NNT : 2017SACLE018

THESE DE DOCTORAT
de
L'UNIVERSITE PARIS-SACLAY
préparée à
L'UNIVERSITE D'EVRY VAL D'ESSONNE

Ecole doctorale N° 577 : Structure et Dynamique des Systèmes Vivants
Spécialité : Sciences de la vie et de la santé

Par

Tristan Cerisy

Rational and evolution-based engineering of *Clostridium phytofermentans*

Evolution dirigée et ingénierie rationnelle chez *Clostridium phytofermentans*

Sous la direction d'Andrew Tolonen et de Marcel Salanoubat

Laboratoire d'acceuil : Laboratoire de Génomique et Biochimie du Métabolisme, Genoscope, UMR8030 Génomique Métabolique, Université d'Evry/CNRS/CEA, France

Thèse présentée et soutenue publiquement le 5 juillet 2017 à Evry, devant un jury composé de :

Bruno Dupuy	Président du Jury	Institut Pasteur	Chercheur
Henri-Pierre Fierobe	Rapporteur	CNRS Marseille	DR
Philippe Soucaille	Rapporteur	INSA Toulouse	Professeur
Nicolas Lopes Ferreira	Examinateur	IFP Energies Nouvelles	Chercheur
Andrew Tolonen	Co-directeur de thèse	Broad Institute	Chercheur
Marcel Salanoubat	Co-directeur de thèse	CNRS Paris-Saclay	DR

REMERCIEMENTS

Je souhaiterais adresser mes sincères remerciements à mon co-directeur de thèse, Andrew Tolonen. Il a accepté de me prendre en tant qu'ingénieur de recherche puis en doctorat. Par sa bienveillance, ses conseils, l'apport scientifique et humain, j'ai appris à tous les niveaux à ses côtés. Thank you ! Je remercie aussi Marcel Salanoubat, mon second co-directeur de thèse, qui m'a accepté dans son UMR. Suite au départ d'Andrew nous avons énormément échangé scientifiquement et humainement, j'ai découvert les responsabilités et la vision d'un directeur, un travail bien différent d'un scientifique. Merci pour ton soutien, tes blagues bien placées, ton aide pour les conférences et aussi tes signatures pour ces milliers de documents administratifs. Merci à MM. Tambourin et Mandon pour leurs soutiens en 2012, autant pour iGEM que pour avoir cru dans l'utilité de ce CDD au Genoscope (avec l'aide de Jean Weissenbach), préambule à cette thèse.

Je remercie mon jury de thèse, M. Henri-Pierre Fierobe, directeur de recherche au CNRS à Marseille, M. Philippe Soucaille, Professeur à l'INSA Toulouse, M. Bruno Dupuy, *Principal Investigator* à l'Institut Pasteur, et M. Nicolas Lopes Ferreira, chef d'équipe à IFPEN, pour avoir accepté d'étudier et de discuter de mon travail.

Je veux remercier très spécialement Magali, notre super technicienne, la force de cette équipe, elle a été essentielle dans les réussites de cette équipe et de ce doctorat. Sans ses qualités humaines et scientifiques, qui vont bien au-delà de ce que l'on demande à la plupart des ingénieurs, nos résultats ne seraient pas ce qu'ils sont. Merci aussi pour tous les moments que l'on a passé avec l'équipe sur Evry, Corbeil, Ris, Paris voir Blois !

Je m'excuse auprès de tous ceux qui m'ont supporté dans leurs bureaux : Christine, Lucille, Virginie, Isa, Anne, Valérie, Véronique, Jean-Louis et Pascal. J'ai changé plusieurs fois de bureau, j'ai côtoyé les ingénieurs, les techniciens et j'ai fini par partager le bureau de Véronique, la chef du LCAB, pour l'écriture de la thèse. Nos débats sur toutes les facettes du monde vont me manquer. J'ai partagé de supers moments avec tous, des fois avec un peu d'EtOH, de belles discussions, et j'ai pu voir le point de vue des différents échelons d'un labo, ce qui a été très enrichissant.

Je remercie les autres équipes avec qui nous avons collaboré dans l'UMR: Nathalie, Ivan et Volker pour le GM3, Virginie pour le Bradford, Emilie pour les RT-qPCR, Sébastien pour la boîte à gants, Nuria pour ses éclaircissement sur le métabolisme de *C. sticklandii*, le personnel de la laverie, Corinne et Karine pour leurs séquençages, parfois à la dernière minute, Alain et Jean-Louis pour l'expertise en enzymologie. En fait je voudrais remercier spécialement Jean-Louis, le MacGyver du labo, la pierre angulaire du laboratoire, cet ingénieur « lab-ninja » est un soutien énorme pour tous les stagiaires et doctorants en quête de résultats ! (malgré ses airs un peu bourru). Merci infiniment.

Je voudrais aussi remercier nos stagiaires : l'incroyable Dutch, Emiel VDK; Ismaël qui a initié le projet d'évolution dirigée avec réussite et talent, maintenant en doctorat à Marseille ; mais aussi Audam et Alba que j'ai encadré et qui ont travaillé de manière extraordinaire, leurs résultats sont publiés ou vont l'être prochainement. Je leur dois beaucoup et je leur souhaite bonne chance pour leurs doctorats en Angleterre. Je n'oublie pas notre collaboratrice Indienne, Sanitha, une doctorante venue six mois, avec Magali elles ont mis en place le protocole permettant d'électroporer *C. phytofermentans*, un duo de choc très efficace entre deux cultures bien différentes, des amies maintenant.

L'ambiance parmi les doctorants, qui ont été de plus en plus nombreux avec les années, fût une force. Et c'est avec un plaisir immense que je les abandonne pour qu'ils terminent le leur ! Un clin d'œil spécial à Tiffany qui a commencé et terminé ce doctorat en parallèle du mien, j'espère que la science continuera à infuser ton esprit de génie. Tu es géniale et j'espère que tu ne quitteras pas la science demain ! Courage à Marion, Marion, Ombeline, Sophia et Thomas. Grande pensée pour Lucille qui nous a précédé et avec qui j'ai passé d'excellents moments ! Elle manque toujours au labo. Je remercie spécialement Sophia qui m'a fait vivre de fabuleux moments à travers cette planète, merci et continue d'être toi, pétillante, drôle et curieuse ! Je remercie aussi Ombeline, qui m'a fait découvrir le badminton (et le tournoi inter-entreprise, de merveilleux moments !) et qui est devenue une vraie amie. Courage à Hela qui va se lancer dans un doctorat prochainement. Je remercie les stagiaires des autres labos avec qui nous avons partagé de grands moments, spécialement Emilie qui est maintenant une amie et qui m'a bien aidé pour la relecture de certaines parties de la thèse ; mais aussi les post-docs Petra et Anestis, avec qui nous avons partagé pintes et sorties parisiennes.

Je ne peux pas oublier Nicolas Pollet qui m'a fait découvrir la science, la vraie dans un laboratoire, ainsi que la rigueur qui va avec, merci ! Merci aussi à Jean-Loup Faulon qui m'a accepté dans le master mSSB et qui a cru en nous dans les aventures iGEM. Je remercie William, Cyrille, Carolina, Hafez, Artémis, Nicolas, Andrew, Tiffany, Karine, Pierre, Iryna, Thomas, Clément etc… de l'équipe iGEM Evry 2012. La création de cette équipe avec William, Cyrille et moi a été le début de mon aventure scientifique. En tant qu'étudiant en 2012 puis en tant qu'advisor en 2013 (Emiel, Guillaume, Audam, Louis, Léni, Hippo, Gab etc.) et 2014 (Cécile, William, Sophia, Noémie, Matthieu etc.), les expériences iGEM m'ont énormément appris en science mais aussi sur les côtés administratifs, communications, managements etc. Merci à tous ceux avec qui j'ai partagé ces aventures et ceux qui nous ont soutenu. Merci à Cyrille de me donner l'opportunité de continuer à développer un projet sur *C. phytofermentans*, j'espère une nouvelle belle aventure.

Merci Marcel de m'avoir permis de faire de l'enseignement. Donner des TP, TD et des cours m'a ouvert l'esprit et j'ai découvert le dur métier d'enseignent-chercheur. Je remercie les enseignants que j'ai accompagné (qui ont été pour certains mes professeurs précédemment) ainsi que les étudiants. Essayons de les encourager à aller toujours plus loin, leur montrer la beauté de la science et des belles aventures qu'elle peut apporter, faisons leur confiance.

A mes colocataires : Laura, Sophie, Pascal, Etienne, Robert et Nicolas avec qui je passe de grands moments depuis plus de 3 ans et qui me supportent, merci pour votre soutien il était primordial. Merci à mes amis du lycée, Xavier, Cyril, Benoit, Julien, Yoann et ceux qui sont arrivés après : Pom, Robin. Merci à mes amis du badminton : Robin, Nicolas, Gilles, Julie, Pom, Julie, Emilie et Ombeline. Nos week-end, ici ou ailleurs sont à chaque fois de supers moments, permettant de penser à autre chose (que la thèse).

Le sport a toujours été indispensable dans ma vie, et la natation, que je pratique depuis 20 ans avec Nicolas, Arthur, Guillaume, Mathieu, Anthony, sous la rigueur des coaches Célestin et Franck, m'a permis 4, puis 3 puis 2 fois par semaine (oui avec les années de thèse cela descend, hélas) de décompresser, de m'évader, c'est ma drogue. Nous la partageons ensemble et je vous remercie d'être là mes amis, en espérant faire encore de nombreux championnats, mais aussi soirées, tennis, courses, vélo, vacances etc. avec vous.

Petite pensée à mon chirurgien, le Dr. Finidori, qui m'a opéré une quinzaine de fois, sans lui, je ne sais pas dans quel état je serais, ni où, ni pourquoi; certaines de ses paroles à la sortie du bloc sont toujours dans ma tête, merci.

Je remercie ma famille, mes parents, qui ont toujours été d'un soutien indéfectible, je suis heureux de partager autant avec vous, même nos disputes et nos réconciliations, prenez soin de votre santé, vous êtes tellement important, je vous dois tout. Je voudrais féliciter ma sœur, Laure, pour qui j'ai une estime et une admiration sans égales, tu me manques beaucoup et je suis très fière d'être ton frère. Je vous aime tous les trois. Je suis heureux d'avoir eu mes grand-mères à mes côtés pendant ce doctorat, Badette tu as rejoint André, Christophe et Jean-Marie. Je pense à vous en cette fin de thèse, vous qui êtes partis et avec qui j'aurais voulu passer encore de longs moments, vous m'avez tellement apporté, vous avez été géniaux ! Je suis très fière d'avoir ma grand-mère Aline à mes côtés pour ma défense. Merci à mes cousins, oncles et tantes, ils sont tous tellement cool.

Voilà ce fût de longs remerciements, à la hauteur de ce que m'ont apporté toutes ces personnes. Toutes ces rencontres, ces échanges, ces conseils, ces aventures font ce que je suis et donc aussi cette thèse, je les remercie toutes et tous ! Cette thèse aura été une aventure inoubliable, à la fois collective et solitaire (pour l'écriture)

Je finirais par remercier une nouvelle fois Andrew, car il a été au-delà de ce que je pouvais rêver en tant que directeur de thèse, j'espère pouvoir collaborer encore avec lui. Je lui souhaite énormément de bonheur à lui, Charlotte et Eric.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

I.1 Clostridium

I.1.1 Characteristics

I.1.1.a History

Research of the Clostridia has a long and important history. *Clostridium botulinum* was isolated and described by the Belgian microbiologist Emile Pierre van Ermengem in 1895, even if symptoms were described 60 years before [1]. At the beginning of XIXth century *Clostridium acetobutylicum* was isolated, described and directly used by Chair Weizmann, a researcher at the University of Manchester (UK), to ferment starchy substances to acetone-butanol-ethanol (ABE) [2]. *Clostridium welchii* (formerly *Clostridium perfringens*) [3] and *Clostridium putrificum* [4] were described in two articles from the 1920's about medical issues of these species for human health. They explain how to differentiate these species from other close bacteria based upon morphological and culture characteristics.

I.1.1.b Basic specificities

In theory, *Clostridium* is a genus which belongs to the domain Bacteria, phylum Firmicutes (from latin *Firmus cutis*: thick skin), Class Clostridia, order Clostridiales, family Clostridiaceae. As other Firmicutes, *Clostridium* possess Gram-positive cells walls. The class Clostridia is composed of obligate anaerobes and Clostridiaceae which mainly form endospores, have a low G+C content, and are rod-shaped. Species and strains of this genus are represented around the world in various environments (guts, soils, tissues, marine water etc.). They have various properties: some are pathogenic, other begnin, some cellulolytic other can fix carbon monoxide. The utilization of nutrients is also very diverse. Although *Clostridium* are different (environment, metabolism, pathogenecity etc.), these bacteria are rod-shaped and appear similar, Figure 1. In that way, before that sequence based phylogeny was widespread, microbiologists assigned numerous new discovered species with these visual and previous basics characteristics to the genus *Clostridium*.

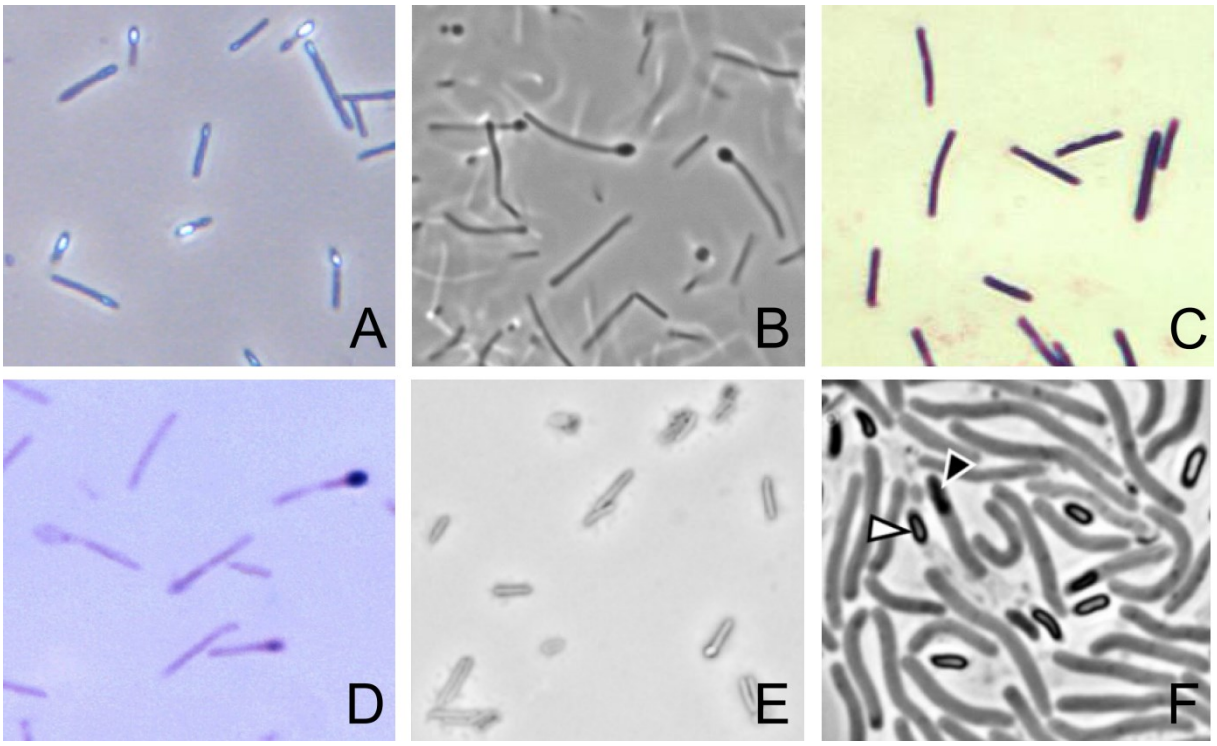

Figure 1 Six microscopic images of *Clostridium*. A *Clostridium beijerinckii,* from [5], B *Clostridium phytofermentans*, C *Clostridium perfringens* from Centers for Disease Control and Prevention, D *Clostridium botulinum* from public domain, E *Clostridium cellulolyticum* from [6], F *Clostridium diffile* from [7].

I.1.1.c Membrane

The goal of the bacterial membrane is to protect the cell from external substances, preserve the cytoplasmic components, and mediate transfer of compounds from inside and outside of the cell. To grow and survive, the cell needs to incorporate nutrients (and others needed compounds) and release others. The cell releases some cytoplasmic compounds to keep the proper osmotic pressure in the cell, to prevent the accumulation of toxic end products of the metabolism, and to liberate proteins into the media (for instance enzymes to degrade polysaccharides to metabolizable sugars by the bacteria).

The membrane of bacteria was historically classed in two categories, Gram-positive and Gram-negative. This classification depends on the result of the Gram stain test, a test published 133 years ago by Hans Christian Gram. But last phylogenic trees[1], based on the comparison of sequence of 16S RNA, modifies this dichotomy and gram-positive and gram negative bacteria

[1] Definition of phylogenic tree or phylogeny: "it is a diagram that depicts the lines of evolutionary descent of different species, organisms, or genes from a common ancestor. Phylogenies are useful for organizing knowledge of biological diversity, for structuring classifications, and for providing insight into events that occurred during evolution.", from Baum *et al.*, 2009[434].

can be present in same phylum, for example Negativicutes (Gram-negative) and Clostridia (Gram-positive) are both present in Firmicutes. The main difference between this two class of bacteria is the presence of one thick peptidoglycan cell wall (Gram-positive) or a thin layer of peptidoglycan (Gram-negative).

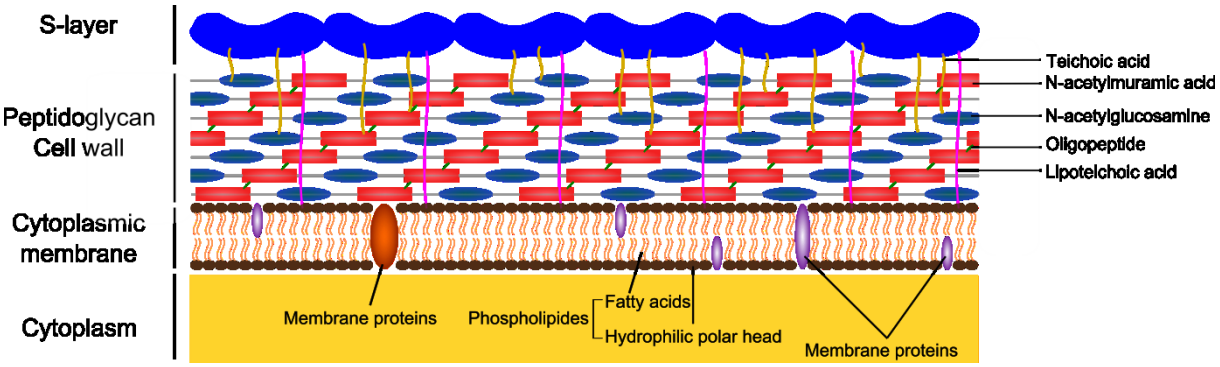

Figure 2 General characteristics of the membrane of gram-positive bacteria.

Components of the membrane are described in the text below, from Gram-positive attributes, with some characteristics described in the literature, which are specific to *Clostridium*, Figure 2.

The cytoplasmic membrane is formed by a phospholipid bilayer. This first membrane is made of hydrophilic polar head (composed of an organic compound and a phosphate group) attached to a fatty acid. Depending on the environment or a stress, the characteristic of the fatty acid chain can be modified [8]: the proportion of saturated/unsaturated fatty acids, the length of the chain, or fatty acids can be branched. This phospholipid bilayer is relatively impermeable to molecules (ions, proteins, etc.), preventing the diffusion of molecules through the membrane. These molecules remain outside or inside the cell and can pass through the phospholipid bilayer *via* different membrane proteins.

These membrane proteins are present within the phospholipid bilayer and allow the transfer of ions , products of metabolism, nutrients, toxic substances, proteins, macromolecules, electrons, etc.[9]. The transport of molecules can be passive or active. With a passive transport, a molecule does not need energy to cross the membrane. In contrast, an active transporter generally uses the energy of a cofactor to drive the passage of one molecule in or out of the cell. Depending on the energy state of the cell, molecules cross the membrane in one direction or the other. These transport proteins can represent 13 to 18 percent of total encoded genes in some gram-positive bacteria [10]. Membranes proteins are organized in different classes (which contain several subclasses) [9]:

- Channels/pores (including cytotoxic pore-forming protein (for instance: it is the toxic substance of *C. perfringens* [11]; or porins)

- Electrochemical potential-driven transporters (uniporter, symporter or antiporters which drive the transport of a substrate using a concentration gradient, with or without the help of another molecule).
- Primary active transporters (including ABC transporter)
- Group translocators (including PTS system)
- Transmembrane electron carriers
- Accessory factors involved in transport (including enzyme EI and HPr of the PTS system)
- Incompletely characterized transport systems (polysaccharide transporter, metal (Fe^{2+}, Mg^{2+}, Pb^{2+}, Hg^{2+}, copper), and ions (Na^{+}, K^{+} or H^{+}, etc.)

This membrane is covered by a peptidoglycan cell, Figure 2, which is made of layers of linked N-acetylmuramic acid and N-acetylglucosamine. These layers are linked with a peptide of five amino acids between N-acetylmuramic acids of two layers. Moreover these layers are also attached together from cytoplasmic membrane to the surface with lipotechoic acid, anchoring the peptidoglycan to the membrane. Teichoic acid links reinforce external peptidoglycan layers and its phosphate groups give a negative charge to the cell wall. This negative charge is important to promote cell growth, to prevent wall damage/lysis and to bind ions (cations) to the cell.

The external layer of the cell in many bacteria is the Surface-layer (S-layer), composed of self-assembled monomeric proteins which form a paracrystalline array coating the cell. The S-layer is increasingly studied because its importance for growth, cell integrity, survival, enzyme display, pathogens, host interaction and immune system for Clostridia [12] is now recognized. This S-layer is not present in model organism, such as *Escherichia coli* and *Bacillus subtilis*, explaining its relative late deep study. The well-studied S-layer of *C. difficile* is formed by a diversity of proteins with at least three tandem of CWB2 domain and a variety of effectors with different functions: peptidoglycan hydrolases, putative adhesins, peptidase etc., [13] and Figure 3.

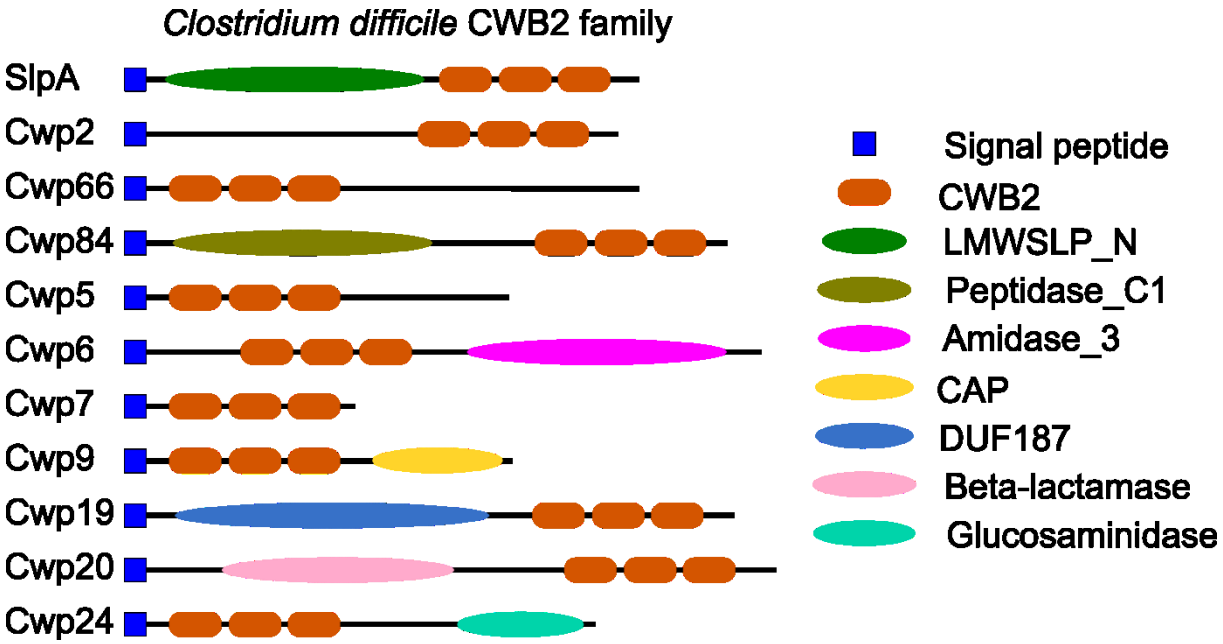

Figure 3 *Clostridium difficile* cell surface protein families. Eleven different domain organization (among 29) of *C. difficile* CW_binding_2 (CWB2) family, simplified from Fagan *et al.*, 2014 [13].

Two types of S-layer anchoring modules were found in Clostridia, the surface layer homology (SLH) domain (found at least in *Clostridium thermocellum*) and the cell wall binding 2 (CWB2) domain (present in *Clostridium tetani* [14], *C. botulinum* [15] and *C. difficile* [16]). In *C. difficile*, the S-layer is composed of high-molecular weight (HWP) and low-molecular (LMW) surface layer protein (SLP). These SLP are produced from different proteolytic cleavage of the precursor S-layer protein SlpA. The S-layer possesses several functions and displays enzymes at the surface of the cell. New functions have to be discovered. The role of the S-layer in the pathogenicity of Clostridia is investigated and studied as a potential therapeutic target [16][17] and a toolkit for nanobiotechnological applications [19].

I.1.1.d Biofilm

Beyond the membrane and the S-layer, another biological component from Clostridia exists: the biofilm. The biofilm congregate Clostridia, forming a close bacteria community which work and communicate together, increasing substrate polymer break down and protecting each other from degradation, oxygen or antibiotics. The biofilm formation is a collective response to a low stress, whereas an intense stress induces sporulation.

Cellulolytic bacteria, such as *C. thermocellum* or *C. phytofermentans* form a biofilm to be anchored to the lignocellulose biomass, improving the degradation and uptake of polysaccharides sugars [20][21][22]. Pathogenic bacteria use biofilms to grab and colonize an

environment, for instance the gut. The biofilm increases the resistance to antibiotics and the immune defense system; improving pathogenesis and persistence [23][24]. The biofilm is made of a matrix of several compounds secreted from Clostridia. This matrix is composed of an aggregate of proteins, extracellular DNA and polysaccharides [21].

I.1.1.e Sporulation

Spore formation is one of the main characteristic of Clostridia. During stress conditions (heat, starvation, antibiotics, desiccation, oxygen stress for anaerobic organism, radiation etc.) bacteria, such as Bacillus and Clostridia, transform their vegetative metabolism to a dormant structure, called spore or endospore [25], Figure 4. Spores are the most resistant cell type to stress conditions previously enumerated [26] and some spores can survive thousands of years before recovering [27]. An alternative non growth state was demonstrated in *C. thermocellum*, the L-form (also called protoplast or cell wall deficient form) [28]. This L-form is more resistant than vegetative state but less than spore. The metabolism of the L-form is slow, but the bacteria have a faster recovery than the spore in favorable conditions. Between *Clostridium* species, spores can have different forms and appendages [29].

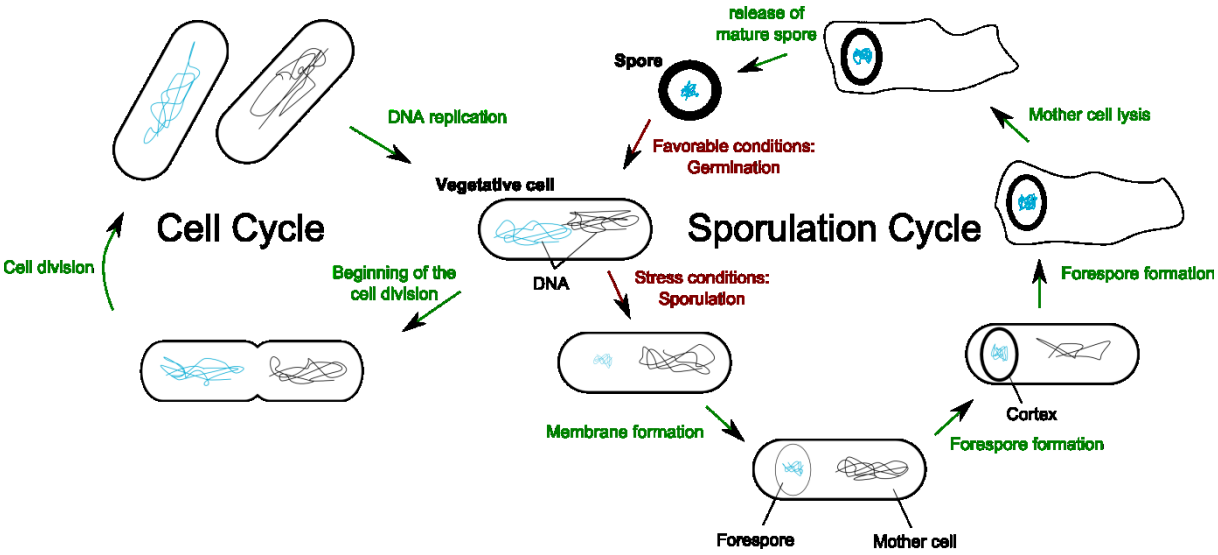

Figure 4 Cell and sporulation cycle of sporulating bacteria.

Spore formation is well studied for two main reasons: the spore resistance is an important health issue, and sporulation can be associated with solventogenesis for industrial strains.

As a spore, Clostridia can survive several years after antibiotic exposition or pasteurization (but not to sterilization)[30]. One way to prevent issue with spore from pathogenic Clostridia is to avoid the germination of the spore (mainly with preservatives). While pasteurization (70 to 80°C) preserves most of food properties, sterilization (110 to 120°C) can alter its nutritive

properties. In the food industry, the main issue is related to the pathogenicity, but the fermentation of food with contaminated Clostridia can alter characteristics of the food product. After the germination of nonpathogenic spore, the vegetative cell can release unwanted products during the fermentation, altering the taste of the food.

In the industry, the sporulation is well studied because *C. acetobutylicum* produce more solvent just before and during the sporulation phase [31], the solventogenesis is coupled to the sporulation. The sporulation abolishment in *C. cellulolyticum* improves the cellulose consumption [32]. These examples show the important role of the sporulation for Clostridia, for pathogenic and industrial issues. Although the *Bacillus* sporulation system is the model and very well known, the *Clostridium* sporulation system is becoming well studied and documented too.

Sporulation stages of *Clostridium* were well defined by Labbé, 2005 [33], But only apparent changes in the shape were described. With the help of the genome sequencing, transcriptomic and gene disruption, the complexity of genes and regulators involved in the sporulation was shown for*: C. difficile* [34][35], *C. acetobutylicum* [36], *C. perfringens* [37].

In short, the sporulation process begin when a *Clostridium* replicates its genome and senses a stress. The master gene of the sporulation is the transcriptional regulator Spo0A. This gene is common to spore forming bacteria. The activation of this gene triggers several sigma factors which will active other genes (beginning the sporulation) and small acid-soluble proteins to protect the spore DNA from damages (UV radiation, heat, enzyme degradation etc.)[38]. This metabolic cascade separates the vegetative cell into two compartments: the mother cell and the forespore (included in the mother cell). Sigma factors from the forespore and the mother cell will be transcriptionally expressed or repressed, continuing the sporulation process. Once the forespore is complete, the mother cell is removed with an autolysis, and the dormant spore is released, Figure 5.

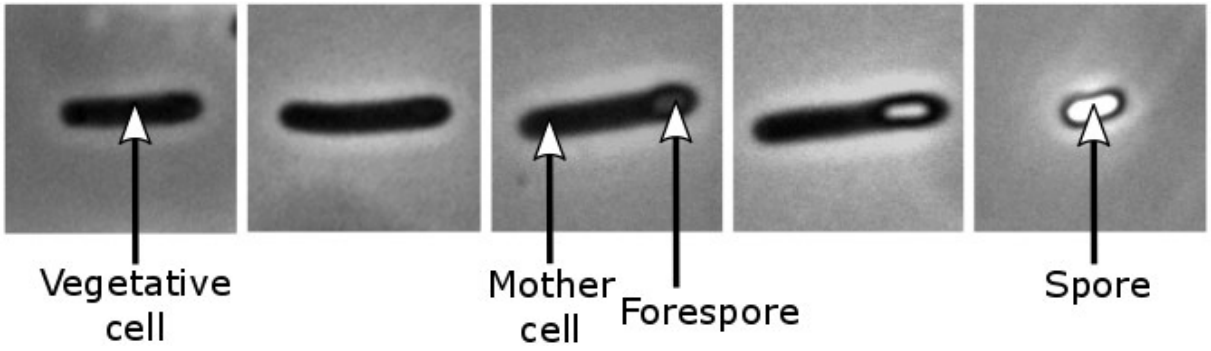

Figure 5 Visual step of sporulation in *C. difficile* with phase contrast microscopy, modified from Pereira *et al.*, 2013 [39].

Narrow differences in the sporulation process was revealed between some Clostridia (Figure 8 of Fimlaid *et al.*, 2013 [35]) and the sporulation model organism *B. subtilis*.

Moreover, proteins at the surface of the *Clostridium* spore, called spore coat proteins can also have an enzyme activity. In *C. difficile*, these enzymes play a role for the coat polymerization and detoxification of H_20_2 [40]. Spores are studied for medical application, against Clostridium pathogenicity, cancer treatments and probiotics. The idea to use spore as a cancer treatment, is relatively simple. *Clostridium* spores can germinate and proliferate only/specifically in hypoxic tissues. Fortunately, only cancer tumors tissues are both, hypoxic and favorable for spore germination conditions. By colonizing the tumor, the non-toxic and non-pathogenic *Clostridium butyricum* (also called *Clostridium oncolyticum* and now belongs to *Clostridium sporogenes*) produce its oncolytic enzymes [41][42]. These experiments were investigated more than 60 year ago and recently receive more and more attention and development [43].

I.1.1.f Germination

When the spore is in a favorable environment, it will germinate to become an active vegetative cell again. The germination is triggered by physical factors and germinant compounds [26]. Germinant receptors can sense several germinants, such as amino acids for the non-proteolytic *C. botulinum* [44]. An interesting review presents germinants of a variety of Clostridia [45], it is also presented that germination of pathogenic can be induce with host compound. For instance *C. difficile* or *C. tetani* germination is induced by cholesterols-derived from bile acid or hormones (progesterone).

Once the germination is induced, several genes are expressed: germination-specific proteases degrade DNA protective proteins (small acid-soluble proteins) and the spore cortex lytic enzymes degrade spore's peptidoglycan. Amino acids released from these degraded proteins are used for protein synthesis and energy metabolism (for proteo- and pepto-lytic Clostridia)[46]. Two different germination receptors are present in Clostridia, the *ger*-type and the *csp*-type. The *ger*-type receptors are activated by germinants, and are present in *C. botulinum*, *C. thermocellum*, *C. tetani* etc. The location of the *ger*-type receptor is proposed to be in the inner membrane, as described for *B. subtilis*. In contrary the *csp*-type present in *C. difficile*, *C. perfringens* or *C. phytofermentans* [47] seems to be localized in the cortex of the spore.

I.1.1.g Carbohydrate-Active Enzymes

Carbohydrate-Active Enzymes (CAZymes) are enzymes which depolymerize, assemble or modify mono-, di, oligo- and poly- saccharides [48]. These particular enzymes are present in all kingdoms (Archea, Bacteria, Eukaryota and Viruses). Glycosyltransferases are involved in the

assembly of polysaccharide, whereas glycoside hydrolases (GH), polysaccharide lyases and carbohydrate esterase are associated with its breakdown. Nevertheless this classification is large and for instance 136 different GH families exist, to date; revealing the diversity of substrates of GHs. The classification is based on amino acid sequence similarities, nevertheless in the same GH family several substrates can be catabolized and several GH family can share the same enzyme activity [49]. In cellulolytic Clostridia, CAZymes and especially GHs are important to deconstruct polysaccharides of the plant biomass. Thus, resulting oligosaccharides are utilized as substrate by cellulolytic Clostridia. Moreover these bacteria possess plentiful of diverse GHs to degrade various and different polysaccharides [50](Appendix 1, [51]). To depolymerize the plant biomass, CAZYmes are secreted outside the cell, freely or attached to a cellulosomes. The cellulosome can be attached or freely present in the medium [52].

I.1.2 Fermentation and general metabolism

I.1.2.a Diversity of Fermentation

The variety of substrates fermented by Clostridiales leads to different metabolic pathways, producing a diversity of chemical compounds, Table 1Table 1 Major metabolic features of *Clostridium* Fermentation, modified from Dürre *et al.*, 2014 [26].. Clostridiales are divided in different main metabolisms: saccharolytic, alcoholic, acid or a proteolytic metabolism [26].

Fermentation pathway	Substrates	Products	Representative species	Cluster [53][54][55]
Homoacetate (Wood-Ljungdahl)	Sugars (i.e., fructose), $CO_2 + H_2$, CO)	Acetate or ethanol	*C. Ljungdahlii, C. autoethanogenum*	I
Cellulolytic	Saccharides and polysaccharides	Ethanol, H_2, acetate, CO_2, lactate	*C. cellulolyticum, C. phytofermentans*	III and XIVa respectively
Proponiate (Acrylyl-Coa)	Lactate	Propionate, acetate, CO_2	*C. propionicum*	XIVb
Proponiate (Succinate decarboxylation)	Succinate	Propionate, CO_2	*C. mayombei*	XI
Butyrate	Sugars (e.g., glucose)	Butyrate, acetate, CO_2, H_2	*C. butyricum*	I
Isopropanol	Hexoses, Pentoses, Starch	Isopropanol	*C. beijerinckii*	I
Acetone/butanol	Hexoses, Pentoses, Starch	Acetone, butanol, ethanol	*C. acetobutylicum*	I

Alcoholotrophic	Various alcohols	Various acids and solvents	*C. aceticum*	XI
Alcoholotrophic	alcohol + acid	Various acids	*C. kluyveri*	I
Proteolytic, peptolytic	Proteins, Peptides	Acetate, propionate, Butyrate, CO_2, $NH4^+$	*C. propionicum*	XIVb
Use of amino acids	Amino acids		*C. tetanomorphum*	I [56]
Stickland reaction	Pairs of amino acids	Acetate, CO_2, $NH4^+$	*C. sticklandii*	XI
Purinolytic	Purines	Acetate, formate, CO_2, $NH4^+$	*C. purinilyticum, C. acidurici*	XII
Pyrimidinolytic	Pyrimidines	Acids, CO_2, $NH4^+$	*C. oroticum*	XIVa

Table 1 Major metabolic features of *Clostridium* Fermentation, modified from Dürre *et al.*, 2014 [26].

I.1.2.b Glycolysis from hexose and pentose

Most *Clostridium* ferment common sugars like glucose or fructose. For instance, *C. acetobutylicum*, *C. butyricum*, or the pathogene *C. difficile* ferment mono- or di-saccharides. In addition to these substrates, cellulolytic *Clostridium* can degrade several polysaccharides to oligosaccharides. Oligosaccharides are chains constituted of 2 to 10 sugars that can be assimilated by cellulolytic *Clostridium*, for instance *C. thermocellum*, *C. phytofermentans* or *C. cellulolyticum*. These oligosaccharides are transported and cleaved inside the cell to monosaccharides. Hexose or pentose monosaccharides, from free monosaccharides or from cleaved di- or oligo-saccharide directly enter in the metabolism via two different routes. Directly to the glycolysis for hexose, pentose goes to the glycolysis via intermediates: glyceraldehyde-3-phosphate or fructose-6-phosphate.

Two different metabolic pathways are present in most of pentose metabolizing *Clostridium*: the pentose phosphate pathway and the phosphoketolase pathway [57], Figure 6. The pentose phosphate pathway transforms xylulose-5-phosphate and ribose-5-phosphate to the intermediate glyceraldehyde-3-phosphate and sedoheptulose-7-phosphate with a transketolase. The phosphoketolase pathway produces acetyl-phosphate and glyceraldehyde-3-phosphate, from xylulose-5-phosphate and phosphate. Acetyl-phosphate can be metabolized to acetate (generating ATP) or acetyl-Coa.

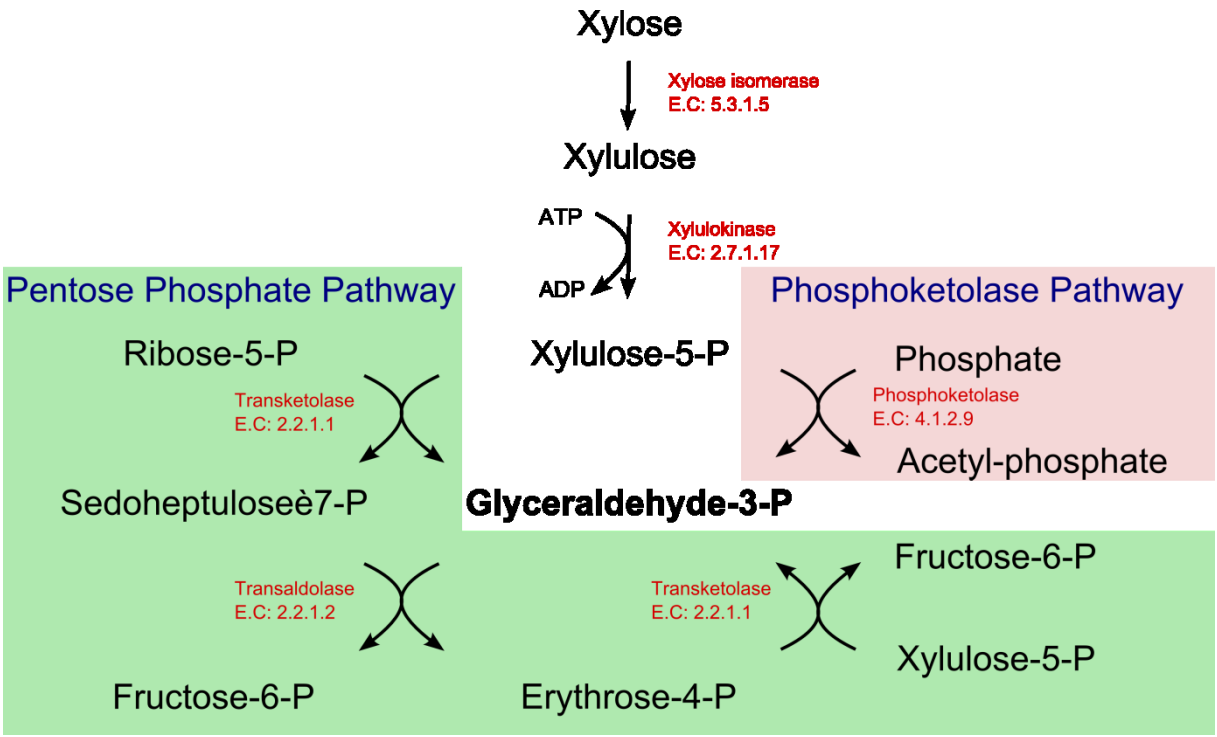

Figure 6 The two pentose pathways in some Clostridia: the pentose phosphate pathway (green) and the phosphoketolase pathway (pink).

I.1.2.c Homoacetogenic pathway

Clostridium with the homoacetogenic pathway can utilize CO_2 or CO as electron acceptor and H_2 as electron donor in autotrophic conditions, or use sugars, like glucose, as carbon source in heterotrophic conditions. Acetate is central in the metabolism of acetogens, it is a product of the fermentation, but also a substrate. These bacteria use the Wood-Ljungdahl pathway[58] to fix the gaseous carbon, Figure 7.

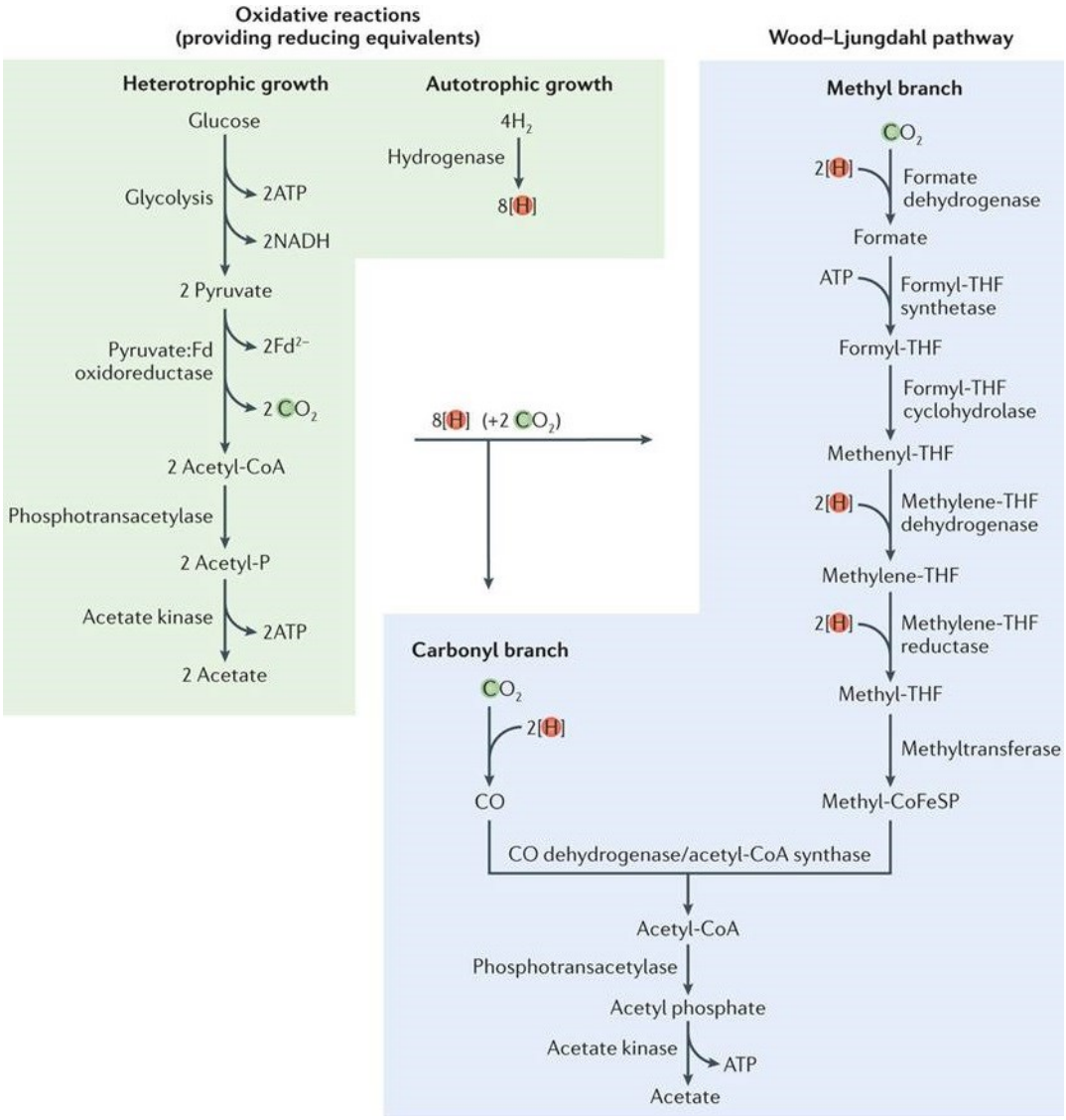

Figure 7 Carbon assimilations in homoactogene *Clostridium* and description of the Wood-Ljungdhal pathway (blue) from Schuchmann *et al.*, 2014 [59].

This metabolic pathway was discovered by teams of Harland Goff Wood and Lars Gerhard Ljungdahl and they worked on *Clostridium thermoaceticum* (now reassigned as *Moorella thermoacetica*). The first ^{14}C-tracer study in biology was used to understand the CO_2 fixation of this bacterium [60]. In this experiment, it was revealed that *M. thermoacetica* produce acetate from CO_2. The fermentation of glucose occurs with CO_2 labelled with ^{13}C carbons. Three different mass of acetate were found with equimolar ratio, acetate with mass 60 contains only unlabeled carbons, acetate with mass 61 possess one fixed ^{13}C carbon and the acetate with mass 62 is made only of fixed ^{13}C carbon from CO_2, Figure 8.

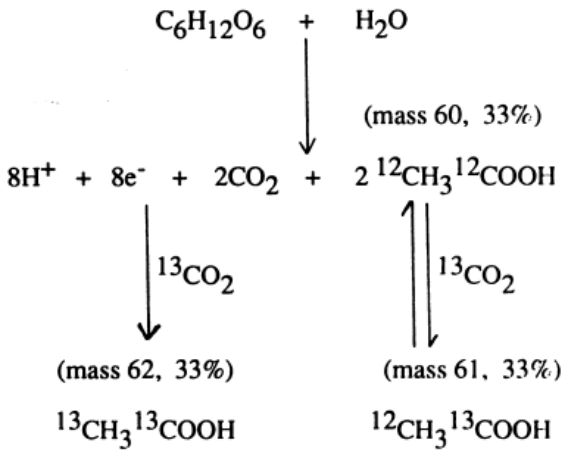

Figure 8 Evidence by mass analysis that there is total synthesis of acetate from $^{13}CO_2$ *by C. thermoaceticum* from Wood *et al.*, 1991 [61].

Organisms at the Life's origin probably used this pathway to fix atmospheric carbon [62].

I.1.2.d Alcoholotrophic pathways

Clostridium possessing alcoholotrophic pathways can ferment alcohol and acids. For instance *Clostridium kluyveri* which was isolated in 1937 in Delft, The Netherlands [63]., ferments ethanol and acetate (or succinate) to butyrate, caproate, methane and H_2, Figure 9. At the end of the growth, butanol or hexanol can be produced [64]. Acids and alcohol are the energy source of this strain and *C. kluyveri* was isolated

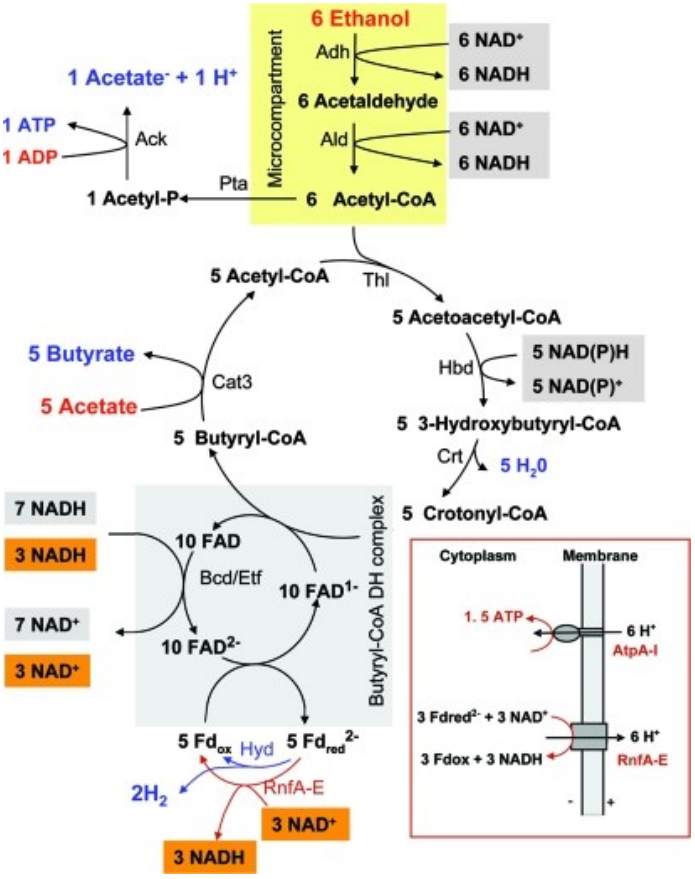

Figure 9 Ethanol-acetate fermentation of *C. kluyveri,* from Seedorf *et al.*, 2008 [64].

I.1.2.e Propionate pathway

The propionate pathway was investigated in some bacteria because some studies revealed the positive effect of short-chain fatty acids (SCFAs: acetate, propionate and butyrate) on human health. SCFAs produced by gut microbiome are absorbed by the host but also seem involved to counteract colorectal cancer and inflammation (butyrate) [65]. The potential propionate effects imply anti-lipogenic, cholesterol-lowering, anti-inflammatory and anti-carcinogenic action [66][67]. The propionate was investigated and three different metabolic pathways were discovered: the succinate, the acrylate and the propanediol pathway, Figure 10. *Clostridium propionic* and *Clostridium mayombei* was isolated from black mud of the San Francisco Bay (USA) [68] and the gut of African soil-feeding termite, respectively [69]. These bacteria were characterized to partially ferment substrates to propionate. *C. mayombei* is also a homoacetogenic bacteria, this bacteria uses the succinate pathway to produce propionate. Whereas *C. propionicum*, a protein/peptide-fermenting bacterium [70], utilizes the acrylate pathway for propionate production.

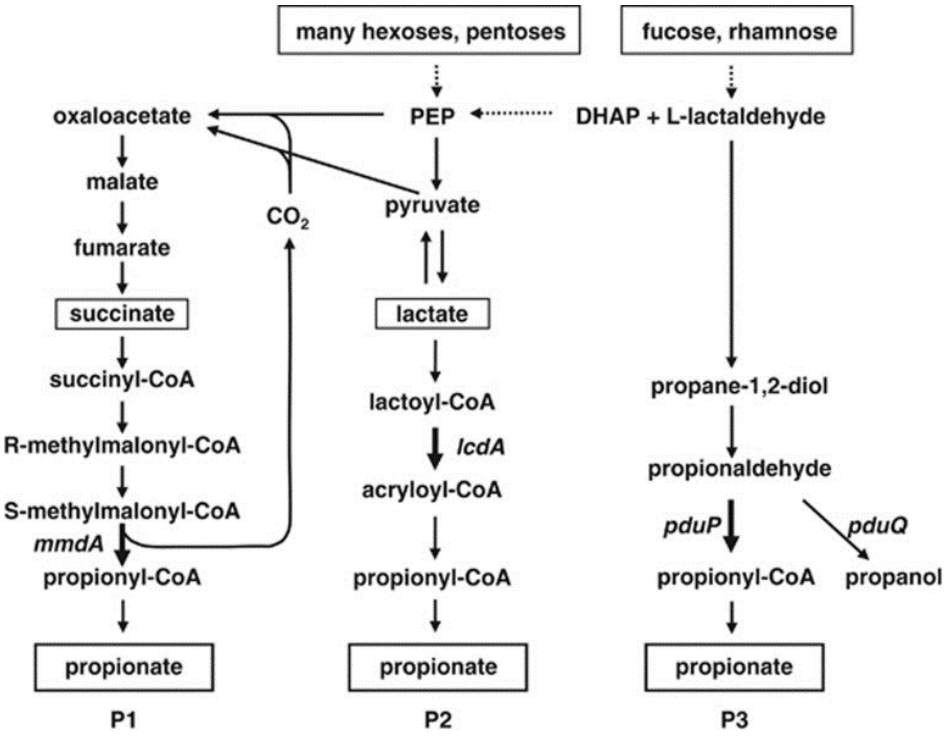

Figure 10 Main pathways for propionate formation in bacteria. (P1), Succinate pathway; (P2), acrylate pathway; (P3), propanediol pathway from Reichardt *et al.*, 2014 [71].

I.1.2.f Amino acids fermentation

Some *Clostridium* can use amino acids, or chains of amino acids (peptides, proteins) as substrates. These substrates are present in large amount in the intestine, faeces. Proteins are degraded to smaller peptides and free amino acids by proteolytic and peptolytic bacteria (including some *Clostridium* [72][73][74][75]). Then, amino acids are assimilated by these bacteria. Bacteria developed enzymes and metabolic pathways to utilize these substrates as energy, carbon and/or nitrogen sources [76], Figure 11.

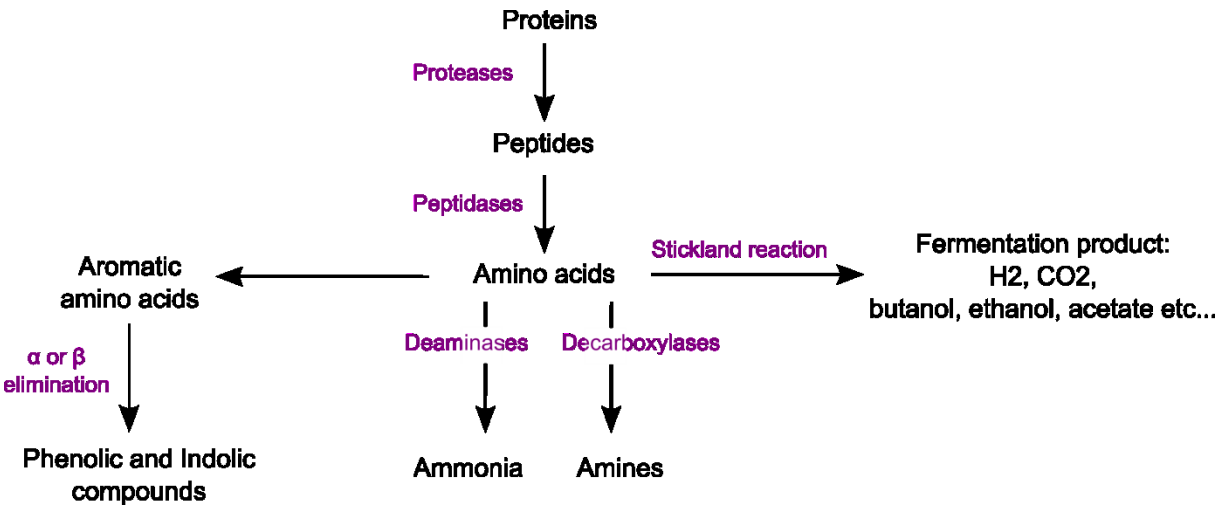

Figure 11 Amino acid utilization by Clostridia.

These metabolic pathways were characterized for some Clostridia. Ammonia or amine are respectively produced from the deamination [77] and decarboxylation [78] of amino acids. In addition, aromatic amino acids are modified into phenolic and indole compounds [79].

But the fermentation of amino acids occurs in Clostridia *via* the Stickland reaction. In this reaction, one amino acid is oxidized using another amino acid as electron acceptor [80], Figure 12. The main product of this fermentation is acetate, but one of the intermediate is pyruvate, an important metabolite in the central carbon metabolism. Amino acids can react as electron donor, electron acceptor or both [81]. Amino acid modified *via* the Stickland reaction can be used as carbon and energy source by the *Clostridium spp*, for instance *C. difficile* or *C. sticklandii* [82][83].

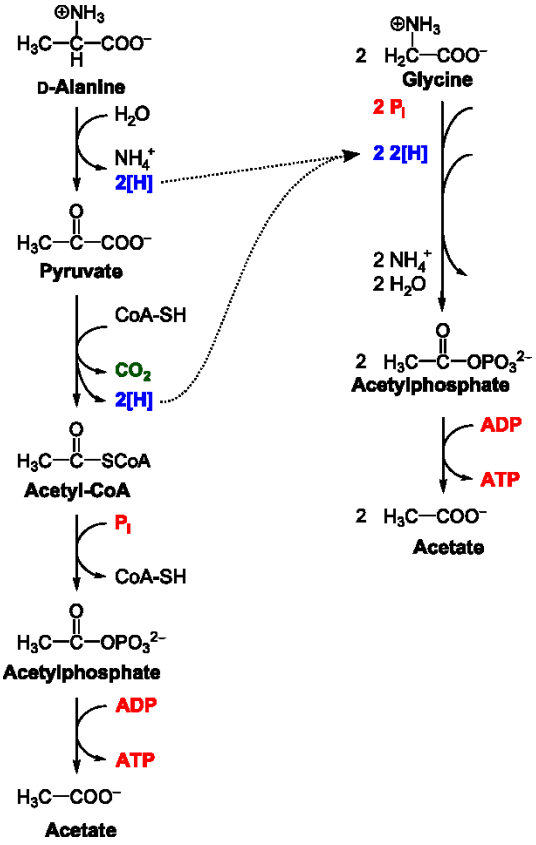

Figure 12 Example of the Stickland reaction with alanine as electron donor and glycine as electron acceptor, image modified from: https://en.wikipedia.org/wiki/Stickland_fermentation.

