de turbidité effectuées toutes les 30 secondes. La culture a été transférée entre les chambres de croissance à fréquence fixe permettant de nettoyer la chambre vide avec de l'hydroxyde de sodium 5 N. Les bactéries ont été acclimatées à des concentrations croissantes de butanol en utilisant le mode "*medium-swap*", un chemostat modifié, où le temps de génération est fixé au préalable. Au long de l'évolution, les concentrations en butanol ont été augmenté graduellement (0-0,1-0,2-0,3-0,5-0,75-1-1,25-1,5-2-2,5 %). Des dilutions du milieu stressant à haute concentration de butanol (*e.g.* 1,5 %) ont été effectuées toutes les 30 minutes si la densité cellulaire dépassait un seuil (mesuré à DO_{880} de 30, environ équivalent à une DO_{600} de 0,4), tandis que le milieu relaxant à faible concentration de butanol (*e.g.* 1 %) a été utilisé dans le cas contraire. Une fois que les densités cellulaires se sont stabilisées à une valeur constante dans le milieu stressant, le GM3 a été basculé en mode turbidostat, utilisant exclusivement le milieu stressant jusqu'à ce que le temps de génération de la culture se rapproche de 3,75h, correspondant au temps de génération de la souche WT en condition contrôle. En mode turbidostat, 20 % du volume de culture a été remplacé par du milieu frais à chaque fois que la densité cellulaire atteignait une OD_{880} de 30. Des prélèvements ont été réalisés puis conservés à -80°C pour des analyses ultérieures.

1.3. Analyse de la croissance de L. phytofermentans

Pour observer les profils de croissance de différentes souches de *L. phytofermentans* dans diverses conditions (présence ou absence de butanol, expression d'un système

CRISPR ± induction, ou en présence de différentes sources de carbone). En ayant, la courbe de croissance d'une souche dans une condition donnée on peut comparer les souches et conditions entre elles en regardant le profil de la courbe ainsi que le temps de génération de celle-ci. Pour réaliser ces expériences, des cultures en fin de phase exponentielle ont été utilisé pour inoculer les différentes conditions. Les différents consommables et milieux ont été préparés la veille, pour être réduits dans une boîte à gants (Mbraun LABmaster Glove Box Mega 2). La plaque 96 puits (TC Plate 96 Well, Suspension,F Sarstedt) a été remplie avec 200µL de milieu GS2 inoculé ou non et scellée avec un pad autocollant (Tape Pads, Qiagen), puis cultivée à 30°C sans agitation dans un lecteur de plaque (CLARIOstar® BMG LabTech) pendant une durée variant selon l'expérience de 72 à 96 heures. La croissance bactérienne a été calculée en mesurant toutes les 20min la DO_{600nm} en procédant à une agitation de 30 sec à 300 RPM avant chaque mesure.

1.4. Quantification de la dégradation de la cellulose

Pour quantifier le phénotype cellulolytique, *L. phytofermentans* a été cultivé en présence de cellulose comme seule source de carbone. Environ 150 mg de cellulose (Whatman® qualitative filter paper, Grade 1) pour 10 mL de culture ont été pesés et autoclavés directement dans des tubes à sec. Le milieu GS2 a été ensuite inoculé au 1/10ème avec une culture diluée la veille. Les cultures ont été incubées dans une jarre anaérobie à 30°C pendant toute la durée de l'expérience. Chaque semaine, les tubes de cultures indépendants ont été récupérés, la cellulose restante a été filtrée à l'aide d'un filtre de papier Whatman préalablement pesé. Les filtres contenant les résidus de cellulose dégradée ont été séchés à 55°C pendant 48 heures, puis pesés. La différence entre la masse initiale de cellulose et celle obtenue après dégradation a été calculée, permettant ainsi de quantifier l'activité cellulolytique de L. phytofermentans. Une partie des surnageants a été récupérée pour analyser les concentrations des produits de la fermentation par HPLC.

2. Techniques de biologie moléculaire et génie génétique

2.1. Plasmides utilisés dans cette étude

Tous les plasmides utilisés et construits au cours de cette étude sont répertoriés dans les **Tableaux 8 et 9**.

Tableau 8 : Plasmides de départs utilisés dans cette thèse

Plasmides	Description / But	Source
pAB_PalkB_sfGFP	*PalkB-sfGFP, Biosenseur butanol plasmide output*	[163]
pCK01_PalkS_alkS	*Palks-alkS(L401G), Biosenseur butanol plasmide input*	[163]
pSol5	*Amplification du gène ccdB*	[164]
pQmod3C-GG	*Plac-RFP,, catP, pCB102 Gram(+) origin, ColE1 Gram(−) origin, MCS*	[165]
pQdC12a	*PGusA2-TetO2/1-dCas12a, cassette miniPthl-tetR, ARNg, pQmod3C-GG*	[165]
pQnl_Pcphy23	*erm, pAMβ1 Gram(+) origin, pUC Gram(−) origin, Pcphy23-Nanoluc*	[165]
DVA-AB, BC, CD, DE, EF	*pSB1A2, LacZ, BsaI/BbsI free, extrémités cohésives variable*	[166]

Tableau 9 : Plasmides construits dans cette thèse

Le numéro du plasmide est le numéro regroupant des plasmides de même famille ayant des variants au niveau de l'insert. Les plasmides en gras sont les plasmides plateforme avant clonage des inserts. Les codes à 8 chiffres de pTZ011 représentent l'ordre des CDS contenu dans les opérons. (e.g 1234-5678 correspond à l'assemblage des inserts dans l'ordre pTZ001, pTZ002, pTZ003, pTZ004 sous le contrôle du promoteur pCons17 puis pTZ005-pTZ006-pTZ007-pTZ008.

N°	*Backbone*	Insert et Description
pTZ001	pSB1A2, LacZ	DVA-AtoB-AB
pTZ001	pSB1A2, LacZ	DVA-AtoB-BC
pTZ001	pSB1A2, LacZ	DVA-AtoB-CD
pTZ001	pSB1A2, LacZ	DVA-AtoB-DE
pTZ002	pSB1A2, LacZ	DVA-AdhE2-AB
pTZ002	pSB1A2, LacZ	DVA-AdhE2-BC
pTZ002	pSB1A2, LacZ	DVA-AdhE2-CD
pTZ002	pSB1A2, LacZ	DVA-AdhE2-DE
pTZ003	pSB1A2, LacZ	DVA-Crt-AB
pTZ003	pSB1A2, LacZ	DVA-Crt-BC
pTZ003	pSB1A2, LacZ	DVA-Crt-CD
pTZ003	pSB1A2, LacZ	DVA-Crt-DE
pTZ004	pSB1A2, LacZ	DVA-Hbd-AB
pTZ004	pSB1A2, LacZ	DVA-Hbd-BC
pTZ004	pSB1A2, LacZ	DVA-Hbd-CD

pTZ004	pSB1A2, LacZ	DVA-Hbd-DE
pTZ005	pSB1A2, LacZ	DVA-Term+pThl-EB
pTZ006	pSB1A2, LacZ	DVA-Bcd-BC
pTZ006	pSB1A2, LacZ	DVA-Bcd-CD
pTZ006	pSB1A2, LacZ	DVA-Bcd-DF
pTZ007	pSB1A2, LacZ	DVA-EtfA-BC
pTZ007	pSB1A2, LacZ	DVA-EtfA-CD
pTZ007	pSB1A2, LacZ	DVA-EtfA-DF
pTZ008	pSB1A2, LacZ	DVA-EtfB-BC
pTZ008	pSB1A2, LacZ	DVA-EtfB-CD
pTZ008	pSB1A2, LacZ	DVA-EtfB-DF
pTZ010	**pQmod3-CamR**	**pCons17-Plac-RFP flanked by BsaI AF sites catP, pCB102 Gram(+) origin, ColE1 Gram(−) origin, MCS**
pTZ011	pQmod3-CamR	pCons17-1243-5687
pTZ011	pQmod3-CamR	pCons17-1324-5687
pTZ011	pQmod3-CamR	pCons17-2134-5687
pTZ011	pQmod3-CamR	pCons17-2143-5687
pTZ011	pQmod3-CamR	pCons17-2314-5687
pTZ011	pQmod3-CamR	pCons17-2341-5687
pTZ011	pQmod3-CamR	pCons17-2413-5687
pTZ011	pQmod3-CamR	pCons17-2431-5687
pTZ011	pQmod3-CamR	pCons17-3142-5687
pTZ011	pQmod3-CamR	pCons17-3421-5687
pTZ012	pSC101-KanR	Palks-alkS(L401G)-PalkB-sfGFP
pTZ014	**pSB1A2, LacZ**	**DVA_Cas12a spacer cloning BbsI (AGAT-Spacer-GTTT)**
pTZ015	pSB1A2, LacZ	Spr_1326(2)-1327(2)-0089(2)
pTZ015	pSB1A2, LacZ	Spr_3925(2)-2463(2)-1029(2)
pTZ015	pSB1A2, LacZ	Spr_1326(2)-1327(2)
pTZ015	pSB1A2, LacZ	Spr_1232(2)-1117(2)
pTZ015	pSB1A2, LacZ	Spr_1326(2)-1327(2)-1232(2)-1117(2)
pTZ016	**pQmod3-CamR**	**PGusA2-TetO2/1-dCas12a, miniPthl-tetR, cassette ARNg *ccdB*-BsaI, pQmod3C-GG**
pTZ017	pQmod3-CamR	Spr_1326(2)-1327(2)-0089(2)
pTZ017	pQmod3-CamR	Spr_3925(2)-2463(2)-1029(2)
pTZ017	pQmod3-CamR	Spr_1326(2)-1327(2)
pTZ017	pQmod3-CamR	Spr_1232(2)-1117(2)
pTZ017	pQmod3-CamR	Spr_1326(2)-1327(2)-1232(2)-1117(2)

2.2. Méthode d'amplification et de clonage

2.2.1. Amplification par PCR

Les réactions PCR pour les clonages ont été effectuées avec la polymérase à haute-fidélité Q5 (New England Biolabs, M049L) en utilisant le protocole du fournisseur. La température d'hybridation des amorces a été calculée à l'aide de l'outil en ligne (https://tmcalculator.neb.com/). Les amorces sont répertoriées dans le **Tableau 11**. Elles ont été commandées auprès de Sigma-Aldrich.

Les réactions d'amplification dans le cas de vérification de constructions ont été réalisées par PCR en utilisant le OneTaq® Quick-Load® 2X Master Mix (New England Biolabs, M0486L). Les produits de PCR ont été analysés par électrophorèse sur gel d'agarose pour confirmer la taille des fragments. Si nécessaire ceux-ci ont été purifiés avec le kit Qiagen, QIAquick PCR Purification et ont été utilisés pour un clonage Golden Gate, Gibson ou clonage restriction classique.

2.2.2. Assemblage par Golden Gate

Pour un volume de mélange réactionnel de 10 µL, un minimum de 100 fmol de chaque fragment d'ADN a été mélangé avec 1 µL de tampon T4 DNA ligase, 1 µL d'enzymes BsaI (R3733S) / BbsI (R3539S) / Esp3I (R0734S) et T4 DNA ligase (M0202L) (NEB), et la quantité appropriée d'eau. Les tubes contenant les mélanges réactionnels ont ensuite été placés dans un thermocycleur avec le protocole suivant :

- Étape 1 : 37°C - 20 minutes
- Étape 2 : 37°C - 1,5 minutes
- Étape 3 : 16°C - 3 minutes
    - Étapes 2-3 : répétés 5 à 25 fois (selon le nombre d'inserts)
- Étape 4 : 50°C - 5 minutes
- Étape 5 : 80°C - 10 minutes

Le temps total de la réaction est compris entre 57,5 et 111,5 minutes. Les échantillons ont ensuite été conservés à 4°C jusqu'à leur récupération pour transformation chez *E. coli* en utilisant 5 µl du mix obtenu.

2.3. Transformation bactérienne

2.3.1. Transformation par choc thermique de E. coli

La préparation des cellules E. coli rendues chimiquement compétentes ainsi que leur transformation par l'ADN plasmidique d'intérêt sont effectuées selon les protocoles standards décrits par Sambrook et Russell (2006) [167]

2.3.2. Électroporation de *L. phytofermentans*

Pour préparer à l'électroporation les cellules de *L. phytofermentans*, 10 mL de culture en fin de phase exponentielle ont été dilués dans 40 mL de milieu frais. Cette culture a été incubée jusqu'à atteindre une DO_{600} entre 0,6 et 0,7. Les cellules ont été centrifugées à 7000 g pendant 5 min à 4°C, lavées une fois avec 4 mL de tampon ETM glacé (composé de 1 mM de MgCl2, 256 mM de sucrose et 5 mM de PBS, pH 7,4), puis remises en suspension dans 200 µL de tampon ETM. Un mélange de 50 µL de cellules électrocompétentes et de l'ADN plasmidique (1 µg d'ADN, avec un volume maximal de 10 µL) a été transféré dans une cuvette d'électroporation de 0,1 cm d'écart pré-réfrigérée, et incubée sur la glace pendant 5 minutes. Les cellules ont ensuite été électroporées à l'aide d'un Genepulser II Bio-Rad réglé à 1,2 kV, 25 µF et immédiatement diluées dans 1 mL de milieu GS2-glucose et enfin incubées en anaérobie à 30°C pendant 16 heures. Les cellules ont été étalées sur des boîtes de milieu GS2-agar contenant l'antibiotique nécessaire à la sélection et cultivées entre 4 et 6 jours à 30°C. Les transformants obtenus ont été cultivés en milieu GS2 liquide contenant l'antibiotique approprié. La présence d'ADN plasmidique dans a été confirmée par PCR.

3. Approche de biologie synthétique

3.1. Adaptation et optimisation du biosenseur AlkS

L'adaptation du biosenseur basé sur le facteur de transcription AlkS [163] a été réalisé en utilisant les plasmides pAB_PAlkB_sfGFP (Addgene Plasmid #166503) et pCK01_PalkS_alkS (Addgene Plasmid #166502) en amplifiant les fragments d'intérêt de chacun des 2 plasmides par PCR. D'une part, pAB_PAlkB_sfGFP a été amplifié avec les amorces oTZ170 et oTZ171 pour y ajouter des sites Esp3I, ce fragment contient le gène de

résistance *KanR*, ainsi que *sfGFP* dont l'expression est contrôlée par pAlkB. D'autre part, pCK01_PalkS_alkS a été utilisé comme matrice pour extraire l'origine de réplication pSC101 et le gène *alkS* ainsi que son promoteur, les amorces oTZ172 et oTZ173 ont été utilisé pour ajouter les sites Esp3I. Le plasmide recombinant obtenu (pTZ012) a été vérifié par séquençage de Sanger puis testé pour s'assurer de sa fonctionnalité en réalisant une plaque de culture avec et sans 10mM de butanol et en mesurant la fluorescence émise.

3.2. Construction des voies de biosynthèses du butanol

Les clonages de la voie de synthèse modulaire du butanol ont été réalisés en utilisant des plasmides accepteurs issus du kit CIDAR MoClo (Kit #1000000059) [166]. Dans un premier temps, les codons des gènes ont été harmonisés afin d'améliorer le repliement des enzymes et d'éliminer d'éventuels sites de restriction BsaI et BbsI à l'aide de l'outils https://codonharmonizer.systemsbiology.nl/. Des séquences de RBS (*Ribosome Binding Sites*) avec un TIR (*Translation Initiation Rate*) défini à 15 000 ont été généré pour chaque CDS (*Coding DNA Sequence*) codant les enzymes de la voie de synthèse du butanol avec l'outils https://www.denovodna.com/. Pour chaque gène, des amorces spécifiques ont été utilisées pour ajouter par PCR les sites de restriction BbsI ainsi que les extrémités cohésives (A, B, C, D ou E en 5' et B, C, D, E ou F en 3'). Une réaction Golden Gate a été effectuée en utilisant les enzymes BbsI et T4 Ligase, permettant de cloner les fragments PCR de chaque gène dans les plasmides accepteurs L0-DVA ayant les extrémités cohésives correspondantes (AB, BC, CD, DE ou EF). Après transformation, les colonies blanches ont été repiquées, validées par PCR, puis la totalité de l'insert a été vérifiée par séquençage.

Dans une seconde réaction Golden Gate, les plasmides L0-DVA contenant les gènes d'intérêt (pTZ001–pTZ008) ont été utilisés comme donneurs, et le plasmide pQ3_Cam_AF (pTZ010) comme accepteur, avec les enzymes BsaI et T4 Ligase. Le pQ3_Cam_AF a été modifié à partir du plasmide pQmod3C-GG (*Addgene Plasmid #191350*) en réalisant une PCR avec les amorces oTZ111 et oTZ114. Cette PCR a permis d'amplifier et d'ajouter en amont de la cassette d'expression de la RFP un promoteur fort et constitutif, pCons17, flanqué par les sites de restriction SacI et NheI. Le fragment PCR et le plasmide pQmod3C-GG ont été digérés et assemblés par restriction-ligation pour insérer la cassette, modifiant

les sites de clonage Golden Gate afin de générer un plasmide accepteur avec des extrémités cohésives BsaI adapté au MoClo.

Les plasmides finaux contenant la voie complète de synthèse du butanol (pTZ011) ont été vérifiés par PCR avec les amorces oTZ110 et oTZ115. Les plasmides extraits des colonies positives ont ensuite été séquencés par la technologie Oxford Nanopore pour confirmer la précision de l'assemblage et l'absence de mutations non désirées dans l'entièreté du plasmide.

3.3. Clonage de guides multiplexés dans pQdC12a

Des améliorations ont été apportées au système de clonage de guides dans le plasmide pQdC12a pour optimiser le processus de clonage. Tout d'abord, la plateforme de clonage initialement basée sur le site de restriction BbsI pour l'insertion de spacers par Golden Gate a été remplacée par un gène *ccdB* flanqué de sites BsaI, plus performant que BbsI pour cette méthode. Cette modification a été réalisée en réalisant une PCR sur le plasmide pSol5 avec les amorces oTZ120 et oTZ121 pour ajouter des sites BsaI et BbsI se faisant face pour insérer *ccdB*. Le plasmide obtenu a été séquencé par séquençage de Sanger avec les amorces oTZ003 et oTZ115.

Pour la construction des plasmides CRISPRi, des spacers spécifiques ont été clonés dans la nouvelle plateforme de clonage (pTZ016). Ces spacers ont été conçus pour être insérés à partir d'amorces possédant des extrémités cohésives compatibles avec celles générées par la digestion du plasmide accepteur. Les amorces ont d'abord été phosphorylées à l'aide de la T4 Polynucléotide Kinase (M0201L, NEB), puis hybridées deux par deux en les incubant à 95°C pendant 5 minutes, suivies d'un refroidissement lent sur un bain sec pendant 30 minutes. Pour le clonage d'un seul spacer, une réaction Golden Gate a été effectuée avec les fragments d'ADN phosphorylés en utilisant BsaI et T4 DNA Ligase. Pour le clonage de plusieurs spacers (de 4 à 8), un sous-clonage a été privilégié. Le plasmide DVA_AB, fourni par Douglas Densmore (Addgene plasmid #66039), a été modifié pour adapter les extrémités cohésives générées par Golden Gate. Cette modification a été réalisée par PCR avec les amorces oTZ259-oTZ260 et oTZ261-oTZ262, suivie d'un assemblage par Golden Gate avec l'enzyme Esp3I. Le plasmide a été séquencés par

séquençage de Sanger avec les amorces oTZ109 et oTZ110, et le plasmide résultant (pTZ014), a été utilisé pour le sous-clonage des spacers et de leurs Direct Repeats.

Les amorces hybridées ont été utilisées comme fragments d'ADN pour la réaction Golden Gate, avec BbsI-HF et T4 DNA Ligase, en utilisant pTZ014. Le produit de la réaction Golden Gate a été transformé dans des cellules chimio-compétentes d'E. coli MDS42. Les clones obtenus ont été vérifiés par PCR sur colonie, puis les plasmides extraits ont été confirmés par séquençage de Sanger avec les amorces oTZ109 et oTZ110. Les spacers ont ensuite été assemblés des pTZ015 vers le plasmide pTZ016 par une autre réaction Golden Gate utilisant BsaI. Les plasmides obtenus (pTZ017) ont été vérifiés par digestion avec BamHI, et la séquence des spacers a été confirmée une nouvelle fois par séquençage de Sanger avec les amorces oTZ003 et oTZ115.

4. Tests fonctionnels et analyse d'expression génique

4.1. Quantification de la fluorescence et traitement des données

La fonctionnalité du biosenseur modifié (pTZ012) a été évaluée en ensemençant 500 µL de milieu LB-kanamycine dans des plaques à 96 puits profonds. Ces plaques ont été incubées à 37°C pendant 16-18h à 350 RPM. Pour maintenir les cellules en phase exponentielle de croissance et assurer un état d'équilibre pour les mesures, deux sous-cultures successives ont été réalisées au cours des mesures. À chaque sous-culture, 5 µL de la culture précédente ont été transférés dans une nouvelle microplaque noire à fond transparent 96 puits contenant 195 µL de LB-Kan par puits, ainsi que les différentes concentrations de butanol à tester. La densité optique à 600 nm (DO600) et la fluorescence de la sfGFP (réglée à un gain de 40) ont été mesurées à l'aide d'un lecteur multimode (CLARIOstar® BMG LabTech). Pour chaque expérience, certains puits ont été laissés non ensemencés pour servir de blancs pour l'DO600, tandis que d'autres puits ont été inoculé avec des cultures d'E. coli possédant un plasmide n'exprimant pas de protéine fluorescente pour servir de référence d'autofluorescence. Les valeurs d'OD600 des puits vierges ont été soustraites des mesures d'OD600 des échantillons en fonction du temps. Les données d'OD600 ont ensuite été tracées en fonction du temps pour identifier les phases de

croissance. La fluorescence par cellule (sfGFP/OD600) a été calculée pour chaque point à l'aide de la formule suivante :

4.2. Mesure de la luminescence

Pour mesurer l'expression de la luciférase exprimée par *Lachnoclostridium phytofermentans* exprimant le plasmide pQnl_Pcphy23 possédant le gène *NanoLuc*, les différentes souches ont été cultivées jusqu'à la fin de la phase exponentielle. La densité cellulaire a été mesurée, puis les cellules ont été lysées en ajoutant un volume de réactif PopCulture (EMD Millipore) contenant 25 µg/mL de lysozyme, à un ratio de 1:10. Les cellules ont ensuite été incubées à température ambiante pendant 10 minutes. Le lysat cellulaire obtenu a été mélangé à un volume égal de réactif NanoLuc (Promega N1110), et la luminescence a été mesurée à 454 nm à l'aide d'un spectrophotomètre (Safas Xenius XMA). Les mesures de luminescence, réalisées en triplicat pour chaque culture, ont été normalisées en fonction de la DO_{600}. Les mesures de la luciférase avec le plasmide pQdC12a ont été effectuées en diluant une culture jusqu'en phase exponentielle tardive, suivie d'une dilution dans un milieu supplémenté en antibiotiques et 100ng/mL d'anhydrotetracycline (aTc). Cette culture a été prélever en phase exponentielle tardive pour mesurer l'expression de la luciférase.

4.3. Mesure de l'expression des ARNm par qRT-PCR

Une fois que les souches de *L. phytofermentans* portant les plasmides exprimant le système CRISPRi ont été obtenues, une culture en phase exponentielle a été diluée au 1/10ème dans 10 mL de milieu GS2 contenant du thiamphénicol, avec ou sans ajout de 100 ng/mL d'aTc. Après 6 heures, les cultures ont été récupérées pour l'extraction des ARN totaux. L'expression des ARNm de *L. phytofermentans* a été mesurée par une reverse transcription suivie d'une PCR quantitative (qRT-PCR). Brièvement, les ARN totaux ont été extraits de 10 mL de cultures en début de phase exponentielle (OD_{600} = 0,3) à l'aide du réactif TRI (Sigma 93289). Dix microgrammes d'ARN ont été traités avec 4 unités de Turbo DNase (Ambion AM2238) durant 30 minutes à 37°C. Un microgramme d'ARN a ensuite été reverse-transcrit en ADN complémentaire (ADNc) simple brin à l'aide du kit Applied Biosystems 4368814. L'amplification par PCR quantitative (qPCR) à été réalisé avec le kit

KAPA KK4621, en utilisant une MJ Research DNA Engine Opticon II. Les valeurs d'expression ont été calculées en moyenne sur trois cultures indépendantes, en utilisant la méthode C_t (cycle de seuil) et normalisées par rapport aux niveaux d'ARN ribosomique 16S.

4.4. Répression métabolique par CRISPRi

Les souches *L. phytofermentans* possédant les plasmides CRISPRi (pTZ017) ont été mise en culture en effectuant une dilution au 1/100ème de cultures en phase exponentielle dans une plaque 96 puits (TC Plate 96 Well, Suspension,F Sarstedt) pour mesurer leur croissance et quantifier les produits de fermentation de chaque puits. Les cultures ont été maintenues à 30°C avec ou sans aTc 100 ng/mL durant 96h pour assurer la consommation complète du glucose dans le milieu. La plaque est ensuite conservée à -20°C ou analyser directement par HPLC.

5. Technique de chimie analytique

5.1. Quantification des produits de fermentation par HPLC

Pour analyser les métabolites d'intérêts tels que, le glucose, le lactate, l'acétate et l'éthanol, des surnageants filtrés à 0,2 µm ont été utilisés. Les produits de fermentation ont été mesurés par Chromatographie Liquide Haute Performance (HPLC) Shimadzu Prominence LC20/SIL-20AC, équipé d'un détecteur d'indice de réfraction RID-10A et d'une colonne Aminex HPX-87H (300 mm × 7,8 mm) Biorad. La colonne a été maintenue à 55°C avec une phase mobile constituée de 5 mM H_2SO_4, à un débit de 0,5 mL/min. Le volume d'injection pour les échantillons a été de 10 µl. Des standards purifiés des produits de fermentation ont été utilisés pour définir les temps d'élutions et générer des courbes de régression linéaire. Ces courbes ont été utilisées pour quantifier les produits de fermentation dans les surnageants de culture dans les différentes expériences. Les échantillons ont été centrifugé à 4000 RPM durant 5 min et les surnageants ont été filtrés à l'aide de plaque filtrante de 0,2 µm (Cytiva, AcroPrep™). Les produits de fermentation sont collectés dans une plaque PCR jupée (FRAMESTAR, 4TITUDE), celle-ci est scellé par un pad autocollant perçable (film Zone-Free™, SEALPLATE).

5.2. *Analyse des acides gras cellulaires*

Les acides gras cellulaires ont été analysés dans des cultures de la souche WT et de la souche évoluée (PHY24.0#6), avec ou sans butanol. La biomasse cellulaire de cultures de 250 mL contenant ±0,75 % de butanol en phase exponentielle tardive à été récoltée en centrifugeant à 5min à 12000 RPM à 4°C. Les culots cellulaires ont été resuspendu dans de l'isopropanol 100%. Les acides gras cellulaires ont été caractérisés et quantifiés par le service d'identification DSMZ (Leibniz, Allemagne). En résumé, les mélanges de FAME ont été séparés à l'aide du système d'identification microbienne Sherlock (MIS) (Microbial ID, USA), utilisant un chromatographe en phase gazeuse Agilent modèle 6890N équipé d'une colonne capillaire en silicone phényl-méthyle à 5 % (0,2 mm x 25 m), d'un détecteur à ionisation de flamme et d'un échantillonneur automatique (Agilent modèle 7683A). Les pics chromatographiques ont été intégrés, et les noms ainsi que les pourcentages des acides gras ont été déterminés à l'aide du logiciel standard Sherlock MIS (Microbial ID, USA). Les plasmalogènes ont été quantifiés en tant qu'acides gras diméthyl-acétal. (DMA). Les lipides totaux ont été détectés à l'aide d'acide molybdatophosphorique, et les groupes fonctionnels spécifiques ont été identifiés à l'aide de réactifs de pulvérisation spécifiques pour des groupes fonctionnels définis.

6. Analyses génomiques et bio-informatiques

6.1. *Séquençage génomique*

Une culture de *L. phytofermentans* a été centrifugé pour récupérer les culots cellulaires puis l'ADN génomique a été extrait à l'aide du kit GenElute™ Bacterial Genomic DNA (Sigma). Les ADNg ont ensuite été quantifié à l'aide d'un fluoromètre (Qubit 4 Fluorometer Thermofisher) pour fournir à la plateforme de séquençage Plasmidsaurus 1 µg d'ADNg à la concentration de 20 ng/µL. Le service utilisé est un séquençage génomique hybride alliant Illumina et Oxford Nanopore Technologies (ONT) garantissant une précision suffisante pour identifier avec certitude des mutations SNP et des grands réarrangements par Nanopore. De plus ces deux technologies permettent un assemblage *de novo* sans la nécessité de fournir un génome de référence.

6.2. Outils bio-informatiques

L'analyse des mutations a été réalisée à l'aide de plusieurs outils bio-informatiques. La plateforme MAGE (*Microbial Genome Annotation & Analysis Platform*) développée par le LABGeM du Genoscope a été employée pour l'analyse comparative des génomes, permettant d'identifier les variants de gènes et de comprendre les voies métaboliques impliquées. En compléments, les logiciels MAUVE et BreSeq ont servi à réaliser des alignements de séquences génomiques pour visualiser les mutations SNPs ainsi que les InDels de plus grande taille. L'ensemble de ces outils ont permis de dresser la liste des mutations de type SNP et InDel.

RÉSULTATS

1. Évolution de *L. phytofermentans* pour la résistance au butanol

La fermentation de la biomasse végétale joue un rôle crucial dans le cycle du carbone des écosystèmes, en participant à l'assimilation des fibres végétales et en contribuant à l'équilibre des sols. Lorsqu'elle est exploitée à des fins industrielles, cette fermentation présente un potentiel significatif pour la production de biocarburants écologiques à faible coût, représentant ainsi une alternative prometteuse aux carburants fossiles traditionnels. Parmi ces biocarburants, le bio-butanol émerge comme un candidat de choix pour remplacer le bioéthanol, notamment en raison de ses propriétés physico-chimiques qui se rapprochent de celles de l'essence. En effet, le butanol possède un pouvoir calorifique élevé similaire à celui de l'essence, une faible pression de vapeur, est entièrement miscible avec l'essence, présente une faible hygroscopicité, est compatible avec les matériaux des moteurs, le rendant particulièrement attrayant pour une utilisation dans les moteurs à essence sans nécessiter de modifications majeures.[168].

Cependant, la viabilité économique et écologique de la production de bio-butanol dépend fortement de l'amélioration de la résistance des organismes impliqués dans sa production. Une approche innovante dans ce contexte est la production industrielle de butanol via le *Consolidated BioProcessing* (CBP) qui consiste en l'utilisation des micro-oragnismes pour la conversion directe de la biomasse en produits d'intérêt tels que les biocarburants. Cette méthode pourrait permettre une production durable et économiquement viable tout en recyclant efficacement la biomasse lignocellulosique [169]. Les stratégies envisagées incluent la production directe de butanol par des souches cellulolytiques génétiquement modifiées ou la mise en œuvre de co-cultures entre souches solvantogènes et cellulolytiques. Dans ces deux cas, il est impératif d'améliorer la résistance au butanol des micro-organismes utilisés pour optimiser le rendement de la fermentation.

Bien que la résistance au butanol ait fait l'objet de nombreuses études visant à élucider les mécanismes sous-jacents, il est désormais établi que cette résistance est un phénomène complexe et multifactoriel. La toxicité observée lors des fermentations de type

ABE est principalement attribuée à l'effet chaotropique du butanol sur les membranes lipidiques des cellules [72]. Ainsi, le développement de micro-organismes capables de tolérer des concentrations élevées de butanol est une priorité pour améliorer l'efficacité de ce processus.

Dans cette étude, nous avons porté notre attention sur l'évolution du phénotype de résistance au butanol chez L. phytofermentans, un *Clostridia* capable de fermenter la biomasse lignocellulosique en produisant principalement de l'éthanol, de l'acétate et du H_2 [38]. La production maximale de butanol par fermentation avec Clostridium acetobutylicum atteint environ 13 g/L en batch [70], notre objectif est donc d'induire une évolution adaptative de *L. phytofermentans* en présence de concentrations croissantes de butanol (variant de 0,1 à 2,5 %). Cette approche vise à générer une souche avec un phénotype de résistance accrue, à isoler un clone performant et à procéder à une caractérisation approfondie de ses capacités de fermentation.

1.1. Processus d'évolution dirigé et sélection

Pour diriger l'évolution de *L. phytofermentans* vers une résistance accrue au butanol, un GM3 [170] anaérobie a été utilisé pour effectuer des dilutions de cultures de manière automatisée. Deux modes de culture ont été employés : le mode *Medium Swap* (en gris) et le mode turbidostat (en couleur) (**Figure 27**). Le mode *Medium Swap* effectue des dilutions à fréquence fixe, en fonction de la densité optique. Une densité cellulaire élevée déclenche l'ajout d'un milieu stressant (concentration élevée de butanol), ce qui réduit la densité de la culture, entraînant ensuite l'ajout d'un milieu relaxant (faible concentration en butanol). Une fois les densités cellulaires stabilisées dans le milieu stressant, le taux de croissance à cette concentration de butanol a été amélioré en mode turbidostat.

Les concentrations de butanol ont été progressivement augmentées, en commençant par de très faibles concentrations tolérées par *L. phytofermentans* WT dans le GM3 (de 0,1 à 0,75 %). À ces concentrations, l'adaptation de la culture n'a pas nécessité l'utilisation du mode *Medium Swap*. Cependant, à partir de 1 % de butanol, il a été nécessaire d'adopter ce mode de culture pour permettre l'évolution. L'adaptation à 2,5 % a requis une adaptation plus prolongée en mode *Medium Swap*, en alternant entre le milieu

relaxant (2 %) et le milieu stressant (2,5 %). Comme le montre le graphique, au début du passage à 2,5 % au jour 210, des tests en mode turbidostat ont été réalisé, mais sans succès. Il a donc été décidé d'adapter la culture en mode *Medium Swap* durant 120 jours supplémentaires avant de la transférer en turbidostat, afin de maximiser la tolérance à 2,5 % de butanol. Une fois stabilisées dans le milieu stressant, les souches de *L. phytofermentans* ont montré un taux de croissance amélioré à la concentration la plus élevée de butanol en mode turbidostat.

La capacité des souches sélectionnées à croître à des concentrations élevées de butanol ouvre des perspectives prometteuses pour leur utilisation dans des procédés industriels, où la tolérance à des inhibiteurs tels que d'autres alcools est cruciale. Les prochaines étapes consisteront à caractériser les mécanismes de résistance développée et à évaluer leur potentiel pour la production de biocarburants en corrélant des données génomiques et phénotypiques.

Après la phase d'évolution, divers prélèvements ont été sauvegardés et des clones ont été isolés pour des analyses approfondies. D'une part, un séquençage génomique a été réalisé sur un clone sélectionné pour son phénotype de résistance au butanol et sa capacité à dégrader la cellulose. Ce séquençage a permis d'identifier les mutations survenues au cours de l'évolution.

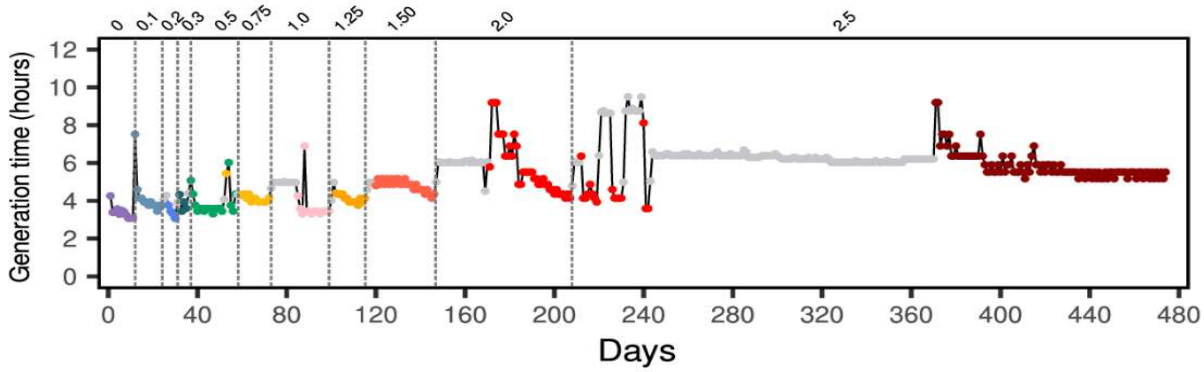


Figure 27 : Temps de génération de L. phytofermentans durant l'évolution en GM3.
Le milieu GS2 a été utilisé en ajoutant des concentrations croissantes de butanol (valeurs au-dessus du graphique allant de 0 à 2.5%). Les points gris représentent les périodes de medium swap avec des temps de génération fixes. Les points de couleurs représentent le temps de génération journalier moyen (h) durant la croissance en turbidostat.Caractérisation génotypique

1.2. Sélection des souches résistantes

Après 480 jours, un prélèvement a été réalisé à la fin de l'évolution pour effectuer des isolations clonales sur milieu GS2-Agar, afin d'obtenir une diversité de souches aux phénotypes variés. Ensuite, les clones isolés ont été mis en culture sur plaques dans un lecteur de microplaques pour observer leur temps de génération et tester leur résistance à différentes concentrations de butanol (0, 0,5 et 1 %). À ce moment, les clones ne résistant pas à 1 % de butanol ont été écartés. En parallèle, des cultures sur cellulose pure ont été effectuées pour valider la conservation des phénotypes cellulolytiques. De la même manière que pour la résistance au butanol, les clones ne dégradant plus la cellulose ont été écartés. Enfin, une visualisation par microscopie optique a permis d'obtenir un premier aperçu de la physiologie des clones.

Dans cette thèse, le clone PHY24.0#6 rassemble les différents aspects cités plus tôt. En effet, son temps de génération est plus élevé que celui de la souche WT, il résiste à 1 % de butanol, a conservé la dégradation de la cellulose et possède une physiologie peu filamenteuse, adaptée à une utilisation industrielle. C'est pour ces raisons que ce clone a été sélectionné pour effectuer les analyses approfondies détaillées dans cette thèse. Dans les parties suivantes, les résultats de l'étude concernant cette souche seront analysés tant au niveau génomique par un séquençage NGS qu'au niveau phénotypique par une caractérisation physiologique et métabolique.

L'étude du processus d'évolution de *L. phytofermentans* n'est pas détaillée dans cette thèse mais fait l'objet d'un projet en cours au laboratoire.

1.3. Caractérisation génotypique par séquençage génomique

Le génome du clone **PHY24.0#6** a été extrait et séquencé à l'aide de deux technologies complémentaires : le séquençage *long read* avec un séquenceur d'Oxford Nanopore Technologies et le séquençage *short read* avec la technologie Illumina. Cette approche hybride a permis d'affiner l'assemblage initial réalisé avec Nanopore, aboutissant à une séquence consensus de haute qualité grâce à un assemblage *de novo* (**Figure 28**). À l'échelle macroscopique, une légère différence de taille du génome a été observée, avec 4 843 703 pb pour la souche PHY24.0#6 contre 4 847 571 pb pour la souche de référence *L.*

phytofermentans ISDg^T, soit une réduction de 3868 pb. Malgré 480 jours d'évolution en culture continue, la structure générale du génome de la souche évoluée ne semble pas avoir subi de modifications majeures.

L'analyse des mutations a été réalisée à l'aide d'outils bio-informatiques, notamment la plateforme MAGE (*Microbial Genome Annotation & Analysis Platform*) développée par le LABGeM du Genoscope, ainsi que les logiciels MAUVE et BreSeq. Ces outils ont permis de dresser la liste des mutations de type SNP et InDel de plus grande taille.

L'analyse des SNP a révélé un nombre important de mutations accumulées au cours de l'évolution. Au total, 509 SNP ont été identifiés dans le génome de PHY24.0#6 en comparaison avec le génome de référence *L. phytofermentans* ISDg^T. Parmi les mutations, 126 sont synonymes, c'est-à-dire qu'elles ne modifient pas la séquence en acides aminés des protéines codées par les ORF concernés, tandis que 281 sont non-synonymes et peuvent potentiellement affecter la structure des protéines correspondantes. De plus, 21 *Single-nucleotide polymorphism* (SNP) entraînent une mutation non-sens dans 21 ORF (*Open Reading Frames*) distincts. Toutefois, cela ne signifie pas nécessairement que toutes les protéines tronquées ainsi générées sont totalement inactives. L'impact fonctionnel de ces mutations dépend de l'emplacement du codon stop prématuré dans l'ORF et de la position du domaine actif de la protéine. En outre, la proximité d'un codon start par rapport à un codon stop lié à une mutation pourrait permettre la production de fragments protéiques encore partiellement fonctionnels. L'ensemble des SNP sont trié en fonction du processus biologique prédit auxquels ils appartiennent, ceux-ci sont présenté dans le **Tableau 12** en annexe.

Quant aux SNP de type InDel, ils se caractérisent par l'insertion ou la délétion d'une base. Cela provoque un décalage du cadre de lecture des ORF, aboutissant à la production de protéines altérées ou pouvant être tronquées en raison de codons stop dans un autre cadre de lecture. Bien que les résultats de l'analyse des variants n'aient pas permis d'établir de lien direct entre la résistance au butanol et les modifications génomiques observées, il est indéniable que de nombreux événements mutateurs se sont produits. Ces mutations, qu'elles soient dues au déplacement d'éléments mobiles tels que les transposases ou à des

erreurs de réplication conservées et transmises aux générations suivantes, témoignent de l'impact de l'évolution sur le génome de la souche.

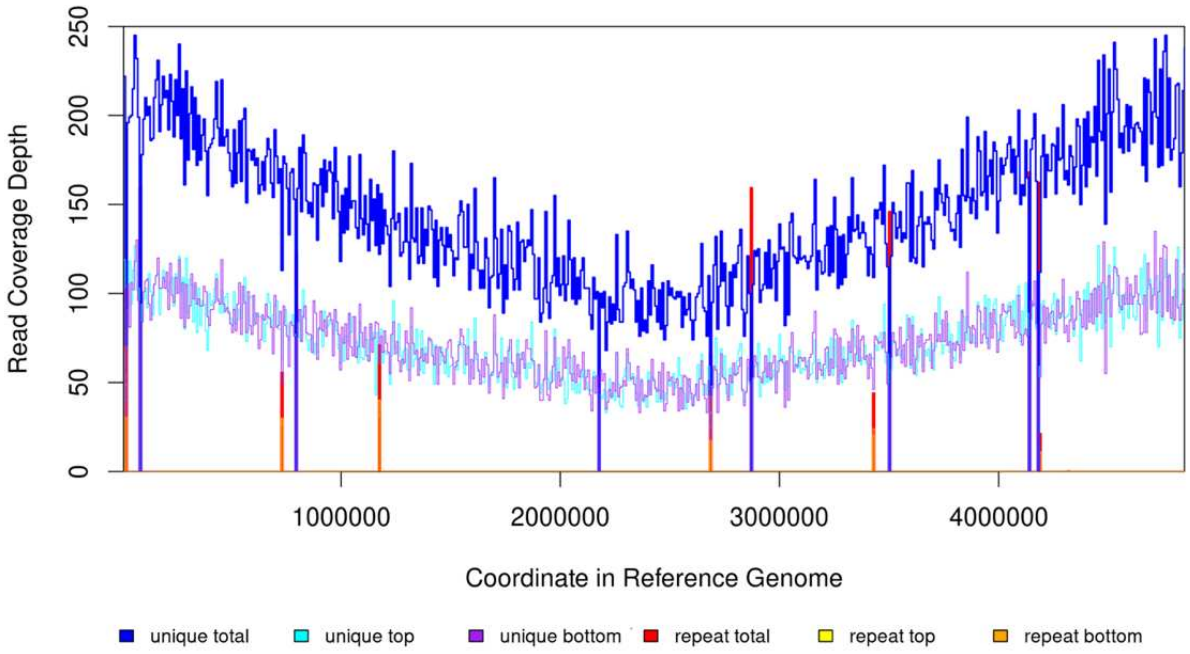


Figure 28 : Alignement génomique fait avec le logiciel BreSeq.
Alignement générer en utilisant les reads obtenus par séquençage du génome de PHY24.0#6 et en l'alignant sur le génome de L. phytofermentans ISDg. Les reads indiquent la couverture du génome et de facto l'abondance de chaque portion du génome. La courbe bleu foncé représente les séquences uniques dans le génome incluant le turquoise et le violet qui chacun représente les reads appareillés avec les brins forward et reverse respectivement. Les pics rouges, oranges et jaunes sont les séquences répétées contenues dans le génome.

1.3.1. Exemples de mutations

Dans cette partie, plusieurs mutations affectant des gènes chez *L. phytofermentans* PHY24.0#6 ont été explorées. Bien que les mutations abordées ici ne représentent qu'une infime partie des altérations génétiques observées, elles sembleraient impacter des gènes putatifs impliqués dans divers mécanismes tels que la stabilité génomique, l'expression génique et le métabolisme des sucres. Ces analyses fournissent un aperçu des mécanismes par lesquels ces mutations pourraient influencer la fitness de cette souche bactérienne. Les mutations discutées dans cette partie sont représentées par la **Figure 29**.

1.3.1.1. ***Cphy_2380* : Impact sur la réparation de l'ADN**

Le gène ***Cphy_2380*** (*mutL*) est essentiel dans la réparation des erreurs d'appariement des bases durant la réplication de l'ADN. Une mutation spécifique, caractérisée par l'insertion d'une thymine (T) à la position 1607/2088 nt, provoque un décalage du cadre de lecture, aboutissant à la division du gène en deux nouveaux ORF

potentiellement non fonctionnels. Cette mutation pourrait expliquer le grand nombre de SNP observés dans cette souche, résultant de l'accumulation d'erreurs non corrigées lors de la réplication d'ADN bactérien.

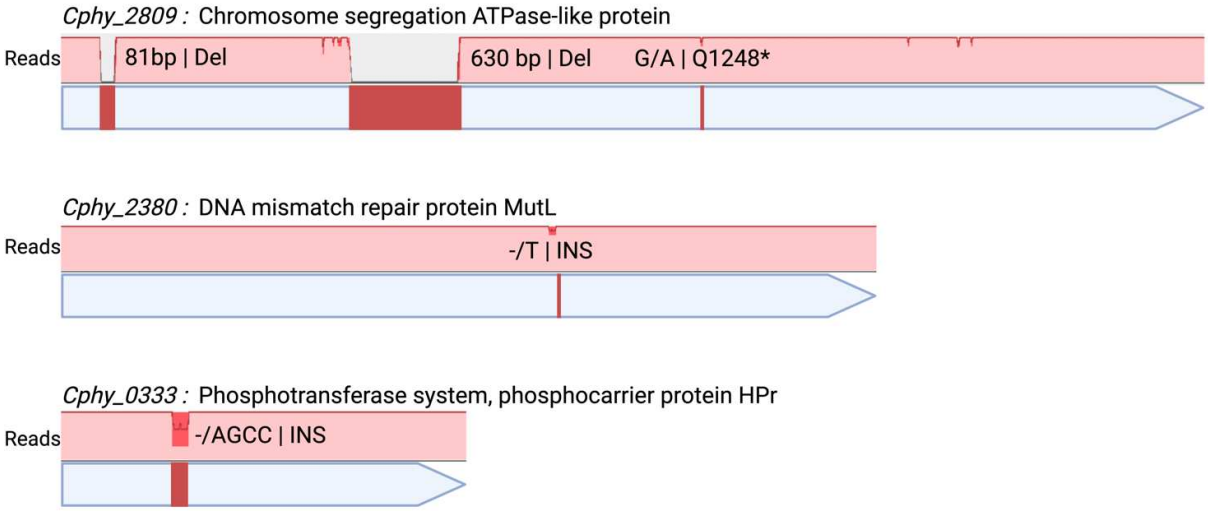


Figure 29 : Visualisation des mutations dans les gènes Cphy_2809, 2380 et 0333.
Les alignements de séquences génomiques ont été réalisées avec le génome de référence de L. phytofermentans ISDg en utilisant le logiciel MAUVE.

La fonction du gène *mutL* est bien établie dans la réparation des erreurs d'appariement de l'ADN (*Mismatch Repair*, MMR). Ce système est crucial pour corriger les erreurs de réplication. Une altération de ce gène peut entraîner une défaillance du système MMR, augmentant ainsi le taux de mutations et par conséquent, l'instabilité génomique. Chez les bactéries, la protéine mutL agit en tandem avec mutS pour détecter et réparer ces erreurs. Des mutations dans ces gènes peuvent conduire à une hypermutabilité, comme observé dans d'autres systèmes bactériens, où la perte de protéines MMR fonctionnelles peut multiplier le taux de mutation par 1 000. De plus, les endonucléases mutL forment des homodimères, et les mutations dans ce gène pourraient également perturber la dimérisation et la fonction, aggravant ainsi la déstabilisation du génome[171-173].

1.3.1.2. *Cphy_2809* : Impact sur la ségrégation chromosomique

Le gène ***Cphy_2809*** codes une ATPase essentielle dans la ségrégation chromosomique. Plusieurs mutations ont été identifiées dans ce gène, incluant deux délétions significatives et un SNP non-sens. Une première délétion de 81 paires de bases entraîne la perte de 27 acides aminés sans perturber le cadre de lecture, tandis qu'une seconde délétion plus importante de 630 pb supprime 210 acides aminés supplémentaires.

Un SNP non-sens à la position 3031 bp introduit un codon stop prématuré, tronquant la protéine à 1248 acides aminés, bien en dessous des 2228 prévus. En conséquence, la protéine mutée Cphy_2809 ne contient que 1011 acides aminés, soit une réduction de près de 55 %. Cette grande délétion engendrerait très probablement une perte de fonctionnalité, compromettant le processus de ségrégation chromosomique et, par extension, la stabilité génomique de cette souche[174].

1.3.1.3. *Cphy_0333* : Impact sur le Système Phosphotransférase (PTS) et l'Import des Sucres

Le gène ***Cphy_0333***, codant la protéine **HPr** du système phosphotransférase (PTS), a subi une insertion de 4 paires de bases (AGCC) à la position 74/258 bp. Cette insertion provoque un décalage du cadre de lecture et la formation d'un codon stop prématuré, interrompant ainsi la traduction de **HPr**, qui est essentielle au transport et à la phosphorylation des sucres dans le PTS. De plus, cette mutation entraîne l'apparition d'un nouveau codon start dans un cadre de lecture différent, initiant la traduction d'une protéine de fonction inconnue. Cela suggère la formation de deux nouvelles protéines dont les rôles restent à déterminer. La position critique de cette insertion rend peu probable la conservation de la fonction originale de **HPr**.

La protéine **HPr** joue un rôle central dans le PTS en transférant le groupe phosphate du phosphoénolpyruvate (PEP) à l'Enzyme II spécifique au sucre transporté, un processus crucial pour l'importation et la phosphorylation des sucres, nécessaires à la conservation de l'énergie [175]. Cependant, *L. phytofermentans* possède trois autres paralogues du gène ***HPr*** dans son génome, à savoir ***Cphy_1769***, ***Cphy_2300***, *et* ***Cphy_3549***. Les deux dernières sont exprimées à des niveaux similaires à ***Cphy_0333*** selon les données transcriptomiques disponibles, ce qui pourrait palier partiellement l'inactivation de ***Cphy_0333***. Néanmoins, une question fondamentale se pose quant à la nécessité du système PTS dans cette bactérie. Le système PTS consomme du PEP pour phosphoryler le glucose. Dans un environnement riche en glucose, *L. phytofermentans* pourrait avoir évolué pour optimiser ainsi l'utilisation du PEP pour d'autres processus métaboliques. Une mutation dans ***HPr*** pourrait donc être sélectionnée si des transporteurs de glucose non-PTS, comme les transporteurs ABC, sont déjà utilisé. En effet, *L. phytofermentans* possède une abondance

de transporteurs ABC, ce qui lui permet d'importer une grande variété de sucres sans nécessiter leur phosphorylation immédiate par le PTS. Cette adaptation suggère que, bien que l'inactivation de ***Cphy_0333*** puisse avoir un impact profond sur le PTS, *L. phytofermentans* a la capacité de compenser cette perte grâce à l'utilisation intensive des transporteurs ABC.

Afin de compléter l'analyse génotypique, il est essentiel d'obtenir une comparaison de l'expression des gènes. Une étude transcriptomique est en cours dans le but d'observer les différences d'expression entre la souche WT et la souche PHY24.0#6, en présence et en absence de butanol dans le milieu de culture. Cette analyse permettra d'établir des corrélations entre les mutations identifiées et les phénotypes observés.

Les répercussions fonctionnelles concrètes des mutations sur les capacités adaptatives de la souche évoluée seront détaillées dans la section suivante. Cette approche intégrée, alliant génotypage et phénotypage, offrira une perspective plus globale sur les mécanismes adaptatifs développés par la souche PHY24.0#6 au cours du processus évolutif.

1.4. Caractérisation phénotypique de la souche évoluée

Cette section présente une caractérisation phénotypique du clone PHY24.0#6 (appelé dans cette thèse PHY24), en examinant sa résistance/tolérance au butanol et à d'autres alcools. Son métabolisme a été étudié, en passant par sa capacité à dégrader la cellulose, une caractéristique intrinsèque de *L. phytofermentans*, ainsi que son profil de fermentation avec différents substrats. Une investigation sur la modification acquise de la composition chimique et structurelle de la paroi a été menée pour élucider un des mécanismes possibles de résistance au butanol.

1.4.1. Le phénotype de résistance au butanol de PHY24

Pour valider l'acquisition du phénotype de résistance au butanol suite à l'évolution dirigée, le clone PHY24 a été cultivé dans une plaque de 96 puits contenant du milieu GS2 avec des concentrations croissantes de butanol. Les souches WT et PHY24 ont été incubées dans un lecteur de plaques jusqu'à la sporulation complète, qui a été atteinte entre 24 et 36 heures selon les conditions expérimentales. L'incubation a été réalisée à 30°C, sans

agitation. Avant chaque mesure de la croissance bactérienne à une densité optique de 600 nm (OD_{600}), la plaque a été agitée pour resuspendre les cellules sédimentées. La densité optique a été mesurée à intervalles réguliers pour évaluer la croissance. Pour corriger les valeurs de croissance bactérienne, la moyenne de plusieurs puits contenant du milieu GS2 non inoculé a été soustraite en tant que blanc. Huit concentrations de butanol ont été testées : 0%, 0,5%, 0,75%, 1%, 1,5%, 2%, 2,5% v/v. Pour chaque concentration, la croissance a été évaluée dans cinq puits indépendants (n=5), et la moyenne des valeurs obtenues a été calculée pour chaque condition. Afin de quantifier la tolérance au butanol de PHY24, les valeurs de EC50 (Concentration Efficace à 50%) ont été calculées pour *L. phytofermentans* WT et PHY24. Le taux de croissance exponentielle (μ) a été déterminé pendant la phase exponentielle de croissance, permettant de calculer le temps de doublement (Td) à l'aide des formules suivante :

$$\mu = \frac{Ln(\frac{DO(t)}{DO(0)}}{t} \qquad et \qquad Tg = \frac{Ln(2)}{\mu}$$

où DO(0) est la densité optique initiale, DO(t) est la densité optique à un temps t et t est le temps écoulé.

À une concentration de 0,75% v/v de butanol, la souche WT montre un léger retard de croissance de 2 à 3 heures par rapport à la condition sans butanol, ce qui reflète une sensibilité modérée au butanol. Cependant, lorsque la concentration de butanol est augmentée à 1% v/v, la croissance de la souche WT est fortement compromise, au point qu'elle est incapable de poursuivre sa croissance. Ce résultat met en évidence la toxicité du butanol à cette concentration pour la souche WT (**Figure 30A-B**).

En revanche, la souche PHY24 présente un seuil d'inhibition au butanol nettement plus élevé. Elle parvient à croître à une densité optique comparable à celle de la souche WT même à une concentration de 2% v/v de butanol, ce qui indique une tolérance accrue. Cependant, au-delà de cette concentration, le butanol devient trop toxique pour permettre la croissance de la culture de PHY24, suggérant une limite à l'adaptation acquise. Les **Figure 30A-B** illustrent ces résultats, en se concentrant sur les concentrations de butanol les plus

pertinentes pour une meilleure clarté et lisibilité des graphiques. L'analyse du temps de génération (Tg) (**Figure 30C**) des deux souches en condition control (sans butanol) révèlent que l'évolution dirigée a non seulement amélioré la résistance au butanol de PHY24, mais a également accéléré sa vitesse de croissance. Le temps de génération moyen pour la souche WT a été mesuré à 4,185 heures, tandis que celui de PHY24 a été significativement réduit à 2,183 heures. Cette réduction indique que PHY24 croît plus rapidement que la souche WT, ce qui constitue un avantage considérable pour des applications industrielles où une croissance rapide est souhaitable.

Les courbes dose-réponse de *L. phytofermentans* WT et PHY24 **(Figure 30D**) montrent toutes deux une pente abrupte, caractéristique d'une transition rapide entre la croissance normale et l'inhibition complète en réponse au butanol. Les valeurs EC50 pour les deux souches sont respectivement de 0,7486% pour la souche WT et de 1,566% pour PHY24. Ces valeurs représentent les concentrations de butanol nécessaires pour inhiber 50% de la population bactérienne et illustrent quantitativement la résistance accrue de PHY24. Le doublement de l'EC50 chez PHY24 par rapport à la souche WT témoigne de l'amélioration significative de la résistance au butanol acquise par cette souche.

Pour confirmer que les différences observées dans le temps de génération entre la souche WT et le clone PHY24 sont significatives, un test t non apparié a été réalisé. Les résultats de ce test montrent une valeur de $p < 0,0001$, indiquant une différence hautement significative entre les deux groupes. Ce résultat statistique robuste renforce l'idée que l'évolution dirigée a non seulement conféré une résistance accrue au butanol, mais a également optimisé la croissance de la souche PHY24 en l'absence de ce stress.

Ces résultats mettent en évidence les effets bénéfiques de l'évolution dirigée sur la souche PHY24, non seulement en termes de résistance au butanol, mais aussi en améliorant sa performance de croissance générale, démontrant ainsi l'efficacité de l'évolution dirigée pour conférer une résistance accrue au butanol.

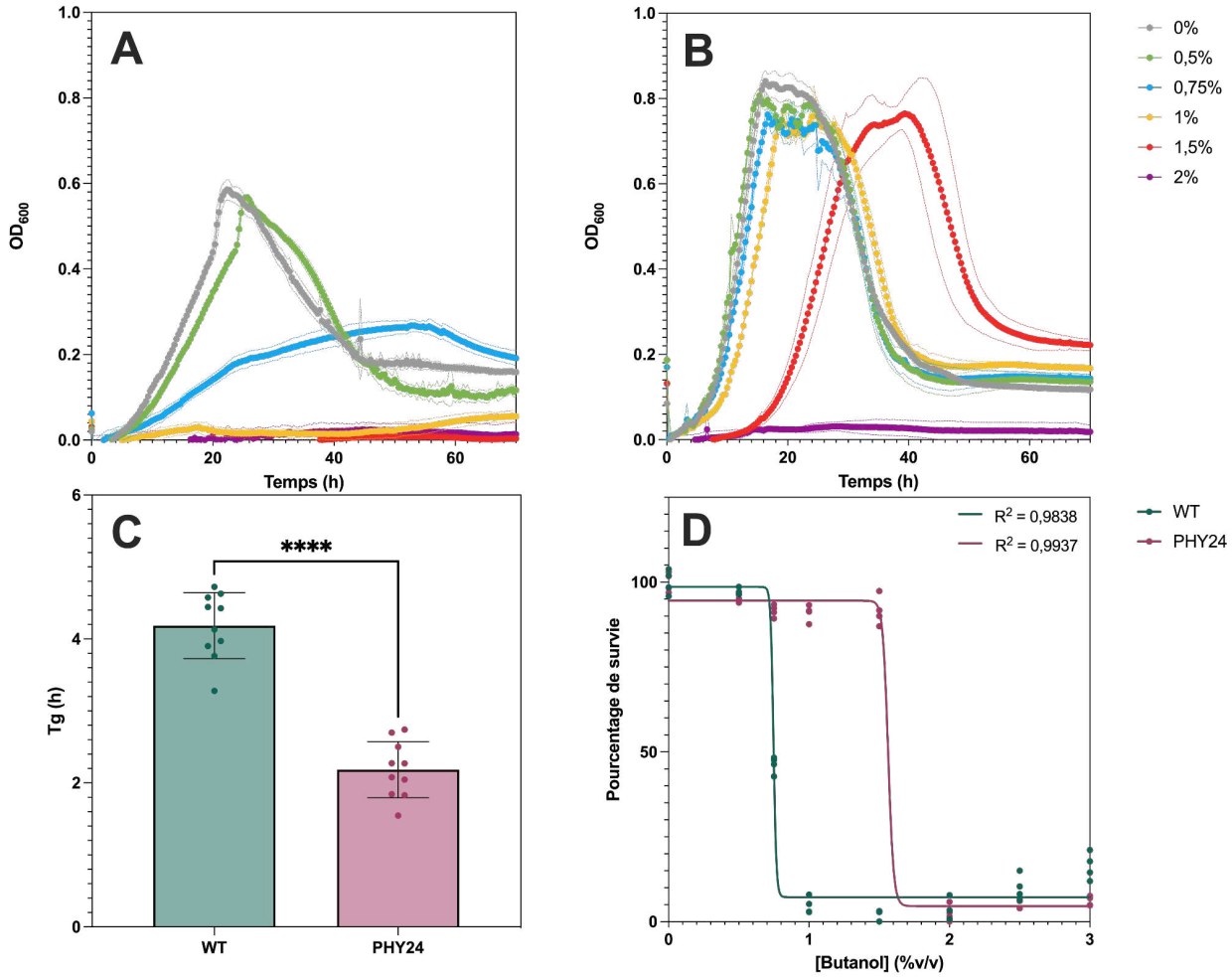


Figure 30 : Courbe de croissance de L. phytofermentans WT et PHY24.0#6 en présence de butanol.
(A) WT et (B) PHY24 en présence de différentes concentrations de butanol. n=4. (C) Comparaison des temps de génération des souches WT et PHY24. Les barres représentent les temps de génération moyens (± SEM) mesurés durant la phase exponentielle de croissance en utilisant 10 mesures de OD600. Les différences entre les groupes ont été analysées à l'aide d'un test t non apparié. p < 0.0001 (D) Courbes dose-réponse (EC50) des souches WT et PHY24 en réponse à des concentrations croissantes de butanol.

1.4.2. La tolérance acquise à d'autres inhibiteurs

Des études antérieures ont mis en évidence que l'évolution adaptative en laboratoire (ALE), visant l'acquisition de phénotypes résistants à des inhibiteurs spécifiques, peut entraîner le développement d'une résistance élargie, incluant d'autres molécules que celle initialement ciblée. Dans cette étude, la résistance croisée au butanol et à d'autres alcools primaires a été évaluée par des tests de croissance sur plaque. Des concentrations variées de quatre alcools primaires ont été testées : l'éthanol (1%, 2%, et 3% v/v), le propanol (0,5%, 1%, et 2% v/v), le pentanol (0,5%, 1%, et 2% v/v), et l'hexanol (0,5%, 1%, et 2% v/v). Ces concentrations ont été ajoutées dans trois puits distincts de plaque de 96 puits (n=3). La croissance de la souche WT et du clone évolué PHY24 a été mesurée dans les mêmes conditions que celles utilisées pour le butanol (30°C, sans agitation). La moyenne des valeurs obtenues pour chaque condition a été calculée afin de déterminer l'impact de chaque alcool sur la croissance des deux souches.

Il a été observé que, de manière similaire à la résistance au butanol, la croissance de la souche PHY24 est améliorée dans toutes les conditions en présence des autres alcools primaires (**Figure 31**). Les courbes de croissance, présentées dans les panels A à H, indiquent que la souche WT tolère l'éthanol et le propanol jusqu'à des concentrations respectives de 2% et 1% v/v. En revanche, les alcools primaires à chaîne plus longue, tels que le pentanol et l'hexanol, ne sont pas tolérés même à une concentration de 0,5% v/v. La souche PHY24, issue de l'évolution dirigée pour la résistance au butanol, a montré une résistance accrue à d'autres alcools, à l'exception de l'hexanol, qui reste non toléré. En effet, l'éthanol, le propanol et le pentanol sont tolérés jusqu'à des concentrations respectives de 3%, 2% et 0,5% v/v.

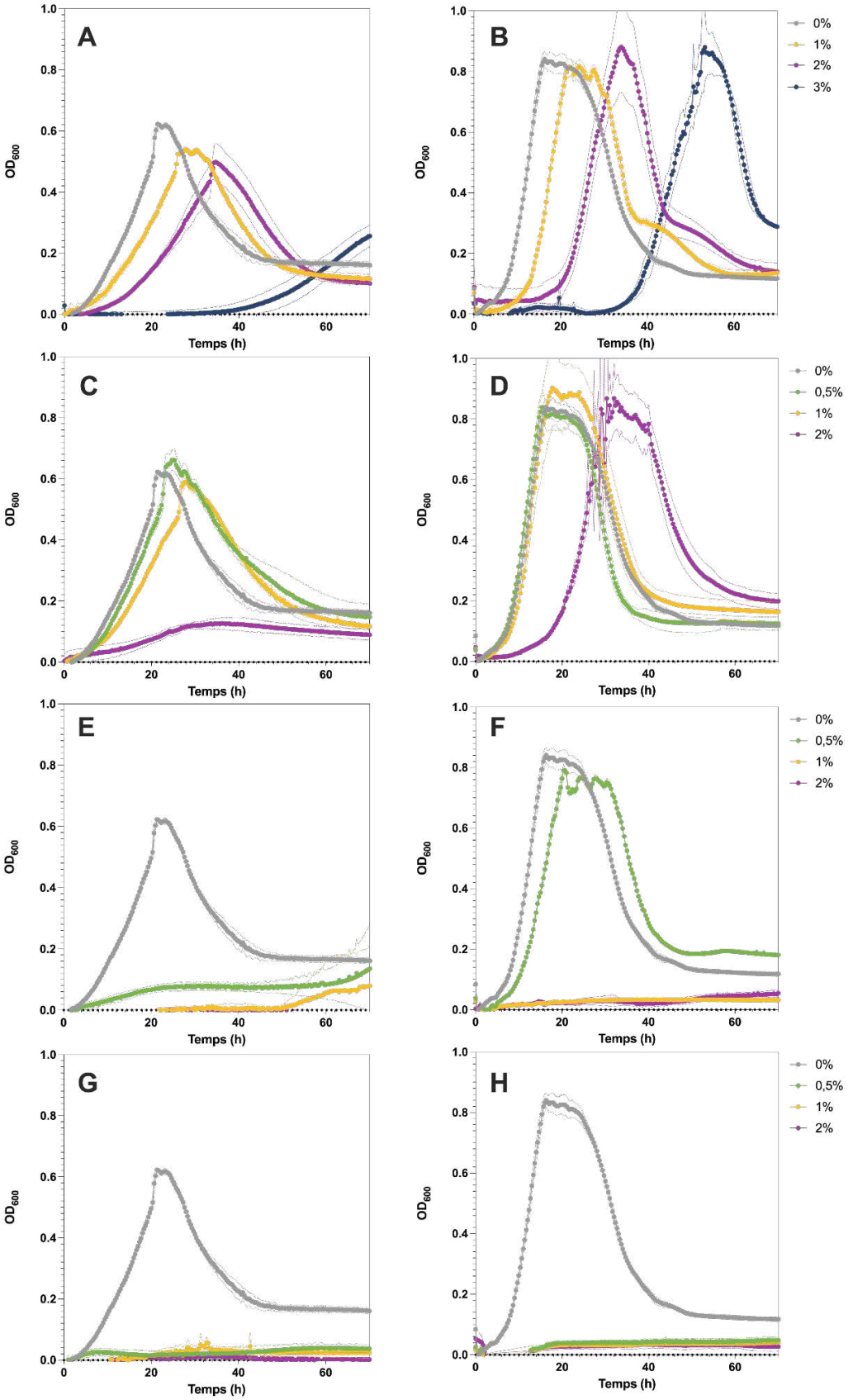


Figure 31 : Courbes de croissance de L. phytofermentans en présence d'alcools primaires.
WT (A, C, E et F) et PHY24 (B, D, F et H) en présence de différentes concentrations d'alcools primaires Éthanol (A, B), Propanol (C, D), Pentanol (E, F), et Hexanol (G, H). n=4

1.4.3. Conservation du phénotype cellulolytique

Comme exposé dans les résultats précédents, malgré le caractère dirigé de cette évolution, des phénotypes non désirés sont apparus. Cela s'est avéré bénéfique dans le cas de la résistance aux alcools primaires à chaîne courte (éthanol, propanol et butanol). Cependant, une des limitations de cette évolution est la nécessité d'une source de carbone soluble ; dans notre cas, le glucose a été utilisé. Le système GM3, composé de tubulure et utilisant la turbidité pour quantifier la croissance bactérienne afin de réaliser les dilutions, a rendu l'utilisation de cellulose impossible. Pour valider la conservation du phénotype cellulolytique, différents clones issus du GM3 ont été sélectionnés après isolation clonale en validant les deux phénotypes d'intérêt : la résistance au butanol et la dégradation de la cellulose. Le clone PHY24.0#6 a été sélectionné car il possède ces deux phénotypes.

Cette partie exposera de manière quantitative la capacité à dégrader la cellulose et le métabolisme fermentaire qui en découle, en comparaison avec la souche WT. Pour ce faire, les souches WT et PHY24 ont été mises en culture avec comme seule source de carbone de la cellulose sous forme de filtres en papier Whatman. Les cultures ont été incubées à 30°C, sans agitation, dans une jarre anaérobie. Trois réplicats biologiques indépendants ont été effectués pour chaque point de prélèvement. Pour chaque réplicat, une partie du surnageant a été analysée par HPLC pour quantifier les métabolites produits, tandis que le reste a été filtré pour récupérer la cellulose. La cellulose restante a été séché puis pesée pour quantifier sa dégradation en fonction du temps tous les 7 jours, jusqu'à 35 jours après inoculation.

En présence de cellulose comme unique source de carbone, l'évolution de la dégradation de la cellulose au fil du temps, de 7 à 35 jours, montre une différence notable entre les souches PHY24 et WT (**Figure 32A**). La souche PHY24 dégrade la cellulose à un rythme plus lent, atteignant seulement 70 % de dégradation de la cellulose initiale, contre 85 % pour la souche WT. Ce retard dans la dégradation est cohérent avec les analyses de la production des métabolites de fermentation, qui montrent une corrélation claire entre la consommation de cellulose et la production d'éthanol. En effet, la souche WT atteint son plateau de production d'éthanol après 14 jours (**Figure 32B**), tandis que PHY24 met 7 jours

de plus pour atteindre des niveaux équivalents (**Figure 32C**). De manière similaire, la production d'acides lactique et acétique est également réduite et retardée chez PHY24 par rapport à la souche WT. Ces résultats suggèrent que la souche évoluée présente un ralentissement métabolique en présence de cellulose. Ce phénomène pourrait s'expliquer par le fait que PHY24 a évolué en utilisant exclusivement du glucose comme source de carbone pendant toute la période d'évolution dirigée, ce qui n'a pas favorisé l'expression de cellulases, réduisant ainsi son efficacité à dégrader la cellulose.

Malgré cette réduction de performance en présence de cellulose, ces résultats confirment que la souche PHY24 a conservé son phénotype cellulolytique. Cela démontre que, bien que son métabolisme ait été partiellement réorienté vers l'efficacité en présence de glucose, elle n’a pas complètement perdu ses fonctionnalités cellulolytiques essentielles. Cette observation est cruciale pour son utilisation dans des applications industrielles, où la dégradation de la biomasse reste un objectif fondamental.

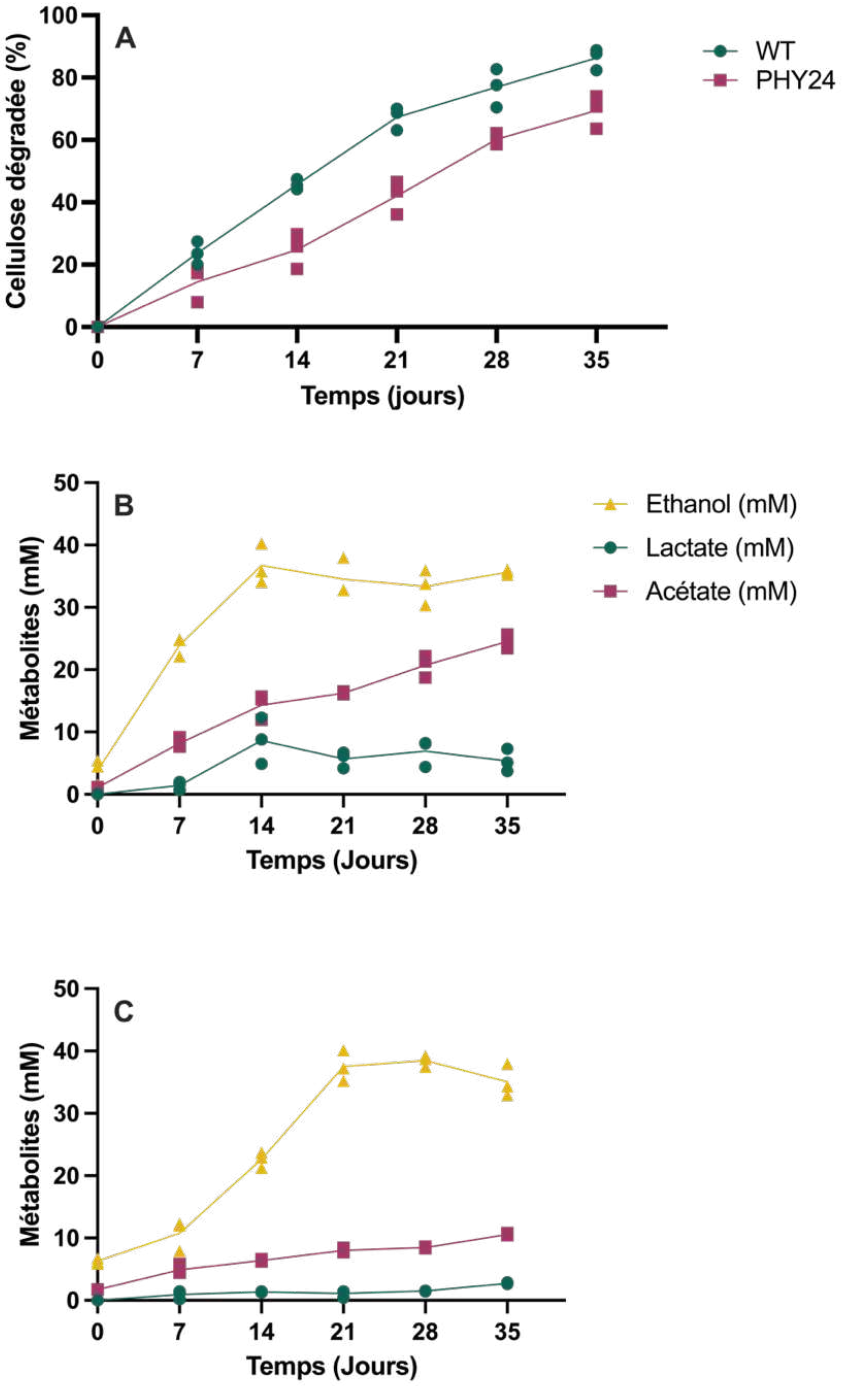


Figure 32 : Métabolisme cellulolytique de L. phytofermentans WT et PHY24.
(A) Évolution de la dégradation de la cellulose au cours du temps par L. phytofermentans WT et PHY24, chaque point de la courbe représente les 3 réplicats ayant été utilisé pour mesurer dégradation de la cellulose. (B-C) Analyse du métabolisme fermentaire de L. phytofermentans WT (B) et PHY24 (C)

1.4.4. Analyse du potentiel industriel de PHY24 en bioréacteur

L'objectif principal de cette étude est d'élucider les mécanismes sous-jacents à la résistance au butanol, tout en générant une souche capable de se développer efficacement

dans des volumes plus importants, tels que les bioréacteurs. Dans cette section, les résultats obtenus lors de la fermentation du glucose dans des bioréacteurs anaérobies de 2 L seront présentés. Les bioréacteurs, contenant du GS2-glucose à 0,5%, ont été inoculés avec 50 mL de cultures saturées des souches WT et PHY24. Les conditions de culture ont été rigoureusement contrôlées, avec un pH maintenu à 7 grâce à l'ajout de NaOH ou de KOH, et une température stable de 30°C assurée par un cryostat. Une faible agitation a été appliquée pour garantir une répartition homogène du NaOH et KOH dans le milieu de culture. Les *inoculums* ont été ajoutés en fin de journée (0h), le premier prélèvement a été effectué 14 heures plus tard, suivi de prélèvements toutes les 2 heures durant la journée, avec un dernier prélèvement 40 heures après inoculation. L'OD_{600} a été mesurée pour chaque échantillon, puis ceux-ci ont été analysés par HPLC pour quantifier la production des métabolites de fermentation (lactate, acétate et éthanol) et pour suivre la consommation de glucose

Les cinétiques de fermentation (**Figure 33**) mettent en évidence les différences entre la souche WT et la souche PHY24 dans des conditions de culture en bioréacteur de 2 L. L'analyse des valeurs de OD_{600} montre que *L. phytofermentans* WT atteint son maximum de croissance 24 heures après l'inoculation, avec une DO_{600} de 1,70, tandis que la souche PHY24 atteint un pic de 1,94 en seulement 16 heures. Cette différence de vitesse de croissance est principalement attribuée à la rapidité de consommation du glucose par les deux souches. En effet, la souche WT consomme l'intégralité des 27,775 mM de glucose en 24 heures, alors que la souche PHY24 ne présente plus de glucose détectable après seulement 18 heures. Cette consommation plus rapide de glucose par PHY24 se traduit également par une production maximale d'éthanol plus précoce. Il a été observé que la production d'éthanol est étroitement liée à la présence de glucose ; une fois le glucose totalement consommé, la production d'éthanol cesse, voire diminue. Cette diminution de la concentration en éthanol d'environ 15 mM dans les deux conditions pourrait être expliquée par une évaporation induite par le flux de N_2 utilisé pour maintenir l'anaérobiose dans le bioréacteur. Enfin, aucune différence significative n'a été observée dans la production d'acides lactique et acétique entre les deux souches.

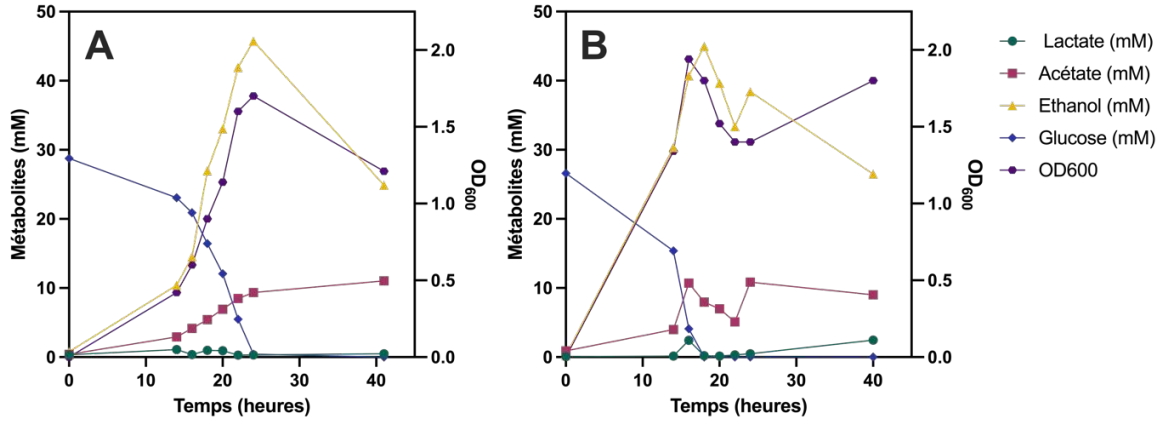


Figure 33 : Profils de fermentation de L. phytofermentans en bioréacteur.
WT (A) et PHY24 (B) avec du glucose comme source de carbone. Les cercles vert correspondent au lactate, les carrés bordeaux à l'acétate, les triangles jaunes à l'éthanol, les losanges bleu au glucose et les hexagones violets à la OD_{600} .

1.4.5. Analyse des acides gras cellulaire

L'objectif de cette expérience est d'analyser la composition en acides gras de la membrane cellulaire de la souche PHY24, afin de vérifier l'hypothèse selon laquelle la résistance de cette souche au butanol est liée à une augmentation de la proportion d'acides gras insaturés dans la membrane plasmique. La littérature scientifique suggère que les souches naturellement productrices de butanol réagissent à ce solvant en augmentant la proportion d'acides gras insaturés dans leur membrane plasmique, un mécanisme essentiel pour maintenir la fluidité et l'intégrité membranaire sous l'effet du stress induit par le butanol. Des études antérieures sur *Clostridium acetobutylicum* ont montré que l'augmentation du ratio d'acides gras insaturés/saturés contribue à stabiliser la membrane en présence de butanol, permettant ainsi aux cellules de préserver des fonctions cellulaires critiques, telles que le transport actif et la rétention de la force proton-motrice, malgré la fluidisation membranaire provoquée par ce solvant [176].

Pour déterminer si la résistance accrue de la souche PHY24 au butanol est due en partie à des modifications de sa membrane plasmique, une analyse de la composition en acides gras des membranes cellulaires a été réalisée en utilisant les services d'analyse de la DSMZ à Leibniz, en Allemagne. Des cultures de 250 mL des souches WT et PHY24 ont été préparées dans du milieu GS2-glucose à 0,5 %, avec ou sans addition de 0,75 % de butanol (concentration à laquelle la souche WT se développe). La biomasse obtenue a ensuite été

resuspendue dans de l'isopropanol à 100 %, une étape critique pour la stabilisation des lipides avant l'analyse.

L'analyse des abondances d'acides gras saturés, insaturés et hydroxylés (**Figure 34A**) a révélé que la proportion d'acides gras insaturés est significativement plus élevée chez la souche PHY24, aux dépens des acides gras saturés.

Aucune différence significative n'a été observée entre les conditions avec et sans butanol, suggérant que cette concentration de butanol n'est pas suffisante pour induire un stress, ou que la proportion d'acides gras insaturés a déjà atteint un seuil maximal. En accord avec des expériences précédentes chez d'autres *Clostridia (172)*, chez la souche WT, une augmentation de la proportion d'acides gras insaturés était attendue en présence de butanol ; cependant, une diminution de ces derniers a été observée, tandis que la proportion d'acides gras hydroxylés a augmenté en réponse au butanol. Concernant la longueur des chaînes carbonées des acides gras (**Figure 34B**), il a été observé que PHY24 présente une proportion plus élevée de C18 par rapport à la souche WT.

Les proportions des plasmalogènes ont été déterminées en quantifiant les produits de dérivatisation en diméthyl-acétal (DMA) (**Figure 34C**). Ces phospholipides particuliers jouent un rôle essentiel chez les bactéries, notamment dans la protection contre le stress oxydatif, la stabilité membranaire et potentiellement la signalisation cellulaire, grâce à leur structure unique avec une liaison éther en position sn-1, qui confère une résistance accrue à l'oxydation [177]. Aucune différence notable dans la proportion globale des plasmalogènes entre les souches WT et PHY24 n'a été observée. Cependant, l'ajout de butanol a diminué la fraction de plasmalogènes insaturés chez *L. phytofermentans* WT, sans effet similaire sur la souche PHY24. Notamment, l'évolution dirigée semble avoir inversé les ratios plasmalogènes saturés/insaturés : sans butanol, les plasmalogènes saturés représentent 32 % chez la souche WT, contre 66 % chez PHY24. Ce changement signale une réorganisation significative de la composition en phospholipides de la membrane plasmique.

L'analyse croisée des données relatives aux acides gras saturés/insaturés et à la longueur de leur chaîne carbonée (**Figure 34D**) montre que les ratios acides gras insaturés/saturés sont de 0,378 pour la souche WT en l'absence de butanol et de 0,850

pour la souche PHY24 dans les mêmes conditions. De même, les ratios C18/C16 sont de 0,267 pour la souche WT et de 0,490 pour la souche PHY24. Ces résultats suggèrent que l'évolution a induit une augmentation de la longueur des chaînes carbonées ainsi qu'une insaturation accrue vers des acides gras C18 insaturés, tout en réduisant la proportion des acides gras C16 saturés.

Une analyse plus détaillée de la composition des acides gras cellulaires est présentée dans le .

Tableau 10. En plus des observations tirées de la **Figure 34**, il est noté que l'acide gras saturé hydroxylé C18:0 3OH apparaît uniquement en présence de butanol dans les cultures des deux souches. De plus, certains acides gras, tels que C15:1 et C20:1, sont détectés exclusivement chez la souche PHY24 en présence de butanol, bien qu'ils représentent chacun moins de 1% des acides gras totaux. En outre, l'acide gras C17:1 constitue 2% des acides gras totaux chez PHY24 lorsque la culture est réalisée en présence de 0,75% de butanol.

En complément de l'étude des acides gras, une analyse de la structure de la paroi est en cours en procédant à la visualisation de celle-ci par microscopie électronique à transmission (MET).

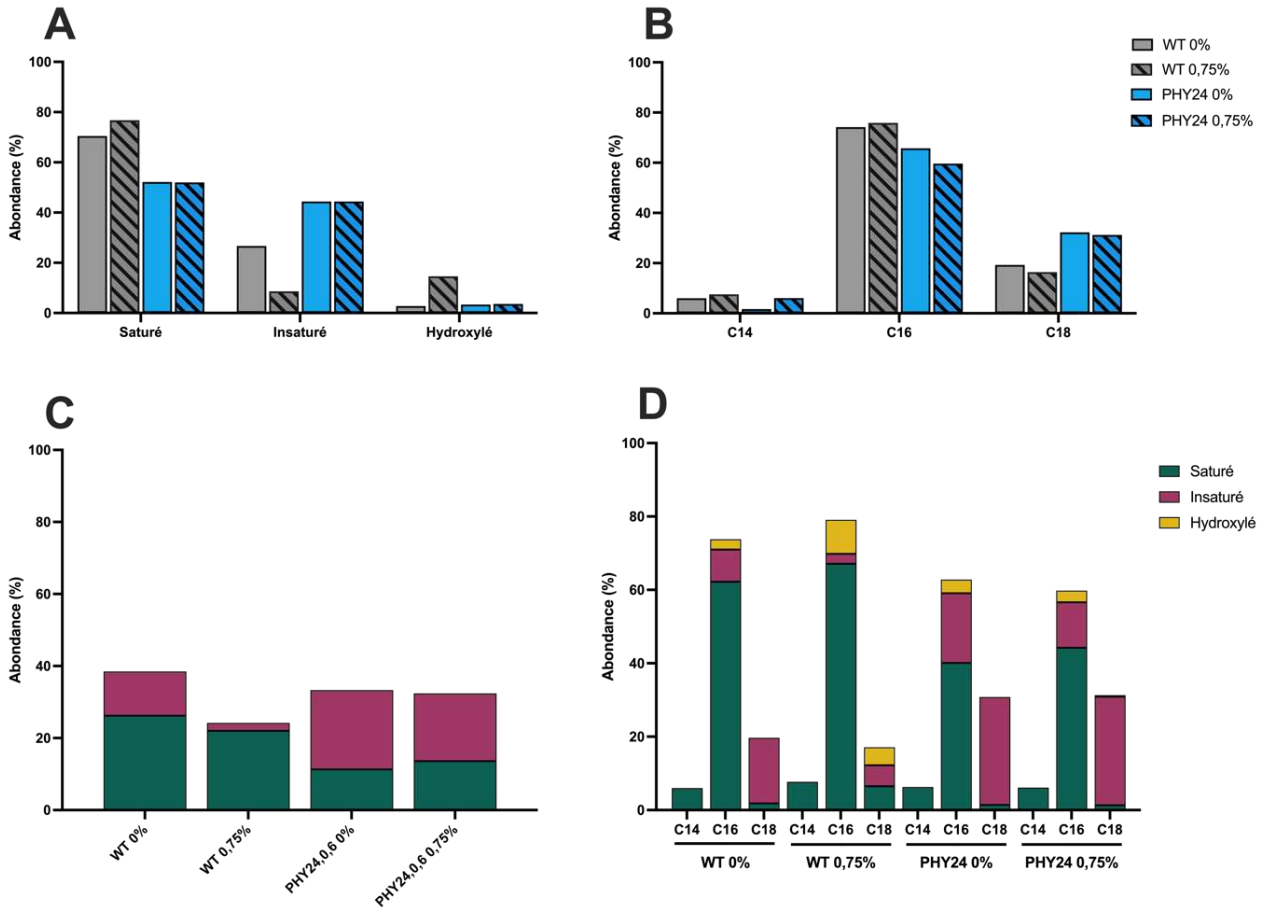


Figure 34 : Abondance relative des acides gras cellulaires.
(A) En fonction de leur état de saturation (saturé, insaturé ou hydroxylé) et de (B) la taille de leur chaine carbonée. (C) La proportion des plasmalogènes en fonction de leur état de saturation et (D) les acides gras cellulaires totaux sont représentés en fonction de la longueur de la chaine carboné et de leur état de saturation ou d'hydroxylation pour les souches WT et PHY24 avec et sans butanol 0,75%. Les plasmalogènes ont été quantifié en mesurant le pourcentage de DMA dans les échantillons.

Tableau 10 : Abondance des différents acides gras selon leur état de saturation et leur taille.

Acides gras	WT 0%	WT 0,75%	PHY24 0%	PHY24 0,75%
Saturé				
12:0	0,5	0,1	0,2	0,3
14:0	6	7,7	6,3	6,1
16:0	62,2	67,1	40,1	44,2
18:0	1,8	6,5	1,4	1,3
Hydroxylé				
16:0 3OH	2,8	9,3	3,7	3,2
18:0 3OH	ND	4,9	ND	0,4
Insaturé				
15:1	ND	ND	ND	0,3
16:1	8,8	2,7	19	12,4
17:1	ND	ND	ND	2
18:1	17,9	5,7	29,4	29,6
20:1	ND	ND	ND	0,2

ND : Non Déterminé

1.5. Discussion et Perspectives

L'évolution dirigée réalisée sur *L. phytofermentans* a conduit à l'obtention de la souche PHY24, qui se distingue de la souche sauvage par une tolérance accrue au butanol, une vitesse de croissance améliorée, une densité cellulaire maximale supérieure et une conservation des capacités cellulolytiques. Ces résultats démontrent le potentiel de l'évolution dirigée comme une méthode efficace pour optimiser des traits phénotypiques complexes chez des micro-organismes d'intérêt industriel, particulièrement ceux dont l'ingénierie est difficile. Cependant, ces résultats suscitent également de nouvelles questions et ouvrent des perspectives intéressantes pour des recherches futures visant à affiner et à approfondir notre compréhension des mécanismes sous-jacents.

L'augmentation de la tolérance au butanol observée chez PHY24 représente une avancée majeure, mais les mécanismes moléculaires exacts responsables de cette résistance demeurent partiellement élucidés. L'analyse génomique a révélé des mutations dans des gènes clés, notamment ceux impliqués dans la réparation de l'ADN et la ségrégation chromosomique, ce qui pourrait améliorer la capacité d'adaptation évolutive de la souche. D'autres expériences d'ALE chez *E. coli* ont montrer que l'inactivation du *Mismatch-Repair System* permet une augmentation de l'adaptabilité dans des conditions de cultures non favorable comme en présence d'antibiotique et de métaux lourds [178]. L'hypothèse émergente à ce projet est que via la mutation de *mutL*, *L. phytofermentans* a pu évoluer plus efficacement et ce caractère hypermutateur a été conservé car il devrait présenter un avantage environnement pour la bactérie. Il est probable que ces mutations entraîneraient l'inactivation transcriptionnelle ou traductionnelle de nombreux gènes, permettant ainsi à la cellule de concentrer ses ressources énergétiques sur des fonctions vitales pour sa survie en présence de butanol.

Pour mieux comprendre les implications de ces mutations dans la tolérance accrue au butanol, des analyses -omics en présence de butanol serait pertinentes pour une caractérisation plus approfondie des réponses au stresse induites par le butanol. A court terme des analyses transcriptomiques sont en train d'être mises place pour révéler les réseaux de régulation génique et les voies métaboliques qui sont down ou up régulées.

La capacité de PHY24 à conserver ses fonctions cellulolytiques malgré le processus d'évolution dirigée constitue un résultat encourageant, suggérant que cette fonction essentielle n'a pas été altérée. Une hypothèse plausible expliquant cette préservation réside dans l'utilisation exclusive de glucose comme source de carbone. En effet, en présence de glucose, l'expression des cellulases est inhibée, ce qui pourrait avoir évité que les gènes les codant soient affectés par des mutations au cours du processus évolutif. À l'inverse, l'utilisation du cellobiose aurait entraîner l'activation et donc plausiblement entrainé l'apparition de mutation dans des gènes codant des cellulases. Même si *L. phytofermentans* ne nécessite que d'une β-glucosidase pour métaboliser la cellobiose, les circuits de régulation activent l'expression de nombreuses cellulases. Bien que cela puisse sembler contre-intuitif, l'utilisation du glucose comme seule source de carbone durant l'évolution pourrait avoir permis de préserver les gènes cellulolytiques, évitant ainsi leur inactivation qui aurait donné un avantage sélectif du fait de l'économie d'énergie réalisé.

Cependant, il est crucial de vérifier que cette capacité se maintient dans des conditions industrielles plus complexes, où plusieurs facteurs de stress peuvent être présents simultanément. Jusqu'à présent, seule de la cellulose pure a été utilisée comme substrat. Il serait donc intéressant d'évaluer la performance de PHY24 dans des bioréacteurs utilisant divers substrats lignocellulosiques d'intérêt industriel, tels que la paille ou du son de blé. Par ailleurs, étant donné que les mutations pourraient affecter la stabilité génétique de la souche, il serait pertinent de mesurer la stabilité génétique et phénotypique de PHY24 au cours de cycles de fermentation prolongés, afin de garantir que ses performances ne se dégradent pas avec le temps dans un contexte de production continue.

Les analyses de la composition lipidique ont montré que la souche PHY24 présente une augmentation notable de la proportion d'acides gras insaturés ainsi qu'une réorganisation des plasmalogènes, ce qui suggère un rôle clé de la membrane cellulaire dans l'acquisition de la tolérance au butanol. Toutefois, la dynamique précise de ces modifications membranaires mérite une exploration plus approfondie pour mieux comprendre leur contribution à cette tolérance. Par ailleurs, d'autres éléments structuraux de la paroi cellulaire, tels que le surface layer et le peptidoglycane, qui jouent un rôle crucial dans la protection des bactéries en conditions défavorables, pourraient également avoir

subi des altérations au cours de l'évolution. Ces aspects devraient être étudiés en détail afin de mieux cerner les mécanismes sous-jacents à la résistance accrue de PHY24.

Les résultats obtenus dans cette étude sont prometteurs pour le développement de souches microbiennes robustes, capables de résister à des conditions industrielles. D'autres études récentes ont utilisé des approches d'évolution adaptative en laboratoire (ALE) pour améliorer la robustesse et la croissance chez les *Clostridia*, en employant des chemostats et des dilutions manuelles [179]. De notre côté, il est indéniable que l'utilisation de GM3 a permis une évolution dirigée de haute qualité, orientée vers des phénotypes d'intérêt spécifique. Les recherches futures basées sur ces approches pourraient aboutir à des améliorations significatives des procédés de production de biocarburants à partir de biomasse, rendant ces processus plus efficaces, durables et économiquement viables.

.

2. Synthèse *de novo* du butanol chez *L. phytofermentans*

L'ingénierie métabolique des micro-organismes pour la production durable de biocarburants représente une voie de recherche prometteuse dans le contexte des défis énergétiques et environnementaux actuels. Parmi les biocarburants, le butanol se distingue par ses propriétés physico-chimiques avantageuses, telles qu'une densité énergétique élevée et une miscibilité avec l'essence, en faisant une alternative attrayante aux carburants fossiles [72]. Cependant, la production biologique de butanol à partir de matières premières renouvelables reste un défi majeur, en raison de son besoin en biomasse dont la production entre en concurrence avec les cultures vivrières [180].

Dans ce contexte, *L. phytofermentans*, une bactérie anaérobie capable de dégrader une large gamme de polysaccharides végétaux en produits de fermentation, émerge comme un candidat prometteur pour la bioproduction de butanol à partir de biomasse lignocellulosique [38]. Pourtant, l'introduction et l'optimisation d'une voie de synthèse de butanol dans cet organisme non modèle s'avèrent complexes, en grande partie à cause de notre compréhension encore limitée de sa biologie et de la régulation de son métabolisme.

Les défis liés à la manipulation génétique de *L. phytofermentans* sont particulièrement prononcés. Cette bactérie présente une résistance naturelle aux transformations génétiques, compliquant ainsi les efforts d'ingénierie métabolique. De plus, la complexité de son métabolisme central ajoute une couche supplémentaire de difficulté, rendant nécessaire une approche méthodique pour rediriger les flux métaboliques vers la production de butanol [181].

Pour surmonter ces obstacles, une voie métabolique hétérologue pour la production de butanol a été conçue et introduite en appliquant les principes de la biologie synthétique, en particulier ceux de la modularité et du cycle DBTL (*Design-Build-Test-Learn)* [182]. Les résultats de cette étude apportent des éclaircissements sur les défis et les possibilités associés à l'ingénierie de *L. phytofermentans* en tant que châssis industriel pour la production de butanol, tout en mettant en lumière des pistes d'amélioration pour optimiser ce processus.

2.1. Construction d'une voie de combinatoire de biosynthèse du butanol

2.1.1. Choix et clonage des *parts* de la voie de synthèse

La voie de synthèse pour la production de butanol chez *L. phytofermentans* a été élaborée en sélectionnant des enzymes reconnues pour leur implication directe et efficace dans la production de butanol chez les *Clostridia*. Les enzymes sélectionnées, à savoir Ec_AtoB, Ck_Hbd1, Ca_Crt, Ca_Bcd-EtfAB, et Ca_AdhE2 (**Figure 35**), ont été choisies en fonction de leur performance rapportée dans la littérature, constituant ainsi une combinaison optimale pour la production de butanol [183].

Pour l'assemblage de cette voie, j'ai utilisé la méthode de clonage combinatoire MoClo (*Modular Cloning*). Le MoClo, une technique de biologie synthétique, permettant un assemblage précis et efficace de plasmides, facilitant ainsi la création de systèmes biologiques complexes. Cette méthode repose sur l'utilisation d'enzymes de restriction de type IIS, qui reconnaissent des séquences d'ADN asymétriques et clivent en dehors de leur site de reconnaissance, permettant un assemblage dit « *scarless* » des fragments d'ADN. Développé par l'équipe de Douglas M. Densmore sous le nom de CIDAR MoClo, ce système utilise les enzymes de restriction BsaI et BbsI [166]. Le MoClo inclut un ensemble de fragments génétiques standardisés, tels que des promoteurs, des *Ribosome-Binding Site* (RBS), des gènes rapporteurs et des terminateurs, permettant leur assemblage par *Golden Gate*. Un des avantages majeurs de cette méthode réside dans son efficacité et sa modularité, offrant la possibilité d'explorer un large éventail de configurations à un coût abordable en peu de temps.

Dans ce projet, la construction d'un opéron combinatoire a été réalisée afin de déterminer l'ordre optimal des CDS (*Coding DNA Sequences*) pour maximiser la production de butanol. Il est bien établi qu'un CDS en position initiale d'un opéron est généralement plus exprimé que celui en position terminale, en raison de sa position relative dans l'ARNm, de la dégradation progressive de l'ARNm et des structures secondaires influençant la stabilité de celui-ci. Cependant, cette observation ne s'applique pas en présence de promoteurs internes dans l'opéron [184].

L'utilisation de la méthode MoClo permet en théorie d'éviter l'insertion de mutations indésirables dans le plasmide, un problème rencontré avec la méthode de clonage Gibson, en particulier lors de l'utilisation de fragments amplifiés par PCR. Les gènes codant les enzymes de la voie de biosynthèse du butanol ont été synthétisés, en supprimant les sites de restriction BsaI et BbsI, les promoteurs internes, et en harmonisant l'usage des codons pour une traduction efficace chez *L. phytofermentans*.

Le projet a été conçu pour suivre le cycle DBTL (*Design-Build-Test-Learn*), permettant plusieurs itérations pour optimiser les paramètres génétiques ou de culture et ainsi améliorer le rendement en butanol. Lors de la phase de *Design*, une approche modulaire et standardisée a été adoptée. Chaque gène a été cloné indépendamment dans un plasmide de sous-clonage (L0-DVA-AB/BC/CD/DE/EF) avec son propre RBS, généré à l'aide de l'algorithme *RBS Calculator* de *De novo DNA* [185], fixant le *Translation Initiation Rate* (TIR) à 15,000 unités arbitraires. Cette valeur a été choisie par comparaison avec d'autres projets d'ingénierie métabolique, tels que celui de *Carbonell et al. (2018)*, où tous les RBS (Ribosome Binding Site) ont été fixés à un TIR de 15,000 [111]. En l'absence de données spécifiques sur l'effet des RBS synthétiques chez les *Clostridia*, et plus précisément chez *L. phytofermentans*, cette valeur a été conservée.

Un premier opéron a été conçu pour exprimer les gènes *Ec_AtoB*, *Ca_AdhE2*, *Ca_Crt*, et Ck_Hbd1 sous le contrôle du promoteur synthétique fort et constitutif pCons17. Un second opéron contenant les gènes *Ca_Bcd-EtfAB* sous la dépendance du promoteur pThl. Ces gènes ont été organisés dans un opéron distinct, car ces trois gènes sont exprimés conjointement dans leur contexte naturel et forment un complexe protéique hétérododécamérique.

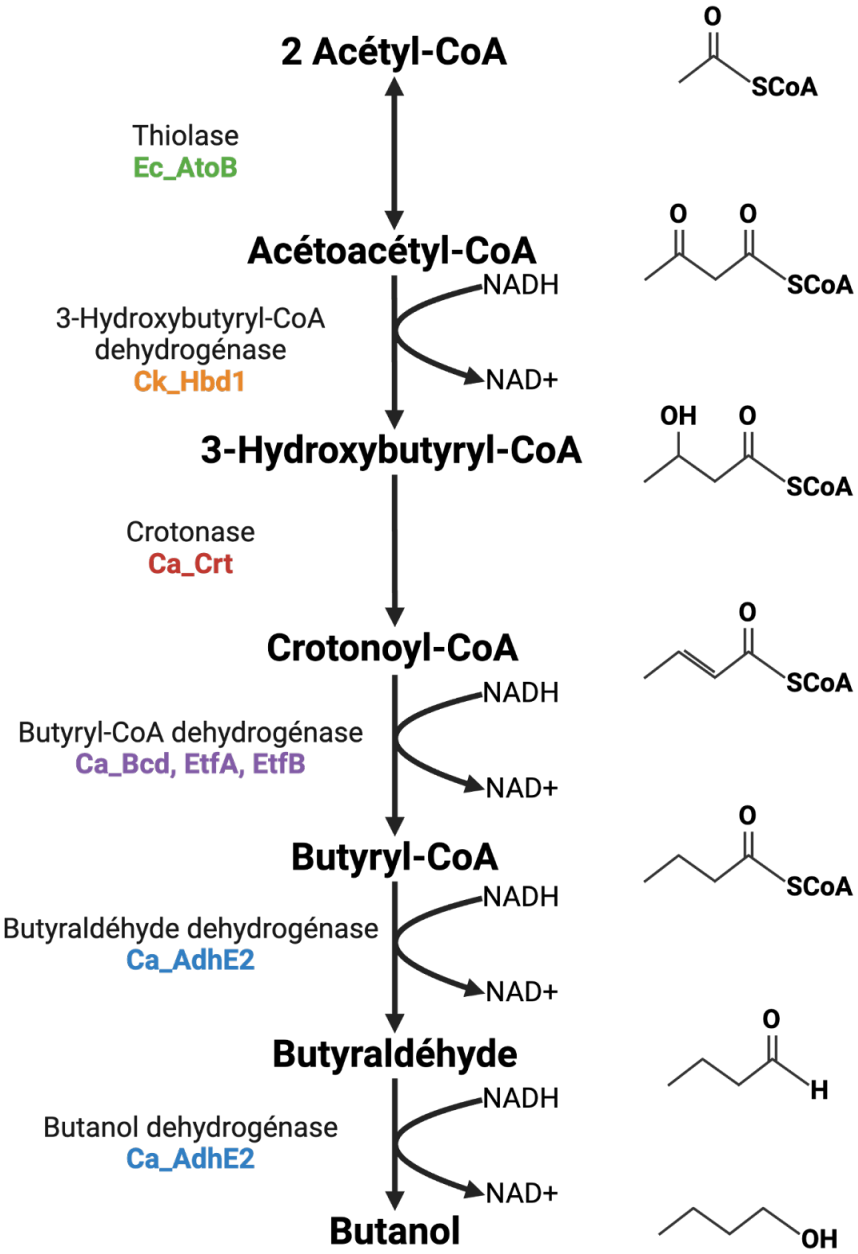


Figure 35 : Voie de synthèse optimisée pour la production de butanol.
E. coli atoB acetoacetyl-CoA thiolase/synthase, C. kluyveri hbd1 NADP+-dependent 3-hydroxybutyryl-CoA dehydrogenase, C. acetobutylicum crt crotonase, bcd-etfAB butyryl-CoA dehydrogenase et adhE2 bifunctional NAD+-dependent aldehyde and alcohol dehydrogenase

2.1.2. Assemblage des composants de la voie

La phase de construction s'est appuyée sur la technique de Golden Gate, utilisant successivement deux enzymes de restriction de type IIS. Dans un premier temps, les gènes et leurs séquences RBS ont été amplifiés par PCR, permettant ainsi l'ajout des différents sites de clivage des enzymes BbsI ainsi que des sites de liaison correspondants (**Figure 36A**). Un

premier assemblage Golden Gate a ensuite été réalisé entre les plasmides accepteurs de niveau L0, chacun comportant des sites de liaison spécifiques (AB - BC - CD - DE - EF).

Une fois les plasmides intermédiaires obtenus, ceux-ci ont été assemblés de manière modulaire dans le plasmide d'expression pQmod3_CmR_RFP_AF, contenant des sites de restriction BsaI ainsi que des sites de liaison AF flanquant la cassette RFP. L'assemblage des cinq fragments dans ce plasmide a conduit à l'excision de la cassette RFP, facilitant ainsi la sélection des clones par criblage blanc-rouge (**Figure 36B**). Les plasmides finaux, au nombre de 24, ont résulté de l'assemblage combinatoire des différents gènes de la voie de synthèse (**Figure 36C**).

Sur les 24 variants possibles, 23 ont été séquencés par la technologie Oxford Nanopore, permettant de séquencer l'intégralité des plasmides et de détecter d'éventuels dimères ou mutations dans des régions souvent non séquencées du plasmide. Parmi ces variants, 11 ont présenté une ou deux mutations non synonymes dans les gènes *Ec_AtoB* et/ou *Ca_AdhE2*, tandis que les 12 autres ont montré une séquence conforme à celle attendue sous entendant que ces deux gènes sont toxiques lors du clonage chez *E. coli*.

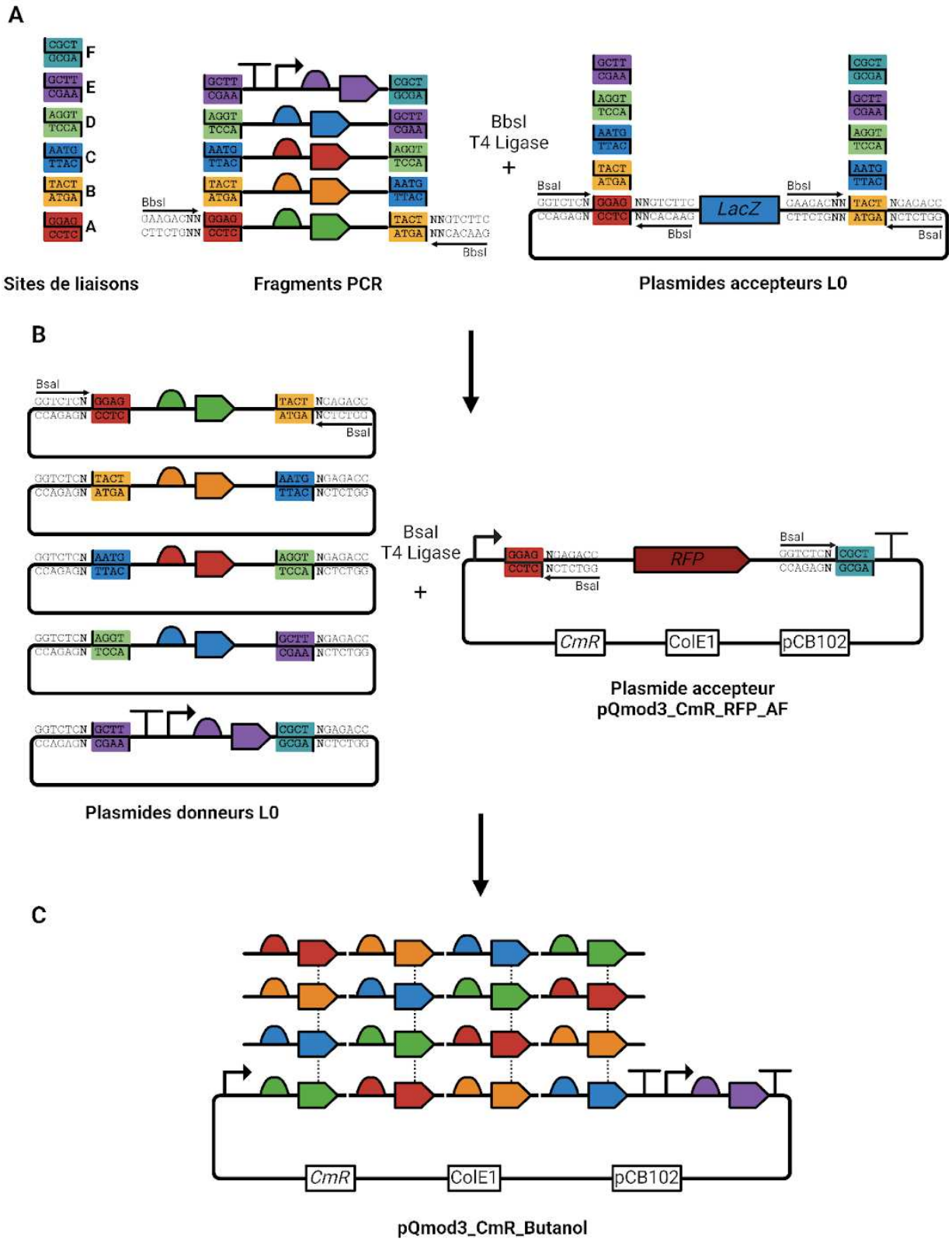


Figure 36 : Stratégie de clonage de la voie de synthèse du butanol.
(A) Les sites de liaison Golden Gate sont ajoutés par PCR aux extrémités des gènes composant la voie de synthèses. Les produits PCR sont insérés par Golden Gate dans les plasmides accepteurs L0 compatibles. (B) Chaque gène est contenu dans une des 4 versions du plasmide accepteur L0 (AB, BC, CD et DE). Dans le cas de Bcd-EtfAB (représenté en violet), il est placé en position EF. Un second Golden Gate est effectué en utilisant 5 plasmides donneurs L0 et le plasmide accepteur pQmod3_CmR_RFP_AF. (C) Pour un souci de lisibilité, les 24 possibilités des assemblages combinatoires ne sont pas toutes représentées, seules 4 combinaisons le sont.

2.2. Quantification du butanol par chimie analytique et détection in vivo

2.2.1. Expression d'une voie de synthèse du butanol chez *L. phytofermentans*

La phase de test est d'une importance cruciale dans un projet d'ingénierie métabolique, en particulier lorsqu'il s'agit de produire un composé *de novo*. Cette étape est déterminante, car l'absence de production du composé ciblé, comme le butanol dans notre cas, complique l'identification des paramètres à modifier pour induire une production ne serait-ce que minime.

Dans ce contexte, six variants distincts de la voie de synthèse ont été introduites par électroporation dans *L. phytofermentans* afin de tester leur capacité à produire du butanol et de comparer les performances de chaque construction. Les cultures correspondantes ont été réalisées en utilisant du glucose à 0,5 % comme source de carbone dans des volumes de 10 mL de GS2, ainsi que de la cellulose à 10 g/L dans des volumes de 50 mL. Ces cultures ont été incubées à 30°C pendant 7 jours pour le GS2-glucose et 35 jours le GS2- cellulose.

À la fin des périodes d'incubation, des échantillons de culture ont été prélevés pour une analyse métabolique détaillée. Les préparations d'échantillons ont consisté en une centrifugation des cultures pour éliminer la biomasse cellulaire, suivie d'une filtration du surnageant à travers des filtres de 0,2 µm avant l'analyse par HPLC.

Malgré les optimisations apportées à la voie de synthèse du butanol, les résultats ont montré qu'aucune production de butanol n'a pu être détectée dans toutes les cultures testées, que ce soit en présence de glucose ou de cellulose. Ces observations indiquent que les plasmides introduits n'ont pas suffi à initier la production de butanol dans les conditions expérimentales utilisées, soulignant ainsi la complexité du système et la nécessité d'explorer d'autres approches ou d'autres conditions expérimentales pour atteindre l'objectif de production.

2.2.2. Adaptation d'un biosenseur pour la détection *in vivo* du butanol chez *E. coli*

En complément de la quantification du butanol par HPLC, une approche alternative basée sur l'utilisation d'un biosenseur a été explorée. Ce système de détection *in vivo,*

reposant sur le régulateur transcriptionnel AlkS, a été développé pour offrir une méthode complémentaire à la quantification par chimie analytique. Le biosenseur est constitué d'une souche d'E. coli portant deux plasmides (**Figure 37**), dont l'un exprime le facteur de transcription AlkS. Ce dernier, en se complexant au butanol, se fixe sur le promoteur PalkB, recrutant ainsi l'ARN polymérase ADN-dépendante permettant l'expression du gène rapporteur *sfGFP* (superfolder Green fluorescente protein) émettant de la fluorescence à 510 nm après une excitation à 488 nm [163].

Dans cette thèse, ce système a été adapté pour être contenu sur un seul plasmide. Cette modification a été nécessaire pour permettre la co-expression des voies de synthèse du butanol préalablement construites et du plasmide biosenseur, facilitant ainsi le criblage des plasmides en utilisant un lecteur de plaques pour mesurer la fluorescence. L'adaptation du biosenseur a été motivée du fait de l'incompatibilité entre le plasmide producteur de butanol et ceux utilisés pour la détection. Les plasmides pQmod3-pCons17-butanol étant ColE1 et *CmR*, un clonage par Golden Gate a été réalisé pour reconstruire le biosenseur sur un unique plasmide. L'ORI pSC101 et le gène *alkS* sous le contrôle de PalkS ont été amplifiés par PCR à partir de *pCK01_palkS_alkS*, tandis que *PalkB_sfGFP* et *KanR* ont été extraits de *pAB_PalkB_sfGFP_KanR* (**Figure 38D**).

Après construction et validation par séquençage, le nouveau plasmide biosenseur a été testé chez E. coli dans un lecteur de plaques avec diverses concentrations de butanol dans les cultures : 10, 5, 2,5, 1,25, 0,625, 0,312, 0,156, 0,078, 0,039, 0,019, et 0 mM, dans du milieu LB à 37°C pendant 16 heures. Les résultats ont montré que les différentes concentrations de butanol n'affectent pas la croissance de la souche biosenseur (**Figure 38A**). Par ailleurs, le biosenseur a réagi au butanol en fonction du temps, dû à l'expression de sfGFP (**Figure 38B**), et la courbe de corrélation entre la fluorescence et la concentration en butanol suit un modèle sigmoïde (**Figure 38C**). Cette tendance suggère l'existence d'un point d'inflexion, indiquant un seuil de détection et un seuil de saturation du système. Le seuil de détection du butanol a été déterminé à environ 0,625 mM, où la fluorescence mesurée était d'environ 0,78 a.u., contre 0,44 a.u. sans ajout de butanol. Le seuil de saturation n'a pas pu être déterminé, car la toxicité du butanol pour *E. coli* a réduit la

croissance au-delà de 20 mM, indiquant que la plage optimale de détection du butanol par ce biosenseur se situe entre 0,625 mM et 20 mM.

Contrairement au biosenseur, la quantification du butanol par HPLC a montré une relation linéaire (**Figure 38E**), offrant ainsi une meilleure précision à la fois à faible et haute concentration.

Des analyses statistiques ont confirmé une corrélation significative entre la concentration de butanol dans le milieu de culture et la fluorescence émise. L'EC50 de la courbe obtenue est de 7,117 avec un coefficient de Hill de 1,398, des résultats comparables à ceux obtenus avec le système à deux plasmides décrit par Bahls et al. (2022) [163].

Pour évaluer la capacité du biosenseur à détecter la production de butanol, *E. coli* a été co-transformé avec les plasmides biosenseur et de production de butanol, permettant ainsi la production et la détection dans une seule expérience. Malheureusement, malgré les tests réalisés dans différentes conditions de culture propices à la production de butanol, notamment en milieu LB à 37°C et en TB supplémenté en glucose à 30°C, aucune des souches testées n'a montré une fluorescence significativement supérieure à celle du contrôle sans voie de synthèse. Ces résultats suggèrent que les conditions expérimentales utilisées n'étaient pas adéquates pour permettre la détection et/ou la production de butanol à des niveaux détectables par le biosenseur

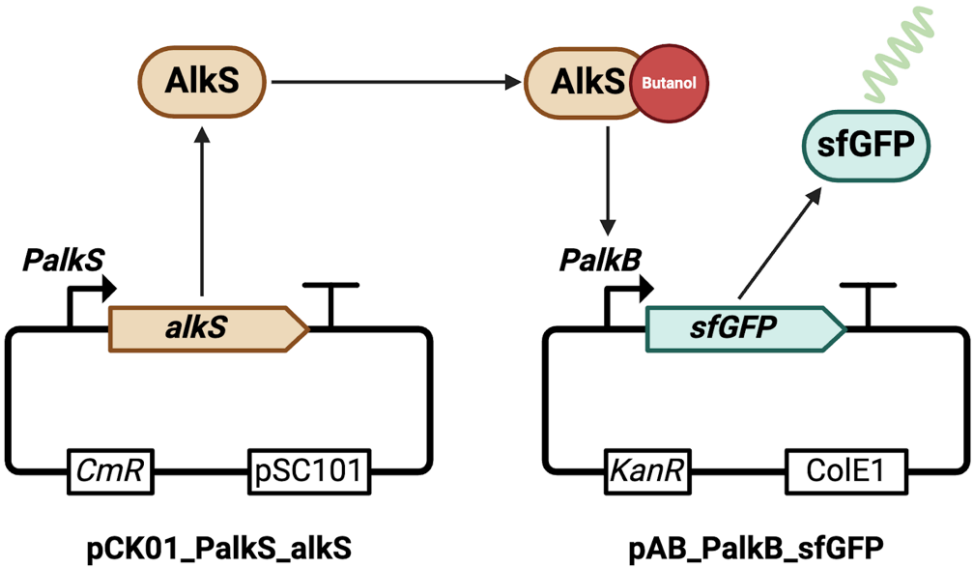


Figure 37 : Biosenseur original séparé sur deux plasmides.
Ce biosenseur crée et publié dans Bahls et al. 2022. Addgene (Plasmid #166502 & #166503) est constitué de 2 plasmides, PCK01_PalkS_alkS qui produit de manière constitutive le facteur de transcription AlkS, en présence de butanol celui-ci est capable d'activé le promoteur PalkB pour permettre l'expression de sfGFP.

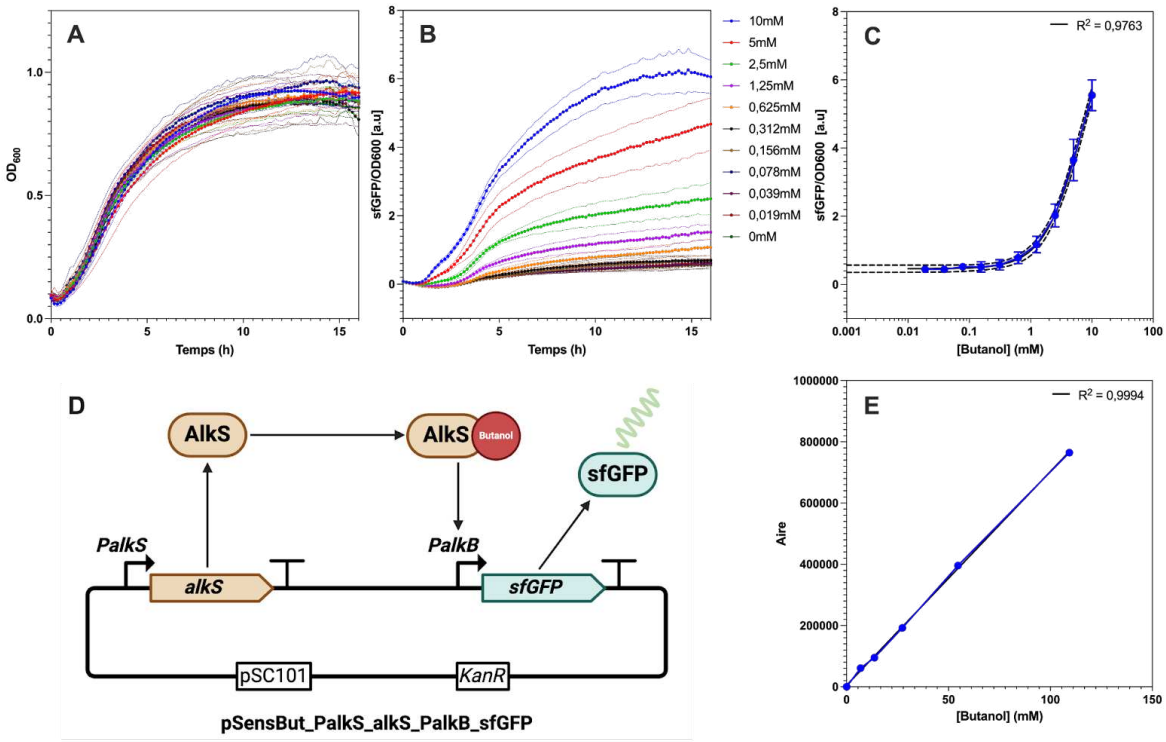


Figure 38 : Adaptation du biosenseur sur un seul plasmide.
(A) courbe de la souche E. coli EcTZ344 avec les différentes concentrations en butanol, le blanc a été soustrait. (B) Évolution de la fluorescence au cours du temps chez EcTZ344. Les courbes représentent la moyenne de 5 puits et les SD sont montrés en pointillés. (C) Courbe dose-réponse de la souche EcTZ344 en présence de butanol en effectuant une régression non-linéaire (D) Représentation schématique du nouveau plasmide contenant alkS et la sfGFP sous la dépendance du promoteur PalkB. (EC50 = 7,117 et coefficient de Hill = 1,398). (E) Courbe de calibration par quantification HPLC en effectuant une régression linéaire (R2 = 0,9994)

2.3. Discussion et Perspectives

L'intégration d'une voie de synthèse pour la production de butanol chez *L. phytofermentans* constitue un défi majeur, en grande partie en raison de la complexité intrinsèque de la biologie de cet organisme, dont la génétique n'est pas encore totalement maîtrisée. Bien que des approches méthodiques aient été déployées pour concevoir, assembler et tester diverses variantes de la voie de synthèse, aucune production de butanol n'a été détectée, que ce soit à partir de glucose ou de cellulose. Ces résultats posent des questions cruciales quant à la faisabilité de l'ingénierie métabolique dans des organismes non conventionnels, où les mécanismes de régulation de l'expression génique sont encore mal compris et difficilement contrôlables.

L'absence de production de butanol observée dans les cultures pourrait s'expliquer par plusieurs facteurs. D'une part, l'expression hétérologue des enzymes impliquées dans la voie de synthèse pourrait ne pas avoir atteint des niveaux nécessaires pour permettre une conversion efficace des précurseurs en butanol. D'autre part, l'instabilité potentielle des ARN messagers et la formation de structures secondaires complexes pourraient avoir conduit à une expression sous-optimale des gènes impliqués dans la voie métabolique. Pour élucider ces questions, une des perspectives à court terme de ce projet serait de valider l'expression des différents gènes par RT-qPCR, afin de déterminer si l'absence de production de butanol chez *L. phytofermentans* est liée à une expression transcriptionnelle insuffisante.

De plus, pour évaluer la fonctionnalité des enzymes impliquées dans cette voie métabolique, nous prévoyons de les produire en utilisant une méthode in vitro *cell-free*. Cette approche permettra la production et la détection du butanol tout en évitant la toxicité des intermédiaires de cette voie.

Il est également important de noter que bien que *L. phytofermentans* ne produise pas naturellement d'intermédiaires de la voie du butanol, certaines de ses enzymes sont prédites comme étant capables de catalyser certaines réactions de cette voie. Par exemple, l'alignement de la séquence protéique de l'enzyme Cphy_1286 avec des bases de données protéiques suggère que cette enzyme pourrait avoir une activité Trans-2-enoyl-CoA

réductase (NAD(+)), catalysant ainsi la conversion du crotonyl-CoA en butyryl-CoA. De plus, des investigations dans notre équipe ciblant les alcool déshydrogénases, en particulier Cphy_3925, une enzyme bifonctionnelle (acétaldéhyde déshydrogénase/alcool déshydrogénase), ont révélé sa capacité in vitro à convertir le butyryl-CoA en butyraldéhyde, puis en butanol. Ces résultats suggèrent que les alcools déshydrogénase de *L. phytofermentans* possèdent une promiscuité enzymatique qui pourrait être exploitée pour optimiser la production de butanol. Une analyse plus approfondie du métabolisme de cet organisme pourrait permettre de déterminer la voie de synthèse optimale pour la production de butanol.

En outre, la complexité de la régulation métabolique chez *L. phytofermentans*, ainsi que les interactions entre les voies métaboliques natives et celles introduites, pourraient avoir compromis la production de butanol. Par exemple, l'équilibre redox, qui est très conservé chez les *Clostridia*, constitue un facteur limitant. En effet, la production concurrente d'éthanol et de lactate, servant à consommer les cofacteurs oxydés tels que le NADH et le NADPH, entre directement en compétition avec la nouvelle voie de synthèse du butanol, qui requiert quatre molécules de NADH par molécule de butanol produite.

Pour orienter les flux métaboliques vers la voie de biosynthèse du butanol, plusieurs stratégies méritent d'être explorées. Premièrement, la réduction ciblée du génome de *L. phytofermentans* en éliminant des voies métaboliques énergivores non essentielles constitue une approche prometteuse. Une souche de *L. phytofermentans* présentant une large délétion d'un îlot phagique a été générée dans une publication précédente [186]. Cette délétion de 38,9 kb dans le génome de *L. phytofermentans* démontre que des modifications génomiques de grande ampleur sont réalisables chez cette souche, ouvrant ainsi des perspectives pour accroître la disponibilité des précurseurs et cofacteurs nécessaires à la synthèse du butanol. L'expression de la voie du butanol dans un génome simplifié pourrait donc favoriser une meilleure intégration métabolique et une production accrue.

En outre, il est pertinent d'envisager des approches alternatives pour la production de butanol. L'une des plus prometteuses pourrait être l'utilisation de cocultures, où *L. phytofermentans* serait associé à un autre micro-organisme naturellement capable de

produire du butanol ou d'autres produits d'intérêt. Les cocultures permettent de tirer parti des capacités métaboliques complémentaires de différents organismes, réduisant ainsi la nécessité de modifications génétiques complexes. Par exemple, nous explorons actuellement la coculture de *C. acetobutylicum*, un *Clostridium* reconnu pour sa capacité à produire du butanol, avec *L. phytofermentans*, afin de tirer parti de sa capacité à dégrader la biomasse végétale. Cette approche pourrait non seulement simplifier le processus de production, mais aussi améliorer l'efficacité globale du système.

En conclusion, bien que les efforts réalisés pour introduire et tester une voie de synthèse du butanol chez *L. phytofermentans* n'aient pas encore permis de détecter une production mesurable, ces résultats offrent des informations précieuses sur les limites actuelles de l'ingénierie métabolique dans des organismes non conventionnels. L'exploration d'alternatives naturelles, telles que les cocultures, pourrait fournir une solution viable pour surmonter les défis posés par la complexité génétique et métabolique de ces systèmes biologiques. Les recherches futures devraient se concentrer sur l'optimisation de ces approches, tout en approfondissant notre compréhension de la biologie de *L. phytofermentans*, pour éventuellement parvenir à une production efficace de butanol ou d'autres biocarburants dans ce type d'organismes. Par ailleurs, des approches innovantes basées sur des modifications transitoires du métabolisme, telles que l'utilisation de système CRISPRi, sera également explorées pour améliorer notre compréhension du métabolisme de *L. phytofermentans*, dans l'objectif de générer des souches optimisées et compatibles avec les exigences industrielles.

s.

3. Régulation génique et prototypage métabolique par CRISPRi

Bien que de nombreux outils génétiques permettant de moduler l'expression des gènes soient disponibles pour certains modèles bien étudiés, tels que *Escherichia coli*, ces technologies font cruellement défaut pour une grande variété d'autres taxons bactériens ayant des applications biotechnologiques, les Clostridia en font partie. La bioproduction, notamment dans le cadre de l'ingénierie métabolique pour la production *de novo* de composés d'intérêt, requiert une compréhension approfondie de la génétique et du métabolisme du micro-organisme hôte. Pour maximiser le potentiel métabolique de *L. phytofermentans*, il est crucial de rediriger les flux métaboliques vers les composés cibles. Des techniques de modification génomiques telles que FLP, Phage Lambda Red [187] et CRISPR-Cas [188] sont utilisées pour induire des recombinaisons homologues dans le génome de l'hôte. Cependant, chez les *Clostridia*, ces recombinaisons homologues montrent une efficacité très limitée [132]. Par conséquent, des approches alternatives, telles que le Targetron, basé sur l'insertion site-spécifique d'introns mobiles du groupe II, bien que disponibles, se révèlent souvent fastidieuses et sont associées à un taux élevé d'erreurs *off-target* [41].

Afin d'atteindre le rendement suffisant du composé d'intérêt, il est nécessaire de tester diverses combinaisons de délétions génétiques, une démarche qui requiert d'effectuer successivement de nombreuses inactivations de gènes. Ce processus, bien que méthodiquement rigoureux, est extrêmement chronophage et expose à des risques d'impasses, notamment dans la sélection des gènes à inactiver.

Contrairement aux méthodes classiques d'ingénierie génomique, la technique de *CRISPR interference* (CRISPRi) repose sur l'utilisation d'effecteurs CRISPR/Cas dérivés de Cas9 (dCas9) ou Cas12a (dCas12a) dont l'activité nucléase a été neutralisée. Ces effecteurs permettent de réprimer l'expression de gènes cibles sans nécessiter la modification préalable des promoteurs ou la suppression des gènes cibles eux-mêmes [189].

Pour réaliser une répression transcriptionnelle ciblée chez *L. phytofermentans*, nous avons choisi la protéine Cas12a du *Lachnospiraceae bacterium ND2006* (LbCas12a) pour

plusieurs raisons. Premièrement, comme elle provient d'un membre de la famille des Lachnospiraceae, il est plus probable qu'elle puisse être exprimée avec succès chez *L. phytofermentans*. Deuxièmement, son choix est lié à la séquence PAM qu'elle utilise pour reconnaître l'ADN cible. En effet, l'analyse de la distribution des sites PAM de LbCas12a (5′-TTTV-3′) [190] dans le génome de *L. phytofermentans* a révélé que, bien que ces sites soient globalement moins fréquents, ils sont plus abondants dans les régions promotrices fortement exprimées. Cela rend LbCas12a particulièrement attrayante pour cibler ces régions via CRISPRi. Nous développons actuellement un script pour identifier les régions cibles les plus efficaces afin de maximiser la répression des gènes visés par CRISPRi

LbCas12a présente donc un potentiel prometteur pour la répression transcriptionnelle chez *L. phytofermentans*. Cette partie de notre étude exposera la mise en place du système ainsi que sa preuve de concept, suivie des améliorations qui lui ont été apportées. Enfin, l'utilisation de CRISPRi sera discutée en tant qu'outil puissant pour l'étude et l'ingénierie du métabolisme. Cette approche pourrait permettre un prototypage rapide et efficace, en simulant différentes combinaisons de gènes à inactiver pour une souche industrielle future, avec une répression de l'expression pouvant atteindre 99 % [150]. En somme, cette méthode offre une voie prometteuse pour accélérer le développement de souches industrielles et approfondir la compréhension des mécanismes cellulaires encore inexplorés.

3.1. Preuve de concept du système CRISPRi chez L. phytofermentans

Bien que LbCas12a soit considérée comme moins toxique que la Cas9 pour les *Clostridia* [146], elle reste particulièrement toxique lorsqu'elle est exprimée chez *E. coli*, bactérie dont nous avons besoin lors des étapes de clonage. Pour remédier à cela, nous avons cherché à atténuer cette toxicité potentielle en développant un plasmide CRISPRi basé sur une dLbCas12a régulée par le système TetR, nommé pQdC12a. Pour cela, le plasmide pQmod3C a été modifié afin d'y inclure plusieurs éléments : une cassette miniPthl-TetR, une version inductible de dLbCas12a sous le contrôle de l'aTc, ainsi qu'un ARNg dont la séquence cible peut être facilement changé par clonage Golden Gate en utilisant l'enzyme BbsI (**Figure 39A**).

Pour maximiser l'efficacité de dLbCas12a, la structure de l'ARNg prend en compte l'impact des éléments terminaux sur l'activité de LbCas12a. En effet, l'insertion d'un terminateur Rho-indépendant immédiatement après le guide a été précédemment rapportée comme pouvant nuire à l'activité de la LbCas12a [191]. Pour contourner ce problème, les 25 premières paires de bases (pb) de la répétition terminale du réseau LbCas12a de type sauvage ont été insérés en aval du guide, séparant ainsi le guide du terminateur *fdx* par une distance de 55 pb, incluant la cicatrice de clonage.

Afin de tester l'efficacité de la répression génique médiée par dLbCas12a chez *L. phytofermentans* nous avons utilisé le plasmide pQdC12a pour réprimer le gène rapporteur NanoLuc en ciblant la boîte -10 du promoteur Pcphy23 (**Figure 39A**). L'expression de ce gène a été mesurée en ajoutant un second plasmide pQnl_Pcphy23 en parallèle de pQdC12a. Nous avons observé une répression efficace de l'expression de NanoLuc, confirmant ainsi la fonctionnalité et l'efficacité du système CRISPRi basé sur dLbCas12a dans ce contexte.

En présence d'aTc, une répression significative de l'expression de *NanoLuc* a été obtenue en utilisant le guide ciblant la boîte -10 de Pcphy23 (répression par un facteur de 5, $p < 0,0001$) par rapport à un plasmide pQdC12a ne possédant pas de guide spécifique (**Figure 39B**). Une répression moindre de l'expression de NanoLuc a également été observée en l'absence d'aTc, en utilisant le guide g-nl1 (répression par un facteur de 3,5, $p = 0,0005$), probablement en raison d'une fuite d'expression de dLbCas12a par TetR. Cependant, la fuite du promoteur PGusA2-TetO2/1 sans l'ajout d'aTc démontre la robustesse du système CRISPRi. En effet, cela signifie qu'une faible quantité de dLbCas12a permet une forte répression de l'expression de Nanoluc. Les résultats obtenus après induction confirment que le CRISPRi médié par dLbCas12a réprime efficacement l'expression du gène rapporteur in vivo chez *L. phytofermentans*.

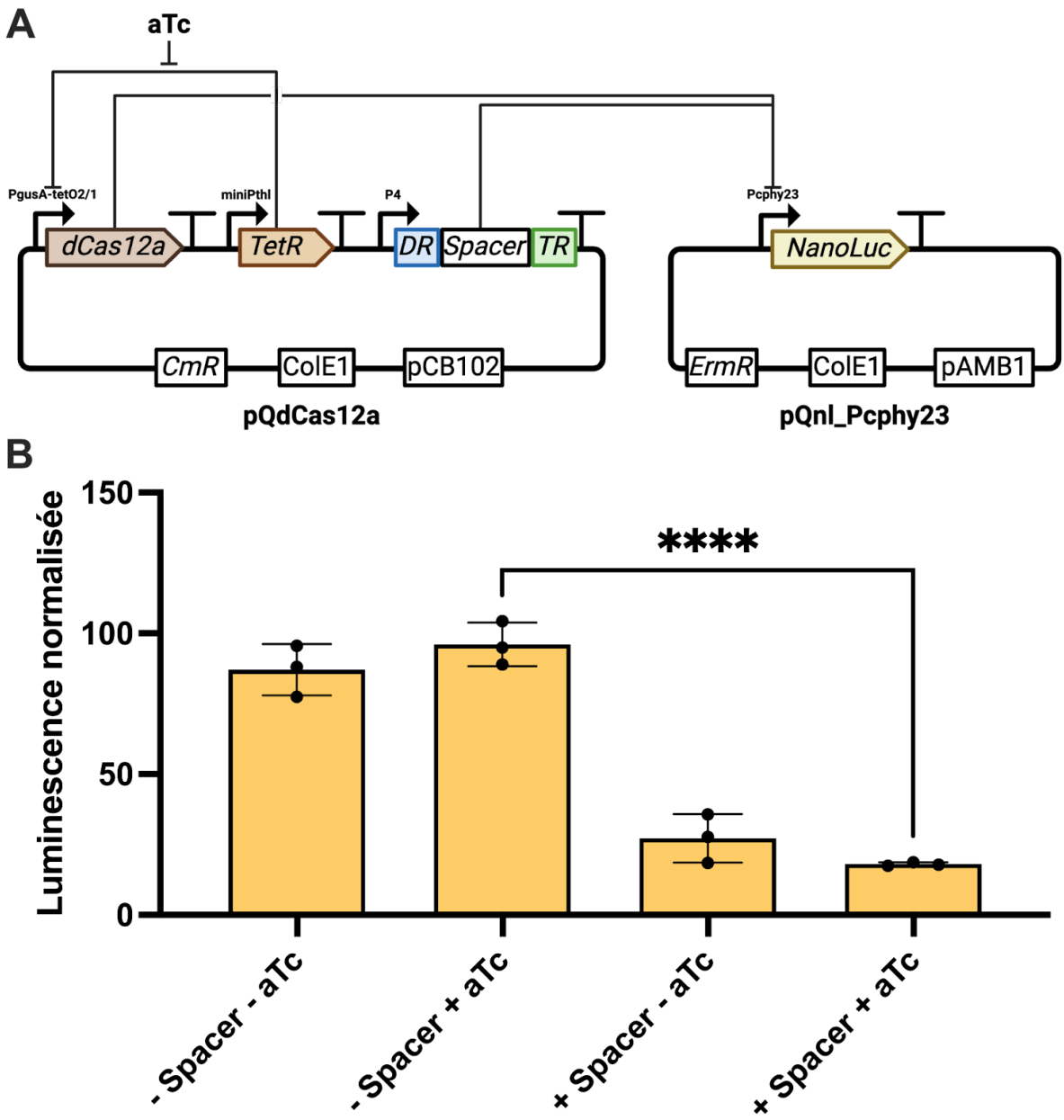


Figure 39 : Répression transcriptionnelle du Nanoluc chez L. phytofermentans par CRISPRi.
(A) Schéma du système de CRISPRi/Nanoluc sur 2 plasmides. L'induction de l'expression de la protéine dCas12a est faite en ajoutant de l'aTc pour lever l'inhibition transcriptionnelle faite par TetR sur le promoteur PgusA-O2-1. (B) Quantification de la luminescence émise par la luciférase produite par Nanoluc. Les différences entre les groupes induitent à l'aTc ont été analysées à l'aide d'un test t non apparié. $p < 0.0001$

3.2. *Amélioration de la stratégie de clonage de guide pour le multiplexage*

Les travaux antérieurs ont démontré l'efficacité d'un guide unique ciblant l'expression du gène *NanoLuc* [165]. En plus de sa compatibilité avec l'expression chez *L. phytofermentans* grâce à son activité RNase intrinsèque, LbCas12a se distingue par sa capacité à être aisément adaptée au multiplexage, contrairement à la Cas9. Les applications

du multiplexage en CRISPRi sont variées. Nous présenterons ici la stratégie adoptée pour optimiser le débit de production de plasmides contenant des guides avec plusieurs *spacers* ciblant différents gènes.

En raison des répétitions directes (DR) contenues dans la matrice de *spacers*, la synthèse de fragments avec plus de deux DR s'avère impossible. Nous avons donc opté pour une alternative : l'utilisation d'amorces spécifiquement conçus, incluant la séquence d'un *spacer* ainsi qu'une partie de la DR précédente en amont et le début de la DR suivante en aval du guide. Cette approche permet de décaler légèrement les extrémités cohésives de chaque DR, facilitant ainsi le clonage de six 3 à 12 *spacers* en une seule réaction *Golden Gate*. L'objectif principal de la modification de la plateforme de clonage de guides est de simplifier considérablement le clonage des *spacers* pour une utilisation en multiplex avec dLbCas12a, exploitant ainsi sa caractéristique intrinsèque de maturation des guides, s'avérant un avantage majeur par rapport à dCas9.

En dépit de la présence de la mutation inactivatrice D832A de la nucléase RuvC, le plasmide pQdC12a dans une souche de clonage *E. coli* MDS42 ralentit significativement la croissance et engendre des mutations non désirées dans le plasmide. La toxicité de dLbCas12a chez *E. coli* a conduit à la nécessité de sous-cloner les guides pour faciliter l'assemblage des fragments de guides.

Initialement, les guides sont construits à partir d'oligonucléotides contenant la séquence spécifique du *spacer* ainsi qu'une portion de la DR flanquant le *spacer* (**Figure 40**). Les oligonucléotides sont hybridés deux à deux avec leurs homologues pour former des fragments d'ADN double brin d'environ 40 paires de bases, avec des extrémités cohésives permettant leur assemblage par ligation lors d'une réaction *Golden Gate* avec le plasmide pTZ014. Ensuite, le plasmide est validé par séquençage, et la matrice de guides est transférée par *Golden Gate* dans le plasmide accepteur pQdC12a-*ccdB*. En raison de la toxicité de dLbCas12a chez *E. coli*, il est impératif d'effectuer ces clonages en cultivant *E. coli* à des températures où dLbCas12a est la moins active, c'est-à-dire entre 24 et 28°C. La stratégie complète de construction de la plateforme de clonage de guides est décrite dans la section Matériaux & Méthodes «3.3»

Bien que cette stratégie demande plus de manipulation qu'un simple clonage de guide, elle permet de cloner de nombreuses combinaisons de *spacers* pour validation par séquençage, garantissant ainsi la validité des guides. Sans ce sous-clonage, chaque miniprep ou criblage nécessiterait des temps d'incubation plus longs en raison des températures de culture plus basses. En définitive, cette méthode permet de gagner du temps tout en assurant la validité des guides clonés dans pQdC12a-ccdB.

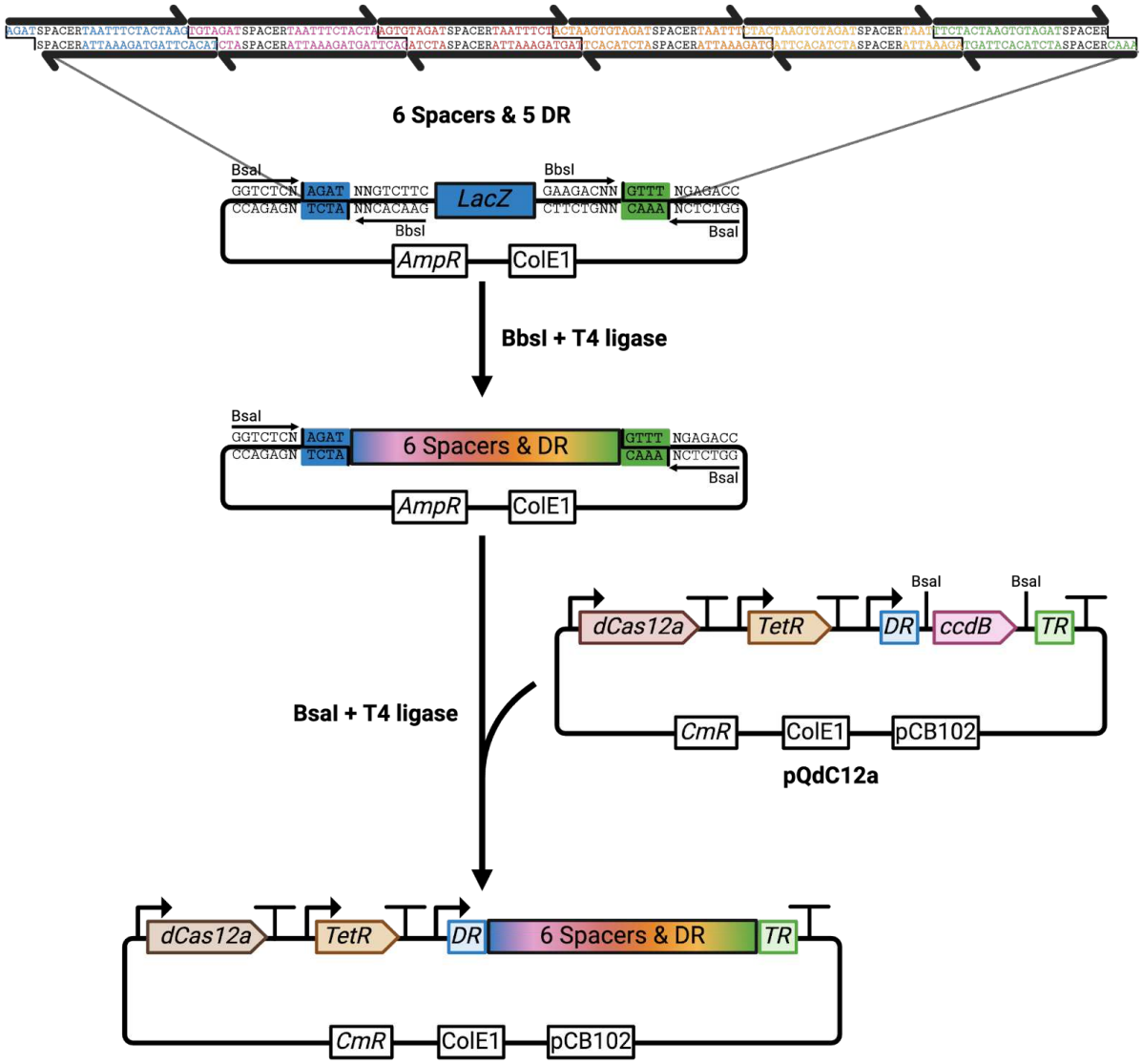


Figure 40 : Amélioration de la plateforme de clonage de guide et multiplexage du CRISPRi.
Les séquences des spacers sont contenus sur des amorces, hybridé deux à deux, phosphorylé puis assemblées dans un plasmide de sous clonage par Golden Gate en utilisant BbsI. Un second Golden gate avec BsaI permet d'assembler les spacers dans le plasmide pQdC12a pour une expression chez L. phytofermetans.