I.1.2.g Nucleic acid fermentation

Purinolytic or pyrimidinolytic Clostridia can metabolize purines and pyrimidines. These compounds are heterocyclic aromatic organic compounds. Noticeable purine and pyrimidine are nucleic acids which compose DNA and RNA. Purines nucleobases are adenine and guanine and pyrimidine nucleobase are cytosine, thymine and uracil. Other noticeable purine: xanthine, uric acid, and for pyrimidine: uridine triphosphaye.

C. cylindrosporum, *C. purinilyticum* and *C. acidurici* are known purinolytic Clostridia [84]. *C. acidurici* was characterized to use purine (except adenine) as carbon, nitrogen and energy source [85]. Moreover glucose and sugar cannot be used as sole carbon source by this strain. Main products of this fermentation are acetate, CO_2 and NH_3. The fermentation occurs after the purine breakdown with several enzymes, glycine is the intermediate produced by the fermentation. Nevertheless the metabolic pathway is not completely resolved [85].

The pyrimidinolytic activity of some Clostridia has been studied before, but this activity was not reported since 1980 in the general metabolism of *Clostridium* [86][87], casting doubt on this property.

I.1.2.h Diversity of fermentative pathways

Often, Clostridiales can possess several fermentative pathways. For instance *C. acidurici* is both, purinolytic and homoacetogenic [85]. Moreover most of *Clostridium* can at least metabolize hexoses with the glycolysis (or Embden-Meyerhof-Parnas pathway) like glucose as a carbon sources, other pathways are specific to family or strain. The glycolysis leads to pyruvate which is mainly metabolized to ethanol or acetate. All this diversity of substrates generates a variety of products from acids (acetate, formate, lactate, propionate and butyrate), gas (H_2, CO_2) and alcohols (ethanol, butanol, isopropanol, hexanol etc.).

I.1.2.i Energy metabolism

The variety of fermentation among Clostridia leads to a diversity of energy metabolisms. Generally, the transport and fermentation of substrate consume and produce energetic cofactors such as ATP, NADH and NADPH. An interesting alternative energy metabolism is shared between some Clostridia. This system is based on an electron transfer agent: the ferredoxin protein composed of iron and sulfur atoms. The reduced ferredoxin plays an important role to transfer an electron to NAD^+ or $NADP^+$ with oxidoreductases (NADH or Rnf type). It was also demonstrated that the Na^+ or H^+ translocation via the Rnf system and ferredoxin reduce NAD^+. This system also generates NADH and the Na^+/H^+ gradient created between both sides of the membrane leads to the ATP synthesis via the F1F0 ATPase and the Rnf system, Figure 13.

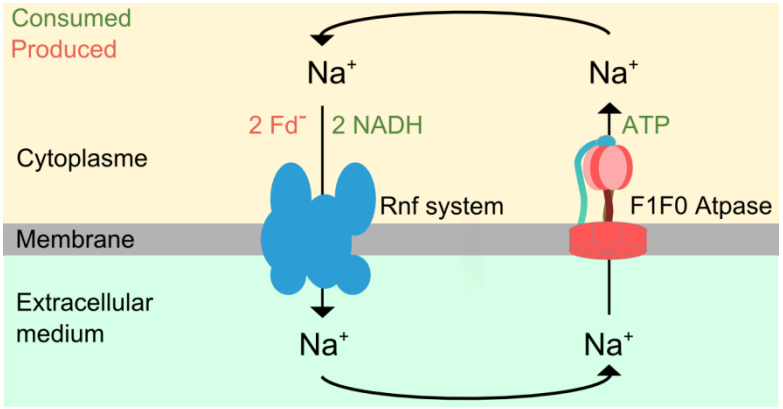

Figure 13 ATP synthesis and NADH regeneration with Na^+/H^+ gradient with the help of the Rnf system and the F1F0 ATPase.

This cycle was proposed or investigated in *C. kluyveri*, *C. sticklandii*, *C. ljungdahlii* or *C. phytofermentans* [64][88][89][90]. The energy metabolism of *Clostridium* is discussed later on the section related to the sugar transport of *Clostridium phytofermentans*.

I.1.3 Phylogeny

I.1.3.a History

The diversity of strains named *Clostridium* was revealed in a phylogenetic study of Clostridia in 1975 using 23S rRNA:DNA hybridization [91], the diversity increased with new discovered species and added with more or less strong scientific evidence to this genus. In 1994 [53], it was proposed to create 19 clusters inside the *Clostridium* genus. This phylogenetic tree based on 16S rRNA (Figure 2 of Collins *et al.*, 1994 [53]) conclude to assign cluster I species only to the genus *Clostridium* [53][54]. It revealed that several species were misclassified and named *Clostridium* whereas they belong to another class or family [53], concluding that new family or genus should be created for these misclassified species. In 1999, It was not possible to reclassify these species, as long as a taxons with these phenotypic properties are not available for these misclassified species [92].

I.1.3.b Reorganization of the phylogeny

The increasing number of new species identified increased the number of various species assigned in the *Clostridium* genus. In 2002 this variability is revealed with the presence of gram-negatives, non-spore formers, cocci, and non-anaerobes' in *Clostridium* taxonomy[93]. In 2009 the specificity of cluster I species was reinforced with genomic and protein sequence comparison (phylogenomic) [94]. An alternative taxonomy was proposed based on last phylogenetic trees and previous proposal [95] in 2009. The new taxonomy took place in 2013 in NCBI database, then some *Clostridium* species were separated between different families but still belong to the same order: Clostridiales[54]. For instance *Ruminiclostridium cellulolyticum* (Formerly *Clostridium cellulolyticum*) or *Peptoclostridium difficile* (formerly *Clostridium difficile*) belongs now to families *Ruminococcaceae* and *Peptostreptococcaceae* respectively, Figure 14.

This new taxonomy was in agreement with the actual phylogenic tree, and close to clusters demonstrated 20 years before by Collins *et al.*, [53]. The renaming of species raised an issue in the Clostridium community. During the Clostridium XIV conference (2016, Hanover, USA), several researchers have reminded to Yutin and Galperin their difficulties to deal with these modifications in databases. Modifying names can create difficulties for people to have access to previous articles if they are using the new name. Fortunately, databases are still using both names and for instance, in the KEGG database, typing "*Clostridium phytofermentans*" leads to *Lachnoclostridium phytofermentans* directly and the old name is written in parenthesis.

Unfortunately no advices were given and researchers are doing different name typing in articles: they are using names, the old one or the new one. Adding confusion and misunderstanding for newcomers in the community and difficulties for the bibliography. I propose to add, at least, both names in the abstract, helping the person who his looking for a strain, regardless if the old or new name is written. Observation: Both names (*Clostridium phytofermentans* and *Lachnoclostridium phytofermentans*) can be used in databases but an issue is possible for the research of articles related to this strain. For instance in pubmed, writing "*Clostridium phytofermentans*" return 39 results compare to three for "*Lachnoclostridium phytofermentans*". Moreover one of the three article [96] is not indexed when "*Clostridium phytofermentans*" is written [97], it reveals the problem of renaming strains, the bibliography becomes more complicated. But writing both names in the abstract allow to be indexed in both "search" in pubmed, it was tested with our last article [98], which is returned after writing either "*Clostridium phytofermentans*" or "*Lachnoclostridium phytofermentans*".

Focusing on the phylogeny of *C. phytofermentans*, this strain was placed in the cluster XIVa in the phylogenic tree of Clostridia. This cluster was renamed and now this strain belongs to the same order as other famous *Clostridium* (*C. acetobutylicum*, *C. difficile*, *C. perfringens*, *C. cellulolyticum*): Clostridiales. From the publication of Yutin *et al.*, 2013 [54], *C. phytofermentans* is in the new family Lachnospiraceae, genus *Lachnoclostridium* and the strain is named *Lachnoclostridium phytofermentans* (nevertheless, as an habit, I decided to use *C. phytoferementans* for the thesis manuscript). Surprisingly, *C. phytofermentans* is closely related to bacteria from various environment: human gut (most abundant firmicutes belongs to the cluster XIVa [99]), rice paddy soils, earth worm intestines etc.

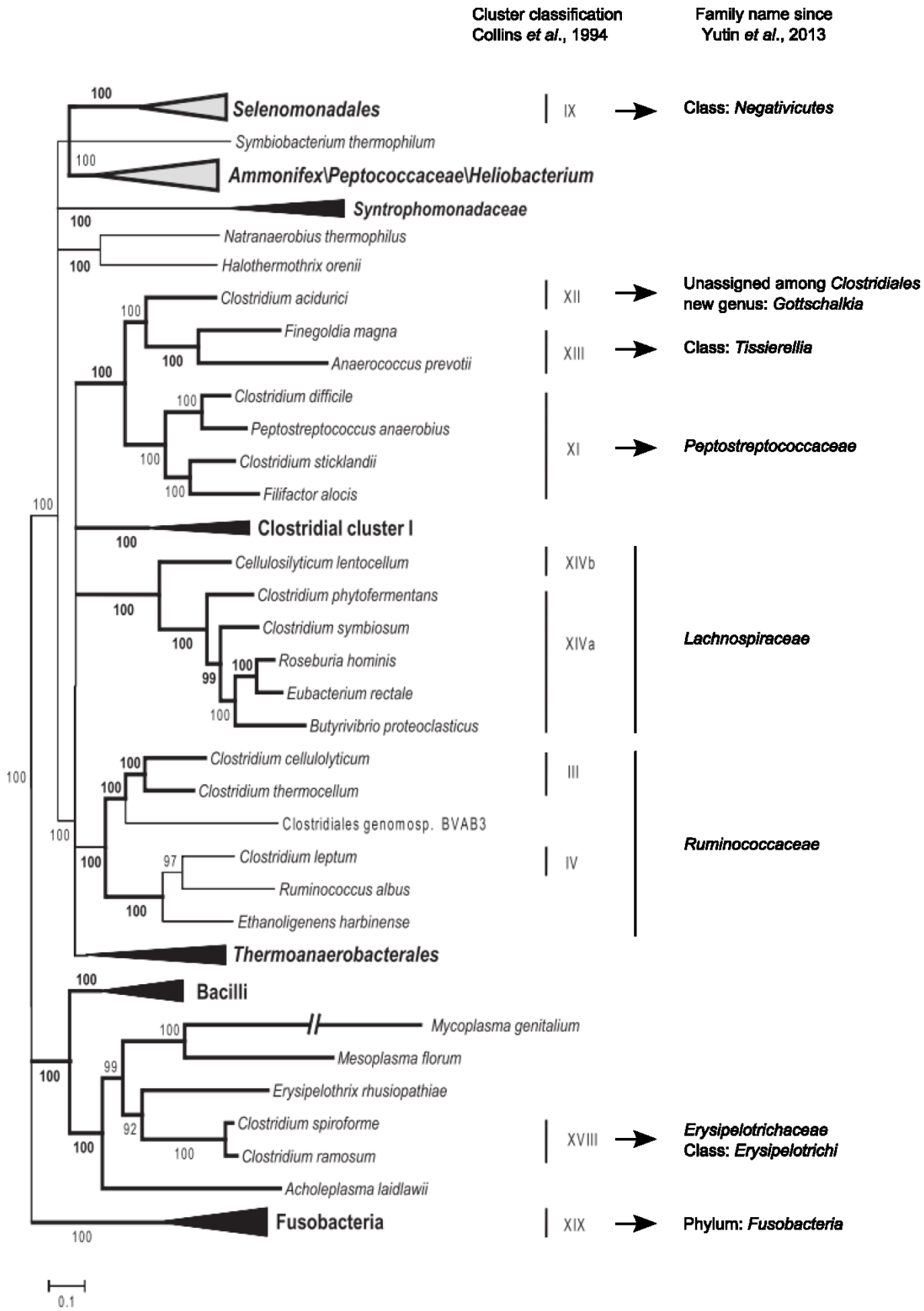

Figure 14 A ribosomal proteins-based phylogenetic tree of the *Firmicutes*. Modification of the Fig. 1 from *Yutin et al., 2013* [54]to assign *Clostridium* clusters to Phylum, Class, Order or Family.

I.1.4 Pathogenic and intestinal *Clostridium*

Pathogenic and Industrial Clostridia were first described more than a century ago. In the taxonomy previously presented, all pathogenic and industrial strains belong to the class Clostridia and the order Clostridiales. In addition, the phylogenic tree of Clostridiales shows that pathogenic and non-pathogenic Clostridiales are mixed in different families and genera; the taxonomy does not separate pathogens and non-pathogens. For instance, the pathogenic *C. difficile* and not pathogenic *C. sticklandii* (studied for its original fermentation of amino acids) belong to the family *Peptostreptococcaceae* (previously named *Clostridium* cluster XI); the not pathogenic *C. acetobutylicum* and the pathogenic *C. perfringens* belong to the *Clostridium* family (cluster I). Although these species are phylogenetically close, conferences about Clostridia are mainly separated between pathogenic Clostridia (ClostPath conferences) and Clostridia with industrial potential (*Clostridium* conferences). This concise introduction on pathogens/industrial Clostridia is interesting for at least one reason: (genetic) tools and methods for pathogenic Clostridia are often transposable between all Clostridia. For instance, plasmids with the origin of replication pIP404 replicate in *C. perfringens*, *C. ljungdahlii* and *C. cellulolyticum* [100][101][102]; or the origin of replication pCD6 works in *C. difficile, C. beijerinckii* and *C. pasteuranium* [103][104].

The main issue of pathogenic Clostridia is the resistance to chemicals and temperature, mainly as spore. Chemicals and temperature are used to disinfect in hospitals, which have a small effect on Clostridia, it explains part of the problem of Clostridia for nosocomial disease. Patient are mainly affected via spores [105]. The danger of pathogenic Clostridia is due to their toxins which have different effects, depending on the *Clostridium*. Moreover in a same species, strains can be more or less toxic; if toxin genes are not present, the strain does not lead to symptoms. Main pathogenic Clostridia are presented below.

C. difficile was firstly named *Bacillus difficilis* because of difficulties to isolate the strain from new-born infants [106]. This strain is an important nosocomial pathogenic bacterium [107] which colonize the gut. Antibiotic treatment increase the risk of *Clostridium difficile* infection (CDI) by killing lot of bacteria of the microbiome, allowing *C. difficile* to colonize new empty spaces, Figure 15. The colonization of the gut by this bacterium can lead to severe disease syndromes, such as acute diarrhea, abdominal pain, fever, no symptoms even the death. In other cases no symptoms are detected if the strain does not possess toxin genes. The pathogenicity is due to the production of toxins by *C. difficile*, Toxin A, B and/or binary toxin (TcdA, TcdB and

CDT respectively). The two first toxins disrupt the cytoskeleton of the cell host and lead to cell death and an inflammatory response [108].

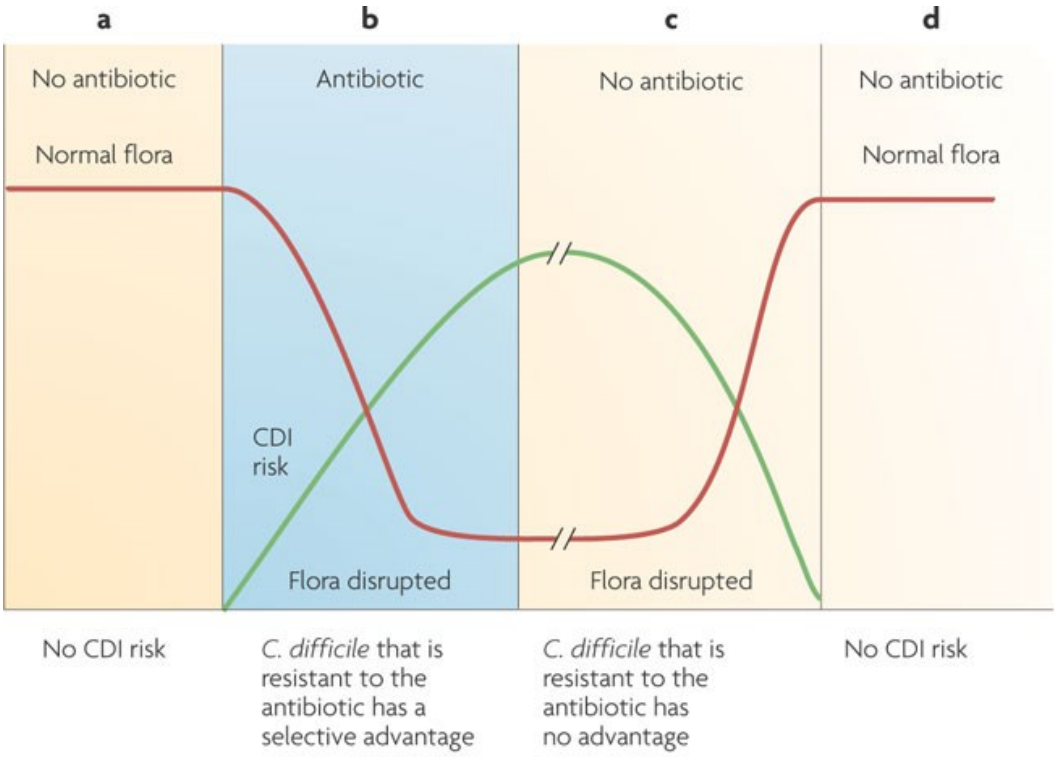

Figure 15 The effect of antibiotics on the normal gut flora and the risk of Clostridium difficile infection (CDI), from Rupnik *et al.*, 2009 [107].

TcdA and TcdB belong to the large clostridial toxins family, (LCT) produced by *C. perfringens* or or *Clostridium noyvi*, whereas CDT is related to the C2 toxin of *C. botulinum*.

C. perfringens can be found in soil, gut, food and other environments. This bacteria possesses more than 16 different toxins [109], leading to several widespread diseases in human and animals: diarrhea, gas gangrene or necrotic enteritis. It is the main pathogenic Clostridia related to gas gangrene and the toxin of *C. perfringens* is produced during the sporulation whereas vegetative cells are not toxic [110].

C. botulinum and *C. tetani* toxins are neurotoxic. Main botulism intoxication occurs with food contaminated with *C. botulinum* toxins, toxins are produced during the late-logarithmic growth of the bacterium [111]. To avoid contamination and to kill *C. botulinum* spores, the agro-food industry modified heat processes and food conditioning. Although the botulinum toxin (BTX) is one of the most powerful biological toxin, it is applied for diverse medical applications, for instance: strabismus, dystonias, migraine, allergic rhinitis, etc. [112][113]. A cosmetic application for facial wrinkle is also well publicized.

After the BTX, the toxin from *C. tetani* is the second most powerful toxin, with a lethal dose of ~1ng/kg [114]. Spore of *C. tetani* are widely present in diverse environment (soil, water, etc.). The infection occurs when spore end up in a wound, a favorable environment for the germination of the bacteria with no oxygen. During the growth of *C. tetani* colonies, toxins are released. Toxins enter in the blood flow and reach the central nervous system, developing continuous muscle contractions which can lead to the death.

Clostridia species are not only related to pathogens, these bacteria are also essential for the homeostasis of the gut microflora[115]. Clostridia comprise between 20 to 40% of gut microbiome and participate to the maintenance of gut function by interacting between intestinal cells and other gut bacteria [116]. But several different strains of these pathogenic Clostridia exist with different degree of pathogenicity and some non-pathogenic strain of these species are present in guts [117].

I.1.5 Industrial *Clostridium*

I.1.5.a Solventogenic strains

The first industrial Clostridia was *C. acetobutylicum*. Properties of acetone, butanol and ethanol production by this bacterium were revealed by Chaim Weizmann in UK in 1915. The produced acetone was converted to cordite, an explosive. This process was mainly used by UK and Canada during the First World War, and used by several countries during the Second World War to bypass the importation of this compound from Germany. Chair Weizmann refused personal honors and rewards for this discovery but he pushed to establish a home for the jewish people after the Second World War. This place was Palestine and Israel was created with Chaim Weizmann as the first President of the new state [118]. Nowadays, acetone and butanol mainly come from the petrochemical industry [119], but were still produced from the fermentation of starch via *C. acetobutylicum* in Russia or South Africa until 1980's. The second important solventogenic Clostridia is *C. beijerinckii* [120], an interesting feature of the strain *C. beijerinckii* 6423 is the production of isopropanol from the intracellular acetone, mainly with the help of the secondary alcohol dehydrogenase [121]. Around 1950's, acetone-butanol-ethanol (ABE) fermentation supplied 65% of the butanol worldwide. This is a bulk chemical, this compound and derivatives are involved in diverse industries: fuel, perfume, pharmaceutical, plastic, detergents, textiles, resins etc. [122].

C. acetobutylicum metabolizes different products depending on the phase of the culture. During the initial growth acetate, H_2, $C0_2$ and butyrate are produced (acidogenic phase); whereas during the stationary growth solvent are produced from acetyl-CoA: acetone, ethanol, butanol (solventogenic phase) [2][123]. This solventogenic phase starts with low pH and temperature, stress, beginning of sporulation and these events can trigger each other; acids previously produced are re-assimilated [124].

I.1.5.b Syngas strains

Some Clostridia are able to indirectly ferment syngas to commodities. These homoacetogenic Clostridia utilize the Wood-Ljungdahl pathway (previously described) with CO, CO_2 and H_2 as energy and carbon source [125] to produce Acetyl-Coa. Acetyl-CoA can be converted into biomass; excreted as acetate which can be reduced to acetaldehyde and then ethanol. Although model organisms are *C. ljungdahlii* [126] and *C. autoethanogenum* [127], several other species exists, for instance: *C. carboxidivorans* [128], *C. coskatii*, *C. ragsdalei* [129]. Historically, *C. aceticum* and *C. thermoaceticum* [60] were first acetogenic Clostridia studied. *C. autothenogenum* and *C. ljungdahlii* have closed sequence genome. Although they are named differently, these two strains cannot be phylogenically distinguished into two different species [130].

The carbon fixation of CO/CO_2 via the Wood-Ljungdahl pathway of acetogenic organisms is proposed to be the metabolism at the origin of life [131][132]. Billions years before O_2 appearance, these organisms were probably first autotrophs, using CO, CO_2 and H_2 as carbon, energy source and electron acceptor [133], Figure 16. Although the Wood-Ljungdahl pathway seems a good candidate for bioenergetic pathway and carbon fixation at the origin of life, this view was challenged by Nitschke and Russel, 2013 [134]. This team proposes that another metabolic pathway existed before the Wood-Ljungdahl pathway. They claim that the last result on the geochemistry of the planet reveal a richer range of redox substrates than previously supposed. They proposed that the Wood-Ljungdahl pathway is derived from a previous biomass-accumulating metabolism, which would be a quinone-based electron transfer system.

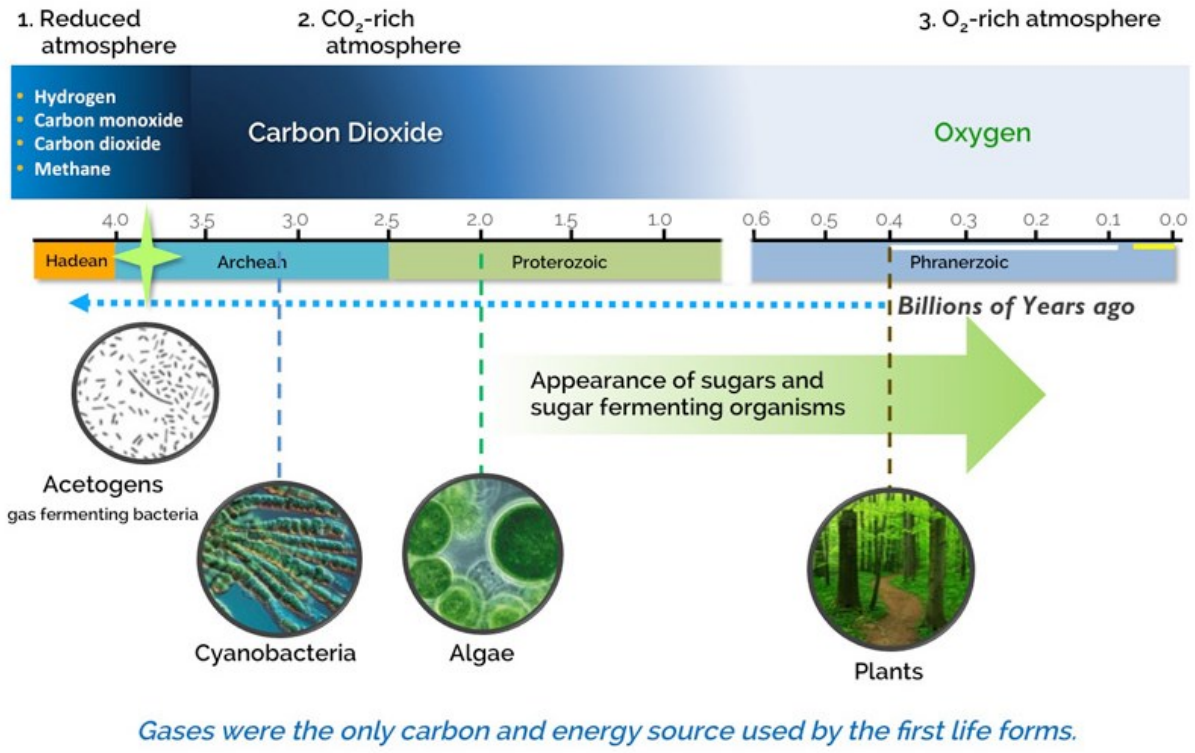

Figure 16 The Wood-Ljungdahl pathway probably at the origin of life, from: http://www.lanzatech.com/innovation/technical-overview/.

Several companies are investigating the syngas fermentation for the production of biofuels and chemicals. These companies are Synata Bio (previously Coskata Inc (*C. coskatii*)), INEOS Bio (*C. ljungdahlii*) and Lanzatech (*C. autothenogenum).* The most advance company (at least with recent release media) is Lanzatech which possesses several pilot and demonstration plants in several continents. Their technology can use syngas from various stocks: mainly from industrial waste gas (steel, PVC, ferroalloys) and woody biomass syngas. The first goal of Lanzatech was to recycle the CO and CO2 produced from steel plant to produce biofuel, such as ethanol. This company is also improving their strain to produce different valuable chemical compounds from syngas.

I.1.5.c Cellulolytic

The large scale production of biofuels and chemicals from the biomass is an ultimate goal of several scientific teams. In this part I will present only cellulolytic Clostridia investigated to achieve this objective. Nevertheless other interesting cellulolytic living organisms are investigated for this purpose. These organisms are fungi (in example: Trichoderma reesei [135]), protozoa [136] or bacteria. These organisms were isolated for instance from forest soil [137] or insect gut [138]. Some of these bacteria cannot be cultivated

(yet) and metagenomics analysis were applied to discover new very efficient cellulases and hemicellulases from gut and rumen of animal which mainly eat plants [139].

Three model cellulolytic Clostridia are *C. cellulolyticum*, *C. thermocellum and C. phytofermentans*. Each of these bacteria possess particular features described below. But other scientific groups try to add cellulolytic activities to non-cellulolytic Clostria which are already tested at the industrial scale, such as *C. acetobutylicum*. Although *C. acetobutylicum* possesses several cellulases, this bacteria is not able to grow on cellulose [140]. To solve this issue *C. acetobutylicum* was engineered to deconstruct and ferment the biomass to solvents with the help of a designed mini-cellulosome [141]. This strategy was not successful yet, even at the lab scale for several reasons: the cellulase expression of was not high enough, toxicity of cellulases, secretion of the heterologous proteins outside the cell and I also think that the transport of oligosaccharides from cellulose degradation need specific transporters.

The goal is to use cellulolytic Clostridia to produce biofuel and chemicals from biomass in one batch with (mainly) one bacterium without any important pretreatment. This technology is named consolidated bioprocessing (CBP) and is presented in another chapter.

C. thermocellum (or *Ruminiclostridium thermocellum*) [142][143] and *C. cellulolyticum* (or *Ruminiclostridium cellulolyticum*) [144] were isolated from sewage soil and compost respectively. These bacteria degrade the lignocellulosic plant biomass with an external complex which harbors cellulases, this system is called cellulosomes. The cellulosomes is made of several structural subunits called scaffoldin, linked together with dockerin and cohesin modules [145]. Thus a dockerin-containing enzyme is linked to the extremity of the cellulosomes, Figure 17, [146]. The cellulolytic enzyme attached to the end of a dockerin is a carbohydrate active enzyme (CAZyme). The proximity of these CAZyme plays a synergic role for the degradation of cellulose and hemicellulose of plant biomass. Moreover, the cellulosomes is not necessary attached to the bacterial cell wall, it can be freely released [147].

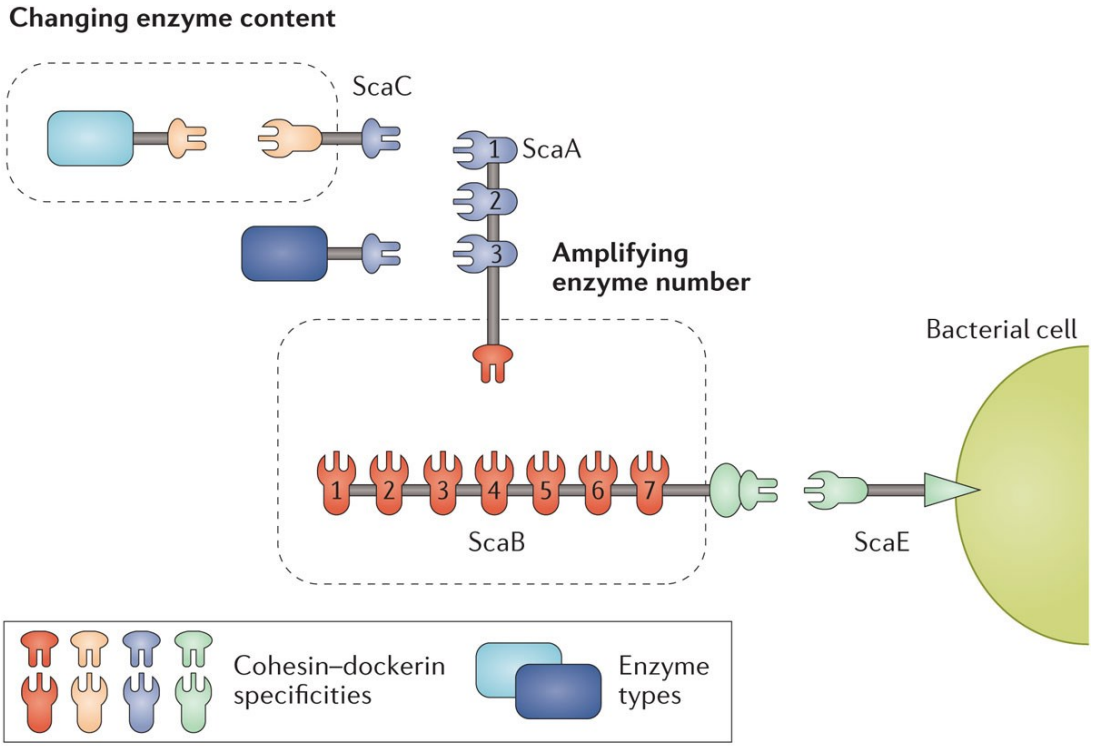

Figure 17 Example of the cellulosome of *Ruminococcus flavefaciens*, from Artzi *et al.*, 2017 [148]. The scaffoldin E (ScaE) is attached to the bacterial cell wall. The ScaE cohesin is linked to the dockerin of ScaB which possesses various cohesin, and then the cohesin can specifically attach a ScaA dockerin. ScaA dockerins can adhere to diverse dockerin-containing enzyme or another scaffoldin (ScaC) which can itself binds to dockerin-containing enzyme.

Usually, the first scaffoldin contains a carbohydrate-binding domain (CBM). CBMs of the cellulosome attach the bacteria to cellulose or hemicellulose substrates [149], moreover different CBMs exist and are carbohydrate specific. The role of the CBM in the scaffolfin arrangement and the carbohydrate degradation is also important [148].

C. thermocellum was first isolated in 1926 by Viljoen *et al.* [142] and described 30 years by McBee *et al.*, [143]. It was demonstrated that *C. thermocellum* is a thermophile bacteria with an optimal pH conditions between 6.7 and 7.0 and an optimal temperature of 55°C. This bacterium is able to grow on various substrates: glucose, cellobiose cellulose and hemicellulose. Fermentation products on these carbon sources are succinate, lactate, CO_2, ethanol etc. [150]. Nevertheless, *C. thermocellum* naturally produces a low quantity of ethanol which does not reach more than 3 g l^{-1}. This cellulolytic *Clostridium* is probably the most studied among cellulolytic Clostridia and numerous studies were released showing the genomic, transcriptomic, proteomic and genetic engineering tool adapted to this strain.

C. cellulolyticum was isolated from decayed grass[144]. Like *C. thermocellum* this bacteria possess a cellulosome [151] and fermentation products from carbohydrates are similar. Nevertheless, *C. cellulolyticum* is a mesophilic bacterium with an optimal growth between 32 and 35°C, it is the main difference with *C. thermocellum*. Although these two species possess cellulosomes, their cohesin and dockerin are not compatible [152]. Nevertheless, cellulosome was engineered and the compatibility of cellulosomal (but also not cellulosomal) enzymes between species was achieved reality [153][154]. This technique was named designer cellulosomes.

C. phytofermentans (or *Lachnoclostridium phytofermentans*) was isolated from forest soil [155]. This Clostridia is the organism studied in this thesis and a deeper presentation is developed in the next chapter.

Other cellulolytic Clostridia were described (but less studied): *C. cellulovorans* [156], *C. clariflavum* [157], *C. papyrosolvens* [158] or *C. cellulosi* [159] were isolated from various environment: wood fermenter, thermophilic methanogenic bioreactor, rumen or paper mill respectively. Cellulolytic Clostridia mainly belongs to the cluster III [53] and reassigned to the family Ruminococcaceae and genus *Ruminiclostridium*, except *C. phytofermentans* (cluster XIVa reassigned to the family Lachnospiraceae and genus *Lachnoclostridium*) [54].

I.1.5.d What are the desired fermentation products?

As previously presented, the ABE fermentation was developed to produce mainly acetone, a precursor of the cordite, an explosive [160] and butanol with *C. acetobutylicum*. This strain generally produces acetone, butanol and ethanol at a weight ratio near 3:6:1 from starch or molasse. Butanol is a solvent uses in different chemical processes and can also be used as a biofuel.

Like butanol, the ethanol is a potential biofuel to replace petroleum based fuels. The production of ethanol at large scale from renewable resources is a goal for many groups and industries [161]. An engineered *C. thermocellum* is able to produce 5.4 g l-1 of isobutanol from cellulose in 75 hours at 50°C [147].

Isopropanol can be produced from *C. beijerinckii* [162][163], or engineered Clostridia [164]. Such as butanol, isopropanol is already used as a solvent in several industries. It is also a fuel additive and cleaning agent.

Butyric acid is naturally produced and harvested from several Clostridia, for instance *C. butyricum* [165] or *C. tyrobutyricum* [166]. Butyric acid is used in various industries: perfumes,

food flavor and textile to improve the resistance to heat and sunlight. Derivative of butyric acid are also involved in the pharmaceutical industry for vasoconstrictor drugs, cancer, gastrointestinal diseases, antioxidants or anesthetics [167][168].

Hydrogen is an intermediary product of the fermentation of some Clostridia [169][170]. As a general fuel (automobile, thermal energy generation, power station), hydrogen does not produce any pollutant, instead water is released. Although hydrogen has a higher energy yield than petrochemical fuels some issues still exist with the production, the storage and the safety [171]. Hydrogen is also used in the industry for crude oil processing and nitrogen fertilizers.

Botulinum toxins (several different toxin exist) are one of the most powerful biological toxin [172]. Nevertheless the neurotoxin type A from *C. botulinum* is applied for famous aesthetic medicine, but also for strabismus, spasmodic torticollis, improvement of facial lines, severe hyperhidrosis, chronic migraine etc. [173]. In a production process, the toxin is harvested from *C. botulinum*. The bacteria growths in a first medium, thus a second medium is inoculated with the first culture, overexpressing the production of the toxin; this strategy is protected with a patent [174].

Collagenases from the pathogenic bacteria *Clostridium histolyticum* (gas gangrene) can be used against several diseases, for instance: Dupuytren's contracture [175], peyronie (sexual dysfunction) [176], wound healing [177], intervertebral disc herniation [178], etc. [179] but also in laboratory for the separation of cell culture [180]. Collagenase protein breaks peptide bonds in collagen without hurting cell membrane. These collagenases are mainly produced by recombinant *Escherichia coli* [181].

The production of the compound 1,4 butanediol by *C. acetobutylicum* was investigated [182]. This solvent belongs to a world market exceeding 1 million tons and is involved in the manufacturing of plastics, elastic fibers, polyurethanes etc. [183].

Lanzatech is investigating and developing the production of numerous chemicals, Figure 18, with the fermentation of syngas in their bug: *C. autoethanogenum* [184].

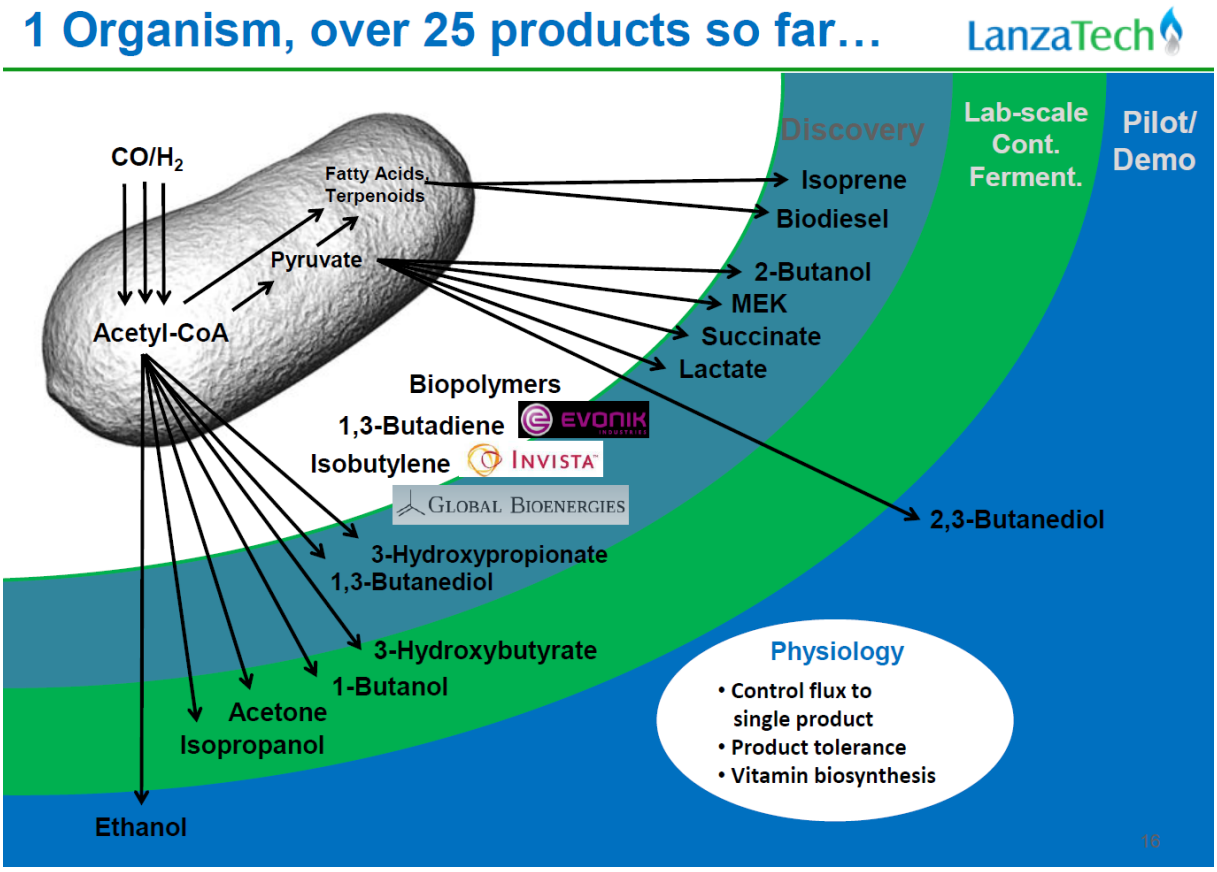

Figure 18 Chemical compounds investigated for a production in *C. autoethanogenum* after genetic manipulation of the strain, by Lanzatech [185].

The company is beginning the production of ethanol and 2,3-butanediol at the industrial scale. 2,3-butanediol can be converted to 1,3-butadiene in one chemical step [186]. Millions of tons of 1,3-butadiene are used each year to mainly produce rubber and resins (i.e. for automobile tire) [187] and the ethanol is reserved for the oil industry.

With recent progress in synthetic biology (genome engineering, metabolic pathways, modelling etc.) with development of new tools in Clostridia the question is more “from which substrate do we want to produce a chemical?” than “what do we want to produce ?”.

I.1.6 *Clostridium phytofermentans*

I.1.6.a History and characteristics of the strain

C. phytofermentans was reported in 2002 by Warnick *et al.* [155], the strain was isolated from the “bed of an intermittent stream in a forested site near Quabbin Reservoir in

Massachusetts (USA)" and several features of the strain were described. Although the gram staining is negative this bacterium belongs to the class Clostridia. The strain is mesophilic with an optimal temperature growth between 35-37°C, is motile (with one or two flagella) with a straight rod shape which turns into spherical terminal spore under stress or at stationary phase, Figure 19. *C. phytofermentans* can utilize various mono-, di- and poly- saccharides to grow, from different components of the plant biomass: cellulose, hemicellulose, pectin or storage polysaccharide (starch). In details, fermented carbohydrates by *C. phytofermentans* are: galactose, arabinoxylan, glucose, glucomannan, lactose, gentiobiose, maltose, galacturonic acid, pectin, ribose, starch, xylan, xylose, cellobiose, homogalacturonan, galactan, xyloglucan, cellulose and complex plant biomass residues (corn stover, sugarcane bagasse, switchgrass), whereas the growth is low on galactomannan, arabinan, mannose, rhamnose, arabinose and fructose. The strain does not grow on glycerol, pyruvate, sucrose, trehalose or tryptone as substrate. Moreover, pyrimidines, purines, vitamins and tryptone are essential for a normal growth. This is one the largest diversity of mono-, di- and poly- saccharides fermented by a bacterium. Moreover products of fermentation from cellulose are valuable compounds: ethanol (as primary product of the fermentation), acetate, CO_2 and H_2; lactate and formate are minor end products.

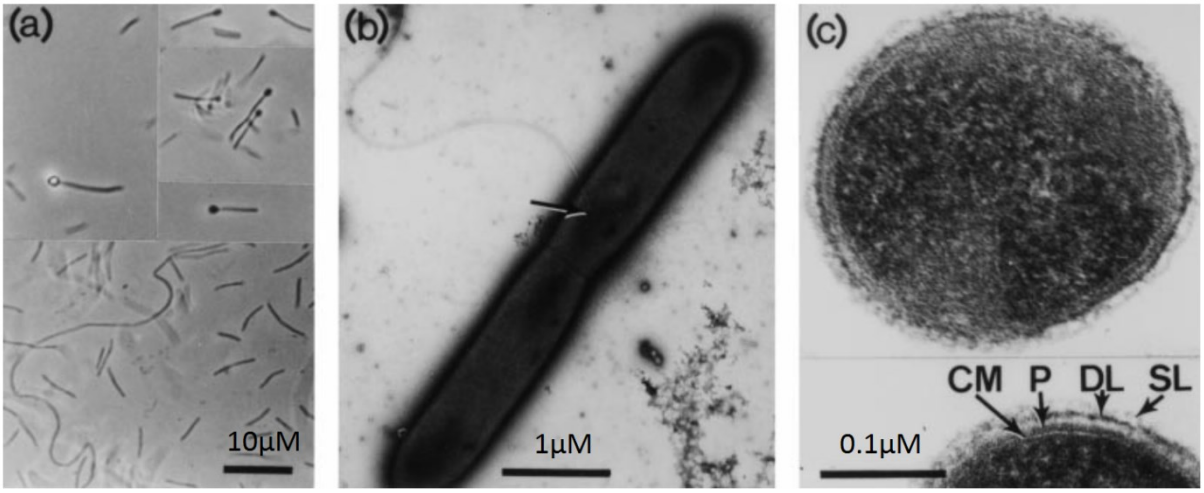

Figure 19 Morphology of the strain *C. phytofermentans* ISDg^T, from Warnick *et al.*, 2002 [155]. A one-day-old culture with a phase-contract micrographs. B two-day-old culture with transmission electron micrograph, the arrow indicates the attachment of single flagellum. C One week on plates of GS2 with transmission electron micrograph.

The genome *of C. phytofermentans* is publicly available (it was the first among cluster XIVa) and has one circular chromosome of 4,8 Mbp, a G+C content of 35% and 3926 CDS (our study reveals that more CDS are transcribed and translated (Appendix 2, [188]). A general overview of the strain was released in the article of petit *et al.*, 2015 [89] with the analysis of the genome and transcriptomes on several substrates.

The diversity of carbohydrates decomposed by C. phytofermentans comes from its variety of CAZYmes and sugar transporters. Among Clostridia, *C. phytofermentans* possess one of the largest number of CAZymes and ABC transporter with 174 genes and 148 genes respectively [89](Appendix 1, [51]). This ability to decompose and ferment the majority of the lignocellulose biomass in this manner is rare for a mesophilic bacterium. It makes *C. phytofermentans* an interesting chassis to study plant biomass decomposition, but also an interesting chassis to develop the strain for hypothetical industrial purposes.

Although *C. phytofermentans* does not possess any cellulosomes compare to other cellulolytic Clostridia (*C. cellulolyticum* or *C. thermocellum*) the bacterium attaches itself to plant substrates. The mechanism involved in this process is still unknown, but it is an advantage to the cell for degrading plant polysaccharides [189].

The energetic metabolism of *C. phytofermentans* is very well investigated in the supporting information 7 of [89], with interesting conjectures with the Rnf and F1F0 ATPase system to regenerate cofacteurs, the production of ethanol and the central place of the ferredoxin as important reducing agent in the general metabolism.

Even if *Clostridia* are obligate anaerobes, the tolerance for oxygen varies between *Clostridium*. For instance *Clostridium cellulolyticum* or *Clostridium difficile* have to be manipulated essentially under strict anoxic conditions, whereas *Clostridium phytofermentans* can survive at ambient atmosphere. *C. phytofermentans* can be spread on Petri dishes or diluted outside a glovebox; it is an important advantage for manipulating this strain. Moreover this strain is not pathogenic, thus manipulating *C. phytofermentans* does not require specific processes or authorizations.

I.1.6.b Previous important studies on C. phytofermentans

After the article of Warnick *et al.*, in 2002 to describe the strain, no important articles were released with a link to *C. phytofermentans* for 5 years. In 2007 an article compared six mesophilic Clostridia on their abilities to ferment cellulose, 5 of them are cellulolytic and the 6th is *C. acetobutylicum* as control. Among cellulolytic *Clostridium* tested (*C. populeti*, *C. cellulolyticum*, *C. celerecrescens*, *C. cellobioparum and C. phytofermentans*), *C. cellulolyticum* is the highest H_2 producer. Whereas *C. phytofermentans* ferments more efficiently avicel (microcrystalline cellulose) into ethanol than other strains [190].

Abilities of *C. phytofermentans* on plant biomass were used by several research groups for different purposes. Number of articles developed the characterization of CAZyme and enzymes of this strain. The characterization could be done *in vitro* or *in vivo*, by adding the gene to

another organism and improves its abilities to degrade some polysaccharides, Table 2. CAZymes where improved by modifying these genes by error prone PCR. Genes were inserted in other bacteria: *B. subtilis* [191] or *C. thermocellum* [192] and screened for best mutants, these articles called this method "directed evolution".

***C. phy* gene**	**CAZyme Classification and/or activity**	**Characterization: *in vitro* or expression in another organism**	**References**
cphy1219	xylose isomerase	in *Saccharomyces cerevisiae*	Brat *et al*., 2009 [193], 2013 [194]; Demeke *et al*., 2013 [195]
cphy0577	GH112 galactosyl-β1→3-N-acetyl-hexosamine phosphorylase	*in vitro*	Nakajima *et al*., 2009 [196]
cphy3030	GH112 galactosyl-β1→3-N-acetyl-hexosamine phosphorylase	*in vitro*	
cphy1920	GH112 galactosyl-β1→4-rhamnose phosphorylase	*in vitro*	
cphy3367	GH9 endo-β-1,4-glucanase	*in vitro*, in *C. phytofermentans*, in *C. cellulolyticum*	Tolonen *et al*., 2009 [197]; Zhang *et al*., 2010 [198]; Liao *et al*., 2011 [199]; Ravachol *et al*., 2015 [200]
cphy3368	GH48 cellobiohydrolase	*in vitro*	Zhang *et al*., 2010 [201]; Tolonen *et al*., 2014 (Appendix 1, [51])
cphy1163	GH5 endo-β-1,4-glucanase	*in vitro* and *C. thermocellum* (the enzyme was improved for thermostability)	Liu *et al*., 2010 [192]; Tolonen *et al*., 2014 (Appendix 1, [51])
cphy1019	GH65 3-O-α-D-glucopyranosyl-L-rhamnose phosphorylase	*in vitro*	Nihira *et al*., 2012(a) [202]
cphy1874	GH65 nigerose phosphorylase	*in vitro*	Nihira *et al*., 2012(b) [203]
cphy3160	GH2 acitvity on xyloglucan	*in vitro*	Tolonen *et al*., 2014 (Appendix 1, [51])
cphy2058	GH5 endo-β-1,4-glucanase	*in vitro*	
cphy3202	GH5 main acitvity on glucomannan and xyloglucan	*in vitro*	
cphy0624	GH10 main activity on xylan and arabinoxylan	*in vitro*	
cphy1510	GH10 main activity on xylan and arabinoxylan	*in vitro*	
cphy2108	GH10 main activity on xylan and arabinoxylan	*in vitro*	
cphy3010	GH10 main activity on xylan and arabinoxylan	*in vitro*	
cphy2105	GH11 main activity on xylan and arabinoxylan	*in vitro*	
cphy1640	GH12 main activity on xyloglucan	*in vitro*	
cphy2304	GH13 low activity on starch	*in vitro*	
cphy2344	GH13 activity on starch	*in vitro*	
cphy2350	GH13 activity on starch	*in vitro*	
cphy1071	GH26 main activity on galactomannan and glucomannan	*in vitro*	
cphy2128	GH26 main activity on galactomannan and glucomannan	*in vitro*	
cphy2276	GH26 main activity on galactomannan and glucomannan	*in vitro*	
cphy2567	GH28 low activity on rhamnogalacturonan I	*in vitro*	
cphy0218	GH31 low activity on cellulose	*in vitro*	
cphy1877	GH31 low activity on xylan	*in vitro*	
cphy1169	GH51 low activity on arabinan	*in vitro*	
cphy3586	GH53 main activity on galactan	*in vitro*	
cphy0776	GH88	*in vitro*	
cphy1929	GH94 low activity on cellulose	*in vitro*	
cphy1718	GH130 low activity on homogalacturonan	*in vitro*	
cphy1687	CE4 low activity on homogalacturonan	*in vitro*	

cphy1612	PL1, PL9 main activity on homogalacturonan	*in vitro*	
cphy1888	PL9 main activity on homogalacturonan	*in vitro*	
cphy2919	PL9 main activity on homogalacturonan	*in vitro*	
cphy3869	PL9 main activity on homogalacturonan	*in vitro*	
cphy1713	CBM32	*in vitro*	
cphy1799	GH18 exochitinase	*in vitro* and in *C. phytofermentans*	Tolonen et al., 2015 (Appendix 3, [204])
cphy1800	GH18 endochitinase	*in vitro* and in *C. phytofermentans*	Tolonen *et al*., 2015 (Appendix 3, [204])
cphy1178	aldehyde dehydrogenase	*in vitro*	Tuck *et al*., 2016 [205]
cphy0694	GH94 1,2-β-oligoglucan phosphorylase	*in vitro*	Nakajima *et al*.,2017 [97]

Table 2 List of enzymes characterized in *C. phytofermentans* in the literature.

General and original studies on *C. phytofermentans* improve the understanding of the metabolism, mainly related on plant biomass. Some studies were based on proteomics or transcriptomics [22] (Appendix 1, [51]) and revealed proteins/RNAs overproduced on several polysaccharides, given a map of important proteins and genes related to the biomass decomposition and its specific metabolism. These results lead to study overexpressed chitinases (Appendix 3, [204]) and CAZymes (Appendix 1, [51]), and ABC transporters characterized in this thesis. The biofilm formation was studied in *C. phytofermentans* [21], it can be an issue for the development of this strain for industrial purposes. This biofilm is an extracellular matrix of DNA and proteins, which can be cleaved with proteases and DNAses. An article predicts small RNAs in several Clostridia (*C. phytofermentans* included), but without assigning any specific role [206]. Recently the analysis of transcription start sites (TSS) of *C. phytofermentans* (Appendix 2, [188]) revealed at the base pair resolution: new non–coding RNA, transcription units, riboswhitches, sequences specific to the regulation depending on plant substrates. Some other features were discovered in *C. phytofermentans*. For instance when fucose and rhamnose are utilized as substrate by the strain, visible bacterial microcompartments are produced to ferment intermediate of the two substrates [207].

Some teams directly study *C. phytofermentans* in industrial environment (or mimicking it): with various polysaccharides composition close to harvested plant biomass or in bioreactor conditions. Best parameters to produce ethanol, or other compounds, from plant biomass are investigated. Different types of biomass, pretreatments and technologies are developed and tested. They do not go deeply into the general metabolism of the bacteria, the genomic and regulation of genes is not studied and *C. phytofermentans* wild-type is used, without any genetic improvement. Nevertheless, it is an important and crucial work to understand how *C. phytofermentans* behaves in these environments compare to laboratory, especially for the

development of this strain for industrial purposes. These studies are quickly presented in the following paragraph.

Interestingly, two closed protocols were proposed to characterize the quality of the biomass for biological conversion; the biomass is evaluated depending on the fermentation efficiency by *C. phytofermentans* [208][209]. A similar approach was used to characterize the biomass quality of waste water algal biomass, showing an ethanol production more important with this substrate than with glucose only, for the same weight per volume (2% (w/v)) [210]. *C. phytofermentans* was also used to test the digestibility and the fermentation of 25 different genotypes of the grass plant *Miscanthus* [211]. The rapid growth of *Miscanthus* is interesting for the biomass production and the cellulosic biofuel industry. A comparison of solid-state (without the use of free water) and submerged state fermentation of delignified switchgrass by *C. phytofermentans* reveals that the normal submerged-state method is more efficient. Nevertheless the solid-state fermentation is interesting by limiting environmental impact of cellulosic ethanol plant by using much less water [212]. *C. phytofermentans* was tested on CBP after the pretreatment of corn stover with ammonia fiber expansion (AFEX^{tm}). As a result they prove that this strain does not need nutrients supplementation and the ethanol production can reach 2.8 g l^{-1} in 125 ml serum vial, it is 72% of the yield by Simultaneous Saccharification & Co-Fermentation (SSCF) with *S. cerevisiae*. Compared to SSCF, CBP does not need to produce and purified enzymes separately, a costly process [213], Figure 20. In another study from the same team, high solid loading and AFEX^{tm} pretreatment were used; acetate produced during the fermentation seems to be the limiting compound of the sugar conversion [214].

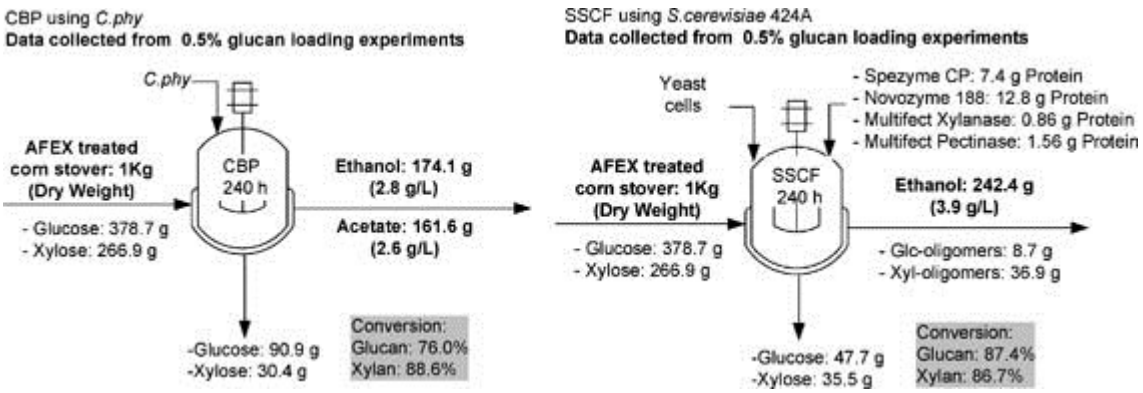

Figure 20 Mass balance comparison between CBP by *C. phytofermentans* and SSCF by *S. cerevisiae* 424A on AFEX-CS, from Jin *et al.*, 2011 [213].

Symbiotic co-cultures with C. phytofermentans and two types of yeasts (*S. cerevisiae* or *Candida molischiana*) were tested in a bioreactor [215]. The semi aerobic condition is controlled, thus yeast "protects" C. phytofermentans from the oxygen and in return bacteria liberate soluble sugars from the cellulose. The maintenance of this population shows the role of the symbiosis.

The production of ethanol reaches 22 g l^{-1} from the symbiotic population, whereas the production from mono culture of *C. phytofermentans* or *S. cerevisiae* is near 6 and 9 g l^{-1}, respectively.

The genetic manipulation of *C. phytoferm*entans was only achieved by our team, so far. The insertion of plasmids was used in several of our studies ([197] (Appendix 4, [216])(Appendix 3, [204]) and this thesis). Genomic intron insertion to disrupt targeted gene was developed using a plasmid harboring Ll.LtrB introns retargeted to the gene ([197] (Appendix 3, [204]) and this thesis).

Directed evolution is another method to modify the genome of a strain and it does not require the expertise of DNA transfer. Moreover it is a method without *a priori*, because modifications that must be applied to the strain to get new features are (mainly) unknown and unpredictable. This method was used by Mukherjee *et al.*, 2014 [217] to improve the ability of *C. phytofermentans* to grow on cellobiose, cellulose or xylan. This team did serial manual dilution with one of these three substrates as sole carbon source. Evolved strains were sequenced and few single nucleotide polymorphism (SNPs) were found in genes or intergenic genes related to the substrate used (CAZyme, kinase, transporters, etc.). We used a similar manual method to evolve *C. phytofermentans* to resist increasing ethanol concentration (Appendix 4, [216]). The mutant is able to grow with 7% of ethanol compare to 4% for the wild-type. After the sequencing of the evolved strain, only twelve mutations were revealed. These mutations are involved in the ethanol production, the energetic metabolism and the membrane synthesis; the ethanol production was restored by adding a plasmid expressing a heterologous alternative ethanol production pathway from *Zymomonas mobilis*. A recent article presents a directed evolution method using an automat. The automat prevents biofilm formation by transferring the continuous culture between two chambers each 12 hours, while the culture grows in a chamber the second chamber is washed; moreover the culture is regularly diluted with fresh medium. In this study the *C. phytofermentans* wild-type culture is improved by gradually increasing the concentration of ferulate, a phenolic inhibitor from the biomass degradation. Final adapted strains are able to grow at 6 g l^{-1} of ferulate, whereas the wild-type is not able to normally grow at 2 g l^{-1} of ferulate. These resistant mutants got less than twelve point mutations. These mutations are involved in gene regulation, metabolism and the cell surface (S-layer). Association of directed evolution and genetic manipulation can be interesting to study and improve Clostridia for different purposes.

In addition of scientific articles, several patents were released related to *C. phytofermentans* and/or the biomass conversion with this Clostridia. Most of patent inventors come from The

University of Massachusetts or Qteros, a company which had negociated an exclusive worldwide licence agreement with this university to exploit *C. phytofermentans* for industrial applications.

The first patent was intented in 2006 [218] by the team of the University of Massachusetts that discovered and described *C. phytofermentans* for the first time. They claim the use of *C. phytofermentans* to hydrolyze and ferment various plant biomasses into different types of alcohol (ethanol, isopropanol, n-propanol, n-butanol), this patent also protects co-fermentation of the biomass with *C. phytofermentans* and other bacteria or yeast.