3.3. Design et stratégie expérimentale

Une fois la stratégie de clonage établie, le choix des gènes à réprimer a été réalisé dans l'objectif d'induire des modifications dans le métabolisme central de *Lachnoclostridium phytofermentans (***Figure 41*A*)**. Cette approche a été spécifiquement conçue réprimer la production des produits majeurs de fermentation, avec une attention particulière portée sur l'optimisation des flux métaboliques (**Figure 41B**).

La sélection des *spacers* a été réalisée afin de cibler chaque gène à deux localisations distinctes. Le premier *spacer* a été positionné dans la région promotrice des gènes, conformément aux résultats préalables indiquant que le ciblage des promoteurs par dLbCas12a est efficace pour la répression du gène rapporteur NanoLuc [165]. En effet, ces données démontrent que l'inhibition de la transcription peut être efficacement réalisée via l'interférence avec la région promotrice. La détermination des *spacers* ciblant les régions promotrices été faite en réalisant des prédiction de séquence promotrice [192] et à l'aide des *Transcription Start Site* (TSS) [193].

En complément, la littérature suggère que, dans les stratégies basées sur CRISPRi, l'augmentation du nombre de *spacers* ciblant un même gène renforce la répression transcriptionnelle de celui-ci. Sur cette base, une approche multipliant les points d'intervention a été adoptée, avec un deuxième *spacer* ciblant le début de la séquence codante. Cette double intervention permettrait non seulement de bloquer l'initiation de la transcription au niveau du promoteur, mais aussi de stopper toute transcription résiduelle qui aurait pu échapper à la première inhibition.

Les différentes souches possédant le système CRISPRi (**Figure 41**) ont été obtenues par électroporation de plasmide pQdC12a, contenant des ensembles de *spacers* variables (de 4 à 8 *spacers*). L'objectif principal de cette approche est d'induire une répression métabolique maximale, afin de rediriger efficacement les flux métaboliques. Concernant la répression de la production d'acétate et de lactate, l'étude visait à explorer la possibilité d'augmenter la production d'éthanol chez *L. phytofermentans* en modifiant son métabolisme central. Inversement, une répression de la production d'éthanol a également été explorée pour évaluer les effets d'une inhibition sur une voie métabolique hautement

conservée, afin de mieux comprendre l'impact potentiel de cette manipulation sur les autres voies fermentaires.

Les différentes souches présentées dans la **Figure 41A** n'ont pas encore toutes été caractérisées en termes de profil de croissance, d'efficacité de la répression transcriptionnelle des gènes ciblés via RT-qPCR, et de profil métabolique par HPLC. Par conséquent, seules les souches **LpTZ072** et **LpTZ080** seront présentées et analysées dans la section suivante.

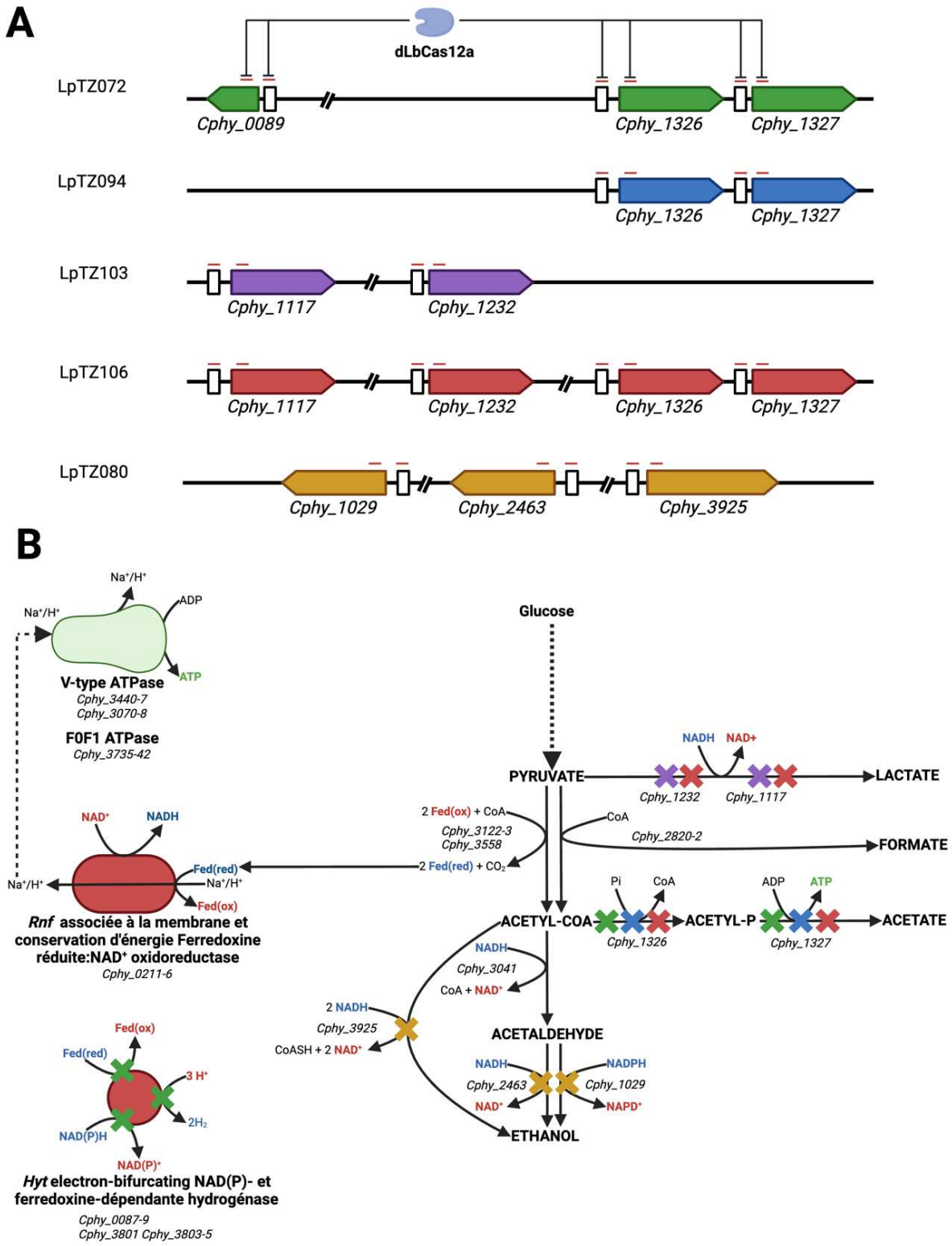


Figure 41 : Répression métabolique effectuer chez L. phytofermentans.
(A) Représentation des zones ciblées par le système CRISPRi, utilisant des spacers spécifiques pour réprimer l'expression des gènes. Les numéros ***LpTZ*** *correspondent aux différentes souches de* L. phytofermentans*, chacune contenant le plasmide pQdC12a intégrant les spacers nécessaires à la répression transcriptionnelle des gènes indiqués. (B) Schéma simplifié du métabolisme de* L. phytofermentans*. Les croix colorées indiquent les réactions théoriquement inhibées en cas de répression des gènes correspondants par CRISPRi. Les équivalents réduits (NADH/NADPH et Ferrédoxine réduite) sont représentés en bleu, tandis que les formes oxydées (NAD+/NADP+ et Ferrédoxine oxydée) sont en rouge, et l'ATP est en vert.*

3.4. Répression métabolique par CRISPRi

Dans un premier temps, le ciblage de la production d'acétate a été effectuer. En effet, l'inhibition de la production d'acétate est fréquemment utilisée dans diverses études pour stimuler la synthèse d'autres métabolites chez les *Clostridia* solvantogènes [25]. Pour ce faire, nous avons ciblé les gènes ***Cphy_1326*** et ***Cphy_1327***, codant respectivement la phosphotransacétylase (pta) et l'acétate kinase (ackA), les deux enzymes clés dans la formation de l'acétate. En complément, l'équilibre redox de la cellule a également été modifié en ciblant le gène ***Cphy_0089***, codant une sous-unité de l'hydrogénase impliquée dans la bifurcation des électrons via $NAD(P)^+$ et la ferrédoxine.

En ciblant simultanément les gènes ***Cphy_0089***, ***Cphy_1326***, et ***Cphy_1327***, il est attendu que cette stratégie permette une optimisation de la répartition des flux métaboliques en faveur de la production d'éthanol du fait de la forte expression des alcools déshydrogénase chez *L. phytofermentans*. L'inactivation de ***Cphy_0089*** devrait entraîner une réduction de la production de H_2, accompagnée d'une diminution des niveaux de $NAD(P)^+$ et de ferrédoxine oxydée (Fed(ox)). En parallèle, une augmentation de la disponibilité en protons (H^+), en NAD(P)H, et en ferrédoxine réduite (Fed(red)) est prévue. Cette redistribution des électrons pourrait conduire à une réduction accrue de NAD(P)H, favorisant la production d'éthanol et de lactate, et par conséquent augmenter le rendement de ces métabolites.

De plus, l'inactivation coordonnée des gènes ***Cphy_1326*** et ***Cphy_1327*** devrait entraîner une réduction significative de la production d'acétate, réduisant ainsi la génération d'ATP via cette voie. Parallèlement, l'augmentation des niveaux de Fed(red) et de protons pourrait stimuler l'activité du complexe RNF, contribuant ainsi à une production accrue d'ATP via les ATPases. Cette production d'ATP supplémentaire pourrait compenser la perte d'ATP associée à la répression de la voie de production de l'acétate, maintenant ainsi un équilibre énergétique stable au sein des cellules.

Dans un second temps, en se basant sur des données transcriptomiques de la littératures *(39)*, nous avons ciblé trois gènes codant les alcools déshydrogénases les plus exprimés dans la cellule afin d'étudier les conséquences de leur répression sur le

métabolisme fermentaire. Cela permettrait de vérifier si la cellule compenserait en produisant d'autres métabolites, comme l'acétate. Les gènes ***Cphy_3925***, codant une enzyme bifonctionnelle aldéhyde/alcool déshydrogénase, ainsi que ***Cphy_2463*** et ***Cphy_1029***, deux gènes codant des alcools déshydrogénases ont été ciblé de la même manière que les gènes produisant de l'acétate. ***Cphy_1029*** a été validé par des analyses in vitro, montrant une préférence pour le NADPH, contrairement aux deux autres enzymes qui utilisent exclusivement le NADH comme cofacteur.

Afin de valider la fonctionnalité du système CRISPRi pour moduler le métabolisme de *Lachnoclostridium phytofermentans*, différentes souches ont été cultivées puis diluées en présence ou en absence d'aTc pour suivre leur croissance à l'aide d'un lecteur de plaque. Une fois les courbes de croissance obtenues (**Figure 42**), les métabolites ont été quantifiés par HPLC à partir des mêmes plaques de culture (**Figure 42C**). En parallèle, ces mêmes souches ont été mise en culture pour quantifier la répression transcriptionnelle des gènes ciblés via RT-qPCR (**Figure 42B**).

Les courbes de croissance des souches **LpTZ072** et **LpTZ080** révèlent une dynamique de croissance similaire à celle de *L. phytofermentans* WT présentée dans les sections précédentes de cette thèse. Cependant, en présence d'aTc, induisant l'expression de la protéine **dLbdCas12a**, une réduction du taux de croissance et de la densité optique maximale (OD_{600}) a été observée. Cette diminution pourrait être liée à la surexpression de la protéine dLbCas12a et son action sur la répression des gènes du métabolisme centrale**.** Toutefois, les résultats de RT-qPCR (**Figure 42B**) montrent que la répression des gènes ciblés est très efficace, atteignant en moyenne 90 % de répression. La répression de gènes impliqués dans le métabolisme central, tels que ceux ciblés dans cette étude, pourrait expliquer la réduction marquée de la croissance des souches.

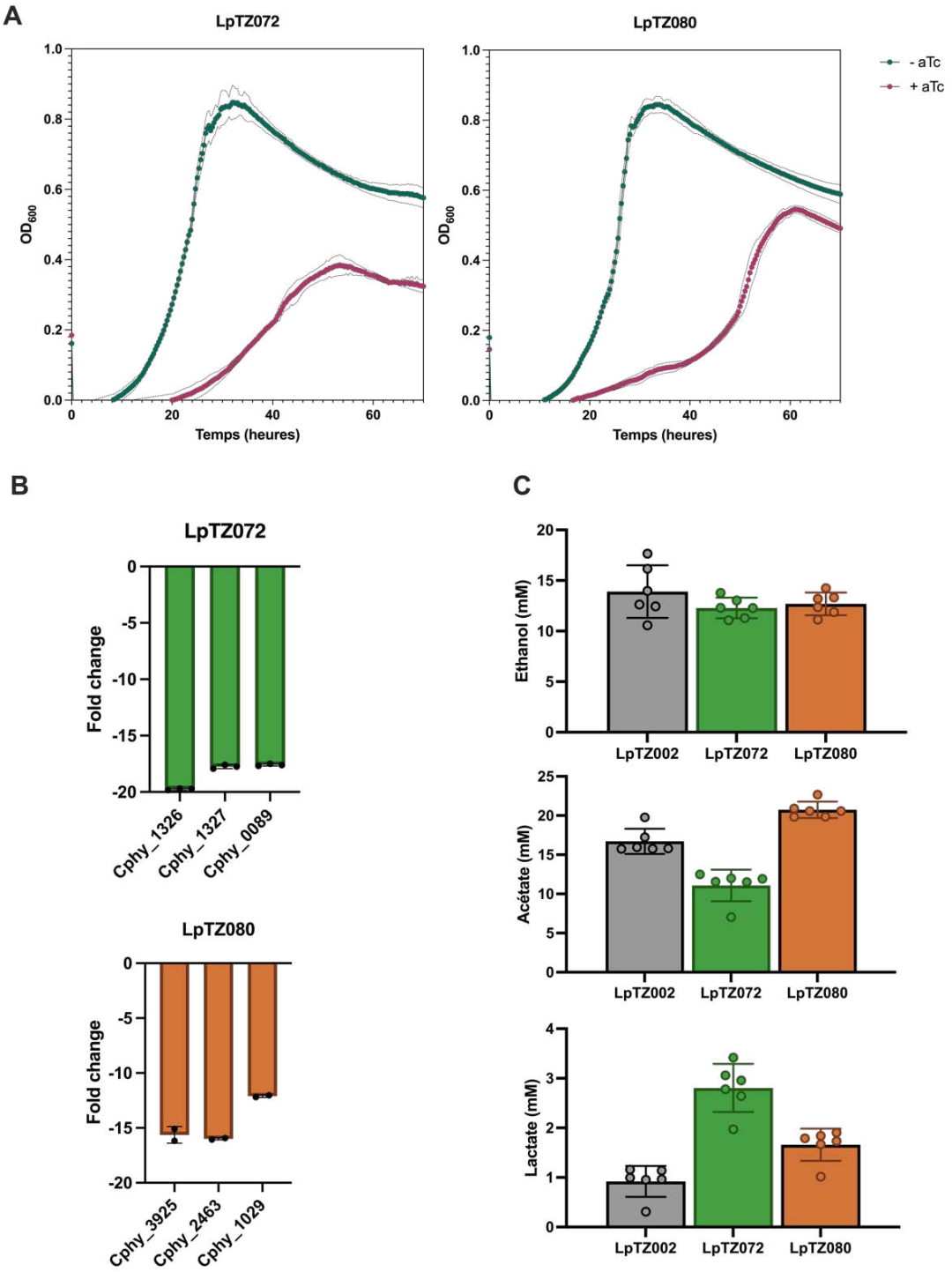


***Figure 42 : Impact du CRISPRi sur les gènes du métabolisme central de* L. phytofermentans.**
*(A) Courbes de croissance des souches **LpTZ072** et **LpTZ080**, mesurées en fonction du temps (OD600), avec et sans induction de **dLbCas12a** par l'ajout d'aTc. Chaque point représente la moyenne de 6 réplicats indépendants (n=6). (B) Analyse de l'expression des gènes ciblés dans les deux souches par qRT-PCR dans des conditions d'induction par l'aTc. La répression de l'expression génique est exprimée en fold change par rapport à la souche contrôle **LpTZ002 en présence d'aTc**. (C) Rendement de production des métabolites de fermentation (éthanol, acétate, lactate) mesuré par HPLC dans des conditions d'induction par l'aTc. Ces résultats sont issus de la plaque de culture réalisé au préalable en (A) pouvoir suivre la croissance de chaque puits et associer un profile de fermentation avec une courbe de croissance. Une comparaison des différentes souches est possible car la quantité de glucose restante a été quantifié et chacun des puits mesurés n'en contenait plus. Les valeurs sont exprimées comme moyenne de 6 réplicats indépendants (n=6)*

Bien que les résultats de RT-qPCR présentés soient préliminaires, ils démontrent clairement chez les deux souches que la répression transcriptionnelle induite par ce système est efficace. Des optimisations sont en cours pour affiner les conditions de culture afin d'améliorer l'extraction des ARNm pendant la phase exponentielle de croissance. En effet, ces souches sont affectées par l'expression de **dLbCas12a** et par les effets des répressions transcriptionnelles, ce qui rend leur croissance plus difficile et complexe pour la réalisation d'expériences.

La quantification des métabolites a permis de mettre en évidence une corrélation entre la répression transcriptionnelle validée par RT-qPCR et les profils de fermentation obtenus par HPLC. En comparaison avec la souche contrôle **LpTZ002**, la souche **LpTZ072**, ciblant la production d'acétate et de H_2, a montré une diminution de la concentration en acétate dans le milieu de culture d'environ 1,5 fois (16,71 mM contre 11,08 mM). De manière surprenante, malgré le grand nombre d'aldéhyde et d'alcool déshydrogénases présentes dans le génome de *L. phytofermentans*, les flux de carbone ont été redirigés vers la production de lactate plutôt que d'éthanol. La concentration en lactate dans cette souche a été quantifiée à 2,80 mM contre 0,92 mM, soit une augmentation d'un facteur 3. En revanche, la concentration en éthanol chez **LpTZ072** n'a pas été significativement augmentée.

Dans le cas de la souche **LpTZ080**, les données obtenues par HPLC soulèvent des questions. En dépit de l'efficacité de la répression transcriptionnelle des gènes ***Cphy_3925***, ***Cphy_2463***, et ***Cphy_1029***, le rendement en éthanol ne semble pas avoir été affecté. Toutefois, une légère augmentation de la production d'acétate et de lactate a été observée. Ces résultats suggèrent plusieurs hypothèses. D'une part, il est possible que d'autres alcools déshydrogénases dans la cellule compensent la diminution de l'expression des gènes réprimés, permettant ainsi une production d'éthanol comparable à celle du contrôle. D'autre part, il est également envisageable que la répression ne soit pas complète et qu'une quantité résiduelle d'enzymes soit suffisante pour maintenir une production d'éthanol similaire.

L'hypothèse la plus plausible est que cette répression n'affecte que la vitesse de production d'éthanol, également appelée productivité volumétrique. En conséquence, cela pourrait rediriger davantage de flux de carbone vers la production d'acétate et de lactate, comme le suggèrent les résultats présentés dans la **Figure 42C**. Cependant, étant donné que la mesure des métabolites a été réalisée à un point final au bout de plusieurs jours de cultures, il n'est pas possible de valider cette hypothèse de ralentissement de la production d'éthanol sans des analyses dynamiques supplémentaires.

3.5. Discussion et Perspectives

L'implémentation d'un système CRISPRi chez *Lachnoclostridium phytofermentans* a été réalisée dans le but de simplifier les modifications métaboliques, tout en s'affranchissant des outils de modification génomique classiques, souvent complexes et lourdes à mettre en place, en particulier chez une bactérie peu étudiée dans ce domaine. À ce jour, très peu de publications en dehors de notre équipe décrivent des méthodes d'ingénierie génétique pour cette espèce. Le développement d'outils moléculaires spécifiques pour *L. phytofermentans* représente donc un défi considérable, tant par la lenteur que par les difficultés rencontrées.

Dans cette étude, l'utilisation du système CRISPRi a démontré un réel potentiel pour moduler le métabolisme de manière efficace. Bien que ce système ne mime pas une délétion génomique, qui entraînerait une perte totale de fonction, il offre néanmoins un aperçu précieux des mécanismes compensatoires mis en place par la cellule pour contrer la répression des gènes ciblés. Cette approche permet d'explorer les voies métaboliques et d'observer les ajustements cellulaires en réponse à une modulation de l'expression génique, ouvrant ainsi la voie à des stratégies d'ingénierie plus concrètes.

L'amélioration de la plateforme de clonage a joué un rôle clé dans l'efficacité de la stratégie mise en place. Le design spécifique des spacers, en chevauchant les *Direct Repeats*, a permis de cloner jusqu'à 8 spacers, avec un potentiel théorique pouvant aller jusqu'à 12 spacers en une seule réaction. Cette stratégie de clonage, en plus d'être efficace, offre la possibilité de créer des librairies de spacers permettant le test simultané de plusieurs gènes. Cependant, la faible efficacité de transformation chez *L. phytofermentans* a limité l'exploration d'un large éventail de cibles, restreignant ainsi notre capacité à tester une grande diversité de possibilités.

Parmi les souches caractérisées, **LpTZ072** (réprimant la production d'acétate et de H_2) et **LpTZ080** (réprimant la production d'éthanol) ont démontré l'efficacité du système CRISPRi, tant au niveau de la répression transcriptionnelle qu'en termes de modulation métabolique. Le mode de culture et de screening utilisé dans cette étude représente une avancée par rapport aux approches précédentes *(162)*. L'utilisation de plaques de culture

96 puits a permis non seulement de suivre la croissance des souches de manière plus précise, mais aussi d'améliorer la reproductibilité des expériences. De plus, les mêmes plaques utilisées pour suivre la croissance ont servi à l'analyse des métabolites d'intérêt (éthanol, acétate, et lactate), corrélant ainsi les profils de croissance avec la production des métabolites d'intérêt. Cette approche à plus haut débit, en augmentant le nombre de réplicats biologiques, a contribué à réduire la variabilité expérimentale, garantissant ainsi une plus grande robustesse et fiabilité des résultats obtenus.

Ces avancées ouvrent de nouvelles perspectives pour l'optimisation du métabolisme de cette bactérie, aussi bien à court qu'à long terme. À court terme, la finalisation de cette étude par l'obtention de données supplémentaires concernant les autres plasmides mis en place permettra de mieux comprendre l'impact de la répression de gènes tels que ***Cphy_0089*** dans la redirection des flux métabolique par rapport à une souche ne réprimant que *Cphy_1326* et *Cphy_1327*.

De plus, la répression des gènes codant des lactate déshydrogénases pourrait révéler une augmentation des rendements en acétate et en éthanol. L'hypothèse principale est que la répression conjointe des gènes impliqués dans la production d'acétate (***Cphy_1326*** et ***Cphy_1327***) et de lactate (***Cphy_1117*** et ***Cphy_1232***) pourrait permettre d'augmenter significativement la productivité en éthanol chez *L. phytofermentans*. La validation de cette hypothèse fournirait des bases solides pour valider la répression de plusieurs voies métaboliques simultané.

À long terme, le potentiel du système CRISPRi dans l'étude de *L. phytofermentans* dépasse largement l'optimisation du métabolisme. Ce système offre de nombreuses perspectives pour explorer des mécanismes biologiques plus fondamentaux, tels que la sporulation, la division cellulaire, ainsi que la régulation génique en réponse à l'adaptation aux substrats carbonés disponibles. Grâce à sa modularité, le CRISPRi permettrait d'examiner divers aspects de la biologie de cette bactérie, encore peu étudiée, et de la positionner comme un modèle pertinent pour la compréhension des processus fermentaires au sein des *Lachnospiraceae*.

CONCLUSION

Pour conclure de manière générale sur ce projet de différentes approches ont été mises en œuvre pour faire avancer les connaissances liées à la production de butanol par *Lachnoclostridium phytofermentans*, directement ou indirectement. Certaines de ces approches ont connu du succès, d'autres non.

Le premier axe de recherche visait à développer une souche évoluée de *L. phytofermentans* capable de croître dans des conditions de stress dues au butanol. Après 480 jours d'évolution, caractérisés par un phénomène d'hypermutation, plus de 500 mutations ont été observées dans le génome séquencé. Ces mutations ont contribué à augmenter les capacités d'adaptation de la souche aux conditions stressantes de l'expérience.

En pratique, en plus d'avoir amélioré la tolérance au butanol tout en conservant le phénotype de dégradation de la cellulose, un nouveau phénotype est apparu. La souche étudiée, PHY24.0#6, présente un temps de génération réduit de moitié et une densité de croissance maximale augmentée d'un facteur 1,3 dans des conditions de culture similaires à celles de son évolution. Un des mécanismes liés à cette résistance a été mis en évidence en analysant la membrane plasmique de cette souche évoluée. Il a été prouvé que cette bactérie a acquis une membrane plasmique composée d'acides gras en moyenne plus longs et plus insaturés que la souche de référence *L. phytofermentans* ISDg^{T}.

Dans le même esprit que la première partie, une approche de biologie synthétique a été mise en place pour produire du butanol cellulosique par une souche recombinante de *L. phytofermentans* possédant une voie de synthèse appropriée. La construction d'une voie de synthèse *de novo* à partir de données rationnelles a été entreprise avec succès, cependant les tests de production chez *L. phytofermentans n'ont pas abouti par une production détectable*. Cependant, un biosenseur spécifique du butanol a été adapté dans un seul plasmide pour à terme détecter le butanol chez les *Clostridia*.

Enfin, la partie la plus exploratoire de cette thèse a été la construction d'un système CRISPRi utilisable en multiplex chez *L. phytofermentans* pour le prototypage et l'étude du

métabolisme. Un protocole de clonage optimisé des guides a été mis en place pour cloner jusqu'à 8 spacers (12 étant théoriquement possibles avec un design similaire) dans un plasmide exprimant dLbCas12a, capable de réprimer spécifiquement l'expression des gènes d'intérêt. Dans cette thèse, le métabolisme central a été modulé en ciblant des voies métaboliques très conservées comme la production d'éthanol et d'acétate. Des répressions d'expression génique jusqu'à 20 fois inférieures au contrôle ont été observées. De plus, cette approche a permis de réprimer un trio de gènes dont la délétion serait impossible (*Cphy_1326-1327-0089*), car l'inactivation totale de cette combinaison de gènes serait létale.

Cet outil moléculaire ouvre un vaste champ de possibles pour l'étude du métabolisme, mais aussi des mécanismes plus fondamentaux tels que la division cellulaire et la sporulation chez *L. phytofermentans*. Ce système serait également transposable à d'autres modèles de *Clostridia*.

Au cours de ces trois années de thèse, j'ai participé à la publication de plusieurs articles dans des revues à comité de lecture. L'un d'eux, directement lié à cette thèse, porte sur la preuve de concept du système CRISPRi chez *L. phytofermentans*, ouvrant ainsi de nouvelles perspectives pour l'ingénierie métabolique de cette bactérie. Enfin, notre équipe a publié une revue sur l'utilisation des bactéries de la famille des *Lachnospiraceae* à des fins de bioproduction industrielle et comme traitement contre des pathologies intestinales et métaboliques. Une dernière publication a mis en évidence les implications de l'interaction entre le L-tryptophane et le cuivre pour réduire la génotoxicité de la colibactine sécrété chez une souche d'*E. coli* pks+. Cette étude apporte des éclairages sur les mécanismes moléculaires sous-jacents. Ces publications sont disponibles en annexes de cette thèse.

En parallèle des travaux expérimentaux réalisés dans cette thèse, j'ai eu l'opportunité d'effectuer une activité complémentaire. En tant que doctorant consultant, j'ai pu participer à la rédaction d'un dossier stratégique pour appuyer la création d'une biofondrie au sein du GIP Génopole. Ce projet avait pour objectif d'accroître la capacité de recherche en optimisant les processus biotechnologiques, grâce à la standardisation et l'automatisation des étapes clés du développement expérimental.

Cette expérience s'est révélée particulièrement enrichissante, car elle m'a permis d'acquérir des compétences essentielles en gestion de projet scientifique et en communication interdisciplinaire. En contribuant à la rédaction du rapport, j'ai approfondi ma compréhension des enjeux liés à l'industrialisation des processus de recherche en biotechnologie. J'ai également affiné ma capacité à assimiler des concepts complexes et à les transformer en propositions adaptées aux besoins du projet.

Le dossier a été non seulement utilisé en interne, mais a également donné lieu à une valorisation externe avec la publication d'un article de vulgarisation scientifique dans le journal Industrie & Technologies (n°1066, février 2024, disponible en annexe). Cet article a permis de présenter au grand public les enjeux et les avantages d'une biofondrie, tout en sensibilisant les acteurs industriels aux avancées réalisées dans le domaine des biotechnologies au sein du GIP Génopole.

Cette expérience a eu un impact direct sur ma perception des sciences, en m'offrant une perspective plus large sur les applications industrielles de la biotechnologie. Elle a également renforcé la valorisation de mes recherches en les intégrant dans un cadre plus global d'innovation et de transfert technologique.

RÉFÉRENCES

(1) Muñoz-Gómez, S. A.; Roger, A. J. Leaving Negative Ancestors Behind. *eLife* **2016**, *5*, e20061. https://doi.org/10.7554/eLife.20061.

(2) Doi, R. H.; Kosugi, A.; Murashima, K.; Tamaru, Y.; Han, S. O. Cellulosomes from Mesophilic Bacteria. *J. Bacteriol.* **2003**, *185* (20), 5907–5914. https://doi.org/10.1128/jb.185.20.5907-5914.2003.

(3) Rohde, M. The Gram-Positive Bacterial Cell Wall. *Microbiol. Spectr.* **2019**, *7* (3), 7.3.10. https://doi.org/10.1128/microbiolspec.GPP3-0044-2018.

(4) Al-Hinai, M. A.; Jones, S. W.; Papoutsakis, E. T. The Clostridium Sporulation Programs: Diversity and Preservation of Endospore Differentiation. *Microbiol. Mol. Biol. Rev.* **2015**, *79* (1), 19–37. https://doi.org/10.1128/mmbr.00025-14.

(5) Dürre, P. Physiology and Sporulation in Clostridium. *Microbiol. Spectr.* **2014**, *2* (4), 10.1128/microbiolspec.tbs-0010–2012. https://doi.org/10.1128/microbiolspec.tbs-0010-2012.

(6) Sasi Jyothsna, T. S.; Tushar, L.; Sasikala, C.; Ramana, C. V. Paraclostridium Benzoelyticum Gen. Nov., Sp. Nov., Isolated from Marine Sediment and Reclassification of Clostridium Bifermentans as Paraclostridium Bifermentans Comb. Nov. Proposal of a New Genus Paeniclostridium Gen. Nov. to Accommodate Clostridium Sordellii and Clostridium Ghonii. *Int. J. Syst. Evol. Microbiol.* **2016**, *66* (3), 1268–1274. https://doi.org/10.1099/ijsem.0.000874.

(7) Petit, E.; Coppi, M. V.; Hayes, J. C.; Tolonen, A. C.; Warnick, T.; Latouf, W. G.; Amisano, D.; Biddle, A.; Mukherjee, S.; Ivanova, N.; Lykidis, A.; Land, M.; Hauser, L.; Kyrpides, N.; Henrissat, B.; Lau, J.; Schnell, D. J.; Church, G. M.; Leschine, S. B.; Blanchard, J. L. Genome and Transcriptome of Clostridium Phytofermentans, Catalyst for the Direct Conversion of Plant Feedstocks to Fuels. *PLOS ONE* **2015**, *10* (6), e0118285. https://doi.org/10.1371/journal.pone.0118285.

(8) Liu, Z.-Y.; Jia, D.-C.; Zhang, K.-D.; Zhu, H.-F.; Zhang, Q.; Jiang, W.-H.; Gu, Y.; Li, F.-L. Ethanol Metabolism Dynamics in Clostridium Ljungdahlii Grown on Carbon Monoxide. *Appl. Environ. Microbiol.* **2020**, *86* (14), e00730-20. https://doi.org/10.1128/AEM.00730-20.

(9) Dürre, P. Handbook on Clostridia. 499.

(10) Collins, M. D.; Lawson, P. A.; Willems, A.; Cordoba, J. J.; Fernandez-Garayzabal, J.; Garcia, P.; Cai, J.; Hippe, H.; Farrow, J. A. The Phylogeny of the Genus Clostridium: Proposal of Five New Genera and Eleven New Species Combinations. *Int. J. Syst. Bacteriol.* **1994**, *44* (4), 812–826. https://doi.org/10.1099/00207713-44-4-812.

(11) Yutin, N.; Galperin, M. Y. A Genomic Update on Clostridial Phylogeny: Gram-Negative Spore Formers and Other Misplaced Clostridia. *Environ. Microbiol.* **2013**, *15* (10), 2631–2641. https://doi.org/10.1111/1462-2920.12173.

(12) Lawson, P. A.; Rainey, F. A. Proposal to Restrict the Genus Clostridium Prazmowski to Clostridium Butyricum and Related Species. *Int. J. Syst. Evol. Microbiol.* **2016**, *66* (2), 1009–1016. https://doi.org/10.1099/ijsem.0.000824.

(13) Purdy, D.; O'Keeffe, T. A. T.; Elmore, M.; Herbert, M.; McLeod, A.; Bokori-Brown, M.; Ostrowski, A.; Minton, N. P. Conjugative Transfer of Clostridial Shuttle Vectors from Escherichia Coli to Clostridium Difficile through Circumvention of the Restriction Barrier. *Mol. Microbiol.* **2002**, *46* (2), 439–452. https://doi.org/10.1046/j.1365-2958.2002.03134.x.

(14) Pyne, M. E.; Bruder, M. R.; Moo-Young, M.; Chung, D. A.; Chou, C. P. Harnessing Heterologous and Endogenous CRISPR-Cas Machineries for Efficient Markerless Genome Editing in Clostridium. *Sci. Rep.* **2016**, *6*, 25666. https://doi.org/10.1038/srep25666.

(15) Garnier, T.; Cole, S. T. Complete Nucleotide Sequence and Genetic Organization of the Bacteriocinogenic Plasmid, pIP404, from *Clostridium Perfringens*. *Plasmid* **1988**, *19* (2), 134–150. https://doi.org/10.1016/0147-619X(88)90052-2.

(16) Leang, C.; Ueki, T.; Nevin, K. P.; Lovley, D. R. A Genetic System for Clostridium Ljungdahlii: A Chassis for Autotrophic Production of Biocommodities and a Model Homoacetogen. *Appl. Environ. Microbiol.* **2013**, *79* (4), 1102–1109. https://doi.org/10.1128/AEM.02891-12.

(17) Jennert, K. C. B.; Tardif, C.; Young, D. I.; Young, M. Gene Transfer to Clostridium Cellulolyticum ATCC 35319. *Microbiol. Read. Engl.* **2000**, *146 Pt 12*, 3071–3080. https://doi.org/10.1099/00221287-146-12-3071.

(18) Lessa, F. C.; Mu, Y.; Bamberg, W. M.; Beldavs, Z. G.; Dumyati, G. K.; Dunn, J. R.; Farley, M. M.; Holzbauer, S. M.; Meek, J. I.; Phipps, E. C.; Wilson, L. E.; Winston, L. G.; Cohen, J. A.; Limbago, B. M.; Fridkin, S. K.; Gerding, D. N.; McDonald, L. C. Burden of Clostridium Difficile Infection in the United States. *N. Engl. J. Med.* **2015**, *372* (9), 825–834. https://doi.org/10.1056/NEJMoa1408913.

(19) Hempel, S.; Newberry, S. J.; Maher, A. R.; Wang, Z.; Miles, J. N. V.; Shanman, R.; Johnsen, B.; Shekelle, P. G. Probiotics for the Prevention and Treatment of Antibiotic-Associated Diarrhea: A Systematic Review and Meta-Analysis. *JAMA* **2012**, *307* (18), 1959–1969. https://doi.org/10.1001/jama.2012.3507.

(20) Zaplana, T.; Miele, S.; Tolonen, A. C. Lachnospiraceae Are Emerging Industrial Biocatalysts and Biotherapeutics. *Front. Bioeng. Biotechnol.* **2024**, *11*. https://doi.org/10.3389/fbioe.2023.1324396.

(21) Weber, C.; Farwick, A.; Benisch, F.; Brat, D.; Dietz, H.; Subtil, T.; Boles, E. Trends and Challenges in the Microbial Production of Lignocellulosic Bioalcohol Fuels. *Appl.*

Microbiol. Biotechnol. **2010**, *87* (4), 1303–1315. https://doi.org/10.1007/s00253-010-2707-z.

(22) Yang, Z.; Leero, D. D.; Yin, C.; Yang, L.; Zhu, L.; Zhu, Z.; Jiang, L. *Clostridium* as Microbial Cell Factory to Enable the Sustainable Utilization of Three Generations of Feedstocks. *Bioresour. Technol.* **2022**, *361*, 127656. https://doi.org/10.1016/j.biortech.2022.127656.

(23) Vees, C. A.; Neuendorf, C. S.; Pflügl, S. Towards Continuous Industrial Bioprocessing with Solventogenic and Acetogenic Clostridia: Challenges, Progress and Perspectives. *J. Ind. Microbiol. Biotechnol.* **2020**, *47* (9–10), 753–787. https://doi.org/10.1007/s10295-020-02296-2.

(24) Weizmann, C. Improvements in the Bacterial Fermentation of Carbohydrates and in Bacterial Cultures for the Same. GB191504845A, March 6, 1919. https://patents.google.com/patent/GB191504845A/en (accessed 2024-03-12).

(25) Nguyen, N.-P.-T.; Raynaud, C.; Meynial-Salles, I.; Soucaille, P. Reviving the Weizmann Process for Commercial N-Butanol Production. *Nat. Commun.* **2018**, *9* (1), 3682. https://doi.org/10.1038/s41467-018-05661-z.

(26) Bengelsdorf, F. R.; Beck, M. H.; Erz, C.; Hoffmeister, S.; Karl, M. M.; Riegler, P.; Wirth, S.; Poehlein, A.; Weuster-Botz, D.; Dürre, P. Bacterial Anaerobic Synthesis Gas (Syngas) and CO2 + H2 Fermentation. In *Advances in Applied Microbiology*; Sariaslani, S., Gadd, G. M., Eds.; Academic Press, 2018; Vol. 103, pp 143–221. https://doi.org/10.1016/bs.aambs.2018.01.002.

(27) Younesi, H.; Najafpour, G.; Mohamed, A. R. Ethanol and Acetate Production from Synthesis Gas via Fermentation Processes Using Anaerobic Bacterium, *Clostridium Ljungdahlii*. *Biochem. Eng. J.* **2005**, *27* (2), 110–119. https://doi.org/10.1016/j.bej.2005.08.015.

(28) Shen, Y.; Brown, R.; Wen, Z. Enhancing Mass Transfer and Ethanol Production in Syngas Fermentation of *Clostridium Carboxidivorans* P7 through a Monolithic Biofilm Reactor. *Appl. Energy* **2014**, *136*, 68–76. https://doi.org/10.1016/j.apenergy.2014.08.117.

(29) Abubackar, H. N.; Veiga, M. C.; Kennes, C. Syngas Fermentation for Bioethanol and Bioproducts. In *Sustainable Resource Recovery and Zero Waste Approaches*; Elsevier, 2019; pp 207–221. https://doi.org/10.1016/B978-0-444-64200-4.00015-3.

(30) Ragsdale, S. W.; Pierce, E. Acetogenesis and the Wood–Ljungdahl Pathway of CO2 Fixation. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2008**, *1784* (12), 1873–1898. https://doi.org/10.1016/j.bbapap.2008.08.012.

(31) *LanzaTech Produces Ethylene from CO2, Changing the Way We Make Products Today – LanzaTech*. https://lanzatech.com/lanzatech-produces-ethylene-from-co2-changing-the-way-we-make-products-today/ (accessed 2024-03-12).

(32) Mazzoli, R.; Olson, D. G. *Clostridium Thermocellum*: A Microbial Platform for High-Value Chemical Production from Lignocellulose. In *Advances in Applied Microbiology*; Gadd, G. M., Sariaslani, S., Eds.; Academic Press, 2020; Vol. 113, pp 111–161. https://doi.org/10.1016/bs.aambs.2020.07.004.

(33) Akinosho, H.; Yee, K.; Close, D.; Ragauskas, A. The Emergence of Clostridium Thermocellum as a High Utility Candidate for Consolidated Bioprocessing Applications. *Front. Chem.* **2014**, *2*. https://doi.org/10.3389/fchem.2014.00066.

(34) *Cellulosome Systems | Ed Bayer's Group*. https://www.weizmann.ac.il/Biomolecular_Sciences/Bayer/research-activities/cellulosome-systems (accessed 2024-03-12).

(35) Artzi, L.; Bayer, E. A.; Moraïs, S. Cellulosomes: Bacterial Nanomachines for Dismantling Plant Polysaccharides. *Nat. Rev. Microbiol.* **2017**, *15* (2), 83–95. https://doi.org/10.1038/nrmicro.2016.164.

(36) Moraïs, S.; Morag, E.; Barak, Y.; Goldman, D.; Hadar, Y.; Lamed, R.; Shoham, Y.; Wilson, D. B.; Bayer, E. A. Deconstruction of Lignocellulose into Soluble Sugars by Native and Designer Cellulosomes. *mBio* **2012**, *3* (6), 10.1128/mbio.00508-12. https://doi.org/10.1128/mbio.00508-12.

(37) Vanderstraeten, J.; da Fonseca, M. J. M.; De Groote, P.; Grimon, D.; Gerstmans, H.; Kahn, A.; Moraïs, S.; Bayer, E. A.; Briers, Y. Combinatorial Assembly and Optimisation of Designer Cellulosomes: A Galactomannan Case Study. *Biotechnol. Biofuels Bioprod.* **2022**, *15* (1), 60. https://doi.org/10.1186/s13068-022-02158-2.

(38) Warnick, T. A.; Methé, B. A.; Leschine, S. B. Clostridium Phytofermentans Sp. Nov., a Cellulolytic Mesophile from Forest Soil. *Int. J. Syst. Evol. Microbiol.* **2002**, *52* (Pt 4), 1155–1160. https://doi.org/10.1099/00207713-52-4-1155.

(39) Boutard, M.; Cerisy, T.; Nogue, P.-Y.; Alberti, A.; Weissenbach, J.; Salanoubat, M.; Tolonen, A. C. Functional Diversity of Carbohydrate-Active Enzymes Enabling a Bacterium to Ferment Plant Biomass. *PLoS Genet.* **2014**, *10* (11), e1004773. https://doi.org/10.1371/journal.pgen.1004773.

(40) Tolonen, A. C.; Haas, W.; Chilaka, A. C.; Aach, J.; Gygi, S. P.; Church, G. M. Proteome-wide Systems Analysis of a Cellulosic Biofuel-producing Microbe. *Mol. Syst. Biol.* **2011**, *7* (1), 461. https://doi.org/10.1038/msb.2010.116.

(41) Cerisy, T.; Iglesias, A.; Rostain, W.; Boutard, M.; Pelle, C.; Perret, A.; Salanoubat, M.; Fierobe, H.-P.; Tolonen, A. C. ABC Transporters Required for Hexose Uptake by Clostridium Phytofermentans. *J. Bacteriol.* **2019**, *201* (15), 10.1128/jb.00241-19. https://doi.org/10.1128/jb.00241-19.

(42) Petit, E.; LaTouf, W. G.; Coppi, M. V.; Warnick, T. A.; Currie, D.; Romashko, I.; Deshpande, S.; Haas, K.; Alvelo-Maurosa, J. G.; Wardman, C.; Schnell, D. J.; Leschine, S. B.; Blanchard, J. L. Involvement of a Bacterial Microcompartment in the Metabolism

of Fucose and Rhamnose by Clostridium Phytofermentans. *PloS One* **2013**, *8* (1), e54337. https://doi.org/10.1371/journal.pone.0054337.

(43) Kerfeld, C. A.; Aussignargues, C.; Zarzycki, J.; Cai, F.; Sutter, M. Bacterial Microcompartments. *Nat. Rev. Microbiol.* **2018**, *16* (5), 277–290. https://doi.org/10.1038/nrmicro.2018.10.

(44) Zuroff, T. R.; Xiques, S. B.; Curtis, W. R. Consortia-Mediated Bioprocessing of Cellulose to Ethanol with a Symbiotic Clostridium Phytofermentans/Yeast Co-Culture. *Biotechnol. Biofuels* **2013**, *6* (1), 59. https://doi.org/10.1186/1754-6834-6-59.

(45) Park, H.; Patel, A.; Hunt, K. A.; Henson, M. A.; Carlson, R. P. Artificial Consortium Demonstrates Emergent Properties of Enhanced Cellulosic-Sugar Degradation and Biofuel Synthesis. *NPJ Biofilms Microbiomes* **2020**, *6* (1), 59. https://doi.org/10.1038/s41522-020-00170-8.

(46) Brat, D.; Boles, E.; Wiedemann, B. Functional Expression of a Bacterial Xylose Isomerase in Saccharomyces Cerevisiae. *Appl. Environ. Microbiol.* **2009**, *75* (8), 2304–2311. https://doi.org/10.1128/AEM.02522-08.

(47) Cunha, J. T.; Soares, P. O.; Romaní, A.; Thevelein, J. M.; Domingues, L. Xylose Fermentation Efficiency of Industrial Saccharomyces Cerevisiae Yeast with Separate or Combined Xylose Reductase/Xylitol Dehydrogenase and Xylose Isomerase Pathways. *Biotechnol. Biofuels* **2019**, *12*, 20. https://doi.org/10.1186/s13068-019-1360-8.

(48) Leschine, S.; Warnick, T. A. Systems and Methods for Producing Biofuels and Related Materials. US20100151546A1, June 17, 2010. https://patents.google.com/patent/US20100151546A1/en (accessed 2024-03-13).

(49) Gill, D. M. Bacterial Toxins: A Table of Lethal Amounts. *Microbiol. Rev.* **1982**, *46* (1), 86–94. https://doi.org/10.1128/mr.46.1.86-94.1982.

(50) Frevert, J. Pharmaceutical, Biological, and Clinical Properties of Botulinum Neurotoxin Type A Products. *Drugs RD* **2015**, *15* (1), 1–9. https://doi.org/10.1007/s40268-014-0077-1.

(51) *FDA approves Botox to treat chronic migraines*. https://doi.org/10.1177/0976500X20110303.

(52) Tortora, G. J.; Funke, B. R.; Case, C. L. *Microbiology: An Introduction, Global Edition*; Pearson Higher Ed, 2020.

(53) *XIAFLEX® (collagenase clostridium histolyticum)*. https://www.xiaflex.com/ (accessed 2024-03-18).

(54) *Official website | Collagenase SANTYL Ointment*. https://santyl.com/ (accessed 2024-03-18).

(55) Gold, M. H. Treatment of Bruising after Injection to Reduce Cellulite. *J. Cosmet. Dermatol.* **2023**, *22* (1), 55–63. https://doi.org/10.1111/jocd.15471.

(56) *Home*. Solar Foods. https://solarfoods.com/ (accessed 2024-03-18).

(57) *Superbrewed Food*. Superbrewed Food. https://www.superbrewedfood.com (accessed 2024-03-18).

(58) *IFPEN | Tableau de bord biocarburants 2020*. IFPEN. https://www.ifpenergiesnouvelles.fr/article/tableau-bord-biocarburants-2020 (accessed 2024-03-18).

(59) Xiu, S.; Shahbazi, A. Bio-Oil Production and Upgrading Research: A Review. *Renew. Sustain. Energy Rev.* **2012**, *16* (7), 4406–4414. https://doi.org/10.1016/j.rser.2012.04.028.

(60) Rocha, W.; Mesa-Pérez, J.; L.A.B, C. EMERGY EVALUATION OF BIO-OIL PRODUCTION USING SUGARCANE BIOMASS RESIDUES AT FAST PYROLYSIS PILOT PLANT IN BRAZIL Alonso-Pippo,; 2006.

(61) *Alternative Fuels Data Center: Biodiesel Blends*. https://afdc.energy.gov/fuels/biodiesel_blends.html (accessed 2024-03-18).

(62) *Transport biofuels – Renewable Energy Market Update 2021 – Analysis*. IEA. https://www.iea.org/reports/renewable-energy-market-update-2021/transport-biofuels (accessed 2024-03-18).

(63) *E-CO_2Met – Electricity & CO_2 to Methanol*. Fraunhofer Center for Chemical-Biotechnological Processes CBP. https://www.cbp.fraunhofer.de/en/reference-projects/e-co2met.html (accessed 2024-08-26).

(64) *About Methanol*. Methanol Institute. https://www.methanol.org/about-methanol/ (accessed 2024-03-15).

(65) Wang, C.; Li, Y.; Xu, C.; Badawy, T.; Sahu, A.; Jiang, C. Methanol as an Octane Booster for Gasoline Fuels. *Fuel* **2019**, *248*, 76–84. https://doi.org/10.1016/j.fuel.2019.02.128.

(66) Devi, A.; Bajar, S.; Sihag, P.; Sheikh, Z. U. D.; Singh, A.; Kaur, J.; Bishnoi, N. R.; Pant, D. A Panoramic View of Technological Landscape for Bioethanol Production from Various Generations of Feedstocks. *Bioengineered* **2024**, *14* (1), 81–112. https://doi.org/10.1080/21655979.2022.2095702.

(67) Balat, M.; Balat, H.; Öz, C. Progress in Bioethanol Processing. *Prog. Energy Combust. Sci.* **2008**, *34* (5), 551–573. https://doi.org/10.1016/j.pecs.2007.11.001.

(68) *Cold Start 101 - Ethanol Producer Magazine*. https://web.archive.org/web/20081024161137/https://ethanolproducer.com/article.jsp?article_id=3717 (accessed 2024-03-19).

(69) Ghadikolaei, M. A.; Wong, P. K.; Cheung, C. S.; Zhao, J.; Ning, Z.; Yung, K.-F.; Wong, H. C.; Gali, N. K. Why Is the World Not yet Ready to Use Alternative Fuel Vehicles? *Heliyon* **2021**, *7* (7), e07527. https://doi.org/10.1016/j.heliyon.2021.e07527.

(70) Lee, S. Y.; Park, J. H.; Jang, S. H.; Nielsen, L. K.; Kim, J.; Jung, K. S. Fermentative Butanol Production by Clostridia. *Biotechnol. Bioeng.* **2008**, *101* (2), 209–228. https://doi.org/10.1002/bit.22003.

(71) Ramey, D. Production of Butyric Acid and Butanol from Biomass DOE Grant Final Report, 2017.

(72) Dürre, P. Biobutanol: An Attractive Biofuel. *Biotechnol. J.* **2007**, *2* (12), 1525–1534. https://doi.org/10.1002/biot.200700168.

(73) Scully, S. M.; Orlygsson, J. Chapter 5 - Biological Production of Alcohols. In *Advanced Bioprocessing for Alternative Fuels, Biobased Chemicals, and Bioproducts*; Hosseini, M., Ed.; Woodhead Publishing Series in Energy; Woodhead Publishing, 2019; pp 83–108. https://doi.org/10.1016/B978-0-12-817941-3.00005-X.

(74) *Introduction to Biofuels | Let's Talk Science*. https://letstalkscience.ca/educational-resources/stem-in-context/biofuels-alternative-energy-source (accessed 2024-03-18).

(75) *L'épuisement des ressources | EDF FR*. https://www.edf.fr/groupe-edf/espaces-dedies/jeunes-enseignants/pour-les-jeunes/lenergie-de-a-a-z/le-developpement-durable/lepuisement-des-ressources (accessed 2024-03-18).

(76) Kim, S.; Dale, B. E. Global Potential Bioethanol Production from Wasted Crops and Crop Residues. *Biomass Bioenergy* **2004**, *26* (4), 361–375. https://doi.org/10.1016/j.biombioe.2003.08.002.

(77) Rosendal, P.; Nielsen, B. H.; Lange, N. K. Stability of Bacterial Alpha-Amylase in the Starch Liquefaction Process. *Starch - Stärke* **1979**, *31* (11), 368–372. https://doi.org/10.1002/star.19790311104.

(78) Antrim, R. L. Liquefaction of Granular-Starch Slurries Using Alpha-Amylase in the Presence of Carbonate Ion. US5180669A, January 19, 1993. https://patents.google.com/patent/US5180669A/en (accessed 2024-03-18).

(79) Klopfenstein, T. J.; Erickson, G. E.; Bremer, V. R. BOARD-INVITED REVIEW: Use of Distillers by-Products in the Beef Cattle Feeding Industry. *J. Anim. Sci.* **2008**, *86* (5), 1223–1231. https://doi.org/10.2527/jas.2007-0550.

(80) Lennartsson, P. R.; Erlandsson, P.; Taherzadeh, M. J. Integration of the First and Second Generation Bioethanol Processes and the Importance of By-Products. *Bioresour. Technol.* **2014**, *165*, 3–8. https://doi.org/10.1016/j.biortech.2014.01.127.

(81) *From first- to third-generation biofuels: Challenges of producing a commodity from a biomass of increasing complexity | Animal Frontiers | Oxford Academic*.

https://academic.oup.com/af/article/3/2/6/4638639?login=true (accessed 2024-08-26).

(82) Dahman, Y.; Syed, K.; Begum, S.; Roy, P.; Mohtasebi, B. 14 - Biofuels: Their Characteristics and Analysis. In *Biomass, Biopolymer-Based Materials, and Bioenergy*; Verma, D., Fortunati, E., Jain, S., Zhang, X., Eds.; Woodhead Publishing Series in Composites Science and Engineering; Woodhead Publishing, 2019; pp 277–325. https://doi.org/10.1016/B978-0-08-102426-3.00014-X.

(83) *Miscanthus (Miscanthus x giganteus) for Biofuel Production – Farm Energy*. https://farm-energy.extension.org/miscanthus-miscanthus-x-giganteus-for-biofuel-production/ (accessed 2024-08-26).

(84) Abdel-Hamid, A. M.; Solbiati, J. O.; Cann, I. K. O. Chapter One - Insights into Lignin Degradation and Its Potential Industrial Applications. In *Advances in Applied Microbiology*; Sariaslani, S., Gadd, G. M., Eds.; Academic Press, 2013; Vol. 82, pp 1–28. https://doi.org/10.1016/B978-0-12-407679-2.00001-6.

(85) Pereira, P. H. F.; Rosa, M. de F.; Cioffi, M. O. H.; Benini, K. C. C. de C.; Milanese, A. C.; Voorwald, H. J. C.; Mulinari, D. R. Vegetal Fibers in Polymeric Composites: A Review. *Polímeros* **2015**, *25*, 9–22. https://doi.org/10.1590/0104-1428.1722.

(86) Banner, A.; Toogood, H. S.; Scrutton, N. S. Consolidated Bioprocessing: Synthetic Biology Routes to Fuels and Fine Chemicals. *Microorganisms* **2021**, *9* (5), 1079. https://doi.org/10.3390/microorganisms9051079.

(87) Foyle, T.; Jennings, L.; Mulcahy, P. Compositional Analysis of Lignocellulosic Materials: Evaluation of Methods Used for Sugar Analysis of Waste Paper and Straw. *Bioresour. Technol.* **2007**, *98* (16), 3026–3036. https://doi.org/10.1016/j.biortech.2006.10.013.

(88) Bajpai, P. Structure of Lignocellulosic Biomass. In *Pretreatment of Lignocellulosic Biomass for Biofuel Production*; Bajpai, P., Ed.; Springer: Singapore, 2016; pp 7–12. https://doi.org/10.1007/978-981-10-0687-6_2.

(89) Fischer, A. J.; Maiyuran, S.; Yaver, D. S. Industrial Relevance of Trichoderma Reesei as an Enzyme Producer. In *Trichoderma reesei: Methods and Protocols*; Mach-Aigner, A. R., Martzy, R., Eds.; Methods in Molecular Biology; Springer US: New York, NY, 2021; pp 23–43. https://doi.org/10.1007/978-1-0716-1048-0_2.

(90) Dahnum, D.; Tasum, S. O.; Triwahyuni, E.; Nurdin, M.; Abimanyu, H. Comparison of SHF and SSF Processes Using Enzyme and Dry Yeast for Optimization of Bioethanol Production from Empty Fruit Bunch. *Energy Procedia* **2015**, *68*, 107–116. https://doi.org/10.1016/j.egypro.2015.03.238.

(91) Tomás-Pejó, E.; Oliva, J. M.; Ballesteros, M.; Olsson, L. Comparison of SHF and SSF Processes from Steam-Exploded Wheat Straw for Ethanol Production by Xylose-Fermenting and Robust Glucose-Fermenting Saccharomyces Cerevisiae Strains. *Biotechnol. Bioeng.* **2008**, *100* (6), 1122–1131. https://doi.org/10.1002/bit.21849.

(92) Thygesen, A.; Vahlgren, L.; Frederiksen, J. H.; Linnane, W.; Thomsen, M. H.; Thygesen, A.; Vahlgren, L.; Frederiksen, J. H.; Linnane, W.; Thomsen, M. H. SSF Fermentation of Rape Straw and the Effects of Inhibitory Stress on Yeast. In *Bioethanol*; IntechOpen, 2012. https://doi.org/10.5772/24445.

(93) Chen, H.; Wang, L. Chapter 7 - Microbial Fermentation Strategies for Biomass Conversion. In *Technologies for Biochemical Conversion of Biomass*; Chen, H., Wang, L., Eds.; Academic Press: Oxford, 2017; pp 165–196. https://doi.org/10.1016/B978-0-12-802417-1.00007-7.

(94) Xu, Q.; Singh, A.; Himmel, M. E. Perspectives and New Directions for the Production of Bioethanol Using Consolidated Bioprocessing of Lignocellulose. *Curr. Opin. Biotechnol.* **2009**, *20* (3), 364–371. https://doi.org/10.1016/j.copbio.2009.05.006.

(95) Blanchard, J.; Leschine, S.; Petit, E.; Fabel, J. Methods and Compositions for Improving the Production of Fuels in Microorganisms. US20090286294A1, November 19, 2009. https://patents.google.com/patent/US20090286294A1/en (accessed 2024-03-18).

(96) Doughman, S. D.; Krupanidhi, S.; Sanjeevi, C. B. Omega-3 Fatty Acids for Nutrition and Medicine: Considering Microalgae Oil as a Vegetarian Source of EPA and DHA. *Curr. Diabetes Rev.* **2007**, *3* (3), 198–203. https://doi.org/10.2174/157339907781368968.

(97) Ferrell, J.; Sarisky-Reed, V. *National Algal Biofuels Technology Roadmap*; DOE/EE-0332; EERE Publication and Product Library, Washington, D.C. (United States), 2010. https://doi.org/10.2172/1218560.

(98) Sheehan, J.; Dunahay, T.; Benemann, J.; Roessler, P. *Look Back at the U.S. Department of Energy's Aquatic Species Program: Biodiesel from Algae; Close-Out Report*; NREL/TP-580-24190; National Renewable Energy Lab. (NREL), Golden, CO (United States), 1998. https://doi.org/10.2172/15003040.

(99) A Promising Oil Alternative: Algae Energy. January 6, 2008. http://www.washingtonpost.com/wp-dyn/content/article/2008/01/03/AR2008010303907.html (accessed 2024-03-18).

(100) Brennan, L.; Owende, P. Biofuels from Microalgae—A Review of Technologies for Production, Processing, and Extractions of Biofuels and Co-Products. *Renew. Sustain. Energy Rev.* **2010**, *14* (2), 557–577. https://doi.org/10.1016/j.rser.2009.10.009.

(101) Mata, T. M.; Martins, A. A.; Caetano, Nidia. S. Microalgae for Biodiesel Production and Other Applications: A Review. *Renew. Sustain. Energy Rev.* **2010**, *14* (1), 217–232. https://doi.org/10.1016/j.rser.2009.07.020.

(102) *BIODIESEL FROM ALGAE*. Biodiesel from Algae. http://biodieselfromalgae2.weebly.com/biodiesel-from-algae.html (accessed 2024-03-18).

(103) Chen, C.-Y.; Kao, A.-L.; Tsai, Z.-C.; Chow, T.-J.; Chang, H.-Y.; Zhao, X.-Q.; Chen, P.-T.; Su, H.-Y.; Chang, J.-S. Expression of Type 2 Diacylglycerol Acyltransferse Gene DGTT1 from Chlamydomonas Reinhardtii Enhances Lipid Production in Scenedesmus Obliquus. *Biotechnol. J.* **2016**, *11* (3), 336–344. https://doi.org/10.1002/biot.201500272.

(104) Ahmad, I.; Sharma, A. K.; Daniell, H.; Kumar, S. Altered Lipid Composition and Enhanced Lipid Production in Green Microalga by Introduction of Brassica Diacylglycerol Acyltransferase 2. *Plant Biotechnol. J.* **2015**, *13* (4), 540–550. https://doi.org/10.1111/pbi.12278.

(105) Yang, J.; Pan, Y.; Bowler, C.; Zhang, L.; Hu, H. Knockdown of Phosphoenolpyruvate Carboxykinase Increases Carbon Flux to Lipid Synthesis in *Phaeodactylum Tricornutum*. *Algal Res.* **2016**, *15*, 50–58. https://doi.org/10.1016/j.algal.2016.02.004.

(106) *Meet The Microbes Powering A Circular Carbon Economy – LanzaTech*. https://lanzatech.com/meet-the-microbes-powering-a-circular-carbon-economy/ (accessed 2024-03-18).

(107) Mavrommati, M.; Daskalaki, A.; Papanikolaou, S.; Aggelis, G. Adaptive Laboratory Evolution Principles and Applications in Industrial Biotechnology. *Biotechnol. Adv.* **2022**, *54*, 107795. https://doi.org/10.1016/j.biotechadv.2021.107795.

(108) Couce, A.; Limdi, A.; Magnan, M.; Owen, S. V.; Herren, C. M.; Lenski, R. E.; Tenaillon, O.; Baym, M. Changing Fitness Effects of Mutations through Long-Term Bacterial Evolution. *Science* **2024**, *383* (6681), eadd1417. https://doi.org/10.1126/science.add1417.

(109) Mutzel, R.; Marliere, P. Method and Device for Selecting Accelerated Proliferation of Living Cells in Suspension. US6686194B1, February 3, 2004. https://patents.google.com/patent/US6686194B1/en (accessed 2024-05-22).

(110) Cerisy, T.; Souterre, T.; Torres-Romero, I.; Boutard, M.; Dubois, I.; Patrouix, J.; Labadie, K.; Berrabah, W.; Salanoubat, M.; Doring, V.; Tolonen, A. C. Evolution of a Biomass-Fermenting Bacterium To Resist Lignin Phenolics. *Appl. Environ. Microbiol.* **2017**, *83* (11), e00289-17. https://doi.org/10.1128/AEM.00289-17.

(111) Carbonell, P.; Jervis, A. J.; Robinson, C. J.; Yan, C.; Dunstan, M.; Swainston, N.; Vinaixa, M.; Hollywood, K. A.; Currin, A.; Rattray, N. J. W.; Taylor, S.; Spiess, R.; Sung, R.; Williams, A. R.; Fellows, D.; Stanford, N. J.; Mulherin, P.; Le Feuvre, R.; Barran, P.; Goodacre, R.; Turner, N. J.; Goble, C.; Chen, G. G.; Kell, D. B.; Micklefield, J.; Breitling, R.; Takano, E.; Faulon, J.-L.; Scrutton, N. S. An Automated Design-Build-Test-Learn Pipeline for Enhanced Microbial Production of Fine Chemicals. *Commun. Biol.* **2018**, *1* (1), 1–10. https://doi.org/10.1038/s42003-018-0076-9.

(112) Waldby, C.; Gray, P.; Griffiths, P.; Vickers, C.; Meek, S.; Small, I. *Synthetic Biology in Australia: An Outlook to 2030. Australian Council of Learned Academies Expert Working Group Horizon Scanning Project.*; 2018.

(113) Pasteur, L. Quelques résultats nouveaux relatifs aux fermentations acétique et butyrique., 1862.

(114) Dürre, P. Fermentative Butanol Production: Bulk Chemical and Biofuel. *Ann. N. Y. Acad. Sci.* **2008**, *1125*, 353–362. https://doi.org/10.1196/annals.1419.009.

(115) Visioli, L. J.; Enzweiler, H.; Kuhn, R. C.; Schwaab, M.; Mazutti, M. A. Recent Advances on Biobutanol Production. *Sustain. Chem. Process.* **2014**, *2* (1), 15. https://doi.org/10.1186/2043-7129-2-15.

(116) Pfromm, P. H.; Amanor-Boadu, V.; Nelson, R.; Vadlani, P.; Madl, R. Bio-Butanol vs. Bio-Ethanol: A Technical and Economic Assessment for Corn and Switchgrass Fermented by Yeast or *Clostridium Acetobutylicum*. *Biomass Bioenergy* **2010**, *34* (4), 515–524. https://doi.org/10.1016/j.biombioe.2009.12.017.

(117) Young, M.; Minton, N. P.; Staudenbauer, W. L. Recent Advances in the Genetics of the Clostridia. *FEMS Microbiol. Rev.* **1989**, *5* (4), 301–325. https://doi.org/10.1111/j.1574-6968.1989.tb03402.x.

(118) Gaida, S. M.; Liedtke, A.; Jentges, A. H. W.; Engels, B.; Jennewein, S. Metabolic Engineering of Clostridium Cellulolyticum for the Production of N-Butanol from Crystalline Cellulose. *Microb. Cell Factories* **2016**, *15* (1), 6. https://doi.org/10.1186/s12934-015-0406-2.

(119) Sleat, R.; Mah, R. A.; Robinson, R. Isolation and Characterization of an Anaerobic, Cellulolytic Bacterium, Clostridium Cellulovorans Sp. Nov. *Appl. Environ. Microbiol.* **1984**, *48* (1), 88–93. https://doi.org/10.1128/aem.48.1.88-93.1984.

(120) Bao, T.; Hou, W.; Wu, X.; Lu, L.; Zhang, X.; Yang, S.-T. Engineering Clostridium Cellulovorans for Highly Selective N-Butanol Production from Cellulose in Consolidated Bioprocessing. *Biotechnol. Bioeng.* **2021**, *118* (7), 2703–2718. https://doi.org/10.1002/bit.27789.

(121) Yang, X.; Xu, M.; Yang, S.-T. Metabolic and Process Engineering of *Clostridium Cellulovorans* for Biofuel Production from Cellulose. *Metab. Eng.* **2015**, *32*, 39–48. https://doi.org/10.1016/j.ymben.2015.09.001.

(122) Heap, J. T.; Pennington, O. J.; Cartman, S. T.; Carter, G. P.; Minton, N. P. The ClosTron: A Universal Gene Knock-out System for the Genus Clostridium. *J. Microbiol. Methods* **2007**, *70* (3), 452–464. https://doi.org/10.1016/j.mimet.2007.05.021.

(123) Kovács, K.; Willson, B. J.; Schwarz, K.; Heap, J. T.; Jackson, A.; Bolam, D. N.; Winzer, K.; Minton, N. P. Secretion and Assembly of Functional Mini-Cellulosomes from Synthetic Chromosomal Operons in Clostridium Acetobutylicum ATCC 824. *Biotechnol. Biofuels* **2013**, *6* (1), 117. https://doi.org/10.1186/1754-6834-6-117.

(124) Mingardon, F.; Perret, S.; Bélaïch, A.; Tardif, C.; Bélaïch, J.-P.; Fierobe, H.-P. Heterologous Production, Assembly, and Secretion of a Minicellulosome by Clostridium Acetobutylicum ATCC 824. *Appl. Environ. Microbiol.* **2005**, *71* (3), 1215–1222. https://doi.org/10.1128/AEM.71.3.1215-1222.2005.

(125) Chanal, A.; Mingardon, F.; Bauzan, M.; Tardif, C.; Fierobe, H.-P. Scaffoldin Modules Serving as "Cargo" Domains to Promote the Secretion of Heterologous Cellulosomal Cellulases by Clostridium Acetobutylicum. *Appl. Environ. Microbiol.* **2011**, *77* (17), 6277–6280. https://doi.org/10.1128/AEM.00758-11.

(126) Wen, Z.; Wu, M.; Lin, Y.; Yang, L.; Lin, J.; Cen, P. Artificial Symbiosis for Acetone-Butanol-Ethanol (ABE) Fermentation from Alkali Extracted Deshelled Corn Cobs by Co-Culture of Clostridium Beijerinckii and Clostridium Cellulovorans. *Microb. Cell Factories* **2014**, *13* (1), 92. https://doi.org/10.1186/s12934-014-0092-5.

(127) Steen, E. J.; Chan, R.; Prasad, N.; Myers, S.; Petzold, C. J.; Redding, A.; Ouellet, M.; Keasling, J. D. Metabolic Engineering of Saccharomyces Cerevisiae for the Production of N-Butanol. *Microb. Cell Factories* **2008**, *7*, 36. https://doi.org/10.1186/1475-2859-7-36.

(128) Lan, E. I.; Liao, J. C. ATP Drives Direct Photosynthetic Production of 1-Butanol in Cyanobacteria. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (16), 6018–6023. https://doi.org/10.1073/pnas.1200074109.

(129) Nielsen, D. R.; Leonard, E.; Yoon, S.-H.; Tseng, H.-C.; Yuan, C.; Prather, K. L. J. Engineering Alternative Butanol Production Platforms in Heterologous Bacteria. *Metab. Eng.* **2009**, *11* (4), 262–273. https://doi.org/10.1016/j.ymben.2009.05.003.

(130) Ferreira, S.; Pereira, R.; Wahl, S. A.; Rocha, I. Metabolic Engineering Strategies for Butanol Production in Escherichia Coli. *Biotechnol. Bioeng.* **2020**, *117* (8), 2571–2587. https://doi.org/10.1002/bit.27377.

(131) Pyne, M. E.; Bruder, M.; Moo-Young, M.; Chung, D. A.; Chou, C. P. Technical Guide for Genetic Advancement of Underdeveloped and Intractable Clostridium. *Biotechnol. Adv.* **2014**, *32* (3), 623–641. https://doi.org/10.1016/j.biotechadv.2014.04.003.

(132) Heap, J. T.; Ehsaan, M.; Cooksley, C. M.; Ng, Y.-K.; Cartman, S. T.; Winzer, K.; Minton, N. P. Integration of DNA into Bacterial Chromosomes from Plasmids without a Counter-Selection Marker. *Nucleic Acids Res.* **2012**, *40* (8), e59–e59. https://doi.org/10.1093/nar/gkr1321.

(133) Huang, H.; Chai, C.; Li, N.; Rowe, P.; Minton, N. P.; Yang, S.; Jiang, W.; Gu, Y. CRISPR/Cas9-Based Efficient Genome Editing in Clostridium Ljungdahlii, an Autotrophic Gas-Fermenting Bacterium. *ACS Synth. Biol.* **2016**, *5* (12), 1355–1361. https://doi.org/10.1021/acssynbio.6b00044.

(134) Bruder, M. R.; Pyne, M. E.; Moo-Young, M.; Chung, D. A.; Chou, C. P. Extending CRISPR-Cas9 Technology from Genome Editing to Transcriptional Engineering in the Genus Clostridium. *Appl. Environ. Microbiol.* **2016**, *82* (20), 6109–6119. https://doi.org/10.1128/AEM.02128-16.

(135) Wang, Y.; Zhang, Z.-T.; Seo, S.-O.; Choi, K.; Lu, T.; Jin, Y.-S.; Blaschek, H. P. Markerless Chromosomal Gene Deletion in Clostridium Beijerinckii Using CRISPR/Cas9 System. *J. Biotechnol.* **2015**, *200*, 1–5. https://doi.org/10.1016/j.jbiotec.2015.02.005.

(136) Xu, T.; Li, Y.; Shi, Z.; Hemme, C. L.; Li, Y.; Zhu, Y.; Van Nostrand, J. D.; He, Z.; Zhou, J. Efficient Genome Editing in Clostridium Cellulolyticum via CRISPR-Cas9 Nickase. *Appl. Environ. Microbiol.* **2015**, *81* (13), 4423–4431. https://doi.org/10.1128/AEM.00873-15.