Two years after the same team patented some gene sequence of metabolic pathways (rnf operon, NADH ferredoxin oxidoreductase, pyruvate ferredoxin oxidoreductase, acetaldehyde dehydrogenase, ethanol dehydrogenase and hydrogenase) which modulate the ethanol (or other fuel) production of C*. phytofermentans* from biomass. This patent protect these genes for an utilization in other organisms [219]. A similar patent was accepted to protect the utilization of some hydrolases, ABC transporter and transcriptional regulator of *C. phytofermentans* and related species in other microorganisms [220].

Reducing the sporulation of *C. phytofermentans* by modifying sporulation genes to improve fuel production was also patented [221]. The optimization of the medium by adding nitrogen source, fatty-acid compounds or pH modifier in a fed-batch strategy to improve the production of fermentive end product (such as ethanol) in *Clostridium* sp. is patented by Qteros [222]. This other patent from Qteros claims that using genetically modified *C. phytofermentans* to produce antioxidants (such as yeast catalases) to improve the fermentation of biomass to various alcohol is patented [223].

The same enterprise presented an interesting patent about evolved *C. phytofermentans* strains (NRRL Accession Numbers NRRL B-50436 (Q. 12) or NRRL B-50437 (Q. 13)) which produce up to 45 g l^{-1} of ethanol at 90% theorical yield from cellulosic biomass [224]. These strains come from the selection of random mutated strain, mutated with the addition of N-methyl- N'-nitro-N-nitroso-guanidine (NTG) to create a diverse mutant pool. It is a different directed evolution strategy.

Qteros have another patent showing that genetically modified *C. phytofermentans* mutants have 5 to 10 times more cellulase activity in presence of an inhibitor molecule (glucose) than the wild-type. Naturally, glucose is an inhibitor for production of cellulases, glucose concentration increases during the hydrolysis of cellulosic substrates repressing the cellulase expression, thus the growth and fermentation of *C. phytofermentans* decrease. Mutant Q.17, Q.18, Q.19 and Q20 have a normal cellulase activity in presence of glucose, improving the fermentation and growth

rate. This was achieved by the deregulation of cellulase synthesis by inactivating the carbon catabolite repression mechanism in these mutants; mutations involved in these interesting phenotypes are presented in the patent [225].

The last patent directly linked to *C. phytofermentans* [226] presents an ethanol production improvement with vitamin supplementation (mainly thiamine, but also nicotinic acid). Added thiamine improves the Acetyl-CoA synthesis and reduces the lactic acid production. It increases the ethanol pathway, Figure 21, which is downstream the Acetyl-CoA metabolic pathway. They claim that adding or modifying metabolic pathways to increase vitamin concentration in *C. phytofermentans* to enhance the ethanol production is related to their patent.

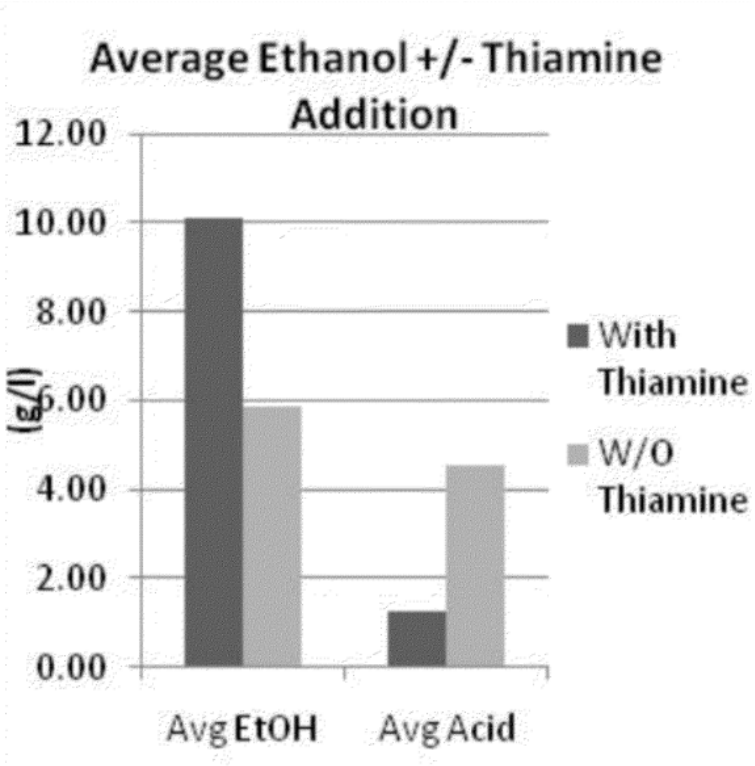

Figure 21 Ethanol and acid production in presence or absence of thiamine, from the patent US 20120107888 A1 [227].

The last patent is related to the article from Brat *et al.,* 2009 [193]. The xylose isomerase from *C. phytofermentans* was added to *S. cerevisiae*, bringing the capacity to ferment xylose to the yeast [228].

All these general and particular studies related to the metabolism, genetic tools, evolution, industrial conditions and patents increase developments and understandings of *C. phytofermentans* but it also expands knowledges related to other Clostridia.

I.2 Clostridium Genetics

As previously described, Clostridia are involved in many areas: medical, environmental or industrial. In order to improve knowledge and to develop applications related to Clostridia, genetic tools have to flourish. In this chapter, already available general genetic tools already available are presented, and perspectives of new tools are discussed. Moreover, lists of promoters, reporters, plasmid‘s origin of replications, genetic engineering tools used in Clostridia are presented. Some tools can probably be transposable between different Clostridia‘s species. Several reviews already did part of this work, I propose to add up to date results and tools from a broad range of Clostridia. Futhermore this chapter could serve as a basis for writing a review on this topic.

The important contribution from Pyne *et al.*, 2014 [229] to genetically engineer Clostridia is a must-to-read review in this field. Lists of different techniques of transformation (conjugation and electroporation), lists of reporters, promoters in various Clostridia are presented with some genetic engineering tools. Nevertheless, new tools were constantly developed since 2014 and this review is becoming out of date for some methods. Another interesting review presents recent genetic engineering tools, notably CRISPR-Cas9 utilization in Clostridia, however this review mainly focuses on advances related to *C. acetobutylicum* [230].

Basic tools, such as replicative plasmids, promoters, reporters, resistance cassette, transformation methods, and tools of genetic engineering were developed and worked in Clostridia. Nonetheless, these tools are far from easy to use, compared to *E. coli* or *B. subtilis* chassis. These technical difficulties prevent the use of Clostridia as chassis for industry, student lab work, do-it-yourself biology (biohacking) or medical; making these strains largely under exploited. For instance, no-easy-to use reporters are available, such as GFP in *E. coli*. This difficulty impacts the high-throughput screening of promoters for Clostridia. The number of promoters (inducible, repressible or with various strength) is not sufficient; this number must increase to bring Clostridia to the synthetic biology community. Genetic regulatory networks (for instance the repressilator [231]) could be designed in Clostridia, allowing the development of new tools for the Clostridia community in return. Transformation difficulties are becoming unraveled and several strains can now be electroporated. Nevertheless, some issues would not be solved; for instance, the anaerobic specificity of Clostridia is an important problem, working in these conditions remains difficult and required special equipment.

I.2.1 Transformation and Plasmids

The first step to precisely modify a bacterium is to transform it. The transformation allows the insertion of chains of nucleotides (DNA or RNA) inside the cell mainly *via* three techniques: conjugation, chemical transformation or electroporation. It can be used for instance to insert a plasmid to express a heterologous protein or to precisely modify the genome. The conjugation is more laborious and time-consuming than the electroporation, but this last method is not transposable for all Clostridia yet. The chemical transformation is not developed and used in Clostridia, perhaps the efficiency was not as efficient as the two other techniques.

Some Clostridia possess restriction-modification (RM) systems which cleave foreign DNA inserted into the cell. RM systems have been found in several strains, for instance: *C. acetobutylicum*, *C. pasteurianum*, *C. difficile* or *C. thermocellum* but some Clostridia, such as *C. phytofermentans* do not have these RM systems. To avoid this DNA degradation by RM systems, two options are available: the DNA can be protected by a methylation specific to the RM system of the strain [197][232][233], or RM systems can be inactivated [234][6].

RM systems and electroporation protocols in Clostridia are detailed in the review Pyne *et al.*, 2014 [229]. Heat-shock treatment to transform Clostridia was successfully investigated in *C. acetobutylicum* [235], but was not applied since 1984.

I.2.1.a Conjugation

Conjugation is a natural process which transfers genetic material from one bacterium to another. The link between two bacteria is a pilus produced by the donor bacterium *via* the DNA sequence of the fertility factor (or F-factor).

Historically the transformation of DNA in Clostridium was achieved with conjugation protocols, for instance from Streptococci to *C. acetobutylicum* [236]. A modified *Escherichia coli* was constructed to contain all elements for the conjugation and the shuttle plasmid must possess an essential conjugative element, the oriT (example in *C. perfringens* [237]). The conjugation is still used for Clostridia whithout established electroporation protocol, as *C. difficile* (depending on the strain) [238] or *C. phytofermentans* [197].

I.2.1.b Electroporation

Bacterial transformation *via* electroporation is a technique which forces DNA to enter into the cell with an electrical pulse. First protocols established to transform Clostridia

occurred in 1988 with *C. acetobutylicum* [239] and *C. perfringens* [240]. Protocols to transform Clostridia by electroporation are various and depend on species. To establish a new protocol for a *Clostridium*, several parameters can be modified to improve the transformation efficiency: the growth phase of the culture, parameters of the electrical field (voltage, faraday and ohm), quantity of DNA, size of the cuvette, time of recovery, resuspension buffer. A list of previous electroporation protocol was also done by Pyne *et al.*, 2014 [229][241]. We are currently testing and developing a protocol to electroporate *C. phytofermentans* with our collaborator Sanitha Mary (SRM University, India).

I.2.1.c Origin of replication

Plasmids are essential toold in genetic engineering to overexpress protein or to modify the genome, for instance. *Clostridium* species are different, thus, origin of replication, promoter and resistance cassette are not necessarily compatible between them. The list of origins of replication (ori) already working in Clostridium are reported in the Table 3, below.

ori's strain	Origin of replication	Plasmid	Maintenance	References
Bacillus subtilis	pIM13	pECII, pIMP, pSY6, pSOS94	*C. ljungdahlii, C. cellulolyticum, C. pasteurianum*	Jennert *et al*., 2000 [232]; Leang *et al*., 2013 [101]; Pyne *et al*., 2013 [241]
Enterococcus faecalis	pAMβ1	pQexp, pAT19	*C. ljungdahlii, C. phytofermentans, C. cellulolyticum, C. beijerinckii*	Trieu-Cuot *et* al., 1991 [242]; Jennert *et al*., 2000 [232]; Tolonen *et al*., 2009 [197]; Leang *et al*., 2013 [101]
Clostridium perfringens	pIP404	pCL2, pJIR705ai	*C. ljungdahlii, C. cellulolyticum, C. perfringens**	Jennert *et al*., 2000 [232]; Leang *et al*., 2013 [101]
C. botulinum	pBP1	pMTL82151	*C. ljungdahlii*	Leang *et al*., 2013 [101]
C. butyricum	pCB102	pMTL83151	*C. ljungdahlii, C, butyricum, C. cellulolyticum, C. pasteurianum*	Collins *et al*., 1985 [243]; Jennert *et al*., 2000 [232]; Pyne *et al*., 2013 [241]; Leang *et al*., 2013 [101]
Lactococcus lactis	pGK12	pWV01	*C. cellulolyticum*	Jennert *et al*., 2000 [232]
C. difficile	pCD6	pMTL84151	*C. difficile*, C. pasteurianum*	Purdue *et al*., 2002 [244]; Heap *et* al., 2009 [245]; Pyne *et al*., 2013 [241]
Caldicellulosiruptor bescii	pBAS2	pMU1117	*C. thermocellum* (hyperthermophilic plasmid)	Groom *et al*.,2016 [246]
Staphylococcus aureus	pUB110	pUB110	*C. thermocellum, C. acetobutylicum*	Lin *et al*., 1984 [235]; McKenzie *et al*., 1986 [247]; Olson *et al*., 2012 [248]; Lee *et al*., 2015 [249]
Lactococcus lactis	pWV01ts	pMTLts	*C. ljungdahlii* (temperature	Molitor *et al*., 2016 [250]

			sensitive plasmid)	

Table 3 List of origin of replication used in Clostridium. * indicates pathogenic Clostridia.

Because temperature sensitive plasmids in Clostridia were not available, a team computationally designed and tested a temperature sensitive plasmid: pMU102-M166A, [251]. Furthermore, in Clostridia an interesting strategy was developed by the group of Dr. Minton, who established a modular system for plasmids construction. This modularity inspired from synthetic biology tools, divide the plasmid into four parts: the Gram+ ori, the resistance cassette, the Gram- ori and the application. This modular system is now exploited by a spin-off company created by this group: CHAIN biotech. The number of parts is substantial but should be extended; a library of characterized promoters will be an important plus. The impact of this system on Clostridia community is important, numerous teams are using this system.

I.2.1.d Resistance cassette

Like origins of replication, resistance cassettes are essentials. Possessing several various cassettes allowing resistance to different antibiotics is important for the toolbox of Clostridia. Mainly two antibiotic resistance cassettes are used in Clostridia to maintain a plasmid: *ermB* and *cat* which respectively confer a resistance to erythromycin and chloramphenicol/thiamphenicol. Nevertheless, the concentration of utilization differs between species.

Antibiotic resistance can be very different between species, a list of working antibiotic resistance gene and related antibiotic concentration on several Clostridia was established, Table 4.

Antibiotic resistance gene	Antibiotic	Species	Concentration	References
ermB	erythromycin	*C. cellulolyticum*	20 µg ml^{-1}	Cui *et al.*, 2014 [252]
ermB	erythromycin	*C. acetobutylicum*	100 µg ml^{-1} (liquid), 40 µg ml^{-1} (plate)	Heap *et al.*, 2009 [245], Al-Hinai *et al.*, 2012 [253]
ermB	erythromycin	*C. difficile 630ΔErm, C. beijerinckii*	10 µg ml^{-1}	Heap *et al.*, 2009 [245],
ermB	lincomycin	*C. difficile R20291*	20 µg ml^{-1}	Heap *et al.*, 2009 [245]
ermB	erythromycin	*C. beijerinckii*	10 µg ml^{-1}	Heap *et al.*, 2009 [245]
ermB	erythromycin	*C. botulinum*	40 µg ml^{-1}	Heap *et al.*, 2009 [245]
ermB	erythromycin	*C. phytofermentans*	200 µg ml^{-1} (liquid), 40 µg ml^{-1} (plate)	Tolonen *et al.*, 2009 [197]
ermB	Clarithromycin	*C. ljungdahlii, C. pasteuranium*	4 µg ml^{-1}	Leang *et al.*, 2013 [101], Pyne*et al.*, 2013 [241]
catP	thiamphenicol	*C. acetobutylicum,*	15 µg ml^{-1}	Heap *et al.*, 2009 [245], Ehsaan *et al.*, [254]

		C. difficile R20291, C. difficile 630ΔErm, C. botulinum		
cat	thiamphenicol	*C. thermocellum*	6 to 48 µg ml^{-1} (depending on T°)	Tripathi *et al.,* [255],
cat	thiamphenicol	*C. cellulolyticum*	10 µg ml^{-1}	Blouzard *et al.*, 2010[256]
cat	thiamphenicol	*C. perfringens*	10 µg ml^{-1}	Han *et al.*, 2015 [257]
neo	neomycin	*C. thermocellum*	250 µg ml^{-1}	Olson *et al.*, 2010 [258]
aad9	spectinomycin	*C. beijerinckii*	750 µg ml^{-1}	Heap *et al.*, 2009 [245]
aad9	spectinomycin	*C. botulinum*	600 µg ml^{-1}	Heap *et al.*, 2009 [245]
tetA	tetracycline	*C. difficile R20291*	10 µg ml^{-1}	Heap *et al.*, 2009 [245]
tetM	tetracycline	*C. cellulolyticum*	5 µg ml^{-1}	Celik *et al.*, 2013 [259]
bcrA/B	bacitracin	*C. perfringens*	128 µg ml^{-1}	Han *et al.*, 2015 [257]

Table 4 Antibiotic resistance cassette and concentration for some Clostridia.

Several other antibiotic resistance genes were recently tested and used: *aad9*, *tet* (A and M) and *neo* which allow a resistance to spectinomycin, tetracycline and neomycin, respectively. A new antibiotic cassette, *bcrA/B* was predicted in *C. perfringens* to confer resistance to bacitracin [260]. This cassette was then isolated and tested on bacitracin sensitive *C. perfringens* strain [257]. This last gene could be tested in other Clostridia, increasing the number of resistance cassettes available.

Instead of using resistance gene/antibiotic system, the maintenance of a DNA sequence can be achieved by building a strain sensitive or auxotroph to a compound. For instance, auxotrophic strains for pyrimidine were built. Two genes of the pyrimidine pathway, *pyrE* (orotate phosphoribosyltransferase) or *pyrF* (Orotidine 5'-phosphate decarboxylase.), can be inactivated for this purpose. *ΔpyrE* or *ΔpyrF* strains can grow supplemented with uracil or by adding a plasmid harboring the missing gene [255]. The addition of 5-fluoroorotic acid (5-FOA) can be used as negative selection for *ΔpyrE* or *ΔpyrF* strains containing a plasmid harboring *pyrE* or *pyrF*, because the 5-FOA is metabolized to 5-flurouracila, a toxic compound, when the complete pyrimidine biosynthesis pathway is present.

This *pyrE*/*pyrF* system for positive and negative selection was also used for plasmid curing [252] or to improve genomic recombination [238]. A similar negative selection was established with the gene *codA* in *C. difficile* [261], *hpt* and *tdk* in *C. thermocellum* [262]. To maintain a plasmid with a *pyrE* or *pyrF* marker in a deficient strain, heterologous gene is preferred to avoid homologous recombination with the native gene.

I.2.1.e Promoters

To overexpress a protein or to finely tune the expression of a heterologous metabolic pathway, promoters are essential. Promoters can induce, repress the gene expression with a chemical compound/substrate; or the transcription can be quite constant, and strengths expression can be low to very high. The expression of the gene following the promoter can be precisely regulated playing with promoter's specificity. A wide library of various promoters could be useful to tune the transcription as desired.

Name	Origin of the promoter	Recipient	Characteristics	References
fdx	*C. phytofermentans, C. pasteurianum,, C. sporogenes*	*C. phytofermentans, C. cellulolyticum, C. difficile*	Strong expression	Tolonen *et al.*, 2009 [197]; Higashide *et al.*, 2011 [263]; Ng *et al.*, 2013 [238]
ptb	*C. beijerinckii*	*C. acetobutylicum, C. cellulolyticum*	Depend on the growth phase, high expression	Girbal *et al.*, 2003 [264]; Feustel *et al.*, 2004 [265]; Cui *et al.*, 2014 [252]
sol	*C. acetobutylicum*	*C. acetobutylicum*	Depend on the growth phase, but weak expression	Feustel *et al.*, 2004 [265]
adc	*C. acetobutylicum*	*C. acetobutylicum*	Early growth phase, high expressiion	Girbal *et al.*, 2003[264]; Feustel *et al.*, 2004 [265]
bdhA	*C. acetobutylicum*	*C. acetobutylicum*	Low expression	Feustel *et al.*, 2004 [265]
bdhB	*C. acetobutylicum*	*C. acetobutylicum*	high expression (close to *adc*)	Feustel *et al.*, 2004 [265]
thl	*C. acetobutylicum*	*C. acetobutylicum, C. cellulolyticum*	Depend on the growth phase, higher expression than *ptb*	Tummala *et al.*, 1999 [266]; Gaida *et al.*, 2016 [267]
eno	*C. thermocellum*	*C. thermocellum*	moderate expression	Olson *et al.*, 2015 [268]
Sig promoters	*C. difficile*	*C. difficile*	Depends on the sporulation phase	Pereira *et al.*, 2013 [39]
cbp_2	*C. thermocellum*	*C. thermocellum*	moderate expression	Tripathy *et al.*, 2010 [269]; Olson *et al.*, 2015 [268]
cpe	*C. perfringens*	*C. perfringens*	Expression increased at the beginning of the stationary phase	Melville *et al.*, 1994 [270]
txeR	*C. difficile,*	*C. difficile, C. perfringens*	Expression in stationary phase	Mani *et al.*, 2002 [271]
gapDH	*C. thermocellum*	*C. thermocellum*	moderate expression	Olson *et al.*, 2010 [258]; Olson *et al.*, 2015 [268]
815, 926, 2638, 2926	*C. thermocellum*	*C. thermocellum*	moderate expression	Olson *et al.*, 2015 [268]
thl-2lacO	*C. acetobutylicum*	*C. acetobutylicum*	*thl* promoter with lac operator to	Perret *et al.*, 2004 [272]

			repress expression in *E. coli*	
pagA	*Bacillus anthracis*	*C. phytofermentans*	Higher expression in late log phase	Leibig *et al.*, 2008 [273]; this thesis
fac: fusion of lacI (E. coli) and ptb	*E. coli* and *C. acetobutylicum*	*C. sporogenes* (high induction), *C. acetobutylicum* (low induction)	Induction with IPTG	Heap *et al.*, 2007 [274]
bgaR-P_{bgaL1}	*C. perfringens*	*C. perfringens, C. ljungdahlii*	Induction with lactose	Hartman *et al.*, 2010 [275]; Banerjee *et al.*, 2014 [276]
cipP	*C. cellulolyticum*	*C. cellulolyticum*	Induction with cellulose	Li *et al.*, 2014[32]
celL	*C. thermocellum*	*C. thermocellum*	Induction with laminaribiose (low induction)	Mearls *et al.*, 2015 [277]
cpr	*C. difficile*	*C. difficile*	Induction with nisin	*McBride et* al., 2011 [278]; Edwards *et al.*, 2015 [279]
xyl	*Staphylococcus xylosus*	*C. acetobutylicum*	Induction with xylose	Girbal *et al.*, 2003[264]
xylR-P_{xylB}	*C. difficile*	*c. perfringens*	Induction with xylose	Nariya *et al.*, 2011 [280]
ARAi	*C. acetobutylicum*	*C. cellulolyticum*	Induction with arabinose (strong induction)	Cui et al., 2015 [281]
RecA	*Clostridium*	*Clostridium*	Induction with radiation (improve expression by 30%)	Nuyts *et al.*, 2001 [282]
2tet01	*E. coli* and *C. acetobutylicum*	*C. acetobutylicum*	Induction with anhydrotetracycline (aTc)	Dong *et al.*, 2012 [283]
IPL12	From a Variant Library created from the *tet3no* promoter	*Clostridium autoethanogenum*	Induction with anhydrotetracycline (aTc)	Nagaraju *et al.*, 2016 [284]

Table 5 List of promoters used in Clostridia. In black constitutive promoters; in purple inducible promoters.

The number of native or engineered promoters are increasing since the review of Pyne *et al.*, 2014 [229], Table 5. For instance, Nagaraju *et* al., 2016 built and tested a small library of inducible promoters like the low tetracycline-inducible promoter *tet3no*. Variants were tested and a strong tetracycline-inducible promoter was found: *IPL12*. In another article, Olson *et al.*, 2015 [268] screened several promoters strength. In this last article, the comparison of activities in *E. coli* and *C. thermocellum* of the reporter protein is interesting and primordial, and three rules of the ideal promoter for metabolic engineering were proposed.

First rule: "Low expression of the gene of interest in cloning strains of *E. coli*", it can be crucial when the cloned protein is toxic. For instance, glycoside hydrolases are toxic in *E. coli*, making cloning laborious. Moreover, most of used *Clostridium* promoters are recognized by the cloning strain. To avoid this issue during the cloning, Perret *et al.*, 2004 [272] developed a promoter with two lac operators in a *thl* promoter, and the protein expression is repressed by an *E. coli* strain containing $lacI^q$ repressor gene. When new promoters are developed and tested in Clostridia, the strength of the promoter in *E. coli* would be investigated too.

The second rule: the promoter must have a "Consistently high expression, independent of the genetic context"; in my point of view, this depends on the objectives of the promoter expression and regulation.

"Low homology to the chromosome (in the case of native promoters, shorter is better)", this third rule can be useful to avoid homologous recombination with the plasmid and the chromosome.

But various promoters are needed in Clostridia: inducible, repressible or with different strength, this idea started by Olson *et al.*, 2015 [268]. To create various promoters with these characteristics, promoters can be designed based on RNA-seq (Appendix 1, [51]) or more precisely at the base pair level from transcriptional start site (TSS) analysis (Appendix 2, [188]). Instead of chemical inducers, substrate inducers could be designed. For instance, the expression of a glycoside hydrolase would be finely tuned depending on the substrate and the transcription would be induced depending on the presence of the substrate. Characterized promoters are also important to bring Clostridia and its interesting properties to the synthetic biology community.

To have a precise library, these promoters would be tested with a unique reporter, to know the level of expression, induction or repression but also to compare results between labs. For this reason, the next chapter presents previously used and characterized promoters.

I.2.1.f Reporters

Reporters are interesting tools to characterize a promoter in various conditions or it can be used for protein localization in a fused protein system. An easy-to-use reporter allows a high-throughput screening of promoters. Identification of promoters using TSS data and screening of best candidates was previously done in *B. thuringiensis* with a β-galactosidase reporter [285]. In other chassis, fluorescent proteins, like the Green Fluorescent Protein (GFP), are mainly used as reporters instead of active enzymes; for instance to screen a synthetic promoters library in *E. coli* [286]. Interestingly, fluorescent proteins expression can be followed in real-time, compared to other reporters. Nevertheless oxygen is required by GFP-like protein

to become fluorescent. It is an issue for its utilization in Clostridia, which grow in anaerobic conditions. In this chapter, I will present various reporters mainly used in other Clostridia, with some advantages and disadvantages.

catP is the chloramphenicol resistance gene and one the most used reporter. In 1994, Matsushita *et al*. [287] proposed this gene as reporter to screen promoters. The chloramphenicol acetyltransferase activity is tested with the same protocol as Shaw *et al*., 1975 [288] and redescribed in 2015 [289]. In this work, the activity fold change is close to ten between induced and non-induced conditions. The chloramphenicol acetyltransferase activity of the cell lysate is followed by the increasing absorption at 412 nm with a spectrophotometer. The extraction of the cell lysate, preparation and the spectrophotometry make the chloramphenicol acetyltransferase assay laborious for the screening of many promoters. Nevertheless, chloramphenicol acetyltransferase assay was successfully tested in *C. acetobutylicum* [289], *C. sporogenes* [289], *C. perfringens* [290], *C. autoethanogenum* [284], *C. cellulolyticum* [291] or *C. difficile* [292].

lacZ encodes the β-galactosidase, an enzyme which cleaves β –glycosidic bonds. This assay can be performed if the bacteria have no high β-galactosidase activity measured. The cleavage of o-nitrophenyl-β-d-galactoside by LacZ produces a yellow product which can be measured in absorbance at 420 nm with a spectrophotometer. A thermostable LacZ can come from *Geobacillus stearothermophilus* to be used in *C. thermocellum* [293][268], or from *Thermoanaerobacterium thermosulfurogenes* in *C. acetobutylicum* [266][264].

adhB gene from *Thermoanaerobacter pseudethanolicus* expresses an NADPH alcohol dehydrogenase. The activity is measured with the disappearance of NADPH in absorbance at 340 nm with a spectrophotometer; the activity is tested on cell lysate. This reporter was tested in *C. thermocellum* and compared to the *lacZ* reporter, the activity of these two reporters are correlated to the strength of tested promoters [268]. But the authors did not explain advantages/disadvantages nor give advice to choose one reporter among *lacZ* and *adhB*.

gusA encodes for a β-Glucuronidase. The activity of this enzyme is measured with a sensitive fluorimetric assay compared to spectrophotometry for the previous reporter. This method was developed for the first time in 1994 in *C. perfringens* [270], and the gene came from *E. coli*. It was also used in *C. difficile* [294], *C. beijerinckii* [295], *C. acetobutylicum* [264]. Nevertheless, the endogenous activity of GusA in *C. perfringens* was an issue to get precise results [296]. The GusA activity increased by 80-fold at maximum induction compare to just two fold above background level in the absence of inducer.

phoZ is an alkaline phosphatase gene. The cleavage of 5-Bromo-4-chloro-3-indolyl phosphate (XP) or *p*-nitrophenyl phosphate (*p*NP) by PhoZ produces a color-forming precipitate which can be quantified. The alkaline phosphatase assay uses this property. *p*NP can be performed anaerobically whereas the assay with XP needs oxygen and this last assay is more suitable on Petri dishes. This method was developed in *C. difficile* with a *phoZ* gene from *Enterococcus faecalis* [297][298]. The preparation to quantify the activity seems as laborious as other enzymatic activity assay previously presented, but the induction can be quantified with a 50 fold change between induction and no induction compare to an over expression near 100 with RT-qPCR, making this approach suitable for promoter screening.

Fluorescent reporters were recently developed and reported in Clostridia. In theory, the main advantage of these reporters is the preparation, assays do not need extract cell lysate and it is possible to measure the fluorescence directly from cells with no or minor preparation. Famous fluorescent reporters used in other bacteria, like GFP, RFP (Red Fluorescent Protein), YFP (Yellow Fluorescent Protein), mCherry, etc. need oxygen to become fluorescent. To solve this issue in Clostridia, cells are exposed to oxygen, but real-time measurements are not possible. Several studies fixed cells before the exposition to oxygen; this technique enables protein localization studies. For instance with this method a mCherry (codon optimized for Clostridia) was successfully tested in *C. difficile* [299], and in the same manner a cyan fluorescent protein was also used in *C. difficile* [300] or YFP in *C. perfringens* [296]. Moreover, most of Clostridia possess a high intrinsic fluorescence which can perturb the precision of fluorescence acquisitions. Here is a list of fluorescent reporters used in Clostridia.

Flavin mononucleotide-based fluorescent reporter proteins like LOV (light, oxygen or voltage sensing) domains or FbFP (flavin mononucleotide (FMN)–based fluorescent proteins) were revealed as a reporter for *in vivo* fluorescence without oxygen in 2007 by Drepper *et al.*, 2007 [301]. This technology was used in *C. difficile*, *C. sordellii*, *C. acetobutylicum* by Buckley *et al.*, 2014 [302]. Nevertheless, the fold change after induction is just 2.5 to 5.6 more important, compare to more than 15 with RT-qPCR on the same experiment. It was also tested on *C. cellulolyticum* [6][32], *C. ljungdahlii* [303] or *C. acetobutylicum* [304].

lucB is a gene which encodes the luciferase, a fluorescent protein from *Photinus pyralis* which was investigated in *C. perfringens* [305] and for the second time in 2004 in *C. acetobutylicum* [265], but no other team worked with this reporter, making doubt on the advantage of this reporter.

SNAP and **CLIP** could be used to follow gene expression at the single cell level with microscopy, to precisely measure the expression of a single bacteria in an heterologous population by flow

cytometry or for immunoblotting [306]. SNAP and CLIP work in a similar manner. The SNAP-tag protein reacts with a modified fluorophore substrate: benzylguanine (SNAP) or benzylcytosine (CLIP) resulting in covalent labeling of the tag. Although this technique has no noisy background, the fluorescence is not possible in real time, and the SNAP substrate must be added before measuring the SNAP-taq expression. Nevertheless, the preparation seems less laborious than previous methods and the fusion protein with the tag can be used for protein localization. This method was mainly used in *C. difficile* to precisely solve sporulation mechanisms [39][307]. This technique could be investigated to quantify the promoter expression.

Measuring mRNA expression could be another possibility to know the strength of a promoter. But RT-qPCR experiments are long and laborious for high-throughput promoter screening and the activity of expressed protein is missing, it can be a bias. But RT-qPCR can be performed to prove the efficiency of new reporters [268].

In the future, RNA-based fluorescent reporter could be tested in Clostridia [308].

To compare all reporters, it should be interesting to use one constitutional promoter, one inducible promoter and promoterless constructions with all reporters mentioned above, plus RT-qPCR. In this way, reporters could be compared: doubts, specificities, advantages and difficulties about each technique (preparation, materials etc.) would be presented. Among reporters with enzymatic activity, *gusA* or *phoZ* genes seem more precise, fold change are more important compared to the background. The choice would depend on the native GusA activity of the strain.

I.2.2 Homologous recombination

The homologous recombination (HR) was the first method to modify Clostridia at the genomic level. The HR is a natural event, which recombines two DNA molecules at homologous sequences. Breaking DNA close to homologous sequences and/or select for positive recombination events with antibiotic resistance cassette, or counter-selection methods are useful methods to improve HR efficiency.

1.2.2.a Single-crossover homologous recombination

The single-crossover HR is a recombination enabling the insertion of a plasmid in a desired *locus*, Figure 22. This method was previously done in *C. acetobutylicum* for gene

inactivation with plasmid integration [309], and the efficiency was improved using non-replicative plasmid [310].

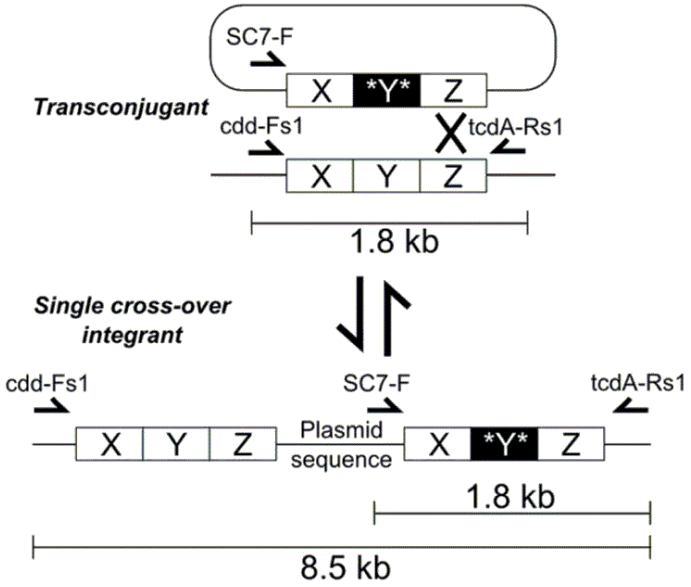

Figure 22 Schema of single cross-over integration into Clostridia genome, modified from Cartman *et al.*, 2012 [261].

Nevertheless, this technique integrates the plasmid and the antibiotic resistance cassette, preventing the use of the cassette for other plasmid integration.

1.2.2.b Double-crossover homologous recombination or Allele Coupled Exchange

To avoid the insertion of the whole plasmid, an elaborated method based on the HR and selection was established, the double-crossover HR. This method was improved by enabling the selection of mutants of this recombination event; this method is sometimes called Allele Coupled Exchange (ACE). It was successfully used to delete genomic region of few kilo base pair, Figure 23, or to insert exogenous DNA.

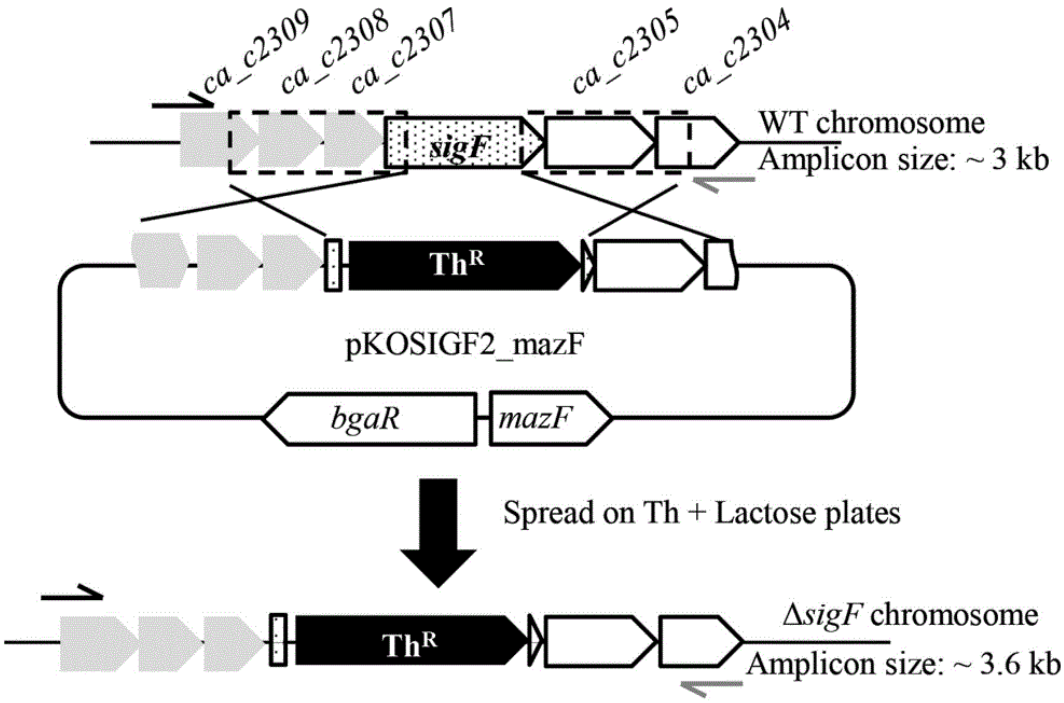

Figure 23 Example of double-crossover homologous recombination in Clostridium to delete a targeted gene, modified from Al-Hinai *et al.*, 2012 [253].

Several counter selection systems were used: the MazF toxin in *C. acetobutylicum* or *C. ljungdahlii* [311]; or four different but close counter-selection systems : the *pyrE* system in *C. difficile* [312], the pyrF in C. thermocellum [313], the *upp* gene in C. acetobutylicum [314] or with the *codA* gene in *C. difficile* [261]. A similar approach using two counter-selection markers was also tested in *C. thermocellum* to delete DNA sequences [315]. The ACE method was tested in a multistep strategy and over 40 kb of exogenous DNA were integrated into *C. acetobutylicum* [316]. Nonetheless, the prerequisite of these counter-selection systems is the creation of a mutant with the counter-selection gene deleted. This mutation can lead to growth defect, like for *C. thermocellum ΔpyrF* [313]. This growth defect is not an advantage for industrial purposes, but the tool is very useful to study metabolic modifications.

In order to have no antibiotic resistance cassette in the genome after the recombination, resistance marker can be surrounded with FRT sequences [317][318]. The expression of the FLP recombinase enables the deletion of the resistant marker.

Between single- and double- crossover strategies, a team tested another strategy: a single crossover recombination followed by a double crossover HR. The frequency of double crossover HR increased with double strand breaks. The expression of I-Sce1 endonuclease enables double strand breaks at I-SceI sites present in the plasmid, (integrated into the chromosome after the single crossover recombination). Moreover, the resistant marker is surrounded by I-SceI sites

and is removed during the double crossover recombination. This technique was used in *C. acetobutylicum* and *C. beijerinckii* [319].

The ACE method is explained with more details in a recent review [320]. Double crossover strategy was recently improved with the Crispr-Cas9 tool; it will be discussed in a next chapter.

I.2.3 Forward genetics

Forward genetics enables to determine the function of DNA sequences randomly mutated. For instance, the phenotype of a mutated strain is studied and the mutation is linked to a function. This random mutation approach followed by phenotype analysis was one of the first genetic technologies applicable to describe the function of a DNA sequence.

Forward genetics tools are available among Clostridia, for instance by creating random library of mutants and to analyze their phenotypes. The random insertion of transposon was used in some Clostridia. Historically, many random transposon insertions were described in Clostridium, especially in pathogens. Transfers *via* natural conjugation of antibiotic resistant transposon were studied between bacteria (Clostridia included [321][322]); the transfer of antibiotic resistance cassette is an issue for the treatment of infections, especially nosocomial infections in hospitals.

The ability of some transposon to randomly integrate genomes was used and developed as a tool for forward genetics. A good random transposon must integrate the transposon just once, for instance the *mariner*-based transposon was tested in *C. difficile* and 98.3% of mutants had a single insertion [323]. The mutant library of this study was screened for non-sporulating clones, and a new gene, the germination-specific protease gene *cspBA* was characterized. This approach was improved with conditional plasmid maintenance, tested in *C. acetobutilicum* and *C. sporogenes* [289] and with an inducible promoter [324]. Random transposon insertions were also achieved in *C. cellulolyticum* [256]. Several different transposons were tested and described for random insertion, nevertheless none of them is completely random and for instance the *mariner*-based transposon target TA sequences.

I.2.4 Reverse Genetics

Reverse genetics is a way to investigate the function of a gene. In the reverse genetics approach, a DNA sequence is targeted and modified, the phenotype of the strain is analyzed to deduce or propose a role to the DNA sequence. In Clostridia this reverse genetics approach was used by gene inactivation with intron insertion. The number of genome sequence available enables to target specific DNA sequence and verify the role or the function of predicted genes.

The group II intron insertion technology is a reverse genetics tool, also called Targetron, applied in 2005 in *C. perfringens* [325], then replicated in *C. acetobutilicum* [326]. In parallel, the Clostron technology was tested in 4 Clostridia (*C. acetobutylicum*, *C. difficile*, *C. sporogenes*, *C. botulinum*) [327]. Both technologies are based on group II introns retargeting.

Group II intron is made of RNAs and retrotransposable elements; the RNA is reverse spliced into a DNA target site with the help of the Intron-Encoded Protein. After the insertion, proteins can remove the insertion [328]. The group II intron LI.LtrB from *Lactococcus lactis* was modified to prevent the intron removal, and an algorithm was released to retarget intron insertion to desired *loci* [329]. This algorithm is available in www.targetrons.com or www.clostron.com (free). To retarget an intron, the binding sequence of the intron with the DNA target-site recognition is modified with algorithm results, to give customized sequences of EBS1, EBS2, IBS1 and IBS2 (Figure 24 **A** and **B**). The mechanism is briefly described in the Figure 24 **C**.

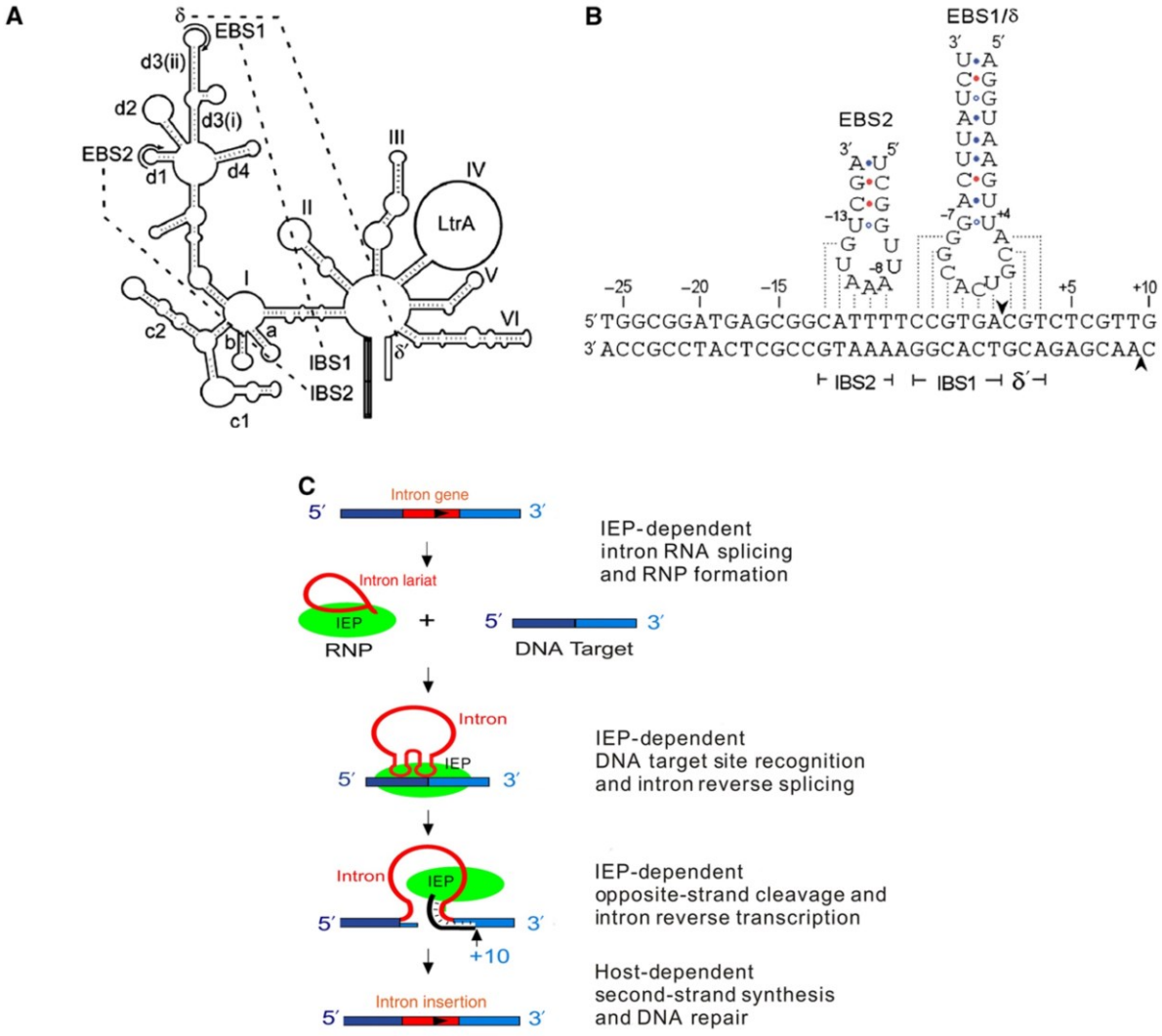

Figure 24 Group II intron technology, from Enyeart *et al.*, 2013 [330]. A Schematic structure of the LI.LtrB intron (RNA). B Base-pair contact involved in DNA target-site recognition. C General mechanism of intron splicing and targeting.

The main difference between Clostron and Targetron is the use of selection marker in Clostron. Albeit the intron insertion efficiency is very high with Targetron, the team of Minton who developed this technology added a resistance cassette inside the intron to select for intron insertion. It is inconvenient for incremental intron insertion because the selection marker can be used only once. To solve this issue, the Clostron technology was improved by surrounding the marker with FRT site, and the resistance cassette could be removed with the expression of the FLP protein [5]. In my point of view, the Targetron technology, without any selection marker and an efficiency between 25% to 62% [326], is sufficient. Compared to Clostron, no other manipulations are required, and no marker has to be removed. Nevertheless most of insertion with group II introns in Clostridia occurred with the Clostron technology, list of Clostron insertion in Clostridia in Table 8 of Pyne *et al.*, 2014 [229]. For instance, Clostron was used for

gene inactivation to guide metabolic pathways, or to block restriction modification systems (and accept unmethylated DNA).

I.2.5 CRISPR-Cas

The well-known genome editing technology CRISPR-cas, developed with success for eukaryotic cells is now adapted to be a new tool in Clostridia too.

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) and CRISPR-associated (Cas) genes is a bacterial defense system. This adaptive immune system enables the bacteria to target and break foreign DNA. This system was studied and guided to break desired DNA sequence [331] in prokaryotic and eukaryotic cells. In bacteria, the double strand break of this tool is mainly lethal because bacterial recombineering system is not efficient as in eukaryotes. The CRISPR-cas technology was modified and adapted to work in bacteria.

This technology was applied to improve the HR method with a selection against cells without HR events *C. beijerinckii* in2015 [332]. The CRISPR/Cas9 from *streptococcus pyogenes* was targeted to break the DNA between two regions of 1 kb homologous to the same sequence in the plasmid, Figure 25. Advantages of this technology compared to the normal HR:

- The selection for the event, which increase the number of positive clone for the deletion *via* HR without plasmid integration.
- No resistance marker remaining in the chromosome (and no further work to remove one.
- The homology sequence and CRISPR/Cas9 elements is present in one plasmid

Perhaps the length of the homologous sequence is a disadvantage.

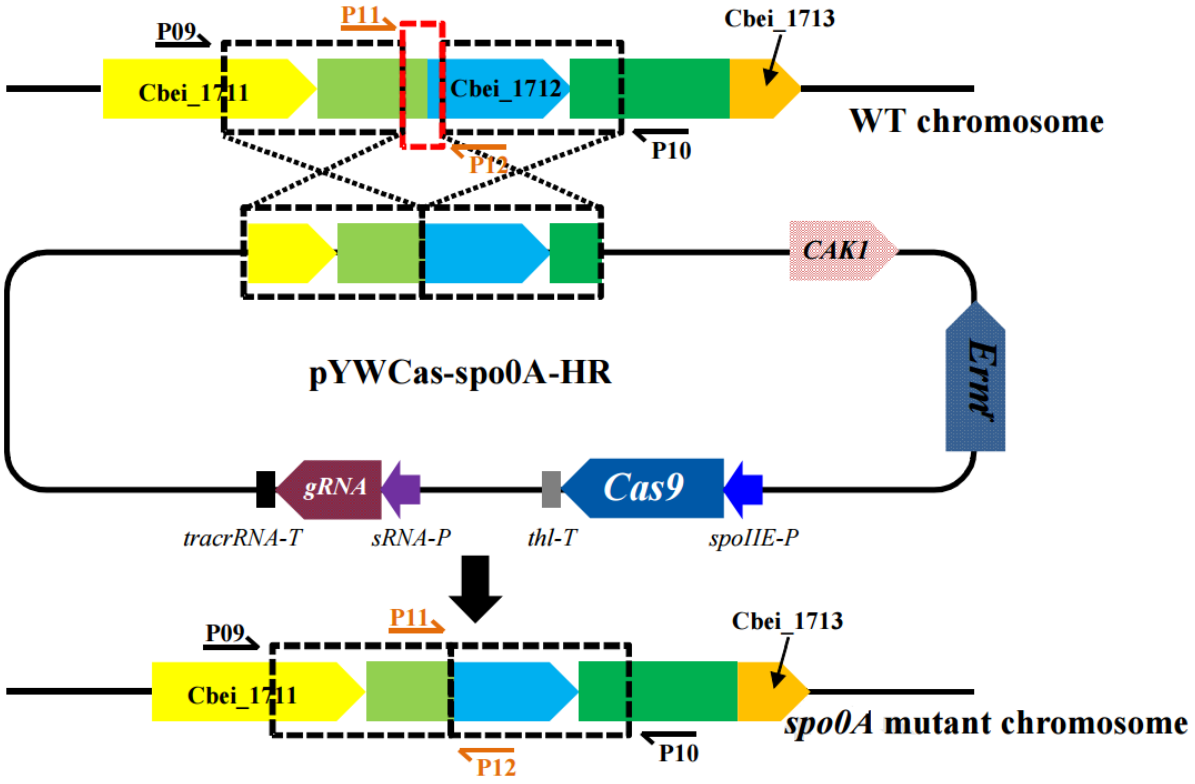

Figure 25 CRISPR/Cas9 in *C. beijerinckii*. Sequences in the black dash rectangles of the plasmid pYWCas-spo0A-HR have a homology of 1kb long with the black dash rectangles corresponding to sequences of the chromosome of *C. beijerinckii*. The sequence surrounded by the red dash rectangle is targeted by the gRNA of CRISPR/Cas9 system. Modified from Wang *et al.*, 2015 [332].

A close strategy was applied in *C. cellulolyticum* few months after [333]. This second team used an engineered Cas9 protein from *S. pyogenes*: Cas9 nickase (Cas9n). This Cas9ne induce single-nick HR instead of double strand breaks. This system seems more efficient than with the wild-type enzyme and only 0.2 kb of homology are required for the HR. This method was applied for a precise 23 bp deletion and insertion up to 1.7 kb. Insertion or deletion with this method was achieved with only one plasmid with HR sequences and CRISPR/Cas9n, nevertheless they were not able to integrate more than 1.7 kb, although 3 and 6 kb were tested. This tool was also applied in *C. acetobutilicum* [334].

Following achievements: 1.5 kb deletion in *C. beijerinckii* and single nucleotide modifications [335], nevertheless in this article they report some issues with plasmid integration. The same team reported the utilization of another customized Cas9 protein (dCas9), which lacks endonuclease activity but the enzyme with the help of other CRISPR element can bind to the DNA. This ability was used for gene repression in *C. beijerinckii* [336][334]. This repression strategy was also applied in *C. acetobutylicum*, *C. pasteurianum* [337]. Four deletions were

achieved in *C. ljungdahlii*, with one near 3 kb [338] and two deletions were successively achieved in the hyper-butanol-producing Clostridium saccharoperbutylacetonicum [339].

Three quarters of Clostridia has an endogeneous CRISPR-Cas system. A team compared the recombination efficiency between the endogeneous (Type I) CRISPR-Cas3 system of *C. pasteurianum* and the heterologous *S. pyogenes* (Type I) CRISPR-Cas9 system (used in other studies of CRISPR-Cas in Clostridia) in *C. pasteurianum*. The efficiency was near 100% for the endogeneous system compared to 25% for the heterologous system, albeit the two systems have small differences in the mechanism, endogeneous system is an interesting possibility and improvement for the CRISPR-Cas system in Clostridia. Obviously, exploiting endogeneous CRISP-Cas system is not possible for Clostridia which do not have this system, like *C. phytofermentans*.

The control of the Cas9 expression seems important for most studies, a good inducible promoter is essential to improve the efficiency, as reported for *C. autoethanogenum* [284]. A long paragraph of the review of Xue *et al.*, 2017 [340] related to *C. acetobutilicum* describes the CRISPR/cas technology in Clostridia.

Although the CRISPR/Cas technology is recent, first articles were released two years ago in Clostridia, this genome editing tool (insertion, deletion, point mutation, repression) is really powerful and can be done in one step most of the time. Numerous of teams are applying, developing and improving this technology for new purposes in Clostridia.

I.2.6 Random mutagenesis, selection and directed evolution

Another powerful method exists to study and/or improve Clostridia: the evolution. Only techniques used on Clostridia related to this method are presented here. The goal of this method is to guide bacteria to get genetic modification, inducing new properties/abilities/specificities; moreover most of techniques do not need prerequisites of genome sequences or transformation.

Historically the tolerance for solvents was improved by serial enrichment procedure. A culture begins with a small solvent concentration and after some dilutions the solvent concentration is increased. Strains are forced to evolve to higher concentration of solvent, and after several round of increasing concentration of solvent, tolerant strains are selected. With this technique

solvent tolerance is increased but the solvent yield is constant or decreased. This strategy was developed in 1980's in *C. thermocellum* [341], *C. acetobutylicum* [342][343] and we reused it in one of our recent publication in *C. phytofermentans* (Appendix 4, [216]). In the same manner, *C. thermocellum* was evolved for faster growth on avicel [262].

To increase the evolution speed and the number of mutations, various random mutagenesis protocols were established, followed by a selection or a mutant analysis. This strategy was developed for the first time in Clostridia in 1985 [344] and several mutagenic mechanisms were tried: U.V radiation, ethyl methane sulphonale (EMS), *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (NTG), mitomycin C, hydrogen peroxide, nalidixic acid, metronidazole. They concluded that EMS and NTG were best mutagenesis agents tested in *C. acetobutylicum* and *C. pasteurianum*, mutants were investigated by analyzing mutant auxotrophy for amino acids.

Using this random mutagenesis strategy and selection, several strains were improved with different mutagenesis techniques, Table 6.

Specie	Mutagenesis technique	Achievements	References
C. acetobutylicum and C. pasteurianum	*N*-methyl-*N'*-nitro-*N*-nitrosoguanidine	Auxotrophy for amino acids	Bowring *et al.*, 1985 [344]
C. beijerinckii	*N*-methyl-*N'*-nitro-*N*-nitrosoguanidine	Enhanced butanol production	Formanek *et al.*, 1997 [345]
C. acetobutylicum	Serial enrichment procedure and genome shuffling	Enhanced butanol tolerance and butanol yield	Mao *et al.*, 2010 [346]; Gao *et al.*, 2014 [347]
C. acetobutylicum	Nitrogen Ion Beam Implantation	Enhanced butanol tolerance and butanol yield	Liu *et al.*, 2012 [348]
C. acetobutylicum	*N*-methyl-*N'*-nitro-*N*-nitrosoguanidine	Higher solvent production (ABE)	Jang *et al.*, 2013 [349]
C. acetobutylicum	Inactivation of the mutS/L operon, and control of the operon expression in a plasmid	Control of mutation with an inducible promoter for butanol tolerance	Luan *et al.*, 2013 [350]
C. acetobutylicum	Atmospheric and room temperature plasma	Enhanced butanol tolerance and butanol yield	Li *et al.*, 2014 [351]
C. beijerinckii	*N*-methyl-*N'*-nitro-*N*-nitrosoguanidine followed by genome shuffling	Enhanced isopropanol tolerance and isopropanol yield	Gérando *et al.*, 2016 [163]

C. phytofermentans	GM3 automat	Enhanced lignocellulosic inhibitors tolerance	Cerisy *et al.*, 2017 [98]

Table 6 Mutagenesis technique used in Clostridia.

Techniques developed in this table are mainly based on mutagenesis elements (compound or conditions). But two techniques are different: genome shuffling which is based on the recombination of entire genomes and the evolution with the GM3 automat. The GM3 automat enables to maintain the culture in log phase and to fill in automatically the culture with pulses of stressing or relaxing medium, depending on the behavior of the culture. This technique is presented in the result section of this thesis.

I.2.7 Antisense RNA

The antisense RNA technology was developed in 1990'. The goal of this technology is to engineer an RNA which would decrease specifically a protein expression. The engineered RNA represses a protein expression by binding the RNA corresponding to the targeted protein. This engineered RNA is named antisense RNA (asRNA). The repression is not total, but important enough to have quantifiable phenotypic modifications and the repression can be tuned with inducible promoters. Nevertheless, due to many factors the downregulation is hard to design to be efficient, and many parameters have to be modified to succeed. This unpredictability had prevented the expansion of this technology. asRNA were applied in 2000' in several Clostridia. In 1999, asRNA were tested in *C. acetobutylicum* to downregulate butyrate kinase and phosphotransbutyrylase expression, increasing the lactate production but with a lower solvent production [352]. This team downregulates the expression of four other proteins in *C. acetobutylicum* [353][354]. asRNA were also used in: *C. cellulolyticum* to show the important role of a cellulase in the cellulolytic system [355], *C. perfringens* to decrease the resistance of spores to heat and UV radiations [356] or in *C. saccharoperbutylacetonicum* to control the electron flow [357].

A long paragraph of the review of Pyne *et al.*, 2014 [229] describes asRNA in Clostridia. But with recent advances with CRISPR-Cas9 in Clostridia, especially with CRISPR-dCas9 (or CRISPRi) variant for gene repression [334][336][337], asRNA would probably not be used in the future. CRISPR-dCas9 is easier to retarget than asRNA to downregulate protein expressions.

I.2.8 Perspectives

Now, in 2017, many various genetic tools exist in Clostridia, as previously presented, but it is still much less than in *E. coli*. Tracks can be pursued, for instance we know that Clostridia have a number of bacteriophages [358], leading to important failure in industrial ABE fermentation [359][360]. Bacteriophages of Clostridia could be more studied, mechanisms between host and phages can be interesting to fight some pathogens, new genetic tools could be also discovered and developed like in *E. coli* with the Keio collection [361] and λ phage proteins.

In my point of view, two tools are urgently required: promoter libraries to develop Clostridia for actual and future purposes, with inducible, repressible or constant promoters with various strength. The second tool is a good an easy to use reporter.

Later in this thesis, I will describe another genetic tool: Genome Editing via Transposon and Recombinases (GETR) [330], this tool was not tested in Clostridia before. This genetic tool could be used for broad genomic manipulation; if it will work as expected, GETR enable the deletion or insertion of long DNA sequences (more than 10 kb) in Clostridia. It will bring new possibilities to Clostridia: minimal genomes, integration of complete metabolic pathways in one step etc.

Some tools were developed in the synthetic biology community to prevent contamination and spreading of modified bacteria. These tools, called kill-switch, are important to protect the environment against our modified bacteria, but also to reassure people of the safety of our work. This can be applied for industrial and medical purposes.

With the developments of -omics (genomics, proteomics, transcriptomics, metabolomics) the quantity of data are increasing and a tiny volume of them are exploited. Mathematic models could utilize more of these data and could help to drive future experiment to improve the understanding of metabolic pathways. Few articles are linking model and experimental data [362][363]. A mini review present current developments of metabolic modeling in Clostridia [364], interactions between modeling and real data could improve both fields.