(137) Swarts, D. C.; Jinek, M. Cas9 versus Cas12a/Cpf1: Structure–Function Comparisons and Implications for Genome Editing. *WIREs RNA* **2018**, *9* (5), e1481. https://doi.org/10.1002/wrna.1481.

(138) Deltcheva, E.; Chylinski, K.; Sharma, C. M.; Gonzales, K.; Chao, Y.; Pirzada, Z. A.; Eckert, M. R.; Vogel, J.; Charpentier, E. CRISPR RNA Maturation by Trans-Encoded Small RNA and Host Factor RNase III. *Nature* **2011**, *471* (7340), 602–607. https://doi.org/10.1038/nature09886.

(139) Anzalone, A. V.; Koblan, L. W.; Liu, D. R. Genome Editing with CRISPR-Cas Nucleases, Base Editors, Transposases and Prime Editors. *Nat. Biotechnol.* **2020**, *38* (7), 824–844. https://doi.org/10.1038/s41587-020-0561-9.

(140) Mali, P.; Yang, L.; Esvelt, K. M.; Aach, J.; Guell, M.; DiCarlo, J. E.; Norville, J. E.; Church, G. M. RNA-Guided Human Genome Engineering via Cas9. *Science* **2013**, *339* (6121), 823–826. https://doi.org/10.1126/science.1232033.

(141) McCarty, N. S.; Graham, A. E.; Studená, L.; Ledesma-Amaro, R. Multiplexed CRISPR Technologies for Gene Editing and Transcriptional Regulation. *Nat. Commun.* **2020**, *11* (1), 1281. https://doi.org/10.1038/s41467-020-15053-x.

(142) Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **2012**, *337* (6096), 816–821. https://doi.org/10.1126/science.1225829.

(143) Paul, B.; Montoya, G. CRISPR-Cas12a: Functional Overview and Applications. *Biomed. J.* **2020**, *43* (1), 8–17. https://doi.org/10.1016/j.bj.2019.10.005.

(144) Zetsche, B.; Heidenreich, M.; Mohanraju, P.; Fedorova, I.; Kneppers, J.; DeGennaro, E. M.; Winblad, N.; Choudhury, S. R.; Abudayyeh, O. O.; Gootenberg, J. S.; Wu, W. Y.; Scott, D. A.; Severinov, K.; van der Oost, J.; Zhang, F. Multiplex Gene Editing by CRISPR-Cpf1 through Autonomous Processing of a Single crRNA Array. *Nat. Biotechnol.* **2017**, *35* (1), 31–34. https://doi.org/10.1038/nbt.3737.

(145) Bandyopadhyay, A.; Kancharla, N.; Javalkote, V. S.; Dasgupta, S.; Brutnell, T. P. CRISPR-Cas12a (Cpf1): A Versatile Tool in the Plant Genome Editing Tool Box for Agricultural Advancement. *Front. Plant Sci.* **2020**, *11*. https://doi.org/10.3389/fpls.2020.584151.

(146) Hong, W.; Zhang, J.; Cui, G.; Wang, L.; Wang, Y. Multiplexed CRISPR-Cpf1-Mediated Genome Editing in Clostridium Difficile toward the Understanding of Pathogenesis of C. Difficile Infection. *ACS Synth. Biol.* **2018**, *7* (6), 1588–1600. https://doi.org/10.1021/acssynbio.8b00087.

(147) Zhang, Y.; Kubiak, A. M.; Bailey, T.; Claessen, L.; Hittmeyer, P.; Dubois, L.; Theys, J.; Lambin, P. Development of a CRISPR-Cas12a System for Efficient Genome Engineering in Clostridia. *Microbiol. Spectr.* **2023**, *11*. https://doi.org/10.1128/spectrum.02459-23.

(148) Meliawati, M.; Schilling, C.; Schmid, J. Recent Advances of Cas12a Applications in Bacteria. *Appl. Microbiol. Biotechnol.* **2021**, *105* (8), 2981–2990. https://doi.org/10.1007/s00253-021-11243-9.

(149) Qi, L. S.; Larson, M. H.; Gilbert, L. A.; Doudna, J. A.; Weissman, J. S.; Arkin, A. P.; Lim, W. A. Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **2013**, *152* (5), 1173–1183. https://doi.org/10.1016/j.cell.2013.02.022.

(150) McAllister, K. N.; Sorg, J. A. CRISPR Genome Editing Systems in the Genus Clostridium: A Timely Advancement. *J. Bacteriol.* **2019**, *201* (16), e00219-19. https://doi.org/10.1128/JB.00219-19.

(151) Zhang, R.; Xu, W.; Shao, S.; Wang, Q. Gene Silencing Through CRISPR Interference in Bacteria: Current Advances and Future Prospects. *Front. Microbiol.* **2021**, *12*, 635227. https://doi.org/10.3389/fmicb.2021.635227.

(152) Komor, A. C.; Kim, Y. B.; Packer, M. S.; Zuris, J. A.; Liu, D. R. Programmable Editing of a Target Base in Genomic DNA without Double-Stranded DNA Cleavage. *Nature* **2016**, *533* (7603), 420–424. https://doi.org/10.1038/nature17946.

(153) Wang, X.; Liu, Z.; Li, G.; Dang, L.; Huang, S.; He, L.; Ma, Y.; Li, C.; Liu, M.; Yang, G.; Huang, X.; Zhou, F.; Ma, X. Efficient Gene Silencing by Adenine Base Editor-Mediated Start Codon Mutation. *Mol. Ther.* **2020**, *28* (2), 431–440. https://doi.org/10.1016/j.ymthe.2019.11.022.

(154) Kuscu, C.; Parlak, M.; Tufan, T.; Yang, J.; Szlachta, K.; Wei, X.; Mammadov, R.; Adli, M. CRISPR-STOP: Gene Silencing through Base-Editing-Induced Nonsense Mutations. *Nat. Methods* **2017**, *14* (7), 710–712. https://doi.org/10.1038/nmeth.4327.

(155) Tian, K.; Hong, X.; Guo, M.; Li, Y.; Wu, H.; Caiyin, Q.; Qiao, J. Development of Base Editors for Simultaneously Editing Multiple Loci in Lactococcus Lactis. *ACS Synth. Biol.* **2022**, *11* (11), 3644–3656. https://doi.org/10.1021/acssynbio.1c00561.

(156) Li, Q.; Seys, F. M.; Minton, N. P.; Yang, J.; Jiang, Y.; Jiang, W.; Yang, S. CRISPR–Cas9D10A Nickase-Assisted Base Editing in the Solvent Producer Clostridium Beijerinckii. *Biotechnol. Bioeng.* **2019**, *116* (6), 1475–1483. https://doi.org/10.1002/bit.26949.

(157) Xia, P.-F.; Casini, I.; Schulz, S.; Klask, C.-M.; Angenent, L. T.; Molitor, B. Reprogramming Acetogenic Bacteria with CRISPR-Targeted Base Editing via Deamination. *ACS Synth. Biol.* **2020**, *9* (8), 2162–2171. https://doi.org/10.1021/acssynbio.0c00226.

(158) Green, M. R.; Sambrook, J. Isolation and Quantification of DNA. *Cold Spring Harb. Protoc.* **2018**, *2018* (6). https://doi.org/10.1101/pdb.top093336.

(159) Umenhoffer, K.; Fehér, T.; Balikó, G.; Ayaydin, F.; Pósfai, J.; Blattner, F. R.; Pósfai, G. Reduced Evolvability of Escherichia Coli MDS42, an IS-Less Cellular Chassis for Molecular and Synthetic Biology Applications. *Microb. Cell Factories* **2010**, *9* (1), 38. https://doi.org/10.1186/1475-2859-9-38.

(160) Bernard, P.; Couturier, M. Cell Killing by the F Plasmid CcdB Protein Involves Poisoning of DNA-Topoisomerase II Complexes. *J. Mol. Biol.* **1992**, *226* (3), 735–745. https://doi.org/10.1016/0022-2836(92)90629-X.

(161) Bertani, G. Studies on Lysogenesis. I. The Mode of Phage Liberation by Lysogenic Escherichia Coli. *J. Bacteriol.* **1951**, *62* (3), 293–300. https://doi.org/10.1128/jb.62.3.293-300.1951.

(162) Hanahan, D. Studies on Transformation of Escherichia Coli with Plasmids. *J. Mol. Biol.* **1983**, *166* (4). https://doi.org/10.1016/s0022-2836(83)80284-8.

(163) Bahls, M. O.; Platz, L.; Morgado, G.; Schmidt, G. W.; Panke, S. Directed Evolution of Biofuel-Responsive Biosensors for Automated Optimization of Branched-Chain Alcohol Biosynthesis. *Metab. Eng.* **2022**, *69*, 98–111. https://doi.org/10.1016/j.ymben.2021.10.014.

(164) Vialetto, E.; Miele, S.; Goren, M. G.; Yu, J.; Yu, Y.; Collias, D.; Beamud, B.; Osbelt, L.; Lourenço, M.; Strowig, T.; Brisse, S.; Barquist, L.; Qimron, U.; Bikard, D.; Beisel, C. L. Systematic Interrogation of CRISPR Antimicrobials in Klebsiella Pneumoniae Reveals Nuclease-, Guide- and Strain-Dependent Features Influencing Antimicrobial Activity. *Nucleic Acids Res.* **2024**, *52* (10), 6079–6091. https://doi.org/10.1093/nar/gkae281.

(165) Rostain, W.; Zaplana, T.; Boutard, M.; Baum, C.; Tabuteau, S.; Sanitha, M.; Ramya, M.; Guss, A.; Ettwiller, L.; Tolonen, A. C. Tuning of Gene Expression in Clostridium Phytofermentans Using Synthetic Promoters and CRISPRi. *ACS Synth. Biol.* **2022**, *11* (12), 4077–4088. https://doi.org/10.1021/acssynbio.2c00385.

(166) Iverson, S. V.; Haddock, T. L.; Beal, J.; Densmore, D. M. CIDAR MoClo: Improved MoClo Assembly Standard and New E. Coli Part Library Enable Rapid Combinatorial Design for Synthetic and Traditional Biology. *ACS Synth. Biol.* **2016**, *5* (1), 99–103. https://doi.org/10.1021/acssynbio.5b00124.

(167) Sambrook, J.; Russell, D. W. Preparation and Transformation of Competent E. Coli Using Calcium Chloride. *CSH Protoc.* **2006**, *2006* (1), pdb.prot3932. https://doi.org/10.1101/pdb.prot3932.

(168) Jones, D. T.; Woods, D. R. Acetone-Butanol Fermentation Revisited. *Microbiol. Rev.* **1986**, *50* (4), 484–524. https://doi.org/10.1128/mr.50.4.484-524.1986.

(169) Lynd, L. R.; van Zyl, W. H.; McBride, J. E.; Laser, M. Consolidated Bioprocessing of Cellulosic Biomass: An Update. *Curr. Opin. Biotechnol.* **2005**, *16* (5), 577–583. https://doi.org/10.1016/j.copbio.2005.08.009.

(170) Mutzel, R.; Marliere, P. Method and Device for Selecting Accelerated Proliferation of Living Cells in Suspension. US6686194B1, February 3, 2004. https://patents.google.com/patent/US6686194B1/en (accessed 2024-08-26).

(171) Fukui, K.; Shimada, A.; Iino, H.; Masui, R.; Kuramitsu, S.; Fukui, K.; Shimada, A.; Iino, H.; Masui, R.; Kuramitsu, S. Biochemical Properties of MutL, a DNA Mismatch Repair Endonuclease. In *DNA Repair - On the Pathways to Fixing DNA Damage and Errors*; IntechOpen, 2011. https://doi.org/10.5772/23758.

(172) Han, X.-P.; Yang, X.-W.; Liu, J. MutS and MutL Sliding Clamps in DNA Mismatch Repair. *Genome Instab. Dis.* **2023**, *4* (1), 1–11. https://doi.org/10.1007/s42764-022-00094-x.

(173) Zhou, H.; Zhang, L.; Xu, Q.; Zhang, L.; Yu, Y.; Hua, X. The Mismatch Repair System (mutS and mutL) in Acinetobacter Baylyi ADP1. *BMC Microbiol.* **2020**, *20* (1), 40. https://doi.org/10.1186/s12866-020-01729-3.

(174) Gogou, C.; Japaridze, A.; Dekker, C. Mechanisms for Chromosome Segregation in Bacteria. *Front. Microbiol.* **2021**, *12*, 685687. https://doi.org/10.3389/fmicb.2021.685687.

(175) Lorca, G. L.; Chung, Y. J.; Barabote, R. D.; Weyler, W.; Schilling, C. H.; Saier, M. H. Catabolite Repression and Activation in Bacillus Subtilis: Dependency on CcpA, HPr, and HprK. *J. Bacteriol.* **2005**, *187* (22), 7826–7839. https://doi.org/10.1128/JB.187.22.7826-7839.2005.

(176) Peabody, G. L.; Kao, K. C. Recent Progress in Biobutanol Tolerance in Microbial Systems with an Emphasis on Clostridium. *FEMS Microbiol. Lett.* **2016**, *363* (5), fnw017. https://doi.org/10.1093/femsle/fnw017.

(177) Řezanka, T.; Křesinová, Z.; Kolouchová, I.; Sigler, K. Lipidomic Analysis of Bacterial Plasmalogens. *Folia Microbiol. (Praha)* **2012**, *57* (5), 463–472. https://doi.org/10.1007/s12223-012-0178-6.

(178) Kang, M.; Kim, K.; Choe, D.; Cho, S.; Kim, S. C.; Palsson, B.; Cho, B.-K. Inactivation of a Mismatch-Repair System Diversifies Genotypic Landscape of Escherichia Coli During Adaptive Laboratory Evolution. *Front. Microbiol.* **2019**, *10*. https://doi.org/10.3389/fmicb.2019.01845.

(179) Ingelman, H.; Heffernan, J. K.; Harris, A.; Brown, S. D.; Shaikh, K. M.; Saqib, A. Y.; Pinheiro, M. J.; de Lima, L. A.; Martinez, K. R.; Gonzalez-Garcia, R. A.; Hawkins, G.; Daleiden, J.; Tran, L.; Zeleznik, H.; Jensen, R. O.; Reynoso, V.; Schindel, H.; Jänes, J.; Simpson, S. D.; Köpke, M.; Marcellin, E.; Valgepea, K. Autotrophic Adaptive

Laboratory Evolution of the Acetogen *Clostridium Autoethanogenum* Delivers the Gas-Fermenting Strain LAbrini with Superior Growth, Products, and Robustness. *New Biotechnol.* **2024**, *83*, 1–15. https://doi.org/10.1016/j.nbt.2024.06.002.

(180) Lynd, L. R.; Laser, M. S.; Bransby, D.; Dale, B. E.; Davison, B.; Hamilton, R.; Himmel, M.; Keller, M.; McMillan, J. D.; Sheehan, J.; Wyman, C. E. How Biotech Can Transform Biofuels. *Nat. Biotechnol.* **2008**, *26* (2), 169–172. https://doi.org/10.1038/nbt0208-169.

(181) Papoutsakis, E. T. Engineering Solventogenic Clostridia. *Curr. Opin. Biotechnol.* **2008**, *19* (5), 420–429. https://doi.org/10.1016/j.copbio.2008.08.003.

(182) Nielsen, J.; Keasling, J. D. Engineering Cellular Metabolism. *Cell* **2016**, *164* (6), 1185–1197. https://doi.org/10.1016/j.cell.2016.02.004.

(183) Vögeli, B.; Schulz, L.; Garg, S.; Tarasava, K.; Clomburg, J. M.; Lee, S. H.; Gonnot, A.; Moully, E. H.; Kimmel, B. R.; Tran, L.; Zeleznik, H.; Brown, S. D.; Simpson, S. D.; Mrksich, M.; Karim, A. S.; Gonzalez, R.; Köpke, M.; Jewett, M. C. Cell-Free Prototyping Enables Implementation of Optimized Reverse β-Oxidation Pathways in Heterotrophic and Autotrophic Bacteria. *Nat. Commun.* **2022**, *13* (1), 3058. https://doi.org/10.1038/s41467-022-30571-6.

(184) *Internal transcription termination widely regulates differential expression of operon-organized genes including ribosomal protein and RNA polymerase genes in an archaeon | Nucleic Acids Research | Oxford Academic*. https://academic.oup.com/nar/article/51/15/7851/7223592?login=true (accessed 2024-07-29).

(185) Reis, A. C.; Salis, H. M. An Automated Model Test System for Systematic Development and Improvement of Gene Expression Models. *ACS Synth. Biol.* **2020**, *9* (11), 3145–3156. https://doi.org/10.1021/acssynbio.0c00394.

(186) Cerisy, T.; Rostain, W.; Chhun, A.; Boutard, M.; Salanoubat, M.; Tolonen, A. C. A Targetron-Recombinase System for Large-Scale Genome Engineering of Clostridia. *mSphere* **2019**, *4* (6), e00710-19. https://doi.org/10.1128/mSphere.00710-19.

(187) Jensen, S. I.; Lennen, R. M.; Herrgård, M. J.; Nielsen, A. T. Seven Gene Deletions in Seven Days: Fast Generation of Escherichia Coli Strains Tolerant to Acetate and Osmotic Stress. *Sci. Rep.* **2015**, *5* (1), 17874. https://doi.org/10.1038/srep17874.

(188) Choudhury, A.; Fenster, J. A.; Fankhauser, R. G.; Kaar, J. L.; Tenaillon, O.; Gill, R. T. CRISPR/Cas9 Recombineering-mediated Deep Mutational Scanning of Essential Genes in Escherichia Coli. *Mol. Syst. Biol.* **2020**, *16* (3), e9265. https://doi.org/10.15252/msb.20199265.

(189) Tak, Y. E.; Kleinstiver, B. P.; Nuñez, J. K.; Hsu, J. Y.; Horng, J. E.; Gong, J.; Weissman, J. S.; Joung, J. K. Inducible and Multiplex Gene Regulation Using CRISPR–Cpf1-Based Transcription Factors. *Nat. Methods* **2017**, *14* (12), 1163–1166. https://doi.org/10.1038/nmeth.4483.

(190) Gao, L.; Cox, D. B. T.; Yan, W. X.; Manteiga, J. C.; Schneider, M. W.; Yamano, T.; Nishimasu, H.; Nureki, O.; Crosetto, N.; Zhang, F. Engineered Cpf1 Variants with Altered PAM Specificities. *Nat. Biotechnol.* **2017**, *35* (8), 789–792. https://doi.org/10.1038/nbt.3900.

(191) Liao, C.; Slotkowski, R. A.; Achmedov, T.; Beisel, C. L. The Francisella Novicida Cas12a Is Sensitive to the Structure Downstream of the Terminal Repeat in CRISPR Arrays. *RNA Biol.* **2019**, *16* (4), 404–412. https://doi.org/10.1080/15476286.2018.1526537.

(192) Coppens, L.; Wicke, L.; Lavigne, R. SAPPHIRE.CNN: Implementation of dRNA-Seq-Driven, Species-Specific Promoter Prediction Using Convolutional Neural Networks. *Comput. Struct. Biotechnol. J.* **2022**, *20*, 4969–4974. https://doi.org/10.1016/j.csbj.2022.09.006.

(193) Boutard, M.; Ettwiller, L.; Cerisy, T.; Alberti, A.; Labadie, K.; Salanoubat, M.; Schildkraut, I.; Tolonen, A. C. Global Repositioning of Transcription Start Sites in a Plant-Fermenting Bacterium. *Nat. Commun.* **2016**, *7* (1), 13783. https://doi.org/10.1038/ncomms13783.

ANNEXES

Tableau 11 : Primers utilisé pour des PCR dans cette thèse

N°	Séquence	Objectif
oTZ003-seq-TetR-F	ggttgcgtattggaagatcaagag	Primer de séquençage et de PCR de vérification
oTZ109-VF	gtgccacctgacgtctaagaaac	
oTZ110-VR	gtattaccgcctttgagtgagctg	
oTZ111pCs17RFP-A-F	tatatagagctctaaaaaaattattgacaattttaaataaaaatgatataatacccgaacacaggagtgagacctcaatacgcaaaccgcctc	Ajout de pCons17 et changement des overhang en amplifiant la casette RFP ajout SacI et NheI pour Restriction ligation
oTZ114-RFP-F-R	gtcgctagcggaagaagcgggagaccatagttataaacgcagaaaggccca	
oTZ115-VR-pQmodMoclo	ctttctattcagcactgttatgccttttg	Primer de séquençage et de PCR de véirification
oTZ120-ccdB-F	tttgaagacatagatagagaccagcggatcctacctgacgct	Amplification de *ccdB*, ajout de sites BsaI pour plateforme de clonage de guide et BbsI pour l'insertion de ce fragment PCR
oTZ121-ccdB-R	aaagaagacataaactgagacctgaattgcgcgcaattaaccctc	
oTZ170-bsens1-F	tttcgtctcacgtgcgaaaaaggatggatataccgaaaaaatcg	Amplification de PalkB_sfGFP_KanR et ajout de Esp3I pour Golden Gate
oTZ171-bsens1-R	tttcgtctcaagggcggcggatttgtcctactcag	
oTZ172-bsens2-F	tttcgtctcaccctaatattcagcgatttgcccgagc	Amplification de PalkS_alkS_pSC101 et ajout de Esp3I pour Golden Gate
oTZ173-bsens2-R	tttcgtctcacacggttgggtaacgccagggttttc	
oTZ259-LacZ-Esp3I-F	tttcgtctcaagatatgtcttctgcaccatatgcgg	Modification des overhang Golden gate pour sous-cloner des guides CRISPR Cas12a et ajout de Esp3I pour Golden Gate
oTZ260-LacZ-Esp3I-R	tttcgtctcaaaacacgtcttccccgcgcgttg	
oTZ261-Bla-Esp3I-F	tttcgtctcagtttagagacctactagtagcggcc	
oTZ262-Bla-Esp3I-R	tttcgtctcaatcttgagacccactagtctctagaagc	

Tableau 12 : Primers utilisé pour le clonage de guides CRISPR-Cas12a dans cette thèse

N°	Séquence	Guide
oTZ282-1326-1-4spr-F	agatatataattacagaggcagattattaatttctacta	Spr_1326(2)-1327(2)
oTZ283-1326-1-4spr-R	cacttagtagaaattaataatctgcctctgtaattatat	
oTZ284-1326-2-4spr-F	agtgtagatcctgagagtatggacagaagaactaatttctac	
oTZ285-1326-2-4spr-R	cttagtagaaattagttcttctgtccatactctcaggatcta	
oTZ286-1327-1-4spr-F	taagtgtagatggatatattaagttgccccgtgataatttct	
oTZ287-1327-1-4spr-R	tagtagaaattatcacggggcaacttaatatatccatctaca	
oTZ288-1327-2-4spr-F	actaagtgtagatgttattaattgcggaagttcttc	
oTZ289-1327-2-4spr-R	aaacgaagaacttccgcaattaataacatctacact	
oTZ290-1232-1-4spr-F	agatttgatattgatagagaaaaagcgtaatttctacta	Spr_1232(2)-1117(2)
oTZ291-1232-1-4spr-R	cacttagtagaaattacgctttttctctatcaatatcaa	
oTZ292-1232-2-4spr-F	agtgtagatatggtatgattgttagcatttcataatttctac	
oTZ293-1232-2-4spr-R	cttagtagaaattatgaaatgctaacaatcataccatatcta	
oTZ294-1117-1-4spr-F	taagtgtagatcgttacgatagcatatattacattaatttct	
oTZ295-1117-1-4spr-R	tagtagaaattaatgtaatatatgctatcgtaacgatctaca	
oTZ296-1117-2-4spr-F	actaagtgtagatgttggtacttcaacggcgtttag	
oTZ297-1117-2-4spr-R	aaacctaaacgccgttgaagtaccaacatctacact	
oTZ298-1326-8spr-F	agatatataattacagaggcagattattaatttctactaagtgtagatcctgagagtatggacagaagaactaatttctacta	Spr_1326(2)-1327(2)-1232(2)-1117(2)
oTZ299-1326-8spr-R	cacttagtagaaattagttcttctgtccatactctcaggatctacacttagtagaaattaataatctgcctctgtaattatat	
oTZ300-1327-8spr-F	agtgtagatggatatattaagttgccccgtgataatttctactaagtgtagatgttattaattgcggaagttcttctaatttctac	
oTZ301-1327-8spr-R	acttagtagaaattagaagaacttccgcaattaataacatctacacttagtagaaattatcacggggcaacttaatatatccatcta	
oTZ302-1232-8spr-F	taagtgtagatttgatattgatagagaaaaagcgtaatttctactaagtgtagatatggtatgattgttagcatttcataatttct	
oTZ303-1232-8spr-R	tagtagaaattatgaaatgctaacaatcataccatatctacacttagtagaaattacgcttttctctatcaatatcaaatctac	
oTZ304-1117-8spr-F	actaagtgtagatcgttacgatagcatatattacattaatttctactaagtgtagatgttggtacttcaacggcgtttag	
oTZ305-1117-8spr-R	aaacctaaacgccgttgaagtaccaacatctacacttagtagaaattaatgtaatatatgctatcgtaacgatctacact	
oTZ306-3925-1-6spr-F	agatgatattaaactctgtaatacaaataatttctactaag	Spr_3925(2)-2463(2)-1029(2)
oTZ307-3925-1-6spr-R	tacacttagtagaaattatttgtattacagagtttaatatc	
oTZ308-3925-2-6spr-F	tgtagatgagagttaaaagaagcacaagagtaatttctacta	
oTZ309-3925-2-6spr-R	cacttagtagaaattactcttgtgcttcttttaactctcatc	
oTZ310-1029-1-6spr-F	agtgtagatacaaataatctacgatacatagataatttct	
oTZ311-1029-1-6spr-R	tagtagaaattatctatgtatcgtagattatttgtatcta	
oTZ312-1029-2-6spr-F	actaagtgtagattatcatggaaagggttctcttgctaattt	
oTZ313-1029-2-6spr-R	gtagaaattagcaagagaacccttccatgataatctacact	
oTZ314-2463-1-6spr-F	ctactaagtgtagatatggtagtcgctacgaaatcaggtaat	

oTZ315-2463-1-6spr-R	agaaattacctgatttcgtagcgactaccatatctacactta	
oTZ316-2463-2-6spr-F	ttctactaagtgtagatgtaaaaatcagatcgatcagcta	
oTZ317-2463-2-6spr-R	aaactagctgatcgatctgatttttacatctacacttagt	
oTZ318-1326-1-6spr-F	agatatataattacagaggcagattattaatttctactaag	
oTZ319-1326-1-6spr-R	tacacttagtagaaattaataatctgcctctgtaattatat	
oTZ320-1326-2-6spr-F	tgtagatcctgagagtatggacagaagaactaatttctacta	
oTZ321-1326-2-6spr-R	cacttagtagaaattagttcttctgtccatactctcaggatc	
oTZ322-1327-1-6spr-F	agtgtagatggatatattaagttgccccgtgataatttct	
oTZ323-1327-1-6spr-R	tagtagaaattatcacggggcaacttaatatatccatcta	Spr_1326(2)-1327(2)-0089(2)
oTZ324-1327-2-6spr-F	actaagtgtagatgttattaattgcggaagttcttctaattt	
oTZ325-1327-2-6spr-R	gtagaaattagaagaacttccgcaattaataacatctacact	
oTZ326-0089-1-6spr-F	ctactaagtgtagattaaaagtatggattgagaaaggataat	
oTZ327-0089-1-6spr-R	agaaattatcctttctcaatccatacttttaatctacactta	
oTZ328-0089-2-6spr-F	ttctactaagtgtagattgacggaactgcttgccatgtta	
oTZ329-0089-2-6spr-R	aaactaacatggcaagcagttccgtcaatctacacttagt	

Tableau 13 : Mutations intra-géniques de type SNP dans le génome de la souche PHY24.0#6

Les mutations intra-géniques sont détaillées avec leur type (SNP : non synonyme ou non-sens, DEL pour délétions, ou INS pour insertions). Les mutations synonymes ont été retirées, et les mutations intergéniques ne sont pas explorées dans ce tableau. Le score et le nombre de lectures (reads) validant la mutation sont exprimés respectivement comme une valeur comprise entre 0 et 1, et comme le ratio du nombre de lectures validant la mutation sur le nombre total de lectures couvrant cette mutation. Ce tableau est organisé en classifiant les gènes touchés par une ou plusieurs mutations en fonction de leur produit ou fonction prédite, à l'aide de la base de données EGGNOG (Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups).

Gènes	Mutation Intra-génique	Produits / Fonctions
PROCESSUS CELLULAIRES ET SIGNALISATION		
Contrôle du cycle cellulaire, division cellulaire, répartition des chromosomes		
Cphy_1697	T/A|SNP|tv|ATT/AAT|I89N|nonsyn|0.945|85/87	arsenite-activated ATPase ArsA
Cphy_1698	G/A|SNP|ts|AGC/AAC|S153N|nonsyn|0.976|105/105	arsenite-transporting ATPase ArsA
Motilité cellulaire		
Cphy_2702	-/T|INS|0.925|101/112	CheC, inhibitor of MCP methylation / FliN fusion protein
Biogenèse de la paroi cellulaire, de la membrane et de l'enveloppe		
Cphy_0905	A/T|SNP|tv|ACA/TCA|T178S|nonsyn|0.95|107/108	alanine racemase
Cphy_1155	C/T|SNP|ts|GCA/GTA|A605V|nonsyn|0.974|120/120	nucleic acid binding OB-fold tRNA/helicase-type
Cphy_1166	C/T|SNP|ts|CCA/TCA|P87S|nonsyn|0.979|99/100	UDP-glucose 4-epimerase
Cphy_2214	G/A|SNP|ts|GCA/GTA|A246V|nonsyn|0.957|108/109	nucleotide sugar dehydrogenase
Cphy_2372	C/T|SNP|ts|GAT/AAT|D397N|nonsyn|0.962|104/106	penicillin-binding protein transpeptidase
Cphy_2476	C/T|SNP|ts|GGA/GAA|G196E|nonsyn|0.958|86/86	peptidoglycan glycosyltransferase
Cphy_2607	G/A|SNP|ts|CGT/TGT|R152C|nonsyn|0.978|115/115	virulence factor MVIN family protein
Cphy_2811	T/C|SNP|ts|ATA/ATG|I631M|nonsyn|0.972|127/127	glycosyl transferase group 1
Cphy_3002	G/A|SNP|ts|GCT/GTT|A379V|nonsyn|0.966|150/150	cell wall hydrolase SleB
Cphy_3063	T/C|SNP|ts|GTG/GCG|V132A|nonsyn|0.981|164/164	peptidase M23B
Cphy_3109	G/A|SNP|ts|GCA/GTA|A77V|nonsyn|0.961|118/119	glycoside hydrolase family protein
Cphy_3230	G/A|SNP|ts|CGG/TGG|R456W|nonsyn|0.965|159/160	hypothetical protein
Cphy_3499	G/C|SNP|tv|CAG/GAG|Q386E|nonsyn|0.964|158/158	Polysaccharide Biosynthesis Protein
Cphy_3507	A/G|SNP|ts|GTA/GCA|V156A|nonsyn|0.947|169/169	Sugar transferase
Cphy_3607	-/T|INS|0.913|150/169	peptidase M23B
Mécanismes de défense		
Cphy_0018	C/T|SNP|ts|CGA/CAA|R48Q|nonsyn|0.981|211/211	ABC transporter related
Cphy_0115	G/-|DEL|0.98|181/188	ABC-2 type transporter
Cphy_0942	A/-|DEL|0.969|137/146	ABC transporter related
Cphy_1237	-/A|INS|0.96|98/105	ABC transporter related
Cphy_1681	G/A|SNP|ts|TGC/TAC|C335Y|nonsyn|0.988|107/107	ABC transporter related
Cphy_1682	C/G|SNP|tv|ACT/AGT|T181S|nonsyn|0.994|98/98	ABC transporter related
Cphy_2404	C/T|SNP|ts|GGA/GAA|G89E|nonsyn|0.898|92/93	MATE efflux family protein
Cphy_3380	G/A|SNP|ts|CAG/TAG|Q327*|nonsense|0.978|163/163	ABC transporter transmembrane region
Cphy_3598	G/A|SNP|ts|CAT/TAT|H112Y|nonsyn|0.963|167/167	ABC transporter related
Cphy_3781	A/G|SNP|ts|CAC/CGC|H185R|nonsyn|0.977|181/182	ABC transporter related
Mécanismes de transduction de signaux		

Cphy_0091	C/T|SNP|ts|GAC/AAC|D220N|nonsyn|0.961|159/160	protein serine/threonine phosphatase
Cphy_0173	-/A|INS|0.923|163/186	diguanylate cyclase
Cphy_0240	-/G|INS|0.922|203/223	diguanylate cyclase/phosphodiesterase
Cphy_0358	A/-|DEL|0.956|149/159	two component transcriptional regulator
Cphy_0525	-/T|INS|0.949|177/193	two component AraC family transcriptional regulator
Cphy_0928	C/T|SNP|ts|GCT/ACT|A471T|nonsyn|0.957|122/122	AraC family transcriptional regulator
Cphy_1136	C/T|SNP|ts|CGA/TGA|R387*|nonsense|0.962|112/112	diguanylate cyclase/phosphodiesterase with PAS/PAC sensor(s)
Cphy_1393	G/A|SNP|ts|GGT/AGT|G525S|nonsyn|0.969|100/100	histidine kinase internal region
Cphy_1583	C/T|SNP|ts|ACG/ATG|T30M|nonsyn|0.974|100/100	two component AraC family transcriptional regulator
Cphy_1676	G/A|SNP|ts|GCA/ACA|A113T|nonsyn|0.965|90/91 G/A|SNP|ts|TGG/TAG|W201*|nonsense|0.985|104/104	two component transcriptional regulator
Cphy_1677	G/A|SNP|ts|GCG/ACG|A337T|nonsyn|0.948|93/93	integral membrane sensor signal transduction histidine kinase
Cphy_2187	C/T|SNP|ts|GAA/AAA|E220K|nonsyn|0.965|74/75	AraC family transcriptional regulator
Cphy_2827	G/A|SNP|ts|TCA/TTA|S118L|nonsyn|0.984|120/122	Two component transcriptional regulator (Winged helix family
Cphy_2829	C/T|SNP|ts|GCA/ACA|A368T|nonsyn|0.972|115/115	integral membrane sensor signal transduction histidine kinase
Cphy_3518	C/T|SNP|ts|GTC/ATC|V52I|nonsyn|0.969|178/178	Two component transcriptional regulator (Winged helix family
Cphy_3937	G/A|SNP|ts|CCA/CTA|P222L|nonsyn|0.973|226/226	Histidine kinase
Trafic intracellulaire, sécrétion et transport vésiculaire		
Cphy_0402	C/T|SNP|ts|CAG/TAG|Q267*|nonsense|0.977|158/158	Sec-independent protein translocase, TatC subunit
Cphy_0837	-/T|INS|0.934|133/147	AAA ATPase
Cphy_2402	T/C|SNP|ts|TAC/TGC|Y97C|nonsyn|0.992|88/88	Signal peptidase i
Cphy_3647	C/T|SNP|ts|GCG/ACG|A275T|nonsyn|0.992|174/174	preprotein translocase, SecY subunit
Modification post-traductionnelle, renouvellement des protéines, chaperons		
Cphy_0263	C/T|SNP|ts|CCA/CTA|P159L|nonsyn|0.99|153/153	von Willebrand factor type A
Cphy_0326	G/A|SNP|ts|GGA/GAA|G220E|nonsyn|0.973|154/154	ATP-dependent metalloprotease FtsH
Cphy_1418	C/T|SNP|ts|CCA/CTA|P105L|nonsyn|0.976|126/126	glycyl-radical activating family protein
Cphy_1485	A/G|SNP|ts|AAA/GAA|K27E|nonsyn|0.981|107/107	Xanthine dehydrogenase accessory factor
Cphy_1572	C/T|SNP|ts|GCG/GTG|A190V|nonsyn|0.958|122/122	peptidase S8/S53 subtilisin kexin sedolisin
Cphy_2013	T/A|SNP|tv|ATA/AAA|I355K|nonsyn|0.969|100/100	proteinase inhibitor I4 serpin
Cphy_2317	C/T|SNP|ts|GGC/GAC|G171D|nonsyn|0.959|96/98	germination protease
Cphy_2662	G/A|SNP|ts|TCA/TTA|S126L|nonsyn|0.985|111/111	Glutathione peroxidase
Cphy_2817	G/A|SNP|ts|GCT/ACT|A364T|nonsyn|0.913|117/119	peptidase S8/S53 subtilisin kexin sedolisin
Cphy_3500	T/A|SNP|tv|AAA/AAT|K272N|nonsyn|0.978|172/172 G/A|SNP|ts|CAA/TAA|Q170*|nonsense|0.993|154/154	hemolytic protein HlpA-like protein
Cphy_3543	G/A|SNP|ts|GGC/AGC|G350S|nonsyn|0.97|136/136	2-alkenal reductase

Cphy_3819	G/T|SNP|tv|CCG/ACG|P117T|nonsyn|0.979|197/197	ATP-dependent chaperone ClpB
PROCESSUS ET STOCKAGE DE L'INFORMATION		
Réplication, recombinaison et réparation		
Cphy_0112	T/C|SNP|ts|TTA/TCA|L1012S|nonsyn|0.965|161/161	transcription-repair coupling factor
Cphy_0259	G/A|SNP|ts|GCT/ACT|A99T|nonsyn|0.967|178/180	excinuclease ABC subunit B
Cphy_0295	T/C|SNP|ts|ACA/GCA|T568A|nonsyn|0.958|164/165	DNA topoisomerase (ATP-hydrolyzing)
Cphy_0820	A/-|DEL|0.936|124/138 A/-|DEL|0.939|124/137 G/-|DEL|0.93|123/139	hypothetical protein
Cphy_0876	G/A|SNP|ts|GGT/AGT|G469S|nonsyn|0.992|118/118	Single-stranded-DNA-specific exonuclease (RecJ)
Cphy_1034	C/T|SNP|ts|GCT/GTT|A162V|nonsyn|0.971|137/137	NUDIX hydrolase
Cphy_1439	C/T|SNP|ts|CGG/CAG|R230Q|nonsyn|0.978|116/116	transposase IS4 family protein
Cphy_1673	C/T|SNP|ts|GCC/ACC|A94T|nonsyn|0.975|106/106	Excinuclease ABC subunit c
Cphy_2006	G/A|SNP|ts|GGG/AGG|G458R|nonsyn|0.977|109/109	DEAD/DEAH box helicase domain-containing protein
Cphy_2205	A/-|DEL|0.928|100/108	ATP-binding protein
Cphy_2380	-/T|INS|0.913|97/110	DNA mismatch repair protein MutL
Cphy_2493	C/T|SNP|ts|GCT/ACT|A150T|nonsyn|0.938|123/126	Primosomal protein N'
Cphy_2926	G/A|SNP|ts|GAT/AAT|D307N|nonsyn|0.976|127/127	transposase IS116 IS110 IS902 family protein
Transcription		
Cphy_0098	G/A|SNP|ts|CAG/TAG|Q215*|nonsense|0.976|191/192	MerR family transcriptional regulator
Cphy_0410	G/A|SNP|ts|AGA/AAA|R187K|nonsyn|0.987|162/162 C/T|SNP|ts|CCG/CTG|P312L|nonsyn|0.986|157/157	transcription termination factor Rho
Cphy_0510	T/C|SNP|ts|ACA/GCA|T17A|nonsyn|0.943|127/130	XRE family transcriptional regulator
Cphy_0548	A/G|SNP|ts|AGT/GGT|S497G|nonsyn|0.98|155/155	(p)ppGpp synthetase I, SpoT/RelA
Cphy_0954	A/G|SNP|ts|CAG/CGG|Q29R|nonsyn|0.978|127/128	XRE family transcriptional regulator
Cphy_1188	G/A|SNP|ts|ATG/ATA|M867I|nonsyn|0.987|146/146	peptidase S41
Cphy_1229	T/C|SNP|ts|GTA/GCA|V71A|nonsyn|0.981|126/126	MarR family transcriptional regulator
Cphy_1442	T/-|DEL|0.962|110/118	Transcriptional regulator, TetR family
Cphy_1454	T/C|SNP|ts|TGC/CGC|C210R|nonsyn|0.959|118/119	LacI family transcription regulator
Cphy_1554	G/A|SNP|ts|GCA/ACA|A76T|nonsyn|0.965|87/87	XRE family transcriptional regulator
Cphy_1762	-/A|INS|0.936|84/95	TetR family transcriptional regulator
Cphy_1902	-/A|INS|0.922|84/98	LysR family transcriptional regulator
Cphy_2151	G/A|SNP|ts|GGG/AGG|G94R|nonsyn|0.956|92/92	ArsR family transcriptional regulator
Cphy_2278	C/T|SNP|ts|GTG/ATG|V279M|nonsyn|0.962|94/94	LacI family transcription regulator
Cphy_2353	C/T|SNP|ts|CGT/TGT|R33C|nonsyn|0.875|80/83	LacI family transcription regulator
Cphy_2769	G/A|SNP|ts|CGT/TGT|R206C|nonsyn|0.99|112/112	transcriptional regulator GntR family
Cphy_2859	G/A|SNP|ts|TGC/TAC|C251Y|nonsyn|0.983|122/122	helix-turn-helix domain-containing protein

Cphy_2866	C/T|SNP|ts|GAT/AAT|D381N|nonsyn|0.985|139/140	ribonuclease R
Cphy_3397	G/A|SNP|ts|CGC/CAC|R17H|nonsyn|0.972|197/197	AraC family transcriptional regulator
	Traduction, structure et biogenèse des ribosomes	
Cphy_0238	A/G|SNP|ts|AAG/GAG|K303E|nonsyn|0.969|167/168	Elongation factor Tu
Cphy_0255	C/T|SNP|ts|CAG/TAG|Q141*|nonsense|0.972|187/189	sigma 54 modulation protein/ribosomal protein S30EA
Cphy_0822	T/A|SNP|tv|TAT/TTT|Y132F|nonsyn|0.986|137/137	HAD-superfamily hydrolase subfamily IA variant 3
Cphy_2280	T/-|DEL|0.952|87/95	S-adenosylmethionine:tRNA ribosyltransferase-isomerase
Cphy_2318	A/G|SNP|ts|ATT/GTT|I14V|nonsyn|0.977|91/91	30S ribosomal protein S20
Cphy_2510	G/A|SNP|ts|CCA/TCA|P198S|nonsyn|0.954|102/102	Hemolysin A
Cphy_2617	A/G|SNP|ts|GTA/GCA|V283A|nonsyn|0.965|135/135	Alanine-tRNA ligase 1
Cphy_3686	G/A|SNP|ts|ACT/ATT|T689I|nonsyn|0.951|176/177	50S ribosomal protein L33
	MÉTABOLISMES	
	Transport et métabolisme des nucléotides	
Cphy_0104	C/T|SNP|ts|ATG/ATA|M225I|nonsyn|0.975|174/174	Phosphoribosyl pyrophosphate synthase
Cphy_0643	A/G|SNP|ts|CAT/CGT|H70R|nonsyn|0.967|141/141	bifunctional phosphoribosylaminoimidazolecarboxamide formyltransferase
Cphy_0882	C/G|SNP|tv|CAG/GAG|Q275E|nonsyn|0.974|120/121	orotidine 5'-phosphate decarboxylase
Cphy_2329	C/T|SNP|ts|GAA/AAA|E47K|nonsyn|0.969|100/101	deoxyadenosine kinase
Cphy_2390	C/T|SNP|ts|GTT/ATT|V205I|nonsyn|0.946|79/79	GTPase Der
Cphy_2750	C/T|SNP|ts|GCA/GTA|A254V|nonsyn|0.965|123/123	Aspartate carbamoyltransferase
Cphy_2751	G/A|SNP|ts|GGA/AGA|G6R|nonsyn|0.99|122/122	aspartate carbamoyltransferase regulatory subunit
Cphy_3143	C/T|SNP|ts|GAG/AAG|E16K|nonsyn|0.963|156/156	adenosine deaminase
Cphy_3523	A/G|SNP|ts|TTG/TCG|L139S|nonsyn|0.97|167/169	deoxyguanosinetriphosphate triphosphohydrolase-like protein
Cphy_3622	G/A|SNP|ts|CCT/CTT|P352L|nonsyn|0.972|184/185	amidohydrolase 3
	Transport et métabolisme des lipides	
Cphy_0523	C/T|SNP|ts|CCA/CTA|P141L|nonsyn|0.971|142/143	acetyl-CoA carboxylase, carboxyl transferase, alpha subunit
Cphy_0523	G/A|SNP|ts|GTT/ATT|V278I|nonsyn|0.988|136/137	
Cphy_1932	C/T|SNP|ts|CTT/TTT|L15F|nonsyn|0.972|97/97	phosphatase PAP2 superfamily
Cphy_2446	G/A|SNP|ts|GCA/GTA|A58V|nonsyn|0.938|122/122	acyl carrier protein
Cphy_2457	A/G|SNP|ts|TTT/CTT|F177L|nonsyn|0.966|107/108	phospholipase D/transphosphatidylase
	Transport et métabolisme des acides aminés	
Cphy_0024	C/T|SNP|ts|GCG/GTG|A354V|nonsyn|0.963|161/162	putative aminopeptidase 2
Cphy_0630	G/A|SNP|ts|GAG/AAG|E47K|nonsyn|0.978|147/148	diaminopimelate epimerase
Cphy_0949	G/A|SNP|ts|GAG/AAG|E44K|nonsyn|0.989|124/124	lactoylglutathione lyase
Cphy_1100	C/T|SNP|ts|CCA/CTA|P228L|nonsyn|0.991|125/126	peptidase M42 family protein
Cphy_1235	A/G|SNP|ts|GAA/GGA|E6G|nonsyn|0.97|136/137	aminotransferase class I and II
Cphy_1499	C/T|SNP|ts|CGT/TGT|R69C|nonsyn|0.972|113/113 C/T|SNP|ts|CCA/TCA|P264S|nonsyn|0.964|88/88	Diaminopropionate ammonia-lyase
Cphy_1500	C/T|SNP|ts|GCA/GTA|A117V|nonsyn|0.978|123/123	peptidase
Cphy_1538	G/A|SNP|ts|GAT/AAT|D98N|nonsyn|0.977|112/112	glycine cleavage system T protein

Cphy_1684	C/T|SNP|ts|GTA/ATA|V138I|nonsyn|0.981|93/93	SufS subfamily cysteine desulfurase
Cphy_1753	G/A|SNP|ts|GCT/GTT|A217V|nonsyn|0.957|85/86	aminotransferase class V
Cphy_1999	T/A|SNP|tv|CAT/CAA|H251Q|nonsyn|0.97|85/87	oligopeptide/dipeptide ABC transporter, ATPase subunit
Cphy_2116	G/A|SNP|ts|GGC/AGC|G25S|nonsyn|0.992|90/90	glyoxalase/bleomycin resistance protein/dioxygenase
Cphy_2352	C/T|SNP|ts|GAT/AAT|D112N|nonsyn|0.977|97/98	aspartate kinase
Cphy_2903	-/A|INS|0.947|106/113	HisJ family histidinol phosphate phosphatase
Cphy_3364	A/G|SNP|ts|TGT/CGT|C21R|nonsyn|0.972|180/180	transglutaminase domain-containing protein
Cphy_3548	G/A|SNP|ts|CGT/TGT|R263C|nonsyn|0.966|148/149	2-isopropylmalate synthase
Cphy_3707	G/A|SNP|ts|ACA/ATA|T62I|nonsyn|0.98|157/158	transglutaminase domain-containing protein
Cphy_3840	G/A|SNP|ts|GCT/GTT|A138V|nonsyn|0.964|187/188	ABC transporter related
	Transport et métabolisme des carbohydrates	
Cphy_0220	G/A|SNP|ts|GAA/AAA|E309K|nonsyn|0.972|163/163	glycoside hydrolase family 3 protein
Cphy_0333	-/AGCC|INS|0.893|159/191	phosphotransferase system, phosphocarrier protein HPr
Cphy_0430	C/T|SNP|ts|CCA/CTA|P241L|nonsyn|0.983|151/151	Glycosyltransferase 36
Cphy_0775	G/A|SNP|ts|GCA/ACA|A180T|nonsyn|0.978|139/139	extracellular solute-binding protein
Cphy_1069	G/A|SNP|ts|GCG/ACG|A24T|nonsyn|0.958|127/128	hypothetical protein
Cphy_1074	T/A|SNP|tv|GAT/GAA|D176E|nonsyn|0.956|114/115	extracellular solute-binding protein
Cphy_1089	-/T|INS|0.93|101/112	extracellular solute-binding protein, family 1
Cphy_1118	G/A|SNP|ts|GAC/AAC|D58N|nonsyn|0.951|103/104	extracellular solute-binding protein family 1
Cphy_1140	A/G|SNP|ts|CAA/CGA|Q179R|nonsyn|0.976|112/112	GDSL family lipase
Cphy_1262	C/T|SNP|ts|GCA/GTA|A282V|nonsyn|0.979|106/106	dihydroxyacetone kinase subunit DhaK
Cphy_1391	G/A|SNP|ts|GAT/AAT|D120N|nonsyn|0.964|124/125	ABC transporter (Permease)
Cphy_1447	C/T|SNP|ts|GCT/GTT|A257V|nonsyn|0.96|91/91	hypothetical protein
Cphy_1510	G/A|SNP|ts|GAT/AAT|D390N|nonsyn|0.964|85/86	endo-1,4-beta-xylanase (EC 3.2.1.8)
Cphy_1585	G/A|SNP|ts|GGT/GAT|G192D|nonsyn|0.923|94/94	putative solute-binding component of ABC transporter
Cphy_1720	G/A|SNP|ts|TGC/TAC|C243Y|nonsyn|0.988|88/88	glycoside hydrolase family protein
Cphy_1935	G/A|SNP|ts|ACC/ATC|T238I|nonsyn|0.967|76/76	sugar (glycoside-Pentoside-hexuronide) transporter
Cphy_1937	-/T|INS|0.938|85/93	glycoside hydrolase family protein
Cphy_2012	A/G|SNP|ts|GAA/GGA|E137G|nonsyn|0.975|81/81	periplasmic binding protein/LacI transcriptional regulator
Cphy_2028	G/A|SNP|ts|GAA/AAA|E126K|nonsyn|0.967|82/82	glycoside hydrolase family 43
Cphy_2264	G/A|SNP|ts|CCA/CTA|P222L|nonsyn|0.981|107/107 C/T|SNP|ts|TGG/TGA|W211*|nonsense|0.961|116/116	glycosidase PH1107-related
Cphy_2274	G/A|SNP|ts|CCG/CTG|P711L|nonsyn|0.984|95/95	extracellular solute-binding protein
Cphy_2569	C/T|SNP|ts|GCA/ACA|A115T|nonsyn|0.961|107/109	extracellular solute-binding protein
Cphy_2740	G/A|SNP|ts|CCA/CTA|P95L|nonsyn|0.989|128/128	Altronate dehydratase

Cphy_3155	-/T|INS|0.94|138/148	L-fucose isomerase
Cphy_3280	G/A|SNP|ts|CAC/TAC|H74Y|nonsyn|0.974|132/133	ABC-type sugar transport system ATPase component-like protein
Cphy_3313	A/G|SNP|ts|GTC/GCC|V140A|nonsyn|0.963|137/137	Kojibiose phosphorylase
Cphy_3368	T/C|SNP|ts|GAT/GGT|D816G|nonsyn|0.978|146/147	cellulose 1,4-beta-cellobiosidase
Cphy_3388	G/A|SNP|ts|CCA/CTA|P970L|nonsyn|0.973|150/150	glucan endo-1,3-beta-D-glucosidase
Cphy_3509	G/A|SNP|ts|GAA/AAA|E525K|nonsyn|0.966|154/158	Ig domain-containing protein
Cphy_3573	G/A|SNP|ts|GGA/AGA|G121R|nonsyn|0.973|134/134	ROK family protein
Cphy_3862	G/A|SNP|ts|GCA/ACA|A1139T|nonsyn|0.948|175/177	endo-1,4-beta-xylanase
Cphy_3869	G/A|SNP|ts|ACG/ATG|T1298M|nonsyn|0.981|201/201 G/A|SNP|ts|CCA/CTA|P324L|nonsyn|0.983|173/173	fibronectin type III domain-containing protein
Cphy_3906	-/A|INS|0.932|170/194	phosphoenolpyruvate-protein phosphotransferase

Transport et métabolisme des ions inorganiques

Cphy_0201	C/T|SNP|ts|TCG/TTG|S160L|nonsyn|0.984|158/158	Methyl-accepting chemotaxis sensory transducer
Cphy_0543	C/G|SNP|tv|AGC/ACC|S805T|nonsyn|0.977|131/131	magnesium-translocating P-type ATPase
Cphy_0677	G/A|SNP|ts|GCA/ACA|A725T|nonsyn|0.925|145/150	ATPase, P-type (transporting), HAD superfamily, subfamily IC
Cphy_0731	C/T|SNP|ts|TCA/TTA|S63L|nonsyn|0.983|138/138	transport system permease protein
Cphy_0814	A/G|SNP|ts|ATT/GTT|I199V|nonsyn|0.979|136/136	Binding-protein-dependent transport systems, inner membrane component
Cphy_0893	C/T|SNP|ts|GGA/AGA|G475R|nonsyn|0.988|113/113	catalase
Cphy_0895	G/A|SNP|ts|GCT/GTT|A220V|nonsyn|0.956|135/135	ornithine carbamoyltransferase
Cphy_1076	G/A|SNP|ts|GAT/AAT|D231N|nonsyn|0.982|116/116	binding-protein-dependent transport systems inner membrane component
Cphy_1449	C/T|SNP|ts|CAG/TAG|Q121*|nonsense|0.936|114/115	phosphonate ABC transporter, ATPase subunit
Cphy_1482	C/T|SNP|ts|CCA/TCA|P55S|nonsyn|0.978|119/119	ABC transporter, periplasmic molybdate-binding protein
Cphy_1520	G/A|SNP|ts|GCA/ACA|A240T|nonsyn|0.985|100/100	Methyl-accepting chemotaxis sensory transducer
Cphy_1586	G/A|SNP|ts|GTT/ATT|V262I|nonsyn|0.99|109/109	ABC transporter related
Cphy_1653	C/T|SNP|ts|CAA/TAA|Q35*|nonsense|0.96|114/114	hemerythrin-like metal-binding protein
Cphy_1670	G/A|SNP|ts|GAT/AAT|D225N|nonsyn|0.973|119/119	ABC transporter related
Cphy_1827	G/A|SNP|ts|CGA/CAA|R4Q|nonsyn|0.97|77/78	sulfate ABC transporter, inner membrane subunit CysT
Cphy_1868	C/T|SNP|ts|GGA/GAA|G99E|nonsyn|0.999|93/93	FAD-dependent pyridine nucleotide-disulphide oxidoreductase
Cphy_2054	C/T|SNP|ts|GGC/GAC|G524D|nonsyn|0.945|103/103	ABC transporter related
Cphy_2132	C/T|SNP|ts|GTT/ATT|V314I|nonsyn|0.973|76/77	ATPase, P-type (transporting), HAD superfamily, subfamily IC
Cphy_2153	A/-|DEL|0.98|99/103	arsenical-resistance protein
Cphy_2392	-/C|INS|0.923|114/125	Periplasmic binding protein
Cphy_3206	G/A|SNP|ts|CGT/TGT|R423C|nonsyn|0.981|135/136	Methyl-accepting chemotaxis sensory transducer

Cphy_3589	C/T|SNP|ts|GCT/ACT|A22T|nonsyn|0.982|170/170	Binding-protein-dependent transport systems, inner membrane component
	Biosynthèse, transport et catabolisme des métabolites secondaires	
Cphy_1226	A/T|SNP|tv|TTA/TTT|L89F|nonsyn|0.982|120/120	Methyltransferase
Cphy_1501	G/A|SNP|ts|GTA/ATA|V477I|nonsyn|0.979|96/96	N-acyl-D-glutamate deacylase
Cphy_1897	C/T|SNP|ts|GCG/GTG|A6V|nonsyn|0.881|84/88	fad dependent oxidoreductase
Cphy_2911	T/C|SNP|ts|AAT/GAT|N141D|nonsyn|0.941|141/141	cytochrome P450
	Production et conversion d'énergie	
Cphy_0007	G/A|SNP|ts|GTA/ATA|V82I|nonsyn|0.975|185/186	tartrate/fumarate subfamily Fe-S type hydrolyase alpha subunit
Cphy_0088	C/T|SNP|ts|GGT/AGT|G533S|nonsyn|0.985|171/171	NADH dehydrogenase (quinone)
Cphy_0092	T/-|DEL|0.97|173/184	putative PAS/PAC sensor protein
Cphy_0216	G/A|SNP|ts|GGA/AGA|G234R|nonsyn|0.974|175/175	electron transport complex, RnfABCDGE type, B subunit
Cphy_0545	C/T|SNP|ts|GCT/GTT|A185V|nonsyn|0.95|148/151	4Fe-4S ferredoxin iron-sulfur binding domain-containing protein
Cphy_0546	T/C|SNP|ts|GTA/GCA|V402A|nonsyn|0.946|179/180 T/G|SNP|tv|CAT/CAG|H590Q|nonsyn|0.989|143/143	4Fe-4S ferredoxin iron-sulfur binding domain-containing protein
Cphy_0603	G/A|SNP|ts|GGT/GAT|G786D|nonsyn|0.96|141/142	pyruvate flavodoxin/ferredoxin oxidoreductase domain-containing protein
Cphy_1733	C/T|SNP|ts|CGC/TGC|R27C|nonsyn|0.964|101/101	ech hydrogenase, subunit EchD, putative
Cphy_1820	G/A|SNP|ts|GGA/GAA|G385E|nonsyn|0.976|97/97	anaerobic c4-dicarboxylate antiporter
Cphy_1923	C/T|SNP|ts|CAT/TAT|H316Y|nonsyn|0.967|98/98	D-isomer specific 2-hydroxyacid dehydrogenase NAD-binding
Cphy_2044	G/A|SNP|ts|ATG/ATA|M1I|nonsyn|0.973|115/115	hypothetical protein
Cphy_2106	G/A|SNP|ts|TGG/TAG|W186*|nonsense|0.954|80/81	hypothetical protein
Cphy_2255	G/C|SNP|tv|GCT/GGT|A115G|nonsyn|0.961|80/81	FMN-dependent alpha-hydroxy acid dehydrogenase
Cphy_3121	G/A|SNP|ts|CAT/TAT|H122Y|nonsyn|0.972|128/128	Rubrerythrin
Cphy_3123	G/A|SNP|ts|CAG/TAG|Q229*|nonsense|0.945|141/143	pyruvate flavodoxin/ferredoxin oxidoreductase domain-containing protein
Cphy_3750	C/T|SNP|ts|GTA/ATA|V115I|nonsyn|0.974|154/154	ATP synthase subunit beta
Cphy_3925	G/A|SNP|ts|GTA/ATA|V320I|nonsyn|0.985|198/199	bifunctional acetaldehyde-CoA/alcohol dehydrogenase
	Transport et métabolisme des coenzymes	
Cphy_1371	T/A|SNP|tv|AAT/AAA|N363K|nonsyn|0.956|146/148	uroporphyrin-III C-methyltransferase
Cphy_1379	G/A|SNP|ts|GAT/AAT|D473N|nonsyn|0.937|104/104	precorrin-6x reductase
Cphy_1481	T/C|SNP|ts|GTG/GCG|V266A|nonsyn|0.981|106/106	Molybdenum cofactor synthesis domain protein
Cphy_2082	C/T|SNP|ts|ATG/ATA|M377I|nonsyn|0.984|99/99	thiamine biosynthesis protein ThiH
Cphy_2124	A/G|SNP|ts|TAC/TGC|Y161C|nonsyn|0.979|85/85 G/A|SNP|ts|GCT/ACT|A184T|nonsyn|0.966|89/89	aspartate dehydrogenase
Cphy_2779	G/A|SNP|ts|CGC/TGC|R323C|nonsyn|0.981|109/109	biotin--acetyl-CoA-carboxylase ligase

ACS SyntheticBiology

Open Access
This article is licensed under CC-BY-NC-ND 4.0

pubs.acs.org/synthbio
Research Article

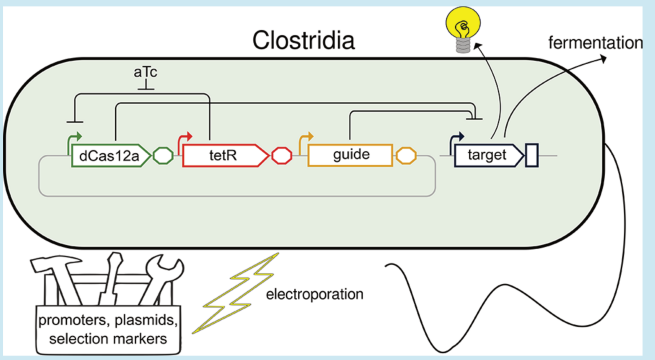
Tuning of Gene Expression in *Clostridium phytofermentans* Using Synthetic Promoters and CRISPRi

William Rostain, Tom Zaplana, Magali Boutard, Chloé Baum, Sibylle Tabuteau, Mary Sanitha, Mohandass Ramya, Adam Guss, Laurence Ettwiller, and Andrew C. Tolonen*

Cite This: *ACS Synth. Biol.* 2022, 11, 4077–4088
Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Control of gene expression is fundamental to cell engineering. Here we demonstrate a set of approaches to tune gene expression in Clostridia using the model *Clostridium phytofermentans*. Initially, we develop a simple benchtop electroporation method that we use to identify a set of replicating plasmids and resistance markers that can be cotransformed into *C. phytofermentans*. We define a series of promoters spanning a >100-fold expression range by testing a promoter library driving the expression of a luminescent reporter. By insertion of *tet* operator sites upstream of the reporter, its expression can be quantitatively altered using the Tet repressor and anhydrotetracycline (aTc). We integrate these methods into an aTc-regulated dCas12a system with which we show *in vivo* CRISPRi-mediated repression of reporter and fermentation genes in *C. phytofermentans*. Together, these approaches advance genetic transformation and experimental control of gene expression in Clostridia.



KEYWORDS: *Clostridia, CRISPRi, Cas12a/Cpf1, fermentation, electroporation, methylome*

Downloaded via 78.194.54.9 on October 11, 2024 at 14:19:48 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

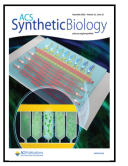
INTRODUCTION

The ability to experimentally control expression of target genes is needed for reliable modification of biological systems. While numerous tools to modulate gene expression are available for a few well-studied models such as *Escherichia coli*, they are lacking in many other bacterial taxa with applications in biotechnology. For example, the Clostridia are Gram-positive, anaerobic, spore-forming bacteria with important roles in industry and health. Clostridia include species that transform plant biomass or carbon dioxide into value-added chemicals, intestinal commensals that metabolize dietary fiber to produce health-promoting metabolites, and important pathogens.[1−4] *Clostridium phytofermentans* (also called *Lachnoclostridium phytofermentans*) is a member of the family Lachnospiraceae that is distinguished by its ability to ferment lignocellulosic biomass with ethanol, hydrogen, and acetate as major products.[5,6]

Advances in genetic manipulation are needed to exploit the therapeutic and industrial potential of additional Clostridia species. In particular, experimental approaches for *in vivo* modulation of gene expression have a few prerequisites. First, a method is needed to deliver foreign DNA into the cell. Among DNA delivery methods, electroporation is advantageous for simultaneous delivery of multiple DNA constructs or linear DNA fragments. Electroporation methods have been demonstrated in a few other model Clostridia, including biomass-fermenting species,[7,8] but typically use an anaerobic glovebox that requires significant infrastructure investment. Second, modulation of gene expression using genetic regulatory circuits often requires more than one plasmid, so a set of compatible resistance markers and plasmid replicons are needed. A set of modular shuttle plasmids, called pMTL plasmids, has been constructed to test replicons and selectable markers in Clostridia.[9] Third, defined regulatory elements are needed to finely control expression of target genes. A set of constitutive promoters of different strengths can be applied to specify transcription levels,[10] and flanking a promoter with operator sites has been applied to modulate transcription in *Clostridium acetobutylicum* using the tet repressor (TetR) and anhydrotetracycline (aTc).[11]

CRISPR interference (CRISPRi) is a method to repress transcription that does not require changing regulatory elements of the target gene.[12] By customizing the guide RNA (gRNA) sequence, the DNase-dead CRISPR effector, *e.g.*, Cas9 (dCas9) or Cas12a (dCas12a), can be programmed

Received: July 20, 2022
Published: November 25, 2022



ACS Publications
© 2022 The Authors. Published by American Chemical Society

https://doi.org/10.1021/acssynbio.2c00385
ACS Synth. Biol. 2022, 11, 4077−4088

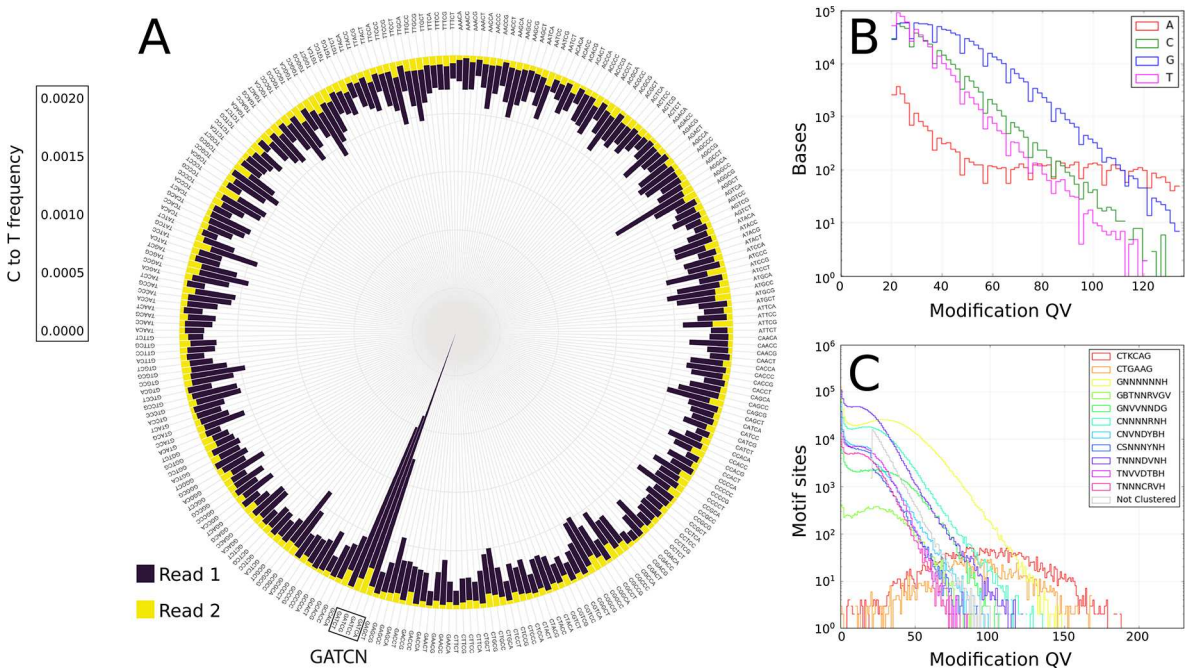


Figure 1. Base modification detection in the *C. phytofermentans* genome by (A) RIMS-seq sequencing and (B, C) SMRT sequencing. (A) RIMS-seq data analysis showing frequency of C-to-T mutations on Read 1 (blue bars) and Read 2 (yellow bars), indicating m5C methylation at the 5′-GATC-3′ motif ($p = 1.23 \times 10^{-4061}$). (B) Kinetic detection histogram of the number of bases at each modification QV. (C) Modified base motif analysis identified two motifs with predicted m6A modification: CTKCAG and CTGAAG, where K is G/T. Modified bases are underlined for both sites. Abbreviations: SMRT, single molecule real-time; RIMS, rapid identification of methylase specificity; QV, quality value; m6A, 6-methyladenosine; m5C, 5-methylcytosine.

to bind any DNA sequence containing a protospacer adjacent motif (PAM) element without cleaving the DNA. Transcriptional repression results because the dCas protein interferes with RNA polymerase by binding either at the promoter to inhibit transcript initiation or downstream to prevent transcript elongation. While Cas9-mediated CRISPRi has been more widely developed in Clostridia,[13−15] Cas12a (also called Cpf1) presents certain advantages.[16] Cas12a does not require a tracrRNA and can process a tandem array of guides using its intrinsic RNase activity.[17,18] Multiplexing targets is thus simpler with Cas12a than with Cas9, which requires RNase III for crRNA cleavage and an independently transcribed gRNA cassette for each guide. Cas12a also appears to be less toxic than Cas9 in many bacteria.[19]

The goal of this study is to develop a suite of approaches for efficient genetic transformation and experimental modulation of gene expression in Clostridia using *C. phytofermentans* as a model. We demonstrate a simple benchtop method for electrotransformation of *C. phytofermentans* that does not require an anaerobic glovebox, and we apply it to systematically test antibiotic resistance cassettes and plasmid origins of replication. We use a luminescent reporter to characterize a collection of synthetic promoters of varying strengths and a TetR-based system for inducible gene expression. We unite these tools into a CRISPRi system for gene repression based on a TetR-regulated dCas12a and show that it can be used for efficient *in vivo* CRISPRi-mediated repression of reporter and metabolic genes. We discuss how these approaches advance *C. phytofermentans* as a model system and can be applied to engineer other Clostridia.

RESULTS

Restriction Modification Systems in *C. phytofermentans*. Electroporation of some Clostridia species requires methylation of vector DNA to circumvent endogenous restriction modification systems (RMSs),[20,21] leading us to investigate the presence of RMSs in *C. phytofermentans*. REBASE[22] predicts that the *C. phytofermentans* genome encodes two type II RMSs, Cphy0266−8 and Cphy2923−5, and one type IV restriction enzyme, Cphy1615. Cphy0267 was predicted to methylate the sequence 5′-CTGAAG-3′, whereas no specificities could be predicted for Cphy2924. Three forms of DNA methylation are common in bacterial genomes: 5-methylcytosine (m5C), N^4-methylcytosine (m4C), and N^6-methyladenine (m6A).[23] We profiled the *C. phytofermentans* genome for these three methylation patterns using two complementary sequencing methods: RIMS-seq identifies m5C as C-to-T mutations on Read 1,[23] and SMRT sequencing identifies m4C and m6A based on nucleotide incorporation rates measured as pulse width and interpulse duration.[24]

RIMS-seq of the *C. phytofermentans* genome revealed m5C modification at 5′-GATC-3′ based on elevated C-to-T mutations ($p = 1.23 \times 10^{-4061}$) (Figure 1A). DNA modification analysis of SMRT sequencing supported the presence of m6A methylation in the *C. phytofermentans* genome based on an increased proportion of adenine bases with elevated modification quality values (QVs) relative to other bases (Figure 1B). SMRT motif analysis[25] localized the m6A modification to 5′-CTKCAG-3′ (97.2% modified, mean score 100.6) and 5′-CTGAAG-3′ (96.4% modified, mean score 95.3) (Figure 1C). Together, sequence analysis and methylome profiling of the *C. phytofermentans* genome suggest that Cphy0267 performs m6A methylation at 5′-CTGAAG-3′ and Cphy2924 performs m5C methylation at 5′-GATC-3′.

https://doi.org/10.1021/acssynbio.2c00385
ACS Synth. Biol. 2022, 11, 4077−4088

On the basis of these methylation activities, we examined the restriction activity of *C. phytofermentans* lysate on plasmid DNA isolated from ER2796, a DNA-methylation-deficient *E. coli* strain.[26] *In vitro* treatment of unmethylated plasmid DNA with *C. phytofermentans* lysate did not result in DNA cleavage (Figure 2A), suggesting that the putative type II and type IV

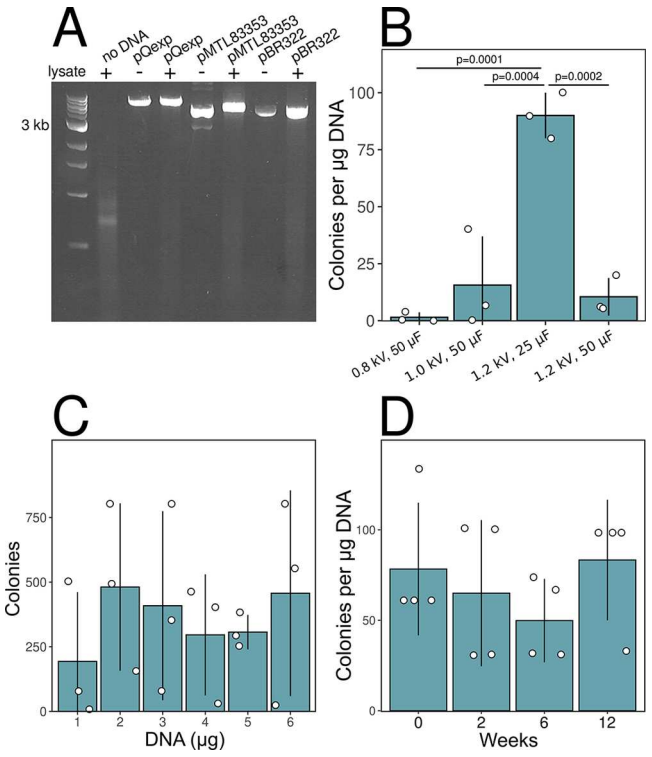


Figure 2. Factors affecting efficacy of *C. phytofermentans* electroporation. (A) Treatment of linearized plasmid DNA with *C. phytofermentans* culture lysate. Three plasmids (pQexp, pMTL83353, and pBR322) isolated from methylation-deficient *E. coli* ER2796 were linearized with *Nhe*I, incubated with (+) or without (−) *C. phytofermentans* culture lysate, and resolved by gel electrophoresis. (B) Electroporation efficiencies (colonies per μg of DNA) of 3 μg of plasmid DNA delivered using different electropulse voltages and capacitances. (C) Colonies after electroporation with 1−6 μg of plasmid DNA. (D) Electroporation efficiencies (colonies per μg of DNA) after storage of competent cells at −80 °C for 0−12 weeks. (B−D) Electrotransformations were performed with pQexp and transformants selected with 40 μg mL^{-1} erythromycin. Data points show each electrotransformation; bars show mean ± SD. Treatment comparisons were analyzed by Tukey's test after confirmation that the data distribution requirements were met using the Shapiro−Wilk test; p values < 0.05 are shown. SD, standard deviation.

restriction systems are inactive. Further, the electroporation efficiencies of *C. phytofermentans* using the methods described below were similar whether pQexp (Table 1) was isolated from *C. phytofermentans*, *E. coli* DH5α, or *E. coli* ER2796, supporting that premethylation of plasmid DNA with *C. phytofermentans* methylases does not improve electroporation (data not shown). We thus concluded that although the *C. phytofermentans* genome is methylated, the associated restriction enzymes are likely inactive and do not impede electrotransformation.