I.3 Biofuel

Nowadays, one of the major issues of our planet and next generations is the global warming. In 2015, the humanity had consumed more than 13.8 Gtoe (Gigatons of oil equivalent) of energy. The growth of the consumption was near 0.5% but the CO_2 decreased by 0.2% thanks to renewables energies [365]. Nevertheless, the energy consumed by the G20 is still dominated by coil, oil, gas; gas and oil represent 82% of the total, Figure 26. 9% of the energy consumed comes from biomass and heat, and the last 10 % part of the energy consumed is from renewable energies: solar, wind, geothermal, hydro but also the nuclear energy (not renewable).

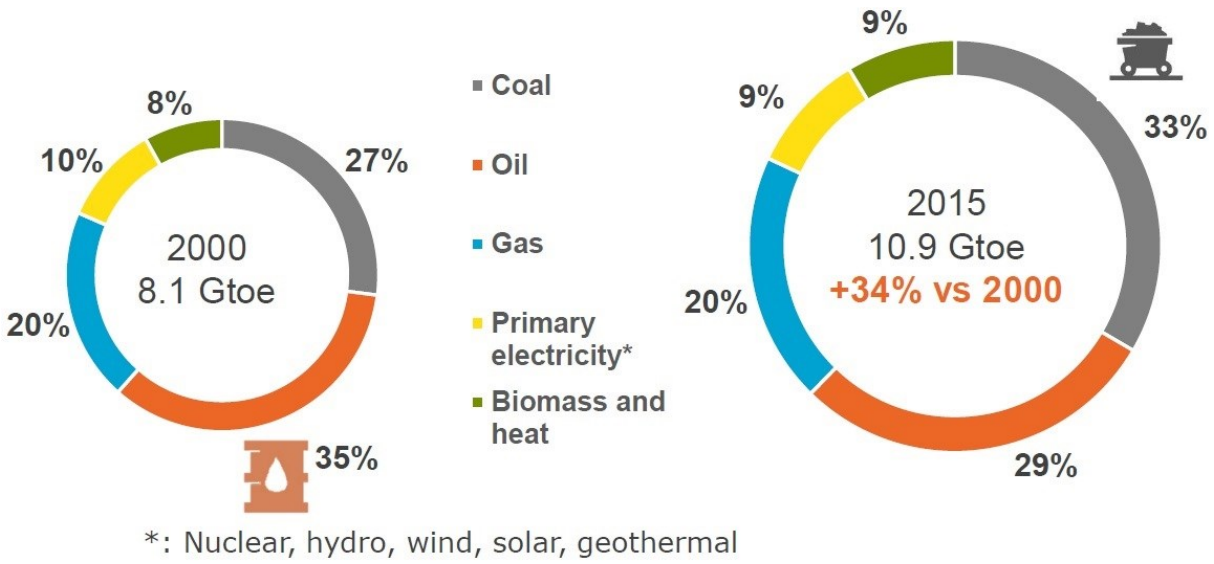

Figure 26 Comparison of the energy mix of G20 countries between 2000 and 2015, From Global Energy Trends, 2016.

The Kyoto protocol, signed in 1997, effective in 2005, commits 192 state parties to recognize that the global warming exists and was (and is still) caused by the CO_2 released by human activities. State parties are engaged to the "stabilization of greenhouse gas concentrations in the atmosphere at a level that would prevent dangerous anthropogenic interference with the climate system" (Art. 2)[366].

In December 2015, 196 parties signed the goal of limiting global warming to less than 2°C to pre-industrial levels, but also continue efforts to limit the temperature increase to 1.5°C. It is the Paris agreement voted at the COP21.

To achieve these objectives, two main paths must be performed: countries have to develop renewables energies to replace fossil energies (oil, gas, coal) which release enormous quantities of CO_2; and they have to decrease their energy consumption. Several alternative energy sources

exist and are in development to take more and more significance in the energy mix of countries: sunlight (solar panel), wind (wind turbine), hydropower, geothermal heat, biomass etc. but with 14% of renewable energy [367] the road ahead is still long to replace fossil energies. In the mix of renewable energies produced, solid biofuels/charcoal are still dominant and renewable energies from new technologies (geothermal, liquid biofuels, solar, tide, wind and biogases) are in minority with around 1.8% of the total of the world total primary energy supply, and 15.3% of renewables energies. Ancient "renewable" energies like hydropower and solid biofuels/charcoal are still prevailing, Figure 27.

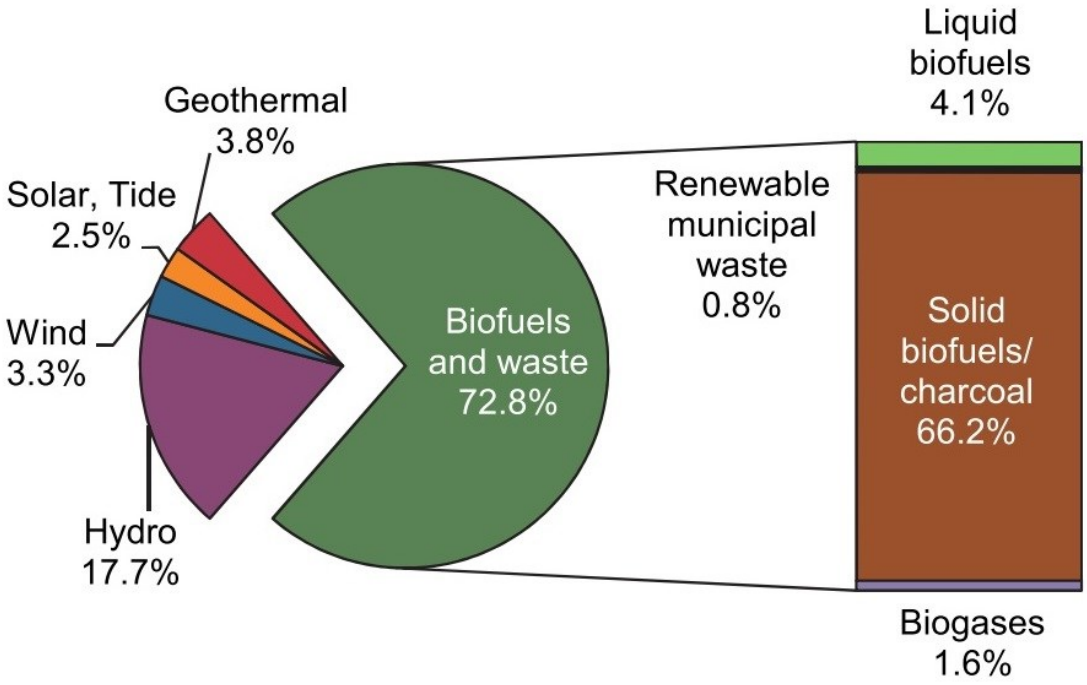

Figure 27 Product shares in world renewable energy supply, From International Energy Agency, Key renewables Trends, 2014 [368].

To develop renewable energies, 135 countries created the International Renewable Energy Agency (IRENA), showing the high level of global interest in advancing renewable energies. Renewable energies possess several goals: decreasing the level of CO_2 released, reaching energy independence, creating new jobs, creating several and different renewable energies locally (depending where you can find the energy source: wind, water, biomass, solar etc…). United Nations created Sustainable Energy for All (SEforALL) with the Special Representative of the UN Secretary-General for Sustainable Energy for All have three objectives: Ensure universal access to modern energy services, double the share of renewable energy in the global energy mix, double the global rate of improvement in energy efficiency [369]. Nevertheless, the new Trump's administration in the United States call into question the need of bioethanol and biodiesel in oil mix, this view can destabilize the development of biofuels industry of all different generations

[370]. Moreover this new administration is opening new coal mine, inconsistent with the COP21 (Paris agreement) signed by USA.

In the next part we will focus on biofuels production from different sources. Depending on the type of carbon source used, the technology is called 1st, 2nd, 3rd or 4th Generation.

I.3.1 1st Generation of biofuels

"The fuel of the future is going to come from fruit like that sumac out by the road, or from apples, weeds, sawdust – almost anything". Henry Ford, Ford Motor Company CEO, 1925. First cars ever built were made to work on biofuels, Rudolph Diesel had the intention to use vegetable oil for his engine and Henry ford designed the 1908 model T to consume ethanol. But the low price to extract fossil oil and the prohibition movement in the USA stopped the research and fossil fuels were encouraged.

The 1973 oil crisis or the gulf war increased price of fossil fuel and renewed interests into biofuels. Sometimes, it allowed biofuel to be competitive with fossil fuel. Countries with important land and appropriate climate like USA and Brazil developed large scale biofuel production.

The main advantage for biofuel is to claim that the CO_2 released is from renewable source, but also it permits (partial) energy independence to a country. The development of jobs related to this industry is also a major point. Nevertheless, the main consideration about this development is the use of arable land for biofuel purposes instead of food production. This competition decreased the size of land dedicated for the human food, increasing the price of cereals, potatoes and finally staple food. Moreover these issues raised ethic issues and countries legislated to devote a limited percentage of land to biofuels. European commission decided to decrease their goal of 10% of 1st generation of biofuels to 7% and they have to prepare advances biofuels (2nd and 3rd generations)[5][6].

Nowadays main types of biofuel are from sugarcane ethanol in Brazil, corn ethanol in US, oilseed rape biodiesel in Germany and palm oil biodiesel in Malaysia[373]. Even if it is one of the major biofuel, soybean biodiesel is considered as the worst feedstock for biofuel production in term of productivity.

If the biofuel feedstock does not fit for food consumption anymore, due to damages, the biofuel produced is considered as 2nd generation of biofuel.

The first generation of biofuels is produced from the fermentation of sugars (mainly extracted from sugarcane) or starch (extracted from grains). For the ethanol production from grains, the substrate is milled and starch is liquefied using enzymatic process[374]. To liquefy the starch slurry, the pH is adjusted, alpha amylase is added and carbonate ions is used to buffer the solution[375]. The alpha-amylase is thermostable and the starch slurry has to be heated, the saccharification is done using glucoamylase. After the saccharification the solution is fermented using the yeast *Saccharomyces cerevisiae*. Co-product are valuable and can be sold as animal feed [376][377].

I.3.2 2nd Generation of biofuels

I.3.2.a Introduction to the second generation

The second generation of biofuels is mainly based on lignocellulosic materials from dedicated non-food cultures (Switchgrasses, *Myscanthus* etc…), inedible parts of plants (straw from agriculture waste), food wastes (including waste vegetable oil) or forest residues [378]. The financing of the second generation of biofuel mainly depends on the success of the first generation. Several pilot and demonstrator constructed and under construction for the second generation are related scientifically and financially to the first generation. Some examples: Projet Futurol is close to Cristanol (1G), Project Liberty is from POET-DSM (POET produces 6.5 billion liters of bioethanol from the first generation), DuPont is developing its own technologies to the second generation and produces 68 billion liters of ethanol from its first generation plants.

The second generation of biofuels can be produced from two main and different technologies [373]:

- Biochemical: cellulose and hemicellulose are metabolized to sugar from extracted enzymes or micro-organisms. Resulting sugars are fermented to produce mainly ethanol. Alternatives metabolic pathways can be added to get other final products.
- Thermochemical: the lignocellulosic feedstock is converted to syngas from torrefaction, pyrolysis/gasification. The final step convert the syngas to hydrocarbons, which can be used as biodiesel or jet fuel [379].

Different variations and alternatives of these technologies exist and are tested at the research or pilot scale. Because this thesis is focusing on C. phytofermentans, I will develop more about

differences into biochemical technologies of the second generation. Moreover it is possible to harvest and add value to different byproduct of the fermentation of 1st and 2nd generation to increase the income of industrial plants. For instance 1/3 of products of the dry milling process from corn are byproducts. Some of these products have a high value. The solid unfermented grain portion is more concentrated in protein, fibers and nitrogen than common silage [380]; allowing to harvest less lands for feed animals.

I.3.2.b Biomass composition

The carbon dioxide from the atmosphere is captured and incorporated by plants with the photosynthesis. Around 200 billion tons of biomass are produced from land plant and less than 2% are utilized by the humanity (for energy and materials) [381]. This biomass is mainly composed of to plant cell wall. This last material is mainly composed of cellulose, hemicellulose and lignin, also called lignocellulosic biomass. Not exhaustive list of polysaccharides in plant cell wall [382][383] (Appendix 1, [51]):

- Cellulose, xylan, xyloglucan, (gluco)mannan, galactomannan, glucuronoxylan, glucuronoarabinoxylan, galactoglucomannan, β-(1→3,1→4)-glucan, arabinoxylan, arabino(glucurono)xylan, heteroxylan; homogalacturonan, rhamnogalacturonan I, xylogalacturonan, rhamnogalacturonan II, arabinogalactan II, arabinan, starch, fructan.

Homosaccharides (for instance cellulose, homogalacturonan, xylan etc.) are composed of one type of sugar in contrary to heterosaccharide (for instance xyloglucan, arabinoxylan, glucomannan etc.) made of different monosaccharides, linked with glycosidic bonds. Glycosidic bonds can be different depending on the the sugar: β(1→4), β(1→6), α(1→4), α(1→6), etc., Table 7. The decomposition and degradation of polysaccharide by bacteria, fungi or yeast occur with released hydrolases of these organisms. Hydrolases cleave polysaccharide into oligosaccharides and sometimes to monosaccharides; these enzymes cut glycosidic bonds between saccharides of the polysaccharides. Hydrolases are (mainly) specific to a glycosidic bonds of a polysaccharide.

Polysaccharide	Natural Abundance (%w/w)	Structure	Sugar Composition
Glucans			
cellulose	9-14% monocot primary cell walls and Arabidopsis leaves 90% in cotton seed hairs	β-1,4-D-glucopyranose chain Hydrogen bonded into microfibrils	D-glucose
carboxymethylcellulose (CMC)	soluble cellulose derivative	β-1,4-D-glucopyranose with carboxymethyl side groups	D-glucose
starch	70% in cereal grains, 5% in aspen/spruce roots	73% amylopectin (α-1,4- and α-1,6 glucose) 27% amylose (α-1,4-glucose)	D-glucose

Hemicelluloses			
xyloglucan	2% in grasses, 20-25% dicot primary walls	β-1,4-D-glucose chain with α-1,6-D-xylose side groups, sometimes β-galactose or α-L-arabinose O2 substitutions	45% D-glucose, 34% D-xylose 18% D-galactose, 3% L-arabinose
xylan	12-18% in hardwoods; 24.5% in corn stalks	β-1,4-D-xylose chain	>90% D-xylose
arabinoxylan	20-50% in grasses	β-1,4-D-xylopyranose backbone O2,O3-linked L-arabinofuranose	59% D-xylose, 41% L-arabinose
mannan	68% of tomato endosperm	β-1,4-mannan (15-20 units/chain)	99% D-mannose
galactomannan	30% of legume seeds 10-30% of conifer secondary walls	β-1-4-D-mannopyranose backbone α-1-6-D-galactopyranose side chains	80% D-mannose, 20% D-galactose
glucomannan	5% of dicot cell walls	β-1,4-D-mannopyranose and D-glucose Some glucose are acetylated. no side chains.	60% mannose, 40% glucose
Pectins			
homogalacturonan (HG)	23% in angiosperm leaves, >50% in fruits	unbranched chain of ~100 1,4-α-D-galacturonic acid, partially methylesterified at C6 carboxyl	D-galacturonic acid
rhamnogalacturonan I (RGI)	9% Arabidopsis leaves ~36% in potato tubers	[->α-D-GalA-1,2-α-L-Rha-1,4->]n ~50% C-4 substituted with AG I or AGII	62% D-galacturonic acid, 20% L-rhamnose, 12% D-galactose, 3.3% L-arabinose, 1% D-xylose
galactan	25% of potato tubers	β-1,4-D-galactose	91% D-galactose, 5% D-galacturonic acid, 2% L-arabinose, 1.8% L-rhamnose, 0.2% D-xylose
arabinogalactan II (AGII)	3% Arabidopsis leaves	β-1,3-galactose backbone β-1,6-galactose and arabinose branches	85% D-galactose; 15% L-arabinose
arabinan	2% Arabidopsis leaves 35% of *Myrothamnus* cell walls	1,5-α-L-arabinan	71% L-arabinose, 26% D-galactose, 3% L-rhamnose

Table 7 Example of some plant polysaccharide characteristics: natural abundance, structure and composition. From supporting information S1 of Boutard *et al.*, 2014, Appendix 1.

The main component of the lignocellulose is the cellulose, a polysaccharide of 2000 to 15000 β-1,4-linked glucose (hexose) molecules [384]. Several cellulose chains are linked together to form microfibrils with hydrogen bonds [385]. In microfibrils, ordered cellulose chain represents the crystalline cellulose whereas the amorphous cellulose consists of a mix of chain cellulose. The second is the hemicellulose, a general term for various polysaccharide which contains 200 to 400 units linked together by glycosidic bonds [150]. These polysaccharides are composed of pentoses (for instance: xylose, arabinose), hexoses (for instance: mannose, glucose, galactose) and sugar acid (for instance: glucuronic acid or galacturonic acid, main components of the pectin), Figure 28. Xylan is the major hemicellulose polysaccharide but other polysaccharide exist (for instance: xyloglucan, glucomannans, galactomannans etc.).

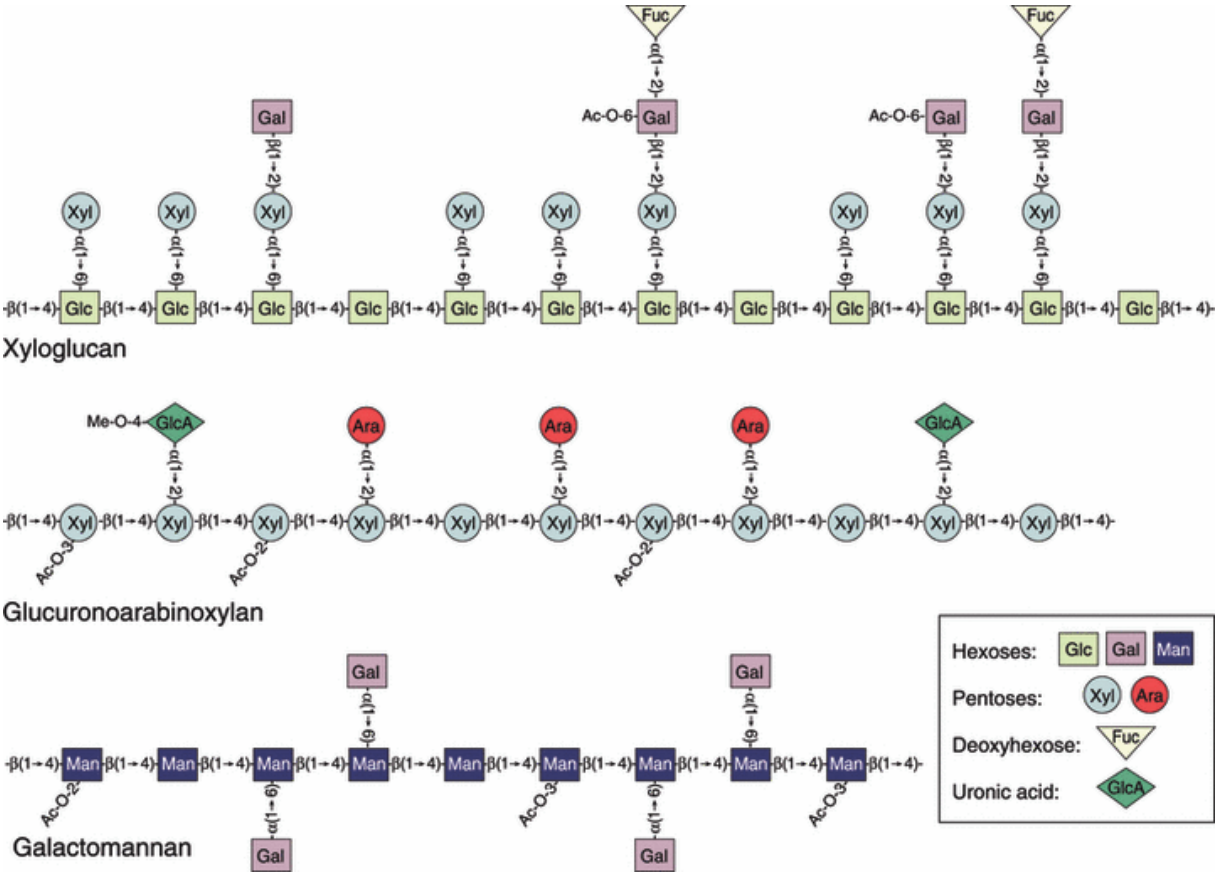

Figure 28 Examples of hemicellulose structures present in cell walls, from Pauly *et al.*, 2008 [386].

The third main component of the lignocellulosic biomass is the lignin. It is an aromatic phenolic polymer that provides structural strength to the plant, like glue. Lignin is mainly composed of p-coumaryl alcohol, coniferyl alcohol and sinapyl alcohol but other components are in between lignin and hemicellulose: caffeic acid, ferulic acid etc. [387]. The lignin polymer has numerous and various different linkages. Because the lignin is hydrophobic (insoluble), recalcitrant to chemicals and biological degradation, this compound is a challenge for the fermentation, at large, scale of the plant biomass [388].

Cellulose, hemicellulose and lignin are intertwined and form a composite material, the lignocellulose [385]. Cellulose chains form microfibrils which are attached to hemicellulose with hydrogen bonds. The lignin fills in gaps between cellulose and hemicellulose creating ester bonds between hemicelluloses, Figure 29. More microfibrils are packed into fibers more the cellulose is crystalline and insoluble. Crystalline cellulose with the protection of lignin is very hard to depolymerize. Without pre-treatment, this cellulose cannot be fermented. The pretreatment cost is one of the main issue for the development of the biomass fermentation (with the price of hydrolyzing enzymes): Heat, chemicals (with further neutralizing chemicals)

and/or mechanical techniques are costly in the process. Moreover it is energy intensive and it produces CO_2, losing the carbon efficiency of the technology [389].

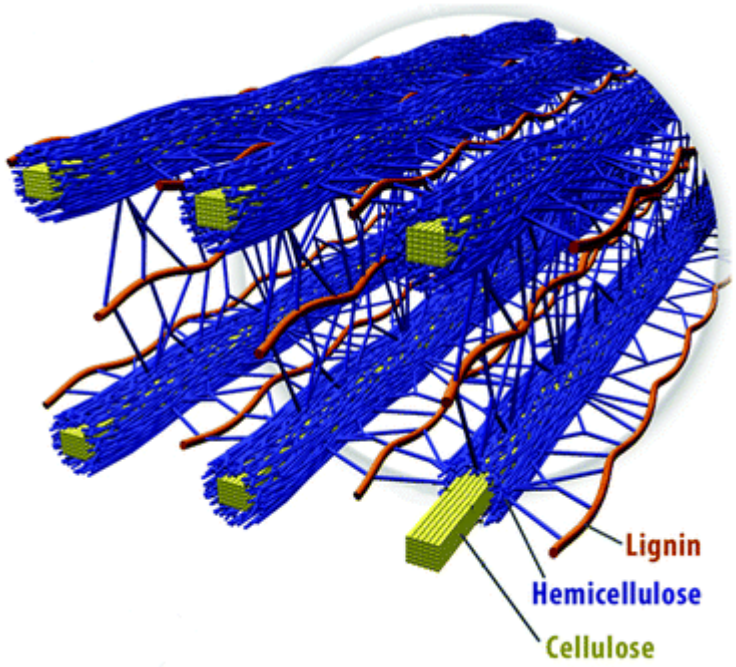

Figure 29 Schematic representation of the lignocellulose, from Brandt *et al.*, 2013 [390].

These three main lignocellulosic components are in various percentages in plants, depending on species, tissue and the environment [391].

I.3.2.c Pretreatment and biochemical technologies of the second generation

The first common step between biochemical technologies is the pretreatment. Different methods are used and developed to degrade the plant cell wall. The goal is to liberate plant polysaccharides for the subsequent hydrolysis followed by the fermentation. These methods utilize different biomass feedstock and create difference of the accessible substrates, cellulose crystallinity and lignin degradation. Main studied pretreatment are [392]:

- Biological: Brown, white and soft-rot fungi with lignin-degrading enzymes like laccases and peroxidases
- Physical: Extrusion or mechanical with a mix of milling, grinding or chipping
- Chemical: Alkali pretreatment (sodium, potassium, calcium, ammonium or lime), acid pretreatment and ionic liquids pretreatment
- Physico-chemical: Steam explosion, liquid hot water, ammonia fiber explosion, wet oxidation

The pretreatment can combine different techonologies. The choice of a pretreatment depends also on the biomass feedstock [393], the cost and following methods used for the hydrolysis and the fermentation.

After the pretreatment, four main technologies have been studied for the hydrolysis and the fermentation of the substrate, Figure 30. The first technique, Separate Hydrolysis and Fermentation (SHF), separates all steps: the enzymatic hydrolysis occurs with externally produced enzymes, hexose and pentose are fermented in different reactors. Nevertheless, the increasing concentration of glucose and cellobiose during the fermentation inhibit cellulases [394].

The Simultaneous Saccharification & Fermentation (SSF) performs the hydrolysis and the hexose fermentation in the same batch. It avoids the inhibition from glucose and cellobiose accumulation, but enzymes and fermentation are not at the optimal conditions. Several studies described a higher and faster ethanol production in SSF than SHF [395][396].

The Simultaneous Saccharification & Co-Fermentation (SSCF) process allows the hydrolysis and fermentation of hexose and pentose in a same batch. For example, one of the strategy was to improve the original strain of the SSF, *Saccharomyces cerevisiae* [397] or *Zymomonas mobilis* [398]. Heterologous genes were added to these genomes to ferment xylose (pentose) in addition to the natural fermentation of hexoses. Abengoa and Dupont did not reveal if their new commercial-scale cellulosic ethanol projects are SSF or SSCF [399] but Poet-DSM, Betarenewables and Futurol (pilot) are developing different strategies related to the SSCF technology.

The last technology is named consolidated bioprocessing (CBP) [400]. CBP is still in research stage and no pilot are developed yet [385]. This technology simplifies the process by combining enzyme production, hydrolysis and the fermentation of hexose and pentose in one reactor. All these steps are realized by one or consortia of organisms (yeast, fungi or bacteria) in one batch. In theory, It would decrease by twofold to fourfold the cost of the lignocellulosic conversion compare to SSCF and increase the efficiency [401][24]. To implement this technology some strain improvements are needed, and two different categories are developed. The category I strategy is engineering a cellulolytic strain such as *Clostridium thermocellum*, *C. phytofermentans* or *T. reesei* [403] to be an important ethanol producer. Vice-versa, the second category [404] implement cellulolytic enzymes in strains with natural high fermentative ability, such as *Saccharomyces cerevisiae* [405], *Zymomonas mobilis* [406] or *C. acetobutylicum* [141].

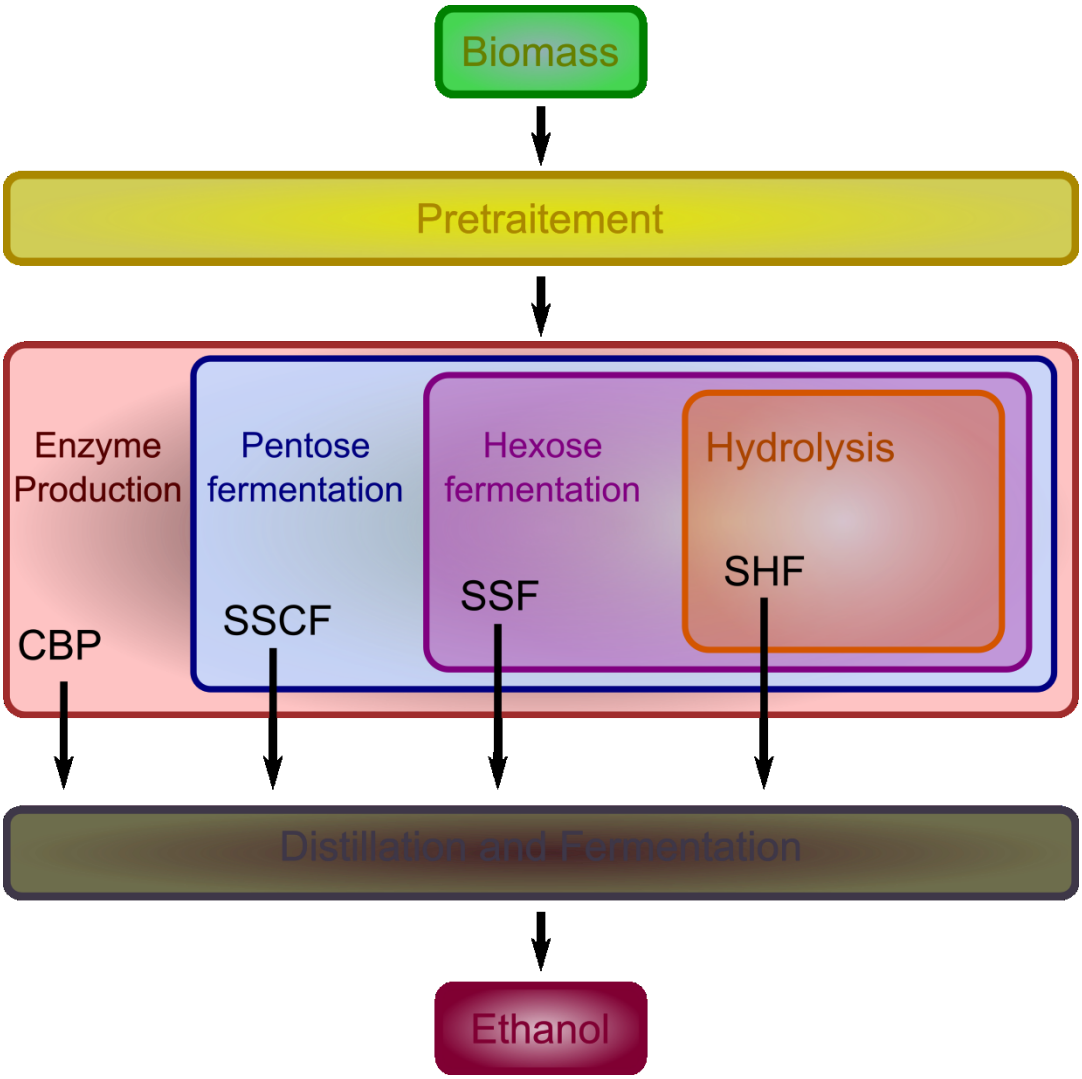

Figure 30 Process difference to produce bioethanol from lignocellulosic feedstock with biochemical technologies. Separate Hydrolysis & Fermentation (SHF), Simultaneous Saccharification & Fermentation (SSF), Simultaneous Saccharification & Co-Fermentation (SSCF) and Consolidated BioProcessing (CBP).

Another way to convert biomass to ethanol using Clostridia is investigated. The biomass is transformed to syngas with gasification, then the syngas is fermented as presented in the previous chapter with *C. autoethanogenum* [407], Lanzatech is developing this technology.

I.3.2.d C. phytofermentans in consolidated bioprocessing

Clostridium phytofermentans has been studied as a promising strain for CBP. The specificity of *C. phytofermentans* compare to other candidate is the secretion of free cellulolytic enzymes, instead of enzymes attached to a cellulosomes [408]. Moreover, *C. phytofermentans* was tested in laboratory conditions for CBP, alone [28][29], or in co-culture with S*. cerevisiae*. All previous studies on *C. phytofermentans* participate to understand and develop this strain for

CBP: *C. phytofermentans* must have a better tolerance to inhibitors resulting of the lignocellulose pretreatment and degradation, resist to increasing concentration of ethanol, must degrade and transport more polysaccharides, understanding of its metabolism must be improved and have different efficient genetic tool to become a powerful strain for CBP.

I.3.3 3rd Generation of biofuels

The third and the fourth generation of biofuels use the CO or CO_2 fixation as a main carbon source. It can be achieve with microalgae or bacteria. The negative conclusion of the feasibility of the algae technology for biofuel production during the 1980s and 1990s decreases the effort on this technology [378][409]. But interesting result on the research at the beginning of the 2000s, related to photo-bioreactor process technology, reactivates the private and public funding to develop a future production of biofuels. In algae, the large production of lipids and carbohydrates (up to 60% of the biomass), can be processed to biodiesel, bioethanol, biogas, hydrogen or valuable co-products [410], Figure 31. Microalgae have several advantages; it should not compete with food or feed production, uses non arable land, and needs mainly light and water. This technology can be implemented in a desert, and waste water can be used and algae grow 20-30 times faster than food crops with a higher production of biofuel sources. Disadvantages: it is not economically viable yet, mainly because microalgae need an important quantity of fertilizer (nitrogen, potassium and phosphorus) to grow and produce biofuels. Moreover the production of fertilizer release large amount of greenhouse gas [411].

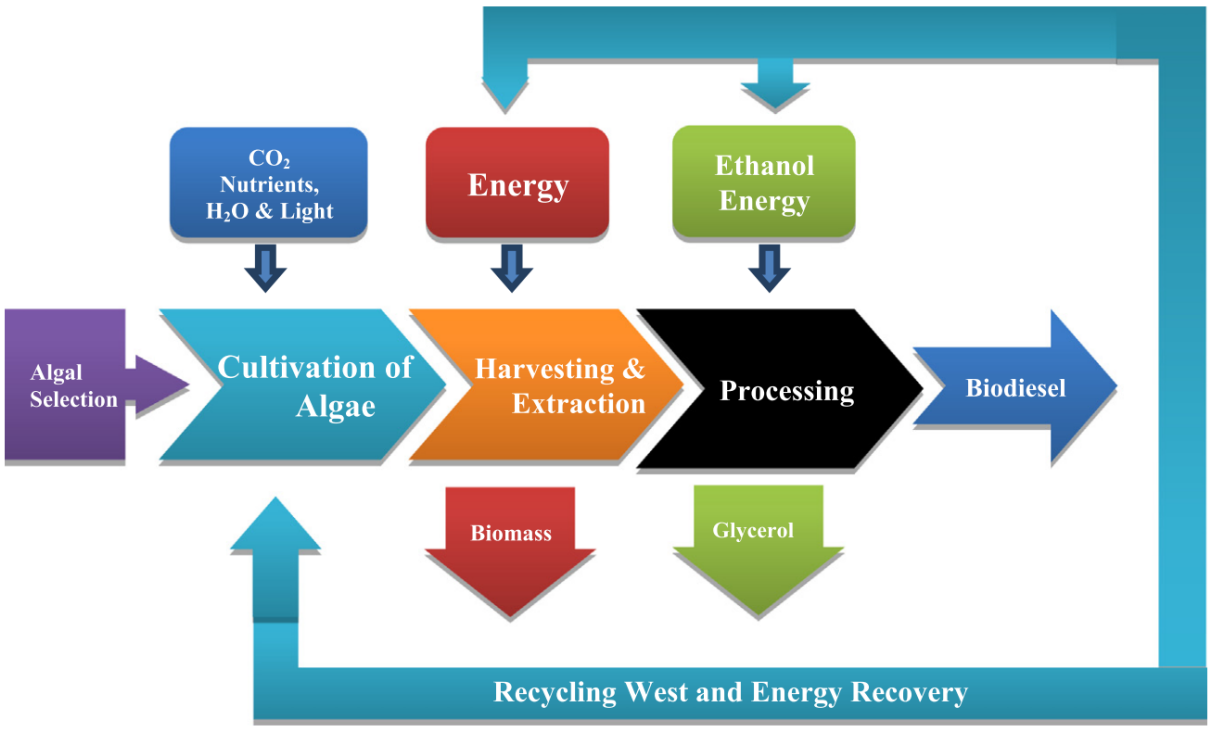

Figure 31 Schematic presentation of various steps of algal biodiesel [412].

Most of research on microalgae are focusing on hydrogen or biodiesel production from lipids[412], but the fermentation of resulting starch by 1st generation plants and polysaccharides by 2nd generation plants is not well developed yet [413]. Combining different technologies to harvest and increase the value of all compounds of microalgae could improve the economic efficiency of this technology, using fewer lands.

I.3.4 4th Generation of biofuels

The fourth generation can represent two technologies. The first one is developing CO_2 or CO fixation mainly using modified bacteria which convert this carbon source as biomass. The second technology, which is compatible (and complementary) with the first one, try to engineer microorganisms to directly produce higher/drop-in biofuels from cheap, widely available and inexhaustible raw materials [414].

The fixation of carbon monoxide or $C0_2$ can also be achieved with bacteria. For example *Clostridium autoethanogenum* or *Clostridium ljungdahlii* (these strains are described in the section I.1.3.b) are able to ferment syngas to ethanol [90][415]. Syngas, as a carbon source, can come from different sources: gasification of biomass, CO_2 recovery platform from heavy industry (including steel industry, a high CO_2 producer), inorganic CO_2, solid waste, transport etc...

Lanzatech, a biotech company, is using *C. autoethanogenum* [416] to industrialize this fermentation to produce biofuels. This company is improving *C. autoethanogenum via* synthetic biology tools: directed evolution, new genome editing tools to modify the strain at the genomic level, modelling of *C.* autoethanogenum, integration of heterologous metabolic pathways. All these tools are used to improve syngas fixation and the production of ethanol and future higher biofuels like butanol, isopropanol, farnesene or other chemicals [417].

Synthetic biology is an emerging new field in biological science; it is more a combination of technologies. One definition from Nature.com: "Synthetic biology is the design and construction of new biological parts, devices, and systems, and the re-design of existing, natural biological systems for useful purposes". One goal is to apply an engineering philosophy to better understand biological entities and to improve them for different purposes. Applications are developed for medicine, pharmacy, biofuels, bioproduction of chemicals etc… [418]. Several existing fields or technology are merged: Mathematics, physics, chemistry, informatics, biology, engineering, humanities, ethics. For instance, biological results are modelled with mathematics and computer science, running the model help the researcher to improve and modify parameters of the biological object. This object is modified and the new experiment give new data to improve the model, an *in silico* prediction of the behavior of the biological entity. It is an virtuous circle. New tools had raised the emergence of synthetic biology: omics (genomics, transcriptomics, proteomics and metabolomics). The global understanding of biological entities, named systems biology, is closed to synthetic biology, but with different objectives. Synthetic and systems biology are linked. Moreover, tools used and developed for these two fields are shared: modeling, directed evolution, omics, genome engineering, metabolic engineering….

This last generation of biofuels, utilize synthetic biology to improve the photon-to-fuel conversion efficiency. It would allow the development of photobiological solar fuel. Another technology develop the electron capture from electrodes: microbial electrosynthesis [414]. These two researches are promising, but have begun recently. Except the CO/CO_2 fixation *via* homoacetogens, only preliminary researches are developed, industrial applications are not available yet.

II. RESULTS

II.1 Sugar transport study in Clostridium phytofermentans

II.1.1 Introduction

We decided to study sugar transport in *C. phytofermentans* after the analysis of RNA-seq data that we performed for the article related to CAZYme activity (Appendix 1, [51]). Transcriptomic data showed an up-regulation of ABC transporters in response to specific sugars, and tiny expression, and low, constitutive expression of the Phosphotransferase System PTS. It was reported in other Clostridia that the PTS is the main transporter system for at least glucose, galactose or cellobiose [419][420]. Although alternative sugar uptake was presented in other cellulolytic Clostridia [421][422], we wanted to investigate the role of the highly expressed ABC transporters in *C. phytofermentans*.

To verify our hypothesis that ABC transporter enables the transport of specific sugars, we inactivated ABC transporter genes by retargeting group II introns. Experiments with this reverse genetic tool were achieved and several mutants were obtained.

This study was recently submitted for publication. My participation to this work consisted of the following: I proposed, designed (with the supervision of Andrew Tolonen) and performed the experiments; I also trained one master student (Alba Iglesias) to retarget more transporters, and I wrote the manuscript with Andrew Tolonen; Sanitha, Mary and Magali Boutard helped us with plasmid transformation, Marcel Salanoubat supervised the project. I did an oral presentation of this work at the Clostridium XIV conference in 2016 at Dartmouth College, NH, USA.

II.1.2 Article

ABC transporters required for uptake of hexose carbohydrates in Clostridium phytofermentans

Tristan Cerisy1,2,3,4*, Alba Iglesias3,4, Magali Boutard1,2, Mary Sanitha5, Marcel Salanoubat1,2,3,4, Andrew C Tolonen1,2,3,4*

1 CEA, Genoscope, 2 rue Gaston Crémieux, 91057 Evry, France

2 CNRS-UMR8030, Évry 91000, France

3 Université Paris-Saclay, Évry 91000, France

4 Université d'Évry, Évry 91000, France

5 Department of Genetic Engineering, SRM University, Tamil nadu 603203, India.

*correspondence: atolonen@genoscope.cns.fr or tristan.cerisy@gmail.com

Abstract

Plant-fermenting Clostridia are critical to the functioning of soil and intestinal ecosystems, and can be applied for industrial transformation of biomass. Here we identify the transporters for uptake of plant hexoses using functional genomics and targeted gene inactivation in *Clostridium* (*Lachnoclostridium*) phytofermentans, a mesophile that ferments lignocellulosic biomass. We analyze the transcription patterns of its more than 170 annotated sugar transporter genes to find those specifically up-regulated on different carbon sources. Inactivation of these genes reveals that single, non-redundant ATP-binding cassette (ABC) transporters are required for uptake of hexoses and hexo-oligosaccharides, and that distinct ABC transporters are used for oligosaccharides versus their constituent monomers. As sugars are not phosphorylated during ABC transport, we identify the intracellular glucose and galactose kinases based on in vitro activities. This study provides mechanistic insights needed to understand the energetics and metabolism of sugars in Clostridia.

Importance

Fermentation of plant biomass by bacteria such as *Clostridium phytofermentans* is important to recycle fixed carbon in the biosphere, provide nutrients in the animal intestine, and process cellulosic biomass in industry. Plants are mostly composed of polysaccharides comprising cellulose, hemicellulose, and pectin, which these bacteria hydrolyze into oligo- and

monosaccharides for transport into the cell. Here, we identify the transporters used *by C. phytofermentans* to uptake abundant hexoses and hexo-oligosaccharides in plant biomass. Even though this bacterium has hundreds of transporter genes, our results show that single ABC transporters are required for the uptake carbon sources. These results reveal mechanisms of sugar transport and metabolism in Clostridia that have implications for engineering strains with expanded sugar metabolism.

Introduction

Fermentation of lignocellulosic biomass by *Clostridium phytofermentans* (1) and other bacteria is central to the functioning of terrestrial and aquatic ecosystems. *C. phytofermentans* is an anaerobic mesophile that metabolizes numerous plant polysaccharides including cellulose, hemicellulose, and pectin (2) (3) (4). This bacterium expresses numerous carbohydrates-active enzymes (CAZymes) (5) to degrade plant polysaccharides into hexoses and pentoses, which are taken into the cell using panoply of transporters. *C. phytofermentans* is predicted to encode 572 transporter genes, including 173 genes for sugar transport (6) (Table S1), but the transporters responsible for uptake of the specific saccharides are unknown. While general predictions about substrate specificity of a transporter can be made based on sequence homology or structure based modeling (7), experimental approaches are generally needed to determine the substrate of a transporter.

Bacteria translocate carbon sources by five different mechanisms (8). The simplest is a uniporter that functions by facilitated diffusion without energy expenditure, such as for glycerol transport in *E. coli* (9). Second, hexose uptake by the phosphotransferase system (PTS) involves transfer of a phosphate from phosphoenolpyruvate (PEP) such that the sugar is both transported and phosphorylated using a single ATP equivalent, providing critical energy savings to anaerobic bacteria. Clostridial species vary in their encoding of PTS systems*: C. acetobutylicum* (10) and *C. beijerinckii* (11) translocate sugars using many PTS, *C. phytofermentans* has a single PTS predicted to transport glucose, and *C. cellulolyticum* (12) and *C. thermocellum* (13) appear to lack functional PTS. Third, bacteria use symporters to assimilate various molecules including hexoses by coupling their uptake with H+ or Na+ molecules (14); *C. phytofermentans* is annotated as encoding 16 symporters. Fourth, ATP-binding cassette (ABC) transporters that power sugar translocation with ATP hydrolysis are widely used in Clostridia; the *C. phytofermentans* genome contains 158 genes predicted to encode sugars ABC transporters. Fifth, some bacteria import phosphorylated sugars using antiporters that couple sugar uptake with export of inorganic phosphate (15); *C.*

phytofermentans has 8 antiporter genes predicted to exchange H+ with dicarboxylate or Na+ ions.

The goal of this study is to examine how plant-fermenting bacteria like *C. phytofermentans* metabolize biomass by identifying the mechanisms for uptake of plant-derived hexo-saccharides. We focus on hexoses and hexo-saccharides abundant in the three main classes of plant polysaccharides: cellulose (glucose), hemicellulose (galactose), and pectin (galacturonic acid). We examine mRNA expression on a panel of 10 plant substrates to identify differentially-expressed transporter genes, which we inactivate by targeted insertion of group II introns (16). We compare growth of wild-type and mutants on different carbon sources to determine the roles of these transporter genes in sugar uptake. As our results show that glucose and galactose are not phosphorylated during transport, we identify and characterize the activities of the intracellular gluco- and galactokinases. Finally, we discuss implications of our results for understanding carbon and energy metabolism in Clostridia.

Results

Identification of putative sugar transporters

The *C. phytofermentans* genome is predicted to encode 572 transporter genes by TransportDB (6) and 485 transporter genes by KEGG (17); 438 genes are common to both databases (Table S1). Among the 173 genes annotated by TransportDB as sugar transporters, 158 genes encode ABC transporters. ABC transporters are comprised of three components: the extracellular binding protein captures the substrate, the transmembrane domain (TMD) protein translocates the substrate across the membrane, and the nucleotide-binding domain (NBD) protein provides the energy by ATP hydrolysis (18). *C. phytofermentans* genes encoding an ABC transporter are generally co-transcribed as an operon to facilitate their expression at similar levels (19).

We observed that distinct ABC transporter genes are transcriptionally up-regulated on each carbon source (Fig 1A, Table S2). A single ABC transporter is expressed at a much higher level than all others on each of the 3 hexoses: glucose and galactose (*cphy2241-3*), and galacturonic acid (*cphy2731-3*). While the PTS is responsible for hexose transport in other clostridia, the genes encoding the PTS in *C. phytofermentans* (Fig S1) are constitutively expressed at very low levels (Table S2), casting doubt on their physiological importance. The *cphy2464-6* operon is most highly expressed transporter on cellobiose, a di-saccharide that is the main product of cellulose degradation by the family 48 cellulase in this organism (20). Among polysaccharides, *cphy3588-90* is highest on galactan and a few ABC transporters are transcribed at similarly high levels on cellulose (*cphy3858-60* and *cphy2464-6*) and homogalacturonan (*cphy2731-3* and

cphy3588-90) (Fig 1A). Collectively, gene expression suggests that a small subset of the numerous transporter genes could play key roles in hexose uptake.

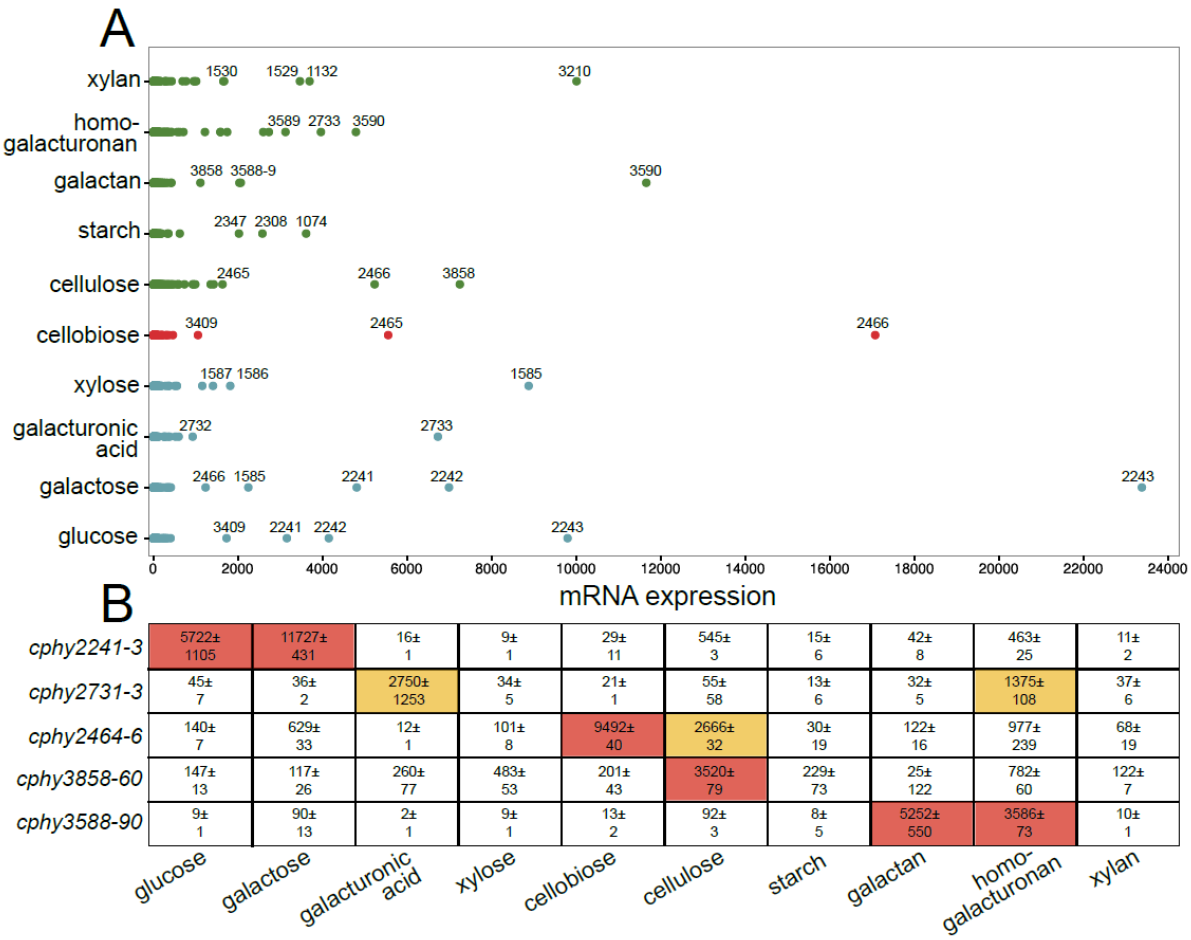

Fig 1 (A) *C. phytofermentans* up-regulates transcription of specific ABC transporters on each carbon source. The mRNA expression (RNA-seq RPKM) of 158 putative sugar ABC transporter genes are shown on monosaccharides (blue), disaccharides (red), and polysaccharides (green). NCBI gene numbers are shown for highly-expressed genes. (B) Expression of the 5 transporters selected for targeted inactivation. The average mRNA expression (mean RPKM± s.d.) of the 3 genes in each transporter operon is shown with expression >3000 highlighted in red and >1000 in yellow.

Gene inactivation and growth of transporter mutants

We examined the function of ABC transporters by targeted gene inactivation. Specifically, we inserted a designed LI.LtrB group II intron into the coding sequence of the transmembrane domain (TMD) gene (Fig 2A), thereby disrupting translation of the mRNA into a functional protein. We targeted TMD genes associated with 5 ABC transporters whose transcription were elevated in response to plant hexoses (Fig 1B): *cphy2241* (glucose and galactose), *cphy2732* (galacturonic acid and homogalacturonan), *cphy2465* (cellobiose and cellulose), *cphy3589* (galactan and homogalacturonan), and *cphy3859* (cellulose).

Following delivery of the group II intron into *C. phytofermentans*, its chromosomal integration into the TMD gene was shown by PCR using primers flanking the programmed insertion site (Fig 2B). Sequencing the PCR products confirmed all introns inserted at the expected genome position. Strains were names to specify the target gene, distance of the intron insertion from the gene start, and an "a" to denote an antisense insertion. Examination of the number of genomic insertion sites in the mutant strains by inverse PCR showed *cphy2241::int164a*, *cphy2732::int662a*, and *cphy3589::int332a* have single genomic insertions at the expected sites (Fig 2C). However, *cphy2465::int293a* and *cphy3859::int586a* each had an additional genomic intron insertion after curing of pQint. As *cphy2465::int293a* had no off-target insertions before transferring the culture to lose pQint (Fig 2C), we used this culture for growth assays. The *cphy3589::int332a* strain had a second genomic insertion in *cphy3242*, an ABC transporter gene that is lowly expressed on all carbon sources (Table S2). The *cphy3242* insertion was present immediately after repeated conjugations, so a culture with this double insertion and pQint was used for growth experiments.

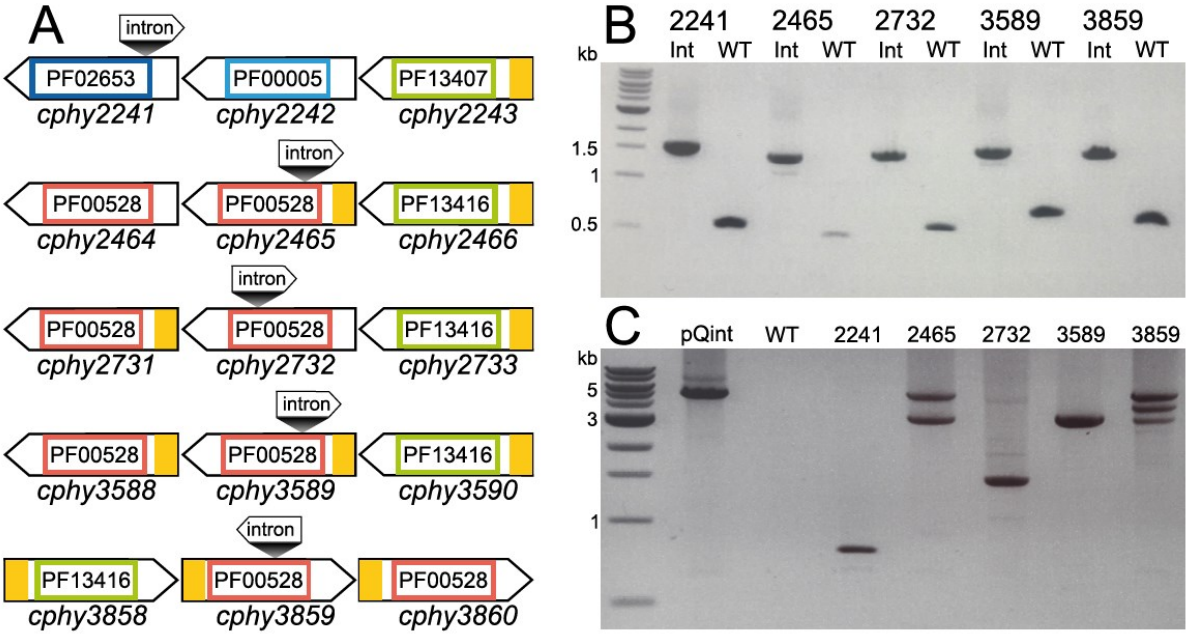

Fig 2 Targeted inactivation of ABC transporter genes. (A) Genomic organization of ABC transporter gene operons and intron insertion sites. Genes show predicted signal peptides (yellow) and PFAM domains [423]: extracellular solute binding (green), ATP-binding (light blue), permease (dark blue), and transmembrane (red). (B) PCR with primers flanking intron insertion sites of target genes in wild-type (WT) and intron insertion strains (Int) confirm insertion of the 1040 bp intron in Int strains. (C) Inverse PCR of genomic DNA from WT and intron insertion strains shows products of expected size for *cphy2241::int164a* (775 bp), *cphy2465::int293a* (3125 bp), *cphy2732::int662a* (1485 bp), *cphy3589::int332a* (3202 bp), and *cphy3859::int586a* (3875 bp). The 5035 bp product from pQint is shown for comparison to the plasmids remaining in the *cphy2465* and *cphy3859* strains. The *cphy3859::int332a* also has a second genomic insertion in *cphy3242*.

We compared growth of *C. phytofermentans* transporter mutants and WT on monosaccharides (glucose, galactose, galacturonic acids), disaccharides (cellobiose) and polysaccharides (galactan). The *cphy2241::int164a* strain completely lost the ability to grow on glucose and galactose and has a small growth deficit on galactan (Fig 3A-F), likely resulting from WT consuming galactose that is residual in galactan or liberated by extracellular galactanase activity. The *cphy2465::int293a* strain specifically lost the ability to grow on cellobiose (Fig 3 G-L), whereas the *cphy3859::int586a* grew normally on all 6 substrates (Fig S2). We compared the cellulose degradation rates of the *cphy2241*, *cphy2465*, and *cphy3859* inactivation strains relative to WT. Initially, cellulose degradation was reduced by 50-75% in all three strains relative to WT (Fig 3M). By day 14, *cphy3859::int586a* re-established cellulose degradation similar to WT, while the other 2 strains still had degraded 25% less cellulose than WT (Fig 3M). We thus propose that extracellular cellulases depolymerize cellulose into a mix of cellodextrins, cellobiose, and glucose that are simultaneously uptaken by the Cphy3858-60, Cphy2464-6, and Cphy2241-3 transporters, respectively. Initially, abundant cellodextrins lead to reduced growth of *cphy3859::int586a* relative to WT. As cellulases accumulate in the medium, cellobiose and glucose become the dominant products, enabling *cphy3859::int586a* to establish growth similar to WT, but resulting in continued growth deficits in the *cphy2465::int293a* and *cphy2241::int164a* strains.

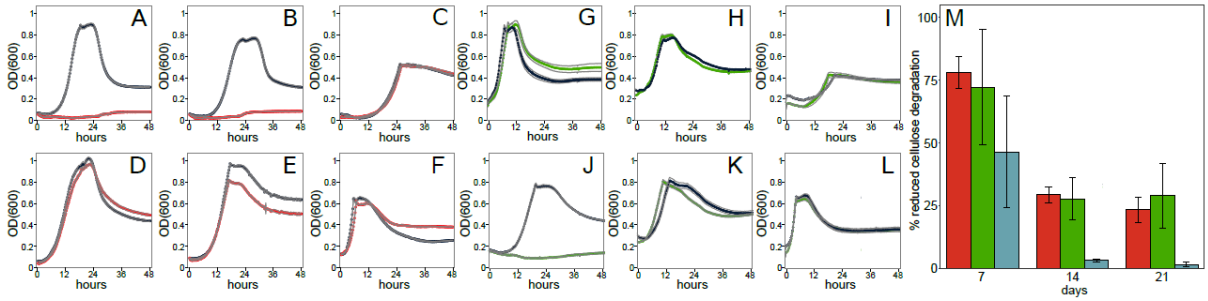

Fig 3 Growth (OD_{600}) on different carbon sources of *C. phytofermentans* WT (dark blue) compared to (A-F) *cphy2241::int164a* (red) and (G-L) *cphy2465::int293a* (green). Carbon sources: glucose (A,G), galactose (B,H), galacturonic acid (C,I), cellobiose (D,J), galactan (E,K), homogalacturonan (F,L). (M) Percent reduction in cellulose degradation relative to WT for *cphy2241::int164a* (red), *cphy2465::int293a* (green), and *cphy3859::int586a* (blue). All data is mean of quadruplicate cultures $\pm$ s.d.

The *cphy3589::int332a* strain cannot grow on galactan (Fig 4A-F), but surprisingly grows normally on homogalacturonan even though *cphy3588-90* are the most highly expressed transporter genes on this substrate. The high expression of *cphy3588-90* on homogalacturonan suggests that a homogalacturonan-associated metabolite inappropriately triggers expression of

the galactan transporter. Growth on homogalacturonan requires the Cphy2731-3 transporter as evidenced by the inability of *cphy2732::int662a* to grow on this substrate (Fig 4L). While *cphy2731-3* are the most highly expressed transporter genes on galacturonic acid, the *cphy2732::int662a* strain grows normally on galacturonic acid (Fig 4G-K). Transcription of transporter genes reveals no obvious candidate for uptake of galacturonic acid (Fig 1, Table S2).