Electroporation of *C. phytofermentans*. Previously, the transfer of plasmid DNA into *C. phytofermentans* was demonstrated by conjugation with *E. coli*.[27−31] Conjugation showed that pQexp, which bears the pAMβ1 replicon and erythromycin resistance gene, stably replicates *C. phytofermentans*.[27] To develop methods for electrotransformation of *C. phytofermentans*, we evaluated the effects of electropulse, DNA concentration, and cell wall-weakening osmolytes on electroporation of pQexp using an exponential-decay wave pulse. All electroporations were performed at the bench without an anaerobic glovebox, making these methods generally accessible to microbiology laboratories.

We varied the pulse voltage and capacitance to find the highest transformation efficiency at 1.2 kV and 25 μF (Figure 2B), which typically yielded 80−100 colonies per μg of DNA and a time constant of ∼3 ms. Similar numbers of colonies were observed at DNA concentrations from 1 to 6 μg, supporting that DNA uptake does not limit the transformation efficiency (Figure 2C). Glycine improves electrotransformation in other Clostridia by incorporating into the cell wall in place of D-alanine, weakening the cell wall by reducing peptidoglycan linkages.[32,33] Supplementing GS2 medium with >100 mM glycine significantly reduced the growth rate, but 50, 100, or 150 mM glycine supplementation did not affect the transformation efficiency (data not shown). We also found that electrocompetent *C. phytofermentans* cells can be stored frozen at −80 °C without loss of transformation efficiency, even when the cells were frozen for 12 weeks (Figure 2D).

Plasmid Replicons and Resistance Markers. The ability to independently select for more than one plasmid in the same cell requires a set of compatible plasmid replicons and antibiotic resistance genes. To identify plasmid replicons that function in *C. phytofermentans*, we used the pMTL plasmid backbone (Figure 3A)[9] to construct plasmids pQmod1−4E, each bearing the erythromycin resistance gene and a different Gram-positive replicon: pIM13, pBP1, pCB102, and pCD6, respectively (Table 1). Plasmids with the pBP1 and pCB102 origins were electroporated with efficiencies similar to that of the pAMβ1 origin used previously (Figure 3B), and transformants with plasmids bearing any of these three origins can be maintained in liquid medium with erythromycin at growth rates similar to wild-type (WT) without selection. The transformation efficiencies were significantly lower with the plasmids bearing a pCD6 origin (p = 0.039) or pIM13 origin (p = 0.037) relative to the pAMβ1 origin (Figure 3B). Cells retain pCD6 plasmids in liquid medium with erythromycin at a growth rate 3 times lower than WT without selection. The pIM13 transformant colonies were unable to grow under selection in liquid medium, suggesting that the pIM13 origin is too unstable to be maintained in *C. phytofermentans* even with antibiotic selection.

To identify selectable markers for *C. phytofermentans*, we constructed pQmod2 plasmids bearing the pBP1 replicon and resistance genes for spectinomycin (*aad9*), thiamphenicol (*catP*), or tetracycline (*tetA*) (Table 1). We found that *aad9*-containing plasmids can be maintained using 600 μg mL^{-1} spectinomycin and *catP* plasmids can be maintained using 40 μg mL^{-1} thiamphenicol (Figure 3C). Although *C. phytofermentans* is sensitive to 15 μg mL^{-1} tetracycline in liquid and solid media, we were unable to isolate transformants with plasmids bearing the *tetA* gene,[34] supporting that this resistance gene is not functional in *C. phytofermentans*.

Having identified three plasmid replicons (pAMβ1, pBP1, and pCB102) and three resistance cassettes (*ermB*, *aad9*, and *catP*) that function in *C. phytofermentans*, we tested whether they can be simultaneously maintained in the same cell. Sequential transformation of *C. phytofermentans* with pQexp

https://doi.org/10.1021/acssynbio.2c00385
ACS Synth. Biol. 2022, 11, 4077−4088

Table 1. Plasmids and Strains Used in This Study

name	description	source
	Plasmids	
pQexp	*erm*, pAMβ1 Gram(+) origin, pUC Gram(−) origin, oriT	ref 27
pBR322	*tet*, *bla*, colE1 Gram(−) origin	New England Biolabs
pMTL82254	*ermB*, *pBP1 Gram(+) origin*, *ColEl Gram(−) origin* + *tra*, *catP reporter*	Chain Biotech
pMTL83353	*aad9*, pCB102 Gram(+) origin, ColE1 Gram(−) origin + *tra*, Pfdx + MCS	Chain Biotech
pMTL84422	*tetA*, pCD6 Gram(+) origin, p15a Gram(−) origin + *tra*, Pthl + MCS	Chain Biotech
pMTL85141	*catP*, pIM13 Gram(+) origin, ColEI, MCS	Chain Biotech
pQmod1E	*ermB*, pIM13 Gram(+) origin, ColE1 Gram(−) origin, MCS	this study
pQmod2E	*ermB*, pBP1 Gram(+) origin, ColE1 Gram(−) origin, MCS	this study
pQmod3E	*ermB*, pCB102 Gram(+) origin, ColE1 Gram(−) origin, MCS	this study
pQmod4E	*ermB*, pCD6 Gram(+) origin, ColE1 Gram(−) origin, MCS	this study
pQmod2C	*catP*, pBP1 Gram(+) origin, ColE1 Gram(−) origin, MCS	this study
pQmod2S	*aad9*, pBP1 Gram(+) origin, ColE1 Gram(−) origin, MCS	this study
pQmod2T	*tetA*, pBP1 Gram(+) origin, ColE1 Gram(−) origin, MCS	this study
pQmod3C	*catP*, pCB102 Gram(+) origin, ColE1 Gram(−) origin, MCS	this study
pQmod3S	*aad9*, pCB102 Gram(+) origin, ColE1 Gram(−) origin, MCS	this study
pQmod3T	*tetA*, pCB102 Gram(+) origin, ColE1 Gram(−) origin, MCS	this study
pQmod2E-GG	pQmod2E, Plac-RFP flanked by *Bsa*I sites	this study (Addgene 191345)
pQmod2S-GG	pQmod2S, Plac-RFP flanked by *Bsa*I sites	this study (Addgene 191346)
pQmod2C-GG	pQmod2C, Plac-RFP flanked by *Bsa*I sites	this study (Addgene 191347)
pQmod3E-GG	pQmod3E, Plac-RFP flanked by *Bsa*I sites	this study (Addgene 191348)
pQmod3S-GG	pQmod3S, Plac-RFP flanked by *Bsa*I sites	this study (Addgene 191349)
pQmod3C-GG	pQmod3C, Plac-RFP flanked by *Bsa*I sites	this study (Addgene 191350)
pQmod4E-GG	pQmod4E, Plac-RFP flanked by *Bsa*I sites	this study (Addgene 191351)
pATmin-GG	*erm*, pAMβ1 Gram(+) origin, pUC Gram(−) origin, Plac-RFP flanked by *Bsa*I sites	this study (Addgene 191352)
pSB1C3-J04450	*cat*, pUC Gram(−) origin, Plac-mRFP1 cassette	Registry of Standard Biological Parts (J04450)
pNL1.1[Nluc]	source of NanoLuc gene	Promega
pQnl_Pcons17	*erm*, pAMβ1 Gram(+) origin, pUC Gram(−) origin, Pcons17-Nanoluc	this study (Addgene 191353)
pQnl_Pcphy1−24	pQnl_cons17 with Pcons17 replaced with promoters Pcphy1 to Pcphy24	this study
pQnl_Pcphy23	*erm*, pAMβ1 Gram(+) origin, pUC Gram(−) origin, Pcphy23-Nanoluc	this study (Addgene 191354)
pQnl_tet	*ermB*, pBP1 Gram(+) origin, ColE1 Gram(−) origin, PGusA2-TetO2/1-Nanoluc, miniPthl-tetR	this study (Addgene 191355)
pY109	source of Cas12a (LbCpf1)	Addgene 84740
pQdC12a	PGusA2-TetO2/1-dCas12a, miniPthl-*tetR*, gRNA cassette, pQmod3C-GG backbone	this study (Addgene 191356)
	Strains	
E. coli NEB 5-alpha	competent *E. coli* cells used in transformations	New England Biolabs
E. coli ER2796	*E. coli* strain deficient in DNA methylation	ref 26
C. phytofermentans ISDg	reference strain	ATCC 700394

(pAMβ1, *erm*), pQmod2S (pBP1, *aad9*), and pQmod3C (pCB102, *catP*) yielded transformants that after five serial transfers in liquid medium with triple antibiotic selection were confirmed by PCR to retain all three plasmids (Figure 3D). To simplify cloning and testing of genetic elements, we built a series of pQmod-GG plasmids containing an IPTG-inducible Plac promoter driving expression of the red fluorescent protein (RFP) gene for red/white selection in *E. coli* (Figure S1). The Plac-RFP cassette is flanked with *Bsa*I sites to facilitate gene replacement by Golden Gate assembly.[35] pQmod-GG plasmids with pBP1 and pCB102 replicons were constructed with all three resistance markers, and plasmids with the pCD6 and pAMβ1 replicons were made with erythromycin resistance (Table 1).

Constitutive and Inducible Promoters. A set of constitutive promoters of varying strengths can be applied to customize the expression of target genes and dissect promoter sequence−function relationships. We designed the synthetic Pcons17 promoter, a benchmark for high expression in *C. phytofermentans*. Pcons17 incorporates the consensus promoter elements (UP element, −35 box, −10 box) defined by genome-wide mapping of transcription start sites in *C. phytofermentans*[36] and computationally optimized sequences between the promoter elements[37,38] and the 5′ UTR containing the ribosome binding site.[39] Pcons17 was cloned upstream of the Nanoluc luciferase reporter[40] in the pQexp backbone, yielding pQnl_Pcons17 (Table 1). Cell-density-normalized luminescence demonstrated high NanoLuc expression from the Pcons17 consensus promoter in pQnl_Pcons17 (Figure 4A).

We created the Pcphy promoter library by systematically permuting the Pcons17 sequence using oligonucleotides with 20 degenerate positions covering all conserved nucleotides in the promoter consensus motif (Figure 4B). The Pcphy library was cloned upstream of the NanoLuc reporter to make the pQnl_Pcphy plasmid library. NanoLuc-based luminescence from different clones demonstrated that promoters in the Pcphy library span a wide range of expression from 4-fold to 2000-fold above the background luminescence of cells containing an empty plasmid (Figure 4A; see Figure S2 for promoter sequences). Sequence motifs in which the frequencies of nucleotides were weighted based on Pcphy-driven NanoLuc expression levels revealed the relative importance of nucleotide conservation across promoter elements (Figure 4C,D). These expression-weighted motifs support that expression is enhanced by an A-rich UP element upstream of the −35 box hexamer (5′-TTGACA-3′) (Figure 4C), which likely interacts with the alpha subunit of RNA polymerase.[41] In addition, there are two conserved positions upstream of the −10 hexamer box (5′-TATAAT-3′) (Figure 4D) that may increase transcription, as in some other Gram-positive bacteria.[42]

The ability to experimentally modulate gene expression using a chemical inducer simplifies expression level optimiza-

https://doi.org/10.1021/acssynbio.2c00385
ACS Synth. Biol. 2022, 11, 4077−4088

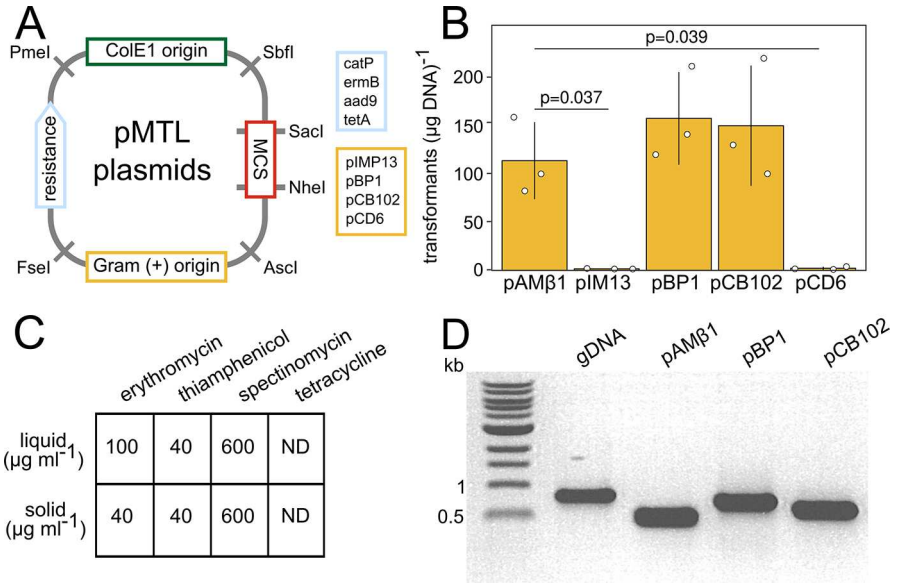


Figure 3. Plasmid replicons and antibiotic resistance genes that function in *C. phytofermentans*. (A) Structure of modular pMTL plasmids showing the MCS, Gram-positive replicon, antibiotic resistance marker, and Gram-negative ColE1 replicon. Restriction enzyme sites between modular units are shown along with the *Sac*I and *Nhe*I sites in the MCS used to insert the Golden Gate RFP cassette in pQmod-GG plasmids. Gram-positive origins and resistance markers tested in *C. phytofermentans* are listed. (B) Efficiencies of *C. phytofermentans* electrotransformation using erythromycin-resistant plasmids with different Gram-positive replicons. Pairwise treatment comparisons were analyzed by Tukey's test after confirmation that the data distribution requirements were met by the Shapiro−Wilk test; significant *p* values relative to pAM*β*1 (pQexp) transformations are shown. (C) Antibiotic concentrations to select for *C. phytofermentans* transformants on solid medium and in liquid culture. (D) PCR of *C. phytofermentans* culture showing stable simultaneous maintenance of three plasmids with different Gram-positive replicons and resistance markers: pQexp (pAM*β*1, *erm*), pQmod2S (pBP1, *aad9*), and pQmod3C (pCB102, *catP*). PCR was performed with primers for genomic DNA (primers 1575F/R) and plasmid origins pAM*β*1 (primers PR46/47), pBP1 (primers PR28/29), and pCB102 (primers PR30/31). Abbreviations: MCS, multiple cloning site; RFP, red fluorescent protein; ND, no data.

tion, especially with toxic gene products that need to be repressed under certain conditions. We demonstrated a system for inducible gene expression in *C. phytofermentans* using the tetracycline repressor (TetR) from Tn10. TetR is a transcriptional repressor whose binding to the TetO operator sequence is inhibited by tetracycline or anhydrotetracycline (aTc). We constructed pQnl_tet (Table 1) containing *tetR* driven by the miniPthl promoter along with the NanoLuc gene under control of the PgusA2-tetO2/1 promoter (Figure 4E), developed for TetR-based gene repression in *C. acetobutylicum*.[11] Upon electroporation of *C. phytofermentans* with pQnl_tet, nanoLuc expression was strongly repressed in the absence of aTc (Figure 4F). NanoLuc expression was induced with 5 ng mL^{-1} aTc and progressively increased up to 200 ng mL^{-1} aTc without affecting growth (Figure 4F). Concentrations of aTc above 200 ng mL^{-1} further increased expression but also reduced growth, likely due to either the metabolic cost of expressing Nanoluc or aTc toxicity.

Gene Repression by CRISPRi. CRISPRi using DNase-dead CRISPR effectors of Cas9 (dCas9) or Cas12a (dCas12a) streamlines transcriptional repression of target genes by eliminating the need to modify their promoters to contain operator sites, as is required in the TetR system. In this study, we selected the Cas12a from Lachnospiraceae bacterium ND2006 (LbCas12a).[18] While CRISPRi based on LbCas12a has not yet been tested in prokaryotes,[19] we reasoned that the protein will likely be active in *C. phytofermentans* because it comes from a bacterium in the same family. As the rarity of Cas9 PAM (5′-NGG-3′) sites has impeded Cas9 targeting in other AT-rich Clostridia genomes,[15] we assessed the distribution of the Cas12a PAM (5′-TTTV-3′)[43] across the *C. phytofermentans* genome. The Cas12a PAM is less common than that of Cas9 across the *C. phytofermentans* genome, but it is more abundant in highly expressed promoter regions (Figure S3). Thus, Cas12a is well-suited to target *C. phytofermentans* promoter regions using CRISPRi.

Plasmids constitutively expressing Cas9 sometimes cannot be transformed into Clostridia due to toxicity of the Cas protein.[44] While Cas12a is reported to be less toxic to Clostridia than Cas9,[45] we mitigated potential Cas12a toxicity by constructing a Tet-repressible dCas12a-based CRISPRi plasmid, pQdC12a. To this end, pQmod3C was modified to carry the following elements: the miniPthl-*tetR* cassette from pQnl_tet, an aTc-inducible dLbCas12a, and a gRNA designed so that the target can be customized by Golden Gate cloning using *Bbs*I (Figure 5A). As placement of a Rho-independent terminator immediately downstream of the guide impairs Cas activity,[46] we included the first 25 bp of the terminal repeat from the wild-type LbCas12a array downstream of the guide, which including the cloning scar separates the guide from the *fdx* terminator by 55 bp. To demonstrate dCas12a-mediated gene repression in *C. phytofermentans*, we targeted pQdC12a to two sites in the Pcphy23 promoter driving NanoLuc expression on pQnl_Pcphy23 (Figure 5A). Guide g-nl1 binds the Pcphy23 promoter at the −10 box, and guide g-nl2 binds at the −35 box (Figure S4). As both guides target the promoter region, we did not expect that binding orientation would affect repression.[47]

We cotransformed *C. phytofermentans* with pQnl_Pphy23 and pQdC12a targeted with either guide g-nl1, guide g-nl2, or a no-guide control. In the presence of aTc, NanoLuc expression was strongly repressed using g-nl1 (19-fold

https://doi.org/10.1021/acssynbio.2c00385
ACS Synth. Biol. 2022, 11, 4077−4088

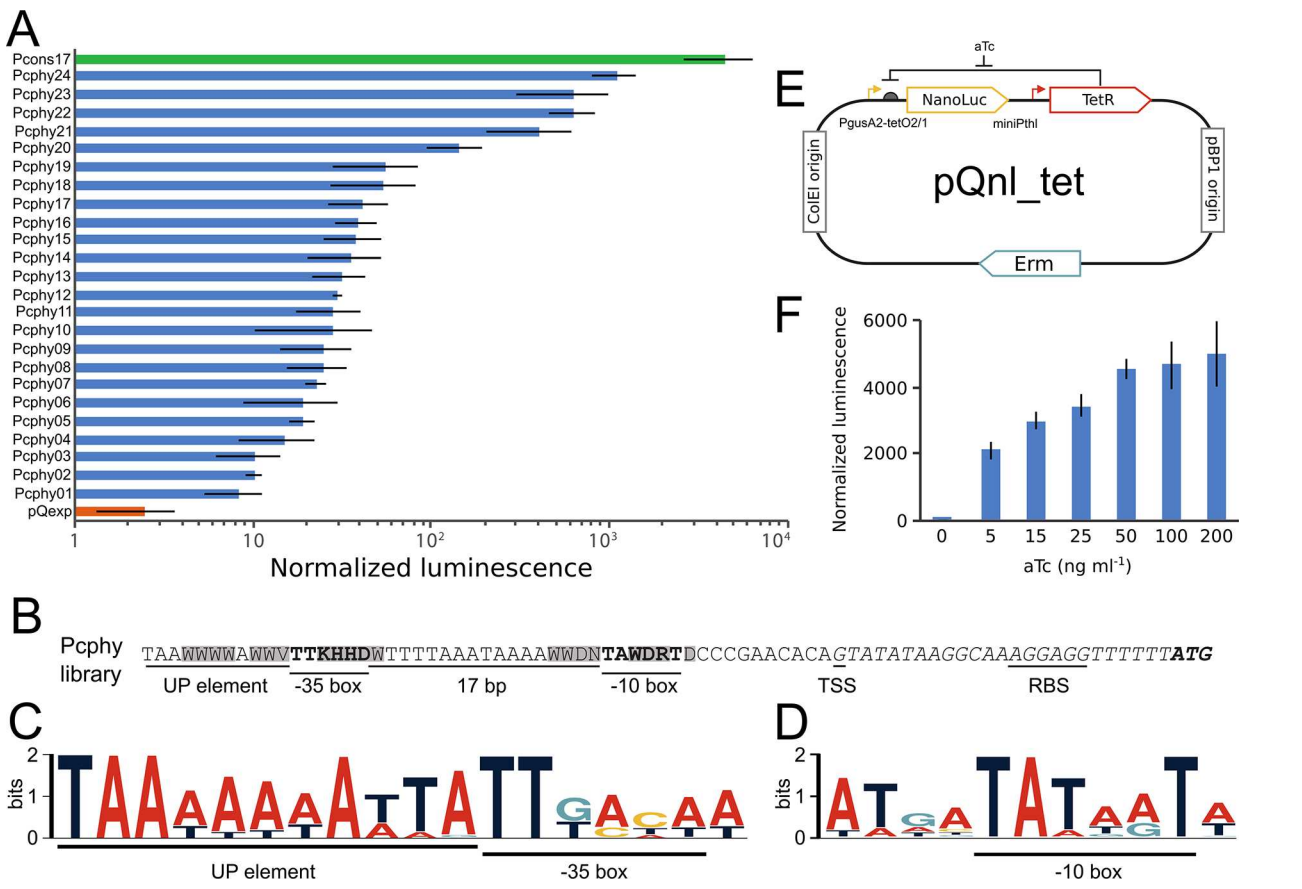


Figure 4. Modulation of gene expression in *C. phytofermentans* using constitutive and inducible promoters. (A) Promoter strengths of Pcons17, Pcphy1−24, and minus-NanoLuc control (pQexp) measured as normalized luminescence. (B) Nucleotide sequence showing degenerate positions used to build the Pcphy promoter library. (C, D) Promoter-strength-weighted sequence motifs of (C) UP element and −35 box and (D) −10 box. (E) Diagram of pQnl_tet plasmid for aTc-regulated gene expression in *C. phytofermentans*. (F) Induction of NanoLuc gene expression in *C. phytofermentans* expressing pQnl_tet measured as normalized luminescence at different aTc concentrations. (A, F) Culture luminescence was normalized to OD_{600}; bars show means ± SD of triplicate cultures. Degenerate nucleotides: W, A/T; R, A/G; K, G/T; V, A/C/G; H, A/C/T; D, A/G/T. Abbreviations: TSS, transcription start site; RBS, ribosome binding site; TetR, tetracycline repressor; aTc, anhydrotetracycline; OD_{600}, optical density at 600 nm; SD, standard deviation.

repression, $p = 1.79 \times 10^{-3}$) or g-nl2 (8-fold repression, $p = 1.94 \times 10^{-3}$) relative to pQdC12a (Figure 5B). Mean NanoLuc expression with g-nl1 was 2.5-fold lower than with g-nl2, suggesting that preventing transcription initiation by targeting the promoter at the −10 box may be more effective than the −35 box ($p = 1.90 \times 10^{-3}$), but these differences could also reflect sequence dependence of gRNA activity. NanoLuc expression was repressed to a lesser extent without aTc using either g-nl1 (6-fold repression, $p = 2.7 \times 10^{-6}$) or g-nl2 (5-fold repression, $p = 3.2 \times 10^{-4}$) relative to pQdC12a (Figure 5B), likely due to leaky Tet repression of dCas12a. Together, these results support that dCas12a-mediated CRISPRi effectively represses *in vivo* target gene expression in *C. phytofermentans*.

We evaluated pQdC12a-mediated CRISPRi of chromosomal genes in *C. phytofermentans* by targeting dCas12a to the *cphy1326* promoter using guide g-cphy1326 (Figure 6A). The *cphy1326* and *cphy1327* genes encode phosphate acetyltransferase and acetate kinase, respectively. These two genes encode the only predicted pathway to convert acetyl-CoA to acetate in *C. phytofermentans*, and their transcription and translation are elevated under conditions of high acetate production,[5,48] leading us to propose that repressing *cphy1326* transcription should reduce acetate formation. Genome-wide mapping of transcription start sites (TSSs) in *C. phytofermentans* indicated that *cphy1326* is expressed from a primary TSS 139 bp upstream of the start codon.[36] Guide g-cphy1326 was designed to bind the −10 box (5′-TATAAT-3′), which is 10 bp upstream of the *cphy1326* TSS.

Targeting pQdC12a to the *cphy1326* promoter using guide g-cphy1326 reduced *cphy1326* transcription relative to the no-guide control (11.5-fold repression, $p = 0.021$) (Figure 6B). While *cphy1326-cphy1327* constitute a putative operon, transcription of *cphy1327* was not repressed, suggesting that it is also independently transcribed. Alternative routes for acetyl-CoA conversion were upregulated in the g-cphy1326 strain based on elevated transcription of genes encoding the primary ethanol dehydrogenase (*cphy3925*), lactate dehydrogenase (*cphy1117*), and pyruvate formate-lyase activating enzyme (*cphy2820*) (Figure 6B). Quantification of fermentation products by high-performance liquid chromatography (HPLC) showed that acetate biosynthesis was reduced in the g-cphy1326 strain (19% reduction, $p = 1.5 \times 10^{-3}$) (Figure 6C). Production of ethanol was not altered in the g-cphy1326 strain (Figure 6D), whereas yields were increased for formate (46% increase, $p = 1.2 \times 10^{-4}$) (Figure 6E) and lactate (75% increase, $p = 1.6 \times 10^{-3}$) (Figure 6F). Together, these results support that *C. phytofermentans* elevated production of

https://doi.org/10.1021/acssynbio.2c00385
ACS Synth. Biol. 2022, 11, 4077−4088

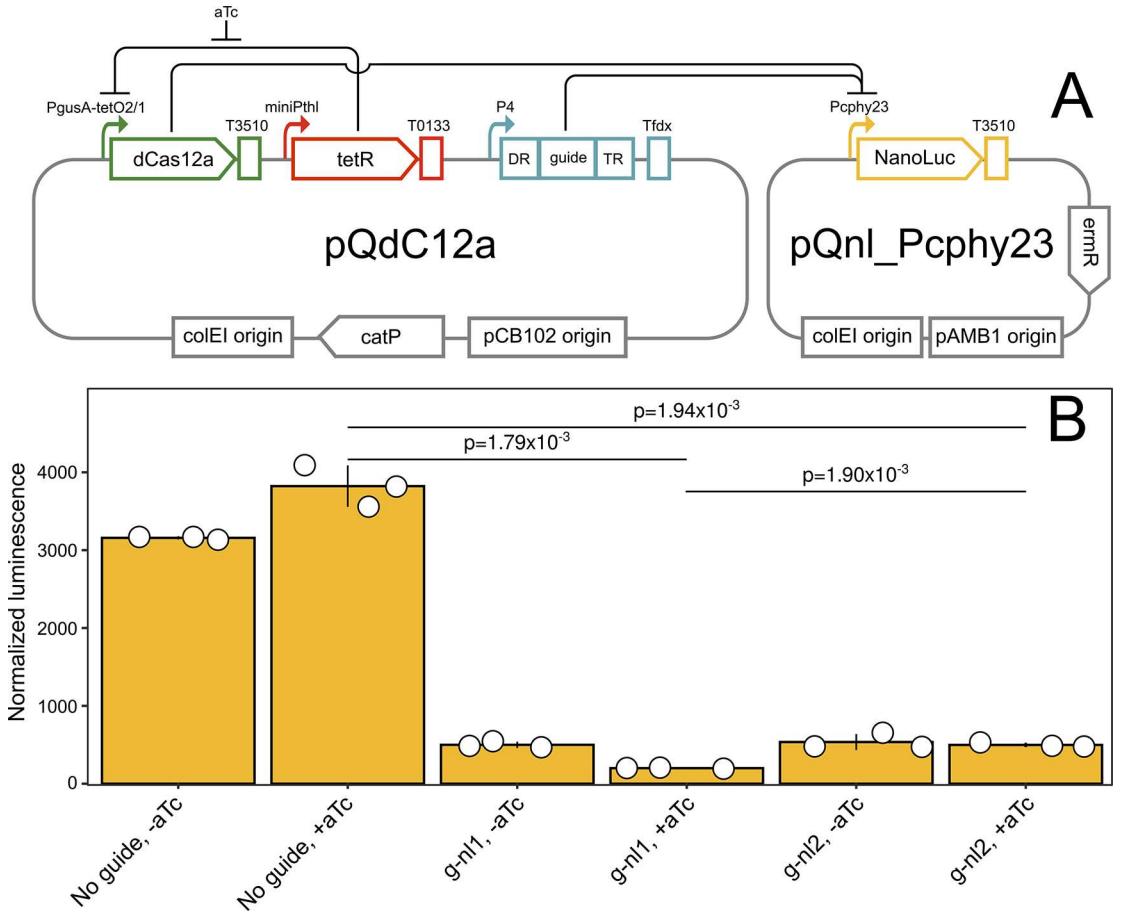


Figure 5. CRISPRi repression of reporter gene expression in *C. phytofermentans*. (A) Two-plasmid system demonstrating dCas12a repression of the NanoLuc reporter. pQdC12a has a tet-repressible dCas12a and gRNA cassette; pQnl_Pcphy23 has the Pcphy23 promoter driving NanoLuc expression. (B) Normalized luminescence of *C. phytofermentans* expressing NanoLuc (pQnl_Pcphy23) and dCas12a targeting Pcphy23 with guide g-nl1, guide g-nl2, or a no-guide control (pQdC12a). Luminescence was measured ±100 ng mL^{-1} aTc and normalized to OD_{600}. Data points are individual cultures, and bars are means ± SD of triplicate cultures. The *p* values show treatment comparisons by two-sided Student's *t* test. Abbreviations: aTc, anhydrotetracycline; T3510, transcription terminator from *C. phytofermentans cphy3510* gene; T0133, transcription terminator from *C. phytofermentans cphy0133* gene; Tfdx, transcriptional terminator from *C. pasteurianum fdx* gene; DR, direct repeat; TR, terminal repeat; OD_{600}, optical density at 600 nm; SD, standard deviation.

alternative fermentation products in response to CRISPRi-mediated repression of acetate production.

DISCUSSION

In this study, we demonstrated a framework for electrotransformation and modulation of gene expression in *C. phytofermentans* using tools that can be applied in other Clostridia. We electrotransformed *C. phytofermentans* using a simple benchtop method that does not require an anaerobic glovebox. The electroporation efficiency was high enough to test plasmids (Figure 3), screen promoters (Figure 4), and construct genetic circuits (Figure 5) but lower than that with other models such as *E. coli*, making improvements in electrotransformation desirable. However, our data support that *C. phytofermentans* electrotransformation was not limited by endogenous RMSs (Figure 2A) or DNA concentration (Figure 2C) and was not improved by cell-wall-weakening agents. We grew *C. phytofermentans* cultures in anaerobic jars and electroporated cells on the bench, making these methods simple to apply in other laboratories. Performing electroporations under more strict anaerobic conditions in a chamber might improve the transformation efficiency.

Although Clostridia are phylogenetically diverse, our results support that their molecular biology is sufficiently conserved for the pMTL plasmid system[9] to be a valuable resource to test functional plasmid elements. We identified three plasmid replicons (pCB102, pAMβ1, and pBP1) and three resistance markers (*catP*, *aad9*, and *ermB*) that function in *C. phytofermentans*. Whereas the pIM13 origin and the *tetA* resistance gene are functional in other Clostridia,[34,49] they are inactive in *C. phytofermentans*. Further, *C. phytofermentans* can be sequentially transformed with all three plasmid replicons using different resistance markers (Figure 3D), supporting that these plasmids could also be simultaneously maintained in other Clostridia. Using the pQmod-GG plasmids, cargo genes can easily be cloned into backbones with these different replicons and resistance markers using Golden Gate cloning and RFP red/white screening (Table 1). Genes can be expressed at different levels on these plasmids using the Pcphy promoter library (Figure 4A) and modulated using TetR-mediated repression (Figure 4F). These tools enable *in vivo* expression of genetic circuits composed of multiple operons whose expression needs to be individually tuned.

We united these tools into a system for CRISPRi-mediated repression in *C. phytofermentans* using a Cas12a protein

https://doi.org/10.1021/acssynbio.2c00385
ACS Synth. Biol. 2022, 11, 4077−4088

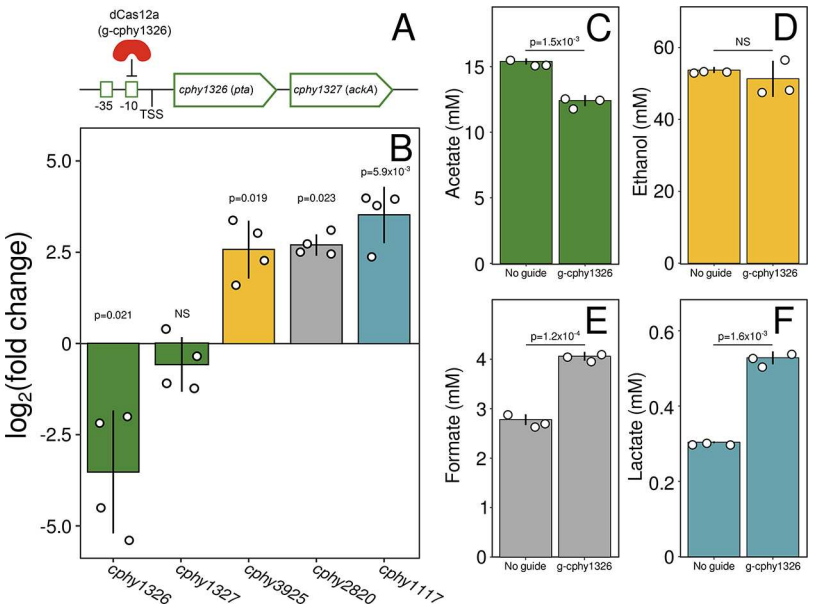


Figure 6. CRISPRi repression of fermentation gene expression in *C. phytofermentans*. (A) dCas12a was targeted to repress transcription of the acetate biosynthesis operon by binding the promoter −10 box upstream of the *cphy1326* (*pta*) gene using guide g-cphy1326. (B) Transcription of fermentation genes shown as $\log_2$(fold change) in the g-cphy1326 strain relative to the no-guide control strain measured by qRT-PCR. (C−F) Yields of fermentation products (C) acetate, (D) ethanol, (E) formate, and (F) lactate in the g-cphy1326 and no-targeting control strains measured by HPLC. (B−F) Colors show data for acetate (green), ethanol (yellow), formate (gray), and lactate (blue). Bars are means ± SD of (B) four or (C−F) three cultures; data points are individual cultures. The *p* values show treatment comparisons by two-sided Student's *t* test. Abbreviations: TSS, transcription start site; pta, phosphate acetyltransferase; ackA, acetate kinase; NS, not significant; SD, standard deviation; HPLC, high-performance liquid chromatography.

(dLbCas12a) from another Lachnospiraceae species, the first demonstration of CRISPR using this protein in bacteria.[19] CRISPRi is particularly useful to study genes that may have essential roles, as when redirecting carbon flux in central metabolism. As a first step toward redirecting carbon flux in *C. phytofermentans*, we showed that repressing genes for acetate biosynthesis reduced the acetate yields and increased the yields of other fermentation products (Figure 6). Our results show that an 11.5-fold repression in *cphy1326* expression translated into only a 19% reduction in acetate formation, which could be augmented either through greater repression of *cphy1326* using multiple guides or simultaneous targeting of multiple genes. Fortunately, dCas12a simplifies multiplex gene repression by CRISPRi due to its ability to process tandem guides without sacrificing efficiency relative to individual guides.[50]

The results of this study provide a foundation for construction of genetic circuits with experimentally modulated gene expression in *C. phytofermentans*. These approaches complement previous technologies to study *C. phytofermentans* genetics using targetron-based gene inactivation,[30] large-scale genome insertion and deletion,[31] and *in vivo* directed evolution.[51] Cas12a could be used to make genomic changes in *C. phytofermentans*, as has been demonstrated in some other Clostridia.[45] In the future, these strategies could be generally applied to other Clostridia, including soil species with important roles in environmental carbon cycling, intestinal commensals, and pathogens.

METHODS AND MATERIALS

Bacterial Cultivation. *C. phytofermentans* ISDg (ATCC 700394) was cultured anaerobically at 37 °C in GS2 medium[52] containing 5 g L^{-1} glucose and antibiotics at selected concentrations to maintain plasmids (Figure 3C). Cultures were incubated anaerobically in sealed jars (BD Jars, 260626) using anaerobic sachets (BD GasPaks, 260678); all experimental manipulations were performed at the bench. *E. coli* was cultured aerobically at 37 °C in LB medium containing 250 μg mL^{-1} erythromycin, 100 μg mL^{-1} spectinomycin, 25 μg mL^{-1} chloramphenicol, or 15 μg mL^{-1} tetracycline to select for plasmids.

Genome Methylation and DNA Restriction. We identified DNA methylation patterns in the *C. phytofermentans* ISDg genome using SMRT sequencing[53] and RIMS-seq.[23] SMRT DNA sequencing was performed at the U.S. Department of Energy Joint Genome Institute (DOE JGI Project ID 1100006), and sequence files are available in the NCBI SRA (accession code SRP123768). RIMS-seq was performed at Genoscope-CEA, and sequences files are available in the European Nucleotide Archive (accession code ERA17973853).

To test whether *C. phytofermentans* cell lysate restricts plasmid DNA, *C. phytofermentans* lysate was prepared by centrifuging an overnight culture and freezing the cell pellet at −80 °C for 2 h. The cell pellet was resuspended in lysis buffer (50 mM phosphate buffer, pH 8, 500 mM NaCl, 15% glycerol, 1 mM pefablock (Sigma 76307)), incubated with 1 mg mL^{-1} lysozyme (Novagen 71230) for 30 min at room temperature, and sonicated (Cole-Parmer Vibracell CV33) four times for 30 s at 28% with 10 s between sonications. Plasmid DNA of pQexp, pMTL83353, and pBR322 (Table 1) were isolated from DNA-methylation-deficient *E. coli* strain ER2796[26] and linearized with *Nhe*I, and 1 μg of plasmid DNA was incubated with and without 250 μg mL^{-1} *C. phytofermentans* lysate overnight at 37 °C. DNA was resolved by electrophoresis to visualize whether incubation with cell lysate resulted in plasmid digestion.

https://doi.org/10.1021/acssynbio.2c00385
ACS Synth. Biol. 2022, 11, 4077−4088

C. *phytofermentans* Electroporation. To prepare *C. phytofermentans* cells for electroporation, 10 mL of late log culture was transferred into 40 mL of fresh medium and grown for 6 h to OD_{600} 0.6−0.7. Cells were centrifuged (6245 *g*, 7 min, 4 °C), washed once with 4 mL of ice-cold ETM buffer (1 mM $MgCl_2$, 256 mM sucrose, 5 mM sodium phosphate buffer, pH 7.4), and resuspended in 200 μL of ETM buffer. A 50 μL aliquot of cells was transferred to a prechilled 0.1 cm gap electroporation cuvette, mixed with plasmid DNA (3 μg of DNA unless noted otherwise), and incubated on ice for 5 min. Cells were electroporated using a Bio-Rad Genepulser II (1.2 kV, 25 μF), immediately diluted with 1 mL of GS2 medium, and incubated anaerobically at 37 °C for 15 h. Cells were plated on GS2 agar plates containing the appropriate antibiotic selection and cultured for 6 days. Transformant colonies were picked and transferred to liquid GS2 medium supplemented with the appropriate antibiotic, and the presence of plasmid DNA in the transformants was confirmed by PCR using 1 μL of log phase culture as template and primers shown in Table S1.

Plasmid Construction. The sequences of all primers used in this study are shown in Table S1. DNA samples and sequences of plasmids from this study have been deposited in Addgene. pMTL plasmids (Table 1) were modified using double digests of either *Asc*I plus *Fse*I to swap Gram-positive replicons or *Fse*I plus *Pme*I to swap antibiotic genes (Figure 3A). To create a set of erythromycin resistance plasmids to test Gram-positive origins, the pIM13 origin of pMTL85241 was replaced with either the pBP1 origin from pMTL82254 to yield pQmod2E, the pCB102 origin from pMTL83353 to yield pQmod3E, or the pCD6 origin from pMTL84422 to yield pQmod4E (Table 1). To test antibiotic resistance genes, the *ermB* gene of pQmod2E was replaced with either the *catP* gene of pMTL85141 to yield pQmod2C, the *aad9* gene of pMTL83353 to yield pQmod2S, or the *tetA* gene of pMTL84422 to yield pQmod2T (Table 1). A set of Golden Gate-compatible plasmids containing an RFP cassette for red/white screening flanked by *Bsa*I sites (pQmod-GG plasmids) was constructed by PCR amplification of the J04450 biobrick RFP cassette from pSB1C3 using primers PR66/PR67 and cloning of the resulting product between the *Sac*I and *Nhe*I sites of the multiple cloning site (Figure S1).

We designed the Pcons17 promoter using the consensus UP, −35, and −10 elements based on *C. phytofermentans* genome-wide transcription start site (TSS) data.[36] The sequences between the −35 and −10 elements were selected by analyzing the *C. phytofermentans* ISDg genome sequence (GenBank CP000885.1) with PePPER[37] to produce a consensus sequence.[38] The 5′ UTR and ribosome binding site (RBS) upstream of NanoLuc were designed to maximize translation.[39] The NanoLuc gene was PCR-amplified (primers PR1/PR2) from pNL1.1[Nluc] (Promega). The vector backbone to express the NanoLuc gene was built by PCR amplification of the *erm* gene (primers PR3/PR4), pAMβ1 replicon (primers PR5/PR6), and pUC replicon (primers PR7/PR8) of pQexp.[27] The fragments were assembled with the NanoLuc PCR fragment by four-part Golden Gate cloning using *Bsm*BI, yielding pQnl_Pcons17 (Table 1). pATmin-GG was constructed by Gibson assembly from PCR of pQnl_Pcons17 with PR70/PR71 and of pQmod2C-GG with PR72/PR73.

The Pcphy plasmid library was constructed by PCR amplification of the entire pQnl_Pcons17 sequence using outward-facing primers (PR48/PR49) to introduce degeneracy to the Pcons17 promoter, followed by *Dpn*I digestion. The PCR product was digested with *Bsa*I, self-ligated, and transformed into *E. coli*. After 75 min recovery in LB medium, the transformed *E. coli* cells were transferred into liquid LB medium containing erythromycin and incubated at 37 °C for 16 h, and plasmid DNA was isolated to yield the pQnl_Pcphy plasmid library. Promoter diversity in the library was measured by transforming it into *E. coli* and sequencing 20 individual clones (primers OK2ML9F and Seq_NanoLuc_rv), which confirmed that no two promoters had identical sequences. *C. phytofermentans* was transformed with the pQnl_Pcphy library, and NanoLuc expression of individual clones was quantified by luciferase assay.

To assemble pQnl_tet for TetR-mediated control of gene expression, the NanoLuc gene was PCR-amplified from pQnl_Pcons17 (primers PR63/PR64), generating a product containing NanoLuc fused to the PGusA2-TetO2/1 promoter. The *tetR* gene (Genbank accession code EEA9527598.1) was purchased as a gBlock DNA fragment (Integrated DNA Technologies, Coralville, IA, USA) and PCR-amplified (primers PR58/PR59) to generate a product of the miniPthl promoter driving *tetR* expression. Subsequently, pQnl_tet was assembled by a three-fragment Golden Gate ligation reaction using *Bsa*I to insert PgusA2-TetO2/1-NanoLuc and miniPthl-*tetR* into pQmod2E-GG.

pQdC12a was assembled by Golden Gate cloning with two fragments of LbCpf1/LbCas12a from pY109,[17] simultaneously incorporating the D832A mutation and ligating to cassettes containing miniPthl-*tetR* and gRNA expression cassettes. The dCas12a gene was fused to the aTc-inducible PgusA2-tetO2/1 promoter[11] and terminator from *cphy3510* by Golden Gate cloning. The gRNA cassette contains the synthetic P4 promoter,[38] extended direct repeat, *Bbs*I sites into which annealed oligos can be inserted to target dCas12a binding, 25 bp from the wild-type Cas12a terminal repeat, and the *fdx* terminator from *C. pasteurianum*.

pQdC12a was targeted to bind the Pcphy23 promoter in pQnl_Pcphy23 using guide sequences that bind either the −10 box (g-nl1) or the −35 box (g-nl2) (Figure S4 and Table S1). To repress *cphy1326* transcription, pQdC12a was targeted using guide g-cphy1326 that binds the −10 box (5-TATAAT-3), which is 10 bp upstream of the *cphy1326* TSS as previously identified by TSS mapping[36] (Table S1). To clone guide sequences, pQdC12a was digested with *Bbs*I, gel-purified, and ligated with annealed guide sequences.

Luciferase Assay. To measure luciferase expression in *C. phytofermentans* containing NanoLuc plasmids, we grew cultures to late log phase, measured the cell density (OD_{600}), and lysed the cells by addition of a 1:10 volume of PopCulture reagent (EMD Millipore) containing 25 μg mL^{-1} lysozyme followed by incubation at room temperature for 10 min. Cell lysate was mixed with an equal volume of NanoLuc Reagent (Promega N1110), and luminescence was measured at 454 nm using a Safas Xenius XMA. Luminescence measurements from triplicate cultures were normalized to the cell density (OD_{600}). Following luminescence measurement of Pcphy clones, plasmids covering a range of luminescence strengths were isolated, and the promoter regions were sequenced after amplification with primers OK2ML9F and Seq_NanoLuc_rv. Luciferase measurements with pQnl_tet and CRISPRi plasmids were conducted by growing a culture through two transfers to late log phase followed by dilution into medium supplemented with antibiotics and aTc as appropriate and

https://doi.org/10.1021/acssynbio.2c00385
ACS Synth. Biol. 2022, 11, 4077−4088

growth to late log phase before measurement of luciferase expression.

Quantitative PCR. We measured *C. phytofermentans* mRNA expression by quantitative reverse transcription PCR (qRT-PCR). Briefly, total RNA was extracted from 5 mL mid log phase (OD_{600} = 0.7) cultures using TRI reagent (Sigma 93289), and 10 μg of RNA was treated with 4 units of Turbo DNase (Ambion AM2238) for 30 min at 37 °C. One microgram of RNA was reverse-transcribed to single-stranded cDNA (Applied Biosystems 4368814). Real-time PCR amplification was conducted by quantitative PCR (qPCR) (KAPA KK4621) with the primers shown in Table S1 using an MJ Research DNA Engine Opticon II machine. Expression values are means of four cultures calculated by the threshold cycle method normalized to 16S rRNA levels.

High-Performance Liquid Chromatography. To quantify the fermentation products, cultures were grown for 5 days to ensure complete consumption of glucose in the medium. Culture supernatants were 0.2 μM filtered and collected. Fermentation products were measured using a Shimadzu Prominence LC20/SIL-20AC HPLC equipped with an RID-10A refractive index detector and a Biorad Aminex HPX-87H 300 mm × 7.8 mm column. The column was maintained at 60 °C with a mobile phase of 5 mM H_2SO_4 at a flow rate of 0.5 mL min^{-1}. The sample injection volume was 20 μL. Purified standards of the fermentation products (Sigma) were used to define elution times and generate dilution curves, which were used to quantify the fermentation products in the culture supernatants.

Sequence Analysis and Statistics. Bioinformatic prediction of putative DNA restriction modification systems in *C. phytofermentans* was conducted using REBASE.[22] Methylation analysis of SMRT sequence data was performed using SMRT Tools (protocol version = 2.2.0, method = RS Modification and Motif Analysis.1).[24] Modification quality values (QVs) were computed as $-10\ \log_{10}(p\ \text{value})$ for a modified base position based on DNA polymerase kinetics measured as interpulse durations. Modified sites were grouped into motifs using SMRT Tools MotifFinder (v1).[25] The RIMS-seq experiment and data analysis were performed as described previously[54] with motif context analysis calculated based on the C to T errors in Read 1 and Read 2 for all NNNCN contexts with N = ATCG. To evaluate the relative importance of nucleotides at different positions across promoter sequences, we generated sequence motifs[54] in which the relative abundances of sequence inputs were based upon the mean strength of each promoter in the Pcphy library (Figure S2). Statistical comparisons were performed by either Tukey's test (multiple comparisons) or Student's *t* test (pairwise comparisons).

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssynbio.2c00385.

Design of Golden Gate plasmids with dropout RFP cassettes for rapid cloning and testing of genetic elements (Figure S1), sequences and expression strengths of promoters in the Pcphy promoter library (Figure S2), relative abundances of protospacer adjacent motifs in the *C. phytofermentans* genome and in promoter regions (Figure S3), binding locations of dCas12a guides targeted to the Pcphy23 promoter driving NanoLuc expression in pQnl_Cphy23 (Figure S4), and primers used in this study (Table S1) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Andrew C. Tolonen − *Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91057 Évry, France;* orcid.org/0000-0001-5907-4504; Email: atolonen@genoscope.cns.fr

Authors

William Rostain − *Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91057 Évry, France*

Tom Zaplana − *Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91057 Évry, France*

Magali Boutard − *Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91057 Évry, France*

Chloé Baum − *Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91057 Évry, France; New England Biolabs, Inc., Ipswich, Massachusetts 01938, United States*

Sibylle Tabuteau − *Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91057 Évry, France*

Mary Sanitha − *Molecular Genetics Laboratory, Department of Genetic Engineering, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur 603 203 TN, India*

Mohandass Ramya − *Molecular Genetics Laboratory, Department of Genetic Engineering, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur 603 203 TN, India*

Adam Guss − *Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831-6038, United States*

Laurence Ettwiller − *New England Biolabs, Inc., Ipswich, Massachusetts 01938, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acssynbio.2c00385

Author Contributions

W.R., M.R., L.E., and A.C.T. conceived the study. W.R., A.G., L.E., and A.C.T. designed the experiments. W.R., T.Z., M.B., C.B., S.T., and M.S. performed the experiments. W.R., C.B., A.G., L.E., and A.C.T. analyzed the results. W.R. and ACT wrote the manuscript with contributions from all authors. All of the authors read and approved the final manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Corinne Cruaud, Shahinaz Gas, and Ioana Popescu for technical expertise. This work was supported by Genoscope-CEA, the Agence Nationale de la Recherche (Grant ANR-16-CE05-0020), Évry Genopole under the "Action Thématique Incitative Genopole" Grant (funding ATIGE 2021 No. 2698), a travel grant from SRM University, and the Oak Ridge National Laboratory at the Center for

Bioenergy Innovation, a U.S. Department of Energy (DOE) Bioenergy Research Center. PacBio methylome data were generated by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, supported by the Office of Science of the U.S. DOE under Contract DE-AC02-05CH11231.

REFERENCES

(1) Vees, C. A.; Neuendorf, C. S.; Pflügl, S. Towards continuous industrial bioprocessing with solventogenic and acetogenic clostridia: challenges, progress and perspectives. *J. Ind. Microbiol. Biotechnol.* **2020**, *47*, 753−787.

(2) Liew, F.; Martin, M. E.; Tappel, R. C.; Heijstra, B. D.; Mihalcea, C.; Köpke, M. Gas Fermentation—A Flexible Platform for Commercial Scale Production of Low-Carbon-Fuels and Chemicals from Waste and Renewable Feedstocks. *Front. Microbiol.* **2016**, *7*, 694.

(3) Guo, P.; Zhang, K.; Ma, X.; He, P. Clostridium species as probiotics: potentials and challenges. *J. Anim. Sci. Biotechnol* **2020**, *11*, 24.

(4) Tolonen, A. C.; Beauchemin, N.; Bayne, C.; Li, L.; Tan, J.; Lee, J.; Meehan, B. M.; Meisner, J.; Millet, Y.; LeBlanc, G.; Kottler, R.; Rapp, E.; Murphy, C.; Turnbaugh, P. J.; von Maltzahn, G.; Liu, C. M.; van Hylckama Vlieg, J. E. T. Synthetic glycans control gut microbiome structure and mitigate colitis in mice. *Nat. Commun.* **2022**, *13*, 1244.

(5) Tolonen, A. C.; Haas, W.; Chilaka, A. C.; Aach, J.; Gygi, S. P.; Church, G. M. Proteome-wide systems analysis of a cellulosic biofuel-producing microbe. *Mol. Syst. Biol.* **2011**, *7*, 461.

(6) Petit, E.; Coppi, M. V.; Hayes, J. C.; Tolonen, A. C.; Warnick, T.; Latouf, W. G.; Amisano, D.; Biddle, A.; Mukherjee, S.; Ivanova, N.; Lykidis, A.; Land, M.; Hauser, L.; Kyrpides, N.; Henrissat, B.; Lau, J.; Schnell, D. J.; Church, G. M.; Leschine, S. B.; Blanchard, J. L. Genome and Transcriptome of *Clostridium phytofermentans*, Catalyst for the Direct Conversion of Plant Feedstocks to Fuels. *PLoS One* **2015**, *10*, No. e0118285.

(7) Jennert, K. C.; Tardif, C.; Young, D. I.; Young, M. Gene transfer to *Clostridium cellulolyticum* ATCC 35319. *Microbiology (Reading, U. K.)* **2000**, *146* (12), 3071−3080.

(8) Olson, D. G.; Lynd, L. R. Transformation of *Clostridium thermocellum* by electroporation. *Methods Enzymol.* **2012**, *510*, 317−330.

(9) Heap, J. T.; Pennington, O. J.; Cartman, S. T.; Minton, N. P. A modular system for Clostridium shuttle plasmids. *J. Microbiol. Methods* **2009**, *78*, 79−85.

(10) Mordaka, P. M.; Heap, J. T. Stringency of Synthetic Promoter Sequences in Clostridium Revealed and Circumvented by Tuning Promoter Library Mutation Rates. *ACS Synth. Biol.* **2018**, *7*, 672−681.

(11) Dong, H.; Tao, W.; Zhang, Y.; Li, Y. Development of an anhydrotetracycline-inducible gene expression system for solvent-producing *Clostridium acetobutylicum*: A useful tool for strain engineering. *Metab. Eng.* **2012**, *14*, 59−67.

(12) Qi, L. S.; Larson, M. H.; Gilbert, L. A.; Doudna, J. A.; Weissman, J. S.; Arkin, A. P.; Lim, W. A. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **2013**, *152*, 1173−1183.

(13) Bruder, M. R.; Pyne, M. E.; Moo-Young, M.; Chung, D. A.; Chou, C. P. Extending CRISPR-Cas9 Technology from Genome Editing to Transcriptional Engineering in the Genus Clostridium. *Appl. Environ. Microbiol.* **2016**, *82*, 6109−6119.

(14) Woolston, B. M.; Emerson, D. F.; Currie, D. H.; Stephanopoulos, G. Rediverting carbon flux in *Clostridium ljungdahlii* using CRISPR interference (CRISPRi). *Metab. Eng.* **2018**, *48*, 243−253.

(15) Fackler, N.; Heffernan, J.; Juminaga, A.; Doser, D.; Nagaraju, S.; Gonzalez-Garcia, R. A.; Simpson, S. D.; Marcellin, E.; Köpke, M. Transcriptional control of *Clostridium autoethanogenum* using CRISP-Ri. *Synth. Biol.* **2021**, *6* (1), ysab008.

(16) Zhao, R.; Liu, Y.; Zhang, H.; Chai, C.; Wang, J.; Jiang, W.; Gu, Y. CRISPR-Cas12a-Mediated Gene Deletion and Regulation in *Clostridium ljungdahlii* and Its Application in Carbon Flux Redirection in Synthesis Gas Fermentation. *ACS Synth. Biol.* **2019**, *8*, 2270−2279.

(17) Zetsche, B.; Gootenberg, J. S.; Abudayyeh, O. O.; Slaymaker, I. M.; Makarova, K. S.; Essletzbichler, P.; Volz, S. E.; Joung, J.; van der Oost, J.; Regev, A.; Koonin, E. V.; Zhang, F. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **2015**, *163*, 759−771.

(18) Tak, Y. E.; Kleinstiver, B. P.; Nuñez, J. K.; Hsu, J. Y.; Horng, J. E.; Gong, J.; Weissman, J. S.; Joung, J. K. Inducible and multiplex gene regulation using CRISPR-Cpf1-based transcription factors. *Nat. Methods* **2017**, *14*, 1163−1166.

(19) Meliawati, M.; Schilling, C.; Schmid, J. Recent advances of Cas12a applications in bacteria. *Appl. Microbiol. Biotechnol.* **2021**, *105*, 2981−2990.

(20) Cui, G.; Hong, W.; Zhang, J.; Li, W.; Feng, Y.; Liu, Y.; Cui, Q. Targeted gene engineering in *Clostridium cellulolyticum* H10 without methylation. *J. Microbiol. Methods* **2012**, *89*, 201−208.

(21) Dong, H.; Zhang, Y.; Dai, Z.; Li, Y. Engineering clostridium strain to accept unmethylated DNA. *PLoS One* **2010**, *5*, No. e9038.

(22) Roberts, R. J.; Vincze, T.; Posfai, J.; Macelis, D. REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.* **2015**, *43*, D298−299.

(23) Baum, C.; Lin, Y.-C.; Fomenkov, A.; Anton, B. P.; Chen, L.; Yan, B.; Evans, T. C.; Roberts, R. J.; Tolonen, A. C.; Ettwiller, L. Rapid identification of methylase specificity (RIMS-seq) jointly identifies methylated motifs and generates shotgun sequencing of bacterial genomes. *Nucleic Acids Res.* **2021**, *49*, No. e113.

(24) Flusberg, B. A.; Webster, D. R.; Lee, J. H.; Travers, K. J.; Olivares, E. C.; Clark, T. A.; Korlach, J.; Turner, S. W. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* **2010**, *7*, 461−465.

(25) Li, T.; Zhang, X.; Luo, F.; Wu, F.-X.; Wang, J. MultiMotif-Maker: A Multi-Thread Tool for Identifying DNA Methylation Motifs from Pacbio Reads. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **2020**, *17*, 220−225.

(26) Anton, B. P.; Mongodin, E. F.; Agrawal, S.; Fomenkov, A.; Byrd, D. R.; Roberts, R. J.; Raleigh, E. A. Complete Genome Sequence of ER2796, a DNA Methyltransferase-Deficient Strain of *Escherichia coli* K-12. *PLoS One* **2015**, *10*, No. e0127446.

(27) Tolonen, A. C.; Chilaka, A. C.; Church, G. M. Targeted gene inactivation in *Clostridium phytofermentans* shows that cellulose degradation requires the family 9 hydrolase Cphy3367. *Mol. Microbiol.* **2009**, *74*, 1300−1313.

(28) Tolonen, A. C.; Cerisy, T.; El-Sayyed, H.; Boutard, M.; Salanoubat, M.; Church, G. M. Fungal lysis by a soil bacterium fermenting cellulose. *Environ. Microbiol.* **2015**, *17*, 2618−2627.

(29) Tolonen, A. C.; Zuroff, T. R.; Ramya, M.; Boutard, M.; Cerisy, T.; Curtis, W. R. Physiology, Genomics, and Pathway Engineering of an Ethanol-Tolerant Strain of *Clostridium phytofermentans*. *Appl. Environ. Microbiol.* **2015**, *81*, 5440−5448.

(30) Cerisy, T.; Iglesias, A.; Rostain, W.; Boutard, M.; Pelle, C.; Perret, A.; Salanoubat, M.; Fierobe, H.-P.; Tolonen, A. C. ABC Transporters Required for Hexose Uptake by *Clostridium phytofermentans*. *J. Bacteriol.* **2019**, *201* (15), No. e00241-19.

(31) Cerisy, T.; Rostain, W.; Chhun, A.; Boutard, M.; Salanoubat, M.; Tolonen, A. C. A Targetron-Recombinase System for Large-Scale Genome Engineering of Clostridia. *mSphere* **2019**, *4* (6), e00710-19.

(32) Bhattacharjee, D.; Sorg, J. A. Factors and Conditions That Impact Electroporation of *Clostridioides difficile* Strains. *mSphere* **2020**, *5* (2), No. e00941-19.

(33) Hammes, W.; Schleifer, K. H.; Kandler, O. Mode of action of glycine on the biosynthesis of peptidoglycan. *J. Bacteriol.* **1973**, *116*, 1029−1053.

(34) Sloan, J.; McMurry, L. M.; Lyras, D.; Levy, S. B.; Rood, J. I. The Clostridium perfringens Tet P determinant comprises two overlapping genes: tetA(P), which mediates active tetracycline efflux, and tetB(P), which is related to the ribosomal protection family of tetracycline-resistance determinants. *Mol. Microbiol.* **1994**, *11*, 403−415.

https://doi.org/10.1021/acssynbio.2c00385
ACS Synth. Biol. 2022, 11, 4077−4088

(35) Potapov, V.; Ong, J. L.; Kucera, R. B.; Langhorst, B. W.; Bilotti, K.; Pryor, J. M.; Cantor, E. J.; Canton, B.; Knight, T. F.; Evans, T. C.; Lohman, G. J. S. Comprehensive Profiling of Four Base Overhang Ligation Fidelity by T4 DNA Ligase and Application to DNA Assembly. *ACS Synth. Biol.* **2018**, *7*, 2665−2674.
(36) Boutard, M.; Ettwiller, L.; Cerisy, T.; Alberti, A.; Labadie, K.; Salanoubat, M.; Schildkraut, I.; Tolonen, A. C. Global repositioning of transcription start sites in a plant-fermenting bacterium. *Nat. Commun.* **2016**, *7*, 13783.
(37) de Jong, A.; Pietersma, H.; Cordes, M.; Kuipers, O. P.; Kok, J. PePPER: a webserver for prediction of prokaryote promoter elements and regulons. *BMC Genomics* **2012**, *13*, 299.
(38) Xu, T.; Li, Y.; Shi, Z.; Hemme, C. L.; Li, Y.; Zhu, Y.; Van Nostrand, J. D.; He, Z.; Zhou, J. Efficient Genome Editing in *Clostridium cellulolyticum* via CRISPR-Cas9 Nickase. *Appl. Environ. Microbiol.* **2015**, *81*, 4423−4431.
(39) Salis, H. M.; Mirsky, E. A.; Voigt, C. A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **2009**, *27*, 946−950.
(40) Mazo-Vargas, A.; Park, H.; Aydin, M.; Buchler, N. E. Measuring fast gene dynamics in single cells with time-lapse luminescence microscopy. *Mol. Biol. Cell* **2014**, *25*, 3699−3708.
(41) Ross, W.; Gosink, K. K.; Salomon, J.; Igarashi, K.; Zou, C.; Ishihama, A.; Severinov, K.; Gourse, R. L. A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science* **1993**, *262*, 1407−1413.
(42) Graves, M. C.; Rabinowitz, J. C. *In vivo* and *in vitro* transcription of the *Clostridium pasteurianum* ferredoxin gene. Evidence for "extended" promoter elements in gram-positive organisms. *J. Biol. Chem.* **1986**, *261*, 11409−11415.
(43) Gao, L.; Cox, D. B. T.; Yan, W. X.; Manteiga, J. C.; Schneider, M. W.; Yamano, T.; Nishimasu, H.; Nureki, O.; Crosetto, N.; Zhang, F. Engineered Cpf1 variants with altered PAM specificities. *Nat. Biotechnol.* **2017**, *35*, 789−792.
(44) Wasels, F.; Jean-Marie, J.; Collas, F.; López-Contreras, A. M.; Lopes Ferreira, N. A two-plasmid inducible CRISPR/Cas9 genome editing tool for *Clostridium acetobutylicum*. *J. Microbiol Methods* **2017**, *140*, 5−11.
(45) Hong, W.; Zhang, J.; Cui, G.; Wang, L.; Wang, Y. Multiplexed CRISPR-Cpf1-Mediated Genome Editing in *Clostridium difficile* toward the Understanding of Pathogenesis of C. difficile Infection. *ACS Synth. Biol.* **2018**, *7*, 1588−1600.
(46) Liao, C.; Slotkowski, R. A.; Achmedov, T.; Beisel, C. L. The *Francisella novicida* Cas12a is sensitive to the structure downstream of the terminal repeat in CRISPR arrays. *RNA Biol.* **2019**, *16*, 404−412.
(47) Zhang, X.; Wang, J.; Cheng, Q.; Zheng, X.; Zhao, G.; Wang, J. Multiplex gene regulation by CRISPR-ddCpf1. *Cell Discovery* **2017**, *3*, 17018.
(48) Boutard, M.; Cerisy, T.; Nogue, P.-Y.; Alberti, A.; Weissenbach, J.; Salanoubat, M.; Tolonen, A. C. Functional diversity of carbohydrate-active enzymes enabling a bacterium to ferment plant biomass. *PLoS Genet.* **2014**, *10*, No. e1004773.
(49) Mermelstein, L. D.; Welker, N. E.; Bennett, G. N.; Papoutsakis, E. T. Expression of cloned homologous fermentative genes in *Clostridium acetobutylicum* ATCC 824. *Bio/Technology* **1992**, *10*, 190−195.
(50) Ao, X.; Yao, Y.; Li, T.; Yang, T.-T.; Dong, X.; Zheng, Z.-T.; Chen, G.-Q.; Wu, Q.; Guo, Y. A Multiplex Genome Editing Method for *Escherichia coli* Based on CRISPR-Cas12a. *Front. Microbiol.* **2018**, *9*, 2307.
(51) Cerisy, T.; Souterre, T.; Torres-Romero, I.; Boutard, M.; Dubois, I.; Patrouix, J.; Labadie, K.; Berrabah, W.; Salanoubat, M.; Doring, V.; Tolonen, A. Evolution of a biomass-fermenting bacterium to resist lignin phenolics. *Appl. Environ. Microbiol.* **2017**, *83* (11), No. e00289-17.
(52) Cavedon, K.; Leschine, S. B.; Canale-Parola, E. Cellulase system of a free-living, mesophilic clostridium (strain C7). *J. Bacteriol.* **1990**, *172*, 4222−4230.
(53) Eid, J.; Fehr, A.; Gray, J.; Luong, K.; Lyle, J.; Otto, G.; Peluso, P.; Rank, D.; Baybayan, P.; Bettman, B.; Bibillo, A.; Bjornson, K.; Chaudhuri, B.; Christians, F.; Cicero, R.; Clark, S.; Dalal, R.; Dewinter, A.; Dixon, J.; Foquet, M.; Gaertner, A.; Hardenbol, P.; Heiner, C.; Hester, K.; Holden, D.; Kearns, G.; Kong, X.; Kuse, R.; Lacroix, Y.; Lin, S.; Lundquist, P.; Ma, C.; Marks, P.; Maxham, M.; Murphy, D.; Park, I.; Pham, T.; Phillips, M.; Roy, J.; Sebra, R.; Shen, G.; Sorenson, J.; Tomaney, A.; Travers, K.; Trulson, M.; Vieceli, J.; Wegener, J.; Wu, D.; Yang, A.; Zaccarin, D.; Zhao, P.; Zhong, F.; Korlach, J.; Turner, S. Real-time DNA sequencing from single polymerase molecules. *Science* **2009**, *323*, 133−138.
(54) Crooks, G. E.; Hon, G.; Chandonia, J.-M.; Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **2004**, *14*, 1188−1190.

Recommended by ACS

PhiReX 2.0: A Programmable and Red Light-Regulated CRISPR-dCas9 System for the Activation of Endogenous Genes in *Saccharomyces cerevisiae*
Fabian Machens, Lena Hochrein, *et al.*
APRIL 04, 2023
ACS SYNTHETIC BIOLOGY
READ

A Suite of New Strain Construction Vectors for Gene Expression Knockdown in Budding Yeast
Christian A. Shively, Robi D. Mitra, *et al.*
JANUARY 17, 2023
ACS SYNTHETIC BIOLOGY
READ

In Vivo Genome Editing in Type I and II Methanotrophs Using a CRISPR/Cas9 System
Bashir L. Rumah, Ying Zhang, *et al.*
JANUARY 23, 2023
ACS SYNTHETIC BIOLOGY
READ

Autoinduction AND Gate Inhibits Cell Lysis to Enhance Protein Production in *Bacillus subtilis* Controlled by Population Density and Cell Physiological State
Kuidong Xu, Song Liu, *et al.*
MARCH 07, 2023
ACS SYNTHETIC BIOLOGY
READ

Get More Suggestions >

https://doi.org/10.1021/acssynbio.2c00385
ACS Synth. Biol. 2022, 11, 4077−4088

frontiers | Frontiers in Bioengineering and Biotechnology

TYPE Review
PUBLISHED 04 January 2024
DOI 10.3389/fbioe.2023.1324396

Check for updates

OPEN ACCESS

EDITED BY
Donatella Cimini,
University of Campania Luigi Vanvitelli, Italy

REVIEWED BY
Zhengming Zhu,
Nanjing Tech University, China
Zhenshang Xu,
Qilu University of Technology, China

*CORRESPONDENCE
Andrew C. Tolonen,
atolonen@genoscope.cns.fr

RECEIVED 19 October 2023
ACCEPTED 05 December 2023
PUBLISHED 04 January 2024

CITATION
Zaplana T, Miele S and Tolonen AC (2024), *Lachnospiraceae* are emerging industrial biocatalysts and biotherapeutics.
Front. Bioeng. Biotechnol. 11:1324396.
doi: 10.3389/fbioe.2023.1324396

COPYRIGHT
© 2024 Zaplana, Miele and Tolonen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Lachnospiraceae are emerging industrial biocatalysts and biotherapeutics

Tom Zaplana, Solange Miele and Andrew C. Tolonen*

Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, University of Evry, Université Paris-Saclay, Evry, France

The *Lachnospiraceae* is a family of anaerobic bacteria in the class Clostridia with potential to advance the bio-economy and intestinal therapeutics. Some species of *Lachnospiraceae* metabolize abundant, low-cost feedstocks such as lignocellulose and carbon dioxide into value-added chemicals. Others are among the dominant species of the human colon and animal rumen, where they ferment dietary fiber to promote healthy gut and immune function. Here, we summarize recent studies of the physiology, cultivation, and genetics of *Lachnospiraceae*, highlighting their wide substrate utilization and metabolic products with industrial applications. We examine studies of these bacteria as Live Biotherapeutic Products (LBPs), focusing on *in vivo* disease models and clinical studies using them to treat infection, inflammation, metabolic syndrome, and cancer. We discuss key research areas including elucidation of intra-specific diversity and genetic modification of candidate strains that will facilitate the exploitation of *Lachnospiraceae* in industry and medicine.

KEYWORDS

***Lachnospiraceae*, microbial biotechnology, microbiome, biocatalyst, live biotherapeutics, Clostridia**

Introduction

The *Lachnospiraceae*, from the ancient Greek "lachnos" (wooly hair) and "spira" (coil, twist), is a family of anaerobic, mesophilic bacteria with Gram-positive ultrastructure. *Lachnospiraceae*, which generally correspond to Clostridia cluster XIVa (Whitman, 2009), inhabit diverse ecosystems. Host-associated species are found in the gastrointestinal tracts of humans (Gosalbes et al., 2011), mice (Meehan and Beiko, 2014), insects (Vera-Ponce de Leon et al., 2022), and ruminants (Seshadri et al., 2018) as well as the human gingival crevice (Antezack et al., 2021). Metagenomic studies have shown *Lachnospiraceae* account for 10%–45% of the total bacteria in feces of healthy adults (Liu C. et al., 2021) and have a life-long association with humans; they colonize the guts of infants and are enriched in the fecal microbiomes of long-living (>90 years old) individuals (Kong et al., 2016). Other *Lachnospiraceae* live in anaerobic soil (Hengstmann et al., 1999) where they recycle plant matter and mediate biological soil disinfestation, a pesticide-free method to control soil-borne pathogens (Huang et al., 2019). *Lachnospiraceae* also inhabit aquatic sediments (Lomans et al., 2001; Dai et al., 2016), Antarctic green snow (Smirnova et al., 2021), wastewater (McLellan et al., 2013), and deep sea hydrothermal vents (Schouw et al., 2016).

Recent interest in the important roles of *Lachnospiraceae* in gut and environmental ecosystems has led to advances in the genomics, cultivation, and genetic manipulation of these bacteria. Here, we examine these advances and explore how they have set the stage for applying *Lachnospiraceae* in industrial and medical biotechnology. Many metagenomic

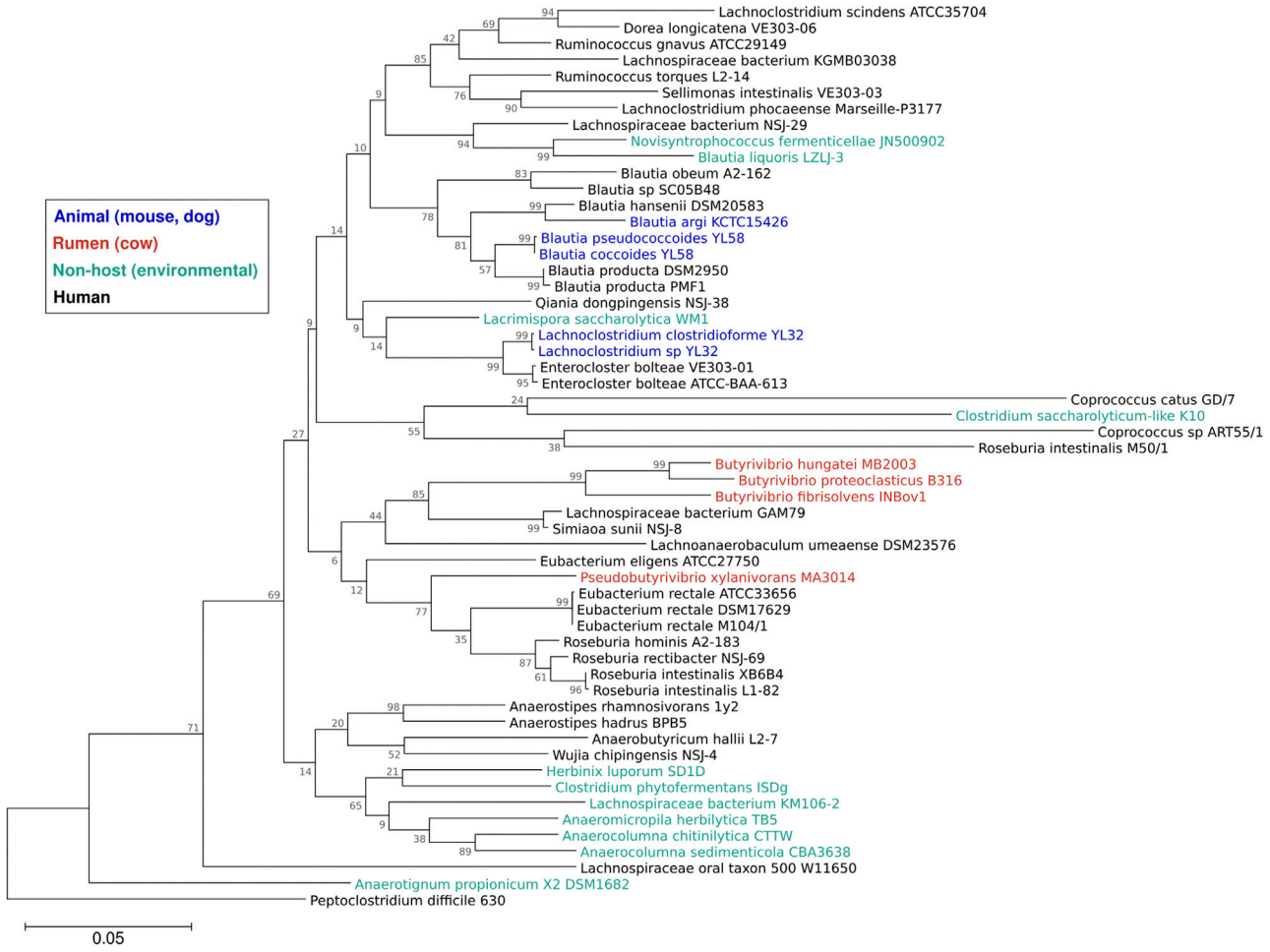


FIGURE 1
Phylogeny of the family *Lachnospiraceae*. Sequences of the 16S rRNA gene from the 57 *Lachnospiraceae* finished genomes in the IMG database were aligned using Muscle (Edgar, 2004) and a maximum likelihood tree (1,000 bootstraps) was constructed using Mega11 software (Tamura et al., 2021) with *Clostridioides (Peptoclostridium) difficile* 630 as the outgroup. Colors show strain isolation sites: animal (dog, mouse) feces (blue), cow rumen (red), human feces or oral cavity (black), or non-host environmental (aqua). Branch numbers are bootstrap percentages, branch lengths are substitution per site.

studies comparing the fecal communities of healthy and diseased subjects have drawn associations between *Lachnospiraceae* and human health. However, associations from fecal metagenomics are sometimes difficult to interpret or conflicting, likely due to differences in study design and strain-specific differences, and have been reviewed previously (Vacca et al., 2020; Liu X. et al., 2021). Thus, we focus on insights gained from preclinical, rodent models and human clinical studies involving administration of live *Lachnospiraceae*. We discuss opportunities and current needs to develop *Lachnospiraceae* to produce value-added biochemicals from low-cost feedstocks and as live biotherapeutic products (LBPs).

Phylogeny and genomes

Our knowledge of *Lachnospiraceae* diversity and genomics has greatly expanded over the past decade. In 2014, the NCBI taxonomy of *Lachnospiraceae* included 24 genera (Meehan and Beiko, 2014), which by 2023 had increased to 118 genera with 1,941 species (NCBI Genome Datasets, 2023). The number of *Lachnospiraceae* genomes has similarly increased. In 2014, NCBI included 30 *Lachnospiraceae* genomes (Meehan and Beiko, 2014), which reached 201 genomes in 2023 (NCBI Genome Datasets, 2023). The Integrated Microbial Genomes & Microbiomes (IMG) portal (Chen I.-M. A. et al., 2023) includes 1,292 *Lachnospiraceae* genomic sequences, of which 57 are finished genomes. Most of the finished *Lachnospiraceae* genomes are of strains isolated from human feces or oral cavity, with strains from the cow rumen, animal feces, and environmental isolates forming clusters in the *Lachnospiraceae* phylogeny (Figure 1).

The GC content of *Lachnospiraceae* genomes varies between 35%–50%, with members of the same genus typically sharing similar GC levels (Sorbara et al., 2020). The distribution of *Lachnospiraceae* genome sizes (mean 4.06 Mb, median 3.59 Mb) is similar across habitats (Figure 2A). *Herbinix luporum* SD1D, a cellulose-degrading strain isolated from a biogas reactor, has the smallest genome of 2.6 Mb encoding 2,632 genes (Koeck et al., 2016). *Lachnoclostridium clostridioforme* YL32, from the mouse gut, has the largest genome of 7.2 Mb with 7,735 genes (Garzetti et al., 2017). Among these finished

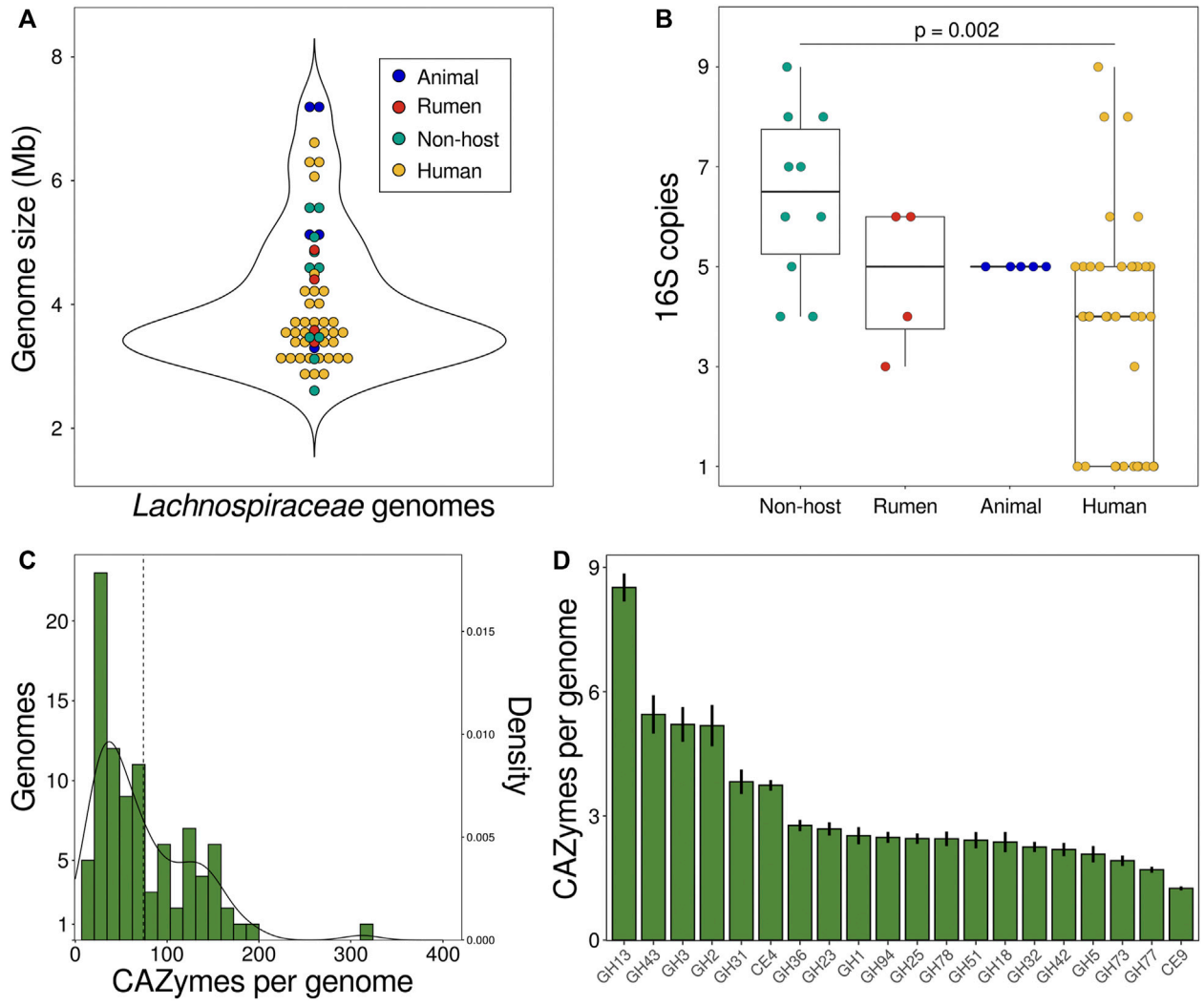


FIGURE 2
Characteristics of *Lachnospiraceae* genomes. **(A)** Genomes sizes and **(B)** number of copies of the 16S gene in 57 finished *Lachnospiraceae* genomes in the IMG database. Each data point is a genome and colors show strain isolation sites: animal (dog, mouse) feces (blue), cow rumen (red), human feces or oral cavity (yellow), or non-host environmental (aqua). **(C)** Number of CAZyme genes encoded by *Lachnospiraceae* genomes. The dashed line is the mean number of CAZymes per genome and the curve shows the kernel density estimation. **(D)** Number of CAZyme genes per genome for the 20 most abundant CAZyme families summed across *Lachnospiraceae* genomes. Bars show the mean ± SEM. Data in **(C,D)** is for degradative enzymes (glycoside hydrolases, carbohydrate esterases, pectin lyases) for 98 species in the CAZy database defined as *Lachnospiraceae* based on NCBI taxonomy. Mb DNA megabases. SEM standard error of the mean.

genomes, non-host species have a higher number of 16S genes than species isolated from humans (Figure 2B). Higher ribosomal gene copy number is linked to increased maximum growth rate and ability to respond to resource changes (Stevenson and Schmidt, 2004), suggesting these traits are under greater selection in non-host species.

Substrate utilization

Species of *Lachnospiraceae* collectively metabolize plant biomass through the assimilation of polysaccharides, peptides, and aromatics as well as subsequent transformation of the fermentation products by acetogens and cross-feeding species (Figure 3). As primary degraders of plant biomass, many *Lachnospiraceae* ferment a variety of complex polysaccharides including glucans, mannans, xylans, galactans, pectins, and arabinans (Boutard et al., 2014). Biomass-fermenting species sometimes grow faster on polysaccharides than on the constituent monosaccharides (Boutard et al., 2016). While typically able to ferment multiple polysaccharides, species of *Lachnospiraceae* are ecologically differentiated by specializing on certain substrates. As such, addition of different glycans results in compound-specific changes to the relative abundances of *Lachnospiraceae* in mixed communities (Tolonen et al., 2022). For example, *Roseburia intestinalis* efficiently metabolizes β-mannans and xylan (Leth et al., 2018; La Rosa et al., 2019), while *Roseburia faecis* ferments arabinogalactan (Sheridan et al., 2016). *Lachnospira* include pectin specialists, called pectinophiles (Cornick et al., 1994), that are stimulated by pectin addition (Bang et al., 2018). *Lachnospiraceae*

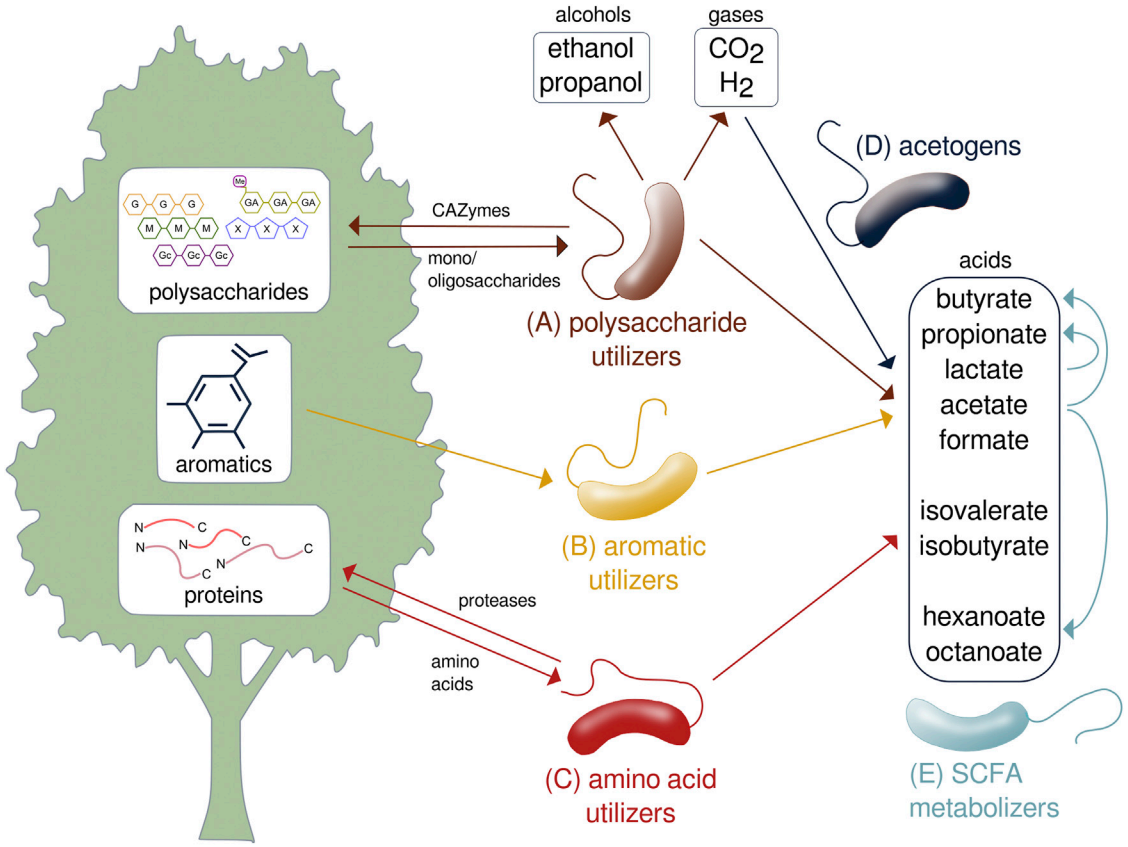


FIGURE 3
Metabolic transformation of lignocellulose performed by species of *Lachnospiraceae*. Specialized species metabolize the **(A)** polysaccharide, **(B)** aromatic, and **(C)** protein components of lignocellulose to produce gases, acids, and alcohols. **(D)** Acetogens transform carbon dioxide and hydrogen to SCFAs. **(E)** SCFAs are further transformed and assimilated by other species. SCFA, short-chain fatty acid; CO_2, carbon dioxide; H_2, hydrogen; CAZyme, carbohydrate-active enzyme.

specialized to metabolize cellulose are found in soil (Warnick et al., 2002; Wolin et al., 2003; Dai et al., 2016) and the rumen (Cai and Dong, 2010).

Genomes of plant-fermenting *Lachnospiraceae* encode numerous carbohydrate-active enzymes (CAZymes), each of which cleave a specific glycosidic linkage to depolymerize complex glycans. *Lachnospiraceae* genomes encode an average of 73 degradative CAZymes (glycoside hydrolases, pectin lyases, carbohydrate esterases) with less than 10% of genomes encoding more than 150 degradative CAZymes (Figure 2C). *Lachnospiraceae* bacterium CE91-St56, closely related to *Eisenbergiella massiliensis*, encodes over 100 more CAZymes than any other strain with 313 CAZymes including 294 glycoside hydrolases (Figure 2C). CE91-St56 was isolated from a supercentenarian (>110 years old) (Sato et al., 2021), highlighting the presence of *Lachnospiraceae* with extensive polysaccharide utilization capabilities in long-lived individuals. *Lachnospiraceae* genomes encode many families of CAZymes, the most abundant of which is GH13 to depolymerize α-glucans including starch and pullulan (Figure 2D). Other abundant CAZyme families enable degradation of substrates such as β-glucans (GH3,5,51,94), β-galactans (GH42), xylan (GH43), pectin (GH78), and chitin (GH18) (Figure 2D).

Some *Lachnospiraceae* assimilate C1 and C2 carbon sources that are produced during the initial stages of plant biomass fermentation. *Blautia hydrogenotrophica* is a human gut acetogen that grows autotrophically using hydrogen to fix carbon dioxide into acetyl-CoA by the Wood–Ljungdahl pathway (Bernalier et al., 1996) and other *Lachnospiraceae* acetogens consume H2 and CO_2 in the cow rumen (Greening and Leedle, 1989). Species that assimilate other C1 and C2 molecules include *Sporobacterium*, isolated from an olive mill, that grows on methanol (Mechichi et al., 1999). Various *Lachnospiraceae* such as *Anaerobutyricum soehngenii* (previously *Eubacterium hallii*) metabolize acetate to butyrate (Udayappan et al., 2016; Zhang et al., 2019), which represents an important cross-feeding interaction in the gut (Flint et al., 2012).

Other *Lachnospiraceae* are specialized to metabolize peptides, alkanes, and aromatics. *Falcatimonas natans*, isolated from a methanogenic reactor, ferments peptides but not carbohydrates (Watanabe et al., 2016). *Abyssivirga alkaniphila*, a hydrothermal vent species, can grow on straight and branched alkanes (C5 to C25) using thiosulfate as an external electron acceptor (Schouw et al., 2016). *Syntrophococcus sucromutans* can oxidize sugars to acetate using formate or methoxymonobenzenoids as electron acceptors (Krumholz and Bryant, 1986). *Parasporobacterium* and *Sporobacterium* are soil bacteria that grow on methoxylated

TABLE 1 Fermentation products of Lachnospiraceae.

Product	Input	Enzymatic steps (EC numbers)	Representative Species	References
Formic acid	Pyruvate	2.3.1.54	*Roseburia intestinalis*	Hillman et al. (2020)
Acetic acid	Acetyl-CoA	2.3.1.8, 2.7.2.1	*Fusicatenibacter saccharivorans*	Takada et al. (2013)
Propionic acid	Lactate (acrylate pathway)	2.8.3.1, 4.2.1.54, 1.3.1.84	*Coprococcus cactus*	Sheridan et al. (2022)
	L-fucose (propanediol pathway)	5.3.1.25, 2.7.1.51, 4.1.2.17, 1.1.1.77, 4.2.1.28, 1.2.1.87, 2.8.3.1	*Roseburia inulinivorans*	Scott et al. (2006)
	L-rhamnose (propanediol pathway)	5.3.1.14, 2.7.1.15, 4.1.2.19, 1.1.1.77, 4.2.1.28, 1.2.1.87, 2.8.3.1	*Lachnoclostridium phytofermentans*	Petit et al. (2013)
	3-oxopropionate (myo-inositol pathway)	2.8.3.5, 1.1.1.35, 4.2.1.116, 1.3.1.95, 2.8.3.1	*Anaerostipes rhamnosivorans*	Bui et al. (2021)
L-lactic acid	Pyruvate	1.1.1.27	*Lachnoclostridium phytofermentans*	Tolonen et al. (2011)
Butyric acid	Acetyl-CoA, (BCoAT)	2.3.1.9, 1.1.1.157, 4.2.1.55, 1.3.1.86, 2.8.3.8	*Roseburia intestinalis*	Vital et al. (2014)
	Acetyl-CoA (Butyrate kinase)	2.3.1.9, 1.1.1.157, 4.2.1.55, 1.3.1.86, 2.3.1.17, 2.7.2.7	*Eubacterium ventriosum*	Vital et al. (2014)
	Lysine	5.4.3.2, 5.4.3.3, 1.4.1.11, 2.3.1.247, 4.3.1.14, 1.3.1.86, 2.8.3.8	*Lachnospiraceae sp F0167*	Vital et al. (2014)
	Glutarate	1.1.99.2, 2.8.3.12, 4.2.1.167, 7.2.4.5, 1.3.1.86, 2.8.3.8	*Clostridiales* sp SS3/4	Vital et al. (2014)
	4-aminobutyrate	1.1.1.61, 2.8.3-, 4.2.1.120, 1.3.1.86, 2.8.3.8	*Anaerostipes caccae*	Vital et al. (2014)
Isobutyric acid	Valine	1.4.1.23/2.6.1.66, 4.1.1.72, 1.2.1.3	*Falcatimonas natans*	Watanabe et al. (2016)
Succinate	Pyruvate	2.7.1.40, 4.1.1.49, 1.1.1.37, 4.2.1.2, 1.3.5.1	*Blautia wexlerae*	Hosomi et al. (2022)
Isovaleric acid	Leucine	1.4.1.9/2.6.1.6, 4.1.1.72, 1.2.13	*Falcatimonas natans*	Watanabe et al. (2016)
5-aminovaleric acid	Proline	1.21.4.1	*Dorea longicatena*	Lopez et al. (2020)
Indoleacetic acid	Tryptophan	2.6.1.57, 4.1.1.74, 1.2.3.7	*Anaerobutyricum soehngenii*	Russell et al. (2013)
Hexanoic, octanoic acid	Acetyl-CoA	2.3.1.16, 1.1.1.35, 4.2.1.17, 1.3.8.7, 3.1.2.20	*Candidatus Weimeria bifida*	Scarborough et al. (2020)
Ethanol	Acetyl-CoA	1.2.1.10, 1.1.1.1	*Lachnoclostridium phytofermentans*	Tolonen et al. (2015b)
1-Propanol	L-fucose	5.3.1.25, 2.7.1.51, 4.1.2.17, 1.1.1.77, 4.2.1.28, 1.1.1.1	*Roseburia inulinivorans*	Scott et al. (2006)
	L-rhamnose	5.3.1.14, 2.7.1.15, 4.1.2.19, 1.1.1.77, 4.2.1.28, 1.1.1.1	*Lachnoclostridium phytofermentans*	Petit et al. (2013)
Hydrogen	Ferredoxin, H+	1.12.7.2	*Roseburia intestinalis*	Dostal et al. (2015)

Shown are the metabolic product, metabolite entering the pathway, enzymatic steps comprising the pathway, and a representative species containing the pathway with a supporting reference. EC number, Enzyme Commission number.

aromatic compounds as sole carbon and energy sources by catabolizing them to short-chain fatty acids (SCFAs) (Lomans et al., 2001). These aromatics are degraded by transforming the side chains and cleaving the aromatic ring using the phloroglucinol pathway (Whitman, 2009), as has recently been shown for *Clostridium scatologenes* (Zhou et al., 2023).

Metabolic products

Lachnospiraceae metabolism yields alcohols, gases, and acids with importance in industry and human health (Table 1). Species producing industrially-relevant alcohols include *Lachnoclostridium phytofermentans*, which ferments cellulose to ethanol at 68% of the maximum theoretical yield (Tolonen et al., 2011) and, similar to *Roseburia inulinivorans*, metabolizes fucose and rhamnose to 1-propanol through a 1,2-propanediol intermediate in polyhedral microcompartments (Scott et al., 2006; Petit et al., 2013). *Lachnospiraceae* produce hydrogen by co-expressing monomeric and multimeric, bifurcating [FeFe] hydrogenases along with energy conserving [NiFe] hydrogenases (Calusinska et al., 2010) to enable yields reaching 2–3 moles of hydrogen per mole glucose (Harvey et al., 2008).

Fermentation of dietary fiber by gut *Lachnospiraceae* yields three main SCFA (acetate, propionate, and butyrate) (Table 1), all of which have been shown to benefit health. Acetate, which is synthesized in two steps from acetyl-CoA, reduces adipose accumulation and improves glucose tolerance (Yamashita et al., 2007). *Lachnospiraceae*

use the acrylate, propanediol, and myo-inositol pathways to synthesize propionate (Reichardt et al., 2014; Bui et al., 2021), which is absorbed from the gut into the bloodstream to regulate cholesterol (Berggren et al., 1996) and reduce visceral and liver fat accumulation (Chambers et al., 2015). Propionate also enhances satiety (Arora et al., 2011), making it a potential way to reduce obesity. Along with *Ruminococcaceae*, *Lachnospiraceae* are the dominant butyrate producers in the human gut (Vital et al., 2014). While butyrate is most often synthesized from acetyl-CoA using butyryl-CoA transferase or butyrate kinase, some *Lachnospiraceae* can produce butyrate from lysine, glutarate, and 4-aminobutyrate (Vital et al., 2014), highlighting their abilities produce butyrate from different nutritional sources. Butyrate is the preferred energy source of colonocytes (Litvak et al., 2018), suppresses pathogens (Walker et al., 2021), and stimulates differentiation of anti-inflammatory T-regulatory cells (Furusawa et al., 2013).

Other organic acids produced by *Lachnospiraceae* include the branched-chain fatty acids isobutyrate and isovalerate that are synthesized from valine and leucine, respectively (Watanabe et al., 2016). *Candidatus* Weimeria bifida metabolizes pentoses to medium-chain fatty acids (hexanoate and octanoate) using the reverse β-oxidation cycle for SCFA chain elongation (Scarborough et al., 2020). Species of *Blautia* metabolize pyruvate to succinate (Hosomi et al., 2022) and, depending on the species, *Lachnospiraceae* can either produce L-lactate (Tolonen et al., 2011) or metabolize DL-lactate to acetate or butyrate using stereospecific lactate dehydrogenases (Sheridan et al., 2022). Gut *Lachnospiraceae* ferment proline to 5-aminovalerate in competition with *Clostridioides difficile*, representing a potential means to prevent colonization by this pathogen (Lopez et al., 2020). *Lachnospiraceae* ferment aromatic amino acids (tryptophan, tyrosine, phenylalanine) to phenolic and indolic acids (Russell et al., 2013), including compounds such as indoleacetic acid that promote intestinal homeostasis by signaling through the aryl hydrocarbon receptor (Roager and Licht, 2018).

Lachnospiraceae also produce other metabolites with health and industrial applications. *Lachnospiraceae* can modify and cleave the heterocyclic C-ring of flavonoids (Braune and Blaut, 2016) yielding molecules such as equol, which is linked to prevention of colorectal cancer (Sugiyama et al., 2013) and aging-related disorders (Mayo et al., 2019). Gut *Lachnospiraceae* generate reactive sulfur species that protect the host from oxidative stress-induced liver injury (Uchiyama et al., 2022). *Lachnospiraceae* produce farnesol (Abdugheni et al., 2022), an isoprene-derived molecule with anti-inflammatory and neuroprotective activities (Sell et al., 2022). Farnesol also has industrial applications as a fragrance ingredient (Lapczynski et al., 2008) and diesel fuel precursor (Rude and Schirmer, 2009). *Butyrivibrio fibrisolvens* synthesizes an exopolysaccharide with potential industrial applications that is rheologically similar to xanthan gum, but is composed of rare sugars including L-altrose and L-iduronic acid (Wachenheim and Patterson, 1992; Ferreira et al., 1997).

Lachnospiraceae are a potential source of other antimicrobial and immunomodulatory compounds. *Blautia obeum* synthesizes a lantibiotic, nisin O, that inhibits pathogens, including *C. difficile* and *Clostridium perfringens* (Hatziioanou et al., 2017). *Lachnospiraceae* produce pyrazines, which are being developed as antimicrobial and anti-fungal drugs (Hassan et al., 2020). Gut *Lachnospiraceae* convert primary bile acids by 7α-dehydroxylation to secondary bile acids that inhibit enteric pathogens and regulate mucosal immunity (Jin et al., 2022). Nonribosomal peptide synthetases (NRPS) of human gut *Lachnospiraceae* produce immunomodulatory secondary metabolites such as di-peptide aldehydes that act as cell-permeable cathepsin inhibitors, which could act as immunosuppressors by blocking antigen processing (Guo et al., 2017).

Cultivation and engineering

Lachnospiraceae are mesophiles that grow at 30°C–45°C and can be cultivated under standard anaerobic conditions using jars with anaerobic sachets or a glove box. A human gut *Lachnospiraceae* biobank (hLchsp) was established as part of the China General Microorganism Culture Collection by isolating strains from healthy adult fecal samples, yielding a collection of 77 species across 33 genera with *Lachnospira*, *Blautia*, and *Roseburia* being the most abundant genera (Abdugheni et al., 2022). Among the seven growth media used for isolations, Yeast Casitone Fatty Acid (YCFA) medium supported the greatest number of *Lachnospiraceae* species, which comprised 19.6% of all bacterial isolates from the fecal samples. Another project to cultivate human gut *Lachnospiraceae* using rich, non-selective media yielded 273 isolates, including all the *Lachnospiraceae* genera that were detected by metagenomic sequences of the donor feces (Sorbara et al., 2020). In contrast to gut species, more oligotrophic, GS2 medium is used to cultivate environmental heterotrophic *Lachnospiraceae* such as *Herbinix hemicellulosilytica* and *L. phytofermentans* (Warnick et al., 2002; Koeck et al., 2015). The acetogen *B. hydrogenotrophica* can be grown on DMSZ 114 or general-acetogen (GA) medium, either as an autotroph using H_2:CO_2 or as a heterotroph by adding a carbon source (Groher and Weuster-Botz, 2016).

Methods for the genetic modification of *Lachnospiraceae* are being developed to study their molecular biology and to engineer optimized strains. Early studies found that conjugative transposons bearing antibiotic resistance genes transfer DNA between *Lachnospiraceae* (Barbosa et al., 1999). Experimental methods have been developed to transfer plasmid DNA into species of *Lachnoclostridium*, *Roseburia*, *Eubacterium*, *Enterocloster*, *Lacrimispora*, and *Blautia* by conjugation with *Escherichia coli* (Tolonen et al., 2009; Cuív et al., 2015; Sheridan et al., 2019; Jin et al., 2022) and by electroporation into species of *Lachnoclostridium* and *Butyrivibrio* (Beard et al., 1995; Rostain et al., 2022) (Figure 4A).

Native plasmids from *Lachnospiraceae* have been used as expression vectors (Hefford et al., 1997) and the pMTL plasmid system (Heap et al., 2009) can be applied to identify plasmid origins and resistance markers that function in strains of interest (Jin et al., 2022; Rostain et al., 2022). Plasmid origins that have been shown to replicate in *Lachnospiraceae* include pCB102 from *Clostridium butyricum*, pBP1 from *Clostridium botulinum*, pAMβ1 from *Enterococcus faecalis*, pWV101 from *Lactococcus lactis*, pIM13 from *Bacillus subtilis*, and pCD6 from *C. difficile*, and these plasmids can be

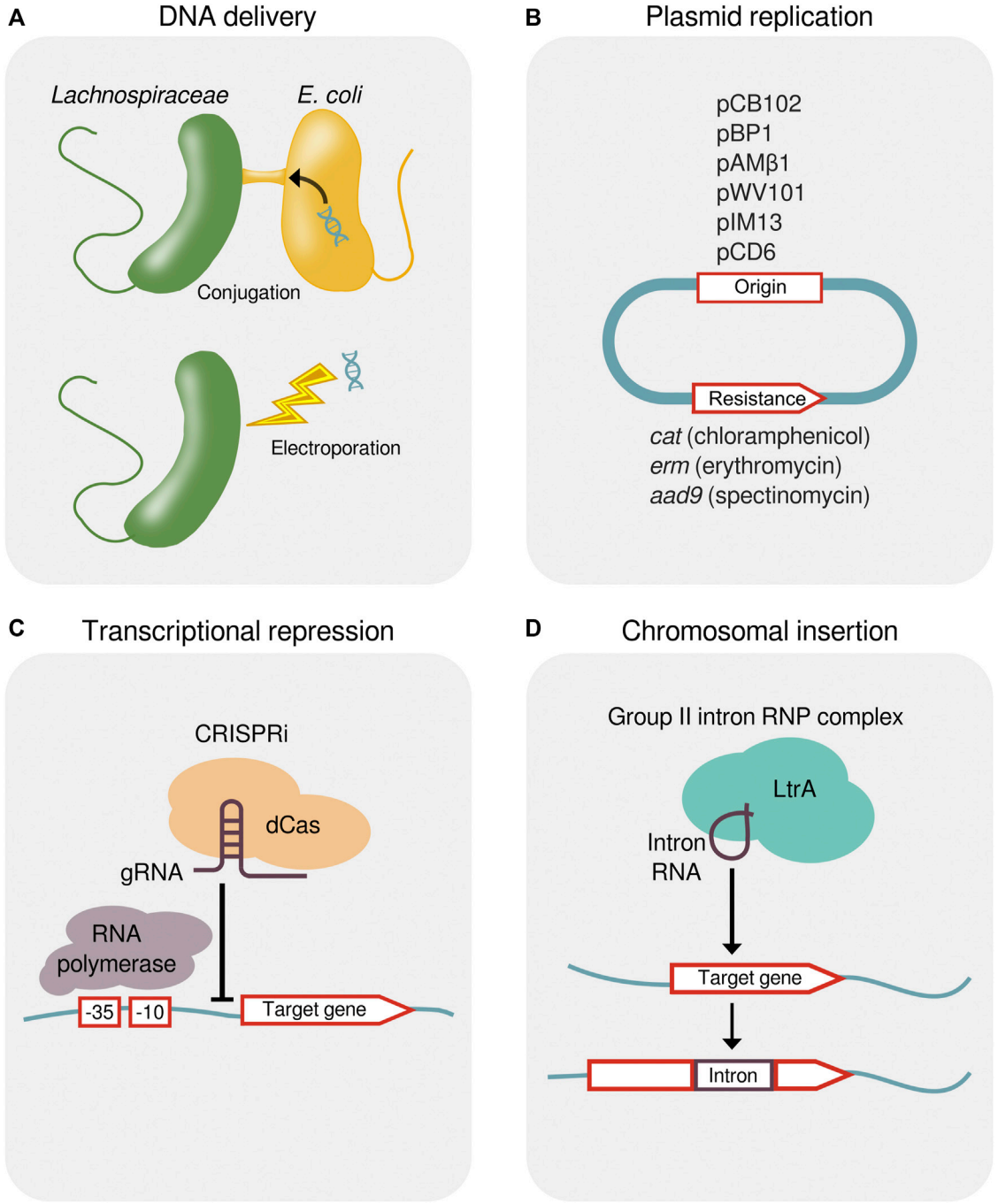


FIGURE 4
Methods for the genetic manipulation of *Lachnospiraceae*. **(A)** Delivery of foreign DNA into *Lachnospiraceae* by either conjugal transfer from *E. coli* or electroporation. **(B)** Plasmid origins that replicate in *Lachnospiraceae* and antibiotic resistance genes for plasmid selection. **(C)** Transcriptional repression by CRISPRi showing a dCas protein targeted by a gRNA to bind upstream of a gene, thereby blocking progression of RNA polymerase. **(D)** Targeted chromosomal insertion using a designed group II intron (targetron). Genomic insertion of the group II intron RNA containing a 13–16 bp target recognition sequence is facilitated by the endonuclease and reverse transcriptase activity of the LtrA protein. CRISPRi, CRISPR interference; dCas, dead CRISPR-associated protein; gRNA, guide RNA; RNP, ribonucleoprotein.

maintained using antibiotic resistance genes *catP* (chloramphenicol), *ermB* (erythromycin), or *aad9* (spectinomycin) (Sheridan et al., 2019; Jin et al., 2022; Rostain et al., 2022) (Figure 4B).

Shuttle vectors containing Gram-negative and Gram-positive origins of replication have permitted the heterologous expression of a β-(1,3–1,4)-glucanase in *Eubacterium rectale* and *R. inulinivorans* (Sheridan et al., 2019) and of a synthetic ethanol formation pathway in *L. phytofermentans* (Tolonen et al., 2015b). Plasmid-based expression of a NanoLuc reporter has identified a library of promoters of varying strength, and reporter gene expression can

be regulated by anhydrotetracycline using a promoter flanked tet repressor sites (Rostain et al., 2022). Clustered Regularly Interspaced Short Palindromic Repeats interference (CRISPRi) using dCas12a has been applied to repress the transcription of chromosomal genes for fermentation, thereby reducing production of butyrate in *Blautia luti* and *Enterocloster boltae* (Jin et al., 2022) and acetate in *L. phytofermentans* (Rostain et al., 2022) (Figure 4C).

Similar to other Clostridia, low rates of DNA transfer and homologous recombination in *Lachnospiraceae* have led to the use of other recombination systems to make targeted chromosomal changes. Designed group II intron called targetrons enabled gene inactivation by targeted chromosome insertion in various *Lachnospiraceae* with efficiencies ranging from 12.5%–100% (Tolonen et al., 2009; Tolonen et al., 2015a; Cerisy et al., 2019a; Jin et al., 2022) (Figure 4D). Multi-gene fragments can be excised and inserted by modifying targetrons to deliver *lox* sites into the genome that act as anchor points for Cre-mediated recombination, which has been applied to delete a 39 kb prophage in *L. phytofermentans* (Cerisy et al., 2019b).

Lachnospiraceae proteins that have been applied as molecular tools in other organisms are strong candidates to advance genome engineering in *Lachnospiraceae*. For example, LbCas12a, which was isolated from *Lachnospiraceae* bacterium ND 2006 (Tak et al., 2017), has been applied for genome editing in eukaryotes including fungi (Chen T. et al., 2023), plants (Kim et al., 2021), flies (Port et al., 2020), and human cells (Zhang et al., 2023). Another Cas12a variant, Lb2Cas12a, derived from *Lachnospiraceae* bacterium MA 2020, has been developed to have enhanced editing activity and broadened protospacer adjacent motif (PAM) recognition in human cells (Tran et al., 2021). A protein with anti-CRISPR function has been identified from *Lachnospiraceae* phage (Forsberg et al., 2019) and a cytosine deaminase from *R. intestinalis* has been fused to CRISPR and Transcription Activator-Like Effector (TALE) proteins to make targeted C-to-T transitions in the genomes of cultured cells and mouse embryos (Guo et al., 2023).

Bioeconomy applications

The capabilities of *Lachnospiraceae* to metabolize the polysaccharides, aromatics, and proteins that compose lignocellulose make them candidates to transform low-cost, sustainable, lignocellulosic feedstocks (i.e., forestry, agricultural, and municipal wastes) into value-added biochemicals (Table 1). For example, *L. phytofermentans*, a species with wide polysaccharide utilization capabilities, ferments corn stover to ethanol with efficiencies similar to commercial enzymes and xylose-fermenting yeast (Jin et al., 2011). Synthesis of longer chain acids from lignocellulose residues by *Ca.* Weimeria bifida (Scarborough et al., 2020) holds potential to develop 'drop in' fuels that are compatible with the current petroleum infrastructure. *F. natans* metabolizes the protein fraction of organic matter to produce isobutyrate and n-butyrate, which are used for artificial fibers, plastics and herbicides as well as isovalerate used for flavoring and perfumes (Agnihotri et al., 2022).

Lachnospiraceae have been identified as biocatalysts for conversion of biomass to hydrogen (Bu et al., 2021), and strains have been isolated that produce 2–3 moles hydrogen per mole of glucose equivalent (Harvey et al., 2008). In addition to its use as a fuel, hydrogen produced from lignocellulosic fermentation can be used as a reductant to fix CO and CO_2 by acetogens through gas fermentation (Figure 3), a process that has potential to convert industrial carbon emissions to useful biochemicals (Liew et al., 2016). *Lachnospiraceae* acetogens compete with methanogens in the cow rumen, representing a means to reduce bovine methane production (Yang et al., 2015).

Alternatively, hydrogen from lignocellulosic fermentation can be used to fix carbon dioxide by methanogens, and *Lachnospiraceae* have been shown to actuate the transformation of cellulose to methane by methanogenic consortia (Dai et al., 2016). Combination of *Lachnospiraceae* with other microorganisms to form synthetic consortia has been generally used to enhanced product formation rates (Zuroff et al., 2013; El Hage et al., 2019; Park et al., 2020), and engineered consortia including biomass-fermenting *Lachnospiraceae* and methanogens or acetogens have potential for the direct conversion of lignocellulose to methane and other value-added biochemicals.

Therapeutic applications: preclinical models

Preclinical *in vivo* models are providing evidence that supplementation with live *Lachnospiraceae* improves gut health and prevents pathogen colonization. For example, in a rat model of irritable bowel syndrome, addition of *Roseburia hominis* increased cecal butyrate content, reduced visceral hypersensitivity, and prevented the decreased expression of occludin (Zhang et al., 2019). Administration of *Blautia producta* directly inhibited growth of vancomycin-resistant *Enterococcus* (Caballero et al., 2017) and a murine *Lachnospiraceae* inhibits *C. difficile* colonization in mice (Reeves et al., 2012).

Lachnospiraceae can alleviate inflammatory and allergic diseases by modulating the immune system through production of antigens presented by innate immune cells and immunomodulatory metabolites. *R. intestinalis* promoted differentiation of regulatory T cells, activation of type 3 innate lymphoid cells, and suppression of inflammation through TLR5 (Shen et al., 2022). *E. rectale* supplementation regulated dendritic cell activation by reducing the frequency of $CD83^+$ cells and improved symptoms in a mouse model of Behçet's disease, a systemic inflammatory condition (Islam et al., 2021). Inoculation of germ-free mice with a 17 strain consortium including 8 *Lachnospiraceae* increased anti-inflammatory, regulatory T-cells ($CD4^+$, FoxP3+) in the colonic lamina propria of multiple mouse lines and alleviated colitis in experimental models (Atarashi et al., 2013). In addition, administration of this 17 strain consortium to mice in an ovalbumin (OVA)-induced allergic diarrhea mouse model reduced diarrhea and OVA-specific serum IgE levels (Atarashi et al., 2013). Additional evidence that *Lachnospiraceae* can mitigate food allergies was demonstrated by colonization of germ-free mice with *Anaerostipes caccae*, which protected against an anaphylactic response upon challenge with β-Lactoglobulin (BLG), a cow's milk allergen, and reduced BLG-specific IgE levels (Feehley et al., 2019).

Lachnospiraceae have provided benefits in preclinical models of metabolic syndrome and diabetes. Administration of *Blautia wexlerae* to male C57BL/6 mice on a high fat diet reduced body weight and multiple diabetes indicators, which was linked to *B. wexlerae* metabolites such as S-adenosylmethionine, acetylcholine, and L-ornithine conferring anti-adipogenesis and anti-inflammatory

TABLE 2 Clinical studies of Lachnospiraceae as live biotherapeutic products.

Principal Investigator; Sponsor	Trial	Intervention	Patient population	Outcome	Trial number, reference
Dr. Elaine Petrof, Queen's University	Open label trial	Synthetic community	rCDI	Both subjects clinically cured at 6 months	NCT01372943 (Petrof et al., 2013)
Dr. Lisa von Moltke, Seres Therapeutics	Phase 3, double-blind, placebo-controlled	SER-109	rCDI	Reduced rCDI at week 8 ($p < 0.001$)	NCT03183128 (Feuerstadt et al., 2022)
Dr. Michele Trucksis, Seres Therapeutics	Phase 2 double-blind, placebo-controlled	SER-109	rCDI	Reduced rCDI (44.1% vs. 53.3% with placebo), but not significant. Engraftment associated with non-recurrence ($p < 0.05$) and increased secondary bile acid concentrations ($p < 0.0001$)	NCT02437487 (McGovern et al., 2022)
Seres Therapeutics	Phase 1b Safety Study	SER-287	Mild to moderate UC	Increased clinical remission at week 8 ($p = 0.024$)	NCT02618187 (Henn et al., 2021)
Dr Eamonn Quigley, Houston Methodist; 4DPharma	Phase 2 double-blind, placebo-controlled	MRx1234 (Blautix)	IBS	Improved bowel habits ($p = 0.007$) and trend to increased overall response ($p = 0.06$)	NCT03721107 (Quigley et al., 2023)
Dr. Darrell Pardi, Mayo Clinic; Vedanta Biosciences	Phase 2 double-blind, placebo-controlled	VE303	rCDI	Reduced rCDI at week 8 ($p = 0.006$)	NCT03788434 (Louie et al., 2023)
Vedanta Biosciences	Phase 1 safety study	VE303	Healthy adults	Well tolerated, engraftment	NCT04236778 (Dsouza et al., 2022)
Dr. Patricia Bloom, University of Michigan; Vedanta Biosciences	Double-blind, placebo-controlled	VE303	Hepatic encephalopathy	Trial in progress	NCT04899115
Vedanta Biosciences	Phase 2, double-blind, placebo-controlled	VE202	Mild to moderate UC	Trial in progress	NCT05370885
Vedanta Biosciences	Phase 1 safety study	VE202	Healthy adults	Well tolerated, strain engraftment	Silber et al. (2022)
Dr. Erik Stroes, University of Amsterdam; Caelus Pharmaceuticals	Phase 2, double-blind placebo controlled	*Anaerobutyricum soehngenii* L2-7	Metabolic Syndrome	Elevated plasma GLP-1 ($p = 0.02$), increased fecal butyrate ($p = 0.06$), reduced glucose variability ($p = 0.05$)	NTR-NL6630, (Koopen et al., 2022)
Dr. James Ryan Atlantia Food Clinical Trials; Caelus Pharmaceuticals	Phase 1/2 dose finding study	*Anaerobutyricum soehngenii* L2-7	Metabolic Syndrome	Abundance of *A. soehngenii* correlated with peripheral insulin sensitivity ($p = 0.05$)	NCT04529473 (Gilijamse et al., 2020)
Oluf B Pedersen, University of Copenhagen	Placebo-controlled crossover study	*Ruminococcus torques* strain ATCC 27756	Overweight adults	Trial in progress	NCT05448274

The synthetic community in the Queen's University study consisted of 33 strains, including 10 *Lachnospiraceae*. SER-109 is purified fecal spores, 36% of genera are *Lachnospiraceae*. SER-287 is purified fecal spores, 44% of genera are *Lachnospiraceae*. MRx1234 (Blautix) is lyophilised *Blautia hydrogenotrophica*. VE303 is a consortium of 8 strains including 5 *Lachnospiraceae*. VE202 is a consortium of 16 *Clostridia* XIVa, IV, XVIII. rCDI, recurrent *Clostridioides difficile* infection; UC, ulcerative colitis; IBS, irritable bowel syndrome.

properties to adipocytes (Hosomi et al., 2022). *B. producta* suppressed lipid accumulation in HepG2 cells and gavage of *B. producta* alleviated hyperlipidemia in mice through production of 12-methylmyristic acid (Wu et al., 2021). Treatment of obese, diabetic *db/db* mice with *A. soehngenii* reduced plasma glucose, epididymal fat, and liver triglycerides and improved peripheral insulin sensitivity (Udayappan et al., 2016).

Mouse studies have also shown beneficial roles of *Lachnospiraceae* in cancer treatment. C57BL/6 mice bearing B16-F10 melanoma or CT-26 colorectal tumors showed reduced tumor growth when administered *Blautia massiliensis* (Goodman et al., 2019). Anti-PD1 mediated tumor control and survivorship of C57BL/6 mice bearing B16-F10 melanoma tumors was improved by administration of *E. rectale*, which was proposed to result from it consuming L-serine, leading to NK cell activation and tumor infiltration (Liu et al., 2023). *E. rectale* treatment also reduced the incidence of lymphoma and reduced TNF levels in sensitized Eμ-Myc mice (Lu et al., 2022). Survivorship and clinical scores of C57BL/6 mice following full body irradiation were increased by prior inoculation with a mix of 23 *Lachnospiraceae*, which resulted from increased hematopoiesis and reduced intestinal epithelial injury (Guo et al., 2020).

Therapeutic applications: clinical trials

Lachnospiraceae are being tested as LBPs in clinical studies for a number of diseases (Table 2), the most advanced of which is the treatment of recurrent *C. difficile* infections (rCDI). In an exploratory study of two rCDI patients at Queen's University, treatment with a

synthetic community of 33 strains, including 10 *Lachnospiraceae*, resulted in full remission in both patients (Petrof et al., 2013). Subsequently, microbial consortia containing *Lachnospiraceae* have been developed to combat rCDI. SER-109, consisting of spores purified from fecal samples, is composed of 77 genera of which 36% are *Lachnospiraceae* (Feuerstadt et al., 2022). In a phase 3 study, SER-109 reduced rCDI at week 8 (Feuerstadt et al., 2022). In 2023, the US Food and Drug Administration approved SER-109 under the commercial name Vowst™ to treat recurrent *C. difficile* infection (Commission of the US FDA, 2023). VE303, a consortium of 8 strains including 5 *Lachnospiraceae*, engrafted into the microbiomes of healthy volunteers to boost production of SCFA and secondary bile acids without any serious adverse events (Dsouza et al., 2022). Phase 2 results in patients supported that rCDI at 8 weeks was reduced by VE303 (Louie et al., 2023). VE303 is being tested in a phase 3 study to treat rCDI starting in 2023, and as an experimental treatment for hepatic encephalopathy (Vedanta Biosciences Inc, 2023).

LBPs that include *Lachnospiraceae* are also being developed for metabolic syndrome and inflammatory bowel disease (Table 2). For example, *A. soehngenii* L2-7 is being tested as a probiotic to improve insulin sensitivity in metabolic syndrome patients. A phase 1/2 study of 24 metabolic syndrome patients correlated *A. soehngenii* engraftment with improved peripheral insulin sensitivity (Gilijamse et al., 2020). A phase 2 study of 12 metabolic syndrome patients showed duodenal infusion of *A. soehngenii* reduced glucose variability and elevated GLP-1, secondary bile acids in plasma, duodenal REGB1 expression, and fecal SCFAs (Koopen et al., 2022). As *B. hydrogenotrophica* consumes intestinal gas (H_2, CO_2), it can treat irritable bowel syndrome (IBS) by reducing intestinal bloating. A phase 2 study of MRx1234 (Blautix), consisting of lyophilized *B. hydrogenotrophica*, improved bowel habits in IBS patients (Quigley et al., 2023).

Looking forward: challenges and opportunities

Lachnospiraceae have numerous potential applications due to their native fermentation of low-cost substrates and importance for intestinal health, but key challenges remain to harness them as industrial biocatalysts and LBPs. In particular, work is needed to isolate and characterize additional *Lachnospiraceae* strains, uncover the genetic basis of their physiological traits, and, ultimately, apply these learning to engineer optimized strains. Recently established *Lachnospiraceae* collections demonstrated methods with which many species can be cultivated (Seshadri et al., 2018; Sorbara et al., 2020; Abdugheni et al., 2022). Additional cultivation efforts could build *Lachnospiraceae* collections from environments such as soil and human patient populations. Further, *Lachnospiraceae* culture collections revealed that isolates from the same species can differ in traits such as substrate utilization and production of antimicrobials (Sorbara et al., 2020). Intraspecific comparative genomics of isolates sharing similar genomes, but with specific physiological differences, is an opportunity to define genotype-phenotype relationships in *Lachnospiraceae*.

Development of microorganisms for biotechnology often requires rewiring native metabolism to improve product yields, highlighting the importance of future research on genetic manipulation of *Lachnospiraceae*. As described above, genetic tools have been ported to *Lachnospiraceae* from other well-studied mesophiles including *B. subtilis* and *L. lactis*. Species of *Lachnospiraceae* have been established as genetically tractable hosts with methods for genetic transformation, plasmid replication and selection, transcriptional repression, and chromosomal insertions (Figure 4); a recent study demonstrates that these methods can be generally applied to many species (Jin et al., 2022). Existing molecular tools derived from *Lachnospiraceae* are promising candidates to advance genome engineering in these bacteria, including Cas proteins with broadened PAM recognition (Tran et al., 2021), anti-CRISPR proteins (Forsberg et al., 2019), and cytosine deaminases (Guo et al., 2023). Moreover, adaptive laboratory evolution can be used to generate *Lachnospiraceae* strains with complex, multigenic traits such as inhibitor tolerance that are intractable by rational genome engineering (Cerisy et al., 2017).

Engineering of *Lachnospiraceae* for industrial production of biochemicals from low-cost feedstocks will need to focus on substrate assimilation and product formation. The rate of lignocellulose solubilization remains a primary obstacle to its utilization as a feedstock (Preethi et al., 2021), which could be addressed by engineering strains with modified expression of CAZymes and associated ABC transporters to accelerate solubilization and uptake of target lignocellulosic substrates. As *Lachnospiraceae* typically produce a mixture of fermentation products, redistribution of metabolic flux by repression or inactivation of genes for undesired products is a valuable approach to streamline and increase fermentation yields.

Genetic manipulation of *Lachnospiraceae* will also be important to define the mechanisms by which they promote intestinal health. Significant advances have been made to elucidate innate and adaptive immune responses modulated by *Lachnospiraceae* (Islam et al., 2021, 1) (Atarashi et al., 2013; Feehley et al., 2019; Liu et al., 2023). It is generally believed that *Lachnospiraceae* modulate host immunity producing SCFAs. Recently, it was shown that differential recognition of *Lachnospiraceae* flagellins by TLR5 contributes to immune tolerance (Clasen et al., 2023). Genetic studies with *Lachnospiraceae* mutants will be useful to further define the mechanisms that underlie host interactions and build strains with customized immunomodulatory properties.

Industrial-scale cultivation of *Lachnospiraceae* for bioproduction and LBPs will necessitate a greater understanding of the genetics and ecology of the phage that infect them. Phage infection is recognized as a persistent threat in industrial microbiology where scale-up of bacterial populations in large bioreactors favors phage outbreaks. For example, phage infection is the main cause of fermentation failures in the dairy industry, leading to intense study of the phage of lactic acid bacteria (Fernández et al., 2017). Prophages are common in the genomes of gut *Lachnospiraceae* (Dikareva et al., 2023). Although hypervirulent phage can drive temporal variation of *R. intestinalis* abundances in the intestine, no phage infecting *Lachnospiraceae* have been deposited in public databases (Cornuault et al., 2020). Thus, isolation and characterization of *Lachnospiraceae* phage are important subjects for future research.

Over the past decade, our understanding of *Lachnospiraceae* molecular biology and physiology has greatly increased. In 2023, Vowst™ became the first FDA-approved oral medication containing live *Lachnospiraceae* (Commission of the US FDA, 2023) and clinical studies are evaluating additional therapeutic

benefits of *Lachnospiraceae* (Table 2). Development of optimized strains by genome engineering will enable us to realize the potential of *Lachnospiraceae* in biotechnology. In the next few years, we expect to see further development of *Lachnospiraceae* to produce useful chemicals and to manage disease through targeted changes to the composition and metabolites produced by the gut microbiome.

Author contributions

TZ: Writing–original draft, Writing–review and editing. SM: Writing–original draft, Writing–review and editing. AT: Conceptualization, Writing–original draft, Writing–review and editing.

Funding

The authors declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by Genoscope-CEA and the Évry Genopole under the "Action Thématique Incitative Genopole" Grant (funding ATIGE 2021 No. 2698).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abdugheni, R., Wang, W.-Z., Wang, Y.-J., Du, M.-X., Liu, F.-L., Zhou, N., et al. (2022). Metabolite profiling of human-originated Lachnospiraceae at the strain level. *iMeta* 1, e58. doi:10.1002/imt2.58

Agnihotri, S., Yin, D.-M., Mahboubi, A., Sapmaz, T., Varjani, S., Qiao, W., et al. (2022). A glimpse of the world of volatile fatty acids production and application: a review. *Bioengineered* 13, 1249–1275. doi:10.1080/21655979.2021.1996044

Antezack, A., Boxberger, M., La Scola, B., and Monnet-Corti, V. (2021). Isolation and description of catonella massiliensis sp. nov., a novel catonella species, isolated from a stable periodontitis subject. *Pathogens* 10, 367. doi:10.3390/pathogens10030367

Arora, T., Sharma, R., and Frost, G. (2011). Propionate. Anti-obesity and satiety enhancing factor? *Appetite* 56, 511–515. doi:10.1016/j.appet.2011.01.016

Atarashi, K., Tanoue, T., Oshima, K., Suda, W., Nagano, Y., Nishikawa, H., et al. (2013). Treg induction by a rationally selected mixture of Clostridia strains from the human microbiota. *Nature* 500, 232–236. doi:10.1038/nature12331

Bang, S.-J., Kim, G., Lim, M. Y., Song, E.-J., Jung, D.-H., Kum, J.-S., et al. (2018). The influence of *in vitro* pectin fermentation on the human fecal microbiome. *Amb. Express* 8, 98. doi:10.1186/s13568-018-0629-9

Barbosa, T. M., Scott, K. P., and Flint, H. J. (1999). Evidence for recent intergeneric transfer of a new tetracycline resistance gene, tet(W), isolated from Butyrivibrio fibrisolvens, and the occurrence of tet(O) in ruminal bacteria. *Environ. Microbiol.* 1, 53–64. doi:10.1046/j.1462-2920.1999.00004.x

Beard, C. E., Hefford, M. A., Forster, R. J., Sontakke, S., Teather, R. M., and Gregg, K. (1995). A stable and efficient transformation system for Butyrivibrio fibrisolvens OB156. *Curr. Microbiol.* 30, 105–109. doi:10.1007/BF00294191

Berggren, A. M., Nyman, E. M., Lundquist, I., and Björck, I. M. (1996). Influence of orally and rectally administered propionate on cholesterol and glucose metabolism in obese rats. *Br. J. Nutr.* 76, 287–294. doi:10.1079/bjn19960032

Bernalier, A., Willems, A., Leclerc, M., Rochet, V., and Collins, M. D. (1996). Ruminococcus hydrogenotrophicus sp. nov., a new H2/CO2-utilizing acetogenic bacterium isolated from human feces. *Arch. Microbiol.* 166, 176–183. doi:10.1007/s002030050373

Boutard, M., Cerisy, T., Nogue, P.-Y., Alberti, A., Weissenbach, J., Salanoubat, M., et al. (2014). Functional diversity of carbohydrate-active enzymes enabling a bacterium to ferment plant biomass. *PLoS Genet.* 10, e1004773. doi:10.1371/journal.pgen.1004773

Boutard, M., Ettwiller, L., Cerisy, T., Alberti, A., Labadie, K., Salanoubat, M., et al. (2016). Global repositioning of transcription start sites in a plant-fermenting bacterium. *Nat. Commun.* 7, 13783. doi:10.1038/ncomms13783

Braune, A., and Blaut, M. (2016). Bacterial species involved in the conversion of dietary flavonoids in the human gut. *Gut Microbes* 7, 216–234. doi:10.1080/19490976.2016.1158395

Bu, J., Wei, H.-L., Wang, Y.-T., Cheng, J.-R., and Zhu, M.-J. (2021). Biochar boosts dark fermentative H2 production from sugarcane bagasse by selective enrichment/colonization of functional bacteria and enhancing extracellular electron transfer. *Water Res.* 202, 117440. doi:10.1016/j.watres.2021.117440

Bui, T. P. N., Mannerås-Holm, L., Puschmann, R., Wu, H., Troise, A. D., Nijsse, B., et al. (2021). Conversion of dietary inositol into propionate and acetate by commensal Anaerostipes associates with host health. *Nat. Commun.* 12, 4798. doi:10.1038/s41467-021-25081-w

Caballero, S., Kim, S., Carter, R. A., Leiner, I. M., Sušac, B., Miller, L., et al. (2017). Cooperating commensals restore colonization resistance to vancomycin-resistant Enterococcus faecium. *Cell Host Microbe* 21, 592–602.e4. doi:10.1016/j.chom.2017.04.002

Cai, S., and Dong, X. (2010). Cellulosilyticum ruminicola gen. nov., sp. nov., isolated from the rumen of yak, and reclassification of Clostridium lentocellum as Cellulosilyticum lentocellum comb. nov. *Int. J. Syst. Evol. Microbiol.* 60, 845–849. doi:10.1099/ijs.0.014712-0

Calusinska, M., Happe, T., Joris, B., and Wilmotte, A. (2010). The surprising diversity of clostridial hydrogenases: a comparative genomic perspective. *Microbiol. Read.* 156, 1575–1588. doi:10.1099/mic.0.032771-0

Cerisy, T., Iglesias, A., Rostain, W., Boutard, M., Pelle, C., Perret, A., et al. (2019a). ABC transporters required for hexose uptake by Clostridium phytofermentans. *J. Bacteriol.* 201, e00241-19. doi:10.1128/JB.00241-19

Cerisy, T., Rostain, W., Chhun, A., Boutard, M., Salanoubat, M., and Tolonen, A. C. (2019b). A targetron-recombinase system for large-scale genome engineering of clostridia. *mSphere* 4, e00710-19. doi:10.1128/mSphere.00710-19

Cerisy, T., Souterre, T., Torres-Romero, I., Boutard, M., Dubois, I., Patrouix, J., et al. (2017). Evolution of a biomass-fermenting bacterium to resist lignin phenolics. *Appl. Environ. Microbiol.* 83, e00289-17. doi:10.1128/AEM.00289-17

Chambers, E. S., Viardot, A., Psichas, A., Morrison, D. J., Murphy, K. G., Zac-Varghese, S. E. K., et al. (2015). Effects of targeted delivery of propionate to the human colon on appetite regulation, body weight maintenance and adiposity in overweight adults. *Gut* 64, 1744–1754. doi:10.1136/gutjnl-2014-307913

Chen, I.-M. A., Chu, K., Palaniappan, K., Ratner, A., Huang, J., Huntemann, M., et al. (2023a). The IMG/M data management and analysis system v.7: content updates and new features. *Nucleic Acids Res.* 51, D723–D732. doi:10.1093/nar/gkac976

Chen, T., Chen, Z., Zhang, H., Li, Y., Yao, L., Zeng, B., et al. (2023b). Development of a CRISPR/Cpf1 system for multiplex gene editing in Aspergillus oryzae. *Folia Microbiol. (Praha).* doi:10.1007/s12223-023-01081-9

Clasen, S. J., Bell, M. E. W., Borbón, A., Lee, D.-H., Henseler, Z. M., de la Cuesta-Zuluaga, J., et al. (2023). Silent recognition of flagellins from human gut commensal bacteria by Toll-like receptor 5. *Sci. Immunol.* 8, eabq7001. doi:10.1126/sciimmunol.abq7001

Commission of the US Food and Drug Administration (2023). FDA approves first orally administered fecal microbiota product for the prevention of recurrence of Clostridioides difficile infection. https://www.fda.gov/news-events/press-announcements/fda-approves-first-orally-administered-fecal-microbiota-product-prevention-recurrence-clostridioides.

Cornick, N. A., Jensen, N. S., Stahl, D. A., Hartman, P. A., and Allison, M. J. (1994). Lachnospira pectinoschiza sp. nov., an anaerobic pectinophile from the pig intestine. *Int. J. Syst. Bacteriol.* 44, 87–93. doi:10.1099/00207713-44-1-87

Cornuault, J. K., Moncaut, E., Loux, V., Mathieu, A., Sokol, H., Petit, M.-A., et al. (2020). The enemy from within: a prophage of Roseburia intestinalis systematically turns lytic in the mouse gut, driving bacterial adaptation by CRISPR spacer acquisition. *ISME J.* 14, 771–787. doi:10.1038/s41396-019-0566-x

Cuív, P. Ó., Smith, W. J., Pottenger, S., Burman, S., Shanahan, E. R., and Morrison, M. (2015). Isolation of genetically tractable most-wanted bacteria by metaparental mating. *Sci. Rep.* 5, 13282. doi:10.1038/srep13282

Dai, Y., Yan, Z., Jia, L., Zhang, S., Gao, L., Wei, X., et al. (2016). The composition, localization and function of low-temperature-adapted microbial communities involved in methanogenic degradations of cellulose and chitin from Qinghai-Tibetan Plateau wetland soils. *J. Appl. Microbiol.* 121, 163–176. doi:10.1111/jam.13164

Dikareva, E., Matharu, D., Lahtinen, E., Kolho, K.-L., De Vos, W. M., Salonen, A., et al. (2023). An extended catalog of integrated prophages in the infant and adult fecal microbiome shows high prevalence of lysogeny. *Front. Microbiol.* 14, 1254535. doi:10.3389/fmicb.2023.1254535

Dostal, A., Lacroix, C., Bircher, L., Pham, V. T., Follador, R., Zimmermann, M. B., et al. (2015). Iron modulates butyrate production by a child gut microbiota *in vitro*. *mBio* 6, e01453–e01415. doi:10.1128/mBio.01453-15

Dsouza, M., Menon, R., Crossette, E., Bhattarai, S. K., Schneider, J., Kim, Y.-G., et al. (2022). Colonization of the live biotherapeutic product VE303 and modulation of the microbiota and metabolites in healthy volunteers. *Cell Host Microbe* 30, 583–598.e8. doi:10.1016/j.chom.2022.03.016

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinforma.* 5, 113. doi:10.1186/1471-2105-5-113

El Hage, R., Hernandez-Sanabria, E., Calatayud Arroyo, M., Props, R., and Van de Wiele, T. (2019). Propionate-producing consortium restores antibiotic-induced dysbiosis in a dynamic *in vitro* model of the human intestinal microbial ecosystem. *Front. Microbiol.* 10, 1206. doi:10.3389/fmicb.2019.01206

Feehley, T., Plunkett, C. H., Bao, R., Choi Hong, S. M., Culleen, E., Belda-Ferre, P., et al. (2019). Healthy infants harbor intestinal bacteria that protect against food allergy. *Nat. Med.* 25, 448–453. doi:10.1038/s41591-018-0324-z

Fernández, L., Escobedo, S., Gutiérrez, D., Portilla, S., Martínez, B., García, P., et al. (2017). Bacteriophages in the dairy environment: from enemies to allies. *Antibiotics* 6, 27. doi:10.3390/antibiotics6040027

Ferreira, F., Kenne, L., Cotta, M. A., and Stack, R. J. (1997). Structural studies of the extracellular polysaccharide from Butyrivibrio fibrisolvens strain CF3. *Carbohydr. Res.* 301, 193–203. doi:10.1016/s0008-6215(97)00097-9

Feuerstadt, P., Louie, T. J., Lashner, B., Wang, E. E. L., Diao, L., Bryant, J. A., et al. (2022). SER-109, an oral microbiome therapy for recurrent Clostridioides difficile infection. *N. Engl. J. Med.* 386, 220–229. doi:10.1056/NEJMoa2106516

Flint, H. J., Scott, K. P., Duncan, S. H., Louis, P., and Forano, E. (2012). Microbial degradation of complex carbohydrates in the gut. *Gut Microbes* 3, 289–306. doi:10.4161/gmic.19897

Forsberg, K. J., Bhatt, I. V., Schmidtke, D. T., Javanmardi, K., Dillard, K. E., Stoddard, B. L., et al. (2019). Functional metagenomics-guided discovery of potent Cas9 inhibitors in the human microbiome. *eLife* 8, e46540. doi:10.7554/eLife.46540

Furusawa, Y., Obata, Y., Fukuda, S., Endo, T. A., Nakato, G., Takahashi, D., et al. (2013). Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature* 504, 446–450. doi:10.1038/nature12721

Garzetti, D., Brugiroux, S., Bunk, B., Pukall, R., McCoy, K. D., Macpherson, A. J., et al. (2017). High-quality whole-genome sequences of the oligo-mouse-microbiota bacterial community. *Genome Announc.* 5, 007588-17. doi:10.1128/genomeA.00758-17

Gilijamse, P. W., Hartstra, A. V., Levin, E., Wortelboer, K., Serlie, M. J., Ackermans, M. T., et al. (2020). Treatment with Anaerobutyricum soehngenii: a pilot study of safety and dose-response effects on glucose metabolism in human subjects with metabolic syndrome. *NPJ Biofilms Microbiomes* 6, 16. doi:10.1038/s41522-020-0127-0

Goodman, B., Sandy, P., Papkoff, J., Gardner, H., and Ponichtera, H. (2019). *Treating cancer using a blautia strain*. US Patent application 2019/0240267 A1.

Gosalbes, M. J., Durbán, A., Pignatelli, M., Abellan, J. J., Jiménez-Hernández, N., Pérez-Cobas, A. E., et al. (2011). Metatranscriptomic approach to analyze the functional human gut microbiota. *PLOS ONE* 6, e17447. doi:10.1371/journal.pone.0017447

Greening, R. C., and Leedle, J. A. Z. (1989). Enrichment and isolation of Acetitomaculum ruminis, gen. nov., sp. nov.: acetogenic bacteria from the bovine rumen. *Arch. Microbiol.* 151, 399–406. doi:10.1007/BF00416597

Groher, A., and Weuster-Botz, D. (2016). General medium for the autotrophic cultivation of acetogens. *Bioprocess Biosyst. Eng.* 39, 1645–1650. doi:10.1007/s00449-016-1634-5

Guo, C.-J., Chang, F.-Y., Wyche, T. P., Backus, K. M., Acker, T. M., Funabashi, M., et al. (2017). Discovery of reactive microbiota-derived metabolites that inhibit host proteases. *Cell* 168, 517–526.e18. doi:10.1016/j.cell.2016.12.021

Guo, H., Chou, W.-C., Lai, Y., Liang, K., Tam, J. W., Brickey, W. J., et al. (2020). Multi-omics analyses of radiation survivors identify radioprotective microbes and metabolites. *Science* 370, eaay9097. doi:10.1126/science.aay9097

Guo, J., Yu, W., Li, M., Chen, H., Liu, J., Xue, X., et al. (2023). A DddA ortholog-based and transactivator-assisted nuclear and mitochondrial cytosine base editors with expanded target compatibility. *Mol. Cell* 83, 1710–1724.e7. doi:10.1016/j.molcel.2023.04.012

Harvey, S., Chambers, A., and Zhang, P. (2008). Overproduction of hydrogen from an anaerobic bacterium. https://apps.dtic.mil/sti/citations/ADA505872.