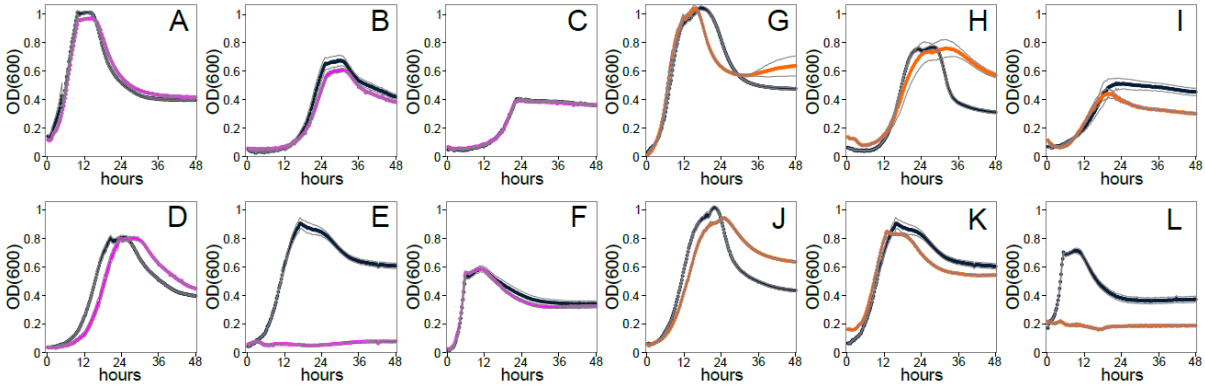

Fig 4 Growth (OD600) comparison on different carbon sources of C. phytofermentans WT (dark blue) to (A-F) *cphy3589::int332a* (magenta) and (G-L) *cphy2732::int662a* (orange). Carbon sources: glucose (A,G), galactose (B,H), galacturonic acid (C,I), cellobiose (D,J), galactan (E,K), homogalacturonan (F,L). All data is mean of quadruplicate cultures ± s.d.

Activities of glucose and galactose kinases

Hexoses must be phosphorylated in order to be metabolized. Unlike the PTS, ABC transporters do not phosphorylate sugars during transport, requiring them to be phosphorylated by cytoplasmic kinases. Glucose is phosphorylated to glucose-6-phosphate by glucokinase. Galactose is phosphorylated to galactose-1-phosphate by galactokinase. Galacturonic acid is isomerized to 3-Deoxy-D-erythro-hex-2-ulosonate and then phosphorylated by 2-keto-3-deoxy-D-gluconate kinase (21), (19). In *C. phytofermentans*, the *cphy2241-3* genes encoding the glucose/galactose ABC transporter are co-localized in the genome with a hexokinase gene *cphy2237*, leading us to question whether this enzyme phosphorylates glucose, galactose, or both. We purified a His-tagged Cphy2237 (Fig S3A) and quantified its kinase activity on glucose and galactose (Fig S2C and E), revealing kinase activity (Km, Kcat) on galactose similar to galactokinases from *E. coli*, *S. cerevisiae* and *H. sapiens* (Tab 1), but with undetectable glucokinase activity. We subsequently determined another putative hexokinase *Cphy0329* (Fig S2A, B and D) phosphorylates glucose with activity (Km, Kcat) similar to glucokinases from other organisms (Tab 1), but has negligible galactokinase activity. Thus, the glucose/galactose ABC transporter genes are co-located in a gene island with galactokinase, whereas glucose phosphorylation requires an enzyme located elsewhere in the genome.

Table 1

Enzyme type	Species	Km (mM)	Kcat (s^{-1})	Kcat/Km ($mM^{-1}.s^{-1}$)	references
GALK (Cphy0329)	*C. phytofermentans*	0.26	273.80	1047.84	This study
GALK	*E. coli*	0.78	91.26	117.00	Meyer *et al.*, 1997
GALK	*T. cruzi*	1.00	1491.67	1491.67	Cáceres *et al.*, 2007
GALK	*L. major*	3.30	316.67	95.96	Cáceres *et al.*, 2007
GALK	*S. cerevisiae*	0.03	18.93	631	Verhees *et al.*, 2002/Maitra, 1970
GALK	*H. sapiens*	0.065	26.72	411	Verhees *et al.*, 2002/Zeng *et al.*, 1998
GLK (Cphy2237)	*C. phytofermentans*	1.19	176.33	147.63	This study
GLK	*E. coli*	0.70	9.66	13.80	Sherman *et al.*, 1963
GLK	*S. cerevisiae*	0.60	53.94	89.90	Verhees *et al.*, 2002/Lavine *et al.*, 1982
GLK	*H. sapiens*	0.12	68.16	568.00	Verhees *et al.*, 2002/Blume *et al.*, 1971

Tab 1 Comparison of glucokinase and galactokinase kinetics of *C. phytofermentans* enzymes with those from other species.

Discussion

Our results show *C. phytofermentans* relies on individual ABC transporters for growth on several of its preferred hexose substrates (glucose, galactose, cellobiose, galactan, and homogalacturonan), revealing an unexpected lack of functional overlap among its nearly 200 sugar transporter genes. It remains to be seen if *C. phytofermentans* encodes non-redundant transporters for a massive diversity of saccharides, many auxiliary transporters that improve but cannot themselves support growth, or functionally redundant transporters for substrates not examined in this study. Also unexpected is that while ATPase activity is essential for ABC transport (18), only one of the required hexose transporters (Cphy2241-3) has an ATPase domain (Fig 2A), supporting promiscuous ATPase subunits are shared among transporters (22). While most carbon sources are translocated by a single, highly-expressed ABC transporter, the most highly expressed transporter on galacturonic acid (Cphy2731-3) is required for growth on homogalacturonan (Fig 4L), but not galacturonic acid (Fig 4I). No other transporters are specifically up-regulated on galacturonic acid, which could account for the lower cell yield relative to glucose and galactose. We similarly did not find a single transporter required for metabolism of cellulose. Rather, *C. phytofermentans* up-regulated operons for multiple transporters (*cphy2464-6*, *cphy3858-60*, and *cphy2241-3*) that each contribute to cellulose metabolism (Fig 3M), supporting extracellular cellulases degrade cellulose into a range of chain lengths that are simultaneously uptaken by distinct ABC transporters.

We assimilated our results into a model of how *C. phytofermentans* transports and metabolizes plant-derived hexoses (Fig 5). Extracellular CAZymes depolymerize cellulose, galactan, and homogalacturonan into a mix of oligo- and monosaccharides that are translocated by substrate-specific ABC transporters. Uptake of oligosaccharides conserves ATP (23) and likely prevents

competing microbes from filching free monosaccharides, but necessitates intracellular CAZymes to split the oligosaccharides. Based upon mRNA expression and sequence homology, we propose galacto-oligosaccharides are cleaved by the GH2 β-galactosidase Cphy0711 and galacturono-oligosaccharides by the GH4 α-galacturonase Cphy2848. Cellodextrins and cellobiose are likely split by three GH94 phosphorylases: Cphy1929, Cphy0430, and Cphy3854. Expression profiles support Cphy0430 transports cellobiose and cellulose, whereas Cphy3854 is specific to cellulose. Cphy1929 is expressed at lower levels, but is active on cellodextrins in vitro (3).

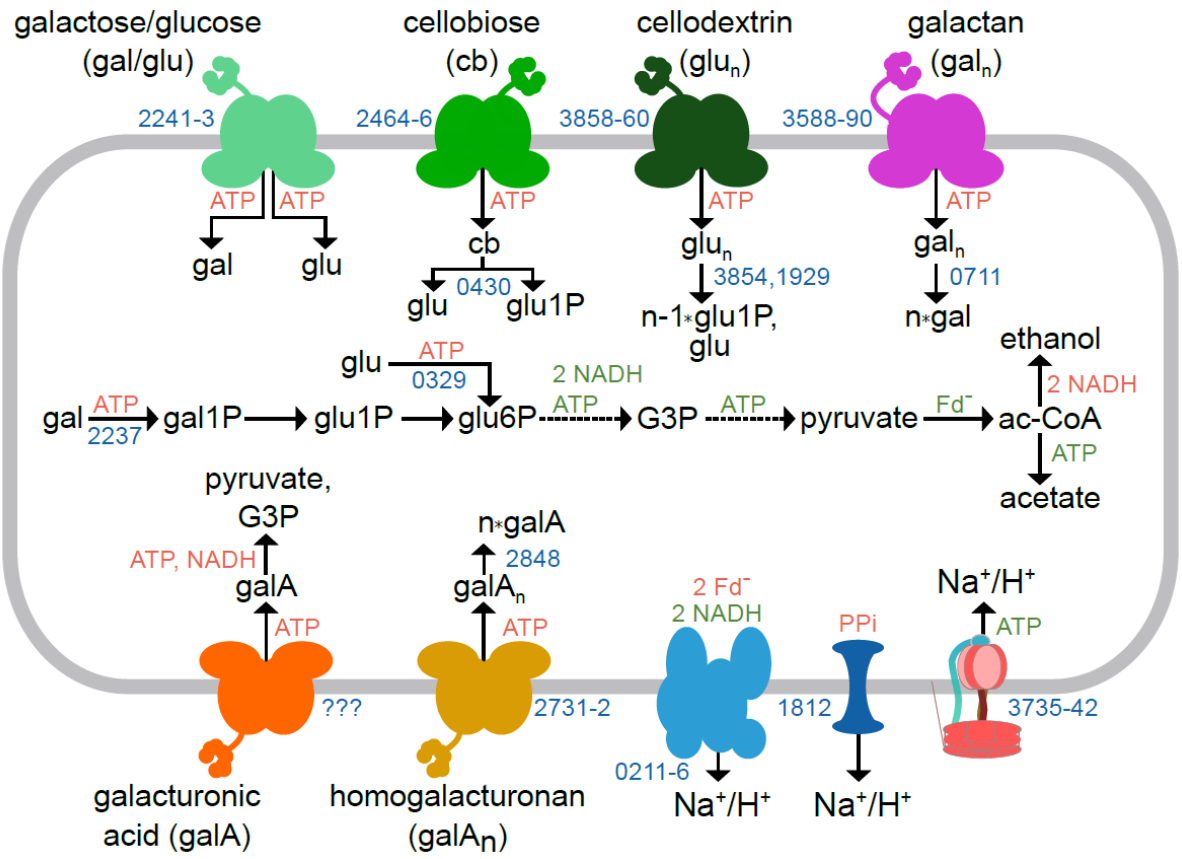

Fig 5 Model of *C. phytofermentans* hexose transport and metabolism. ABC transporters and metabolic enzymes are shown with NCBI numbers (blue) and associated production (green) or consumption (red) of ATP, NADH, and ferredoxin. Abbreviations: glucose-1-phosphate (glu1P), glucose-6-phosphate (glu6P), galactose-1-phosphate (gal1P), glyceraldehyde-3-phosphate (G3P), acetyl-CoA (ac-CoA), ferredoxin (Fd), pyrophosphate (PPi).

The use of ABC transporters for hexose transport by *C. phytofermentans* contrasts with *E. coli* and *Bacillus subtilis*, and even solventogenic Clostridia such as *C. acetobutylicum* and *C. beijerinckii*, which all use PTS for hexose uptake (24). Expression and in vitro evidence supports that other cellulolytic Clostridia similarly uptake hexoses with ABC transporters (25), (26). ABC transport of a sugar molecule requires 1 or 2 ATP (27), (28) and another ATP for intracellular phosphorylation, whereas PTS uptake and phosphorylate sugars using a single ATP equivalent.

In addition, the PTS indirectly facilitates further ATP savings during galactose metabolism in *C. acetobutilicum*, which converts galactose-6-P from the PTS into glyceraldehyde-3-P by the tagatose pathway (29), thereby saving another ATP relative to the Leloir pathway used *by C. phytofermentans*. ABC transport imposes significant cost as transport consumes to 60% of cellular ATP (18). Moreover, *C. phytofermentans* encodes a putative glucose PTS whose low expression supports that it is not functional. This seemingly non-adaptive preference in cellulolytic clostridia may result from ABC transporters having a higher substrate affinity to scavenge sugars in carbon-poor environments. In addition, genes for ABC transporters are often co-located with glycoside hydrolases in the *C. phytofermentans* genome (4), supporting gene cassettes for polysaccharide degradation and ABC transport are co-acquired by horizontal gene transfer.

Metabolism of a molecule of glucose or galactose to pyruvate yields only 1 ATP (Fig 5), yet fermenting these carbon sources produces high ethanol:acetate ratios (2), (4), supporting the need for NADH oxidation outweighs that of ATP production. Assimilation of galacturonic acid to pyruvate is net negative 1 ATP (Fig 5), and concomitantly the primary fermentation product shifts to acetate (19), which boosts ATP production by 2 ATP per sugar equivalent. *C. phytofermentans* can produce additional ATP using the F1F0-ATPase (Cphy3735-42) to harness the ion gradient generated by the Rnf complex (Cphy0211-16), which couples ferredoxin oxidation to NAD+ reduction to pump Na+ or H+ outside the cell (Fig 5). The high expression of these complexes supports their importance is similar to those in some other clostridia (30). *C. phytofermentans* also likely uses a membrane pyrophosphate to augment the Na+ and H+ gradient (31). Together, these ATP-generating adaptations enable cellulolytic clostridia maintain cellular ATP levels while transporting carbon using ABC transporters.

Elucidating how plant-fermenting bacteria uptake and metabolize sugars is important to understand the functioning of soil and intestinal microbial communities and to develop industrial strains for processing of plant biomass. This approach combining functional genomics and targeted gene inactivation can be generally applied to assign function to the numerous transporters encoded by Clostridia. Further, while most efforts to engineer microbes to utilize plant polysaccharides focus on degradative enzymes, our results show oligosaccharide-specific transporters are also required. Transfer of gene cassettes including both CAZymes and the associates transporters may help overcome difficulties often encountered when engineering novel polysaccharide capabilities into bacteria. Generally, understanding the mechanisms of sugar transport will facilitate strategies for metabolic engineering to improve substrate utilization and fermentation efficiency in these organisms.

Methods

Bacterial cultivation

C. phytofermentans ISDg (ATCC 700394) was cultured anaerobically at 37°C in GS2 medium (32) containing 5 g l-1 of the appropriate sugar. Growth on different carbon sources was quantified in log phase cultures that were resuspended in GS2 without sugars and inoculated at 1:10 volume into 100-well microtiter plates (Bioscreen 9502550) sealed in an anaerobic chamber (2% H2, 98% N2) by press-fitting adhesive pads (Qiagen 1018104) (33). Growth (OD600) was measured using a Thermo Scientific Bioscreen C at 37°C with 30 seconds shaking before readings each 15 minutes.

RNA sequencing

Total RNA was extracted from log phase *C. phytofermentans* galactose cultures using TRI reagent (Sigma 93289) and treated at 37ºC for 30 min with 0.2 units Turbo DNase (Ambion AM2238) per µg RNA. RNA was purified by Zymo Concentrator-5 (Zymo Research R1015) (>200 bp capture) into 15 µl water. Five µg total RNA was rRNA depleted by Ribo-Zero (Illumina MRZMB126) and purified by Zymo Concentrator-5 (total capture). cDNA libraries were prepared from 100 ng RNA using the Truseq Stranded mRNA kit (Illumina 15031047) and sequenced on an Illumina MiSeq with paired-end 150 bp reads. Reads were aligned to the *C. phytofermentans* ISDg genome (NCBI NC_010001.1) using Bowtie 2 (34). RNA-seq mRNA expression was calculated as reads per kilo base per million mapped reads (RPKM) using the Bioconductor 'easyRNASeq' package (35) (Table S2). Gene expression measurements from galactose cultures were compared with RNA-seq from 9 other carbon sources, which we previously prepared using the same methods (3).

Targeted gene inactivation

Transporter genes in the *C. phytofermentans* genome were inactivated by targeted insertion of group II introns. The Targetron algorithm (Sigma-Aldrich TA0100) was used to design primers (Table S3) to insert the intron in the antisense orientation relative to gene transcription at the site closest to the gene start. The targeted intron was PCR amplified and inserted between the NdeI and BsrGI sites of pQint (36). Targeted pQint plasmids were transferred into *C. phytofermentans* by conjugation with *Escherichia coli* (36) (37) (38) under anaerobic conditions. Colonies on GS2 plates containing the 40 µg ml-1 erythomycin were picked and the genomic intron insertion was confirmed by PCR and sequencing using primers in Table S2. pQint plasmids were cured by five successive transfers at 1:50 volume into GS2 medium lacking antibiotics. The number of intron insertions in the *C. phytofermentans* genome was then examined by inverse PCR (39), whereby genomic DNA was extracted from 5 ml cultures using the Sigma GenElute Bacterial Genomic DNA kit (NA2110), digested with HindIII (NEB R0104),

and ligated with T4 DNA ligase (NEB M0202). The genomic sequence flanking the intron was amplified by PCR with Q5 polymerase (NEB M0491) with a 7 minute elongation time using outward-facing primers annealing to the intron sequence (Table S3). Genomic sites of intron insertion were determined by sequencing the intron-genome junction regions in the PCR products.

Activities of purified kinases

The *cphy2237* and *cphy0329* genes encoding putative sugar kinases were cloned with C-terminal His-tags into pET-22B(+) (Novagen 69744) by Ligation-Independent Cloning (40) (primers in Table S3). Gene sequences were confirmed by sequencing and plasmids were transformed into *E. coli* BL21(DE3) (Novagen 70235). Expression was induced by adding 500 µM IPTG to cultures at OD600=1.0 in 50 ml TB medium and incubating overnight at 20°C. Cells were pelleted, resuspended in 5 ml of lysis buffer (50 mM phosphate buffer pH=8, 0.5M NaCl, 10 mM imidazole, 15% glycerol, 1 mM pefabloc (Sigma 76307)), and cells were lysed by sonication (Cole-Parmer Vibracell CV33) in the presence of lysozyme (Novagen 71230). His-tagged proteins were purified from 50 ml culture on Ni-NTA spin columns (Qiagen 31014) and quantified by Bradford assay. Purified proteins were visualized on a 12% SDS-PAGE gel (Fig S2A).

Glucokinase and galactokinase activities were quantified in PMMA cuvettes (Sigma-Aldrich Z330388) as absorbance at 340 nm using a SAFAS UVmc1 spectrophotometer at room temperature. Glucokinase activities were measured as NADPH formation using initial glucose concentrations from 2.5 µM to 10 mM in 50 mM Tris-HCl (pH 7.5) containing 5 mM MgCl2, 1 mM ATP, 0.25 mM NADP+, 5 U ml-1 glucose-6-phosphate dehydrogenase (Sigma G7877), and 0.035 µg ml-1 Cphy0329 or 0.0345 mg ml-1 Cphy2237. Galactokinase activities were measured as NADH consumption using initial galactose concentrations from 2.5 µM to 25 mM in 50 mM Tris-HCl (pH 7.5) containing 0.3 mM NADH, 5 mM MgCl2, 1 mM ATP, 5 mM phosphoenolpyruvate, 10 U ml-1 pyruvate kinase (Sigma P1506), 23 U ml-1 lactate dehydrogenase (Sigma L2500), and 0.085 µg ml-1 Cphy2237 or 0.034 mg ml-1 Cphy0329.

References

1. Warnick TA, Methé BA, Leschine SB. 2002. Clostridium phytofermentans sp. nov., a cellulolytic mesophile from forest soil. Int J Syst Evol Microbiol 52:1155–1160.

2. Tolonen AC, Haas W, Chilaka AC, Aach J, Gygi SP, Church GM. 2011. Proteome-wide systems analysis of a cellulosic biofuel-producing microbe. Mol Syst Biol 7:461.

3. 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.

4. Petit E, Coppi MV, Hayes JC, Tolonen AC, Warnick T, Latouf WG, Amisano D, Biddle A, Mukherjee S, Ivanova N, Lykidis A, Land M, Hauser L, Kyrpides N, Henrissat B, Lau J, Schnell DJ, Church GM, Leschine SB, Blanchard JL. 2015. Genome and Transcriptome of Clostridium phytofermentans, Catalyst for the Direct Conversion of Plant Feedstocks to Fuels. PloS One 10:e0118285.

5. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490-495.

6. Ren Q, Chen K, Paulsen IT. 2007. TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. Nucleic Acids Res 35:D274-279.

7. Ecker GF, Stockner T, Chiba P. 2008. Computational models for prediction of interactions with ABC-transporters. Drug Discov Today 13:311–317.

8. Jojima T, Omumasaba CA, Inui M, Yukawa H. 2010. Sugar transporters in efficient utilization of mixed sugar substrates: current knowledge and outlook. Appl Microbiol Biotechnol 85:471–480.

9. Heller KB, Lin EC, Wilson TH. 1980. Substrate specificity and transport properties of the glycerol facilitator of Escherichia coli. J Bacteriol 144:274–278.

10. Voigt C, Bahl H, Fischer R-J. 2014. Identification of PTS(Fru) as the major fructose uptake system of Clostridium acetobutylicum. Appl Microbiol Biotechnol 98:7161–7172.

11. Essalem MEE, Mitchell WJ. 2016. Identification of a glucose-mannose phosphotransferase system in Clostridium beijerinckii. FEMS Microbiol Lett 363.

12. Abdou L, Boileau C, de Philip P, Pagès S, Fiérobe H-P, Tardif C. 2008. Transcriptional regulation of the Clostridium cellulolyticum cip-cel operon: a complex mechanism involving a catabolite-responsive element. J Bacteriol 190:1499–1506.

13. Nochur SV, Jacobson GR, Roberts MF, Demain AL. 1992. Mode of sugar phosphorylation in Clostridium thermocellum. Appl Biochem Biotechnol 33:33–41.

14. Iancu CV, Zamoon J, Woo SB, Aleshin A, Choe J. 2013. Crystal structure of a glucose/H+ symporter and its mechanism of action. Proc Natl Acad Sci U S A 110:17862–17867.

15. Island MD, Wei BY, Kadner RJ. 1992. Structure and function of the uhp genes for the sugar phosphate transport system in Escherichia coli and Salmonella typhimurium. J Bacteriol 174:2754–2762.

16. Karberg M, Guo H, Zhong J, Coon R, Perutka J, Lambowitz AM. 2001. Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria. Nat Biotechnol 19:1162–1167.

17. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 44:D457-462.

18. Rees DC, Johnson E, Lewinson O. 2009. ABC transporters: the power to change. Nat Rev Mol Cell Biol 10:218–227.

19. Boutard M, Ettwiller L, Cerisy T, Alberti A, Labadie K, Salanoubat M, Schildkraut I, Tolonen AC. 2016. Global repositioning of transcription start sites in a plant-fermenting bacterium. Nat Commun 7:13783.

20. Zhang X-Z, Zhang Z, Zhu Z, Sathitsuksanoh N, Yang Y, Zhang Y-HP. 2010. The noncellulosomal family 48 cellobiohydrolase from Clostridium phytofermentans ISDg: heterologous expression, characterization, and processivity. Appl Microbiol Biotechnol 86:525–533.

21. Richard P, Hilditch S. 2009. D-galacturonic acid catabolism in microorganisms and its biotechnological relevance. Appl Microbiol Biotechnol 82:597–604.

22. Mukherjee S, Thompson LK, Godin S, Schackwitz W, Lipzen A, Martin J, Blanchard JL. 2014. Population Level Analysis of Evolved Mutations Underlying Improvements in Plant Hemicellulose and Cellulose Fermentation by Clostridium phytofermentans. PloS One 9:e86731.

23. Muir M, Williams L, Ferenci T. 1985. Influence of transport energization on the growth yield of Escherichia coli. J Bacteriol 163:1237–1242.

24. Mitchell WJ. 2016. Sugar uptake by the solventogenic clostridia. World J Microbiol Biotechnol 32:32.

25. Nataf Y, Yaron S, Stahl F, Lamed R, Bayer EA, Scheper T-H, Sonenshein AL, Shoham Y. 2009. Cellodextrin and laminaribiose ABC transporters in Clostridium thermocellum. J Bacteriol 191:203–209.

26. Ravachol J, de Philip P, Borne R, Mansuelle P, Maté MJ, Perret S, Fierobe H-P. 2016. Mechanisms involved in xyloglucan catabolism by the cellulosome-producing bacterium Ruminiclostridium cellulolyticum. Sci Rep 6:22770.

27. Poolman B, Doeven MK, Geertsma ER, Biemans-Oldehinkel E, Konings WN, Rees DC. 2005. Functional analysis of detergent-solubilized and membrane-reconstituted ATP-binding cassette transporters. Methods Enzymol 400:429–459.

28. van der Does C, Tampé R. 2004. How do ABC transporters drive transport? Biol Chem 385:927–933.

29. Sund CJ, Servinsky MD, Gerlach ES. 2013. Differing Roles for Clostridium acetobutylicum's Galactose Utilization Pathways. Adv Microbiol 03:490.

30. Tremblay P-L, Zhang T, Dar SA, Leang C, Lovley DR. 2012. The Rnf complex of Clostridium ljungdahlii is a proton-translocating ferredoxin:NAD+ oxidoreductase essential for autotrophic growth. mBio 4:e00406-00412.

31. Luoto HH, Baykov AA, Lahti R, Malinen AM. 2013. Membrane-integral pyrophosphatase subfamily capable of translocating both Na+ and H+. Proc Natl Acad Sci U S A 110:1255–1260.

32. Cavedon K, Leschine SB, Canale-Parola E. 1990. Cellulase system of a free-living, mesophilic clostridium (strain C7). J Bacteriol 172:4222–4230.

33. 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. (in press)

34. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359.

35. Delhomme N, Padioleau I, Furlong EE, Steinmetz LM. 2012. easyRNASeq: a bioconductor package for processing RNA-Seq data. Bioinforma Oxf Engl 28:2532–2533.

36. Tolonen AC, Chilaka AC, Church GM. 2009. Targeted gene inactivation in Clostridium phytofermentans shows that cellulose degradation requires the family 9 hydrolase Cphy3367. Mol Microbiol 74:1300–1313.

37. Tolonen AC, Cerisy T, El-Sayyed H, Boutard M, Salanoubat M, Church GM. 2015. Fungal lysis by a soil bacterium fermenting cellulose. Environ Microbiol 17:2618–2627.

38. Tolonen AC, Zuroff TR, Ramya M, Boutard M, Cerisy T, Curtis WR. 2015. Physiology, Genomics, and Pathway Engineering of an Ethanol-Tolerant Strain of Clostridium phytofermentans. Appl Environ Microbiol 81:5440–5448.

39. Ochman H, Gerber AS, Hartl DL. 1988. Genetic applications of an inverse polymerase chain reaction. Genetics 120:621–623.

40. Aslanidis C, de Jong PJ. 1990. Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res 18:6069–6074.

41. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44:D279-285.

Acknowledgments

We thank C Cruaud and K Labadie for DNA and RNA sequencing as well as JL Petit and A Perret for biochemistry advice. This work was funded by the Genoscope-CEA, the Université d'Évry, a CNRS Chaire d'Excellence (ACT), and the Agence Nationale de la Recherche Grant ANR-16-CE05-0020 (ACT).

Data Availability

The authors confirm that all data underlying the findings are fully available without restriction. RNA sequencing files in FASTQ format have been deposited in the European Nucleotide Archive under accessions ERP019898 and ERP019903.

Supplementary Tables

Table S1 *C. phytofermentans* genes predicted to encode transporter proteins. (A) (page 1) 573 genes predicted to encode transporters by TransportDB including 393 genes for ABC transporters genes. Columns show gene name, genomic coordinates, and annotation. (B) (page 2) 485 genes predicted to encode transporters by KEGG. Columns show gene name, KEGG category, and annotation. This table is too big to be inserted in this thesis.

Table S2 mRNA expression of all *C. phytofermentans* genes. (A) (page 1) is the mRNA expression level (RPKM) of all 3,902 genes during log-phase growth on galactose. (B) (page 2) is mRNA expression of the 572 transporter genes from TransportDB during log phase growth on 10 different carbon sources. Genes encoding putative sugar transporters according to TransportDB are shown blue. This table is too big to be inserted in this thesis.

Primer	**Sequence (5' to 3')**	**Description**
Gene inactivation using group II intron (pQint)		
EBSuniv	CGAAATTAGAAACTTGCGTTCAGTAAAC	Universal primer for retargeting intron in pQint.
2241IBS	AAAACATATGATAATTATCCTTAGAAAGCAAGTTAGTGCGCCCAGATAGGGTG	Retarget intron to make antisense insertion in *cphy2241* 164 bp from start codon. Targeton score=10, e-value=0.021.
2241EBS1d	CAGATTGTACAAATGTGGTGATAACAGATAAGTCAAGTTACTTAACTTACCTTTCTTTGT	
2241EBS2	TGAACGCAAGTTTCTAATTTCGATTCTTTCTCGATAGAGGAAAGTGTCT	
OK2241F	CCAACCGACAATTCCACCAA	Primers flanking *cphy2241* to confirm intron insertion.
OK2241R	ACTTGGGGTTTGTGACAGGA	Primers flanking *cphy2241* to confirm intron insertion.
2465IBS	AAAACATATGATAATTATCCTTAGCTAACAGGAGAGTGCGCCCAGATAGGGTG	Retarget intron to make antisense insertion in *cphy2465 293* bp from start codon. Targeton score=8.47, e-value=0.077.
2465EBS1d	CAGATTGTACAAATGTGGTGATAACAGATAAGTCAGGAGAGATAACTTACCTTTCTTTGT	
2465EBS2	TGAACGCAAGTTTCTAATTTCGATTTTAGCTCGATAGAGGAAAGTGTCT	
OK2465F	CACCATCAACATAGGCTGCC	Primers flanking *cphy2465* to confirm intron insertion.
OK2465R	GGGGATTGTCAAAGGTAGGC	Primers flanking *cphy2465* to confirm intron insertion.
2732IBS	AAAACATATGATAATTATCCTTATTTGCCCTGCTTGTGCGCCCAGATAGGGTG	Retarget intron to make antisense insertion in *cphy2732 622* bp from start codon. Clostron (Perutka) score=5.24, e-value=0.
2732EBS1d	CAGATTGTACAAATGTGGTGATAACAGATAAGTCCTGCTTCGTAACTTACCTTTCTTTGT	

2732EBS2	TGAACGCAAGTTTCTAATTTCGGTTGCAAATCGATAGAGGAAAGTGTCT	
OK2732	TCGTCTACCCCTCCTATACGA	Primers flanking *cphy2732* to confirm intron insertion.
OK2732	ACTGCATCTGTATTCCGCTT	Primers flanking *cphy2732* to confirm intron insertion.
3589IBS	AAAACATATGATAATTATCCTTACATGGCAAGAAGGTGCGCCCAGATAGGGTG	Retarget intron to make antisense insertion in *cphy3589 332* bp from start codon. Targeton score=8.43, e-value=0.079.
3589EBS1d	CAGATTGTACAAATGTGGTGATAACAGATAAGTCAAGAAGATTAACTTACCTTTCTTTGT	
3589EBS2	TGAACGCAAGTTTCTAATTTCGGTTCCATGTCGATAGAGGAAAGTGTCT	
OK3589F	AAGCGTTCCATTGGTTGACA	Primers flanking *cphy3589* to confirm intron insertion.
OK3589R	TAGGAGGAGGTGCAGGTTTT	Primers flanking *cphy3589* to confirm intron insertion.
3859IBS	AAAACATATGATAATTATCCTTATTGCACCATCGAGTGCGCCCAGATAGGGTG	Retarget intron to make antisense insertion in *cphy3859 586* bp from start codon. Targeton score=8.01, e-value=0.108.
3859EBS1d	CAGATTGTACAAATGTGGTGATAACAGATAAGTCCATCGATTTAACTTACCTTTCTTTGT	
3859EBS2	TGAACGCAAGTTTCTAATTTCGGTTTGCAATCGATAGAGGAAAGTGTCT	
OK3859F	TGTCTCTCATGGTGCGAAGT	Primers flanking *cphy3859* to confirm intron insertion.
OK3859R	AGTTTTCATAGCAACCCTCCTC	Primers flanking *cphy3859* to confirm intron insertion.
OKtargetF	CGCCAGAAAACAAAAGAAAGCC	Confirm retargeting of pQint by sequencing
OktargetR	CGCTTTCGTTTCGTTCCCAT	Confirm retargeting of pQint by sequencing
Intron_Right	TGGGAAATGGCAATGATAGCG	Amplify ligation product during the inverse PCR
Intron_Left	AACTCATCGCTTTCCAACCG	Amplify ligation product during the inverse PCR
Expression of His-tagged hexose kinases		
2237_F	AAAGAAGGAGATAGGATCATGATACATCCTAGTATAGAAATACTG	Amplify *cphy2237* coding sequence. 5' red sequence is for ligation-independent clonging in pET-22B(+)
2237_R	GTGTAATGGATAGTGATCTTAATGGTGATGGTGATGATGAAGTTTATCCAAATGAATTGCTCC	Amplify *cphy2237* coding sequence. 5' red sequence for ligation-independent clonging in pET-22B(+). Blue sequence is C-terminal His tag.
0329_F	AAAGAAGGAGATAGGATCATGGACAAGTTTTGCTTTGGAATTG	Amplify *cphy0329* coding sequence. 5' red sequence is for ligation-independent clonging in pET-22B(+)
0329_R	GTGTAATGGATAGTGATCTTAATGGTGATGGTGATGATGGCCAATCACTAATTTTGCACAAC	Amplify *cphy0329* coding sequence. 5' red sequence for ligation-independent clonging in pET-22B(+). Blue sequence is C-terminal His tag.
P2116	ATCGAGATCTCGATCCCGCG	Confirm cloning in pET-22(B) by sequencing
P2117	GCAGCAGCCAACTCAGCTTCC	Confirm cloning in pET-22(B) by sequencing

Table S3 Primers used in this study for targeted gene inactivation and purification of hexose kinases. Columns show primer name, sequence (5' to 3'), and description.

Supplementary Figures

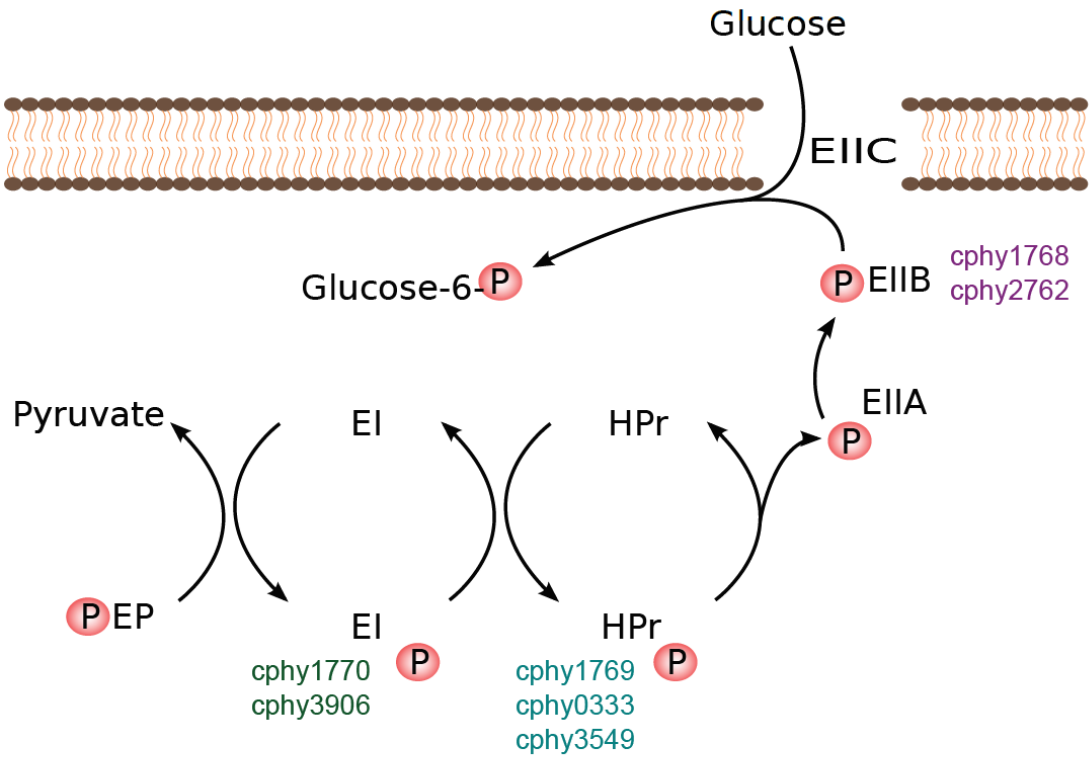

Fig S1 *C. phytofermentans* genes predicted to encode PTS. The operon *cphy1768-70* encodes EI (Cphy1770) to harvest the phosphate from PEP, HPr (Cphy1769) to transfer the phosphate from EI to EIIA, and EIIABC (Cphy1768) (EIIABC). Cphy1768 includes an EIIA domain (residues 600-750) to transfer the phosphate from HPr to EIIBC and the EIIBC domain (residues 4-530) to receive the phosphate from EIIA and transfer it to the sugar. *C. phytofermentans* also encodes a second EI protein (Cphy3906), two other HPr proteins (Cphy0333, Cphy3549), and a second EIIABC (Cphy2762).

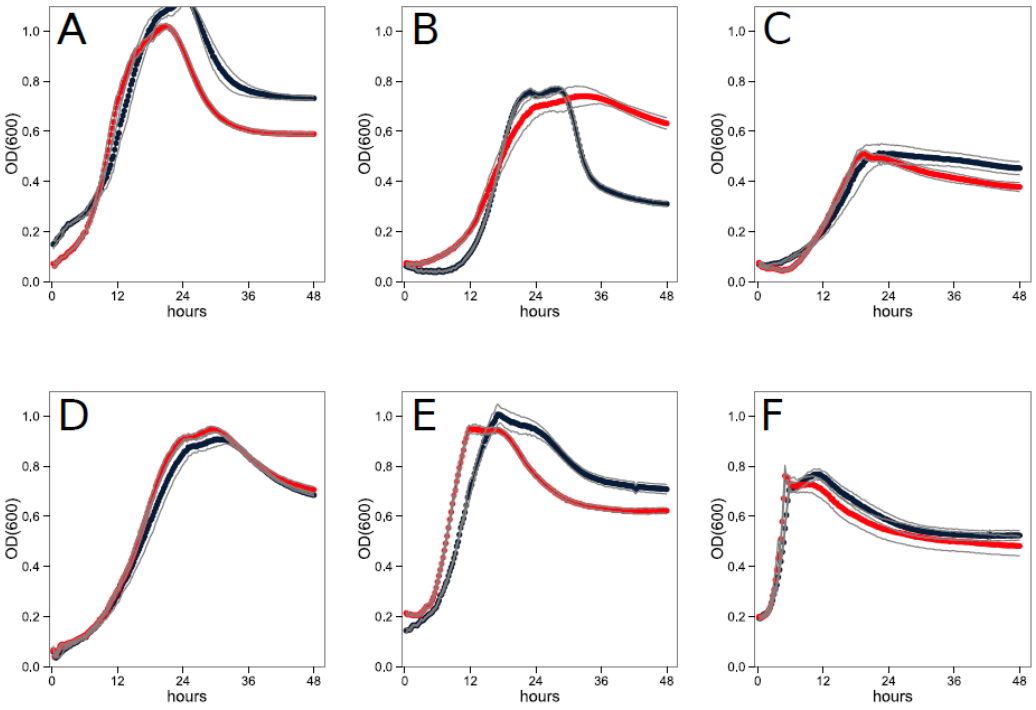

Fig S2 Growth (OD_{600}) on different carbon sources of *C. phytofermentans* WT (black) and *cphy3859*::int586a (red). Carbon sources: glucose (A), galactose (B), galacturonic acid (C), cellobiose (D), galactan (E), homogalacturonan (F). All data is mean of quadruplicate cultures ± s.d.

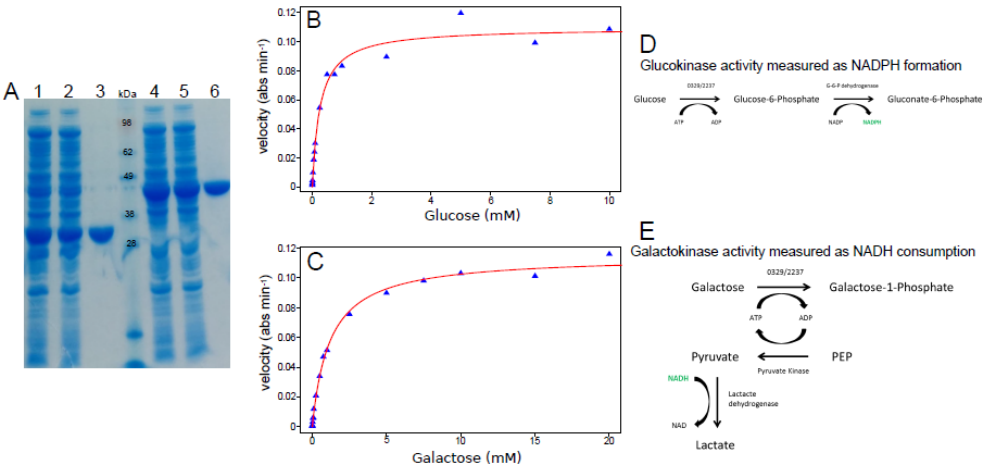

Fig S3 Glucokinase and galactokinase activities of purified Cphy2237 and Cphy0329. (A) Visualization of His-tagged Cphy0329 (lanes 1-3) and Cphy2237 (lanes 4-6) proteins on a 12% SDS-PAGE gel (Nupage bis-Tris novex gel IM-8042). Gel shows whole protein lysates (lanes 1,2,4,5) and purified Cphy0329 (lane 3) and Cphy2237 (lane 6). Molecular weights (kDa) from the ladder (Invitrogen LC5925) confirm the expected sizes of 32.6 kDa for Cphy0329 and 46.3 kDa for Cphy2237. (B) Kinetics of Cphy0329 glucokinase activity measured as NADPH formation (Absorbance 340 nm) at different glucose concentrations. (C) Kinetics of Cphy2237 galactokinase activity measured as NADH consumption (absorbance 340 nm) at different galactose concentrations. Diagrams of the reactions to measure (D) glucokinase and (E) galactokinase activities.

II.1.3 Conclusion and perspectives

Results of this study show the utility of functional genomics and targeted gene inactivation to identify sugar transporters in C. phytofermentans. This strategy can probably be used to characterize other genes in C. phytofermentans metabolism but these predictions must be experimentally validated. For instance, the mutant cphy2732::int662a has a normal growth on galacturonic acid even though transcriptomic results revealed that cphy2731-3 encode the most highly expressed ABC transporter.

Although these results are convincing, the complementation of these different mutants with a plasmid harboring the inactivated transporter would be a supplementary proof, as previously done in *C. difficile* for Spo0A mutants [424]. The mutant with the supplementary plasmid would transport the specific sugar as the WT. Nevertheless the construction of a plasmid with an ABC transporter is challenging, the operon is near 4 kb long without a working promoter.

Looking at the regulation level, we recently released a study about Transcriptional Start Sites (TSS) in *C. phytofermentans* (Appendix 2, [188]), enabling to know the beginning of the transcription at base-pair level and the direction of the transcription. As presented above, the predicted association of the glucose/galactose transporter to the ABC transporter *cphy2241-3* was confirmed with the mutant *cphy2241::int164a*, unable to grow on glucose/galactose. Nevertheless, the analysis of TSS of *C. phytofermentans* confirms the regulation of this operon on glucose and repression on other carbon sources. We found that one TSS is highly expressed on glucose to transcribe the whole operon *cphy2241-3*. In contrast, under other substrate conditions, this RNA is lowly expressed and an antisense TSS is expressed upstream *cphy2241-3*. This antisense TSS is the beginning of an RNA, which does not map on *any C. phytofermentans* peptides [22] or predicted ORF. It suggests its role as a transcriptional regulator of the glucose/galactose ABC transporter operon.

Although results of ABC transporters are convincing, the complementation of these different mutants with a plasmid harboring the inactivated transporter would be a supplementary proof, as previously done in *C. difficile* for Spo0A mutants [424]. The mutant with the supplementary plasmid would transport the specific sugar as the wild-type. One gene is inactivated but we suppose that the whole operon is affected. Do we have to complement one gene or the operon? The construction of a plasmid with an ABC transporter is challenging, the operon of the three genes is near 4 kb long without a working promoter. How to choose the promoter? Is the native promoter a good choice? If yes, what is the size of this promoter? The

paragraph above shows that different TSS of the glucose/galactose ABC transporter depends on substrates, making the choice and the size of the promoter difficult. We decided that these results are sufficient to link an ABC transporter to a specific sugar, but we can imagine that reviewers of the submitted article would ask these experiments.

Using this study, some experiments would be investigated to discover new functions and properties in Clostridia:

Following these results and reusing RNA-Seq data (**Fig 1**), two ABC transporters are highly expressed on pentose-based sugars xylose and xylan, operons *cphy1585-7* and *cphy3208-10* respectively. I would suggest inactivating these two transporters in *C. phytofermentans* to validate their roles on these two substrates.

An ambitious experiment would be to transfer an ABC transporter from *C. phytofermentans* to another *Clostridium* which does not assimilate the sugar specific to the ABC transporter. The growth of *C. acetobutylicum*, which does not possess cellulosomes, is very low on cellulose and hemicellulose and strategies are applied to improve this lack by adding cellulase or cellulosome to degrade these polysaccharides. Nevertheless, this study shows that oligosaccharides from polysaccharides degradation are transported with specific transporter. In my point of view, addind cellulase is not sufficient to give cellulolytic ability to solventogenic Clostridia, specific transporters should be added too.

Last idea: with the (slowly) increasing number of ABC transporter characterized for various substrate in Clostridia, it should be interesting to see the emergence of a verified database of transporter, like the CAZyme database [425]. This database would be useful for the community because annotations of sugar specificity of ABC transporters are mainly vague or wrong.

II.2 Application of GETR to C. phytofermentans

II.2.1 Introduction

Currently available genetic tools in Clostridia were presented in the section *I.3 Clostridium genetics*. In this chapter the *Genome editing via Targetron and Recombinases* (GETR) strategy is described and results of this strategy in *C. phytofermentans* are revealed.

The novelty of this strategy is to merge two well-known techniques: Targetron for the insertion of group II introns at desired *loci* (cf. section I.3.4 Reverse Genetics) and the Cre/*lox* system for the recombination. This strategy was released by Enyeart *et al.*, in 2013 [330] and main achievements are: inversion of 1.2 Mb in *E. coli* and 1.5 Mb in *B. subtilis* - deletion of 120 kb in *E. coli* and 15 kb in *S. aureus* - insertion of an operon of 12kb in *E. coli* (efficiency near 100%) - cut-and-paste in *E. coli* of 120 kb to 1.5 Mb away.

The goal of this strategy is to use group II introns to carry *lox* sites into desired and specific *loci* of the genome. And the recombination is promoted with the expression of the Cre protein inside the cell.

The Cre/lox system is based on site-specific recombination between two DNA sequences. The Cre protein is 38 kDa and comes from the P1 bacteriophage [426], the protein recognizes and splices specific DNA sequences at *lox* sites. *lox* sites are 34 bp long with an asymmetric sequence of 8 bp surrounded by two palindromic sequences of 13 bp, Table 8. This technology was applied in bacteria [427], yeast [428] and mammalian cells [429].

Name	13 bp recognition region	8bp spacer region	13 bp recognition region
loxP	ATAACTTCGTATA	NNNTANNN	TATACGAAGTTAT
lox66	ATAACTTCGTATA	GCATACAT	TATACGAACGGTA
lox71	TACCGTTCGTATA	GCATACAT	TATACGAAGTTAT
lox72	TACCGTTCGTATA	GCATACAT	TATACGAACGGTA
loxFas/66	ATAACTTCGTATA	TACCTTTC	TATACGAACGGTA
lox511/66	ATAACTTCGTATA	GTATACAT	TATACGAACGGTA
loxFas/71	TACCGTTCGTATA	TACCTTTC	TATACGAAGTTAT
lox511/71	TACCGTTCGTATA	GTATACAT	TATACGAAGTTAT

Table 8 Sequences of various *lox* sites.

Depending on the orientation of *loxP* sites, the recombination with the Cre protein can enable a deletion, an inversion or a translocation, Figure 32. Nevertheless, the *loxP* sites can recombine several times if the Cre protein is still present and active, Figure 32. To avoid this reversible recombination, *lox* sites were engineered and tested to give a permanent recombination. For instance *lox66* and *lox71* are modified *lox* sites, which after recombination form the *lox72* site, lowly recognized by the Cre protein [427].

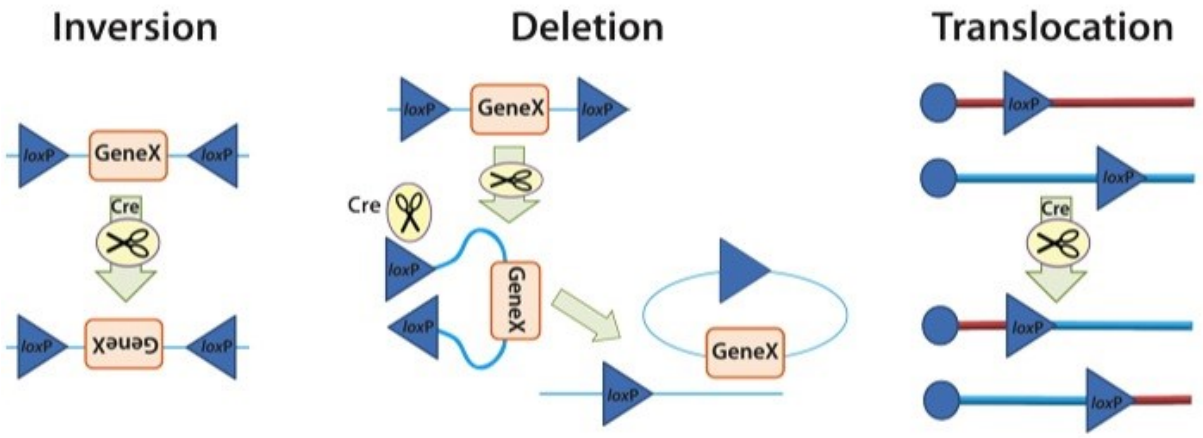

Figure 32 Different recombinations between lox sites with the help of the Cre protein, from http://blog.addgene.org/plasmids-101-cre-lox .

With a supplementary DNA surrounded by *lox* sites in a plasmid, complementary to genomic *lox* sites in the good orientation, the Cre protein can mediate the double recombination and integrate this exogenous DNA in the genome.

II.2.2 Deletion

II.2.2.a Presentation of the deletion strategy with GETR

As a proof of principle and to test the GETR strategy in *C. phytofermentans,* we decided to delete a phage island in *C. phytofermentans*. The DNA sequence to be removed is between *cphy2944* and *cphy2993* and is 38609 bp. It is important to have non-essential genes within the removed sequence, otherwise mutants will not be viable. We supposed that the DNA sequence between *cphy2944* and *cphy2993* is non-essential for several reasons:

- Most of these 49 genes have a very low RNA expression on 13 different substrates (Appendix 1, [51]). The expression of *cphy2944*, *cphy2968*, *cphy2968*; *cphy2969*, *cphy2973* and *cphy2974* is low but the expression of *cphy2976* and *cphy2977* is higher, with an overexpression of *cphy2976* on polysaccharides.
- Expressed genes are annotated phage protein or hypothetical protein.

To insert *lox* sites in the genome, the intron is used as a cargo. The *lox* sequence is inserted at the MluI restriction site of the intron, Figure 33.

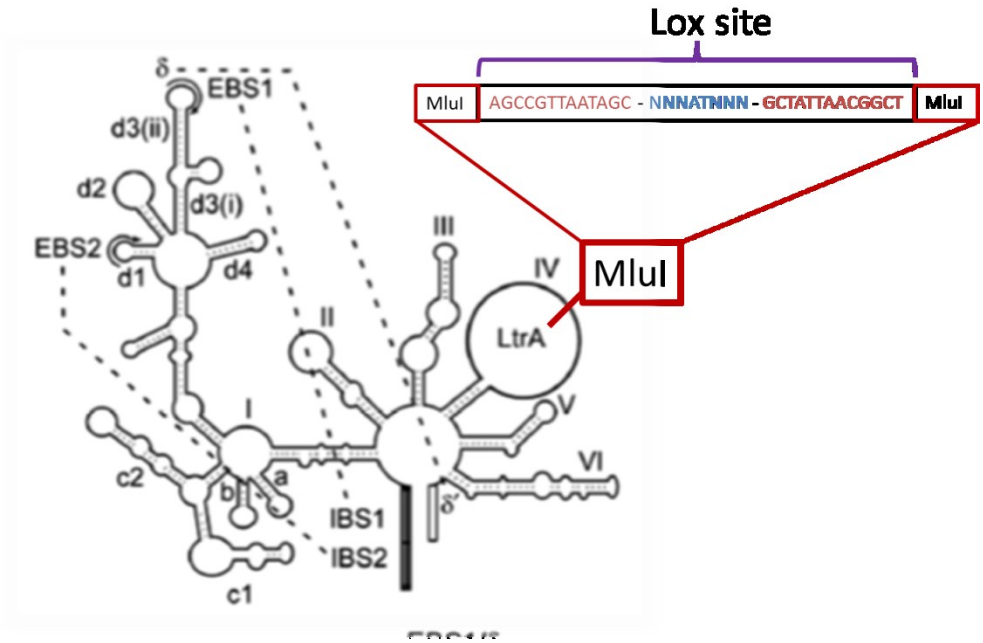

Figure 33 Insertion of *lox* sequence into the intron at MluI restriction site, modified from Enyeart *et al.*, 2013 [330].

The intron with *lox* sites is integrated into the genome after the retargeting of the intron. To retarget a group II intron, the intron can be modified to target a genome *locus*. Algorithms (www.targetrons.com or www.clostron.com) calculate best intron insertion sites into the genome; around 4 sites are available each 500 bp. The algorithm returns the modification to apply to the group II intron to retarget the intron to the selected *locus*.

The strategy to delete a DNA sequence is presented in the Figure 34 with the example of the deletion between *cphy2944* and *cphy2993*.

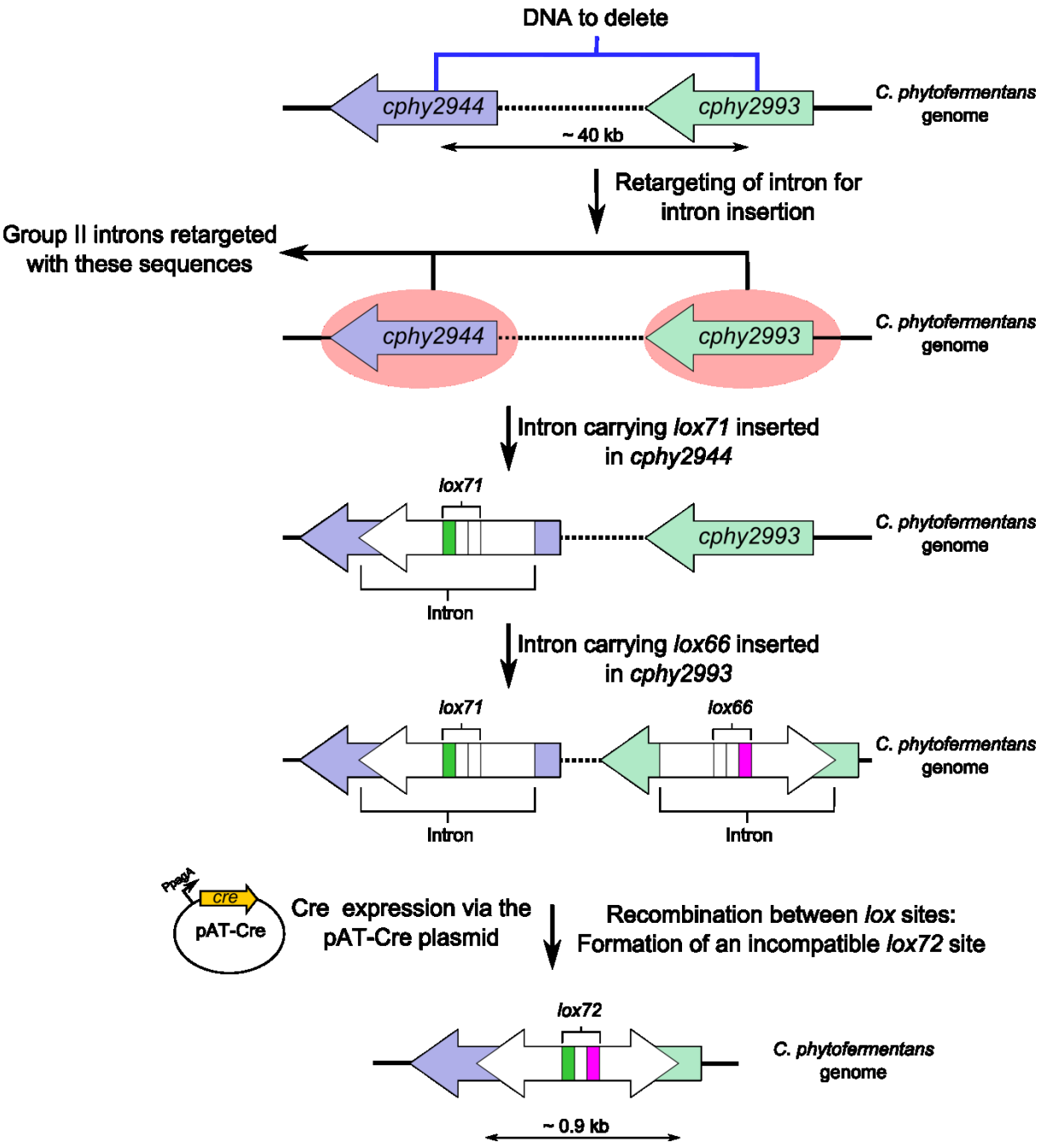

Figure 34 Step by step strategy to delete a broad DNA sequence *via* GETR, example of the deletion between *cphy2944* and *cphy2993*.

II.2.2.b Construction of the mutant

To delete the sequence between *cphy2944* and *cphy2993*, the strategy of the Figure 34 was followed:

- Around 500 bp of *cphy2944* and *cphy2993* were sent to the targetron platform (www.targetron.com) The plasmid with the group II intron (pQint), with the appropriate *lox* site, were retarget using returned primers from targetron (supp Table 1 and supp Table 2).

An intron with *lox71* was retargeted to *cphy2944,* with the plasmid pQintlox71_2944. The transformation of *C. phytofermentans* WT with this plasmid enables the insertion of the intron, and the *lox71*, in *cphy2944*. The plasmid of the mutant was removed by successive dilutions. The mutant is named *cphy2944::intlox71* and was transformed with the plasmid pQintlox66_2993. Clones *cphy2944::intlox71* with the insertion in *cphy2993* of the intron containing *lox66* were selected and verified, Figure 35. The plasmid was removed from the mutant by successive dilution as before. This mutant is named CphyDI and has the two insertion: the *lox71* intron in *cphy2944*, the *lox66* intron in *cphy2993*. These insertions were sequenced to validate orientations and sequences of *lox* sites and introns.

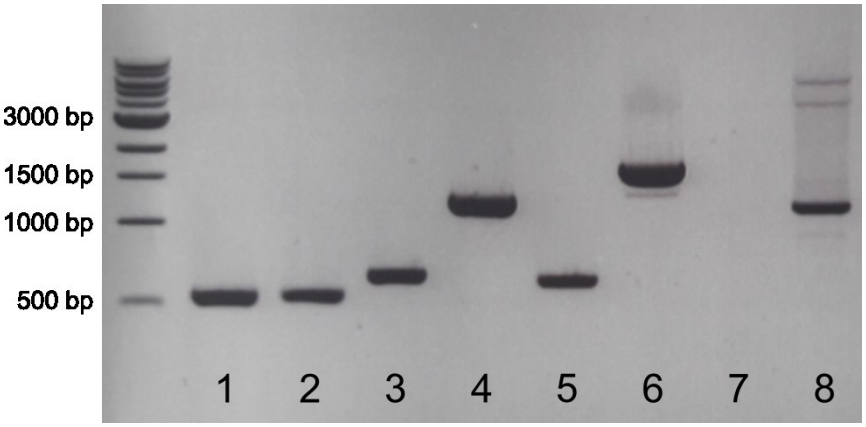

Figure 35 Electrophoresis of PCR product to verify the mutant CphyDI. 1, 3, 5, 7 correspond to *C. phytofermentans* WT. 2, 4, 6, 8 correspond to CphyDI. 1 and 2 are amplified from OK2241 F and R to check if the culture is *C. phytofermentans*. Primers OK2944 F and R are used for 3 and 4 to check the intron insertion; Primers OK2993 F and R are used for 5 and 6 to check the intron insertion. 7 and 8 are inverse PCR from gDNA of *C. phytofermentans* WT and cphy DI with primers Intron_Right and Intron_Left to check the number of intron insertions, no introns are expected for *C. phytofermentans* WT; two introns are expected for cphyDI: 5576 bp for the intron in *cphy2944* and 1311 bp for the intron in *cphy2993.*

The direction of *lox71* and *lox66* is crucial, if the orientation is not appropriate, instead of a deletion an inversion can occurred. Retargeted plasmids and intron insertion were verified by sequencing, data not shown.