Hassan, N. W., Saudi, M. N., Abdel-Ghany, Y. S., Ismail, A., Elzahhar, P. A., Sriram, D., et al. (2020). Novel pyrazine based anti-tubercular agents: design, synthesis, biological evaluation and *in silico* studies. *Bioorg Chem.* 96, 103610. doi:10.1016/j.bioorg.2020.103610

Hatziioanou, D., Gherghisan-Filip, C., Saalbach, G., Horn, N., Wegmann, U., Duncan, S. H., et al. (2017). Discovery of a novel lantibiotic nisin O from Blautia obeum A2-162, isolated from the human gastrointestinal tract. *Microbiol. Read.* 163, 1292–1305. doi:10.1099/mic.0.000515

Heap, J. T., Pennington, O. J., Cartman, S. T., and Minton, N. P. (2009). A modular system for Clostridium shuttle plasmids. *J. Microbiol. Methods* 78, 79–85. doi:10.1016/j.mimet.2009.05.004

Hefford, M. A., Kobayashi, Y., Allard, S. E., Forster, R. J., and Teather, R. M. (1997). Sequence analysis and characterization of pOM1, a small cryptic plasmid from Butyrivibrio fibrisolvens, and its use in construction of a new family of cloning vectors for Butyrivibrios. *Appl. Environ. Microbiol.* 63, 1701–1711. doi:10.1128/aem.63.5.1701-1711.1997

Hengstmann, U., Chin, K. J., Janssen, P. H., and Liesack, W. (1999). Comparative phylogenetic assignment of environmental sequences of genes encoding 16S rRNA and numerically abundant culturable bacteria from an anoxic rice paddy soil. *Appl. Environ. Microbiol.* 65, 5050–5058. doi:10.1128/AEM.65.11.5050-5058.1999

Henn, M. R., O'Brien, E. J., Diao, L., Feagan, B. G., Sandborn, W. J., Huttenhower, C., et al. (2021). A phase 1b safety study of SER-287, a spore-based microbiome therapeutic, for active mild to moderate ulcerative colitis. *Gastroenterology* 160, 115–127.e30. doi:10.1053/j.gastro.2020.07.048

Hillman, E. T., Kozik, A. J., Hooker, C. A., Burnett, J. L., Heo, Y., Kiesel, V. A., et al. (2020). Comparative genomics of the genus Roseburia reveals divergent biosynthetic pathways that may influence colonic competition among species. *Microb. Genom* 6, mgen000399. doi:10.1099/mgen.0.000399

Hosomi, K., Saito, M., Park, J., Murakami, H., Shibata, N., Ando, M., et al. (2022). Oral administration of Blautia wexlerae ameliorates obesity and type 2 diabetes via metabolic remodeling of the gut microbiota. *Nat. Commun.* 13, 4477. doi:10.1038/s41467-022-32015-7

Huang, X., Liu, L., Zhao, J., Zhang, J., and Cai, Z. (2019). The families Ruminococcaceae, Lachnospiraceae, and Clostridiaceae are the dominant bacterial groups during reductive soil disinfestation with incorporated plant residues. *Appl. Soil Ecol.* 135, 65–72. doi:10.1016/j.apsoil.2018.11.011

Islam, S. M. S., Ryu, H.-M., Sayeed, H. M., Byun, H.-O., Jung, J.-Y., Kim, H.-A., et al. (2021). Eubacterium rectale attenuates HSV-1 induced systemic inflammation in mice by inhibiting CD83. *Front. Immunol.* 12, 712312. doi:10.3389/fimmu.2021.712312

Jin, M., Balan, V., Gunawan, C., and Dale, B. E. (2011). Consolidated bioprocessing (CBP) performance of Clostridium phytofermentans on AFEX-treated corn stover for ethanol production. *Biotechnol. Bioeng.* 108, 1290–1297. doi:10.1002/bit.23059

Jin, W.-B., Li, T.-T., Huo, D., Qu, S., Li, X. V., Arifuzzaman, M., et al. (2022). Genetic manipulation of gut microbes enables single-gene interrogation in a complex microbiome. *Cell* 185, 547–562.e22. doi:10.1016/j.cell.2021.12.035

Kim, D., Hager, M., Brant, E., and Budak, H. (2021). Efficient genome editing in wheat using Cas9 and Cpf1 (AsCpf1 and LbCpf1) nucleases. *Funct. Integr. Genomics* 21, 355–366. doi:10.1007/s10142-021-00782-z

Koeck, D. E., Ludwig, W., Wanner, G., Zverlov, V. V., Liebl, W., and Schwarz, W. H. (2015). Herbinix hemicellulosilytica gen. nov., sp. nov., a thermophilic cellulose-degrading bacterium isolated from a thermophilic biogas reactor. *Int. J. Syst. Evol. Microbiol.* 65, 2365–2371. doi:10.1099/ijs.0.000264

Koeck, D. E., Maus, I., Wibberg, D., Winkler, A., Zverlov, V. V., Liebl, W., et al. (2016). Complete genome sequence of Herbinix luporum SD1D, a new cellulose-degrading bacterium isolated from a thermophilic biogas reactor. *Genome Announc.* 4, 006877-16. doi:10.1128/genomeA.00687-16

Kong, F., Hua, Y., Zeng, B., Ning, R., Li, Y., and Zhao, J. (2016). Gut microbiota signatures of longevity. *Curr. Biol.* 26, R832–R833. doi:10.1016/j.cub.2016.08.015

Koopen, A., Witjes, J., Wortelboer, K., Majait, S., Prodan, A., Levin, E., et al. (2022). Duodenal Anaerobutyricum soehngenii infusion stimulates GLP-1 production, ameliorates glycaemic control and beneficially shapes the duodenal transcriptome in metabolic syndrome subjects: a randomised double-blind placebo-controlled cross-over study. *Gut* 71, 1577–1587. doi:10.1136/gutjnl-2020-323297

Krumholz, L. R., and Bryant, M. P. (1986). Syntrophococcus sucromutans sp. nov. gen. nov. uses carbohydrates as electron donors and formate, methoxymonobenzenoids or Methanobrevibacter as electron acceptor systems. *Arch. Microbiol.* 143, 313–318. doi:10.1007/BF00412795

Lapczynski, A., Bhatia, S. P., Letizia, C. S., and Api, A. M. (2008). Fragrance material review on farnesol. *Food Chem. Toxicol.* 46 (Suppl. 11), S149–S156. doi:10.1016/j.fct.2008.06.046

La Rosa, S. L., Leth, M. L., Michalak, L., Hansen, M. E., Pudlo, N. A., Glowacki, R., et al. (2019). The human gut Firmicute Roseburia intestinalis is a primary degrader of dietary β-mannans. *Nat. Commun.* 10, 905. doi:10.1038/s41467-019-08812-y

Leth, M. L., Ejby, M., Workman, C., Ewald, D. A., Pedersen, S. S., Sternberg, C., et al. (2018). Differential bacterial capture and transport preferences facilitate co-growth on dietary xylan in the human gut. *Nat. Microbiol.* 3, 570–580. doi:10.1038/s41564-018-0132-8

Liew, F., Martin, M. E., Tappel, R. C., Heijstra, B. D., Mihalcea, C., and Köpke, M. (2016). Gas fermentation—a flexible platform for commercial scale production of low-carbon-fuels and chemicals from waste and renewable feedstocks. *Front. Microbiol.* 7, 694. doi:10.3389/fmicb.2016.00694

Litvak, Y., Byndloss, M. X., and Bäumler, A. J. (2018). Colonocyte metabolism shapes the gut microbiota. *Science* 362, eaat9076. doi:10.1126/science.aat9076

Liu, C., Du, M.-X., Abuduaini, R., Yu, H.-Y., Li, D.-H., Wang, Y.-J., et al. (2021a). Enlightening the taxonomy darkness of human gut microbiomes with a cultured biobank. *Microbiome* 9, 119. doi:10.1186/s40168-021-01064-3

Liu, N., Chen, L., Yan, M., Tao, Q., Wu, J., Chen, J., et al. (2023). Eubacterium rectale improves the efficacy of anti-PD1 immunotherapy in melanoma via l-serine-mediated NK cell activation. *Research* 6, 0127. doi:10.34133/research.0127

Liu, X., Mao, B., Gu, J., Wu, J., Cui, S., Wang, G., et al. (2021b). Blautia-a new functional genus with potential probiotic properties? *Gut Microbes* 13, 1–21. doi:10.1080/19490976.2021.1875796

Lomans, B. P., Leijdekkers, P., Wesselink, J. J., Bakkes, P., Pol, A., van der Drift, C., et al. (2001). Obligate sulfide-dependent degradation of methoxylated aromatic compounds and formation of methanethiol and dimethyl sulfide by a freshwater sediment isolate, Parasporobacterium paucivorans gen. nov., sp. nov. *Appl. Environ. Microbiol.* 67, 4017–4023. doi:10.1128/AEM.67.9.4017-4023.2001

Lopez, C. A., McNeely, T. P., Nurmakova, K., Beavers, W. N., and Skaar, E. P. (2020). Clostridioides difficile proline fermentation in response to commensal clostridia. *Anaerobe* 63, 102210. doi:10.1016/j.anaerobe.2020.102210

Louie, T., Golan, Y., Khanna, S., Bobilev, D., Erpelding, N., Fratazzi, C., et al. (2023). VE303, a defined bacterial consortium, for prevention of recurrent Clostridioides difficile infection: a randomized clinical trial. *Jama* 329, 1356–1366. doi:10.1001/jama.2023.4314

Lu, H., Xu, X., Fu, D., Gu, Y., Fan, R., Yi, H., et al. (2022). Butyrate-producing Eubacterium rectale suppresses lymphomagenesis by alleviating the TNF-induced TLR4/MyD88/NF-κB axis. *Cell Host Microbe* 30, 1139–1150.e7. doi:10.1016/j.chom.2022.07.003

Mayo, B., Vázquez, L., and Flórez, A. B. (2019). Equol: a bacterial metabolite from the daidzein isoflavone and its presumed beneficial health effects. *Nutrients* 11, 2231. doi:10.3390/nu11092231

McGovern, B. H., Ford, C. B., Henn, M. R., Pardi, D. S., Khanna, S., Hohmann, E. L., et al. (2021). SER-109, an investigational microbiome drug to reduce recurrence after Clostridioides difficile infection: lessons learned from a phase 2 trial. *Clin. Infect. Dis.* 72, 2132–2140. doi:10.1093/cid/ciaa387

McLellan, S. L., Newton, R. J., Vandewalle, J. L., Shanks, O. C., Huse, S. M., Eren, A. M., et al. (2013). Sewage reflects the distribution of human faecal *Lachnospiraceae*. *Environ. Microbiol.* 15, 2213–2227. doi:10.1111/1462-2920.12092

Mechichi, T., Labat, M., Garcia, J. L., Thomas, P., and Patel, B. K. (1999). Sporobacterium olearium gen. nov., sp. nov., a new methanethiol-producing bacterium that degrades aromatic compounds, isolated from an olive mill wastewater treatment digester. *Int. J. Syst. Bacteriol.* 49 (Pt 4), 1741–1748. doi:10.1099/00207713-49-4-1741

Meehan, C. J., and Beiko, R. G. (2014). A phylogenomic view of ecological specialization in the Lachnospiraceae, a family of digestive tract-associated bacteria. *Genome Biol. Evol.* 6, 703–713. doi:10.1093/gbe/evu050

NCBI Genome Datasets (Lachnospiraceae) (2023). Genome, https://www.ncbi.nlm.nih.gov/datasets/genome/?taxon=186803&reference_only=true.

O Sheridan, P., Martin, J. C., Lawley, T. D., Browne, H. P., Harris, H. M. B., Bernalier-Donadille, A., et al. (2016). Polysaccharide utilization loci and nutritional specialization in a dominant group of butyrate-producing human colonic Firmicutes. *Microb. Genom* 2, e000043. doi:10.1099/mgen.0.000043

Park, H., Patel, A., Hunt, K. A., Henson, M. A., and Carlson, R. P. (2020). Artificial consortium demonstrates emergent properties of enhanced cellulosic-sugar degradation and biofuel synthesis. *NPJ Biofilms Microbiomes* 6, 59. doi:10.1038/s41522-020-00170-8

Petit, E., LaTouf, W. G., Coppi, M. V., Warnick, T. A., Currie, D., Romashko, I., et al. (2013). Involvement of a bacterial microcompartment in the metabolism of fucose and rhamnose by Clostridium phytofermentans. *PLoS ONE* 8, e54337. doi:10.1371/journal.pone.0054337

Petrof, E. O., Gloor, G. B., Vanner, S. J., Weese, S. J., Carter, D., Daigneault, M. C., et al. (2013). Stool substitute transplant therapy for the eradication of *Clostridium difficile* infection: 'RePOOPulating' the gut. *Microbiome* 1, 3. doi:10.1186/2049-2618-1-3

Port, F., Starostecka, M., and Boutros, M. (2020). Multiplexed conditional genome editing with Cas12a in Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* 117, 22890–22899. doi:10.1073/pnas.2004655117

Preethi, M., Kumar, G., Karthikeyan, O. P., and Varjani, S. (2021). Lignocellulosic biomass as an optimistic feedstock for the production of biofuels as valuable energy source: techno-economic analysis, Environmental Impact Analysis, Breakthrough and Perspectives. *Environ. Technol. Innovation* 24, 102080. doi:10.1016/j.eti.2021.102080

Quigley, E. M. M., Markinson, L., Stevenson, A., Treasure, F. P., and Lacy, B. E. (2023). Randomised clinical trial: efficacy and safety of the live biotherapeutic product MRx1234 in patients with irritable bowel syndrome. *Aliment. Pharmacol. Ther.* 57, 81–93. doi:10.1111/apt.17310

Reeves, A. E., Koenigsknecht, M. J., Bergin, I. L., and Young, V. B. (2012). Suppression of *Clostridium difficile* in the gastrointestinal tracts of germfree mice inoculated with a murine isolate from the family Lachnospiraceae. *Infect. Immun.* 80, 3786–3794. doi:10.1128/IAI.00647-12

Reichardt, N., Duncan, S. H., Young, P., Belenguer, A., McWilliam Leitch, C., Scott, K. P., et al. (2014). Phylogenetic distribution of three pathways for propionate production within the human gut microbiota. *ISME J.* 8, 1323–1335. doi:10.1038/ismej.2014.14

Roager, H. M., and Licht, T. R. (2018). Microbial tryptophan catabolites in health and disease. *Nat. Commun.* 9, 3294. doi:10.1038/s41467-018-05470-4

Rostain, W., Zaplana, T., Boutard, M., Baum, C., Tabuteau, S., Sanitha, M., et al. (2022). Tuning of gene expression in Clostridium phytofermentans using synthetic promoters and CRISPRi. *ACS Synth. Biol.* 11, 4077–4088. doi:10.1021/acssynbio.2c00385

Rude, M. A., and Schirmer, A. (2009). New microbial fuels: a biotech perspective. *Curr. Opin. Microbiol.* 12, 274–281. doi:10.1016/j.mib.2009.04.004

Russell, W. R., Duncan, S. H., Scobbie, L., Duncan, G., Cantlay, L., Calder, A. G., et al. (2013). Major phenylpropanoid-derived metabolites in the human gut can arise from microbial fermentation of protein. *Mol. Nutr. Food Res.* 57, 523–535. doi:10.1002/mnfr.201200594

Sato, Y., Atarashi, K., Plichta, D. R., Arai, Y., Sasajima, S., Kearney, S. M., et al. (2021). Novel bile acid biosynthetic pathways are enriched in the microbiome of centenarians. *Nature* 599, 458–464. doi:10.1038/s41586-021-03832-5

Scarborough, M. J., Myers, K. S., Donohue, T. J., and Noguera, D. R. (2020). Medium-chain fatty acid synthesis by "Candidatus Weimeria bifida" gen. Nov., sp. nov., and "Candidatus pseudoramibacter fermentans" sp. nov. *Appl. Environ. Microbiol.* 86, 022422-19. doi:10.1128/AEM.02242-19

Schouw, A., Leiknes Eide, T., Stokke, R., Pedersen, R. B., Steen, I. H., and Bødtker, G. (2016). Abyssivirga alkaniphila gen. nov., sp. nov., an alkane-degrading, anaerobic bacterium from a deep-sea hydrothermal vent system, and emended descriptions of Natranaerovirga pectinivora and Natranaerovirga hydrolytica. *Int. J. Syst. Evol. Microbiol.* 66, 1724–1734. doi:10.1099/ijsem.0.000934

Scott, K. P., Martin, J. C., Campbell, G., Mayer, C.-D., and Flint, H. J. (2006). Whole-genome transcription profiling reveals genes up-regulated by growth on fucose in the human gut bacterium "Roseburia inulinivorans.". *J. Bacteriol.* 188, 4340–4349. doi:10.1128/JB.00137-06

Sell, L. B., Ramelow, C. C., Kohl, H. M., Hoffman, K., Bains, J. K., Doyle, W. J., et al. (2022). Farnesol induces protection against murine CNS inflammatory demyelination and modifies gut microbiome. *Clin. Immunol.* 235, 108766. doi:10.1016/j.clim.2021.108766

Seshadri, R., Leahy, S. C., Attwood, G. T., Teh, K. H., Lambie, S. C., Cookson, A. L., et al. (2018). Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. *Nat. Biotechnol.* 36, 359–367. doi:10.1038/nbt.4110

Shen, Z., Luo, W., Tan, B., Nie, K., Deng, M., Wu, S., et al. (2022). Roseburia intestinalis stimulates TLR5-dependent intestinal immunity against Crohn's disease. *eBioMedicine* 85, 104285. doi:10.1016/j.ebiom.2022.104285

Sheridan, P. O., Louis, P., Tsompanidou, E., Shaw, S., Harmsen, H. J., Duncan, S. H., et al. (2022). Distribution, organization and expression of genes concerned with anaerobic lactate utilization in human intestinal bacteria. *Microb. Genom* 8, 000739. doi:10.1099/mgen.0.000739

Sheridan, P. O., Martin, J. C., Minton, N. P., Flint, H. J., O'Toole, P. W., and Scott, K. P. (2019). Heterologous gene expression in the human gut bacteria Eubacterium rectale and Roseburia inulinivorans by means of conjugative plasmids. *Anaerobe* 59, 131–140. doi:10.1016/j.anaerobe.2019.06.008

Silber, J., Norman, J., Kanno, T., Crossette, E., Szabady, R., Menon, R., et al. (2022). Randomized, double-blind, placebo (pbo)-controlled, single- and multiple-dose phase 1 study of ve202, a defined bacterial consortium for treatment of ibd: safety and colonization dynamics of a novel live biotherapeutic product (lbp) in healthy adults. *Inflamm. Bowel Dis.* 28, S65–S66. doi:10.1093/ibd/izac015.106

Smirnova, M., Miamin, U., Kohler, A., Valentovich, L., Akhremchuk, A., Sidarenka, A., et al. (2021). Isolation and characterization of fast-growing green snow bacteria from coastal East Antarctica. *MicrobiologyOpen* 10, e1152. doi:10.1002/mbo3.1152

Sorbara, M. T., Littmann, E. R., Fontana, E., Moody, T. U., Kohout, C. E., Gjonbalaj, M., et al. (2020). Functional and genomic variation between human-derived isolates of Lachnospiraceae reveals inter- and intra-species diversity. *Cell Host Microbe* 28, 134–146.e4. doi:10.1016/j.chom.2020.05.005

Stevenson, B. S., and Schmidt, T. M. (2004). Life history implications of rRNA gene copy number in *Escherichia coli*. *Appl. Environ. Microbiol.* 70, 6670–6677. doi:10.1128/AEM.70.11.6670-6677.2004

Sugiyama, Y., Masumori, N., Fukuta, F., Yoneta, A., Hida, T., Yamashita, T., et al. (2013). Influence of isoflavone intake and equol-producing intestinal flora on prostate cancer risk. *Asian Pac J. Cancer Prev.* 14, 1–4. doi:10.7314/apjcp.2013.14.1.1

Tak, Y. E., Kleinstiver, B. P., Nuñez, J. K., Hsu, J. Y., Horng, J. E., Gong, J., et al. (2017). Inducible and multiplex gene regulation using CRISPR-Cpf1-based transcription factors. *Nat. Methods* 14, 1163–1166. doi:10.1038/nmeth.4483

Takada, T., Kurakawa, T., Tsuji, H., and Nomoto, K. (2013). Fusicatenibacter saccharivorans gen. nov., sp. nov., isolated from human faeces. *Int. J. Syst. Evol. Microbiol.* 63, 3691–3696. doi:10.1099/ijs.0.045823-0

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38, 3022–3027. doi:10.1093/molbev/msab120

Tolonen, A. C., Beauchemin, N., Bayne, C., Li, L., Tan, J., Lee, J., et al. (2022). Synthetic glycans control gut microbiome structure and mitigate colitis in mice. *Nat. Commun.* 13, 1244. doi:10.1038/s41467-022-28856-x

Tolonen, A. C., Cerisy, T., El-Sayyed, H., Boutard, M., Salanoubat, M., and Church, G. M. (2015a). Fungal lysis by a soil bacterium fermenting cellulose. *Environ. Microbiol.* 17, 2618–2627. doi:10.1111/1462-2920.12495

Tolonen, A. C., Chilaka, A. C., and Church, G. M. (2009). Targeted gene inactivation in Clostridium phytofermentans shows that cellulose degradation requires the family 9 hydrolase Cphy3367. *Mol. Microbiol.* 74, 1300–1313. doi:10.1111/j.1365-2958.2009.06890.x

Tolonen, A. C., Haas, W., Chilaka, A. C., Aach, J., Gygi, S. P., and Church, G. M. (2011). Proteome-wide systems analysis of a cellulosic biofuel-producing microbe. *Mol. Syst. Biol.* 7, 461. doi:10.1038/msb.2010.116

Tolonen, A. C., Zuroff, T. R., Ramya, M., Boutard, M., Cerisy, T., and Curtis, W. R. (2015b). Physiology, genomics, and pathway engineering of an ethanol-tolerant strain of Clostridium phytofermentans. *Appl. Environ. Microbiol.* 81, 5440–5448. doi:10.1128/AEM.00619-15

Tran, M. H., Park, H., Nobles, C. L., Karunadharma, P., Pan, L., Zhong, G., et al. (2021). A more efficient CRISPR-Cas12a variant derived from Lachnospiraceae bacterium MA2020. *Mol. Ther. Nucleic Acids* 24, 40–53. doi:10.1016/j.omtn.2021.02.012

Uchiyama, J., Akiyama, M., Hase, K., Kumagai, Y., and Kim, Y.-G. (2022). Gut microbiota reinforce host antioxidant capacity via the generation of reactive sulfur species. *Cell Rep.* 38, 110479. doi:10.1016/j.celrep.2022.110479

Udayappan, S., Manneras-Holm, L., Chaplin-Scott, A., Belzer, C., Herrema, H., Dallinga-Thie, G. M., et al. (2016). Oral treatment with Eubacterium hallii improves insulin sensitivity in db/db mice. *NPJ Biofilms Microbiomes* 2, 16009. doi:10.1038/npjbiofilms.2016.9

Vacca, M., Celano, G., Calabrese, F. M., Portincasa, P., Gobbetti, M., and De Angelis, M. (2020). The controversial role of human gut Lachnospiraceae. *Microorganisms* 8, E573. doi:10.3390/microorganisms8040573

Vedanta Biosciences, Inc (2023). Vedanta Biosciences VE303 program. https://www.vedantabio.com/pipeline/ve303.

Vera-Ponce de Leon, A., Schneider, M. G., Jahnes, B. C., Sadowski, V., Camuy-Vélez, L. A., Duan, J., et al. (2022). Genetic drift and host-adaptive features likely underlie cladogenesis of insect-associated Lachnospiraceae. *Genome Biol. Evol.* 14, evac086. doi:10.1093/gbe/evac086

Vital, M., Howe, A. C., and Tiedje, J. M. (2014). Revealing the bacterial butyrate synthesis pathways by analyzing (meta)genomic data. *MBio* 5, e00889. doi:10.1128/mBio.00889-14

Wachenheim, D. E., and Patterson, J. A. (1992). Anaerobic production of extracellular polysaccharide by Butyrivibrio fibrisolvens nyx. *Appl. Environ. Microbiol.* 58, 385–391. doi:10.1128/aem.58.1.385-391.1992

Walker, A. C., Bhargava, R., Vaziriyan-Sani, A. S., Pourciau, C., Donahue, E. T., Dove, A. S., et al. (2021). Colonization of the *Caenorhabditis elegans* gut with human enteric bacterial pathogens leads to proteostasis disruption that is rescued by butyrate. *PLoS Pathog.* 17, e1009510. doi:10.1371/journal.ppat.1009510

Warnick, T. A., Methé, B. A., and Leschine, S. B. (2002). Clostridium phytofermentans sp. nov., a cellulolytic mesophile from forest soil. *Int. J. Syst. Evol. Microbiol.* 52, 1155–1160. doi:10.1099/ijs.0.02125-0

Watanabe, M., Kaku, N., Ueki, K., and Ueki, A. (2016). Falcatimonas natans gen. nov., sp. nov., a strictly anaerobic, amino-acid-decomposing bacterium isolated from a methanogenic reactor of cattle waste. *Int. J. Syst. Evol. Microbiol.* 66, 4639–4644. doi:10.1099/ijsem.0.001403

Whitman, W. B. (2009). *Systematic bacteriology*. New York, NY, USA: Springer. doi:10.1007/978-0-387-68489-5

Wolin, M. J., Miller, T. L., Collins, M. D., and Lawson, P. A. (2003). Formate-dependent growth and homoacetogenic fermentation by a bacterium from human feces: description of Bryantella formatexigens gen. nov., sp. nov. *Appl. Environ. Microbiol.* 69, 6321–6326. doi:10.1128/AEM.69.10.6321-6326.2003

Wu, C., Xu, W., Yu, W., Li, Z., Zhang, Y., and Zhang, F. (2021). Strain-level screening of human gut microbes identifies Blautia producta as a novel anti-hyperlipidemic probiotic via the production of 12-methylmyristic acid. https://www.researchsquare.com/article/rs-989302/v1.

Yamashita, H., Fujisawa, K., Ito, E., Idei, S., Kawaguchi, N., Kimoto, M., et al. (2007). Improvement of obesity and glucose tolerance by acetate in Type 2 diabetic Otsuka Long-Evans Tokushima Fatty (OLETF) rats. *Biosci. Biotechnol. Biochem.* 71, 1236–1243. doi:10.1271/bbb.60668

Yang, C., Guan, L., Liu, J., and Wang, J. (2015). Rumen fermentation and acetogen population changes in response to an exogenous acetogen TWA4 strain and *Saccharomyces cerevisiae* fermentation product. *J. Zhejiang Univ. Sci. B* 16, 709–719. doi:10.1631/jzus.B1500013

Zhang, J., Song, L., Wang, Y., Liu, C., Zhang, L., Zhu, S., et al. (2019). Beneficial effect of butyrate-producing Lachnospiraceae on stress-induced visceral hypersensitivity in rats. *J. Gastroenterol. Hepatol.* 34, 1368–1376. doi:10.1111/jgh.14536

Zhang, L., Li, G., Zhang, Y., Cheng, Y., Roberts, N., Glenn, S. E., et al. (2023). Boosting genome editing efficiency in human cells and plants with novel LbCas12a variants. *Genome Biol.* 24, 102. doi:10.1186/s13059-023-02929-6

Zhou, Y., Wei, Y., Jiang, L., Jiao, X., and Zhang, Y. (2023). Anaerobic phloroglucinol degradation by Clostridium scatologenes. *mBio* 14, e0109923–23. doi:10.1128/mbio.01099-23

Zuroff, T. R., Barri Xiques, S., and Curtis, W. R. (2013). Consortia-mediated bioprocessing of cellulose to ethanol with a symbiotic Clostridium phytofermentans/yeast co-culture. *Biotechnol. Biofuels* 6, 59. doi:10.1186/1754-6834-6-59

AMERICAN SOCIETY FOR MICROBIOLOGY | mSystems

Physiology and Metabolism | Research Article

L-tryptophan and copper interactions linked to reduced colibactin genotoxicity in *pks+ Escherichia coli*

Charlie Bayne,[1] **Magali Boutard,**[2] **Tom Zaplana,**[2] **Andrew C. Tolonen**[2]

AUTHOR AFFILIATIONS See affiliation list on p. 13.

ABSTRACT Colibactin, a nonribosomal peptide/polyketide produced by *pks+ Enterobacteriaceae*, is a virulence factor and putative carcinogen that damages DNA by interstrand crosslinking (ICL). While the *clb* genes for colibactin biosynthesis have been identified, studies are needed to elucidate the mechanisms regulating colibactin production and activity. Here we perform untargeted metabolomics of *pks+ Escherichia* coli cultures to identify L-tryptophan as a candidate repressor of colibactin activity. When *pks+ E. coli* is grown in a minimal medium supplemented with L-tryptophan *in vitro* ICL of plasmid DNA is reduced by >80%. L-tryptophan does not affect the transcription of *clb* genes but protects from copper toxicity and triggers the expression of genes to export copper to the periplasm where copper can directly inhibit the ClbP peptidase domain. Thus, L-tryptophan and copper interact and repress colibactin activity, potentially reducing its carcinogenic effects in the intestine.

IMPORTANCE Colibactin is a small molecule produced by *pks+ Enterobacteriaceae* that damages DNA, leading to oncogenic mutations in human genomes. Colibactin-producing *Escherichia coli* (*pks+*) cells promote tumorigenesis in mouse models of colorectal cancer (CRC) and are elevated in abundance in CRC patient biopsies, making it important to identify the regulatory systems governing colibactin production. Here, we apply a systems biology approach to explore metabolite repression of colibactin production in *pks+ E. coli*. We identify L-tryptophan as a repressor of colibactin genotoxicity that stimulates the expression of genes to export copper to the periplasm where it can inhibit ClbP, the colibactin-activating peptidase. These results work toward an antibiotic-sparing, prophylactic strategy to inhibit colibactin genotoxicity and its tumorigenic effects in the intestine.

KEYWORDS colibactin, *Escherichia coli*, metabolomics

The human gut microbiome produces numerous metabolites that affect the initiation and progression of disease. Among these metabolites, colibactin is a nonribosomal peptide/polyketide hybrid natural product synthesized by various *Enterobacteriaceae,* including a subset of isolates belonging to the genera *Escherichia*, *Klebsiella*, and *Citrobacter* (1). Colibactin is a procarcinogen that alkylates DNA, causing interstrand crosslinks (ICLs), double-strand breaks, and increased levels of gamma-H2aX in eukaryotic cells (2, 3). The genotoxic effects of colibactin result in host cell blockage at the G2/M transition (2), leading to megalocytosis and impeding regeneration of the intestinal epithelium, which could increase bacterial fitness by prolonging intestinal colonization (4). Mouse models of sepsis (5) and a rat model of meningitis (6) established colibactin as a virulence factor of extraintestinal pathogenic *Escherichia coli* that are associated with urinary tract infections, neonatal meningitis, sepsis, and pneumonia (7). Evidence increasingly supports that colibactin has a role in the etiology of colorectal cancer (CRC): colibactin-producing *E. coli* cells promote tumorigenesis in mouse models

Editor Jeffrey Lawrence Blanchard, University of Massachusetts Amherst, Amherst, Massachusetts, USA

Address correspondence to Andrew C. Tolonen, atolonen@genoscope.cns.fr.

The authors declare no conflict of interest.

Received 26 July 2024
Accepted 6 August 2024
Published 12 September 2024

Copyright © 2024 Bayne et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Downloaded from https://journals.asm.org/journal/msystems on 11 October 2024 by 2a01:e34:ec23:6090:bc02:eb34:142:90b3.

of CRC and are elevated in abundance in CRC patient biopsies (60%) relative to non-CRC controls (20%) (8, 9), and the colibactin mutational signature is present in human cancer genomes (10, 11).

Biosynthesis of colibactin requires a 54 kb *pks* island containing 19 *clb* genes (*clbA* to *clbS*), encoding nonribosomal peptide synthases (NRPS), polyketide synthases (PKS), NRPS/PKS megasynthases, an efflux pump, and accessory and processing enzymes (12). Synthesis of colibactin involves multimodular assembly of precolibactin by NRPS/PKS. Ultimately, active colibactin is produced using a prodrug resistance mechanism based on the cleavage of the N-terminal myristoyl-asparagine from precolibactin by ClbP, a periplasmic peptidase anchored in the inner membrane (13, 14). While colibactin abundance has eluded direct measurement due to its instability, studies have made advances to elucidate the structural basis of colibactin and its DNA cross-linking activity (15, 16). In addition, expression of the *clb* genes has been shown to be regulated by environmental factors including iron (17), polyamines (18), and carbon sources (19).

The goal of this study is to identify metabolites that repress or inhibit colibactin activity using the model *pks*+ *E. coli* ATCC 25922, a strain encoding a *clb* cluster with 99%–100% identity to those of uropathogenic strain CFT073, probiotic Nissle1917, and newborn meningitis strain IHE3034 (20). The cellular potential to produce mature colibactin from precolibactin can be quantified using a fluorogenic probe of ClbP activity (21, 22). Here, we used an activity-based probe that releases fluorescent 7-hydroxy-4-methylcoumarin (7H4M) upon cleavage by ClbP (21) to identify differences in ClbP activity in *E. coli* ATCC 25922 growing in different conditions. We perform a time series of untargeted metabolomics in growth conditions with different ClbP activities, revealing tryptophan as a potential repressor. We examine how L-tryptophan supplementation in *E. coli* ATCC 25922 cultures affects colibactin genotoxicity and genome-wide transcription, identifying a link between L-tryptophan and copper export. We therefore investigate how copper affects ClbP activity *in vivo* and the activity of the purified ClbP peptidase domain. Together, our experiments support that L-tryptophan is linked to copper export and reduced ClbP activity, representing a potential strategy to reduce colibactin genotoxicity in the intestine.

RESULTS

ClbP activity varies across growth conditions

Expression of colibactin genes (*clb* genes) in *pks*+ *E. coli* is affected by growth state and medium composition (23). For example, transcription of *clbR,* encoding a transcriptional activator of colibactin genes, is elevated in minimal (M9) medium relative to rich media (24). We examined the activity of ClbP, the periplasmic colibactin-activating peptidase, in *pks*+ *E. coli* ATCC 25922 growing in different media: chopped meat glucose (CMG) and mega medium (MM29) are rich media and MS is a defined, minimal medium (Fig. 1A). To quantify ClbP activity, we synthesized a ClbP activity-based probe (21) (Fig. S1), hereafter ClbP-17. Upon cleavage by ClbP, ClbP-17 releases 7H4M, which can be monitored by fluorescence (Fig. S2A). We first established that ClbP-17 fluorescence is specific to *pks*+ *E. coli* (Fig. S2B through E) and that the addition of ClbP-17 or 7H4M did not affect *E. coli* growth in CMG, MM29, or MS media (Fig. S3).

During our tests of ClbP-17, we observed that the fluorescence of 7H4M is sensitive to pH in the physiological range with a 15-fold variation in fluorescence between pH 6 and 8 (Fig. S4). Differences in pH during *E. coli* growth would thus influence fluorescence measurements using ClbP-17. The pH in cultures can be measured in real-time based on the fluorescence of BCECF [2,7-bis-(2-carboxyethyl)5-(and 6)-carboxyfluorescein] (25, 26). We confirmed that BCECF does not affect the growth of *E. coli* ATCC 25922 (Fig. S3) and developed this method to quantify the pH dynamics of ATCC 25922 growing in MS, MM29, and CMG media (Fig. S5). The pH of cultures in all three media initially decreased as the cells produced acids during log-phase growth, then increased as cells entered the stationary phase (Fig. 1B). The greatest pH increase occurred in the CMG medium, likely due to enhanced bacterial liberation of ammonia associated with increased amino acid

Downloaded from https://journals.asm.org/journal/msystems on 11 October 2024 by 2a01:e34:ec23:6090:bc02:eb34:142:90b3.

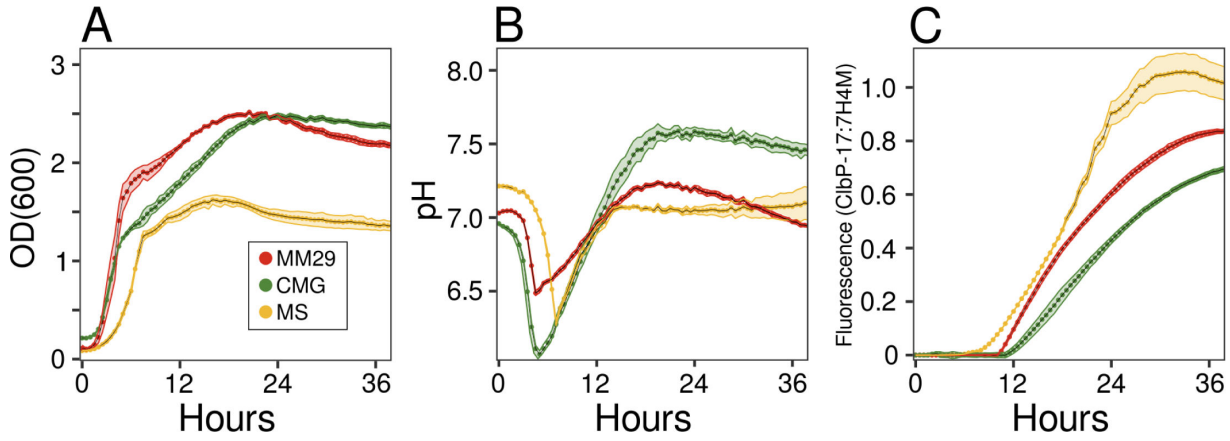


FIG 1 Cultivation of *E. coli* ATCC 25922 in different media. The (A) growth, (B) pH, and (C) ClbP-17:7H4M fluorescence of cultures growing in MS (yellow), MM29 (red), or CMG (green) media. Cultures contained 2 µM BCECF and 100 µM of either ClbP-17 or 7H4M. Growth was measured as optical density at 600 nm (OD_{600}). Culture pH was measured as the ratio of the fluorescence of BCECF at the pH-sensitive point (485 ± 15 nm excitation; 540 ± 20 nm emission) relative to the pH-insensitive isosbestic point (450 ± 15 nm excitation; 540 ± 20 nm emission). Fluorescence of ClbP-17 and 7H4M were measured as 360 ± 40 nm excitation, and 440 ± 20 nm emission. ClbP-17:7H4M is the fluorescence of ClbP-17 divided by the median 7H4M fluorescence of replicate cultures. Data points are means of triplicate cultures with shaded areas showing ±SD. BCECF, [2,7-bis-(2-carboxyethyl)5-(and-6)-carboxyfluorescein]; SD, standard deviation.

metabolism. Changes in pH during growth of *E. coli* cultures mirrored changes in 7H4M fluorescence (Fig. S6), confirming that growth-associated changes in pH drive variation in 7H4M fluorescence. To control for variation in 7H4M fluorescence, we thus present ClbP-17 measurements as the ClbP-17:7H4M ratio by dividing ClbP-17 fluorescence by the median 7H4M fluorescence of triplicate samples incubated under the same conditions in adjacent wells.

Growth rate and cell densities in *E. coli* ATCC 25922 cultures were lower in MS than in MM29 or CMG media (Fig. 1A). In contrast, ClbP-17:7H4M fluorescence was highest in MS medium, intermediate in MM29, and lowest in CMG medium (Fig. 1C). This inverse relationship between growth and ClbP-17:7H4M fluorescence in ATCC 25922 reflects prior results that *clbR* transcription is higher in minimal medium (24). As an expression of *clb* genes is repressed by the carbon storage regulatory protein CsrA (27, 28), we hypothesized that a metabolite that represses ClbP in the CMG medium is absent in the defined medium and present in lower quantities in the MM29 medium.

Metabolomics identifies tryptophan as a potential ClbP repressor

To find candidate metabolites that repress ClbP activity, we performed untargeted metabolomics on a time series of supernatants from ATCC 25922 cultures growing in CMG or MM29 media (Table S1A). We reasoned that the abundance of a metabolite that represses ClbP activity in CMG relative to MM29 medium would be progressively depleted in MM29 medium, but remain replete in CMG. Upon hierarchical clustering of all 1,515 metabolite features across the time series (Fig. 2A), we identified a cluster of 24 metabolites with relative abundance profiles consistent with our proposed profile of a ClbP repressor (Fig. 2B; Fig. S7; Table S1B). Most features in this cluster showed similar retention times near 420 seconds (Fig. 2B and C; Table S1B), suggesting they have comparable hydrophobicities.

To characterize this cluster of metabolite features selectively depleted in MM29 medium, we performed LC-MS/MS on the cluster members (Table S2A) and queried the MS/MS spectra against the mzCloud database to putatively identify seven metabolite features with HighChem HighRes identity scores >50 (Table S2B). We identified two metabolite features, M205T421 and M206T421, for which tryptophan was the top hit in the mzCloud database. The *m/z* of their precursor ions is consistent with the monoisotopic mass of protonated tryptophan (M205T421) or a protonated tryptophan isotopologue (M206T421) (Table S2B). M205T421 and M206T421 have similar retention times of

Downloaded from https://journals.asm.org/journal/msystems on 11 October 2024 by 2a01:e34:ec23:6090:bc02:eb34:142:90b3.

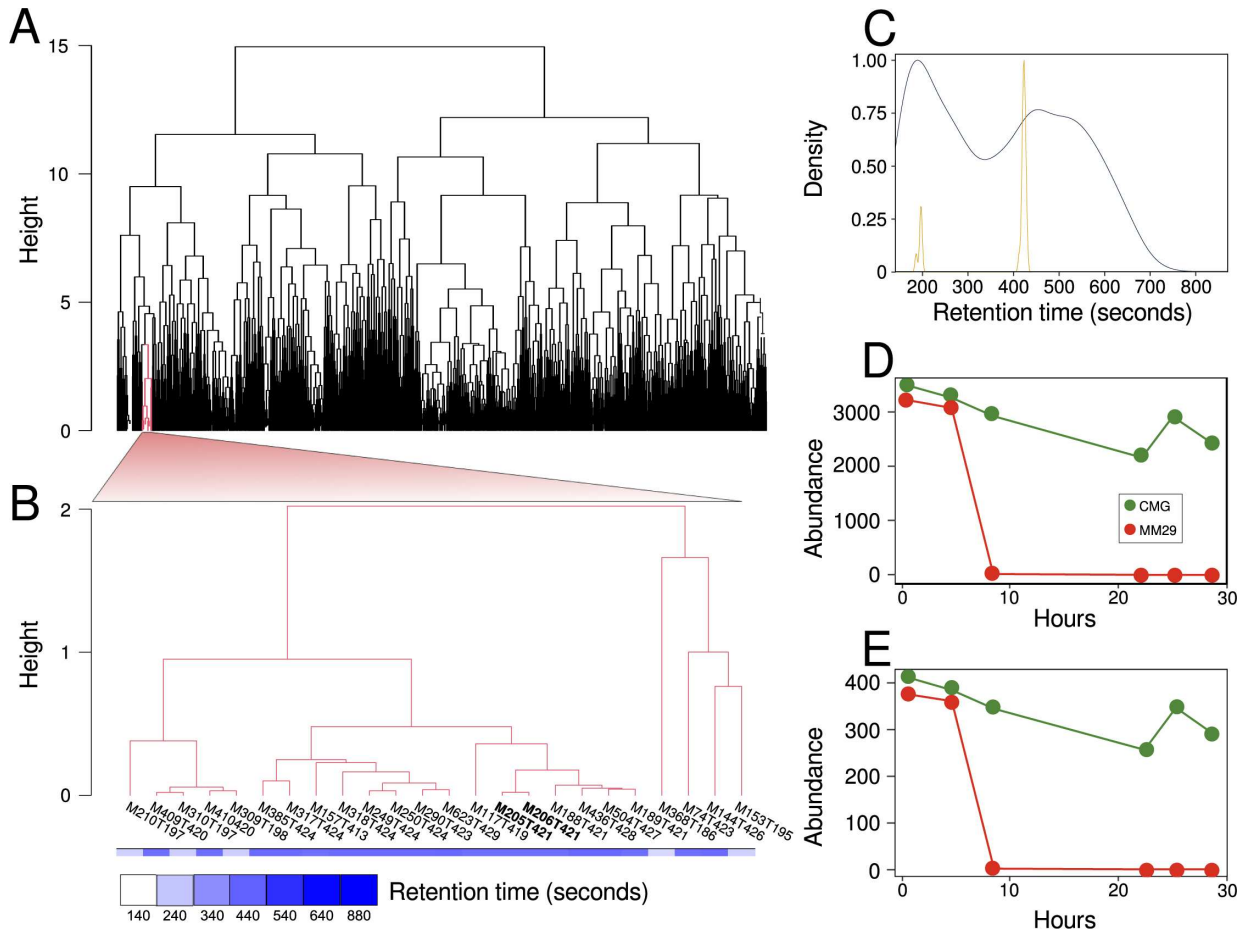


FIG 2 Identification of candidate repressors of ClbP activity in *E. coli* ATCC 25922 cultures using untargeted metabolomics. Metabolite abundances were measured by LC-MS for cultures growing in either MM29 or CMG medium at six-time points (0, 4, 8, 22, 25, and 29 hours) after inoculation. (A) Dendrogram showing hierarchical clustering of the abundance profiles for 1,515 metabolite features based on a dissimilarity matrix produced using dynamic time warping. (B) Dendrogram of metabolite features in the target cluster showing selective depletion in MM29 relative to CMG medium. Colored bars below dendrograms show retention time. (A and B) The height (distance) between cluster members using the complete agglomeration method is shown on the *y*-axis. (C) Kernel density estimation shows the distribution of retention times for all 1,515 metabolites (blue) and 24 metabolites in the target cluster (yellow). (D) Relative abundance profiles of metabolite spectra for (D) M205T421 and (E) M206T421 that both have tryptophan as the top hit in the mzCloud database. Metabolite abundances are based on peak areas of extracted ion chromatograms. LC-MS, liquid chromatography–mass spectrometry.

420.33 and 420.20 seconds (Table S1B) and their relative abundance profiles are highly similar showing depletion in MM29 versus CMG medium (Fig. 2D and E).

Tryptophan reduces ClbP activity and DNA crosslinking

We investigated how L-tryptophan supplementation of *E. coli* ATCC 25922 cultures in defined (MS) medium affected growth and ClbP activity. Supplementation of MS medium with up to 50 mM L-tryptophan did not affect growth (Fig. 3A). However, 7H4M fluorescence was 40% reduced by adding 25 mM L-tryptophan to MS medium without affecting pH (Fig. S8), complicating the use of ClbP-17 to measure the effect of L-tryptophan on ClbP activity. More generally, the sensitivity of 7H4M fluorescence to pH (Fig. S4), L-tryptophan (Fig. S8), and likely other metabolites show the challenges of using the ClbP-17 activity-based probe to investigate ClbP activity.

As an alternative assay to ClbP-17 to assay colibactin, we tested if L-tryptophan supplementation reduced *in vitro* cross-linking (ICL) of DNA. DNA damage by colibactin results from intercalation by the thiazoline-thiazole tail and subsequent alkylation by the electrophilic cyclopropane, resulting in covalent crosslinking of complementary DNA

Downloaded from https://journals.asm.org/journal/msystems on 11 October 2024 by 2a01:e34:ec23:6090:bc02:eb34:142:90b3.

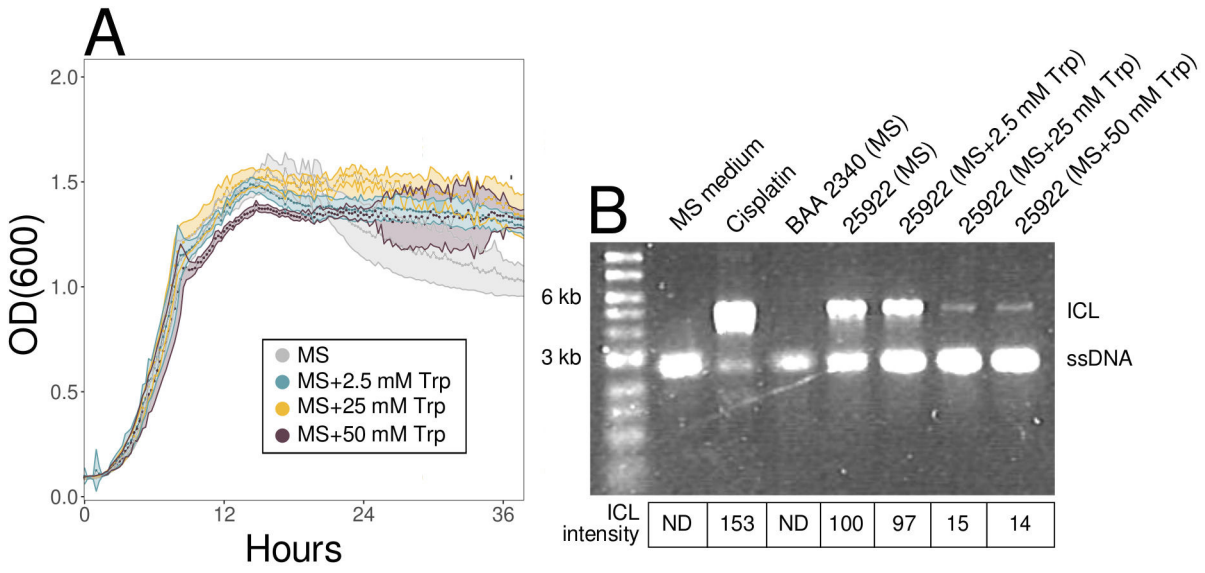


FIG 3 Effect of L-tryptophan supplementation on growth and DNA crosslinking activity of *pks*+ *E. coli* ATCC 25922 cultures. (A) Growth of *E. coli* ATCC 25922 cultures in MS medium (gray) or MS medium supplemented with either 2.5 mM (light blue), 25 mM (yellow), or 50 mM (purple) L-tryptophan. Data points are means of triplicate cultures, shaded regions are ±SD. (B) DNA crosslinking assay performed on *Bam*HI-linearized pUC19 DNA that was incubated for 6 hours at 37°C with either MS medium, *E. coli* ATCC BAA 2340 (*pks*–) in MS medium, *E. coli* ATCC 25922 (*pks*+) in MS medium at different Trp concentrations, or 80 µM cisplatin. pUC19 single-stranded DNA is at 2,686 bp; pUC19 ICL DNA is at 5,372 bp. Percent ICL band intensities relative to ATCC 25922 cultures in MS medium shown below gel image. OD_{600}, optical density at 600 nm; L-tryptophan, Trp; ICL, interstrand crosslinked; ND, not detected.

strands (15, 16, 29). As ICL prevents the denaturation of DNA strands, it can be observed as an apparent doubling of the molecular weight of linearized plasmid DNA on a denaturing agarose gel (29). As expected, treatment with the drug cisplatin resulted in nearly complete ICL (Fig. 3B; Fig. S9A). DNA crosslinking occurred with *E. coli* ATCC 25922 (*pks*+), but not *E. coli* ATCC BAA 2340 (*pks*–), when grown in MS medium (Fig. 3B; Fig. S9A). Quantification of band intensities showed that relative to MS medium alone, supplementation of *E. coli* ATCC 25922 cultures with 25 mM or 50 mM L-tryptophan reduced ICL by 85% and 86%, respectively (Fig. 3B; Fig. S9A). These results support that supplementation of *pks*+ *E.coli* cultures with L-tryptophan reduces colibactin-mediated DNA damage.

Tryptophan elevates gene expression for ammonium assimilation and copper export

To examine the mechanism by which L-tryptophan may reduce colibactin activity, we quantified genome-wide mRNA expression differences between log phase *E. coli* ATCC 25952 cultures growing in MS medium with 25 mM L-tryptophan supplementation (+Trp) versus MS medium alone (–Trp) (Table S3). As expected, *tnaA*, encoding the tryptophanase that splits tryptophan into indole, pyruvate, and ammonia, and the *tnaB* tryptophan permease were highly upregulated in the +Trp treatment (Fig. 4A). To assimilate the liberated ammonia, genes were upregulated to synthesize carbamoyl phosphate (*carA*) and ornithine (*argABCD*) and to incorporate these metabolites into arginine (*argFGH*) (Fig. 4A and B). Also as expected, the *trp* genes for tryptophan biosynthesis were highly repressed in the +Trp treatment (Fig. 4A). However, *clbP* was not differentially expressed in response to L-tryptophan and *clbI* was the only gene among 19 genes in the *clb* island (*clbA-clbS*) that was repressed (adjusted *P*-value < 0.01, Table S3). Thus, reduced colibactin activity upon L-tryptophan supplementation is unlikely to be the result of transcriptional repression of the *clb* genes.

Genes encoding all three systems to mitigate copper toxicity in *E. coli* (30) were highly upregulated in response to L-tryptophan. The *cus* genes (*cusCBA, cusF*, and *cusR*) and

Downloaded from https://journals.asm.org/journal/msystems on 11 October 2024 by 2a01:e34:ec23:6090:bc02:eb34:142:90b3.

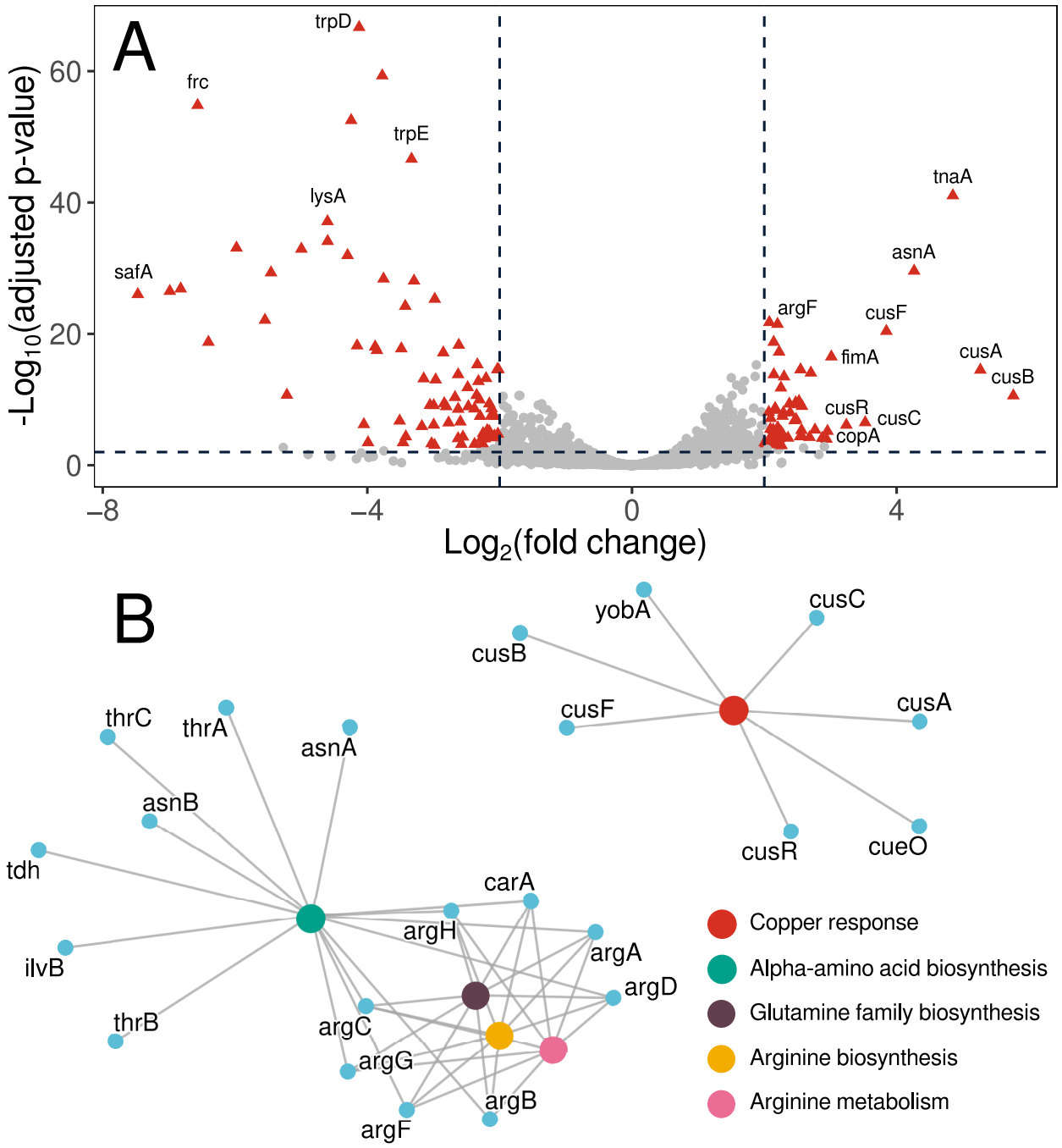


FIG 4 Gene expression of *E. coli* ATCC 25922 cultures growing in MS medium with and without 25 mM L-tryptophan supplementation measured by RNA-seq. (A) Volcano plot of mRNA expression in L-tryptophan supplemented cultures relative to MS medium. The 137 differentially-expressed [>2 $\log_2$(fold change), *P*-value < 0.001] genes are shown as red triangles; all other genes are gray circles. Names of selected genes are shown. (B) Gene networks showing Gene Ontology linkages of functional pathways whose expression increased upon L-tryptophan supplementation (*P*-value < 0.01). The *P*-values were corrected for multiple hypothesis testing using the Benjamini and Hochberg method.

copA were among the 10 most highly upregulated genes in the +Trp treatment; *cueO* was also upregulated to a lesser extent (Fig. 4B; Table S3). The CusCBA efflux system is an RND-type efflux system that transports copper ions from the cytosol to the periplasm and outside of the cell (31), CopA is a P-type ATPase that transports cytoplasmic copper to the periplasm (32), and CueO oxidizes Cu(I) to the less toxic Cu(II) (33).

Downloaded from https://journals.asm.org/journal/msystems on 11 October 2024 by 2a01:e34:ec23:6090:bc02:eb34:142:90b3.

Tryptophan and copper interact to alter growth, colibactin activity

To investigate connections between L-tryptophan and copper export, we measured the growth of ATCC 25922 cultures at different copper chloride concentrations in MS medium (−Trp) and in MS medium with 25 mM L-tryptophan (+Trp). Growth of −Trp cultures was slightly reduced by 300 µM copper (maximum cell density 14% lower relative to 0.3 µM copper) and strongly inhibited by 3 mM copper (maximum cell densities 75% lower relative to 0.3 µM copper), whereas growth in +Trp cultures was not affected by these copper concentrations (Fig. 5A). L-tryptophan thus protects *E. coli* ATCC 25922 from copper toxicity, likely by sequestering copper and reducing its intracellular accumulation.

As L-tryptophan stimulated the expression of the gene to efflux copper from the cell, we examined the effect of adding copper to the media on ClbP-17 fluorescence. In contrast to L-tryptophan, increased copper concentration had little effect on 7H4M fluorescence. Fluorescence of 7H4M was only 4% lower when the copper concentration in the MS medium was increased from 0.3 µM to 300 µM (Fig. S8B), supporting that ClbP-17 can be used to assess the effect of copper on ClbP activity. We observed a reduction in ClbP-17:7H4M fluorescence in ATCC 25922 cultures when copper was increased from 0.3 µM to 30 µM or 300 µM (Fig. 5B). Similarly, copper supplementation inhibited ICL of plasmid DNA by ATCC 25922 (Fig. 5C; Fig. S9B). Quantification of band intensities showed that relative to MS medium alone, ICL of plasmid DNA was reduced by supplementation with either 25 mM tryptophan, 30 µM copper, or 300 µM copper by 94%, 95%, or >99%, respectively (Fig. 5C). Reduction of ClbP-17 fluorescence and ICL activity by copper in *E. coli* ATCC 25922 cultures was particularly interesting as copper has previously been shown to increase both the degradation and the *in vitro* DSB activity of a purified colibactin metabolite (34).

Based on copper-reducing ClbP-17:7H4M fluorescence and ICL of extracellular DNA in ATCC 25922 cultures, we hypothesized that copper may directly inhibit the peptidase function of ClbP. We purified the ClbP peptidase domain (ClbP-pep) using a C-terminal His tag and confirmed its ability to cleave ClbP-17 (Fig. S10). We then measured the effect of increasing copper concentrations on ClbP-17 cleavage by ClbP-pep. ClbP-17 cleavage by ClbP-pep is progressively inhibited by increasing copper concentrations, whereas 7H4M fluorescence is unaffected (Fig. 5D). While the activity profile of ClbP-pep has been found to be different from full-length ClbP (14), our results suggest that the reduction in ClbP activity and ICL in ATCC 25922 cultures supplemented with copper (Fig. 5B and C) is associated with a direct effect of copper inhibiting ClbP-pep (Fig. 5D). Together, these observations suggest that copper export to the periplasm could be responsible for the L-tryptophan-mediated reductions in ClbP activity and DNA damage.

DISCUSSION

Colibactin-encoding *E. coli* cells are present in 60% of humans with familial adenomatous polyposis or CRC and the colibactin mutational signature is present in human cancer patient cells (10, 11), highlighting the importance of developing strategies to reduce colibactin activity. In this study, we identified metabolites that repress ClbP activity (ClbP-17:7H4M fluorescence) and the DNA crosslinking activity (ICL) of colibactin in *pks*+ *E. coli*. Untargeted metabolomics revealed L-tryptophan as a potential ClbP repressor, which was supported by supplementing cultures with L-tryptophan to reduce extracellular DNA crosslinking (Fig. 3B). L-tryptophan supplementation did not affect *clb* gene expression, but stimulated the expression of copper export genes (Fig. 4A) and protected from copper toxicity (Fig. 5A). Increasing copper in ATCC 25922 cultures reduced ClbP-17:7H4M fluorescence (Fig. 5B) and extracellular DNA crosslinking (Fig. 5C), and inhibited the *in vitro* activity of purified ClbP-pep (Fig. 5D).

Similar to ClbP, copper inhibits the activities of other proteases such as *Helicobacter pylori* HtrA (35), the 20S proteasome (36), and the protease from human immunodeficiency virus 1 (37). Copper-mediated inhibition of ClbP could be a mechanism to regulate colibactin production to protect *pks*+ cells from autotoxicity, similar to ClbS

Downloaded from https://journals.asm.org/journal/msystems on 11 October 2024 by 2a01:e34:ec23:6090:bc02:eb34:142:90b3.

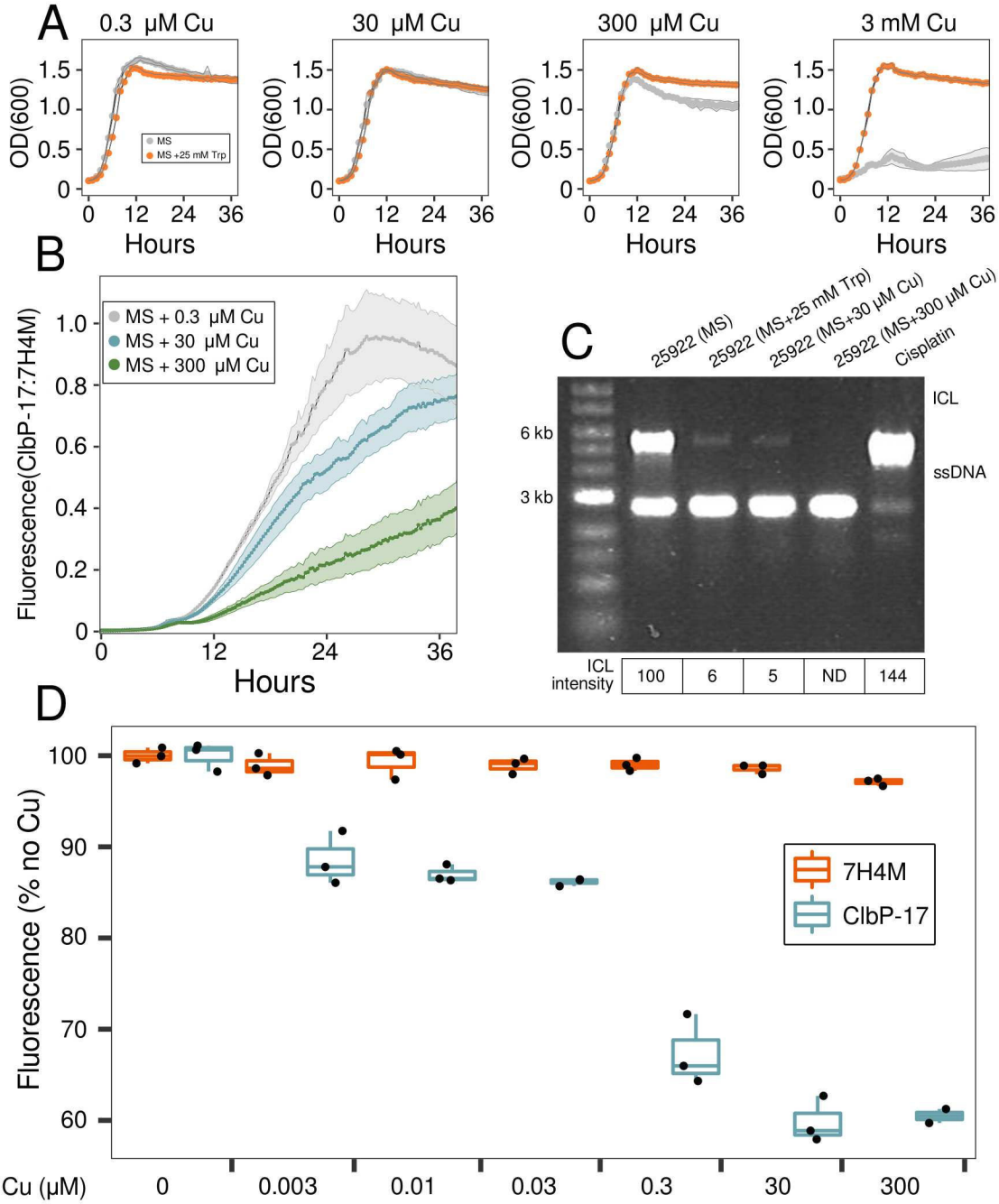


FIG 5 Effect of L-tryptophan and copper supplementation on *E. coli* ATCC 25922 growth, ClbP activity, and DNA cross-linking. (A) Growth of cultures in MS medium (gray) or in MS medium with 25 mM tryptophan (orange) at different copper concentrations (shown above plots). Data points are means of triplicate cultures, shaded areas are ±SD. (B) ClbP-17:7H4M fluorescence of cultures growing in MS medium containing either 0.3 µM (gray), 30 µM (blue), or 300 µM (green) copper chloride. Fluorescence of ClbP-17 and 7H4M was measured as 360 ± 40 nm excitation, and 440 ± 20 nm emission in triplicate cultures. ClbP-17:7H4M is the fluorescence of ClbP-17 divided by the median 7H4M fluorescence. (C) DNA crosslinking assay performed on *BamH*I-linearized pUC19 DNA that was incubated for 6 hours at 37°C with 80 µM cisplatin or with cultures of *E. coli* ATCC 25922 in MS medium (0.3 µM Cu) or MS medium supplemented with either 25 mM Trp or Cu (30 µM, 300 µM). pUC19 single-stranded DNA is at 2,686 bp; pUC19 ICL DNA is at 5,372 bp. ICL band intensities relative to ATCC 25922 cultures in MS medium are shown below the gel image. (D) Activity of purified ClbP peptidase domain (ClbP-pep) in buffer containing different Cu concentrations. ClbP-17 fluorescence (blue) is shown as a percentage of the mean fluorescence in the absence of copper and is compared to 7H4M fluorescence (red) at the same Cu concentrations. Box plots show median and interquartile ranges; data points are individual measurements. (B, D) ClbP-17 and 7H4M fluorescence were measured at 360 ± 40 nm excitation, and 440 ± 20 nm emission. OD_{600}, optical density at 600 nm; SD, standard deviation; Cu, copper; Trp, L-tryptophan; ICL, interstrand crosslinked DNA; ND, not detected.

Downloaded from https://journals.asm.org/journal/msystems on 11 October 2024 by 2a01:e34:ec23:6090:bc02:eb34:142:90b3.

that hydrolyses the cyclopropane warhead of colibactin (38, 39). In a previous study, copper enhanced both the instability and the activity of a colibactin metabolite that acts by a copper-mediated oxidative mechanism to induce DNA double-strand breaks (34). Combined with our data that supplementation of ATCC 25922 cultures with copper reduces DNA crosslinking (Fig. 5C), it appears that copper has a dual role to both limit and enhance the DNA-damaging activity of colibactin metabolites.

We integrated our data into a model of how L-tryptophan and copper interact to repress ClbP activity and DNA crosslinking by *pks+ E. coli* (Fig. 6). In this model, (i) L-tryptophan uptake stimulates the expression of copper exporters. While the mechanism by which L-tryptophan supplementation stimulates the expression of copper exporters is unknown, cellular amino acids sequester copper (40) and, in particular, tryptophan complexes with copper (41). The uptake of tryptophan-copper complexes could lead to intracellular copper accumulation and consequent induction of the copper export genes. Furthermore, the protective effect of L-tryptophan from copper toxicity suggests that either tryptophan increases the efficiency of copper export or that intracellular tryptophan mitigates copper toxicity. (ii) Expression of copper exporters pump copper into the periplasm and extracellular milieu. (iii) Copper in the periplasm could inhibit ClbP, resulting in lower processing of precolibactin into mature colibactin. (iv) Colibactin, perhaps complexed with copper, mediates DNA damage outside the cell. (v) DNA damage by colibactin in the producing cells is mitigated by ClbS.

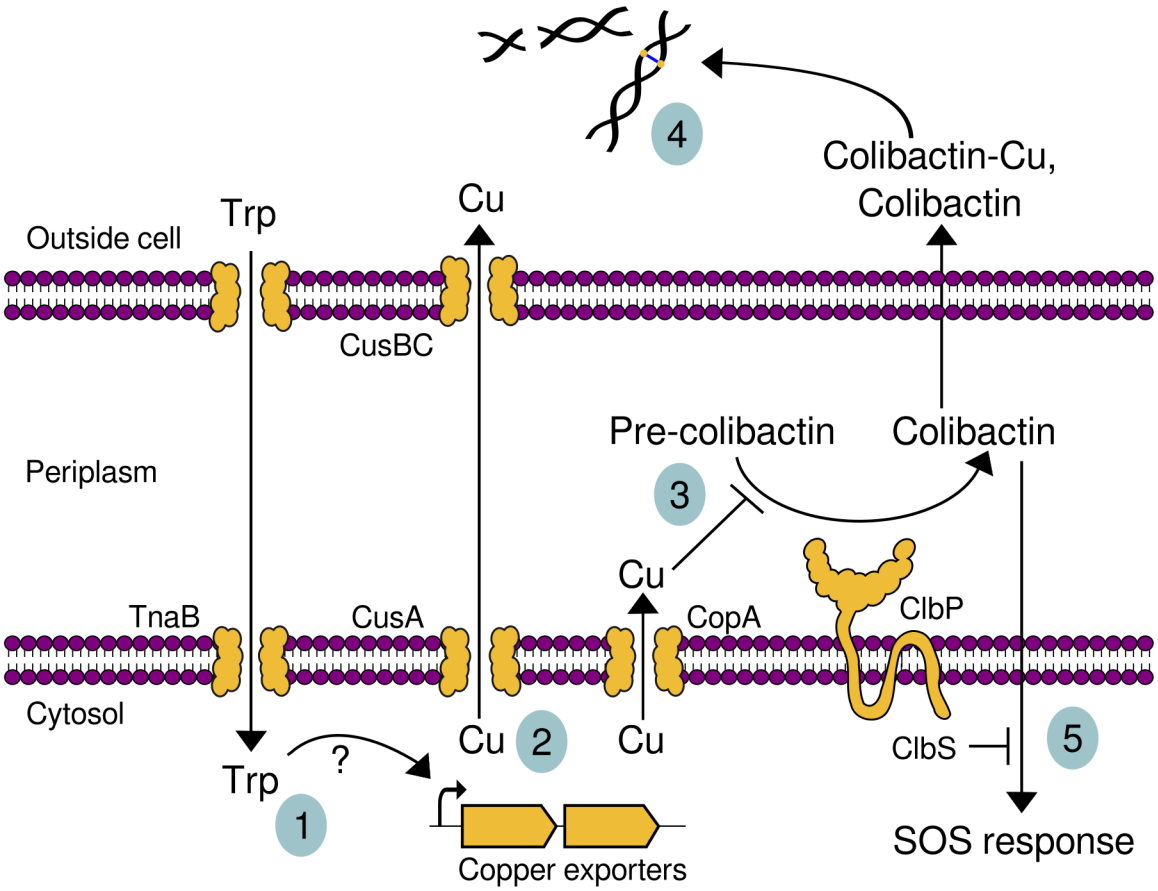


FIG 6 Model for how tryptophan reduces DNA damage by colibactin through copper-mediated repression of ClbP activity. (1) Uptake of tryptophan stimulates transcription of copper exporters by an unknown mechanism. (2) Copper exporters pump copper to the periplasm and outside cells. (3) Periplasmic copper inhibits ClbP activity, reducing the processing of precolibactin to active colibactin. (4) Colibactin and colibactin-copper complexes cause DNA crosslinking and double-strand breaks. (5) ClbS protects the colibactin-producing cell from DNA damage by colibactin, which otherwise activates the SOS response.

Downloaded from https://journals.asm.org/journal/msystems on 11 October 2024 by 2a01:e34:ec23:6090:bc02:eb34:142:90b3.

Elucidating mechanisms regulating colibactin activation and genotoxicity can lead to novel strategies to repress its activity in the gut and ultimately prevent colibactin-mediated carcinogenesis. Pathogenicity of *Enterobacteriaceae* is typically controlled using antibiotics such as aminoglycosides, fluoroquinolones, and B-lactams. However, antibiotics can have negative health consequences by altering the gut microbiome and leading to the emergence of multi-drug resistant strains (42). Studies have identified small molecule inhibitors of the genotoxic activity of colibactin-producing *E. coli* (43, 44). Repression of colibactin activity using L-tryptophan is particularly attractive due to its other beneficial effects such as prevention of *E. coli* biofilm formation (45). In addition, tryptophan metabolites from the gut microbiota stimulate aryl hydrocarbon receptor-mediated IL-22 signaling to reduce mucosal inflammation and protect from *Candida albicans* colonization (46).

Additional research is needed to assess the efficacy of using molecules such as L-tryptophan or copper to repress colibactin production and activity in humans. In this study, we focused on strain ATCC 25922 as a model *pks*+ *E. coli*, but humans harbor various *pks*+ strains that can differ in their metabolic and genetic regulatory features. Moreover, studies are needed to show that metabolite concentrations needed to repress colibactin activity in *pks*+ *E. coli* can be achieved in the gut. Previous studies have found that 4%–6% of tryptophan is metabolized to indoles by the gut microbiota (47) and that fecal indole concentrations in healthy adults were 2.59 mM (48). Intestinal copper concentrations have been estimated to reach 10 µM (49). Thus, the tryptophan and copper concentrations used in this culture-based study may be achievable in the human intestine. Other studies have identified other metabolites such as D-serine, cinnamaldehyde, and medicinal plant extracts that reduce *E. coli clbB* expression and colibactin cytotoxicity (20, 50, 51). Our results complement these studies to advance the development of potent, antibiotic-sparing prophylactics that inhibit colibactin genotoxicity and its tumorigenic effects.

MATERIALS AND METHODS

Microbial growth

Cultures of *E. coli* ATCC 25922 (*pks*+) or BAA 2340 (*pks*–) were grown at 37°C in either MS medium (52) (Table S4), MM29 medium (26), or CMG medium (Anaerobe Systems, CA, USA). Cultures growing in MS medium were supplemented with L-tryptophan (Sigma T0254) or copper chloride dihydrate ($CuCl_2{\cdot}2H_2O$) (Sigma 307483), as appropriate. Growth (optical density at 600 nm [OD_{600}]) and all fluorescence measurements were made using a CLARIOstar microplate reader (BMG Labtech, Ortenberg, Germany).

BCECF method for pH measurements

The pH of cultures was measured using the BCECF method (25, 26). BCECF (C3411 Sigma) at 2 µM was added to cultures and fluorescence at the pH-sensitive point (485 nm excitation; 540 nm emission) relative to the pH-insensitive isosbestic point (450 nm excitation; 540 nm emission) was calculated. Culture pH was computed using a four-parameter log-logistic (4PL) regression that was fit to BCECF fluorescence for each growth medium at experimentally adjusted pH 4–9 (Fig. S5). The 4PL regression was calculated using the R package drc version 3.0-1.

ClbP activity-based probe (ClbP-17)

ClbP-17 (21) was synthesized from N-Boc-D-asparagine using a five-step process yielding the product at 98% purity (WuXi AppTech, Shanghai China) (Fig. S1). Ten mM (100×) stocks of ClbP-17 and 7H4M (Fischer 10377230) were prepared in DMSO and stored at −20°C. ClbP-17 and 7H4M were added to cultures or ClbP-pep at a final concentration of 100 µM. ClbP-17 and 7H4M fluorescence (360 ± 40 nm excitation, 440 ± 20 nm emission) were measured using a CLARIOstar microplate reader (BMG Labtech, Ortenberg,

Downloaded from https://journals.asm.org/journal/msystems on 11 October 2024 by 2a01:e34:ec23:6090:bc02:eb34:142:90b3.

Germany). The ClbP-17:7H4M ratio was calculated for triplicate cultures containing each molecule as the ratio of ClbP-17 fluorescence to the median 7H4M fluorescence.

Liquid chromatography–mass spectrometry

Culture supernatant samples for metabolomics were taken 0, 4, 8, 22, 25, and 29 hours post inoculation by centrifuging cultures at 3000 × *g* for 5 minutes, removing the supernatant, and freezing at −80°C. Metabolite analysis was performed by liquid chromatography (Thermo Dionex Ultimate 3000) coupled with a Thermo Scientific Q Exactive mass spectrometer. Full MS analysis was performed in positive mode using *m/z* range from 50 to 750 *m/z* at a resolution of 35,000 with the following source parameters: spray voltage 4,200 V, capillary temperature 320°C, sheath gas at 40 units, auxiliary gas at 10 units, and S-lens radio frequency level at 50%. Samples were injected onto a Waters Acquity UPLC BEH amide column (2.1 mm × 150 mm, 1.7 µm, 130 Å) with a 60°C column temperature. Mobile phase A was 2.5 mM ammonium formate with 0.1% formic acid in 95:5 acetonitrile:water. Mobile phase B was 2.5 mM ammonium formate with 0.1% formic acid in 5:95 acetonitrile:water. Water, ammonium formate, and formic acid were LC-MS grade (Thermo Fisher, Waltham, MA). Mobile phase B gradient was as follows: 0 minutes: 10%, 2 minutes: 10%, 12 minutes: 60%, 13 minutes: 80%, 15 minutes: 80%. A 2-minute ramp back to 10% and 3-minute equilibration was performed between each sample. The volume of each injection was 4 µL.

To prepare samples for LC-MS, an internal standard (IS) solution containing 2 µM leucine-d10 (Sigma-Aldrich, St Louis, MO) and 2 µM indole-3-acetic acid-d7 (Cambridge Isotope Laboratories, Tewksbury, MA) was prepared in methanol and chilled. An aliquot of each culture sample was combined into a pooled sample, which was treated using the same methods as each culture sample. Each culture sample (30 µL) was mixed with 170 µL of methanol IS solution, vortexed, and centrifuged at 1,800 × *g* for 10 minutes at 4°C. The supernatant (150 µL) was extracted and dried in a speed vacuum for 2 hours without light or heat. Samples were resuspended in 100 µL of 50:50 methanol:water, vortexed, and centrifuged at 1,800 × *g* for 10 minutes at 4°C. An aliquot of each sample (10 µL) was combined into a pooled sample. The remaining sample (80 µL) was extracted and transferred to an LC injection vial.

Full MS data were processed to generate abundance profiles based on MS1-extracted ion chromatograms using XCMS (53). Peak parameters and alignment were determined and processed to achieve good correspondence using the IS solution and multiply injected pooled samples. Results were trimmed against the pooled QC to remove noisy and uninformative features, then Pareto scaled. Metabolite features were clustered based on their abundance profiles across the six-time points in MM29 and CMG media. The 12 abundance values for each metabolite (six-time points, two media) were scaled (per feature) to generate a distance matrix using the dynamic time warping algorithm implemented in the TSClust R package (version 1.3.1) (54). Hierarchical clustering was performed on the distance matrix of metabolite features using the complete agglomeration method as implemented in the hclust function from the stats R package (version 3.1.0). Manually searching the metabolite features identified a target cluster having relative abundance profiles matching a potential colibactin repressor.

The chemical structures of features in the target cluster were investigated by re-injecting the pooled sample and performing LC-MS/MS with parallel reaction monitoring at a resolution of 17,500, AGC target 2e5, and maximum IT 100 ms. Higher energy C-trap dissociation was collected with normalized collision energy (NCE) 30 and NCE 90. Source and LC conditions were the same as for MS1 data. MS1 chromatogram peaks corresponding with the retention time and mass of the target features were identified using Xcalibur version 3.1 (Thermo Fisher). The corresponding MS2 spectrum was then extracted and queried against the mzCloud database (Table S2B).

Downloaded from https://journals.asm.org/journal/msystems on 11 October 2024 by 2a01:e34:ec23:6090:bc02:eb34:142:90b3.

DNA crosslinking assay

The colibactin activity of *E. coli* cultures was tested using a DNA crosslinking assay based on denaturing gel electrophoresis under alkaline conditions to enable migration of single-stranded DNA (29). *E. coli* BAA 2340 (*pks*−) was cultured in MS medium and *E. coli* ATCC 25922 (*pks*+) was cultured in MS medium ± L-tryptophan or copper supplementation, as appropriate. Cultures of 200 µL were grown for 6 hours at 37°C, spiked with 1 µg of *Bam*HI-linearized pUC19 plasmid DNA, and incubated for an additional hour. Cisplatin-treated controls were prepared by incubating 1 µg of *BamH*I linearized pUC19 DNA for 6 hours at 37°C with 80 µM cisplatin in 1× phosphate-buffered saline.

Plasmid DNA from culture supernatants and cisplatin treatments was purified using a Qiagen QIAquick PCR kit (Qiagen, Germany) and resolved by denaturing gel electrophoresis. Briefly, 1% agarose gels were prepared in a 100 mM sodium chloride, 2 mM EDTA (pH 8) solution and soaked in an electrophoresis running buffer (40 mM NaOH, 1 mM EDTA). DNA (200 ng per lane) was resolved by electrophoresis for 45 minutes at 1 V cm^{-1} and then for 2 hours at 2 V cm^{-1}. The gel was soaked in a neutralization solution (150 mM sodium chloride, 100 mM Tris pH 7.4) for 4 hours with the buffer changed after each hour, DNA was stained with Gel Red (Biotium, Fremont, CA), and visualized. Gel band intensities were compared using ImageJ software version 1.54d (55).

RNA sequencing

Cultures of *E. coli* ATCC 25922 were grown in MS medium with and without supplementation with 25 mM L-tryptophan to mid-log phase. RNA was extracted using TRI reagent (Sigma 93289) (56) and 1 µg of total RNA was depleted of rRNA by the NEBNext rRNA depletion kit (New England Biolabs, MA, USA). RNA was purified using NEBNext RNA sample purification beads (New England Biolabs, MA, USA) into 5 µL of water. Quantity and quality of rRNA-depleted RNA were assessed on a Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA) and an Agilent 2100 Bioanalyzer (Agilent Technologies, USA). cDNA libraries were prepared using the NEBNext Ultra II directional RNA library preparation kit (New England Biolabs, MA, USA). RNA fragmentation was performed for 90 seconds at 94°C and incubated for 60 minutes at 42°C for first-strand cDNA synthesis. NEBNext adaptors were diluted five times for adaptor ligation and 16 PCR cycles were applied for enrichment of adaptor-ligated DNA. After profile analysis by a Bioanalyzer and qPCR quantification on MxPro equipment (Stratagene, La Jolla, USA), libraries were sequenced on an Illumina Novaseq 6000 under a PE-50 mode.

Sequencing reads were trimmed of low-quality nucleotides (Q < 20) from both ends of the reads, sequencing adaptors, and primer sequences. Sequences between the second unknown nucleotide and the end of the read were removed and read shorter than 30 nucleotides (after trimming) were discarded. RNA sequencing reads were mapped to the *E. coli* ATCC 25922 genome (NCBI Genbank accession CP009072) using Bowtie 2 (57). Reads per genome feature were counted from BAM files using featureCounts (58) and differentially expressed genes were identified using DEseq2 (59). Gene Ontology terms were used to calculate functional pathway enrichments ($P < 0.01$) using clusterProfiler 4.0 (60).

ClbP purification and activity measurements

The peptidase domain of ClbP (ClbP-pep), as defined in (13), was PCR amplified with a C-terminal His tag using primers clbP_F (5-AAAGAAGGAGATAGGATCATGACAATAATGG AACACGTTAG-3) and clbP_R (5-GTGTAATGGATAGTGATCTTAATGGTGATGGTGATGATGATAT TTGCCAATGCGCAGA-3). The PCR product was cloned by ligation-independent cloning into pET-22B(+) (61, 62) and the forward and reverse sequences were confirmed by sequencing. Plasmids were transformed into *E. coli* BL21(DE3) (Novagen 70235). A culture of *E. coli* BL21(DE3) expressing ClbP-pep was grown in 100 mL TB medium to OD_{600} = 2, induced by adding 500 µM IPTG, and incubated overnight at 20°C. Cells were pelleted and resuspended in lysis buffer (50 mM phosphate buffer pH 8, 0.5 M NaCl, 30 mM imidazole, 15% glycerol, 1 mM pefabloc [Sigma 76307]). Cells were lysed by sonication

Downloaded from https://journals.asm.org/journal/msystems on 11 October 2024 by 2a01:e34:ec23:6090:bc02:eb34:142:90b3.

(Cole-Parmer Vibracell CV33) in the presence of 1 mg mL^{-1} lysozyme (Novagen 71230). ClbP-pep was purified using a Ni-NTA spin column (Qiagen 31014) and quantified by Bradford assay. Purified ClbP-pep was visualized on 12% SDS-PAGE gels (Novex 12% Bis-Tris Gel NP0342BOX) (Fig. S10).

The activity of ClbP-pep was measured by diluting the enzyme to 0.01 µg $µL^{-1}$ in buffer (50 mM Tris, 200 mM sodium chloride, pH 8) containing 100 µM of either ClbP-17 or 7-hydroxy-4-methylcoumarin (Fisher Scientific AC156380250) and copper chloride dihydrate (Sigma 307483), as appropriate. ClbP-pep was then incubated for 120 minutes at 37°C and fluorescence was measured (360 ± 40 nm excitation, 440 ± 20 nm emission) using a CLARIOstar microplate reader (BMG Labtech, Ortenberg, Germany).

ACKNOWLEDGMENTS

We thank A. Perret and M. Giuggio for mass spectrometry expertise and the Genoscope-CEA sequencing platform for RNA sequencing.

C.B. was supported in part by the UCSD Graduate Training Program in Cellular and Molecular Pharmacology through an institutional training grant from the National Institute of General Medical Sciences (T32 GM007752). This work was also supported by the Genoscope-CEA and the Évry Genopole under the "Action Thématique Incitative Genopole" grant (funding ATIGE 2021 No. 2698).

AUTHOR AFFILIATIONS

[1]Department of Pharmacology, University of California, San Diego, California, USA
[2]Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, Evry, France

AUTHOR ORCIDs

Charlie Bayne http://orcid.org/0000-0003-1870-5298
Andrew C. Tolonen http://orcid.org/0000-0001-5907-4504

AUTHOR CONTRIBUTIONS

Charlie Bayne, Conceptualization, Investigation, Methodology, Writing – original draft, Writing – review and editing | Magali Boutard, Investigation, Methodology | Tom Zaplana, Investigation, Methodology | Andrew C. Tolonen, Conceptualization, Funding acquisition, Supervision, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

The RAW files for mass spectrometry analysis were deposited in the MassIVE database (accession MSV000091164). FASTQ files for RNA sequencing were deposited in the European Nucleotide Archive (accession ERP163145).

ADDITIONAL FILES

The following material is available online.

Supplemental Material

Supplemental Figures (mSystems00992-24-s0001.pdf). Figures S1-S10.
Table S1 (mSystems00992-24-s0002.xlsx). Metabolite abundances.
Table S2 (mSystems00992-24-s0003.xlsx). Chromatogram and mass spectrum data.
Table S3 (mSystems00992-24-s0004.xlsx). Differential mRNA expression.
Table S4 (mSystems00992-24-s0005.xlsx). MS medium recipe.

REFERENCES

1. Putze J, Hennequin C, Nougayrède J-P, Zhang W, Homburg S, Karch H, Bringer M-A, Fayolle C, Carniel E, Rabsch W, Oelschlaeger TA, Oswald E, Forestier C, Hacker J, Dobrindt U. 2009. Genetic structure and distribution of the colibactin genomic island among members of the family

Downloaded from https://journals.asm.org/journal/msystems on 11 October 2024 by 2a01:e34:ec23:6090:bc02:eb34:142:90b3.

Enterobacteriaceae. Infect Immun 77:4696–4703. https://doi.org/10.1128/IAI.00522-09
2. Nougayrède J-P, Homburg S, Taieb F, Boury M, Brzuszkiewicz E, Gottschalk G, Buchrieser C, Hacker J, Dobrindt U, Oswald E. 2006. *Escherichia coli* induces DNA double-strand breaks in eukaryotic cells. Science 313:848–851. https://doi.org/10.1126/science.1127059
3. Cuevas-Ramos G, Petit CR, Marcq I, Boury M, Oswald E, Nougayrède J-P. 2010. *Escherichia coli* induces DNA damage *in vivo* and triggers genomic instability in mammalian cells. Proc Natl Acad Sci U S A 107:11537–11542. https://doi.org/10.1073/pnas.1001261107
4. Nowrouzian FL, Oswald E. 2012. *Escherichia coli* strains with the capacity for long-term persistence in the bowel microbiota carry the potentially genotoxic pks island. Microb Pathog 53:180–182. https://doi.org/10.1016/j.micpath.2012.05.011
5. Marcq I, Martin P, Payros D, Cuevas-Ramos G, Boury M, Watrin C, Nougayrède J-P, Olier M, Oswald E. 2014. The genotoxin colibactin exacerbates lymphopenia and decreases survival rate in mice infected with septicemic *Escherichia coli*. J Infect Dis 210:285–294. https://doi.org/10.1093/infdis/jiu071
6. McCarthy AJ, Martin P, Cloup E, Stabler RA, Oswald E, Taylor PW. 2015. The genotoxin colibactin is a determinant of virulence in *Escherichia coli* K1 experimental neonatal systemic infection. Infect Immun 83:3704–3711. https://doi.org/10.1128/IAI.00716-15
7. Smith JL, Fratamico PM, Gunther NW. 2007. Extraintestinal pathogenic *Escherichia coli*. Foodborne Pathog Dis 4:134–163. https://doi.org/10.1089/fpd.2007.0087
8. Arthur JC, Perez-Chanona E, Mühlbauer M, Tomkovich S, Uronis JM, Fan T-J, Campbell BJ, Abujamel T, Dogan B, Rogers AB, Rhodes JM, Stintzi A, Simpson KW, Hansen JJ, Keku TO, Fodor AA, Jobin C. 2012. Intestinal inflammation targets cancer-inducing activity of the microbiota. Science 338:120–123. https://doi.org/10.1126/science.1224820
9. Buc E, Dubois D, Sauvanet P, Raisch J, Delmas J, Darfeuille-Michaud A, Pezet D, Bonnet R. 2013. High prevalence of mucosa-associated *E. coli* producing cyclomodulin and genotoxin in colon cancer. PLoS One 8:e56964. https://doi.org/10.1371/journal.pone.0056964
10. Dziubańska-Kusibab PJ, Berger H, Battistini F, Bouwman BAM, Iftekhar A, Katainen R, Cajuso T, Crosetto N, Orozco M, Aaltonen LA, Meyer TF. 2020. Colibactin DNA-damage signature indicates mutational impact in colorectal cancer. Nat Med 26:1063–1069. https://doi.org/10.1038/s41591-020-0908-2
11. Pleguezuelos-Manzano C, Puschhof J, Rosendahl Huber A, van Hoeck A, Wood HM, Nomburg J, Gurjao C, Manders F, Dalmasso G, Stege PB, et al. 2020. Mutational signature in colorectal cancer caused by genotoxic *pks*$^{+}$ *E. coli*. Nature 580:269–273. https://doi.org/10.1038/s41586-020-2080-8
12. Taieb F, Petit C, Nougayrède J-P, Oswald E. 2016. The enterobacterial genotoxins: cytolethal distending toxin and colibactin. EcoSal Plus 7. https://doi.org/10.1128/ecosalplus.ESP-0008-2016
13. Dubois D, Baron O, Cougnoux A, Delmas J, Pradel N, Boury M, Bouchon B, Bringer M-A, Nougayrède J-P, Oswald E, Bonnet R. 2011. ClbP is a prototype of a peptidase subgroup involved in biosynthesis of nonribosomal peptides. J Biol Chem 286:35562–35570. https://doi.org/10.1074/jbc.M111.221960
14. Brotherton CA, Balskus EP. 2013. A prodrug resistance mechanism is involved in colibactin biosynthesis and cytotoxicity. J Am Chem Soc 135:3359–3362. https://doi.org/10.1021/ja312154m
15. Xue M, Kim CS, Healy AR, Wernke KM, Wang Z, Frischling MC, Shine EE, Wang W, Herzon SB, Crawford JM. 2019. Structure elucidation of colibactin and its DNA cross-links. Science 365:eaax2685. https://doi.org/10.1126/science.aax2685
16. Jiang Y, Stornetta A, Villalta PW, Wilson MR, Boudreau PD, Zha L, Balbo S, Balskus EP. 2019. Reactivity of an unusual amidase may explain colibactin's DNA cross-linking activity. J Am Chem Soc 141:11489–11496. https://doi.org/10.1021/jacs.9b02453
17. Tronnet S, Garcie C, Rehm N, Dobrindt U, Oswald E, Martin P. 2016. Iron homeostasis regulates the genotoxicity of *Escherichia coli* that produces colibactin. Infect Immun 84:3358–3368. https://doi.org/10.1128/IAI.00659-16
18. Chagneau CV, Garcie C, Bossuet-Greif N, Tronnet S, Brachmann AO, Piel J, Nougayrède J-P, Martin P, Oswald E. 2019. The polyamine spermidine modulates the production of the bacterial genotoxin colibactin. mSphere 4:e00414-19. https://doi.org/10.1128/mSphere.00414-19
19. Oliero M, Calvé A, Fragoso G, Cuisiniere T, Hajjar R, Dobrindt U, Santos MM. 2021. Oligosaccharides increase the genotoxic effect of colibactin produced by *pks+ Escherichia coli* strains. BMC Cancer 21:172. https://doi.org/10.1186/s12885-021-07876-8
20. Kaewkod T, Tobe R, Tragoolpua Y, Mihara H. 2021. Medicinal plant extracts protect epithelial cells from infection and DNA damage caused by colibactin-producing *Escherichia coli*, and inhibit the growth of bacteria. J Appl Microbiol 130:769–785. https://doi.org/10.1111/jam.14817
21. Volpe MR, Wilson MR, Brotherton CA, Winter ES, Johnson SE, Balskus EP. 2019. *In vitro* characterization of the colibactin-activating peptidase ClbP enables development of a fluorogenic activity probe. ACS Chem Biol 14:1097–1101. https://doi.org/10.1021/acschembio.9b00069
22. Hirayama Y, Tsunematsu Y, Yoshikawa Y, Tamafune R, Matsuzaki N, Iwashita Y, Ohnishi I, Tanioka F, Sato M, Miyoshi N, Mutoh M, Ishikawa H, Sugimura H, Wakabayashi K, Watanabe K. 2019. Activity-based probe for screening of high-colibactin producers from clinical samples. Org Lett 21:4490–4494. https://doi.org/10.1021/acs.orglett.9b01345
23. Homburg S, Oswald E, Hacker J, Dobrindt U. 2007. Expression analysis of the colibactin gene cluster coding for a novel polyketide in *Escherichia coli*. FEMS Microbiol Lett 275:255–262. https://doi.org/10.1111/j.1574-6968.2007.00889.x
24. Wallenstein A, Rehm N, Brinkmann M, Selle M, Bossuet-Greif N, Sauer D, Bunk B, Spröer C, Wami HT, Homburg S, von Bünau R, König S, Nougayrède J-P, Overmann J, Oswald E, Müller R, Dobrindt U. 2020. ClbR is the key transcriptional activator of colibactin gene expression in *Escherichia coli*. mSphere 5:e00591-20. https://doi.org/10.1128/mSphere.00591-20
25. Heux S, Philippe B, Portais J-C. 2011. High-throughput workflow for monitoring and mining bioprocess data and its application to inferring the physiological response of *Escherichia coli* to perturbations. Appl Environ Microbiol 77:7040–7049. https://doi.org/10.1128/AEM.05838-11
26. Tolonen AC, Beauchemin N, Bayne C, Li L, Tan J, Lee J, Meehan BM, Meisner J, Millet Y, LeBlanc G, Kottler R, Rapp E, Murphy C, Turnbaugh PJ, von Maltzahn G, Liu CM, van Hylckama Vlieg JET. 2022. Synthetic glycans control gut microbiome structure and mitigate colitis in mice. Nat Commun 13:1244. https://doi.org/10.1038/s41467-022-28856-x
27. Wong JJ, Ho FK, Choo PY, Chong KKL, Ho CMB, Neelakandan R, Keogh D, Barkham T, Chen J, Liu CF, Kline KA. 2022. *Escherichia coli* BarA-UvrY regulates the *pks* island and kills Staphylococci via the genotoxin colibactin during interspecies competition. PLoS Pathog 18:e1010766. https://doi.org/10.1371/journal.ppat.1010766
28. Rehm N, Wallenstein A, Keizers M, Homburg S, Magistro G, Chagneau CV, Klimek H, Revelles O, Helloin E, Putze J, Nougayrède J-P, Schubert S, Oswald E, Dobrindt U. 2021. Two polyketides intertwined in complex regulation: posttranscriptional CsrA-mediated control of colibactin and yersiniabactin synthesis in *Escherichia coli*. mBio 13:e0381421. https://doi.org/10.1128/mbio.03814-21
29. Bossuet-Greif N, Vignard J, Taieb F, Mirey G, Dubois D, Petit C, Oswald E, Nougayrède J-P. 2018. The colibactin genotoxin generates DNA interstrand cross-links in infected cells. mBio 9:e02393-17. https://doi.org/10.1128/mBio.02393-17
30. Osman D, Cavet JS. 2008. Copper homeostasis in bacteria. Adv Appl Microbiol 65:217–247. https://doi.org/10.1016/S0065-2164(08)00608-4
31. Franke S, Grass G, Rensing C, Nies DH. 2003. Molecular analysis of the copper-transporting efflux system CusCFBA of *Escherichia coli*. J Bacteriol 185:3804–3812. https://doi.org/10.1128/JB.185.13.3804-3812.2003
32. Rensing C, Fan B, Sharma R, Mitra B, Rosen BP. 2000. CopA: an *Escherichia coli* Cu(I)-translocating P-type ATPase. Proc Natl Acad Sci U S A 97:652–656. https://doi.org/10.1073/pnas.97.2.652
33. Grass G, Rensing C. 2001. CueO is a multi-copper oxidase that confers copper tolerance in *Escherichia coli*. Biochem Biophys Res Commun 286:902–908. https://doi.org/10.1006/bbrc.2001.5474
34. Li Z-R, Li J, Cai W, Lai JYH, McKinnie SMK, Zhang W-P, Moore BS, Zhang W, Qian P-Y. 2019. Macrocyclic colibactin induces DNA double-strand breaks via copper-mediated oxidative cleavage. Nat Chem 11:880–889. https://doi.org/10.1038/s41557-019-0317-7
35. Bernegger S, Brunner C, Vizovišek M, Fonovic M, Cuciniello G, Giordano F, Stanojlovic V, Jarzab M, Simister P, Feller SM, Obermeyer G, Posselt G,

Downloaded from https://journals.asm.org/journal/msystems on 11 October 2024 by 2a01:e34:ec23:6090:bc02:eb34:142:90b3.

Turk B, Cabrele C, Schneider G, Wessler S. 2020. A novel FRET peptide assay reveals efficient *Helicobacter pylori* HtrA inhibition through zinc and copper binding. Sci Rep 10:10563. https://doi.org/10.1038/s41598-020-67578-2
36. Santoro AM, Monaco I, Attanasio F, Lanza V, Pappalardo G, Tomasello MF, Cunsolo A, Rizzarelli E, De Luigi A, Salmona M, Milardi D. 2016. Copper(II) ions affect the gating dynamics of the 20S proteasome: a molecular and in cell study. Sci Rep 6:33444. https://doi.org/10.1038/srep33444
37. Karlström AR, Levine RL. 1991. Copper inhibits the protease from human immunodeficiency virus 1 by both cysteine-dependent and cysteine-independent mechanisms. Proc Natl Acad Sci U S A 88:5552–5556. https://doi.org/10.1073/pnas.88.13.5552
38. Tripathi P, Shine EE, Healy AR, Kim CS, Herzon SB, Bruner SD, Crawford JM. 2017. ClbS is a cyclopropane hydrolase that confers colibactin resistance. J Am Chem Soc 139:17719–17722. https://doi.org/10.1021/jacs.7b09971
39. Bossuet-Greif N, Dubois D, Petit C, Tronnet S, Martin P, Bonnet R, Oswald E, Nougayrède J-P. 2016. *Escherichia coli* ClbS is a colibactin resistance protein. Mol Microbiol 99:897–908. https://doi.org/10.1111/mmi.13272
40. Fung DKC, Lau WY, Chan WT, Yan A. 2013. Copper efflux is induced during anaerobic amino acid limitation in *Escherichia coli* to protect iron-sulfur cluster enzymes and biogenesis. J Bacteriol 195:4556–4568. https://doi.org/10.1128/JB.00543-13
41. Stanila A, Marcu A, Rusu D, Rusu M, David L. 2007. Spectroscopic studies of some copper(II) complexes with amino acids. J Mol Struct 834–836:364–368. https://doi.org/10.1016/j.molstruc.2006.11.048
42. Blaser M. 2011. Antibiotic overuse: stop the killing of beneficial bacteria. Nat New Biol 476:393–394. https://doi.org/10.1038/476393a
43. Cougnoux A, Delmas J, Gibold L, Faïs T, Romagnoli C, Robin F, Cuevas-Ramos G, Oswald E, Darfeuille-Michaud A, Prati F, Dalmasso G, Bonnet R. 2016. Small-molecule inhibitors prevent the genotoxic and protumoural effects induced by colibactin-producing bacteria. Gut 65:278–285. https://doi.org/10.1136/gutjnl-2014-307241
44. Volpe MR, Velilla JA, Daniel-Ivad M, Yao JJ, Stornetta A, Villalta PW, Huang H-C, Bachovchin DA, Balbo S, Gaudet R, Balskus EP. 2023. A small molecule inhibitor prevents gut bacterial genotoxin production. Nat Chem Biol 19:159–167. https://doi.org/10.1038/s41589-022-01147-8
45. Shimazaki J, Furukawa S, Ogihara H, Morinaga Y. 2012. L-Tryptophan prevents *Escherichia coli* biofilm formation and triggers biofilm degradation. Biochem Biophys Res Commun 419:715–718. https://doi.org/10.1016/j.bbrc.2012.02.085
46. Zelante T, Iannitti RG, Cunha C, De Luca A, Giovannini G, Pieraccini G, Zecchi R, D'Angelo C, Massi-Benedetti C, Fallarino F, Carvalho A, Puccetti P, Romani L. 2013. Tryptophan catabolites from microbiota engage aryl hydrocarbon receptor and balance mucosal reactivity via interleukin-22. Immunity 39:372–385. https://doi.org/10.1016/j.immuni.2013.08.003
47. Gao J, Xu K, Liu H, Liu G, Bai M, Peng C, Li T, Yin Y. 2018. Impact of the gut microbiota on intestinal immunity mediated by tryptophan metabolism. Front Cell Infect Microbiol 8:13. https://doi.org/10.3389/fcimb.2018.00013
48. Roager HM, Licht TR. 2018. Microbial tryptophan catabolites in health and disease. Nat Commun 9:3294. https://doi.org/10.1038/s41467-018-05470-4
49. Rensing C, Grass G. 2003. *Escherichia coli* mechanisms of copper homeostasis in a changing environment. FEMS Microbiol Rev 27:197–213. https://doi.org/10.1016/S0168-6445(03)00049-4
50. Hallam JC, Sandalli S, Floria I, Turner NCA, Tang-Fichaux M, Oswald E, O'Boyle N, Roe AJ. 2023. D-Serine reduces the expression of the cytopathic genotoxin colibactin. Microb Cell 10:63–77. https://doi.org/10.15698/mic2023.03.793
51. Kosari F, Taheri M, Moradi A, Hakimi Alni R, Alikhani MY. 2020. Evaluation of cinnamon extract effects on *clbB* gene expression and biofilm formation in *Escherichia coli* strains isolated from colon cancer patients. BMC Cancer 20:267. https://doi.org/10.1186/s12885-020-06736-1
52. Richaud C, Mengin-Lecreulx D, Pochet S, Johnson EJ, Cohen GN, Marlière P. 1993. Directed evolution of biosynthetic pathways. Recruitment of cysteine thioethers for constructing the cell wall of *Escherichia coli*. J Biol Chem 268:26827–26835. https://doi.org/10.1016/S0021--9258(19)74186-6
53. Tautenhahn R, Patti GJ, Rinehart D, Siuzdak G. 2012. XCMS Online: a web-based platform to process untargeted metabolomic data. Anal Chem 84:5035–5039. https://doi.org/10.1021/ac300698c
54. Iscar AM, Escudero LAG, FritzH. 2022. tclust: robust trimmed clustering. https://CRAN.R-project.org/package=tclust.
55. Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9:671–675. https://doi.org/10.1038/nmeth.2089
56. Cerisy T, Souterre T, Torres-Romero I, Boutard M, Dubois I, Patrouix J, Labadie K, Berrabah W, Salanoubat M, Doring V, Tolonen AC. 2017. Evolution of a biomass-fermenting bacterium to resist lignin phenolics. Appl Environ Microbiol 83:e00289-17. https://doi.org/10.1128/AEM.00289-17
57. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923
58. Liao Y, Smyth GK, Shi W. 2019. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic Acids Res 47:e47. https://doi.org/10.1093/nar/gkz114
59. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. https://doi.org/10.1186/s13059-014-0550-8
60. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, Fu X, Liu S, Bo X, Yu G. 2021. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. Innov (Camb) 2:100141. https://doi.org/10.1016/j.xinn.2021.100141
61. Boutard M, Cerisy T, Nogue P-Y, Alberti A, Weissenbach J, Salanoubat M, Tolonen AC. 2014. Functional diversity of carbohydrate-active enzymes enabling a bacterium to ferment plant biomass. PLoS Genet 10:e1004773. https://doi.org/10.1371/journal.pgen.1004773
62. Cerisy T, Iglesias A, Rostain W, Boutard M, Pelle C, Perret A, Salanoubat M, Fierobe H-P, Tolonen AC. 2019. ABC transporters required for hexose uptake by *Clostridium phytofermentans*. J Bacteriol 201:e00241-19. https://doi.org/10.1128/JB.00241-19

Downloaded from https://journals.asm.org/journal/msystems on 11 October 2024 by 2a01:e34:ec23:6090:bc02:eb34:142:90b3.

GENOPOLE

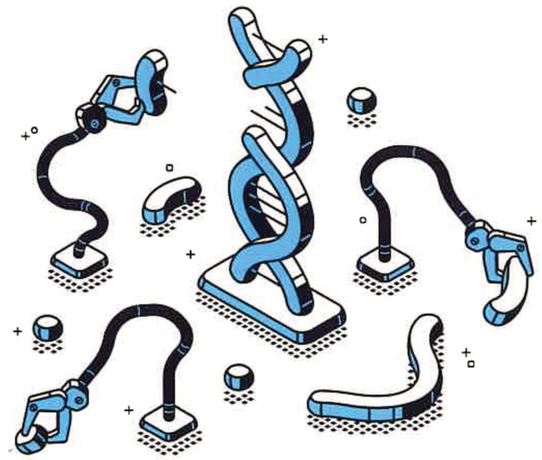
Genopole est le premier biocluster français créé sous l'impulsion des pouvoirs publics en 1998, autour de l'AFM-Téléthon à Évry-Courcouronnes (Essonne). Il est dédié à la recherche en génomique et aux biotechnologies appliquées à la santé, l'agro-alimentaire et l'environnement.
Son objectif : favoriser l'émergence et la croissance de sociétés de biotechnologie, le transfert d'innovations vers le secteur industriel, le développement de la recherche et l'enseignement supérieur en sciences et technologies du vivant.

Genopole rassemble des laboratoires de recherche académique, des sociétés de biotechnologie, des plateformes technologiques dans un bioparc associant l'université d'Évry Paris-Saclay et le Centre hospitalier sud francilien. Il est moteur d'innovations, en particulier dans les domaines stratégiques des biothérapies et de la bioproduction en santé, et de la bioéconomie et décarbonation des activités industrielles.
Genopole est principalement soutenu par l'État, la Région Ile-de-France, le Département de l'Essonne, l'agglomération Grand Paris Sud, la Ville d'Évry-Courcouronnes et l'AFM-Téléthon.
www.genopole.fr

GENOSCOPE - UMR Génomique métabolique

Installé sur le campus 2 de Genopole, le Genoscope, centre national de séquençage, fait partie de l'Institut de biologie François Jacob du CEA. Il contribue à l'exploration des génomes et des métagénomes de la biodiversité et fournit à la communauté scientifique son expertise de pointe, ses capacités de séquençage et d'analyse de données. Le Genoscope concentre son activité de recherche au sein de l'UMR 8030 Génomique Métabolique (CEA / CNRS / Université d'Évry Paris-Saclay), selon une approche interdisciplinaire associant génomique, ingénierie métabolique, biochimie, chimie analytique, bio-informatique.

Après avoir été l'un des acteurs du projet génome humain au début des années 2000, le Genoscope a mis le cap sur la génomique environnementale. L'exploitation des données de séquençage et l'identification des fonctions biologiques qui en résulte ont ouvert de nouveaux horizons de développement dans les biotechnologies industrielles, notamment en biocatalyse et en biologie de synthèse. Dans une perspective de développement durable, l'UMR Génomique métabolique recherche des solutions biologiques pour contribuer à une chimie de synthèse moins polluante et moins consommatrice d'énergie et de carbone fossile.
www.cea.fr/drf/ifrancoisjacob/Pages/Departements/Genoscope.aspx
https://www.cea.fr/drf/ifrancoisjacob/Pages/Departements/Genoscope/Les-laboratoires/UMR-8030-Genomique-Metabolique-du-Genoscope.aspx



Biologie de synthèse
De l'ingénierie cellulaire pour la bioproduction


D.R.

TOM ZAPLANA
Doctorant au Genoscope - UMR Génomique métabolique (CEA, CNRS, université d'Évry Paris-Saclay)


X. RENAULD

VÉRONIQUE LE BOULC'H
Responsable communication scientifique du Genopole


X. RENAULD

JULIEN PICOT
Directeur adjoint délégué aux infrastructures et plateformes du Genopole


D.R.

ANDREW TOLONEN
Chercheur CEA au Genoscope - UMR Génomique métabolique

DÉTOURNER LA VOIE MÉTABOLIQUE D'UNE BACTÉRIE POUR QU'ELLE PRODUISE UN COMPOSÉ D'INTÉRÊT, VOICI L'UNE DES PROUESSES RÉALISABLES AUJOURD'HUI PAR LA BIOLOGIE DE SYNTHÈSE GRÂCE À UNE COMBINAISON D'OUTILS ET DE MÉTHODES À HAUTE PRODUCTIVITÉ.

La biologie de synthèse, ou biologie synthétique, peut se définir comme les principes de l'ingénierie appliqués à la modification génétique d'organismes. C'est une discipline récente, à la frontière de la biologie, de la chimie, de l'informatique et des sciences de l'ingénieur. Elle vise à concevoir des systèmes biologiques complexes dotés de fonctions parfois totalement absentes de la nature. Cela passe par la construction de nouveaux composants biologiques, notamment des segments d'ADN fonctionnels. L'objectif, d'un point de vue fondamental, est de mieux comprendre les mécanismes du vivant et, en termes d'application, de créer de nouvelles fonctions utiles dans les domaines de la biotechnologie, de la médecine, de l'alimentation ou de l'agriculture. Aujourd'hui, les biologistes cherchent à intégrer l'automatisation, la robotique et les techniques à haut débit dans les laboratoires afin d'atteindre la vitesse et l'efficacité nécessaires à des applications à grande échelle. Ces laboratoires robotisés sont appelés biofonderies.
En combinant la puissance de design et de décision de la bio-informatique actuelle et des flux de travail expérimentaux optimisés et rationalisés,

CE QU'IL FAUT RETENIR

Grâce aux progrès de la bio-informatique et à l'essor des technologies d'ingénierie génétique, la biologie de synthèse, qui vise à concevoir et fabriquer des systèmes biologiques artificiels dotés de fonctions nouvelles, s'attache à mettre au point des procédés de bioproduction de composés d'intérêt. Les chercheurs s'appuient sur des biofonderies, des plateformes technologiques complètement automatisées, pour accélérer leurs travaux et faire passer les procédés à l'échelle industrielle.

les chercheurs accélèrent les découvertes scientifiques et favorisent les innovations biotechnologiques. Les développements s'orientent par exemple vers des biocapteurs microbiens pour la détection de molécules cibles ou toxines, vers des tests diagnostiques innovants ou encore des bactéries capables de dépolluer, mais surtout vers de nouveaux procédés industriels de production de composés d'intérêt (carburants, matériaux, médicaments, arômes, protéines alternatives...), par voie biologique, en utilisant des ressources renouvelables comme matières premières.

1 Les bases technologiques de la biologie de synthèse

L'universalité du code génétique démontrée dans les années 1960 à 1970 permettait d'imaginer la possibilité de transférer des voies métaboliques d'un organisme à un autre en lui apportant les gènes codant la synthèse d'enzymes qui catalysent les réactions successives. En effet, les unités génétiques que sont les gènes, construites sur un modèle commun, peuvent être lues par les machineries cellulaires d'organismes différents. L'information est traduite par la production de protéines aux rôles divers (composants de structure, communication cellulaire, transport...), dont les enzymes, catalyseurs naturels de toutes les réactions chimiques nécessaires au fonctionnement du vivant et molécules clés de la biologie de synthèse.

Il faut attendre les années 2000 pour voir ces concepts se concrétiser et la biologie de synthèse émerger. Elle représente à ce moment-là un bouleversement scientifique et technologique. Les progrès en fiabilité et en rapidité réalisés dans le séquençage et la synthèse d'ADN et la baisse du coût qui les accompagne constituent le socle technologique de la biologie de synthèse. En 2002, le séquençage d'ADN coûtait 4 millions de dollars par million de **bases nucléiques**. Vingt ans plus tard, ce coût est tombé à environ 0,006 dollar par million de bases. Identifier les gènes portant des fonctions utiles, synthétiser les séquences d'ADN et les assembler pour construire une fonction nouvelle est devenu possible.

Les progrès de la bio-informatique posent le deuxième pilier de la biologie de synthèse. Explorer la biodiversité bactérienne, largement inconnue encore aujourd'hui, afin de trouver de nouvelles fonctions enzymatiques nécessite l'analyse à grande échelle des génomes, impossible sans des méthodes computationnelles et algorithmiques dédiées à la génomique. De plus, la conception de voies de synthèse s'appuie sur la modélisation et la simulation informatiques et sur le deep learning.

Le développement continu des technologies d'ingénierie génétique, troisième assise de la biologie de synthèse, élargit son champ d'application aux bactéries, levures, algues et plantes. Par exemple, des scientifiques ont réussi à apporter aux chloroplastes (organites dotés d'un génome propre, réalisant la photosynthèse dans les cellules végétales) de lentilles d'eau un système génétique autonome sous forme d'un minichromosome portant une vingtaine de gènes. Ce dernier se réplique et génère un niveau de production d'un composé d'intérêt cinq à dix fois plus élevé qu'avec une insertion des gènes dans le génome. Les lentilles d'eau deviennent ainsi des usines biologiques utilisant comme seules ressources la lumière solaire et l'eau, tout en consommant du CO_2.

Les bases nucléiques sont les composants organiques azotés entrant dans la composition de l'ADN et de l'ARN. L'ADN en comporte quatre sortes : A, T, G et C.

2 Optimiser les systèmes biologiques conçus avec le cycle Design-Build-Test-Learn (DBTL)

Dans le cadre de la conception de nouveaux systèmes biologiques, il est indispensable d'appliquer un cadre d'ingénierie standardisé au génie génétique pour apporter rapidité et reproductibilité au processus. Un élément clé de la démarche est la mise en œuvre d'une procédure itérative dite DBTL, pour Design-Build-Test-Learn (fig. 1). Essentiel en biologie synthétique, le cycle DBTL englobe les étapes de conception des systèmes biologiques, de construction, d'analyse et d'apprentissage. Chaque itération s'appuie sur les connaissances acquises lors de la précédente.

La boucle d'opérations DBTL conduit à une accélération notable de la recherche et de la découverte biotechnologiques. Elle permet d'optimiser les systèmes conçus, et de mieux en comprendre les principes pour développer des applications de biologie synthétique plus sophistiquées et plus efficaces. Elle comporte quatre étapes :

- **Design.** La première phase consiste à conceptualiser le système biologique souhaité et le protocole expérimental. Les chercheurs et les ingénieurs utilisent des outils bio-informatiques et des connaissances préalables pour concevoir des séquences d'ADN, des circuits génétiques (association de séquences d'ADN codant une fonction – les gènes – et de séquences promotrices et régulatrices nécessaires à l'expression de ces gènes) ou des voies biologiques complètes qui coderont les fonctions souhaitées. Les implications éthiques et les mesures de sécurité sont prises en compte pendant la phase de conception.
- **Build.** Les chercheurs procèdent ensuite à la construction physique des séquences d'ADN. Cela implique souvent des techniques telles que la synthèse de l'ADN, le clonage génétique pour multiplier les séquences conçues et l'édition du génome pour introduire l'ADN modifié dans des organismes hôtes, dits « châssis », tels que les bactéries, les levures, les algues...
- **Test.** Une fois le système biologique construit et intégré dans l'organisme hôte, celui-ci est soumis à des tests expérimentaux. Des quantifications par chimie analytique ou bien l'expression de protéines fluorescentes révèlent les rendements de production des molécules d'intérêt. Les données collectées sont utilisées pour évaluer les performances du système synthétique au sein de l'organisme.
- **Learn.** Au cours de la phase d'apprentissage, les données recueillies lors des expériences sont analysées et comparées aux résultats escomptés. Les chercheurs identifient les succès, les limites et les axes d'amélioration. Ces connaissances sont utilisées pour affiner le design et apporter les ajustements nécessaires au système biologique pour l'itération suivante.

3 L'automatisation des procédés : vers la création de biofonderies

Pour passer d'une biologie de synthèse expérimentale au développement de procédés industrialisables, il faut pouvoir accélérer la réalisation des cycles DBTL. C'est le rôle des biofonderies. Ces plateformes technologiques de pointe appliquent des principes d'ingénierie et d'automatisation à la biologie synthétique. Infrastructures centralisées pour la conception, la construction et le prototypage à haut débit de systèmes biologiques,

P. 55
Le processus itératif DBTL

FIG. 1 Le processus itératif DBTL

La construction génétique de la voie métabolique est d'abord conceptualisée par ordinateur, avant que les séquences d'ADN la composant ne soient construites physiquement et intégrées dans la bactérie. Après la sélection des bactéries ayant bien intégré la construction, l'ensemble est testé par séquençage et analyse chimique pour évaluer ses performances. Les données sont enfin analysées, puis on utilise le machine learning pour optimiser la conception de la voie métabolique.

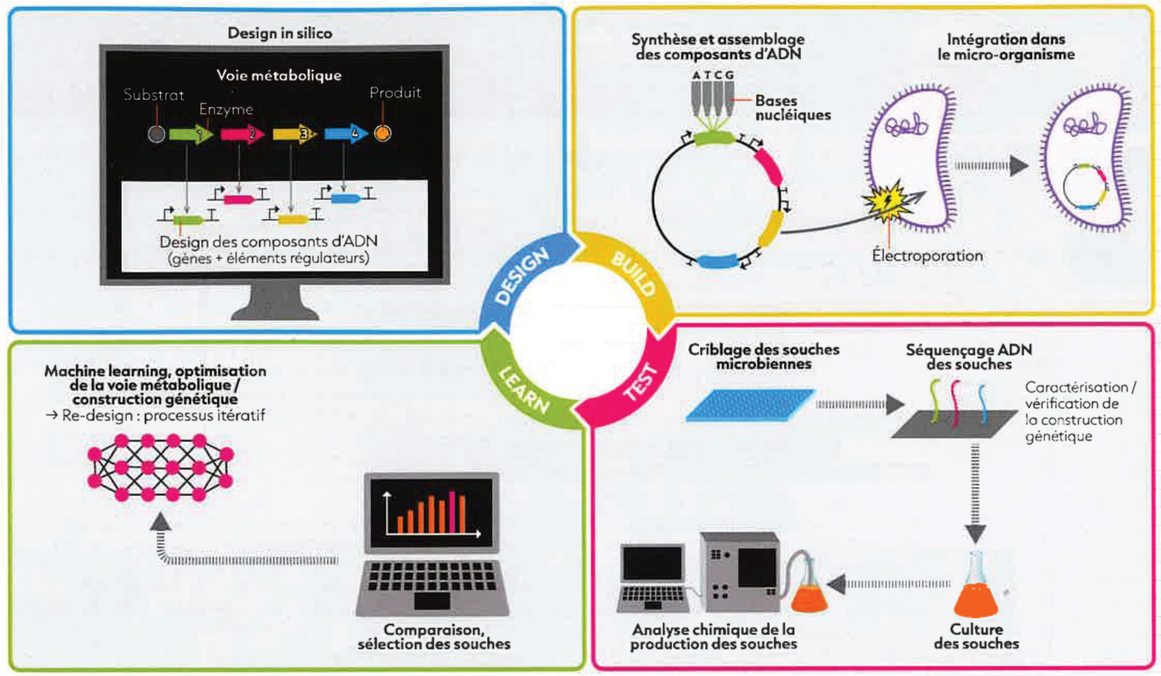


FIG. 2 Les enjeux de la montée en échelle

La montée en échelle de la production d'un composant d'intérêt doit faire face à plusieurs enjeux : tout d'abord limiter la fluctuation en grand volume des paramètres environnementaux (comme la température, le pH ou encore l'oxygène dissout), mais aussi tester la robustesse de la souche face à ces stress environnementaux, optimiser le métabolisme microbien, et donc le rendement, et enfin standardiser la production.

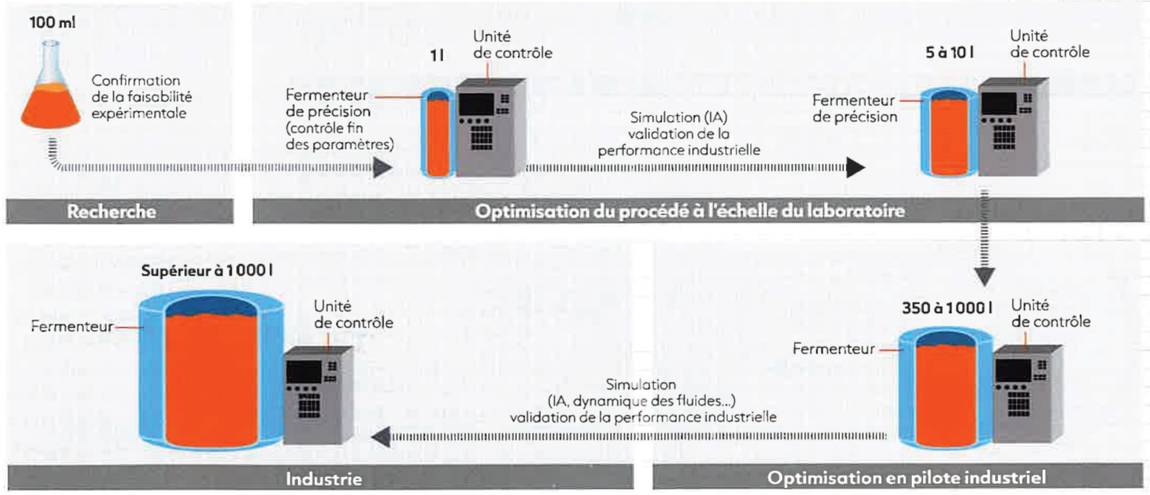


FIG. 3 Le procédé de récupération et purification des bioproduits (downstream process)

Quel que soit l'endroit où se trouve le produit d'intérêt après sa production par les cellules, plusieurs étapes sont nécessaires pour le récupérer. Le mélange est d'abord clarifié, puis, si nécessaire, le produit est extrait des cellules et enfin purifié avec plusieurs niveaux de filtration selon la taille de la molécule d'intérêt, avant d'être stabilisé sous différentes formes.

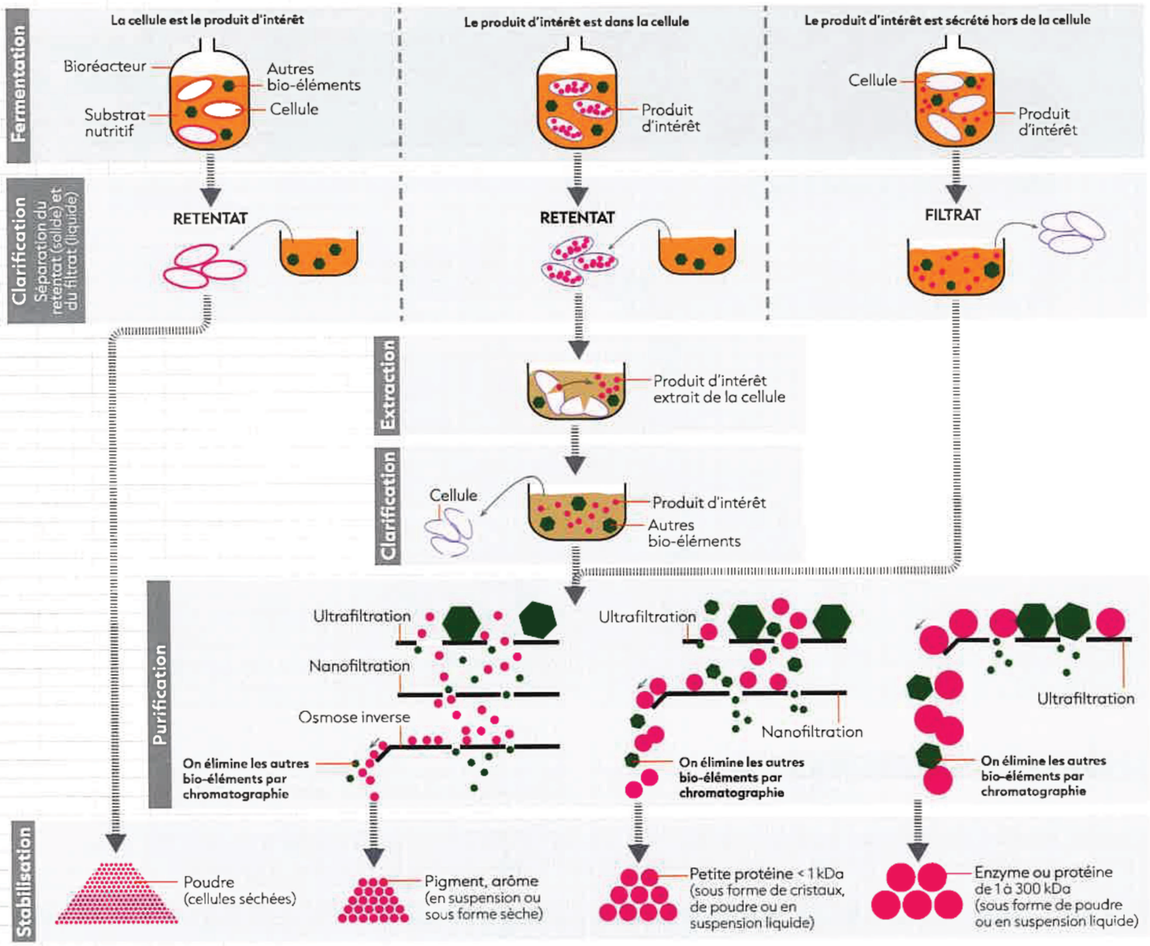
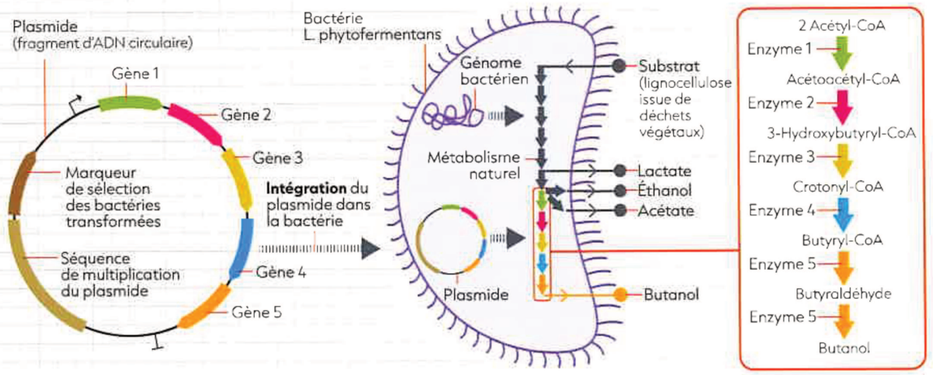


FIG. 4 Insertion de la voie de synthèse du butanol dans la bactérie L. phytofermentans



Une fois la voie métabolique construite génétiquement et insérée dans un plasmide circulaire, elle est intégrée dans la bactérie. La machinerie cellulaire de la bactérie lit alors ce système génétique comme sa propre information génétique : son métabolisme naturel est dévié au profit de la synthèse du butanol.

ces dernières représentent une avancée technologique majeure. Elles offrent une multitude d'avantages, comme l'expérimentation à grande échelle, la normalisation des procédures et la reproductibilité des résultats, ainsi qu'un prototypage rapide des constructions biologiques. Dans la pratique, cette approche a des retombées concrètes.

Ainsi, les biofonderies optimisent les systèmes de production biologique conçus par biologie de synthèse. Cela inclut l'amélioration du taux de production des protéines ciblées, le perfectionnement de voies métaboliques pour la création de molécules à haute valeur ajoutée et le développement de souches industrielles de micro-organismes plus performantes. Plus globalement, elles ont cinq points forts :

- **L'expérimentation à haut débit.** La biologie à haut débit désigne une approche utilisant des technologies automatisées et des méthodes robotisées pour générer de grandes quantités de données biologiques. La finalité d'une biofonderie est d'atteindre le plus haut débit de production et de traitement de ces données. Chaque étape du cycle DBTL est réalisée de manière optimale. Lors du développement d'un procédé de bioproduction d'un composé d'intérêt, cela permet d'explorer et de cribler un vaste champ de possibilités de souches et de constructions génétiques.
- **Le prototypage et les itérations rapides.** Une biofonderie enchaîne les cycles DBTL à haute fréquence. Elle construit et teste donc rapidement les systèmes biologiques conçus. Ainsi, les intervalles entre deux cycles DBTL sont réduits, ce qui est primordial pour accélérer la mise au point des prototypes.
- **La normalisation et la reproductibilité.** Les biofonderies établissent des flux de travail et des protocoles normalisés pour le génie biologique, ce qui garantit la cohérence et la reproductibilité des procédures expérimentales. Cela minimise la variabilité entre les expériences et permet aux chercheurs de comparer et de reproduire les résultats plus efficacement. La normalisation facilite aussi la collaboration et le partage des ressources et des connaissances entre les collaborateurs.
- **L'analyse.** Les biofonderies utilisent des techniques d'analyse et de modélisation informatiques pour interpréter les grands ensembles de données générés par les expériences. En intégrant ces données aux modèles bio-informatiques, les chercheurs peuvent mieux comprendre les systèmes biologiques complexes, prédire leur comportement et optimiser les plans d'expérience. Cette approche « data driven » renforce l'efficacité de la R&D.
- **L'accélération du développement technologique.** Les biofonderies jouent un rôle crucial dans le développement technologique en biologie synthétique. En fournissant une installation centralisée dotée d'équipements et d'expertises de pointe, elles favorisent des avancées méthodologiques et technologiques qui profitent à l'ensemble de la communauté scientifique.

4 Passer à l'échelle industrielle

Les résultats obtenus sur les souches de micro-organismes dans de faibles volumes de culture dans des conditions dites de laboratoire, c'est-à-dire avec des modalités opératoires et des matières premières « idéales », doivent progressivement être portés à des échelles supérieures. Il s'agit de passer à des bioréacteurs (ou « fermenteurs ») ayant des capacités volumétriques de plus en plus grandes pour tester le comportement des souches issues du cycle DBTL sélectionnées lors de la fermentation de précision au laboratoire. Cette montée en échelle progressive, également nommée upstream process ou scale-up (fig. 2), permet de retenir, étape après étape, les meilleures combinaisons entre les souches, les recettes de milieux et les conditions de culture.

Cette phase sert également à tester la faisabilité de l'industrialisation du procédé, car à chaque accroissement de volume de bioréacteur, des phénomènes physiques, chimiques ou biologiques peuvent passer d'un impact négligeable sur le rendement du produit d'intérêt à un effet rédhibitoire.

Les étapes suivantes, dites downstream process (DSP) (fig. 3), vont consister à extraire le produit d'intérêt, le concentrer, le purifier et le stabiliser pour le conditionner. Il existe de très nombreux procédés DSP qui diffèrent en fonction de la nature du produit (biocarburant, protéine, pigment, arôme…).

5 Preuve de concept : l'ingénierie de la bactérie L. phytofermentans pour la production de butanol à partir de biomasse lignocellulosique

La biologie synthétique offre des opportunités innovantes dans le domaine de la valorisation des déchets agricoles en vue de la production de biocarburant durable. Dans cette perspective, l'un des organismes ciblés est Lachnoclostridium phytofermentans, une **bactérie anaérobie** présente dans le sol, dotée de la capacité rare de dégrader les fibres végétales.

L'histoire des biocarburants remonte à une époque antérieure à l'utilisation généralisée de l'essence dérivée du pétrole fossile. Dès l'invention du moteur à combustion interne, l'éthanol, obtenu par la fermentation de sucre par des micro-organismes, a été envisagé comme carburant. Les États-Unis en sont le plus grand producteur au monde, avec 57,5 milliards de litres d'éthanol par an. Malgré la prédominance de l'éthanol, le **butanol** émerge comme un alcool prometteur en raison de ses propriétés physico-chimiques supérieures. Contrairement au bioéthanol, il peut être utilisé pur dans les moteurs, sans modification. Il est aussi plus énergétique. Cependant, les défis liés à la rentabilité de sa production biologique entravent sa compétitivité par rapport à l'industrie chimique.

Une solution réside dans l'application de la biologie synthétique à la bactérie L. phytofermentans. En combinant les avantages du butanol avec les avancées de la biologie synthétique, cette approche offre une nouvelle perspective pour optimiser la production de biocarburant de manière économiquement viable et écologiquement responsable.

Pour parvenir à un tel résultat, il a d'abord fallu passer par l'étape de la conception (Design) du cycle DBTL. Des choix cruciaux doivent être faits sur la voie de synthèse du butanol (fig. 4). Cela inclut la détermination des enzymes, des séquences d'ADN qui les codent, des éléments régulateurs, ainsi que leur mode de clonage et le mode d'expression, en particulier l'intégration des séquences au génome bactérien ou dans un ADN circulaire indépendant (plasmide). Le métabolisme de L. phytofermentans a été modélisé pour identifier les réactions enzymatiques naturelles susceptibles d'interférer avec

P. 55 **Les enjeux de la montée en échelle**

P. 56 **Le procédé de récupération et purification des bioproduits (downstream process)**

Une bactérie anaérobie est une bactérie qui ne se développe que dans un milieu dépourvu d'oxygène.

Le butanol est un alcool primaire qui peut être produit lors de la fermentation alcoolique. Il est considéré comme un biocarburant potentiel.

P. 56 **Insertion de la voie de synthèse du butanol dans la bactérie L. phytofermentans**

la nouvelle voie envisagée. Une compréhension approfondie du métabolisme est effectivement indispensable, notamment en vue de la modification éventuelle du châssis, par l'inactivation de certains gènes naturels, visant à optimiser la production de butanol.

Une fois le projet conceptualisé, sa concrétisation (Build) s'opère par la synthèse de l'ADN et son assemblage à l'aide de techniques de biologie moléculaire telles que Golden Gate et Gibson Assembly. Dans ce cas précis, nous avons utilisé la méthode d'assemblage d'ADN Golden Gate, qui utilise des enzymes de restriction et coupe l'ADN en des points précis, les fragments étant ensuite liés entre eux grâce à une enzyme ligase. La méthode Gibson Assembly est une technique sans enzyme de restriction. Elle repose sur l'action d'une exonucléase, une enzyme qui dégrade l'ADN et rend accessibles les extrémités à une ligase, laquelle réalise une fusion des fragments d'ADN. Dans le cadre d'un tel projet d'ingénierie métabolique, il est fréquent de construire et d'explorer des milliers de combinaisons de séquences enzymatiques et régulatrices de l'activité de ces gènes. En combinant la puissance de production d'une biofonderie avec le nombre considérable de combinaisons d'ADN envisageables, la recherche de la voie de synthèse optimale devient réalisable. L'intégration de cet ADN exogène dans la bactérie s'effectue ensuite par la technique d'électroporation, qui consiste à appliquer un choc électrique créant des pores dans la membrane de la cellule hôte.

Une fois les souches dotées des nouvelles voies de synthèse, elles sont mises en culture et testées pour quantifier le butanol et sélectionner les meilleures productrices. Cette opération s'effectue au moyen d'une technique de chimie analytique telle que la **chromatographie** liquide à haute performance (HPLC). L'étape de quantification génère un ensemble volumineux de données qui enrichit notre compréhension du système biologique en cours de création. À partir des concentrations de butanol obtenues par HPLC, des analyses approfondies déterminent quels paramètres génétiques, conditions de culture ou gènes inactivés favorisent la meilleure production de butanol (Learn). Ces informations serviront de base pour lancer un nouveau cycle DBTL, auquel des ajustements successifs seront apportés. À travers ces itérations, la production de butanol chez L. phytofermentans sera progressivement optimisée.

La chromatographie est une technique servant à séparer les différents composants d'un mélange en les faisant migrer sur une phase stationnaire en utilisant une phase liquide ou gazeuse.

6 Perspectives

Il y a vingt ans, les communautés scientifique, industrielle et institutionnelle s'interrogeaient sur la biologie de synthèse, alors naissante : nouvelle discipline ou prolongation de disciplines existantes ? Science ou technologie ? Toutes s'accordaient cependant sur la promesse d'un large champ d'applications, dans presque tous les domaines : molécules thérapeutiques et vaccins, médicaments produits dans notre organisme et délivrés périodiquement en associant au gène d'intérêt une horloge génétique, biocarburants et plastiques produits par des micro-organismes n'utilisant que des ressources carbonées, colorants, arômes, adjuvants et autres ingrédients produits par des bactéries, des levures, des algues, des plantes… En 2012, l'Office parlementaire d'évaluation des choix scientifiques et technologiques pointait d'ailleurs les « applications potentielles considérables de ce domaine émergent […] oscillant entre recherche fondamentale et applications, entre biologie moléculaire, biologie systémique et modélisation ».

Aujourd'hui, la biologie de synthèse entre dans l'ère industrielle grâce aux progrès conjugués des technologies biomoléculaires, du numérique, de l'IA et de l'automatisation des procédés. Les biofonderies traduisent ce tournant. La Commission européenne a classé, le 3 octobre dernier, les biotechnologies et la biologie de synthèse dans les technologies clés à protéger, avec l'IA, le quantique et les semi-conducteurs, afin de renforcer la sécurité économique dans un contexte de compétition internationale croissante. La France, disposant d'un tissu de laboratoires académiques et de biotechs, est bien positionnée pour développer une filière de l'ingénierie du vivant. Paris a été choisi il y a deux ans pour accueillir l'exposition universelle de biologie synthétique **iGEM**, née au MIT il y a vingt ans.

Pour parvenir à une structuration et une coordination des biofonderies académiques à l'échelle mondiale, une Alliance globale des biofonderies (https://www.biofoundries.org) a été créée en 2019. L'objectif est de partager des expériences et des ressources et d'œuvrer collectivement à surmonter les défis pour répondre aux besoins scientifiques et techniques. Citons par exemple le développement de technologies analytiques plus rapides et leur utilisation en parallèle afin de s'affranchir des goulets d'étranglement actuels dans les biofonderies. Citons également l'utilisation non pas de micro-organismes châssis uniques, mais de consortiums de micro-organismes qui permettraient la co-construction de complexes moléculaires utiles ou l'optimisation des procédés grâce aux interactions bactériennes.

Outre son intérêt écologique et son potentiel économique et industriel, la biologie de synthèse, en utilisant les propriétés illimitées du vivant, ouvre la voie, par exemple, à des molécules que ni la nature, ni la chimie ne sauraient fabriquer. La nouvelle ère industrielle qui se met en place sera probablement marquée par une forte montée en puissance de ces procédés issus des sciences et technologies du vivant. Libérés de la dépendance à la pétrochimie, les bioprocédés pourront à terme s'inclure dans des circuits d'économie circulaire. ■

L'iGEM (https://igem.org/) est une compétition étudiante de biologie de synthèse. Tom Zaplana y a remporté en 2023, avec son équipe Genopole Évry Paris-Saclay, une médaille d'or, une nomination pour le prix Best Software Tool et le prix Best Hardware.



université PARIS-SACLAY

ÉCOLE DOCTORALE
Structure et dynamique des systèmes vivants (SDSV)

Titre : Optimisation de la fermentation de la biomasse végétale par *Lachnoclostridium phytofermentans*

Mots clés : *Clostridia*, Biocarburant, Butanol, Biomasse, Ingénierie métabolique, Évolution dirigé

Résumé : Le développement de sources d'énergie renouvelables nécessite l'optimisation de la conversion de la biomasse végétale en biocarburants tels que le butanol. Les *Clostridia*, et plus particulièrement *Lachnoclostridium phytofermentans*, sont des bactéries anaérobies capables de dégrader la cellulose et de produire des composés d'intérêt, ce qui en fait des candidates idéales pour cette transformation.

Cette thèse vise à repousser les limites de l'ingénierie métabolique de *L. phytofermentans* pour optimiser la production de butanol. Les objectifs principaux sont : (1) développer une souche évoluée capable de résister à de fortes concentrations de butanol, (2) introduire une voie de synthèse du butanol via des approches de biologie synthétique, et (3) implémenter un système CRISPRi pour moduler le métabolisme central de manière précise.

Pour atteindre ces objectifs, plusieurs approches ont été mises en œuvre. *L. phytofermentans* a été soumise à une évolution dirigée sur 480 jours dans un automate de cultures continues (appelé GM3), aboutissant à une souche résistante au butanol. La caractérisation phénotypique et génomique de cette souche a permis d'identifier plus de 500 mutations associées à la résistance. Parallèlement, une voie de synthèse du butanol a été conçue et transformé dans *L. phytofermentans*. Un biosenseur spécifique du butanol a été adapté pour détecter sa production. Enfin, un système CRISPRi basé sur dLbCas12a a été développé, permettant le clonage jusqu'à huit guides pour réprimer spécifiquement des gènes clés du métabolisme.

La souche évoluée, PHY24.0#6, a montré une tolérance accrue au butanol, un temps de génération réduit de moitié et une densité de croissance maximale augmentée de 1,3 fois. L'analyse de la membrane plasmique a révélé une composition en acides gras plus longs et plus insaturés. Bien que la production de butanol n'ait pas été détectée après l'introduction de la voie de synthèse, le biosenseur spécifique a été intégré avec succès. Le système CRISPRi a permis de réprimer l'expression de gènes du métabolisme central jusqu'à 20 fois, y compris des gènes dont la délétion serait létale, comme le trio *Cphy_1326-1327-0089 responsable de la production d'acétate et de dihydrogène*.

Ces@ travaux ont permis d'améliorer la tolérance *de* L. phytofermentans au butanol et de développer des outils moléculaires innovants tels que le système CRISPRi multiplex. Ces avancées contribuent à une meilleure compréhension du métabolisme de cette bactérie et ouvrent des perspectives pour optimiser la production durable de biocarburants à partir de biomasse végétale. Les approches développées pourraient être transposées à d'autres modèles de *Clostridia*, favorisant ainsi le développement de solutions énergétiques innovantes et respectueuses de l'environnement.

université PARIS-SACLAY

ÉCOLE DOCTORALE
Structure et dynamique des systèmes vivants (SDSV)

Title: Optimization of plant biomass fermentation by *Lachnoclostridium phytofermentans*

Keywords: *Clostridia*, Biofuel, Butanol, Biomass, Metabolic engineering, Directed evolution

Abstract: The development of renewable energy sources requires optimization of the conversion of plant biomass into biofuels such as butanol. *Clostridia*, and more specifically *Lachnoclostridium phytofermentans*, are anaerobic bacteria capable of degrading cellulose and producing compounds of interest, making them ideal candidates for this transformation.

This thesis aims to push back the limits of metabolic engineering of *L. phytofermentans* to optimize butanol production. The main objectives are: (1) to develop an evolved strain capable of withstanding high concentrations of butanol, (2) to introduce a butanol synthesis pathway via synthetic biology approaches, and (3) to implement a CRISPRi system to precisely modulate central metabolism.

To achieve these objectives, several approaches were implemented. *L. phytofermentans* was subjected to directed evolution over 480 days in a continuous culture automaton (called GM3), resulting in a butanol-resistant strain. Phenotypic and genomic characterization of this strain identified over 500 mutations associated with resistance. In parallel, a butanol synthesis pathway was designed and introduced into *L. phytofermentans*, and a butanol-specific biosensor was adapted to detect its production. Finally, a CRISPRi system based on dLbCas12a was developed, enabling the cloning of up to eight guides into a plasmid to specifically repress key metabolic genes.

The evolved strain, PHY24.0#6, showed increased tolerance to butanol, a halved generation time and a 1.3-fold increase in maximum growth density. Analysis of the plasma membrane revealed a composition of longer, more unsaturated fatty acids. Although butanol production was not detected after introduction of the synthetic pathway, the specific biosensor was successfully integrated. The CRISPRi system was able to repress the expression of central metabolism genes up to 20-fold, including genes whose deletion would be lethal, such as the *Cphy_1326-1327-0089* trio, responsible for the acetate and dihydrogen production.

This work has made it possible to improve the tolerance of *L. phytofermentans* to butanol, and to develop innovative molecular tools such as the multiplex CRISPRi system. These advances contribute to a better understanding of this bacterium's metabolism and open prospects for optimizing the sustainable production of biofuels from plant biomass. The approaches developed could be transposed to other models of *Clostridia*, thereby fostering the development of innovative, environmentally friendly energy solutions.