The number of insertion in the genome by the group II intron strategy was verified by inverse PCR (iPCR). The inverse PCR is a protocol enabling to amplify all intron inserted in a genome, Figure 36.

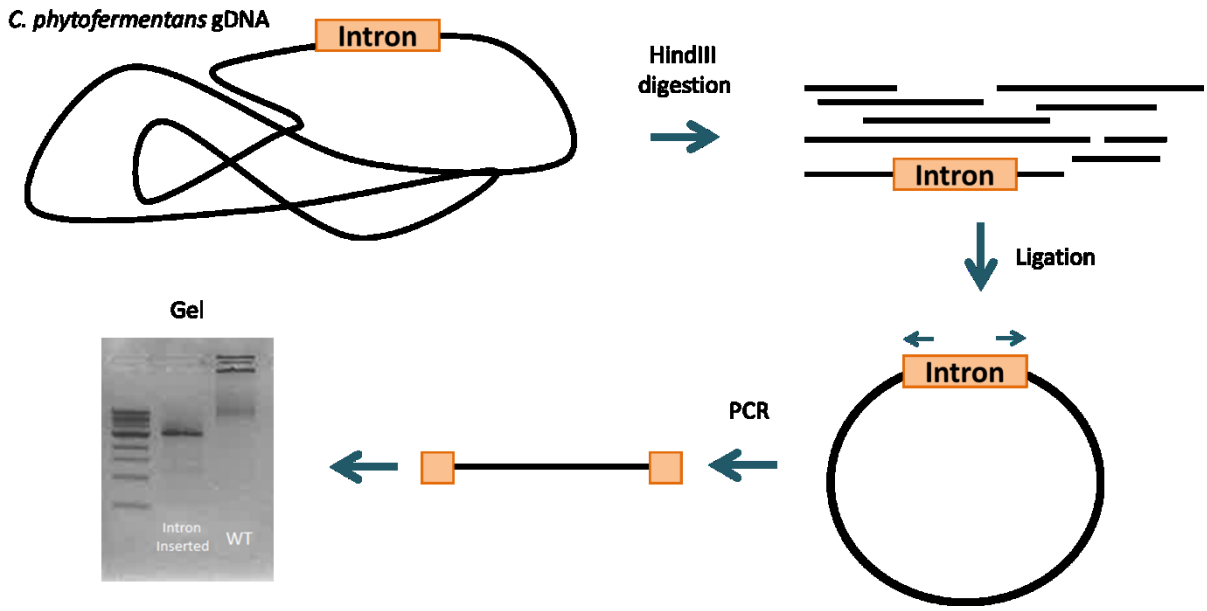

Figure 36 The inverse PCR procedure.

The iPCR of CphyDI revealed that the clone has the two wanted insertion plus an unexpected insertion, Figure 35 lane 7 and 8.

II.2.2.c Recombination event

To delete the prophage region surrounded by the two lox sites, the Cre protein must be present inside the cell. The expression of the Cre protein was challenging, it was not possible to transform different Cre-expressing plasmids by conjugation. Several plasmids were built, with alternative origins of replication, or different promoters to regulate the Cre expression. The toxicity of Cre for *E. coli* can be an issue during the conjugation, preventing the transfer of the plasmid to CphyDI. Recently an electroporation protocol was established by Sanitha Mary and Magali Boutard in *C. phytofermentans*.

The transfer of the pAT-Cre plasmid in CphyDI occurred by electroporation. Several CphyDI were screened for the presence of the pAT-Cre plasmid, the recombination and to check if clones are *C. phytofermentans*. The gel A of the Figure 37 shows that the recombination occurred in four of the 11 clones. Clones A9-2, B2-1, B5-2 and B8 have a band near 400 bp, whereas the expected band for the WT is near 40 kb. The gel B confirms the deletion because no DNA was amplified with primers which amplify a gene inside the deleted sequence. But other mutants, WT and CphyDI (prior to recombination) have the amplification.

Moreover other primers were used to amplify another region of *C. phytofermentans*, confirming that all clones tested are *C. phytofermentans* (data not shown).

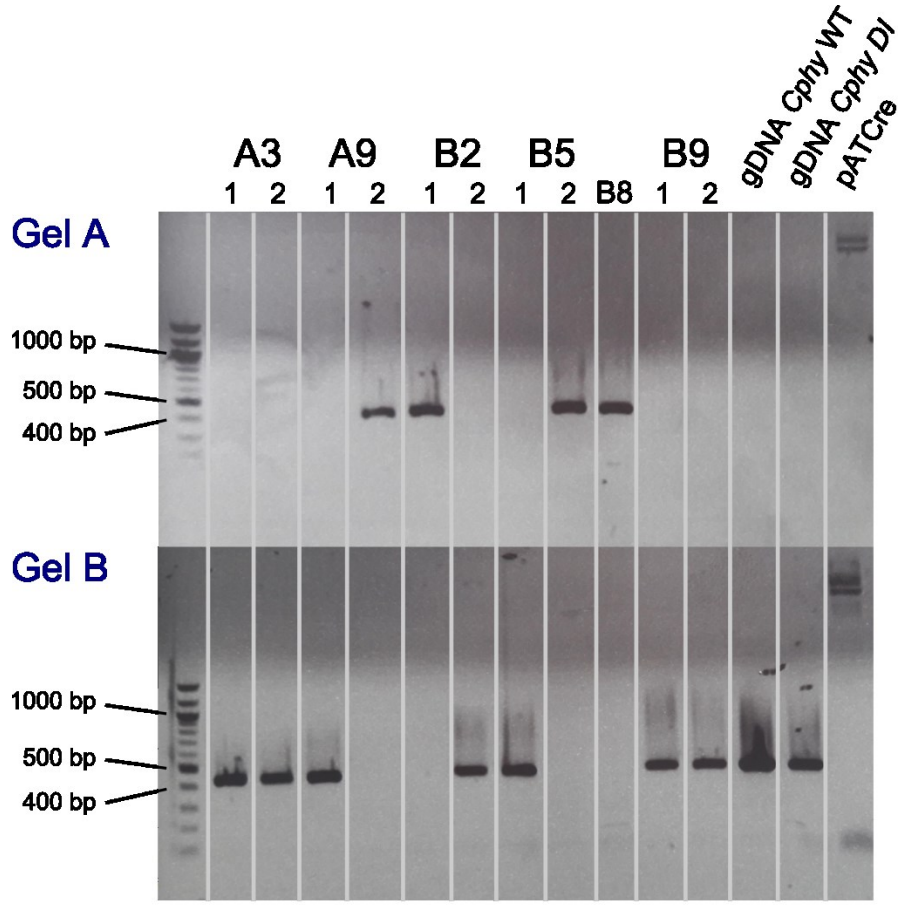

Figure 37 Electrophoresis of PCR products of 11 clones of CphyDI after the transformation with pAT-Cre. The gel A is made with PCR products performed with primers OK2944F and OK2993R which flank the region to delete. The gel B is an made of PCR products performed with primers OK2967 F and R which would amplify genes inside the region to delete. The ladder is the 100 bp ladder from New England Biolabs.

The expected band after the recombination is near 900 bp, but the band observed in the gel is near 400 bp. The recombination is confirmed by sequencing, but the mechanism is not as expected, Figure 38.

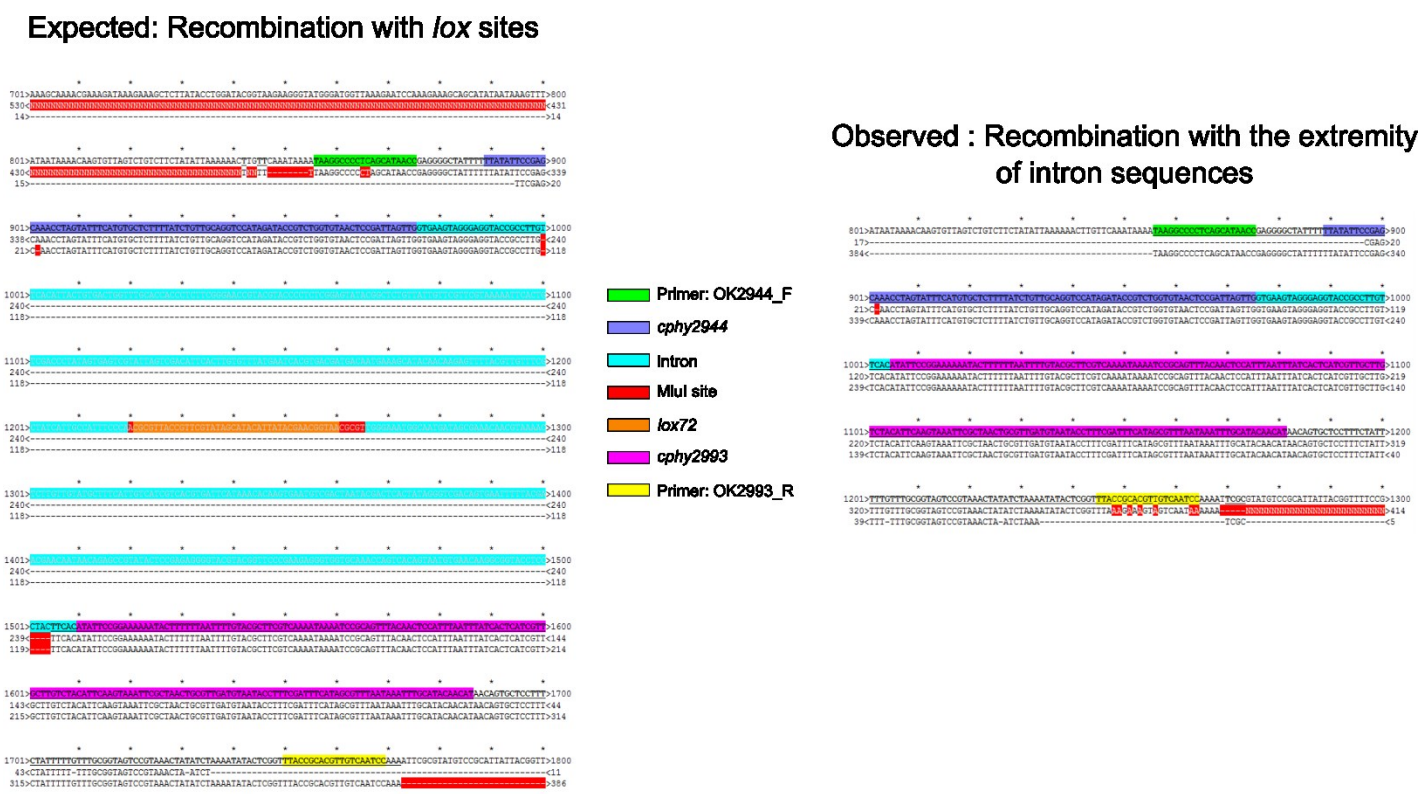

Figure 38 Sequencing results of CphyDI after recombination, comparison between the expected and unexpected matching of sequencing results.

Sequencing results confirmed that the deletion occurred: sequencing results match to the DNA sequence of the gene *cphy2944*, a portion of the intron and then the gene *cphy2993*. Nevertheless an unexpected event happened, after the recombination and only 29 bp of the intron left in the genome. Whereas after the recombination of *lox* sites it was expected to observe 534 bp of intron containing the *lox72* site, from the reassembling of the two parts of introns separated by 38 kb.

This unexpected event occurred in 6 different electroporation and four separate sequencing results confirm the same recombination at the same base pair.

II.2.2.d Hypothesis

The mechanism of this unexpected event is not yet understood. Nevertheless I propose one hypothesis: the direction of intron can influence the recombination, which occurs only with the presence of the Cre protein, Figure 39. *lox* sites and the Cre protein favored this unexpected event, which seems to be a homologous recombination.

A discussion with Peter Enyeart was begun to understand the event. Peter Enyeart is the author of the GETR study. He did not see this mechanism before.

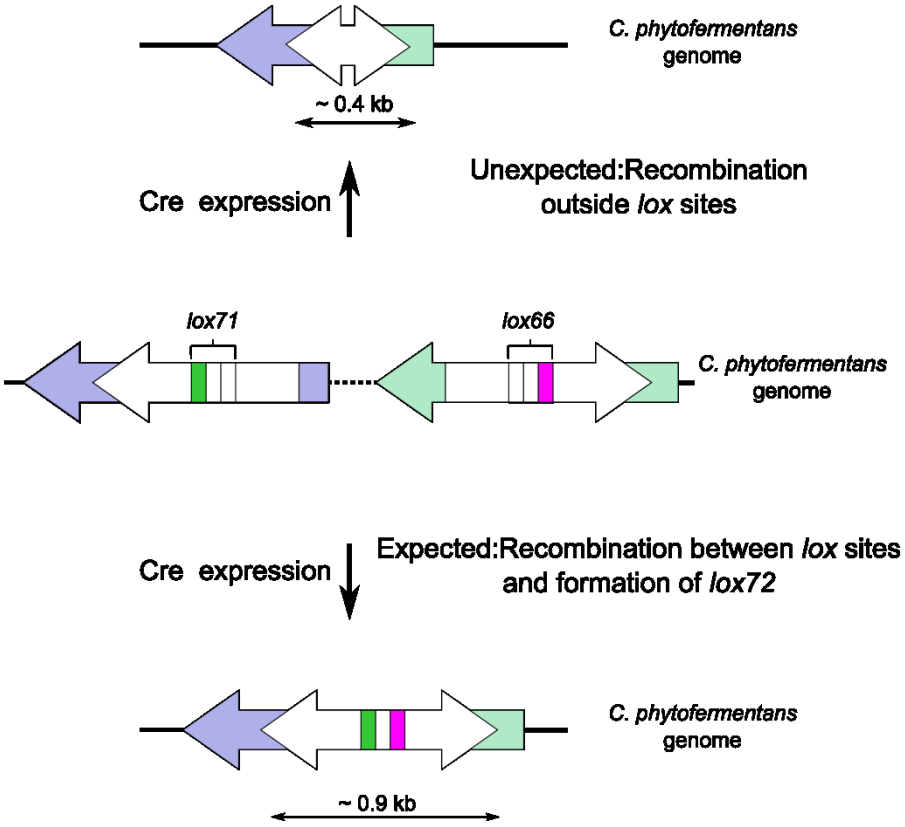

Figure 39 Schema of the expected and unexpected recombination for the deletion between *cphy2944* and *cphy2993*.

The next step is to test the recombination with the two *lox* sites in he same direction as CphyDI but with both introns in the same direction (in CphyDI introns are in opposite directions), this mutant is named CphyDI2 in Figure 40. This second mutant enables to verify if the intron orientation is important in the deletion and if the expected deletion occurs in this configuration.

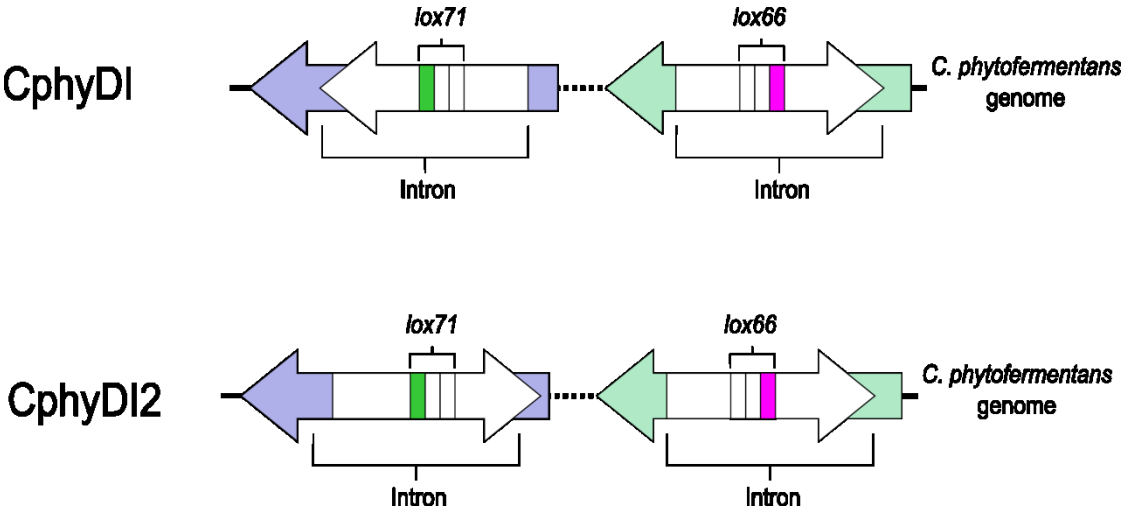

Figure 40 Schema of the two constructions of *C. phytofermentans* to delete the prophage region.

To demonstrate the reproducibility of this event, the two configurations in Figure 40 are designed and will be tested in another region of the genome: 8 kb between *cphy2967* and *cphy2975* would be deleted.

II.2.3 Insertion

To insert a DNA sequence with the GETR strategy, a recombination must occur between two pairs of *lox* sites. The DNA to insert is present in a separated plasmid and is surrounded by two *lox* sites; two *lox* sites have to be previously integrated into the genome with the help of group II introns. One *lox* site in the genome is compatible with the *lox* site of the plasmid; the second *lox* site of the genome is compatible with the second *lox* site of the plasmid. These two pairs of *lox* site have to be incompatible each other, preventing unwanted recombination. Then the expression of the Cre protein enables the double recombination between lox sites and the insertion of the foreign DNA from the plasmid to the chromosome occurs.

To test this insertion technique in *C. phytofermentans*, two *lox* sites were inserted in once in the gene *cphy1575*. The two lox sites were integrated in the intron in the same manner as for *lox71* and *lox66*, at the MluI restriction site. This two lox site are *lox511/71* (spacer 511 and one recognition region 71) and *loxFas/66* (spacer Fas and one recognition region 66), this combination is named 2ML4. These *lox* sites are compatible to *lox511/66* (spacer 511 and one recognition region 66) and loxFas/71 (spacer Fas and one recognition region 71) respectively. The two last *lox* sites are in the plasmid pAT2ML9 and these sites surround an exogenous DNA of 1 kb, Figure 41.

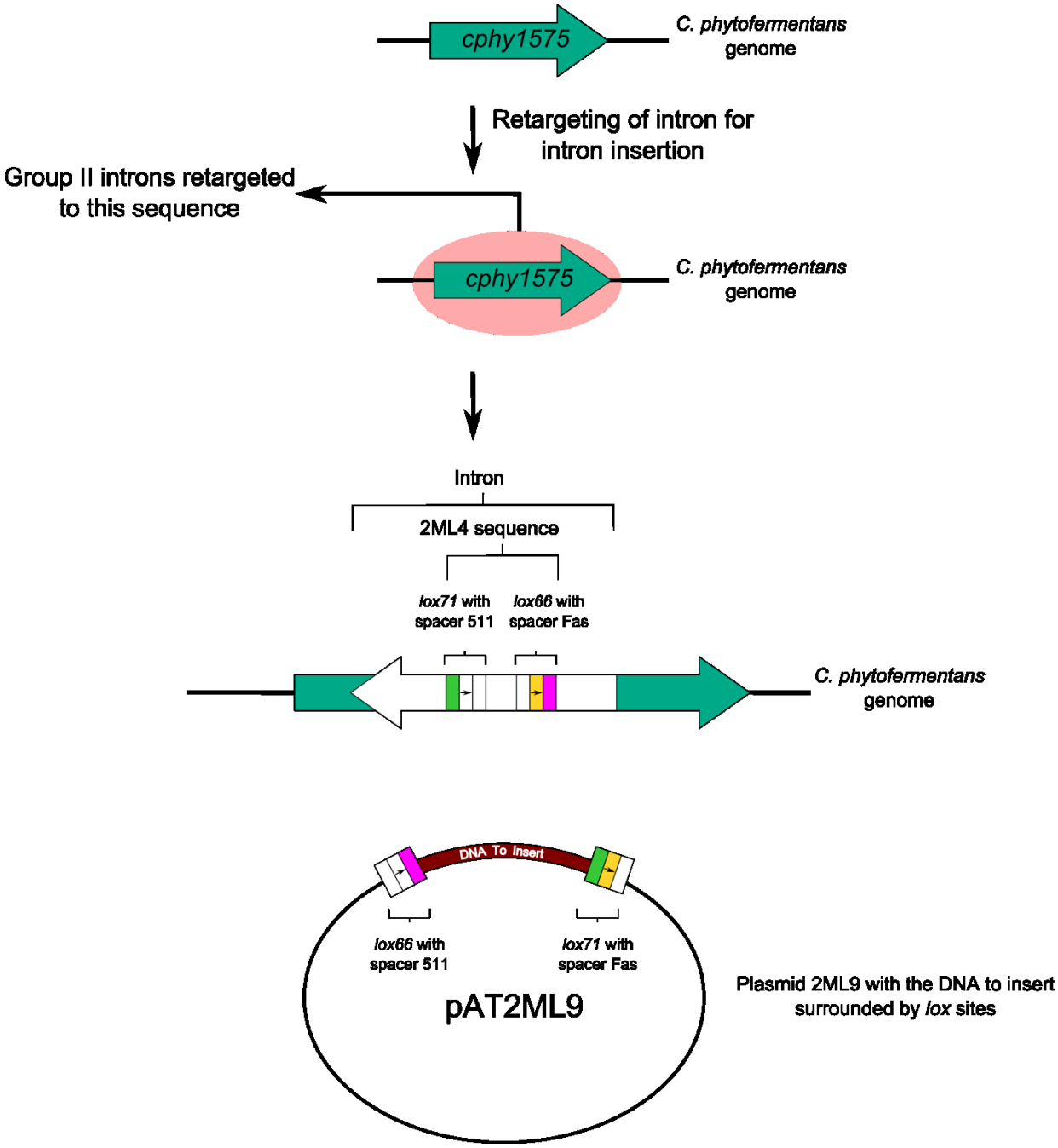

Figure 41 Insertion of exogenous DNA with the GETR strategy, first part. Insertion of *lox* sites in the genome with group II introns. The plasmid with the DNA to insert is present in a separated plasmid; this supplementary DNA is surrounded by *lox* sites.

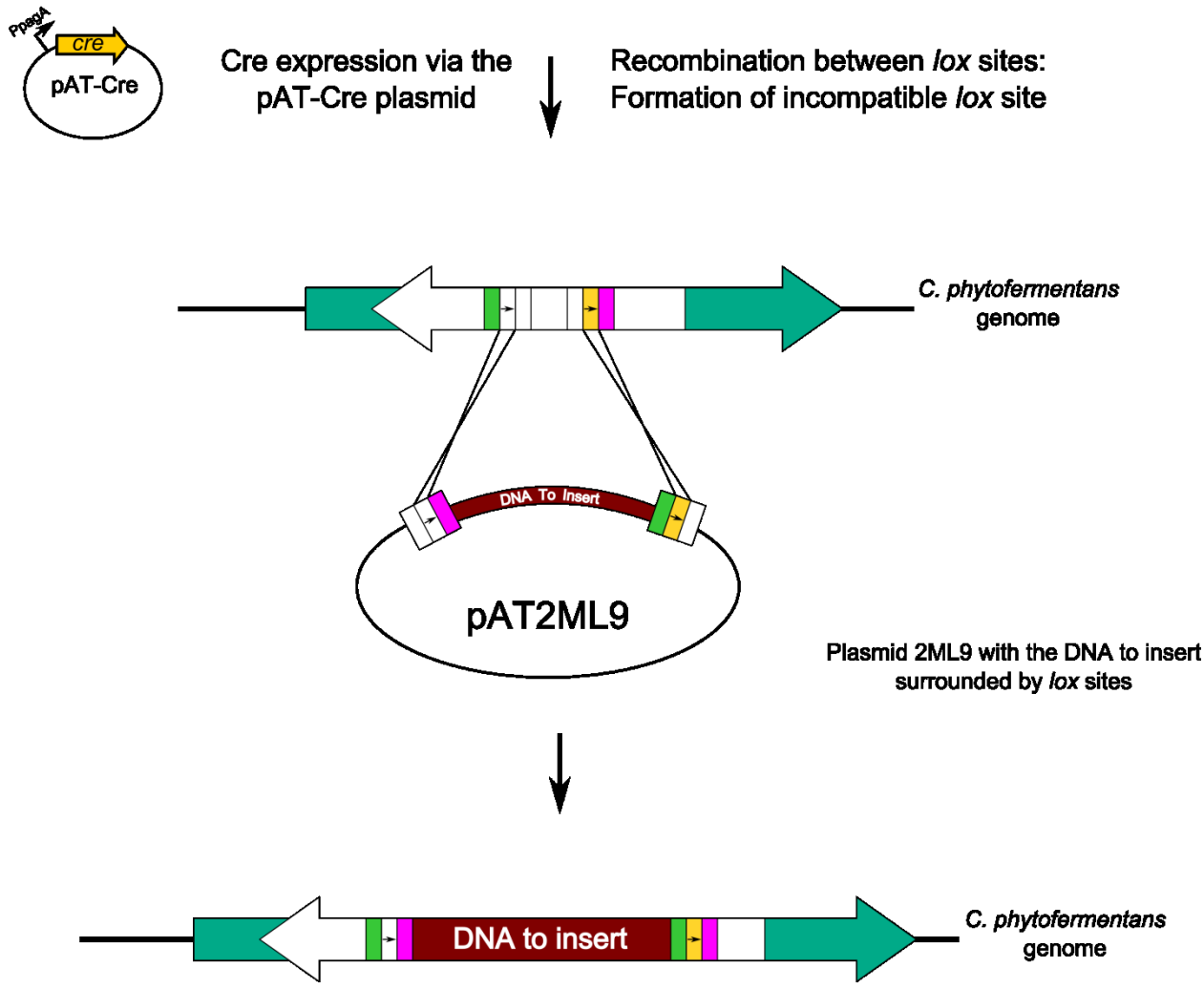

Figure 42 Insertion of exogenous DNA with the GETR strategy, second part. The recombination occurs with the Cre protein, this protein is produced from a second plasmid in the cell. *lox* sites from the recombination are incompatible each other.

One *C. phytofermentans* mutant has the two lox sites in the genome plus the plasmid pAT2ML9. Nevertheless the pAT2ML9 has the same origin of replication and resistant cassette as pAT-Cre, preventing the transformation of pAT-Cre in this mutant. A second origin of replication and a second resistant cassette have to be found and characterized in *C. phytofermentans*. It would enable to have a Cre expressing plasmid compatible with pAT2ML9. The expression of the Cre protein in this mutant would make the double recombination and the insertion of the exogenous DNA possible, Figure 42.

II.2.4.Methods

Culture conditions *C. phytofermentans* ISDg ATCC 700394 was cultured at 30°C, anaerobically in GS2 medium [430] and 0.5% glucose supplemented, if needed, with erythromycin at 40 µg ml^{-1} in plate and 200 µg ml^{-1} in liquid. *E. coli* DH5α was cultured in LB medium; during cloning, plasmids were transformed *via* electroporation. *E. coli* pRK24 was used for the conjugation with *C. phytofermentans* [197]. This *E. coli* strain is selected with the carbenicillin at 50µg ml^{-1} or with both carbenicillin at 50µg ml^{-1} and erythromycin at 200 µg ml^{-1} when a pQint plasmid is present. The protocol for the conjugation from Tolonen *et al.*, 2009 [197] was adapted to increase the efficiency, all steps occurred anaerobically, from the mating to the plating. To remove the plasmid after the insertion, *C. phytofermentans* mutant is diluted at 1/10 five times successively (one dilution each day), then 2 µl are streaked on GS2 plate without antibiotics. Some colonies are inoculated in 5 ml of GS2 media and are checked for plasmid remaining: a PCR with primer specific to the plasmid is achieved. From the preculture, 10 µl are spread on LB plate to check contaminations and 10 µl are spread on GS2 erythromycin at 40 µg ml^{-1} plates to verify that no mutant possess the plasmid and survive.

Plasmid construction: the plasmid pAT-Cre was built from the plasmid pAT19 [242] and the amplification of the PpagA promoter and the Cre protein from the plasmid pRAB1 [327] (generously provided by Dr. Ralph Bertram and Christopher F. Schuster from the University of Tuebingen, Germany) with primers Cre_EcoRI and CreR_XbaI. The PCR product and the plasmid were cut with EcoRI and XbaI restriction enzymes for the cloning.

pQint plasmid was adapted to get *lox* sites. *lox* sites were ordered as two complementary oligomers (1L71F with 1L71R, 1L66F with 1L66R, 2ML4F with 2ML4R and 2ML9F with 2ML9R). Each pairs were annealed together by mixing 10 µL of 200-µM solutions of each of the oligomers with 80 µL of water, holding at 95°C for 20 minutes, then the temperature was decreased slowly to 40°C then cooling at room temperature. Annealed oligomers were ligated directly into the MluI restriction site of pQint plasmids (previously cut with MluI). Five different versions were build, pQintlox71a, pQintlox71s, pQintlox66a and pQintlox66s and pQint2ML4. *lox71* and *lox66* were considered in both directions, “s “ for sense “a” for antisense. It enables to add *lox* sites in the right direction, no matter the direction of intron insertion. These five plasmids were checked by sequencing.

pQint plasmids and derived (with *lox* sites) were retargeted using the protocol[1]:, and [197] with primers in Supp Table 1. Here, the plasmid and the PCR product have to be cut with NdeI and BsrGI instead of HindIII and BsrGI.

Inverse PCR: the genomic DNA is extracted and purified from 5 ml of fresh culture and the GenElute™ Bacterial Genomic DNA Kit Protocol (NA2110 from Sigma-Aldrich). 1µg of DNa is cut with the HindIII restriction enzyme, then purified with QIAquick PCR Purification Kit (ID: 28104 from Qiagen). The purified DNA is ligated with the T4 DNA ligase (NEB) and the following PCR is proceed:

Reaction mix		PCR program		
H_2O	34.5µl	98°C	30 sec	
Q5 buffer	10 µl	98°C	15 sec	
dNTPs	1 µl	64°C	30 sec	X30
Intron_Right	1 µl	72°C	7 min	
Intron_Left	1 µl	72°C	10 min	
Q5 polymerase	0.5 µl	4°C	∞	
DNA (ligation product)	2 µl			

II.2.5 Supplementary data

Supp Table 1 Primers used for this study.

Name	Sequence	Role
1L71F	CGCGTTACCGTTCGTATAGCATACATTATACGAAGTTATA	To build the *lox71*
1L71R	CGCGTATAACTTCGTATAATGTATGCTATACGAACGGTAA	
1L66F	CGCGTATAACTTCGTATAGCATACATTATACGAACGGTAA	To build the *lox66*
1L66R	CGCGTTACCGTTCGTATAATGTATGCTATACGAAGTTATA	
OKMluIF	AATCTTGCAAGGGTACGGAG	To check by PCR and sequencing *lox* site in the intron
OKMluIR	ACGTACCCCTCTCGGAGTAT	
2ML4F	CGCGTGAGAGTACCGTTCGTATAGTATACATTATACGAAGTTATGAGAGTTTA AACGAGAATAACTTCGTATATACCTTTCTATACGAACGGTAGAGAGA	To build the 2ML4 sequence in the intron for the insertion
2ML4R	CGCGTCTCTCTACCGTTCGTATAGAAAGGTATATACGAAGTTATTCTCGTTTAA ACTCTCATAACTTCGTATAATGTATACTATACGAACGGTACTCTCA	
2ML9F	GCGCGCAGCCTGCAGATAACTTCGTATAGTATACATTATACGAACGGTAGAGA GACTAGTGTGATATGCCCTCGAGGAGATACCGTTCGT	To build the 2ML9 sequence in

[1] https://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/General_Information/targetron-user-guide.pdf

2ML9R	GCTGCGCGCGAATTCTATAACTTCGTATAGAAAGGTATATACGAACGGTATCT CCTCGAGGGCATATCACACTAGTCTCTCTACCGTTCG	the plasmid for the insertion
OK2ML9F	TACCGCCTTTGAGTGAGCTG	To verify the 2ML9 sequence in the pAT2ML9 plasmid
OK2ML9R	TACATCACCGACGAGCAAGG	
CreF_EcoRI	cgcgatagGAATTCgaggttcggattcatcctattaaac	To amplify the PpagA promoter and *Cre* gene from pRAB1 [273] to clone it in pAT19
CreR_XbaI	cgcgatagTCTAGAttactaatcgccatcttccagcag	
OK2ML9_F2	CTTCCGGCTCGTATGTTGTG	To verify the PpagA promoter and *Cre* gene from pRAB1 [273] in pAT19
OKIns_R2	CTCTTCGCTATTACGCCAGC	
Intron_Right	TGGGAAATGGCAATGATAGCG	To amplify the intron insertion in the inverse PCR
Intron_Left	AACTCATCGCTTTCCAACCG	
OKtargetF	CGCCAGAAAACAAAAGAAAGCC	To verify the retargeting into pQint
OKtargetR	CGCTTTCGTTTCGTTCCCAT	
EBSUniv	CGAAATTAGAAACTTGCGTTCAGTAAAC	For the retargeting
OK2944F	TAAGGCCCCTCAGCATAACC	To verify the intron insertion in *cphy2944*
OK2944R	CAAGACCTGTTATGCGCTCG	
2944IBS	AAAACATATGATAATTATCCTTAAAGAACCTGCAGGTGCGCCCAGATAGGGTG	To retarget the intron for the gene *cphy2944*
2944EBS1d	CAGATTGTACAAATGTGGTGATAACAGATAAGTCCTGCAGCATAACTTACCTT TCTTTGT	
2944EBS2	TGAACGCAAGTTTCTAATTTCGGTTTTCTTCCGATAGAGGAAAGTGTCT	
OK2993F	TTCCACCTCCTCTCAGCTTG	To verify the intron insertion in *cphy2993*
OK2993R	GGATTGACAACGTGCGGTAA	
2993IBS	AAAACATATGATAATTATCCTTAATAGTCATAATCGTGCGCCCAGATAGGGTG	To retarget the intron for the gene *cphy2993*
2993EBS1d	CAGATTGTACAAATGTGGTGATAACAGATAAGTCATAATCATTAACTTACCTT TCTTTGT	
2993EBS2	TGAACGCAAGTTTCTAATTTCGGTTACTATCCGATAGAGGAAAGTGTCT	
OK1575F	CCGAATAGAACCTGGGGATATGG	To verify the intron insertion in *cphy1575*
OK1575R	TTCTGCTAAGGGTGCCACTC	
1575IBS	AAAACATATGATAATTATCCTTAATATCCTGCGTTGTGCGCCCAGATAGGGTG	To retarget the intron for the gene *cphy1575*
1575EBS1d	CAGATTGTACAAATGTGGTGATAACAGATAAGTCTGCGTTAATAACTTACCTT TCTTTGT	
1575EBS2	TGAACGCAAGTTTCTAATTTCGGTTGATATCCGATAGAGGAAAGTGTCT	

Supp Table 2 Precise site of intron insertion and score of the retargeting.

Loci	Intron insertion: Gene 5' exon - Intron - Gene 3' exon	Score	e-value
cphy2944	ATGCGTGGGGATGATGTTAAGAAGCTGCAG - intron - CAACTAATCGGAGTT	7.62	0.144
cphy2993	TTTTATTTTGACGAAGCGTACAAAATTAAA - intron - AAAGTATTTTTTCCG	9.8	0.026
Cphy1575	TGGCACATTGATAAGAGTATATCTTGCGTT - intron - AATATTTGTCTTATT	8.67	0.066

II.2.6 Conclusion and perspectives

In this chapter a new genomic editing tool was tested. This tool, called GETR was not tried before in other Clostridia. In theory this tool could enable the deletion and insertion of large DNA fragment.

We have successfully deleted 40 kb of a prophage region in *C. phytofermentans*; this achievement was reproducible and is the biggest deletion in Clostridia to date. Nevertheless the deletion occurred with an unexpected mechanism. This mechanism has to be clarified and reproduced elsewhere in the genome to be confirmed and released to the community. But this new genomic tool could be useful for the community, to delete large DNA sequence in one recombination, enabling to investigate a minimal genome project in Clostridia or the deletion of entire operons, etc.

Moreover, if this unexpected event can be reproduced somewhere else in the genome, it could be an interesting new genetic tool, for Clostridia and other bacteria, as no *lox* sites are remaining after the recombination. If the GETR is functional in *C. phytofermentans*, it would open the door to large-scale DNA deletion and insertion, which could be very advantageous to add entire metabolic pathways. To achieve the insertion *via* GETR a second origin of replication and a second antibiotic resistance cassette compatible with *C. phytofermentans* are required.

This work with *lox* sites and group II introns in Clostridia is new, but adding numerous *lox* site can be advantageous. For instance the Sc2.0 project replaced natural chromosome *in S. cerevisiae* by synthetic chromosome [431]. In these synthetic chromosomes, numbers of *loxP* sites were added. The expression of the Cre protein produces numerous and various recombination. This technique, called SCRaMbLE system, shows that many interesting important recombination appeared, with little effect on the phenotype, except if a selection pressure is added. A similar approach could be tested in Clostridia with the insertion of *loxP* thanks to group II introns.

II.3 Directed Evolution of C. phytofermentans

II.3.1 Introduction

Solventogenic, homoacetogenic and cellulolytic Clostridia are studied for the conversion of sugars (with different chain length) or syngas to valuable compounds, mainly to produce biofuels at industrial scales. In industrial and continuous cultures the fermentation can release various inhibitors, depending on the type of substrate. These inhibitors can reduce or prevent the production of the desired compound and the culture growth can be affected. Except in *E. coli* the inhibitor constraint in Clostridia was lightly studied [432]. The goal of this research is to depict the transcriptomic behavior of Clostridia with increasing concentration of inhibitor and force *C. phytofermentans* to evolve and resist to high concentration of inhibitors. Resistant strains are phenotypically and genetically characterized. This directed evolution technique was experimented with a GM3 machine [433], enabling the continuous culture of the strain. Moreover GM3 automat enables to fill in automatically the culture with pulses of stressing or relaxing medium, depending on the behavior of the culture. This enables to drive the evolution to resist to a stressing media. Directed-evolution strategies were also applied to make a culture resistant to its fermentation products (ethanol, butanol etc.).

This strategy takes advantage of the natural evolution of living organisms: selection of occurring mutations. Moreover, this way to make a strain resistant to a (toxic) compound explores solutions probably unknown and unpredictable; compared to the utilization of genetic engineering tool to add resistance capacity to the bacteria. The process of this strategy enables the selection of best evolved strain among millions of evolved bacteria in a batch culture.

This study was recently published in Applied and Environmental Microbiology. My participation to this work consisted in the following: I participated in the conception of the project, in the RNA-Seq analysis, and I performed several experiments (media preparation for the GM3 machine, real-time growth measurement etc.). I wrote some chapters and reviewed the manuscript. I did an oral presentation of this work at the Clostridium XIV conference in 2016 at Dartmouth College, NH, USA.

II.3.2 Article

AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology®

ENVIRONMENTAL MICROBIOLOGY

Evolution of a Biomass-Fermenting Bacterium To Resist Lignin Phenolics

Tristan Cerisy,[a,b,c,d] Tiffany Souterre,[a,b,c,d] Ismael Torres-Romero,[a,b,c,d] Magali Boutard,[a,b] Ivan Dubois,[a,b] Julien Patrouix,[a,b] Karine Labadie,[a,b] Wahiba Berrabah,[a,b] Marcel Salanoubat,[a,b,c,d] Volker Doring,[a,b,c,d] Andrew C. Tolonen[a,b,c,d]

CEA, Genoscope, Évry, France[a]; CNRS-UMR8030, Évry, France[b]; Université Paris-Saclay, Évry, France[c]; Université d'Évry, Évry, France[d]

ABSTRACT Increasing the resistance of plant-fermenting bacteria to lignocellulosic inhibitors is useful to understand microbial adaptation and to develop candidate strains for consolidated bioprocessing. Here, we study and improve inhibitor resistance in *Clostridium phytofermentans* (also called *Lachnoclostridium phytofermentans*), a model anaerobe that ferments lignocellulosic biomass. We survey the resistance of this bacterium to a panel of biomass inhibitors and then evolve strains that grow in increasing concentrations of the lignin phenolic, ferulic acid, by automated, long-term growth selection in an anaerobic GM3 automat. Ultimately, strains resist multiple inhibitors and grow robustly at the solubility limit of ferulate while retaining the ability to ferment cellulose. We analyze genome-wide transcription patterns during ferulate stress and genomic variants that arose along the ferulate growth selection, revealing how cells adapt to inhibitors through changes in gene dosage and regulation, membrane fatty acid structure, and the surface layer. Collectively, this study demonstrates an automated framework for *in vivo* directed evolution of anaerobes and gives insight into the genetic mechanisms by which bacteria survive exposure to chemical inhibitors.

IMPORTANCE Fermentation of plant biomass is a key part of carbon cycling in diverse ecosystems. Further, industrial biomass fermentation may provide a renewable alternative to fossil fuels. Plants are primarily composed of lignocellulose, a matrix of polysaccharides and polyphenolic lignin. Thus, when microorganisms degrade lignocellulose to access sugars, they also release phenolic and acidic inhibitors. Here, we study how the plant-fermenting bacterium *Clostridium phytofermentans* resists plant inhibitors using the lignin phenolic, ferulic acid. We examine how the cell responds to abrupt ferulate stress by measuring changes in gene expression. We evolve increasingly resistant strains by automated, long-term cultivation at progressively higher ferulate concentrations and sequence their genomes to identify mutations associated with acquired ferulate resistance. Our study develops an inhibitor-resistant bacterium that ferments cellulose and provides insights into genomic evolution to resist chemical inhibitors.

KEYWORDS clostridia, evolution, genomics

Fermentation of lignocellulosic biomass by bacteria like *Clostridium phytofermentans* is central to the function of soil, aquatic, and intestinal microbiomes. In addition, industrial fermentation of lignocellulosic biomass into fuels and chemicals could contribute significantly to global energy needs without impacting food production (1). Plant biomass is primarily composed of a macromolecular network of polysaccharides linked with lignin, a polymer of phenylpropanoid subunits with aromatic rings of various degrees of methoxylation (2). Thus, when microorganisms hydrolyze lignocel-

Received 2 February 2017 **Accepted** 24 March 2017

Accepted manuscript posted online 31 March 2017

Citation 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.

Editor Robert M. Kelly, North Carolina State University

Address correspondence to Andrew C. Tolonen, atolonen@genoscope.cns.fr.

T.C. and T.S. are co-first authors.

lulose to access sugars, they also liberate the following three main types of inhibitors: aliphatic acids, furans, and solubilized phenolics. The relative amounts of inhibitors depend on the species and condition of the plant matter (3), but hydrolysates generally contain inhibitors at concentrations that impede the growth of microorganisms (4) by damaging the cell membrane, metabolic enzymes, and nucleic acids (5). The most abundant aliphatic acids are generally acetate, particularly in acetylxylan-rich hardwoods (6), and formate from furan breakdown. The main furans are furfural and hydroxymethylfurfural (5-HMF), which are formed by the dehydration of pentose and hexose sugars, respectively. The most potent inhibitors released during biomass hydrolysis are generally phenolics released from lignin (7).

The resistance of model, sugar-fermenting bacteria, such as *Escherichia coli*, to biomass inhibitors has been well studied for aliphatic acids (8, 9), furans (7), and phenolics (10). However, much less is known about resistance in bacteria like *C. phytofermentans* that hydrolyze and ferment lignocellulose, even though plant inhibitors are important to the ecology of these species. Moreover, development of inhibitor-resistant microorganisms that directly metabolize biomass is needed for consolidated bioprocessing in a single reactor, which is generally regarded as the most economical configuration for microbial transformation of biomass into value-added chemicals (11).

Here, we study and increase resistance to plant-derived inhibitors in *C. phytofermentans*, an anaerobic bacterium in *Clostridium* cluster XIVa that expresses dozens of carbohydrate-active enzymes to degrade lignocellulosic biomass into hexoses and pentoses, which it then ferments to ethanol, H_2, and acetate (12). We initially define the effects of a panel of biomass inhibitors, including phenolics, furans, and aliphatic acids, on *C. phytofermentans* growth. Among these compounds, we focus on ferulic acid, a guaiacyl lignin precursor that is one of the most abundant phenolic inhibitors in woods, grasses, and agriculturally important crops (13). We examine the transcriptional response to ferulate stress by quantifying genome-wide mRNA expression changes. We apply long-term, anaerobic growth selection in a GM3 device (14) to isolate a series of increasingly ferulate-resistant strains. We examined the phenotypes of clones from along the selection and sequenced their genomes to identify positively selected genomic point mutations, small insertions and deletions (indels), and large structural rearrangements. Finally, we discuss how these results improve our understanding of the genetic basis of how bacteria evolve to resist chemical inhibitors.

RESULTS

Native *C. phytofermentans* inhibitor resistance. We measured the growth of *C. phytofermentans* in various concentrations of 12 lignocellulosic inhibitors (see Table S1 in the supplemental material) to gain a general understanding of the relative effects of aliphatic acids, furans, and phenolics (Fig. 1; see also Fig. S1 in the supplemental material). Both aliphatic acids reduce growth; acetate (Fig. 1A) was less toxic than formate (Fig. 1B) on a mass per volume basis, but both acids had similar effects in terms of molarity (Fig. S1). At low furan concentrations, we observed normal growth rates after an extended lag phase (Fig. 1C and D), similar to other bacteria that reduce and detoxify furans (15, 16). Growth lags are proposed to be due to alcohol dehydrogenase (ADH) reducing furan, causing NADH depletion and acetaldehyde accumulation (17). Additionally, the ADH protein Cphy1179 shares 28% amino acid identity with a furfural-reducing, Zn-dependent ADH (18). However, if *C. phytofermentans* detoxifies furans, this mechanism is abruptly overwhelmed at concentrations above 2 g liter^{-1} 5-HMF and 1 g liter^{-1} furfural.

We examined the toxicities of two types of phenolic acids: hydrocinnamic acids (*p*-coumarate and ferulate) and hydroxybenzoic acids (vanillate and 4-hydroxybenzoic acid). We found that hydrocinnamic acids (Fig. 1E and F) are more toxic than hydroxybenzoic acids (Fig. 1G and H), which supports that the propionic group on the benzene ring in hydrocinnamic acids enhances toxicity, likely by affecting how the molecules partition into the membrane. Moreover, we found that phenolic acids are typically less toxic than the corresponding aldehydes (Fig. 1I to K) and catechol (Fig. 1L). For

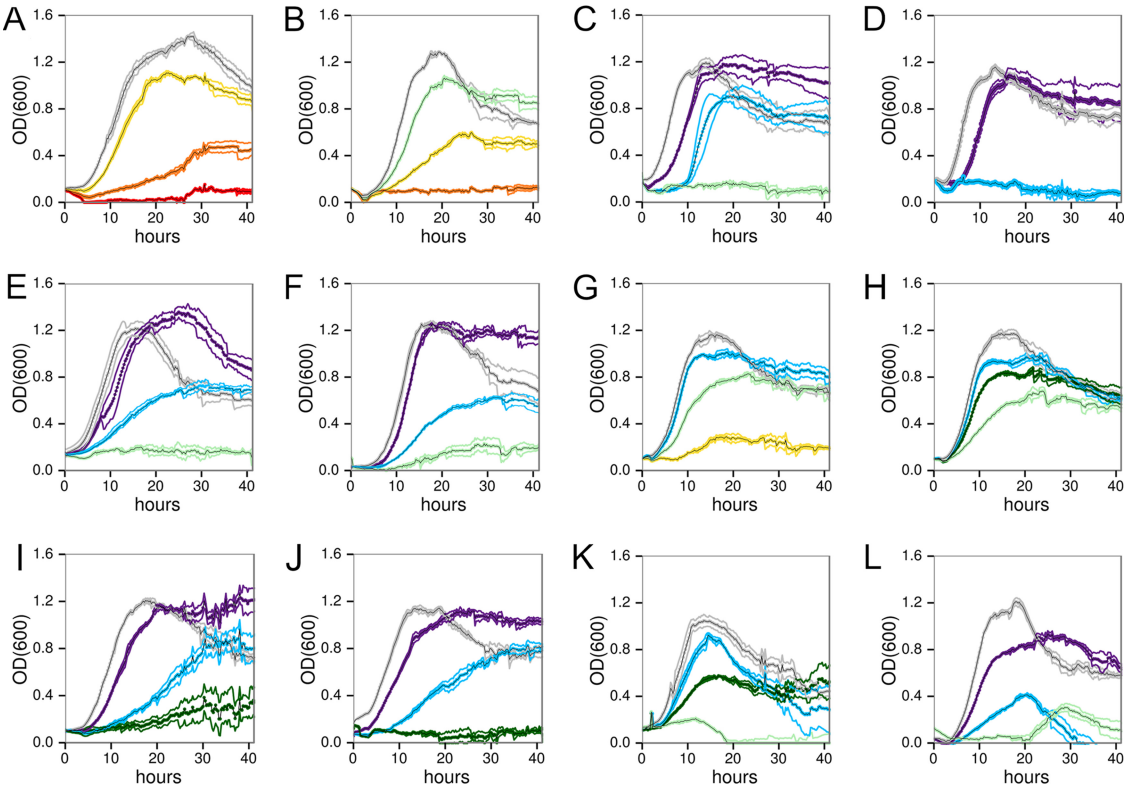

FIG 1 *C. phytofermentans* growth in GS2 glucose medium containing different concentrations of acetate (A), formate (B), 5-hydroxymethylfurfural (C), furfural (D), coumarate (E), ferulate (F), vanillate (G), 4-hydroxybenzoic acid (H), vanillin (I), benzaldehyde (J), syringaldehyde (K), and catechol (L). Colors show inhibitor concentrations (g liter^{-1}): gray (0), purple (1), blue (2), dark green (3), light green (5), yellow (10), orange (20), and red (30). Data show mean cell density (OD_{600}) of 4 cultures ± standard deviation (SD).

example, vanillate (Fig. 1G) is much less toxic than vanillin (Fig. 1I), and 4-hydroxybenzoic acid (Fig. 1H) is similarly less toxic than benzaldehyde (Fig. 1J). The enhanced toxicity of aldehydes is likely due to their reactivity, resulting in formation of adducts with nucleophilic sites on DNA, proteins, and other macromolecules (19).

Genome-wide mRNA expression during ferulate stress. We quantified genome-wide mRNA expression changes at two time (*t*) points ($t = 0.5$ h and $t = 4$ h) following supplementation of mid-log cultures with 2 g liter^{-1} ferulate, which reduced growth (Fig. 2A) similar to that of the initial growth screen (Fig. 1F). Three to five million read pairs were aligned to the genome for each culture (see Table S3A in the supplemental material) to calculate gene expression levels (Table S3B). The number of differentially expressed genes (Table S3C to E) increased from 0 genes before ferulate addition (Fig. 2B) to 78 genes after 30 min (Fig. 2C) and then declined to 47 genes after 4 h (Fig. 2D). The most abundant functional categories of differentially expressed genes at a *t* of 0.5 h relate to the repression of energy production, coenzyme metabolism, and lipids (Fig. 2E). The coenzyme-associated genes enable siroheme biosynthesis, which is repressed in clostridia in response to redox stress (20). Lipid genes include the *fab* gene cluster (*cphy0516-cphy0523*) for fatty acid biosynthesis, which was strongly repressed at a *t* of 0.5 h (Fig. 2E) and recovered by a *t* of 4 h. While cultures continued active growth after sampling, many of the differences between cultures with or without ferulate at a *t* of 4 h indicate that the treatment without ferulate had depleted nutrients in the medium, triggering expression of genes to assimilate alternative carbohydrates (Fig. 2F).

Gene expression at a *t* of 0.5 h shows that abrupt ferulate stress induces expression of genes encoding the efflux pump *cphy1055-cphy1056*, which is similar to *E. coli mdlAB* conferring resistance to organic solvents (21). Many of the genes upregulated at a *t* of 0.5 h are colocated in two genomic regions. The first region encodes *tad* (tight adherence) *cphy0029-cphy0040* genes for Flp-type type IV pilus assembly. Type IV pili

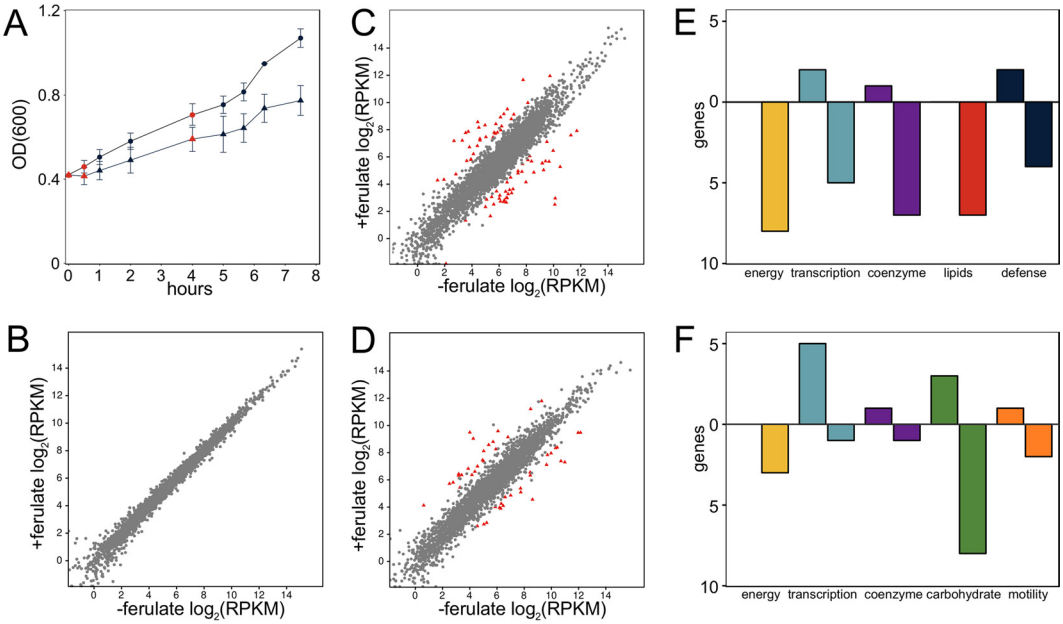

FIG 2 *C. phytofermentans* growth and gene expression during ferulate stress. (A) Growth in log-phase cultures in medium either lacking ferulate (−ferulate, circles) or containing 2 g liter^{-1} ferulate (+ferulate, triangles). Points are mean cell density (OD_{600}) of duplicate cultures ± SD with red points showing times sampled for RNA-seq: $t = 0$ h (immediately before dilution), $t = 0.5$ h, and $t = 4$ h. (B to D) mRNA expression from cultures at $t = 0$ h (B), $t = 0.5$ h (C), and $t = 4$ h (D). Differentially expressed genes are identified by red triangles; unchanged genes are gray circles. (E and F) Five most abundant COG functional categories (64) of differentially expressed genes at $t = 0.5$ h (E) and $t = 4$ h (F). Positive *y* axis is upregulated genes and negative *y* axis is repressed genes.

are widespread in clostridia (22) for adhesion to solid substrates to form protective biofilms (23), reflecting how ferulate represses motility genes in *Clostridium beijerinckii* (24). The other cluster *cphy1838-cphy1845* includes genes for the flavin mononucleotide (FMN)-binding protein WrbA (25) and two NADPH:FMN reductases. NADPH:FMN reductase inactivation confers ferulate resistance in *C. beijerinckii* by an unknown mechanism (26). While this appears to be in opposition to our data showing that NADPH:FMN reductases are upregulated by ferulate, both results support the importance of FMN-mediated oxidoreduction in ferulate resistance. This island also includes genes encoding an acetyltransferase and Cphy1845 that shares 41% amino acid identity and metal coordination with *E. coli* YhhW, which cleaves the plant phenolic quercetin (27). *C. phytofermentans* may thus upregulate genes to transform or detoxify plant phenolics, similar to some ruminal clostridia (28).

Selection and physiology of ferulate-resistant strains. We selected *C. phytofermentans* strains with increased ferulate resistance by cultivation in a GM3 automat, a dual-chamber, continuous-culture device that automates delivery of fresh medium and transfers the evolving cell suspension between twin growth chambers to prevent biofilm formation. During acclimation to increased ferulate in medium-swap mode, cell densities oscillated for 2 to 5 days because high densities triggered pulses of stressing medium (high ferulate) that reduced culture density, which in turn resulted in delivery of relaxing medium (low ferulate) that enabled recovery (Fig. 3A). As such, the ferulate-based selection in medium-swap mode is modulated by the ratio of relaxing and stressing medium. Once cell densities stabilized in the stressing medium, the growth rate at the higher ferulate concentration was improved in turbidostat mode (Fig. 3B). We initiated the growth selection with a stressing medium containing 1 g liter^{-1} ferulate, the highest concentration at which we could establish a stable wild-type (WT) culture in the GM3. After 93 days (~500 generations) of continuous, log-phase growth selection with incrementally higher ferulate, the culture grew with the same 3.75-h generation time in the 3 g liter^{-1} ferulate medium as that of the WT in the absence of ferulate (Fig. 3C). Clones isolated along the growth selection are progressively more

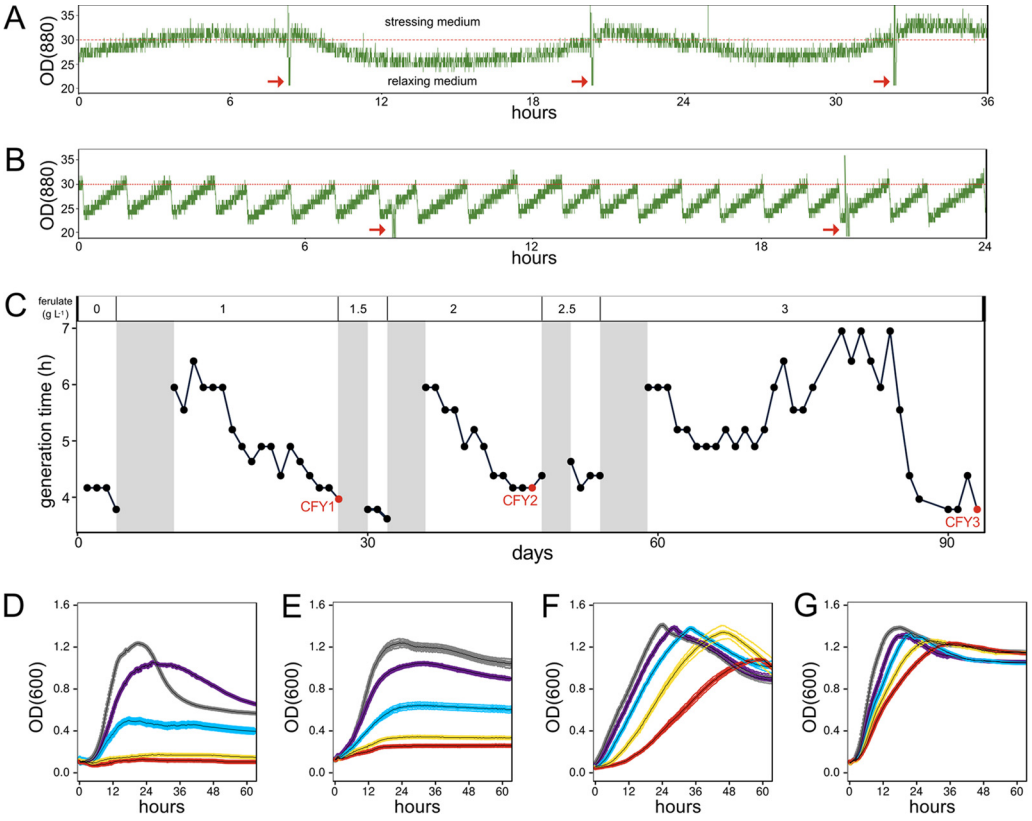

FIG 3 Growth improvement of *C. phytofermentans* GM3 strains in ferulate medium. (A) Cells were acclimated to increased ferulate using medium-swap mode, a chemostat with dilutions of stressing medium if density exceeds a threshold and with relaxing medium otherwise. (B) Growth rate was improved using turbidostat mode in which the culture was diluted each time it reached the threshold. (A and B) Dashed red lines show cell density threshold (OD_{880} of 30), and red arrows show when the growth chamber was sterilized. (C) *C. phytofermentans* growth rate over 93 day GM3 experiment in medium with increasing ferulate concentrations (shown above plot). Shaded areas are periods of medium-swap with fixed 6-h generation time. Black lines show average daily generation time (h) during turbidostat growth selection. Red points are sample times for physiology and genome sequences (CFY1, CFY2, CFY3). (D to G) Batch culture growth (OD_{600}) of wild-type (D) and clones CFY1A (E), CFY2C (F), and CFY3E (G) in GS2 glucose medium containing either 0 (gray), 1 (purple), 2 (blue), 4 (yellow), or 6 (red) g liter^{-1} ferulate. Data show mean cell density (OD_{600}) of 4 cultures ± SD.

ferulate resistant in batch culture (Fig. 3D to G); while no growth was observed above 2 g liter^{-1} ferulate in the WT strain (Fig. 3D), CFY3 clones grow robustly at the ferulate solubility limit (6 g liter^{-1}) (Fig. 3G). We assessed the ferulate resistance of 2 clones from each of the CFY1 (CFY1A and CFY1B) and CFY2 (CFY2C and CFY2D) time points and 4 clones from the CFY3 time point (CFY3E to CFY3H). The duplicate CFY1 and CFY2 clones showed similar ferulate resistance, but CFY3H is much less ferulate resistant than the 3 other clones (see Fig. S2 in the supplemental material), showing that cells in the GM3 culture are heterogeneous with respect to ferulate resistance.

We examined whether selection for ferulate resistance in glucose medium resulted in physiological changes impacting cellulose fermentation and resistance to other inhibitors. CFY3 strains degrade cellulose similar to that in the WT (Fig. 4A) and show accelerated cellulose degradation in medium supplemented with ferulate (Fig. 4B), supporting that the evolved strains are potentially improved candidates for fermentation of lignocellulose. Moreover, the evolved resistance mechanisms extend to other biomass inhibitors, as CFY3 strains are also more resistant to vanillate and acetate (see Fig. S3 in the supplemental material), albeit with considerable variability between strains. We also used mass spectrometry to investigate if ferulate was consumed or transformed in WT and CFY3 cultures, revealing that the ferulate concentration was unaltered with no products corresponding to reduced, demethoxylated, or decarboxylated ferulate (see Fig. S4 in the supplemental material). Thus, even though *C. phytofermentans* upregulates potential phenol-degrading enzymes in response to ferulate,

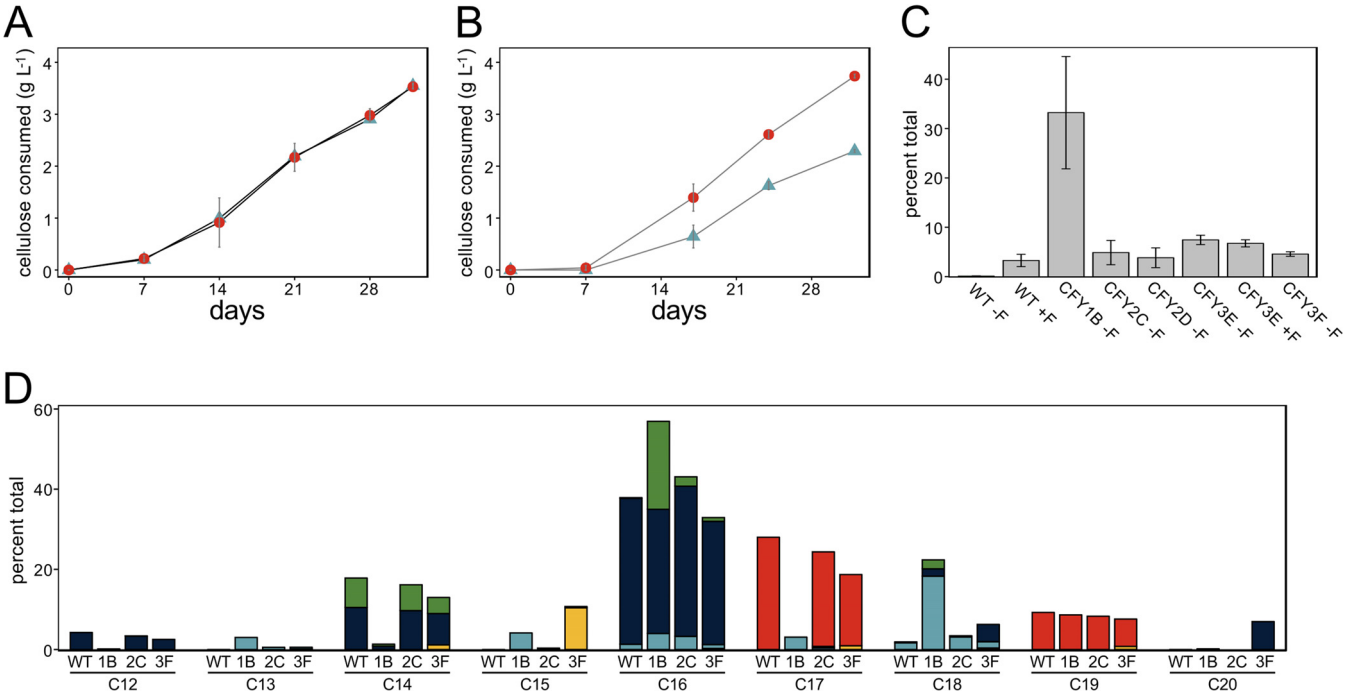

FIG 4 Growth physiology and membrane fatty acid composition of *C. phytofermentans* WT and GM3 strains. (A and B) Rate of cellulose degradation by CFY3E (red circles) and WT (blue triangles) in medium lacking ferulate (A) and medium supplemented with 2 g liter^{-1} ferulate (B). (C) Plasmalogen content expressed as percentage of total fatty acids of WT and GM3 strains grown in the presence (+F) or absence (−F) of ferulate. (D) Cellular fatty acid profiles of log-phase WT, CFY1B, CFY2C, and CFY3F cultures in medium without ferulate. Fatty acids are classified by acyl chain length (C_{12} to C_{20}) and whether acyl chains were saturated (dark blue), unsaturated (light blue), hydroxylated (green), cyclopropane (red), or branched (yellow). (A to C) Data show mean of duplicate cultures ± SD.

the cell adapted to ferulate by reinforcing the cell or excluding this molecule rather than detoxifying it.

As the toxicity of aromatic molecules is often associated with disruption of the cell membrane, we profiled fatty acids (FAs) to determine if ferulate resistance is associated with altered membrane phospholipids (see Table S4 in the supplemental material). We found that when WT was exposed to ferulate, the plasmalogen (vinyl ether phospholipid) content in the membranes increased 18-fold. Moreover, CFY strains retained elevated plasmalogens even in the absence of ferulate (Fig. 4C). In particular, the CFY1B plasmalogen content in the medium without ferulate was 185-fold higher than that of the WT. Related clostridia similarly increase plasmalogens in response to aliphatic alcohol stress (29, 30), likely to fine tune membrane fluidity and protect from redox-mediated damage (31). The distribution of FA chain lengths in WT cells (Fig. 4D) is similar to that of other clostridia but with fewer unsaturated FAs and more cyclopropanes (32), both of which reduce membrane fluidity to protect from solvent stress (33). While the addition of ferulate had little immediate effect on the FA chains of WT cells (see Fig. S5A and B in the supplemental material), the CFY strains showed altered FAs relative to those of the WT in the absence of ferulate (Fig. 4D). The CFY1B FA profile was the most perturbed with increased hydroxylated C_{16} and unsaturated fatty acids, largely $C_{18:1}$, which is associated with increased ethanol tolerance in *E. coli* (34). CFY3F shifted to branched FA (especially C_{15}) and longer chain lengths (C_{18}, C_{20}), which increase membrane rigidity (10) to potentially combat the membrane-fluidizing effects of ferulate. *C. phytofermentans* fatty acids are decorated with a diversity of phospho, glyco, and amino head groups (Fig. S5C). While we did not detect changes in these head groups in the WT response to ferulate or in the CFY strains, we consider it likely that they participate in the response to solvents, similar to some other bacteria (10).

Genomes of ferulate-resistant isolates. We sequenced the genomes of eight CFY1 to CFY3 clones, giving between 106- and 705-fold coverage (see Table S5A in the

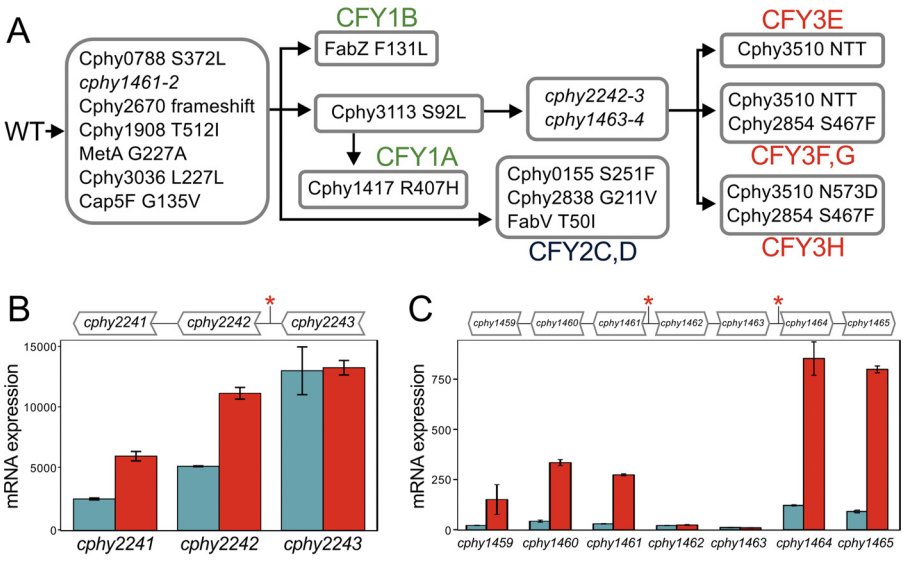

FIG 5 Small-scale genome differences in *C. phytofermentans* GM3 strains. (A) Accumulation of single-nucleotide variants and small indels in the genomes of clones isolated from the CFY1 to CFY3 time points. (B and C) mRNA expression of the *cphy2241-cphy2243* operon (B) and the *cphy1459-cphy1465* genes (C) in WT (blue) and CFY3E (red) strains. Genes are shown above plots with asterisks denoting positions of DNA changes. Expression was measured by qRT-PCR and quantified as $2^{-\Delta CT}$ normalized to 16S rRNA expression; bars show means of triplicate measurements ± SD.

supplemental material) to identify DNA variants relative to the wild type (Table S5B). Seven single-nucleotide variations (SNVs) and short insertions/deletions (indels) are present in all of the CFY genomes (Fig. 5A), which likely fixed in the population during an early selective sweep. These variants caused nonsynonymous changes in 5 proteins, including a homolog of Cap5F (Cphy3503), a protein for biosynthesis of capsular polysaccharides (35) that is associated with biofilm formation (36) and stress resistance (37). Strains subsequently accrued strain-specific mutations consistent with the population exploring alternative mutational pathways to improve ferulate resistance, particularly by modifying sensor kinases that can transduce signals associated with ferulate stress, fatty acid biosynthesis, and the surface layer (S-layer) (Fig. 5A). For example, the CFY1 and CFY2 strains incurred coding variants in 3 genes putatively encoding fatty acid biosynthesis proteins: Cphy3113 for anaerobic synthesis of unbranched fatty acids (38), the fatty acid dehydratase FabZ (Cphy0520), and the reductase FabV (Cphy1286) for the final step in fatty acid elongation. The genomes of CFY3E to CFY3G (high resistance) and CFY3H (low resistance) differ by variants in Cphy3510, the most highly expressed protein in the proteome that is proposed to form the S-layer (39). The S-layer is a protein lattice that provides mechanical stabilization, sites for extracellular protein attachment, and a selective barrier for molecules (40).

Intergenic changes that arose in the CFY genomes affect the expression levels of adjacent genes. For example, a 15-bp sequence between the first two genes of the ABC glucose transporter operon (*cphy2241-cphy2243*) was duplicated in the CFY3 strains (see Fig. S6A and B in the supplemental material). The repeated sequence forms an inverted repeat (IR) similar to repeated extragenic palindrome (REP) sequences, a widespread mechanism in bacteria to tune gene expression by modulating the stability of different mRNA segments within an operon (41). Duplication of this putative REP increases the mRNA secondary structure of the *cphy2243-cphy2242* intergenic region (Fig. S6B), supporting functions similar to those of REP that increase expression by forming stable stem-loop structures that protect mRNA from ribonucleases (42, 43). Similarly, we found that mRNA expression of the two genes downstream of the insertion was elevated (Fig. 5B), which may have increased fitness because the GM3 growth selections were done in glucose medium. The mRNA expression of genes in two colocated operons with upstream point mutations was upregulated in the CFY strains

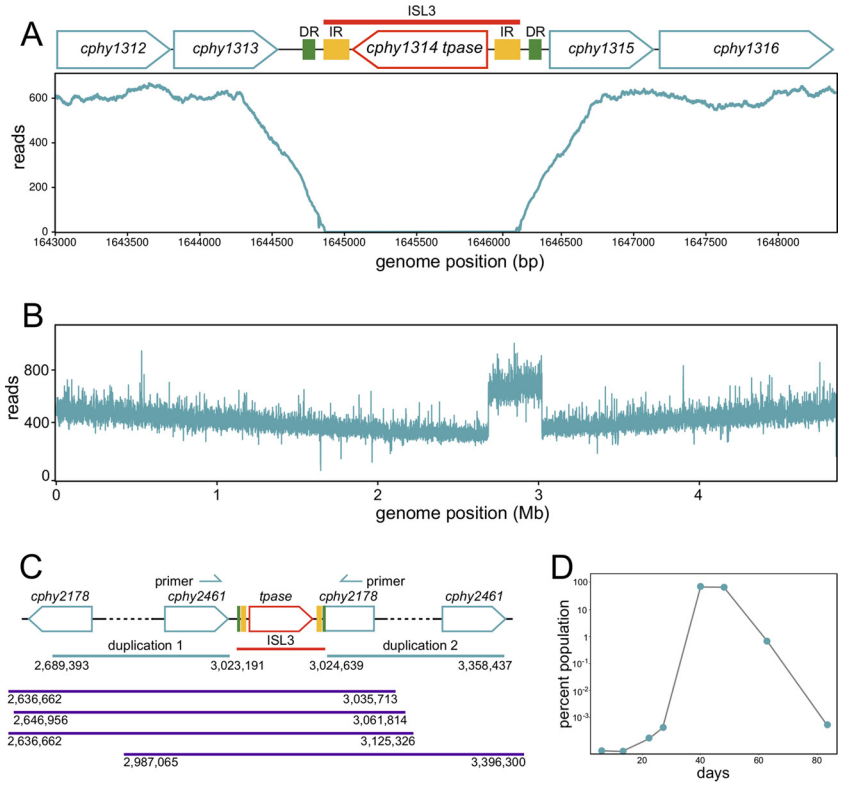

FIG 6 ISL*3*-associated genome changes in GM3 strains. (A) Read coverage showing deletion of the ISL*3-1* element including the *cphy1314* transposase gene (CFY1A shown). (B) Read coverage in CFY2C and CFY2D (shown) reveals the duplication of a 333-kb region from *cphy2178-cphy2461* (genome position 2689393 to 3023191). (C) The *cphy2178-cphy2461* region is a tandem duplication joined by an ISL*3-2* element. Positions of qPCR primers and single-molecule restriction fragments of >400 kb from BioNano optical mapping that span the duplicated region are shown. (D) Relative abundance of the *cphy2178*-ISL*3*-*cphy2461* junction in the GM3 culture from day 6 to 83 as measured by qPCR.

(Fig. 5C). The A-to-G transition upstream of *cphy1464* created a TG dinucleotide 2 bp upstream of the Pribnow hexamer (Fig. S6A) that enhances transcription in other bacteria (44, 45) and is present in the consensus −10 promoter sequence in *C. phytofermentans* (46). We propose that the upregulated operons *cphy1459-cphy1461* and *cphy1464-cphy1465* either enable increased production of malonyl coenzyme A (malonyl-CoA) for fatty acid biosynthesis or neutralize intracellular pH in response to ferulic acid stress through production of ammonium and lactate (7) and bicarbonate buffering (Fig. S6C).

Adaptive function can be imparted by structural changes to the genome resulting from recombination and transposition of insertion sequences (IS elements). *C. phytofermentans* encodes 31 IS elements (see Table S6 in the supplemental material), including 12 ISL*3* comprised of 2 isoforms—8 ISL*3-1* elements and 4 ISL*3-2* elements. IS elements inactivate genes through their transposition and act as the substrates for homologous recombination. In addition, the IRs of all 12 ISL*3* contain a 5′-TTGACA-3′ sequence matching an outward-facing, consensus −35 box from this organism (46) (Table S6), suggesting that ISL*3* could activate expression of adjacent genes (47). An ISL*3-1* was precisely deleted in all CFY genomes as evidenced by reduced read coverage (Fig. 6A) as well as BioNano optical mapping and Sanger sequencing (see Fig. S7A and B in the supplemental material), showing that ISL*3* are active in *C. phytofermentans*. Further, the CFY2C and CFY2D strains share a 333-kb duplication from *cphy2178* to *cphy2461* (276 genes) (Fig. 6B), which we showed by PCR exists as a tandem duplication joined by a novel ISL*3-2* insertion (Fig. 6C; Fig. S7C). We did not observe any extrachromosomal DNA by pulsed-field gel in the CFY2C strain, supporting that the *cphy2178-*

ISL*3-cphy2461* fragment did not excise as a circular molecule. Further, BioNano sequencing of DNA molecules greater than 400 kb spanning the junctions of the duplicated region localizes the rearrangement as a genomic, tandem duplication (Fig. 6C). We quantified the relative abundance of cells bearing this duplication by quantitative PCR of the *cphy2178*-ISL*3-cphy2461* fragment (Fig. 6D). The duplication arose between day 13 and 20 and overtook the population to comprise 68% of cells by day 40; supporting it was the subject of positive selection. Subsequently, this variant declined in the population, representing 1% of cells at day 63, as it was gradually replaced by mutants with higher fitness; it was not present in any of the CFY3 genomes.

DISCUSSION

When plant-fermenting bacteria like *C. phytofermentans* degrade lignocellulosic biomass to access sugars, they also release various biomass-derived inhibitors, including ferulic acid. The abilities of these bacteria to survive exposures to these inhibitors thus influence both their ecology and industrial potential. Fatty-acid (FA) biosynthesis genes, such as *cphy0520* (*fabZ*) and *cphy3113*, were associated with both the transcriptional and evolutionary response to ferulate. In particular, the FA profiles were massively perturbed in CFY1B, which has a strain-specific variant in FabZ, a dehydratase for unsaturated fatty acid synthesis. FabZ inactivation in CFY1B is consistent with the accumulation of hydroxylated C_{16} and decrease of C_{17} cyclopropanes, which are synthesized from $C_{16:1}$. Our results also show important differences in the cellular processes implicated in the mRNA response to short-term stress and the DNA changes enabling long-term resistance. Faced with an abrupt increase in ferulate, the cell slows growth and upregulates transcription of genes for efflux pumps, biofilm formation, and flavoproteins, including two NADPH:FMN reductases that are associated with ferulate stress in *C. beijerinckii* (26). Over longer time periods, natural selection of strains with robust growth in the presence of ferulate resulted in DNA changes associated with metabolism, gene regulation, and the cell surface (S-layer).

In Gram-negative bacteria, the outer membrane protects against influx of toxic compounds. Gram-positive bacteria lack this outer membrane and instead have a thick peptidoglycan wall that cannot exclude solvents. Consequently, Gram-positive bacteria are generally more sensitive to hydrophobic solvents (10). The S-layer is a lattice, often composed of a single protein, that covers many Gram-positive bacteria. It functions as a permeability barrier (40), potentially excluding ferulate, and stabilizes the cell through noncovalent linkage to cell wall polysaccharides using threonine residues (48). A three-residue (asparagine-threonine-threonine) insertion near the C terminus of the S-layer protein Cphy3510 was the sole mutation exclusively present in the three highly resistant CFY3 strains (CFY3E to CFY3G); the CFY3 strain with lower resistance (CFY3H) had an N573D variant in Cphy3510. S-layer proteins differ greatly among bacteria, and neither of the Cphy3510 variants are in known domains; however, our results support that modification of the S-layer may be an effective strategy to improve inhibitor resistance.

In addition to minor genomic changes (SNVs and indels) that alter gene expression or protein activity, both CFY2 genomes contain a tandem duplication of a 333-kb region joined by a novel ISL*3-2* insertion, supporting that this large genome rearrangement was positively selected during ferulate selection. Tandem duplications of large chromosomal regions have been detected in other bacteria (41) and can improve fitness by increasing gene dosage. A tandem duplication of regions joined by an IS element in *E. coli* was proposed to have arisen following insertion of IS elements into each copy of the duplication; when the IS elements recombined with each other, the intervening region was deleted to leave a single, central IS element (49). The duplication observed in CFY2 strains arose early in the experiment, perhaps because stress induced IS element activity. Strains containing the duplication rapidly took over the population, supporting that it enhanced fitness, then gradually declined to represent 1 in 10^5 cells at the end of the experiment (Fig. 6D), likely because this strain was outcompeted by others with higher fitness.

Our approach uses continuous, directed evolution as a framework for real-time study of natural selection by analyzing the succession of microbial strains with progressively higher fitness. Genome analysis of these strains using both high-coverage short reads and long-range optical mapping reveals both the small and large genomic changes that underlie a complex phenotype. When coupled with transcriptome sequencing (RNA-seq) to study the transcriptional response to abrupt change, this approach gives a portrait of how the cell adapts to a given perturbation on different time scales. These results can be applied to prioritize genes to engineer bacterial stress resistance. For example, abrupt ferulate stress could be mitigated by overexpressing efflux pumps and flavoproteins, whereas long-term ferulate resistance could be improved by altering the primary surface layer protein and membrane biosynthesis (*fab* genes) to favor longer fatty acids.

MATERIALS AND METHODS

Cell cultivation. *C. phytofermentans* ISDg (ATCC 700394) was cultured anaerobically in GS2 medium (50). The growth of batch cultures containing inhibitors (see Table S1 in the supplemental material) was measured in 100-well microtiter plates (Bioscreen 9502550) containing 400 μl of GS2 medium with 3 g liter^{-1} glucose supplemented with a given inhibitor neutralized to pH 7. Wells were inoculated with 1:10 volume cells grown to log phase in a medium without inhibitor. The plates were sealed in the anaerobic chamber (2% H_2, 98% N_2) by press-fitting adhesive sheets (Qiagen 1018104) (51) and incubated at 37°C in a Thermo Scientific Bioscreen C. The cell densities (optical density at 600 nm [OD_{600}]) were measured every 15 min with 30 s of shaking before each reading. Cellulose cultures were inoculated into GS2 containing 10 g liter^{-1} cellulose (0.5- by 5-cm strips of Whatman filter paper 1001-090, >98% cellulose content). Cellulose degradation was measured as the dry mass of cellulose remaining in culture by collecting the remaining cellulose on 11-μm-pore-size filters by vacuum filtration and drying it overnight at 65°C (52).

Ferulate-resistant *C. phytofermentans* clones were selected using a GM3 automat (14), a dual-chamber continuous-culture device that maintained anaerobic conditions by flushing cultures with 100% N_2 gas. A 50-ml culture was maintained at 30°C with optical density readings every 30 s and was transferred between growth chambers every 12 h to clean the empty chamber with 5 N sodium hydroxide. Cells were acclimated to increased ferulate using medium-swap mode, a modified chemostat (6-h generation time) with dilutions every 30 min of stressing medium (high ferulate) if the cell density exceeded the density threshold (measured as an OD_{880} of 30, which is equivalent to an OD_{600} of 0.4) and relaxing medium (low ferulate) otherwise. Once cell densities stabilized at a constant cell density in the stressing medium for 24 h, the GM3 was run as a turbidostat using the stressing medium until the culture reattained a 3.75-h generation time, similar to that of the WT strain in medium without ferulate. In turbidostat mode, 20% of the culture volume was replaced with fresh medium each time the cell density reached an OD_{880} of 30. Initially, the stressing medium contained 1 g liter^{-1} ferulate, the highest concentration at which a WT culture can be established in the GM3, and the relaxing medium lacked ferulate. The medium-swap/turbidostat approach was iterated by incrementing the stress medium by 0.5 g liter^{-1} ferulate and replacing the relaxing medium with the previous stressing medium. Samples from the GM3 culture were plated to isolate colonies, called CFY1, CFY2, and CFY3 clones, at the end of the turbidostat selections in 1, 2, and 3 g liter^{-1} ferulate, respectively.

RNA-seq. Log-phase cultures (OD_{600} of 0.8) of WT *C. phytofermentans* ISDg were diluted with 1 volume medium either lacking ferulate or containing 4 g liter^{-1} ferulate (2 g liter^{-1} final concentration). Samples for RNA and cell densities were taken from duplicate cultures for each treatment immediately before ferulate addition and 0.5 and 4 h afterward. Total RNA was extracted using TRI reagent (Sigma 93289), and 20 μg RNA was treated with 4 U Turbo DNase (Ambion AM2238) for 30 min at 37°C. RNA was purified by Zymoclean (Zymo Research R1015) to capture RNAs of >200 bp. Five micrograms of total RNA was depleted of rRNA by Ribo-Zero (Illumina MRZMB126), yielding 200 to 400 ng RNA, and purified by Zymo Concentrator-5 (total capture) into 10 μl of water. cDNA libraries were prepared from 100 ng RNA using the TruSeq stranded mRNA kit (Illumina 15031047) and sequenced on an Illumina HiSeq 2000 sequencer with paired-end 150-bp reads. Reads were aligned to the *C. phytofermentans* ISDg genome (NCBI accession number NC_010001.1) using Bowtie 2 (53). Gene expression was calculated as reads per kilobase of gene per million (RPKM) using the easyRNASeq bioconductor package (54). Differential expression was defined as a greater than 4-fold change in expression and a DESeq (55) *P* value of <0.01 after Bonferroni correction for multiple testing of the 3,902 genes in the genome.

Genome sequencing. Genomes were sequenced for clones isolated from the GM3 samples: CFY1 (2 clones, CFY1A and CFY1B), CFY2 (2 clones, CFY2C and CFY2D), and CFY3 (4 clones, CFY3E to CFY3H). Genomic DNA (15 to 20 μg) was extracted from 3-ml cultures using the Sigma GenElute bacterial genomic DNA kit (NA2110). DNA (100 to 250 ng) was fragmented by Covaris E220 (Covaris, Inc., Woburn, MA, USA) to a 600-bp mean fragment size. The DNA was end-repaired, 3′ A-tailed, and ligated to Illumina compatible adapters using the NEBNext DNA sample prep master mix set 1 (New England BioLabs E6040). Ligation products were purified with 1 volume solid-phase reversible immobilization (SPRI) beads (Beckman Coulter A63880) and amplified by 12 cycles of PCR using the Kapa Hifi HotStart NGS library amplification kit (Kapa Biosystems KK2611) with P5/P7 primers. PCR products were purified (0.8 volume SPRI beads) and run on a 2% agarose gel, and DNA (700 to 800 bp) was excised and purified using the

NucleoSpin extract II DNA purification kit (Macherey-Nagel 740609). cDNA libraries were sequenced using 300-bp paired-end reads on an Illumina MiSeq instrument. Reads were quality filtered by Picard (https://github.com/broadinstitute/picard) and aligned to the *C. phytofermentans* ISDg reference genome (NCBI accession number NC_010001.1) using Bowtie 2 (53). Sequence variants (single nucleotide polymorphisms [SNPs], indels) in the CFY strains relative to the reference genome were identified using GATK (56) as described previously (57). Structural variations were detected using the breseq split-read analysis tool (58). Insertion sequences (IS) were identified using ISfinder (59).

Optical genome mapping. High-molecular-weight DNA of strain CFY2C was extracted in agar plugs, which were solubilized with 0.4 U of GELase (Epicentre G09200) and dialyzed for 45 min. DNA was treated using IrysPrep reagent kit (BioNano Genomics) to prepare nicked, labeled, repaired, and stained (NLRS) DNA. Briefly, 300 ng of DNA was nicked with 10 U Nt.BspQI (NEB R0644S) for 2 h at 37°C. Nicked DNA was incubated for 1 h at 72°C with fluorescently labeled Alexa 546-dUTP and *Taq* polymerase (NEB M0273). Nicks were ligated using *Taq* ligase (NEB M0208) with deoxynucleoside triphosphates (dNTPs). DNA was counterstained with YOYO-1 (Life Technologies). NLRS DNA was loaded into IrysChips (BioNano Genomics), and data were collected on the Irys instrument (BioNano Genomics) until reaching ≥1,000-fold coverage of molecules ≥100 kb.

CFY2C DNA molecules were filtered using BioNano IrysView software (version 2.5.1) retaining molecules ≥100 kb with at least 6 label sites, yielding 32,359 molecules with an N_{50} of 172 kb. The NCBI assembly (NCBI accession number NC_010001) was *in silico* digested with BspQI (5′-GCTCTTC-3′) and used to align and assemble CFY2C molecules using the BioNano assembly pipeline (Pipeline version 4618; RefAligner and Assembler version 4704) with the parameters used for small genomes. Molecules ≥400 kb were aligned against the NCBI assembly with a tandem duplication of bp 2689393 to 3023191 joined by an ISL*3-2* element in order to identify molecules spanning the duplicated zone.

Quantitative PCR. We measured mRNA expression by quantitative reverse transcription-PCR (qRT-PCR) as described previously (60). Briefly, RNA was extracted from log-phase WT and CFY3E cultures as for RNA-seq. RNA was reverse transcribed (Applied Biosystems 4368814), and mRNA expression was quantified by quantitative PCR (qPCR) (KAPA KK4621) with primers in Table S2 in the supplemental material. Expression values are means of triplicate measurements of duplicate cultures calculated by the threshold cycle method as $2^{-\Delta CT}$ (61), normalized to 16S rRNA levels and multiplied by a scaling factor of 10^6. To calculate the relative abundance of DNA variants in the GM3 cultures, genomic DNA was extracted as for genome sequencing from samples taken directly from the GM3 at 8 time points. The abundance of a DNA variant was measured by qPCR (KAPA KK4621) relative to 16S (primers in Table S2). The abundance of the DNA variant in the mixed population was calculated as $2^{-\Delta\Delta CT}$ relative to a CFY clone that bears the variant in 100% of cells.

Mass spectrometry and chromatography. Ferulate concentrations were compared in WT and CFY3E cultures after 5 days of growth in GS2 medium containing 6 g liter^{-1} ferulate by liquid chromatography-electrospray ionization-mass spectrometry (LC-ESI-MS) and tandem mass spectrometry (MS/MS) using a Dionex TCC-3000RS chromatographic system (Thermo Fisher Scientific) coupled to an Orbitrap Elite mass spectrometer (Thermo Electron Corporation) equipped with a heated electrospray ionization (HESI) source. High-pressure liquid chromatography (HPLC) separation was performed on a 5-μm, 4.6- by 150-mm SeQuant ZIC-pHILIC column (Merck) at 40°C with a flow rate of 0.5 ml min^{-1} and a mobile phase of 10 mM $(NH_4)_2CO_3$ pH 9.9 (phase A) and acetonitrile (phase B). Elution was conducted using the following gradient conditions: 2 min at 80% phase B, 20-min linear gradient from 80% to 40% of phase B, 8 min at 40% phase B, 5-min increase to 80% phase B, and 15 min of 80% phase B. The mass spectrometer was operated in ESI negative ion mode using a −4.5 kV ion spray, a 275°C capillary temperature, and a mass resolution of 60,000. Sheath gas, auxiliary gas, and sweep gas flow rates were set to 60, 10, and 2 arbitrary units, respectively. Mass spectra were analyzed using Xcalibur version 2.2 (Thermo Fisher Scientific).

Cellular fatty acids were analyzed in WT and CFY3E cultures with and without ferulate and CFY1B, CFY2C, CFY2D, and CFY3F cultures without ferulate. Late log-phase cells were collected by centrifugation from cultures grown in medium either with 2 g liter^{-1} ferulate or lacking ferulate. Fatty acid methyl esters (FAME) were obtained from 100 mg of cells by saponification, methylation, and extraction (62) and were identified using the DSMZ identification service (Braunschweig, Germany). Briefly, FAME mixtures were separated using the Sherlock microbial identification system (MIS) (Microbial ID, USA): an Agilent model 6890N gas chromatograph with a 5% phenyl-methyl silicone capillary column (0.2 mm by 25 m), a flame ionization detector, and an automatic sampler (Agilent model 7683A). Peaks were integrated and fatty acid names and percentages calculated using Sherlock MIS standard software (Microbial ID, USA). Plasmalogens were quantified as dimethyl acetyl fatty acids. Polar lipids were extracted from 100 mg cells using a chloroform–methanol–0.3% aqueous NaCl mixture (1:2:0.8, vol/vol/vol) by stirring overnight. Cells were centrifuged, and the polar lipids were recovered in the chloroform phase by adjusting the chloroform–methanol–0.3% aqueous NaCl mixture to 1:1:0.9 (vol/vol/vol). Polar lipids were resolved by two-dimensional (2D) silica gel thin-layer chromatography; dimension 1 was chloroform–methanol–water (65:25:4, vol/vol/vol), and dimension 2 was chloroform–methanol–acetic acid–water (80:12:15:4, vol/vol/vol/vol). Total lipids were detected using molybdatophosphoric acid, and specific functional groups were identified using spray reagents specific for defined functional groups (63).

Accession number(s). All data underlying the findings are fully available without restriction. Sequencing files in FASTQ format are available in the European Nucleotide Archive under study accession numbers ERP018602 (RNA-seq) and ERP018603 (whole-genome sequencing).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/AEM.00289-17.

SUPPLEMENTAL FILE 1, PDF file, 6.3 MB.
SUPPLEMENTAL FILE 2, XLS file, 0.1 MB.
SUPPLEMENTAL FILE 3, XLS file, 0.1 MB.
SUPPLEMENTAL FILE 4, XLS file, 2.3 MB.
SUPPLEMENTAL FILE 5, XLS file, 0.1 MB.
SUPPLEMENTAL FILE 6, XLS file, 0.1 MB.
SUPPLEMENTAL FILE 7, XLS file, 0.1 MB.

ACKNOWLEDGMENTS

This work was funded by the Genoscope-CEA, a CNRS Chaire d'Excellence (to A.C.T.), and the Agence Nationale de la Recherche Grant ANR-16-CE05-0020 (to A.C.T.).

We thank P. Marlière for the GM3, E. Darii for mass spectrometry, and the Genoscope-CEA sequencing platform for DNA and RNA sequencing.

We declare no competing financial interests.

REFERENCES

1. Metzger JO, Hüttermann A. 2009. Sustainable global energy supply based on lignocellulosic biomass from afforestation of degraded areas. Naturwissenschaften 96:279–288. https://doi.org/10.1007/s00114-008-0479-4.
2. Klinke HB, Thomsen AB, Ahring BK. 2004. Inhibition of ethanol-producing yeast and bacteria by degradation products produced during pre-treatment of biomass. Appl Microbiol Biotechnol 66:10–26. https://doi.org/10.1007/s00253-004-1642-2.
3. Galbe M, Zacchi G. 2007. Pretreatment of lignocellulosic materials for efficient bioethanol production. Adv Biochem Eng Biotechnol 108:41–65.
4. Zaldivar J, Martinez A, Ingram LO. 1999. Effect of selected aldehydes on the growth and fermentation of ethanologenic *Escherichia coli*. Biotechnol Bioeng 65:24–33. https://doi.org/10.1002/(SICI)1097-0290(19991005)65:1<24::AID-BIT4>3.0.CO;2-2.
5. Mills TY, Sandoval NR, Gill RT. 2009. Cellulosic hydrolysate toxicity and tolerance mechanisms in *Escherichia coli*. Biotechnol Biofuels 2:26. https://doi.org/10.1186/1754-6834-2-26.
6. Jönsson LJ, Alriksson B, Nilvebrant N-O. 2013. Bioconversion of lignocellulose: inhibitors and detoxification. Biotechnol Biofuels 6:16. https://doi.org/10.1186/1754-6834-6-16.
7. Ibraheem O, Ndimba BK. 2013. Molecular adaptation mechanisms employed by ethanologenic bacteria in response to lignocellulose-derived inhibitory compounds. Int J Biol Sci 9:598–612. https://doi.org/10.7150/ijbs.6091.
8. Axe DD, Bailey JE. 1995. Transport of lactate and acetate through the energized cytoplasmic membrane of *Escherichia coli*. Biotechnol Bioeng 47:8–19. https://doi.org/10.1002/bit.260470103.
9. Roe AJ, McLaggan D, Davidson I, O'Byrne C, Booth IR. 1998. Perturbation of anion balance during inhibition of growth of *Escherichia coli* by weak acids. J Bacteriol 180:767–772.
10. Weber FJ, de Bont JA. 1996. Adaptation mechanisms of microorganisms to the toxic effects of organic solvents on membranes. Biochim Biophys Acta 1286:225–245. https://doi.org/10.1016/S0304-4157(96)00010-X.
11. Lynd LR, van Zyl WH, McBride JE, Laser M. 2005. Consolidated bioprocessing of cellulosic biomass: an update. Curr Opin Biotechnol 16:577–583. https://doi.org/10.1016/j.copbio.2005.08.009.
12. Warnick TA, Methé BA, Leschine SB. 2002. *Clostridium phytofermentans* sp. nov., a cellulolytic mesophile from forest soil. Int J Syst Evol Microbiol 52:1155–1160. https://doi.org/10.1099/00207713-52-4-1155.
13. Skerker JM, Leon D, Price MN, Mar JS, Tarjan DR, Wetmore KM, Deutschbauer AM, Baumohl JK, Bauer S, Ibáñez AB, Mitchell VD, Wu CH, Hu P, Hazen T, Arkin AP. 2013. Dissecting a complex chemical stress: chemogenomic profiling of plant hydrolysates. Mol Syst Biol 9:674. https://doi.org/10.1038/msb.2013.30.
14. Mutzel R, Marliere P. June 2000. Method and device for selecting accelerated proliferation of living cells in suspension. WIPO patent WO/2000/034433.
15. Ezeji T, Qureshi N, Blaschek HP. 2007. Butanol production from agricultural residues: impact of degradation products on *Clostridium beijerinckii* growth and butanol fermentation. Biotechnol Bioeng 97:1460–1469. https://doi.org/10.1002/bit.21373.
16. Miller EN, Jarboe LR, Yomano LP, York SW, Shanmugam KT, Ingram LO. 2009. Silencing of NADPH-dependent oxidoreductase genes (*yqhD* and *dkgA*) in furfural-resistant ethanologenic *Escherichia coli*. Appl Environ Microbiol 75:4315–4323. https://doi.org/10.1128/AEM.00567-09.
17. Palmqvist E, Almeida JS, Hahn-Hägerdal B. 1999. Influence of furfural on anaerobic glycolytic kinetics of *Saccharomyces cerevisiae* in batch culture. Biotechnol Bioeng 62:447–454. https://doi.org/10.1002/(SICI)1097-0290(19990220)62:4<447::AID-BIT7>3.0.CO;2-0.
18. Kang C, Hayes R, Sanchez EJ, Webb BN, Li Q, Hooper T, Nissen MS, Xun L. 2012. Furfural reduction mechanism of a zinc-dependent alcohol dehydrogenase from *Cupriavidus necator* JMP134. Mol Microbiol 83:85–95. https://doi.org/10.1111/j.1365-2958.2011.07914.x.
19. LoPachin RM, Gavin T. 2014. Molecular mechanisms of aldehyde toxicity: a chemical perspective. Chem Res Toxicol 27:1081–1091. https://doi.org/10.1021/tx5001046.
20. Sander K, Wilson CM, Rodriguez M, Klingeman DM, Rydzak T, Davison BH, Brown SD. 2015. *Clostridium thermocellum* DSM 1313 transcriptional responses to redox perturbation. Biotechnol Biofuels 8:211. https://doi.org/10.1186/s13068-015-0394-9.
21. Foo JL, Jensen HM, Dahl RH, George K, Keasling JD, Lee TS, Leong S, Mukhopadhyay A. 2014. Improving microbial biogasoline production in *Escherichia coli* using tolerance engineering. mBio 5:e01932-14. https://doi.org/10.1128/mBio.01932-14.
22. Melville S, Craig L. 2013. Type IV pili in Gram-positive bacteria. Microbiol Mol Biol Rev 77:323–341. https://doi.org/10.1128/MMBR.00063-12.
23. Tomich M, Fine DH, Figurski DH. 2006. The TadV protein of *Actinobacillus actinomycetemcomitans* is a novel aspartic acid prepilin peptidase required for maturation of the Flp1 pilin and TadE and TadF pseudopilins. J Bacteriol 188:6899–6914. https://doi.org/10.1128/JB.00690-06.
24. Lee S, Lee JH, Mitchell RJ. 2015. Analysis of *Clostridium beijerinckii* NCIMB 8052's transcriptional response to ferulic acid and its application to enhance the strain tolerance. Biotechnol Biofuels 8:68. https://doi.org/10.1186/s13068-015-0252-9.
25. Grandori R, Khalifah P, Boice JA, Fairman R, Giovanielli K, Carey J. 1998. Biochemical characterization of WrbA, founding member of a new family of multimeric flavodoxin-like proteins. J Biol Chem 273:20960–20966. https://doi.org/10.1074/jbc.273.33.20960.
26. Liu J, Guo T, Shen X, Xu J, Wang J, Wang Y, Liu D, Niu H, Liang L, Ying H. 2016. Engineering *Clostridium beijerinckii* with the Cbei_4693 gene knockout for enhanced ferulic acid tolerance. J Biotechnol 229:53–57. https://doi.org/10.1016/j.jbiotec.2016.04.052.

27. Adams M, Jia Z. 2005. Structural and biochemical analysis reveal pirins to possess quercetinase activity. J Biol Chem 280:28675–28682. https://doi.org/10.1074/jbc.M501034200.
28. Schoefer L, Mohan R, Schwiertz A, Braune A, Blaut M. 2003. Anaerobic degradation of flavonoids by *Clostridium orbiscindens*. Appl Environ Microbiol 69:5849–5854. https://doi.org/10.1128/AEM.69.10.5849-5854.2003.
29. Kolek J, Patáková P, Melzoch K, Sigler K, Řezanka T. 2015. Changes in membrane plasmalogens of *Clostridium pasteurianum* during butanol fermentation as determined by lipidomic analysis. PLoS One 10: e0122058. https://doi.org/10.1371/journal.pone.0122058.
30. Timmons MD, Knutson BL, Nokes SE, Strobel HJ, Lynn BC. 2009. Analysis of composition and structure of *Clostridium thermocellum* membranes from wild-type and ethanol-adapted strains. Appl Microbiol Biotechnol 82:929–939. https://doi.org/10.1007/s00253-009-1891-1.
31. Nagan N, Zoeller RA. 2001. Plasmalogens: biosynthesis and functions. Prog Lipid Res 40:199–229. https://doi.org/10.1016/S0163-7827(01)00003-0.
32. Johnston NC, Goldfine H. 1983. Lipid composition in the classification of the butyric acid-producing clostridia. J Gen Microbiol 129:1075–1081.
33. Zhao Y, Hindorff LA, Chuang A, Monroe-Augustus M, Lyristis M, Harrison ML, Rudolph FB, Bennett GN. 2003. Expression of a cloned cyclopropane fatty acid synthase gene reduces solvent formation in *Clostridium acetobutylicum* ATCC 824. Appl Environ Microbiol 69:2831–2841. https://doi.org/10.1128/AEM.69.5.2831-2841.2003.
34. Ingram LO, Vreeland NS, Eaton LC. 1980. Alcohol tolerance in *Escherichia coli*. Pharmacol Biochem Behav 13(Suppl):S191–S195. https://doi.org/10.1016/S0091-3057(80)80030-X.
35. Kneidinger B, O'Riordan K, Li J, Brisson J-R, Lee JC, Lam JS. 2003. Three highly conserved proteins catalyze the conversion of UDP-*N*-acetyl-D-glucosamine to precursors for the biosynthesis of O antigen in *Pseudomonas aeruginosa* O11 and capsule in *Staphylococcus aureus* type 5. Implications for the UDP-*N*-acetyl-L-fucosamine biosynthetic pathway. J Biol Chem 278:3615–3627. https://doi.org/10.1074/jbc.M203867200.
36. Branda SS, Vik S, Friedman L, Kolter R. 2005. Biofilms: the matrix revisited. Trends Microbiol 13:20–26. https://doi.org/10.1016/j.tim.2004.11.006.
37. Ophir T, Gutnick DL. 1994. A role for exopolysaccharides in the protection of microorganisms from desiccation. Appl Environ Microbiol 60: 740–745.
38. Isabella VM, Clark VL. 2011. Identification of a conserved protein involved in anaerobic unsaturated fatty acid synthesis in *Neiserria gonorrhoeae*: implications for facultative and obligate anaerobes that lack FabA. Mol Microbiol 82:489–501. https://doi.org/10.1111/j.1365-2958.2011.07826.x.
39. Tolonen AC, Haas W, Chilaka AC, Aach J, Gygi SP, Church GM. 2011. Proteome-wide systems analysis of a cellulosic biofuel-producing microbe. Mol Syst Biol 7:461. https://doi.org/10.1038/msb.2010.116.
40. Sára M, Sleytr UB. 1987. Molecular sieving through S layers of *Bacillus stearothermophilus* strains. J Bacteriol 169:4092–4098. https://doi.org/10.1128/jb.169.9.4092-4098.1987.
41. Darmon E, Leach DRF. 2014. Bacterial genome instability. Microbiol Mol Biol Rev 78:1–39. https://doi.org/10.1128/MMBR.00035-13.
42. Newbury SF, Smith NH, Robinson EC, Hiles ID, Higgins CF. 1987. Stabilization of translationally active mRNA by prokaryotic REP sequences. Cell 48:297–310. https://doi.org/10.1016/0092-8674(87)90433-8.
43. Xu C, Huang R, Teng L, Jing X, Hu J, Cui G, Wang Y, Cui Q, Xu J. 2015. Cellulosome stoichiometry in *Clostridium cellulolyticum* is regulated by selective RNA processing and stabilization. Nat Commun 6:6900. https://doi.org/10.1038/ncomms7900.
44. Graves MC, Rabinowitz JC. 1986. *In vivo* and *in vitro* transcription of the *Clostridium pasteurianum* ferredoxin gene. Evidence for "extended" promoter elements in Gram-positive organisms. J Biol Chem 261: 11409–11415.
45. Helmann JD. 1995. Compilation and analysis of *Bacillus subtilis* sigma A-dependent promoter sequences: evidence for extended contact between RNA polymerase and upstream promoter DNA. Nucleic Acids Res 23:2351–2360. https://doi.org/10.1093/nar/23.13.2351.
46. Boutard M, Ettwiller L, Cerisy T, Alberti A, Labadie K, Salanoubat M, Schildkraut I, Tolonen AC. 2016. Global repositioning of transcription start sites in a plant-fermenting bacterium. Nat Commun 7:13783. https://doi.org/10.1038/ncomms13783.
47. Zafarullah M, Charlier D, Glansdorff N. 1981. Insertion of IS*3* can "turn-on" a silent gene in *Escherichia coli*. J Bacteriol 146:415–417.
48. Fagan RP, Fairweather NF. 2014. Biogenesis and functions of bacterial S-layers. Nat Rev Microbiol 12:211–222. https://doi.org/10.1038/nrmicro3213.
49. Raeside C, Gaffé J, Deatherage DE, Tenaillon O, Briska AM, Ptashkin RN, Cruveiller S, Médigue C, Lenski RE, Barrick JE, Schneider D. 2014. Large chromosomal rearrangements during a long-term evolution experiment with *Escherichia coli*. mBio 5:e01377-14. https://doi.org/10.1128/mBio.01377-14.
50. Cavedon K, Leschine SB, Canale-Parola E. 1990. Cellulase system of a free-living, mesophilic clostridium (strain C7). J Bacteriol 172:4222–4230. https://doi.org/10.1128/jb.172.8.4222-4230.1990.
51. 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.
52. Tolonen AC, Chilaka AC, Church GM. 2009. Targeted gene inactivation in *Clostridium phytofermentans* shows that cellulose degradation requires the family 9 hydrolase Cphy3367. Mol Microbiol 74:1300–1313. https://doi.org/10.1111/j.1365-2958.2009.06890.x.
53. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.
54. Delhomme N, Padioleau I, Furlong EE, Steinmetz LM. 2012. easyRNASeq: a bioconductor package for processing RNA-seq data. Bioinformatics 28:2532–2533. https://doi.org/10.1093/bioinformatics/bts477.
55. Anders S, Huber W. 2010. Differential expression analysis for sequence count data. Genome Biol 11:R106. https://doi.org/10.1186/gb-2010-11-10-r106.
56. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S, DePristo MA. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics 43:11.10.1–11.10.33.
57. Tolonen AC, Zuroff TR, Ramya M, Boutard M, Cerisy T, Curtis WR. 2015. Physiology, genomics, and pathway engineering of an ethanol-tolerant strain of *Clostridium phytofermentans*. Appl Environ Microbiol 81: 5440–5448. https://doi.org/10.1128/AEM.00619-15.
58. Barrick JE, Colburn G, Deatherage DE, Traverse CC, Strand MD, Borges JJ, Knoester DB, Reba A, Meyer AG. 2014. Identifying structural variation in haploid microbial genomes from short-read resequencing data using breseq. BMC Genomics 15:1039. https://doi.org/10.1186/1471-2164-15-1039.
59. Siguier P, Gourbeyre E, Chandler M. 2014. Bacterial insertion sequences: their genomic impact and diversity. FEMS Microbiol Rev 38:865–891. https://doi.org/10.1111/1574-6976.12067.
60. Tolonen AC, Cerisy T, El-Sayyed H, Boutard M, Salanoubat M, Church GM. 2015. Fungal lysis by a soil bacterium fermenting cellulose. Environ Microbiol 17:2618–2627. https://doi.org/10.1111/1462-2920.12495.
61. Schmittgen TD, Livak KJ. 2008. Analyzing real-time PCR data by the comparative C(T) method. Nat Protoc 3:1101–1108. https://doi.org/10.1038/nprot.2008.73.
62. Miller LT. 1982. Single derivatization method for routine analysis of bacterial whole-cell fatty acid methyl esters, including hydroxy acids. J Clin Microbiol 16:584–586.
63. Daniels L, Hanson RS, Phillips JA. 2007. Chemical analysis, p 462–503. *In* Reddy CA, Beveridge TJ, Breznak JA, Marzluf GA, Schmidt TM, Snyder LR (ed), Methods for general and molecular microbiology, 3rd ed. ASM Press, Washington, DC.
64. Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res 28:33–36. https://doi.org/10.1093/nar/28.1.33.

II.3.3 Conclusion and perspectives

This work demonstrates an automated strategy for directed evolution of anaerobes, which we apply here to improve inhibitor resistance in *C. phytofermentans* . This anaerobic continuous culture device is an interesting tool to evolve Clostridia in log phase.

This directed-evolution strategy was recently tested to improve the growth of *C. phytofermentans* in carboxymethyl cellulose (CMC) a viscuous cellulose derivative poorly assimilated by *C. phytofermentans*. Nevertheless the viscosity raised an issue in the machine, and the project was stopped. A similar approach to the inhibitor tolerance strategy of this article is applying to increase the resistance of *C. phytofermentans* for butanol, an interesting solvent for the industry.

C. phytofermentans does not grow well on several plant polysaccharides, uncluding arabinan, rhamnogalacturonan or galactomannan (Appendix 1, [51]). It should be interesting to drive the metabolism of *C. phytofermentans* for these substrates with the GM3 machine, increasing the wide and various numbers of polysaccharides used as substrate by this Clostridia. Engineering this strain with genetic tools could be more complicated, as numerous CAZymes and specific transporters should be added.

In my point of view, genetic engineering and directed-evolution strategies would not be contradictory but complementary. Both can be used to improve strains; complementary strategies could reach interesting results for the production and resistance to compounds.

PS: Supplemental material for this article may be found at:

http://aem.asm.org/content/83/11/e00289-17/suppl/DCSupplemental

III. CONCLUSION AND PERSPECTIVES

During this thesis many objectives were achieved, scientifically and personally.

Genesis and proceeding of the thesis

Originally I was recruited in 2013 for the project: hyper-saturated transposon mutagenesis followed by high throughput DNA sequencing to determine the transposon insertion sites that affect biomass fermentation. But several obstacles were quickly not solved; after transformation very few different transposon insertions were integrated, mainly because of the promoter strength which regulate the *mariner* transposon; and the poor transformation efficiency in *C. phytofermentans*. These various issues were too challenging to consider a hyper-saturated transposon mutagenesis project in three years. But I learnt how to work with *C. phytofermentans*, the conjugation protocol was mastered and I became familiar with Clostridia's bibliography.

Andrew Tolonen, Marcel Salanoubat and I decided to switch to another ambitious project: adapting to *C. phytofermentans* a genetic modification strategy recently developed *in S. aureus*, *B. subtilis* and *E. coli*. This strategy, named GETR, integrates or deletes large DNA sequence in bacteria, this was not achieved in Clostridia before. This GETR strategy is based on two technologies: transposons, which perfectly work in Clostridia; and recombination with the Cre-recombinase, a system utilized in all kingdoms. Nevertheless I was blocked for several months in this project, due to the incapacity to transform *C. phytofermentans* with a Cre expressing plasmid. During this temporary stop, A second project was developed using RNA-seq data from our previous article, which were lowly exploited, and the group II intron technology which was fully mastered, thanks to the GETR project. This project aims to characterize some sugar ABC transporters.

Sanitha Mary (a collaborator) and Magali Boutard spent months to develop an electroporation protocol; the transformation was the bottleneck of the GETR project in *C. phytofermentans*. The transformation and maintenance of a Cre-expressing plasmid with the electroporation enable the recombination at *lox* sites of *C. phytofermentans*. Whereas numbers of conjugation were tried with the same plasmid, without any transformant obtained.

In parallel, I participated to other project of our small team (Andrew Tolonen, Magali Boutard and I, plus temporary interns/collaborators) in *C. phytofermentans*: characterization of CAZymes (Appendix 1, [51]), elucidation of the role of chitinases (Appendix 3, [204]), analysis of

an ethanol-tolerant strain (Appendix 4, [216]), the study of TSS (Appendix 2, [188]) or the understanding of mechanisms involved in directed-evolution experiment (Chapter II.3).

Objectives

The goal of this thesis was to build *C. phytofermentans* as a model system for lignocellulosic fermentation by developing new genetic tools for the community (GETR, directed-evolution with the GM3 automat) and to study the sugar transportation system. The publication of the project related to the directed evolution of *C. phytofermentans* to lignocellulosic inhibitors and the submission of results related to the characterization of hexose-based ABC transporter show that part of achievements were reached. Moreover the recent results on the GETR technique in *C. phytofermentans* are very promising and I hope that future developments in this project will lead to an interesting publication. The last achievement could be related to the chapter "I.3 Clostridia genetics"; after improvements, the goal is to publish this chapter as a recent review of genetic tools in Clostridia.

Perspectives

This thesis is a very small part of all interesting project done in Clostridia and I hope this work participate to the development of Clostridia for industrial, environmental or medical purposes. The oral and poster presentation of this work during the Clostridium XIV conference (in 2016 at Dartmouth College, NH, USA) enable rich conversation with members of the Clostridia community.

In my point of view, numerous genetic engineering tools were and are developed for Clostridia (Group II introns, CRISPR-Cas, directed evolution, random mutagenesis etc.). GETR can be part of these useful when wide insertions or deletions are required. Moreover these tools, combined with characterized promoters, reporters, (transporters ?) and accurate mathematic models can bring the Clostridia community in the wider Synthetic Biology community, with natural advantages of Clostridia. The combination of these tools enables the future development of industrial strain of Clostridia, like *C. autoethanogenum* with the company Lanzatech. These tools can reach industrial projects but also medical projects, related to pathogen Clostridia or the microbiome.

The development of *C. phytofermentans* for industrial consolidated bioprocessing projects needs more detailed studies; it can be supported with these new genetic tools and future techniques and instruments. Perhaps *C. phytofermentans* or other cellulolytic Clostridia (for instance *C. thermocellum* or *C. cellulolyticum*) will be improved to be economically viable for the

plant biomass conversion to valuable compounds. By working in these strains, we probably participate to this ambitious achievement.

This work, more precisely GETR and the directed evolution strategy, will be pursued in the team in the project *Phytocell* with the help of an ANR grant. The goal of this project is to produce butanol, a high-order alcohol in *C. phytofermentans* and *C. cellulolyticum,* from plant biomass. To achieve this goal, these strains will be modified to increase butanol tolerance and by expressing a synthetic butanol production pathway.

BIBLIOGRAPHY

[1] F. J. Erbguth, "From poison to remedy: The chequered history of botulinum toxin," *J. Neural Transm.*, vol. 115, no. 4, pp. 559–565, 2008.

[2] D. T. Jones and D. R. Woods, "Acetone-butanol fermentation revisited.," *Microbiol. Rev.*, vol. 50, no. 4, pp. 484–524, 1986.

[3] J. Esty, "The Biology of Clostridium welchii," *J Bacteriol.*, pp. 375–429, 1920.

[4] G. F. Reddish and L. E. O. F. Rettger, "Clostridium Putrificum (B. putrificus Bienstock), a distinct species," no. 1917, pp. 505–510, 1922.

[5] J. T. Heap, S. A. Kuehne, M. Ehsaan, S. T. Cartman, C. M. Cooksley, J. C. Scott, and N. P. Minton, "The ClosTron: Mutagenesis in Clostridium refined and streamlined," *J. Microbiol. Methods*, vol. 80, no. 1, pp. 49–55, 2010.

[6] G. Z. Cui, W. Hong, J. Zhang, W. L. Li, Y. Feng, Y. J. Liu, and Q. Cui, "Targeted gene engineering in Clostridium cellulolyticum H10 without methylation," *J. Microbiol. Methods*, vol. 89, no. 3, pp. 201–208, 2012.

[7] E. E. Putnam, A. M. Nock, T. D. Lawley, and A. Shen, "SpoIVA and sipl are Clostridium difficile spore morphogenetic proteins," *J. Bacteriol.*, vol. 195, no. 6, pp. 1214–1225, 2013.

[8] J. Kolek, P. Patakova, K. Melzoch, K. Sigler, and T. Rezanka, "Changes in membrane plasmalogens of Clostridium pasteurianum during butanol fermentation as determined by lipidomic analysis," *PLoS One*, vol. 10, no. 3, 2015.

[9] M. H. Saier, V. S. Reddy, B. V. Tsu, M. S. Ahmed, C. Li, and G. Moreno-Hagelsieb, "The Transporter Classification Database (TCDB): Recent advances," *Nucleic Acids Res.*, vol. 44, no. D1, pp. D372–D379, 2016.

[10] G. L. Lorca, R. D. Barabote, V. Zlotopolski, C. Tran, B. Winnen, R. N. Hvorup, A. J. Stonestrom, E. Nguyen, L. W. Huang, D. S. Kim, and M. H. Saier, "Transport capabilities of eleven gram-positive bacteria: Comparative genomic analyses," *Biochimica et Biophysica Acta - Biomembranes*, vol. 1768, no. 6. pp. 1342–1366, 2007.

[11] I. M. Gohari, V. R. Parreira, V. J. Nowell, V. M. Nicholson, K. Oliphant, and J. F. Prescott, "A novel pore-forming toxin in type A Clostridium perfringens is associated with both fatal canine hemorrhagic gastroenteritis and fatal foal necrotizing enterocolitis," *PLoS One*, vol. 10, no. 4, 2015.

[12] R. P. Fagan and N. F. Fairweather, "Biogenesis and functions of bacterial S-layers," *Nat. Rev. Microbiol.*, vol. 12, pp. 211–22, 2014.

[13] R. P. Fagan and N. F. Fairweather, "Biogenesis and functions of bacterial S-layers," *Nat. Rev. Microbiol.*, vol. 12, pp. 211–22, 2014.

[14] H. Bruggemann, S. Baumer, W. F. Fricke, A. Wiezer, H. Liesegang, I. Decker, C. Herzberg, R. Martinez-Arias, R. Merkl, A. Henne, and G. Gottschalk, "The genome sequence of \textit{Clostridium tetani}, the causative agent of tetanus disease.," *Proc Natl Acad Sci U S A*, vol. 100, no. 3, pp. 1316–1321, 2003.

[15] M. Sebaihia, M. W. Peck, N. P. Minton, N. R. Thomson, M. T. G. Holden, W. J. Mitchell, A. T. Carter, S. D. Bentley, D. R. Mason, L. Crossman, C. J. Paul, A. Ivens, M. H. J. Wells-Bennik, I. J. Davis, A. M. Cerdeño-Tárraga, C. Churcher, M. A. Quail, T. Chillingworth, T. Feltwell, A. Fraser, I. Goodhead, Z. Hance, K. Jagels, N. Larke, M. Maddison, S. Moule, K. Mungall, H. Norbertczak, E. Rabbinowitsch, M. Sanders, M. Simmonds, B. White, S. Whithead, and J. Parkhill, "Genome sequence of a proteolytic (Group I) Clostridium botulinum strain Hall A and comparative analysis of the clostridial genomes," *Genome Res.*, vol. 17, no. 7, pp. 1082–1092, 2007.

[16] E. Calabi, S. Ward, B. Wren, T. Paxton, M. Panico, H. Morris, a Dell, G. Dougan, and N. Fairweather, "Molecular characterization of the surface layer proteins from Clostridium difficile.," *Mol. Microbiol.*, vol. 40, no. 5, pp. 1187–1199, 2001.

[17] C. M. Ausiello, M. Cerquetti, G. Fedele, F. Spensieri, R. Palazzo, M. Nasso, S. Frezza, and P. Mastrantonio, "Surface layer proteins from Clostridium difficile induce inflammatory and regulatory cytokines in human monocytes and dendritic cells," *Microbes Infect.*, vol. 8, no. 11, pp. 2640–2646, 2006.

[18] A. Ryan, M. Lynch, S. M. Smith, S. Amu, H. J. Nel, C. E. McCoy, J. K. Dowling, E. Draper, V. O'Reilly, C. McCarthy, J. O'Brien, D. Eidhin, M. J. O'Connell, B. Keogh, C. O. Morton, T. R. Rogers, P. G. Fallon, L. A. O'Neill, D. Kelleher, and C. E. Loscher, "A role for TLR4 in clostridium difficile infection and the recognition of surface layer proteins," *PLoS Pathog.*, vol. 7, no. 6, 2011.

[19] U. B. Sleytr, C. Huber, N. Ilk, D. Pum, B. Schuster, and E. M. Egelseer, "S-layers as a tool kit for nanobiotechnological applications," *FEMS Microbiol. Lett.*, vol. 267, no. 2, pp. 131–144, Feb. 2007.

[20] A. Dumitrache, D. M. Klingeman, J. Natzke, M. J. Rodriguez, R. J. Giannone, R. L. Hettich, B. H. Davison, and S. D. Brown, "Specialized activities and expression differences for Clostridium thermocellum biofilm and planktonic cells.," *Sci. Rep.*, vol. 7, p. 43583, Feb. 2017.

[21] T. R. Zuroff, W. Gu, R. L. Fore, S. B. Leschine, and W. R. Curtis, "Insights into Clostridium phytofermentans biofilm formation: aggregation, microcolony development and the role of extracellular DNA," *Microbiology*, vol. 160, no. Pt_6, pp. 1134–1143, 2014.

[22] A. C. Tolonen, W. Haas, A. C. Chilaka, J. Aach, S. P. Gygi, and G. M. Church, "Proteome-wide systems analysis of a cellulosic biofuel-producing microbe," *Mol. Syst. Biol.*, vol. 7, no. 1, pp. 461–461, 2011.

[23] A. Charlebois, M. Jacques, M. Boulianne, and M. Archambault, "Tolerance of Clostridium perfringens biofilms to disinfectants commonly used in the food industry.," *Food Microbiol.*, vol. 62, pp. 32–38, Apr. 2017.

[24] T. Dapa and M. Unnikrishnan, "Biofilm formation by Clostridium difficile.," *Gut microbes*, vol. 4, no. 5. pp. 397–402, 2013.

[25] E. B. Mearls, J. a Izquierdo, and L. R. Lynd, "Formation and characterization of non-growth states in Clostridium thermocellum: spores and L-forms," *BMC Microbiol.*, vol. 12, p. 180, 2012.

[26] P. Dürre, "Physiology and Sporulation in Clostridium," *Microbiol. Spectr.*, vol. 2, no. 4, pp. 1–14, 2014.

[27] W. L. Nicholson, N. Munakata, G. Horneck, H. J. Melosh, and P. Setlow, "Resistance of Bacillus endospores to extreme terrestrial and extraterrestrial environments.," *Microbiol. Mol. Biol. Rev.*, vol. 64, no. 3, pp. 548–72, 2000.

[28] E. B. Mearls, J. a Izquierdo, and L. R. Lynd, "Formation and characterization of non-growth states in Clostridium thermocellum: spores and L-forms," *BMC Microbiol.*, vol. 12, p. 180, 2012.

[29] R. Labbé, "Sporulation (Morphology) of Clostridia," in *Handbook on Clostridia*, CRC Press, 2005, pp. 647–658.

[30] C. J. Doyle, D. Gleeson, K. Jordan, T. P. Beresford, R. P. Ross, G. F. Fitzgerald, and P. D. Cotter, "Anaerobic sporeformers and their significance with respect to milk and dairy products," *International Journal of Food Microbiology*, vol. 197. pp. 77–87, 2015.

[31] P. Dürre and C. Hollergschwandner, "Initiation of endospore formation in Clostridium acetobutylicum," *Anaerobe*, vol. 10, no. 2. pp. 69–74, 2004.
[32] Y. Li, T. Xu, T. J. Tschaplinski, N. L. Engle, Y. Yang, D. E. Graham, Z. He, and J. Zhou, "Improvement of cellulose catabolism in Clostridium cellulolyticum by sporulation abolishment and carbon alleviation.," *Biotechnol. Biofuels*, vol. 7, no. 1, p. 25, 2014.
[33] R. Labbé, "Sporulation (Morphology) of Clostridia," in *Handbook on Clostridia*, CRC Press, 2005, pp. 647–658.
[34] L. Saujet, F. C. Pereira, M. Serrano, O. Soutourina, M. Monot, P. V Shelyakin, M. S. Gelfand, B. Dupuy, a O. Henriques, and I. Martin-Verstraete, "Genome-wide analysis of cell type-specific gene transcription during spore formation in Clostridium difficile," *PLoS Genet*, vol. 9, no. 10, p. e1003756, 2013.
[35] K. A. Fimlaid, J. P. Bond, K. C. Schutz, E. E. Putnam, J. M. Leung, T. D. Lawley, and A. Shen, "Global Analysis of the Sporulation Pathway of Clostridium difficile," *PLoS Genet.*, vol. 9, no. 8, p. e1003660, 2013.
[36] K. V. Alsaker and E. T. Papoutsakis, "Transcriptional program of early sporulation and stationary-phase events in Clostridium acetobutylicum," *J. Bacteriol.*, vol. 187, no. 20, pp. 7103–7118, 2005.
[37] J. Li, D. Paredes-Sabja, M. R. Sarker, and B. A. McClane, "Clostridium perfringens Sporulation and Sporulation-Associated Toxin Production.," *Microbiol. Spectr.*, vol. 4, no. 3, Jun. 2016.
[38] D. Wetzel and R. J. Fischer, "Small acid-soluble spore proteins of Clostridium acetobutylicum are able to protect DNA in vitro and are specifically cleaved by germination protease GPR and spore protease YyaC," *Microbiol. (United Kingdom)*, vol. 161, no. 11, pp. 2098–2109, 2015.
[39] F. C. Pereira, L. Saujet, A. R. Tomé, M. Serrano, M. Monot, E. Couture-Tosi, I. Martin-Verstraete, B. Dupuy, and A. O. Henriques, "The spore differentiation pathway in the enteric pathogen Clostridium difficile.," *PLoS Genet.*, vol. 9, no. 10, p. e1003782, 2013.
[40] P. Permpoonpattana, J. Phetcharaburanin, A. Mikelsone, M. Dembek, S. Tan, M. C. Brisson, R. La Ragione, A. R. Brisson, N. Fairweather, H. A. Hong, and S. M. Cutting, "Functional characterization of Clostridium difficile spore coat proteins," *J. Bacteriol.*, vol. 195, no. 7, pp. 1492–1503, 2013.
[41] A. M. Kubiak and N. P. Minton, "The potential of clostridial spores as therapeutic delivery vehicles in tumour therapy," *Research in Microbiology*, vol. 166, no. 4. pp. 244–254, 2015.
[42] A. M. Mowday, C. P. Guise, D. F. Ackerley, N. P. Minton, P. Lambin, L. J. Dubois, J. Theys, J. B. Smaill, and A. V. Patterson, "Advancing clostridia to clinical trial: Past lessons and recent progress," *Cancers*, vol. 8, no. 7. 2016.
[43] J. Theys and P. Lambin, "Clostridium to treat cancer: dream or reality?," *Ann. Transl. Med.*, vol. 3, no. Suppl 1, p. S21, 2015.
[44] J. Plowman and M. W. Peck, "Use of a novel method to characterize the response of spores of non-proteolytic Clostridium botulinum types B, E and F to a wide range of germinants and conditions.," *J. Appl. Microbiol.*, vol. 92, no. 4, pp. 681–694, 2002.
[45] D. Bhattacharjee, K. N. McAllister, and J. A. Sorg, "Germinants and their receptors in clostridia," *Journal of Bacteriology*, vol. 198, no. 20. pp. 2767–2775, 2016.
[46] D. Wetzel and R. J. Fischer, "Small acid-soluble spore proteins of Clostridium acetobutylicum are able to protect DNA in vitro and are specifically cleaved by germination protease GPR and spore protease YyaC," *Microbiol. (United Kingdom)*, vol. 161, no. 11, pp. 2098–2109, 2015.
[47] Y. Kevorkian, D. J. Shirley, and A. Shen, "Regulation of Clostridium difficile spore germination by the CspA pseudoprotease domain," *Biochimie*, vol. 122, pp. 243–254, 2016.
[48] V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho, and B. Henrissat, "The carbohydrate-active enzymes database (CAZy) in 2013," *Nucleic Acids Res.*, vol. 42, no. D1, pp. 490–495, 2014.
[49] G. Davies and B. Henrissat, "Structures and mechanisms of glycosyl hydrolases," *Structure*, vol. 3, no. 9, pp. 853–859, 1995.
[50] Q. Xu, M. G. Resch, K. Podkaminer, S. Yang, J. O. Baker, B. S. Donohoe, C. Wilson, D. M. Klingeman, D. G. Olson, S. R. Decker, R. J. Giannone, R. L. Hettich, S. D. Brown, L. R. Lynd, E. A. Bayer, M. E. Himmel, and Y. J. Bomble, "Dramatic performance of Clostridium thermocellum explained by its wide range of cellulase modalities.," *Sci. Adv.*, vol. 2, no. 2, p. e1501254, 2016.
[51] M. Boutard, T. Cerisy, P.-Y. Nogue, A. Alberti, J. Weissenbach, M. Salanoubat, and A. C. Tolonen, "Functional Diversity of Carbohydrate-Active Enzymes Enabling a Bacterium to Ferment Plant Biomass.," *PLoS Genet.*, vol. 10, no. 11, p. e1004773, 2014.
[52] Q. Xu, M. G. Resch, K. Podkaminer, S. Yang, J. O. Baker, B. S. Donohoe, C. Wilson, D. M. Klingeman, D. G. Olson, S. R. Decker, R. J. Giannone, R. L. Hettich, S. D. Brown, L. R. Lynd, E. A. Bayer, M. E. Himmel, and Y. J. Bomble, "Dramatic performance of Clostridium thermocellum explained by its wide range of cellulase modalities.," *Sci. Adv.*, vol. 2, no. 2, p. e1501254, 2016.
[53] M. D. Collins, P. A. Lawson, A. Willems, J. J. Cordoba, J. Fernandez-Garayzabal, P. Garcia, J. Cai, H. Hippe, and J. A. E. Farrow ', "The Phylogeny of the Genus Clostridium: Proposal of Five New Genera and Eleven New Species Combinations," no. May, pp. 812–826, 1994.
[54] N. Yutin and M. Y. Galperin, "A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced clostridia," *Environ. Microbiol.*, vol. 15, no. 10, pp. 2631–2641, 2013.
[55] J. Wiegel, R. Tanner, and F. a Rainey, "An Introduction to the Family Clostridiaceae," *Prokaryotes*, pp. 654–678, 2006.
[56] D. G. Cirne, O. D. Delgado, S. Marichamy, and B. Mattiasson, "Clostridium lundense sp. nov., a novel anaerobic lipolytic bacterium isolated from bovine rumen," *Int. J. Syst. Evol. Microbiol.*, vol. 56, no. 3, pp. 625–628, 2006.
[57] L. Aristilde, I. A. Lewis, J. O. Park, and J. D. Rabinowitz, "Hierarchy in pentose sugar metabolism in Clostridium acetobutylicum," *Appl. Environ. Microbiol.*, vol. 81, no. 4, pp. 1452–1462, 2015.
[58] L. G. Ljungdahl, "The autotrophic pathway of acetate synthesis in acetogenic bacteria," *Annu. Rev. Microbiol.*, no. 40, pp. 415–450, 1986.
[59] K. Schuchmann and V. Müller, "Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria," *Nat. Rev. Microbiol.*, vol. 12, no. 12, pp. 809–821, 2014.
[60] H. L. Drake, A. S. G????ner, and S. L. Daniel, "Old acetogens, new light," in *Annals of the New York Academy of Sciences*, 2008, vol. 1125, pp. 100–128.
[61] H. G. Wood, "Life with CO or CO2 and H2 as a source of carbon and energy.," *FASEB J.*, vol. 5, no. 2, pp. 156–163, 1991.
[62] W. Martin, F. Sousa, and N. Lane, "Energy at life's origin," *Science (80-.).*, vol. 344, no. 6188, pp. 1092–1093, 2014.
[63] H. A. Barker, "The production of caproic and butyric acids by the methane fermentation of ethyl alcohol," *Arch. f{ü}r Mikrobiol.*, vol. 8, no. 1, pp. 415–421, 1937.
[64] H. Seedorf, W. F. Fricke, B. Veith, H. Brüggemann, H. Liesegang, A. Strittmatter, M. Miethke, W. Buckel, J. Hinderberger, F. Li, C. Hagemeier, R. K. Thauer, and G. Gottschalk, "The genome of Clostridium kluyveri, a strict anaerobe with unique metabolic features.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, no. 6, pp. 2128–2133, 2008.

[65] H. M. Hamer, D. Jonkers, K. Venema, S. Vanhoutvin, F. J. Troost, and R. J. Brummer, "Review article: The role of butyrate on colonic function," *Alimentary Pharmacology and Therapeutics*, vol. 27, no. 2. pp. 104–119, 2008.
[66] E. Hosseini, C. Grootaert, W. Verstraete, and T. Van de Wiele, "Propionate as a health-promoting microbial metabolite in the human gut.," *Nutr. Rev.*, vol. 69, no. 5, pp. 245–258, May 2011.
[67] N. Reichardt, S. H. Duncan, P. Young, A. Belenguer, C. McWilliam Leitch, K. P. Scott, H. J. Flint, and P. Louis, "Phylogenetic distribution of three pathways for propionate production within the human gut microbiota," *ISME J*, vol. 8, no. 6, pp. 1323–1335, 2014.
[68] A. Poehlein, K. Schlien, N. P. Chowdhury, G. Gottschalk, W. Buckel, and R. Daniel, "Complete Genome Sequence of the Amino Acid-Fermenting *Clostridium propionicum* X2 (DSM 1682)," *Genome Announc.*, vol. 4, no. 2, pp. e00294-16, 2016.
[69] M. D. Kane, A. Brauman, and J. A. Breznak, "Clostridium mayombei sp. nov., an H2/CO2 acetogenic bacterium from the gut of the African soil-feeding termite, Cubitermes speciosus," *Arch. Microbiol.*, vol. 156, no. 2, pp. 99–104, 1991.
[70] B. P. Cardon and H. A. Barker, "Two New Amino-Acid-Fermenting Bacteria, Clostridium propionicum and Diplococcus glycinophilus," *J. Bacteriol.*, vol. 52, no. 6, pp. 629–634, 1946.
[71] N. Reichardt, S. H. Duncan, P. Young, A. Belenguer, C. McWilliam Leitch, K. P. Scott, H. J. Flint, and P. Louis, "Phylogenetic distribution of three pathways for propionate production within the human gut microbiota," *ISME J*, vol. 8, no. 6, pp. 1323–1335, 2014.
[72] M. Nevas, M. Lindström, S. Hielm, K. Johanna Björkroth, M. W. Peck, and H. Korkeala, "Diversity of proteolytic Clostridium botulinum strains, determined by a pulsed-field gel electrophoresis approach," *Appl. Environ. Microbiol.*, vol. 71, no. 3, pp. 1311–1317, 2005.
[73] S. V Seddon and S. P. Borriello, "Proteolytic activity of Clostridium difficile," *J Med Microbiol*, vol. 36, no. 5, pp. 307–311, 1992.
[74] G. T. Attwood, K. Reilly, and B. K. Patel, "Clostridium proteoclasticum sp. nov., a novel proteolytic bacterium from the bovine rumen.," *Int. J. Syst. Bacteriol.*, vol. 46, no. 3, pp. 753–8, 1996.
[75] T. Mechichi, M. L. Fardeau, M. Labat, J. L. Garcia, F. Verhé, and B. K. C. Patel, "Clostridium peptidivorans sp. nov., a peptide-fermenting bacterium from an olive mill wastewater treatment digester," *Int. J. Syst. Evol. Microbiol.*, vol. 50, no. 3, pp. 1259–1264, 2000.
[76] E. A. Smith and G. T. MacFarlane, "Enumeration of amino acid fermenting bacteria in the human large intestine: Effects of pH and starch on peptide metabolism and dissimilation of amino acids," *FEMS Microbiol. Ecol.*, vol. 25, no. 4, pp. 355–368, 1998.
[77] M. D. Flythe and J. B. Russell, "Fermentation acids inhibit amino acid deamination by Clostridium sporogenes MD1 via mechanism involving a decline in intracellular glutamate rather than protonmotive force," *Microbiology*, vol. 152, no. 9, pp. 2619–2624, 2006.
[78] A. MEISTER, H. A. SOBER, and S. V TICE, "Enzymatic decarboxylation of aspartic acid to alpha-alanine.," *J. Biol. Chem.*, vol. 189, no. 2, pp. 577–590, Apr. 1951.
[79] S. R. Elsden, M. G. Hilton, and J. M. Waller, "The end products of the metabolism of aromatic amino acids by clostridia," *Arch. Microbiol.*, vol. 107, no. 3, pp. 283–288, 1976.
[80] N. Fonknechten, S. Chaussonnerie, S. Tricot, A. Lajus, J. R. Andreesen, N. Perchat, E. Pelletier, M. Gouyvenoux, V. Barbe, M. Salanoubat, D. Le Paslier, J. Weissenbach, G. N. Cohen, and A. Kreimeyer, "Clostridium sticklandii, a specialist in amino acid degradation:revisiting its metabolism through its genome sequence.," *BMC Genomics*, vol. 11, p. 555, 2010.
[81] H. de Vladar, "Amino acid fermentation at the origin of the genetic code," *Biol. Direct*, vol. 7, no. 1, p. 6, 2012.
[82] L. Bouillaut, W. T. Self, and A. L. Sonenshein, "Proline-dependent regulation of Clostridium difficile stickland metabolism," *J. Bacteriol.*, vol. 195, no. 4, pp. 844–854, 2013.
[83] C. Janoir, C. Denève, S. Bouttier, F. Barbut, S. Hoys, L. Caleechum, D. Chapetón-Montes, F. C. Pereira, A. O. Henriques, A. Collignon, M. Monot, and B. Dupuy, "Adaptive strategies and pathogenesis of *Clostridium difficile* from in vivo transcriptomics.," *Infect. Immun.*, vol. 81, no. 10, pp. 3757–69, 2013.
[84] P. Durre and J. R. Andreesen, "Purine and glycine metabolism by purinolytic clostridia," *J. Bacteriol.*, vol. 154, no. 1, pp. 192–199, 1983.
[85] K. Hartwich, A. Poehlein, and R. Daniel, "The Purine-Utilizing Bacterium Clostridium acidurici 9a: A Genome-Guided Metabolic Reconsideration," *PLoS One*, vol. 7, no. 12, 2012.
[86] W. H. Taylor, M. L. Taylor, W. E. Balch, and P. S. Gilchrist, "Purification and Properties of Dihydroorotase, a Zinc-Containing Metalloenzyme in Clostridium-Oroticum," *J. Bacteriol.*, vol. 127, no. 2, p. 863–873 ST–PURIFICATION AND PROPERTIES OF DIHYD, 1976.
[87] R. Schafer and A. C. Schwartz, "Degradation of pyrimidine bases in Clostridium sticklandii.," *Arch. Microbiol.*, vol. 124, no. 1, pp. 111–114, Jan. 1980.
[88] N. Fonknechten, S. Chaussonnerie, S. Tricot, A. Lajus, J. R. Andreesen, N. Perchat, E. Pelletier, M. Gouyvenoux, V. Barbe, M. Salanoubat, D. Le Paslier, J. Weissenbach, G. N. Cohen, and A. Kreimeyer, "Clostridium sticklandii, a specialist in amino acid degradation:revisiting its metabolism through its genome sequence.," *BMC Genomics*, vol. 11, p. 555, 2010.
[89] E. Petit, M. V Coppi, J. C. Hayes, A. C. Tolonen, T. Warnick, W. G. Latouf, D. Amisano, A. Biddle, S. Mukherjee, N. Ivanova, A. Lykidis, M. Land, L. Hauser, N. Kyrpides, B. Henrissat, J. Lau, D. J. Schnell, G. M. Church, S. B. Leschine, and J. L. Blanchard, "Genome and Transcriptome of Clostridium phytofermentans, Catalyst for the Direct Conversion of Plant Feedstocks to Fuels.," *PLoS One*, vol. 10, no. 6, p. e0118285, 2015.
[90] M. Köpke, C. Held, S. Hujer, H. Liesegang, A. Wiezer, A. Wollherr, A. Ehrenreich, W. Liebl, G. Gottschalk, and P. Dürre, "Clostridium ljungdahlii represents a microbial production platform based on syngas.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, no. 29, pp. 13087–92, 2010.
[91] J. L. Johnson and B. S. Francis, "Taxonomy of the Clostridia: Ribosomal Ribonucleic Acid Homologies among the Species," *J. Gen. Microbiol.*, vol. 88, no. 2, pp. 229–244, 1975.
[92] E. Stackebrandt, I. Kramer, J. Swiderski, and H. Hippe, "Phylogenetic basis for a taxonomic dissection of the genus Clostridium," *FEMS Immunol. Med. Microbiol.*, vol. 24, no. 3, pp. 253–258, 1999.
[93] S. M. Finegold, Y. Song, and C. Liu, "Taxonomy--General comments and update on taxonomy of Clostridia and Anaerobic cocci.," *Anaerobe*, vol. 8, no. 5, pp. 283–5, 2002.
[94] R. S. Gupta and B. Gao, "Phylogenomic analyses of clostridia and identification of novel protein signatures that are specific to the genus Clostridium sensu stricto (cluster I)," *Int. J. Syst. Evol. Microbiol.*, vol. 59, no. 2, pp. 285–294, 2009.
[95] W. Ludwig, K.-H. Schleifer, and W. Whitman, "Revised road map to the phylumFirmicutes," *Bergey's Manual® Syst. Bacteriol. SE - 1*, pp. 1–13, 2009.

[96] M. Nakajima, N. Tanaka, N. Furukawa, T. Nihira, Y. Kodutsumi, Y. Takahashi, N. Sugimoto, A. Miyanaga, S. Fushinobu, H. Taguchi, and H. Nakai, "Mechanistic insight into the substrate specificity of 1,2-β-oligoglucan phosphorylase from Lachnoclostridium phytofermentans," *Sci. Rep.*, vol. 7, p. 42671, Feb. 2017.
[97] M. Nakajima, N. Tanaka, N. Furukawa, T. Nihira, Y. Kodutsumi, Y. Takahashi, N. Sugimoto, A. Miyanaga, S. Fushinobu, H. Taguchi, and H. Nakai, "Mechanistic insight into the substrate specificity of 1,2-β-oligoglucan phosphorylase from Lachnoclostridium phytofermentans," *Sci. Rep.*, vol. 7, p. 42671, Feb. 2017.
[98] T. Cerisy, T. Souterre, I. Torres-Romero, M. Boutard, I. Dubois, J. Patrouix, K. Labadie, W. Berrabah, M. Salanoubat, V. Doring, and A. Tolonen, "Evolution of a biomass-fermenting bacterium to resist lignin phenolics.," *Appl. Environ. Microbiol.*, Mar. 2017.
[99] M. A. Mahowald, F. E. Rey, H. Seedorf, P. J. Turnbaugh, R. S. Fulton, A. Wollam, N. Shah, C. Wang, V. Magrini, R. K. Wilson, B. L. Cantarel, P. M. Coutinho, B. Henrissat, L. W. Crock, A. Russell, N. C. Verberkmoes, R. L. Hettich, and J. I. Gordon, "Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla," *Proc Natl Acad Sci U S A*, vol. 106, no. 14, pp. 5859–5864, 2009.
[100] T. Garnier and S. T. Cole, "Complete nucleotide sequence and genetic organization of the bacteriocinogenic plasmid, pIP404, from Clostridium perfringens," *Plasmid*, vol. 19, no. 2, pp. 134–150, 1988.
[101] C. Leang, T. Ueki, K. P. Nevin, and D. R. Lovley, "A Genetic system for Clostridium ljungdahlii: A chassis for autotrophic production of biocommodities and a model homoacetogen," *Appl. Environ. Microbiol.*, vol. 79, no. 4, pp. 1102–1109, 2013.
[102] K. C. B. Jennert, C. Tardif, D. I. Young, and M. Young, "Gene transfer to Clostridium cellulolyticum ATCC 35319," *Microbiology*, vol. 146, no. 12, pp. 3071–3080, 2000.
[103] D. Purdy, T. A. T. O'Keeffe, M. Elmore, M. Herbert, A. McLeod, M. Bokori-Brown, A. Ostrowski, and N. P. Minton, "Conjugative transfer of clostridial shuttle vectors from Escherichia coli to Clostridium difficile through circumvention of the restriction barrier," *Mol. Microbiol.*, vol. 46, no. 2, pp. 439–452, 2002.
[104] M. E. Pyne, M. Moo-Young, D. a Chung, and C. P. Chou, "Development of an electrotransformation protocol for genetic manipulation of Clostridium pasteurianum.," *Biotechnol. Biofuels*, vol. 6, no. 1, p. 50, 2013.
[105] R. M., W. M.H., and G. D.N., "Clostridium difficile infection: New developments in epidemiology and pathogenesis," *Nat. Rev. Microbiol.*, vol. 7, no. 7, pp. 526–536, 2009.
[106] I. Hall and E. O'Toole, "Intestinal flora in new-born infants: With a description of a new pathogenic anaerobe, Bacillus difficilis," *Am. J. Dis. Child.*, vol. 49, no. 2, pp. 390–402, 1935.
[107] R. M., W. M.H., and G. D.N., "Clostridium difficile infection: New developments in epidemiology and pathogenesis," *Nat. Rev. Microbiol.*, vol. 7, no. 7, pp. 526–536, 2009.
[108] I. Martin-Verstraete, J. Peltier, and B. Dupuy, "The regulatory networks that control Clostridium difficile toxin synthesis," *Toxins*, vol. 8, no. 5. 2016.
[109] A. Charlebois, M. Jacques, M. Boulianne, and M. Archambault, "Tolerance of Clostridium perfringens biofilms to disinfectants commonly used in the food industry.," *Food Microbiol.*, vol. 62, pp. 32–38, Apr. 2017.
[110] J. Li and B. A. McClane, "Evaluating the Involvement of Alternative Sigma Factors SigF and SigG in Clostridium perfringens Sporulation and Enterotoxin Synthesis," *Infect. Immun.*, vol. 78, no. 10, pp. 4286–4293, 2010.
[111] M. Lindstromm, K. Kiviniemi, and H. Korkeala, "Hazard and control of group II (non-proteolytic) Clostridium botulinum in modern food processing," *Int. J. Food Microbiol.*, vol. 108, no. 1, pp. 92–104, 2006.
[112] E. S. Felber, "Botulinum toxin in primary care medicine.," *J. Am. Osteopath. Assoc.*, vol. 106, pp. 609–614, 2006.
[113] R. Persaud, G. Garas, S. Silva, C. Stamatoglou, P. Chatrath, and K. Patel, "An evidence-based review of botulinum toxin (Botox) applications in non-cosmetic head and neck conditions.," *JRSM Short Rep.*, vol. 4, no. 2, p. 10, 2013.
[114] H. Bruggemann, S. Baumer, W. F. Fricke, A. Wiezer, H. Liesegang, I. Decker, C. Herzberg, R. Martinez-Arias, R. Merkl, A. Henne, and G. Gottschalk, "The genome sequence of \textit{Clostridium tetani}, the causative agent of tetanus disease.," *Proc Natl Acad Sci U S A*, vol. 100, no. 3, pp. 1316–1321, 2003.
[115] L. R. Lopetuso, F. Scaldaferri, V. Petito, and A. Gasbarrini, "Commensal Clostridia: leading players in the maintenance of gut homeostasis," *Gut Pathog.*, vol. 5, no. 1, p. 1, 2013.
[116] D. N. Frank, A. L. St Amand, R. A. Feldman, E. C. Boedeker, N. Harpaz, and N. R. Pace, "Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 34, pp. 13780–5, 2007.
[117] M. Natarajan, S. T. Walk, V. B. Young, and D. M. Aronoff, "A clinical and epidemiological review of non-toxigenic clostridium difficile," *Anaerobe*, vol. 22. pp. 1–5, 2013.
[118] B. Dixon, "Chaim Weizmann: From Fermentation Chemist to President of Israel," *Anaerobe*, vol. 2, pp. 195–196, 1996.
[119] J. M. Landete, ??ngela Peirot??n, E. Rodr??guez, A. Margolles, M. Medina, and J. L. Arqu??s, "Anaerobic green fluorescent protein as a marker of Bifidobacterium strains," *Int. J. Food Microbiol.*, vol. 175, pp. 6–13, 2014.
[120] J. Chen and S. F. Hiu, "Acetone-Butanol-Isopropanol Production By Clostridium Beijerinckii (Synonym, Clostridium Butylicum)," *Biotechnol. Lett.*, vol. 8, no. 5, pp. 371–376, 1986.
[121] J. Lee, Y. S. Jang, S. J. Choi, J. A. Im, H. Song, J. H. Cho, D. Y. Seung, E. Terry Papoutsakis, G. N. Bennett, and S. Y. Lee, "Metabolic engineering of clostridium acetobutylicum ATCC 824 for isopropanol-butanol-ethanol fermentation," *Appl. Environ. Microbiol.*, vol. 78, no. 5, pp. 1416–1423, 2012.
[122] P. Dürre, "Fermentative butanol production: Bulk chemical and biofuel," *Ann. N. Y. Acad. Sci.*, vol. 1125, pp. 353–362, 2008.
[123] S. Dusséaux, C. Croux, P. Soucaille, and I. Meynial-Salles, "Metabolic engineering of Clostridium acetobutylicum ATCC 824 for the high-yield production of a biofuel composed of an isopropanol/butanol/ethanol mixture," *Metab. Eng.*, vol. 18, pp. 1–8, 2013.
[124] F. Monot, J.-M. Engasser, and H. Petitdemange, "Influence of pH and undissociated butyric acid on the production of acetone and butanol in batch cultures of Clostridium acetobutylicum," *Appl. Microbiol. Biotechnol.*, vol. 19, no. 6, pp. 422–426, 1984.
[125] M. Devarapalli and H. K. Atiyeh, "A review of conversion processes for bioethanol production with a focus on syngas fermentation," *Biofuel Res. J.*, vol. 2, no. 3, pp. 268–280, 2015.
[126] H. Younesi, G. Najafpour, and A. R. Mohamed, "Ethanol and acetate production from synthesis gas via fermentation processes using anaerobic bacterium, Clostridium ljungdahlii," *Biochem. Eng. J.*, vol. 27, no. 2, pp. 110–119, Dec. 2005.
[127] J. Abrini, H. Naveau, and E.-J. Nyns, "Clostridium autoethanogenumt, sp. nov., an anaerobic bacterium that produces ethanol from carbon monoxide," *Arch. Microbiol.*, vol. 161, no. 4, pp. 345–351, 1994.
[128] Y. Shen, R. Brown, and Z. Wen, "Enhancing mass transfer and ethanol production in syngas fermentation of Clostridium carboxidivorans P7 through a monolithic biofilm reactor," *Appl. Energy*, vol. 136, pp. 68–76, 2014.
[129] F. R. Bengelsdorf, A. Poehlein, S. Linder, C. Erz, T. Hummel, S. Hoffmeister, R. Daniel, and P. D??rre, "Industrial acetogenic

biocatalysts: A comparative metabolic and genomic analysis," *Front. Microbiol.*, vol. 7, no. JUL, pp. 1–15, 2016.
[130] J. M. Bruno-Barcena, M. S. Chinn, and A. M. Grunden, "Genome Sequence of the Autotrophic Acetogen Clostridium autoethanogenum JA1-1 Strain DSM 10061, a Producer of Ethanol from Carbon Monoxide.," *Genome Announc.*, vol. 1, no. 4, pp. e00628-13, 2013.
[131] F. L. Sousa, T. Thiergart, G. Landan, S. Nelson-Sathi, I. a C. Pereira, J. F. Allen, N. Lane, and W. F. Martin, "Early bioenergetic evolution.," *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 368, no. 1622, p. 20130088, 2013.
[132] W. Martin and M. J. Russell, "On the origin of biochemistry at an alkaline hydrothermal vent.," *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 362, no. 1486, pp. 1887–925, 2007.
[133] S. W. Ragsdale and E. Pierce, "Acetogenesis and the Wood-Ljungdahl pathway of CO2 fixation," *Biochimica et Biophysica Acta - Proteins and Proteomics*, vol. 1784, no. 12. pp. 1873–1898, 2008.
[134] W. Nitschke and M. J. Russell, "Beating the acetyl coenzyme A-pathway to the origin of life.," *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 368, no. 1622, p. 20120258, 2013.
[135] B. S. Montenecourt and D. E. Eveleigh, "Preparation of mutants of Trichoderma reesei with enhanced cellulase production.," *Appl. Environ. Microbiol.*, vol. 34, no. 6, pp. 777–782, Dec. 1977.
[136] G. S. Coleman, "A comparison of the cellulolytic activities of rumen ciliate protozoa grown under different conditions," *Appl. Biochem. Biotechnol.*, vol. 9, no. 4, pp. 345–346, 1984.
[137] R. López-Mondéjar, D. Zühlke, D. Becher, K. Riedel, and P. Baldrian, "Cellulose and hemicellulose decomposition by forest soil bacteria proceeds by the action of structurally variable enzymatic systems.," *Sci. Rep.*, vol. 6, no. April, p. 25279, 2016.
[138] Z. Bashir, V. K. Kondapalli, N. Adlakha, A. Sharma, R. K. Bhatnagar, G. Chandel, and S. S. Yazdani, "Diversity and functional significance of cellulolytic microbes living in termite, pill-bug and stem-borer guts.," *Sci. Rep.*, vol. 3, p. 2558, 2013.
[139] M. Hess, A. Sczyrba, R. Egan, T.-W. Kim, H. Chokhawala, G. Schroth, S. Luo, D. S. Clark, F. Chen, T. Zhang, R. I. Mackie, L. A. Pennacchio, S. G. Tringe, A. Visel, T. Woyke, Z. Wang, and E. M. Rubin, "Metagenomic Discovery of Biomass-Degrading Genes and Genomes from Cow Rumen," *Science (80-.).*, vol. 331, no. 6016, pp. 463–467, 2011.
[140] A. M. López-Contreras, K. Gabor, A. A. Martens, B. A. M. Renckens, P. A. M. Claassen, J. Van Der Oost, and W. M. De Vos, "Substrate-induced production and secretion of cellulases by Clostridium acetobutylicum," *Appl. Environ. Microbiol.*, vol. 70, no. 9, pp. 5238–5243, 2004.
[141] K. Kovács, B. J. Willson, K. Schwarz, J. T. Heap, A. Jackson, D. N. Bolam, K. Winzer, and N. P. Minton, "Secretion and assembly of functional mini-cellulosomes from synthetic chromosomal operons in Clostridium acetobutylicum ATCC 824.," *Biotechnol. Biofuels*, vol. 6, p. 117, 2013.
[142] J. a Viljoen, E. B. Fred, and W. H. Peterson, "The fermentation of cellulose by thermophilic bacteria," *J. Agric. Sci.*, vol. 16, no. 1, pp. 1–17, 1926.
[143] R. H. McBEE, "The characteristics of Clostridium thermocellum.," *J. Bacteriol.*, vol. 67, no. 4, pp. 505–506, Apr. 1954.
[144] E. Petitdemange, F. Caillet, J. Giallo, and C. Gaudin, "Clostridium cellulolyticum sp. nov., a Cellulolytic, Mesophilic: Species from Decayed Grass," *Int. J. Syst. Bacteriol.*, vol. 34, no. 2, pp. 155–159, 1984.
[145] E. A. Bayer, E. Morag, and R. Lamed, "The cellulosome - A treasure-trove for biotechnology," *Trends in Biotechnology*, vol. 12, no. 9. pp. 379–386, 1994.
[146] L. Artzi, E. A. Bayer, and S. Morais, "Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides," *Nat Rev Micro*, vol. 15, no. 2, pp. 83–95, Feb. 2017.
[147] P. P. Lin, L. Mi, A. H. Morioka, K. M. Yoshino, S. Konishi, S. C. Xu, B. A. Papanek, L. A. Riley, A. M. Guss, and J. C. Liao, "Consolidated bioprocessing of cellulose to isobutanol using Clostridium thermocellum," *Metab. Eng.*, vol. 31, pp. 44–52, 2015.
[148] L. Artzi, E. A. Bayer, and S. Morais, "Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides," *Nat Rev Micro*, vol. 15, no. 2, pp. 83–95, Feb. 2017.
[149] N. Franche, C. Tardif, J. Ravachol, S. Harchouni, P.-H. Ferdinand, R. Borne, H.-P. Fierobe, and S. Perret, "Cel5I, a SLH-Containing Glycoside Hydrolase: Characterization and Investigation on Its Role in Ruminiclostridium cellulolyticum," *PLoS One*, vol. 11, no. 8, p. e0160812, Aug. 2016.
[150] H. Akinosho, K. Yee, D. Close, and A. Ragauskas, "The emergence of Clostridium thermocellum as a high utility candidate for consolidated bioprocessing applications," *Front. Chem.*, vol. 2, no. August, pp. 1–18, 2014.
[151] H. P. Fierobe, S. Pagès, A. Bélaïch, S. Champ, D. Lexa, and J. P. Bélaïch, "Cellulosome from Clostridium cellulolyticum: Molecular study of the dockerin/cohesin interaction," *Biochemistry*, vol. 38, no. 39, pp. 12822–12832, 1999.
[152] S. Pagès, A. Bélaïch, J. P. Bélaïch, E. Morag, R. Lamed, Y. Shoham, and E. A. Bayer, "Species-specificity of the cohesin-dockerin interaction between Clostridium thermocellum and Clostridium cellulolyticum: Prediction of specificity determinants of the dockerin domain," *Proteins Struct. Funct. Genet.*, vol. 29, no. 4, pp. 517–527, 1997.
[153] S. Moraïs, E. Morag, Y. Barak, D. Goldman, Y. Hadar, R. Lamed, Y. Shoham, D. B. Wilson, and E. A. Bayera, "Deconstruction of lignocellulose into soluble sugars by native and designer cellulosomes," *MBio*, vol. 3, no. 6, 2012.
[154] J. Stern, S. Moraïs, R. Lamed, and E. A. Bayer, "Adaptor scaffoldins: An original strategy for extended designer cellulosomes, inspired from nature," *MBio*, vol. 7, no. 2, 2016.
[155] T. A. Warnick, B. A. Methé, and S. B. Leschine, "Clostridium phytofermentans sp. nov., a cellulolytic mesophile from forest soil," *Int. J. Syst. Evol. Microbiol.*, vol. 52, no. 4, pp. 1155–1160, 2002.
[156] R. Sleat, R. A. Mah, and R. Robinson, "Isolation and characterization of an anaerobic, cellulolytic bacterium, Clostridium cellulovorans sp. nov.," *Applied and Environmental Microbiology*, vol. 48, no. 1. pp. 88–93, 1984.
[157] L. Artzi, B. Dassa, I. Borovok, M. Shamshoum, R. Lamed, and E. a Bayer, "Cellulosomics of the cellulolytic thermophile Clostridium clariflavum," *Biotechnol. Biofuels*, vol. 7, no. 1, p. 100, 2014.
[158] R. H. Madden, M. J. Bryder, and N. J. Poole, "Isolation and cCharacterization of an anaerobic, cellulolytic bacterium, Clostridium papyrosolvens sp. nov," *Int. J. Syst. Evol. Microbiol.*, vol. 32, no. 1, pp. 87–91, 1982.
[159] H. YANLING, D. YOUFANG, and L. YANQUAN, "Two Cellulolytic Clostridium Species: Clostridium cellulosi sp. nov. and Clostridium cellulofermentans sp. nov.," *Int. J. Syst. Bacteriol.*, vol. 41, no. 2, pp. 306–309, 1991.
[160] B. Dixon, "Chaim Weizmann: From Fermentation Chemist to President of Israel," *Anaerobe*, vol. 2, pp. 195–196, 1996.
[161] A. L. Demain, M. Newcomb, and J. H. Wu, "Cellulase, clostridia, and ethanol," *Microbiol Mol Biol Rev*, vol. 69, no. 1, pp. 124–154, 2005.
[162] H. A. George, J. L. Johnson, and W. E. C. Moore, "Acetone, isopropanol, and butanol production by Clostridium beijerinckii (syn. Clostridium butylicum) and Clostridium aurantibutyricum," *Applied and Environmental Microbiology*, vol. 45, no. 3. pp. 1160–1163, 1983.
[163] H. M. de Gérando, F. Fayolle-Guichard, L. Rudant, S. K. Millah, F. Monot, N. L. Ferreira, and A. M. López-Contreras, "Improving

isopropanol tolerance and production of Clostridium beijerinckii DSM 6423 by random mutagenesis and genome shuffling," *Appl. Microbiol. Biotechnol.*, vol. 100, no. 12, pp. 5427–5436, 2016.
[164] J. Lee, Y. S. Jang, S. J. Choi, J. A. Im, H. Song, J. H. Cho, D. Y. Seung, E. Terry Papoutsakis, G. N. Bennett, and S. Y. Lee, "Metabolic engineering of clostridium acetobutylicum ATCC 824 for isopropanol-butanol-ethanol fermentation," *Appl. Environ. Microbiol.*, vol. 78, no. 5, pp. 1416–1423, 2012.
[165] J. Zigová, E. Šturdík, D. Vandák, and Š. Schlosser, "Butyric acid production by Clostridium butyricum with integrated extraction and pertraction," *Process Biochem.*, vol. 34, no. 8, pp. 835–843, Oct. 1999.
[166] Z. Wu and S. T. Yang, "Extractive fermentation for butyric acid production from glucose by Clostridium tyrobutyricum," *Biotechnol. Bioeng.*, vol. 82, no. 1, pp. 93–102, 2003.
[167] A. Brioschi, G. P. Zara, S. Calderoni, M. R. Gasco, and A. Mauro, "Cholesterylbutyrate solid lipid nanoparticles as a butyric acid prodrug," *Molecules*, vol. 13, no. 2. pp. 230–254, 2008.
[168] T. Gaschott, D. Steinhilber, V. Milovic, and J. Stein, "Tributyrin, a stable and rapidly absorbed prodrug of butyric acid, enhances antiproliferative effects of dihydroxycholecalciferol in human colon cancer cells.," *J. Nutr.*, vol. 131, no. 6, pp. 1839–1843, 2001.
[169] C. H. Chou, C. L. Han, J. J. Chang, and J. J. Lay, "Co-culture of Clostridium beijerinckii L9, Clostridium butyricum M1 and Bacillus thermoamylovorans B5 for converting yeast waste into hydrogen," *Int. J. Hydrogen Energy*, vol. 36, no. 21, pp. 13972–13983, 2011.
[170] M. L. Chong, V. Sabaratnam, Y. Shirai, and M. A. Hassan, "Biohydrogen production from biomass and industrial wastes by dark fermentation," *Int. J. Hydrogen Energy*, vol. 34, no. 8, pp. 3277–3287, 2009.
[171] S. Meher Kotay and D. Das, "Biohydrogen as a renewable energy resource-Prospects and potentials," *Int. J. Hydrogen Energy*, vol. 33, no. 1, pp. 258–263, 2008.
[172] D. M. Gill, "Bacterial toxins: a table of lethal amounts.," *Microbiol. Rev.*, vol. 46, no. 1, pp. 86–94, 1982.
[173] J. Frevert, "Pharmaceutical, biological, and clinical properties of botulinum neurotoxin type A products," *Drugs R. D.*, vol. 15, no. 1, pp. 1–9, 2015.
[174] S. Donovan, "Botulinum toxin production method," US7189541 B2, 13-Mar-2007.
[175] F. T. D. Kaplan, "Collagenase clostridium histolyticum injection for the treatment of Dupuytren's contracture," *Drugs of Today*, vol. 47, no. 9. pp. 653–667, 2011.
[176] M. Gelbard, L. I. Lipshultz, J. Tursi, T. Smith, G. Kaufman, and L. A. Levine, "Phase 2b study of the clinical efficacy and safety of collagenase clostridium histolyticum in patients with peyronie disease," *J. Urol.*, vol. 187, no. 6, pp. 2268–2274, 2012.
[177] K. N. Riley and I. M. Herman, "Collagenase promotes the cellular responses to injury and wound healing in vivo.," *J. Burns Wounds*, vol. 4, p. e8, 2005.
[178] D. Zhang, Y. Zhang, Z. Wang, X. Zhang, and M. Sheng, "Target radiofrequency combined with collagenase chemonucleolysis in the treatment of lumbar intervertebral disc herniation," *Int. J. Clin. Exp. Med.*, vol. 8, no. 1, pp. 526–532, 2015.
[179] H. Alipour, A. Raz, S. Zakeri, and N. Dinparast Djadid, "Therapeutic applications of collagenase (metalloproteases): A review," *Asian Pac. J. Trop. Biomed.*, vol. 6, no. 11, pp. 975–981, 2016.
[180] H. Lodish, A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore, and J. Darnell, *Molecular Cell Biology. 4th edition*. 2000.
[181] H. C. Lin and S. P. Lei, "Recombinant 68,000 dalton collagenase of Clostridium histolyticum," US5418157, 1995.
[182] H. Markus, K; Rentsch, D; Wolfang, B; Lutke-Everloh, T; Bahl, "Metabolic Engineering of Clostridium acetobutylicum ATCC 824 for the production of 1,4-Butanediol," in *14th International Conference on the Genetics, Physiology and Synthetic Biology of Solvent- and Acid-forming Clostridia*, 2016, p. 37.
[183] J. Nielsen, M. Fussenegger, J. Keasling, S. Y. Lee, J. C. Liao, K. Prather, and B. Palsson, "Engineering synergy in biotechnology.," *Nat. Chem. Biol.*, vol. 10, no. 5, pp. 319–22, 2014.
[184] S. (LanzaTech) Simpson, "Circular economy through C1 fermentation," 2014. [Online]. Available: https://www.bio.org/sites/default/files/WorldCongress/Sean Simpson.pdf.
[185] S. (LanzaTech) Simpson, "Circular economy through C1 fermentation," 2014. .
[186] X. Liu, V. Fabos, S. Taylor, D. W. Knight, K. Whiston, and G. J. Hutchings, "One-Step Production of 1,3-Butadiene from 2,3-Butanediol Dehydration," *Chem. - A Eur. J.*, vol. 22, no. 35, pp. 12290–12294, 2016.
[187] N. L. Morrow, "The industrial production and use of 1,3-butadiene," *Environ. Health Perspect.*, vol. 86, pp. 7–8, 1990.
[188] M. Boutard, L. Ettwiller, T. Cerisy, A. Alberti, K. Labadie, M. Salanoubat, I. Schildkraut, and A. C. Tolonen, "Global repositioning of transcription start sites in a plant-fermenting bacterium," *Nat. Commun.*, vol. 7, p. 13783, 2016.
[189] a C. Tolonen, E. Petit, J. L. Blanchard, T. Warnick, and S. B. Leschine, "Technologies to study plant biomass fermentation using the model bacterium *Clostridium phytofermentans*," *Biol. Convers. Biomass Fuels Chem.*, no. 7, pp. 114–139, 2014.
[190] Z. Ren, T. E. Ward, B. E. Logan, and J. M. Regan, "Characterization of the cellulolytic and hydrogen-producing activities of six mesophilic Clostridium species," *J. Appl. Microbiol.*, vol. 103, no. 6, pp. 2258–2266, 2007.
[191] S. Ahmad, H. Ma, M. W. Akhtar, Y. H. P. Zhang, and X. Z. Zhang, "Directed Evolution of Clostridium phytofermentans Glycoside Hydrolase Family 9 Endoglucanase for Enhanced Specific Activity on Solid Cellulosic Substrate," *Bioenergy Res.*, vol. 7, no. 1, pp. 381–388, 2014.
[192] W. Liu, X. Z. Zhang, Z. Zhang, and Y. H. P. Zhang, "Engineering of clostridium phytofermentans endoglucanase Cel5A for improved thermostability," *Appl. Environ. Microbiol.*, vol. 76, no. 14, pp. 4914–4917, 2010.
[193] D. Brat, E. Boles, and B. Wiedemann, "Functional expression of a bacterial xylose isomerase in Saccharomyces cerevisiae," *Appl. Environ. Microbiol.*, vol. 75, no. 8, pp. 2304–2311, 2009.
[194] D. Brat and E. Boles, "Isobutanol production from d-xylose by recombinant Saccharomyces cerevisiae," *FEMS Yeast Res.*, vol. 13, no. 2, pp. 241–244, 2013.
[195] M. M. Demeke, H. Dietz, Y. Li, M. R. Foulquié-Moreno, S. Mutturi, S. Deprez, T. Den Abt, B. M. Bonini, G. Liden, F. Dumortier, A. Verplaetse, E. Boles, and J. M. Thevelein, "Development of a D-xylose fermenting and inhibitor tolerant industrial Saccharomyces cerevisiae strain with high performance in lignocellulose hydrolysates using metabolic and evolutionary engineering," *Biotechnol. Biofuels*, vol. 6, no. 1, p. 89, 2013.
[196] M. Nakajima, M. Nishimoto, and M. Kitaoka, "Characterization of Three β-Galactoside Phosphorylases from Clostridium phytofermentans: DISCOVERY OF d-GALACTOSYL-β1→4-l-RHAMNOSE PHOSPHORYLASE," *J. Biol. Chem.*, vol. 284, no. 29, pp. 19220–19227, 2009.
[197] A. C. Tolonen, A. C. Chilaka, and G. M. Church, "Targeted gene inactivation in Clostridium phytofermentans shows that cellulose degradation requires the family 9 hydrolase Cphy3367," *Mol. Microbiol.*, vol. 74, no. 6, pp. 1300–1313, 2009.
[198] X. Z. Zhang, N. Sathitsuksanoh, and Y. H. P. Zhang, "Glycoside hydrolase family 9 processive endoglucanase from Clostridium phytofermentans: Heterologous expression, characterization, and synergy with family 48 cellobiohydrolase," *Bioresour.*

Technol., vol. 101, no. 14, pp. 5534–5538, 2010.
[199] H. Liao, X.-Z. Zhang, J. A. Rollin, and Y.-H. P. Zhang, "A minimal set of bacterial cellulases for consolidated bioprocessing of lignocellulose.," *Biotechnol. J.*, vol. 6, no. 11, pp. 1409–1418, Nov. 2011.
[200] J. Ravachol, R. Borne, I. Meynial-Salles, P. Soucaille, S. Pagès, C. Tardif, and H.-P. Fierobe, "Combining free and aggregated cellulolytic systems in the cellulosome-producing bacterium Ruminiclostridium cellulolyticum.," *Biotechnol. Biofuels*, vol. 8, p. 114, 2015.
[201] X. Z. Zhang, Z. Zhang, Z. Zhu, N. Sathitsuksanoh, Y. Yang, and Y. H. P. Zhang, "The noncellulosomal family 48 cellobiohydrolase from Clostridium phytofermentans ISDg: Heterologous expression, characterization, and processivity," *Appl. Microbiol. Biotechnol.*, vol. 86, no. 2, pp. 525–533, 2010.
[202] T. Nihira, H. Nakai, and M. Kitaoka, "3-O-alpha-D-glucopyranosyl-L-rhamnose phosphorylase from Clostridium phytofermentans.," *Carbohydr. Res.*, vol. 350, pp. 94–97, Mar. 2012.
[203] T. Nihira, H. Nakai, K. Chiku, and M. Kitaoka, "Discovery of nigerose phosphorylase from Clostridium phytofermentans.," *Appl. Microbiol. Biotechnol.*, vol. 93, no. 4, pp. 1513–1522, Feb. 2012.
[204] A. C. Tolonen, T. Cerisy, H. El-Sayyed, M. Boutard, M. Salanoubat, and G. M. Church, "Fungal lysis by a soil bacterium fermenting cellulose," *Environ. Microbiol.*, vol. 17, no. 8, pp. 2618–2627, 2015.
[205] L. R. Tuck, K. Altenbach, T. F. Ang, A. D. Crawshaw, D. J. Campopiano, D. J. Clarke, and J. Marles-Wright, "Insight into Coenzyme A cofactor binding and the mechanism of acyl-transfer in an acylating aldehyde dehydrogenase from Clostridium phytofermentans," *Sci. Rep.*, vol. 6, no. October 2015, p. 22108, 2016.
[206] Y. Chen, D. C. Indurthi, S. W. Jones, and E. T. Papoutsakis, "Small RNAs in the genus clostridium," *MBio*, vol. 2, no. 1, 2011.
[207] E. Petit, W. G. LaTouf, M. V Coppi, T. A. Warnick, D. Currie, I. Romashko, S. Deshpande, K. Haas, J. G. Alvelo-Maurosa, C. Wardman, D. J. Schnell, S. B. Leschine, and J. L. Blanchard, "Involvement of a bacterial microcompartment in the metabolism of fucose and rhamnose by Clostridium phytofermentans.," *PLoS One*, vol. 8, no. 1, p. e54337, 2013.
[208] S. J. Lee, T. A. Warnick, S. B. Leschine, and S. P. Hazen, "A high-throughput biological conversion assay for determining lignocellulosic quality," *Methods Mol. Biol.*, vol. 918, pp. 341–349, 2012.
[209] S. J. Lee, T. A. Warnick, S. Pattathil, J. G. Alvelo-Maurosa, M. J. Serapiglia, H. McCormick, V. Brown, N. F. Young, D. J. Schnell, L. B. Smart, M. G. Hahn, J. F. Pedersen, S. B. Leschine, and S. P. Hazen, "Biological conversion assay using Clostridium phytofermentans to estimate plant feedstock quality," *Biotechnol. Biofuels*, vol. 5, no. 1, p. 5, 2012.
[210] A. A. Fathima, M. Sanitha, T. Kumar, S. Iyappan, and M. Ramya, "Direct utilization of waste water algal biomass for ethanol production by cellulolytic Clostridium phytofermentans DSM1183," *Bioresour. Technol.*, vol. 202, pp. 253–256, 2016.
[211] R. M. F. da Costa, S. J. Lee, G. G. Allison, S. P. Hazen, A. Winters, and M. Bosch, "Genotype, development and tissue-derived variation of cell-wall properties in the lignocellulosic energy crop Miscanthus.," *Ann. Bot.*, vol. 114, no. 6, pp. 1265–77, 2014.
[212] A. Jain, C. K. Morlok, and J. M. Henson, "Comparison of solid-state and submerged-state fermentation for the bioprocessing of switchgrass to ethanol and acetate by Clostridium phytofermentans," *Appl. Microbiol. Biotechnol.*, vol. 97, no. 2, pp. 905–917, 2013.
[213] M. Jin, V. Balan, C. Gunawan, and B. E. Dale, "Consolidated bioprocessing (CBP) performance of Clostridium phytofermentans on AFEX-treated corn stover for ethanol production," *Biotechnol. Bioeng.*, vol. 108, no. 6, pp. 1290–1297, 2011.
[214] M. Jin, C. Gunawan, V. Balan, and B. E. Dale, "Consolidated bioprocessing (CBP) of AFEX???-pretreated corn stover for ethanol production using Clostridium phytofermentans at a high solids loading," *Biotechnol. Bioeng.*, vol. 109, no. 8, pp. 1929–1936, 2012.
[215] T. R. Zuroff, S. B. Xiques, and W. R. Curtis, "Consortia-mediated bioprocessing of cellulose to ethanol with a symbiotic Clostridium phytofermentans/yeast co-culture.," *Biotechnol. Biofuels*, vol. 6, no. 1, p. 59, 2013.
[216] A. C. Tolonen, T. R. Zuroff, M. Ramya, M. Boutard, T. Cerisy, and W. R. Curtis, "Physiology, Genomics, and Pathway Engineering of an Ethanol-Tolerant Strain of Clostridium phytofermentans.," *Appl. Environ. Microbiol.*, vol. 81, no. 16, pp. 5440–8, 2015.
[217] S. Mukherjee, L. K. Thompson, S. Godin, W. Schackwitz, A. Lipzen, J. Martin, and J. L. Blanchard, "Population Level Analysis of Evolved Mutations Underlying Improvements in Plant Hemicellulose and Cellulose Fermentation by Clostridium phytofermentans," *PLoS One*, vol. 9, no. 1, p. e86731, Jan. 2014.
[218] S. Leschine and T. A. Warnick, "Systems and methods for producing biofuels and related materials." Google Patents, 23-Mar-2010.
[219] J. Blanchard, S. Leschine, E. Petit, and J. Fabel, "Methods and Compositions for Improving the Production of Fuels in Microorganisms." Google Patents, 19-Nov-2009.
[220] J. Blanchard, S. Leschine, E. Petit, J. Fabel, and M. Schmalisch, "Methods and compositions for improving the production of products in microorganisms." Google Patents, 17-May-2011.
[221] J. Blanchard, S. Leschine, E. Petit, and J. Fabel, "Methods and Compositions for Regulating Sporulation." Google Patents, 29-Apr-2010.
[222] S. Parekh, K. Karim, and J. Kilbane, "Production of fermentive end products from clostridium sp." Google Patents, 18-Jan-2012.
[223] G. Richards, "Methods and compositions for biomass treatment." Google Patents, 16-Jun-2011.
[224] T. Bowman, K. Gray, P. O'mullan, and I. Warner, "New biocatalysts and primers for production of chemicals." Google Patents, 08-Nov-2012.
[225] P. O'mullan, J. Patel, M. Schmalisch, B. Wolner, P. Reeder, and M. A. Ali, "Biocatalysts synthesizing deregulated cellulases." Google Patents, 06-Dec-2012.
[226] M. Schmalisch, W. G. Latouf, B. Wolner, and P. O'mullan, "Modulation of fermentation products through vitamin supplementation." Google Patents, 03-Oct-2013.
[227] M. Schmalisch, W. G. Latouf, B. Wolner, and P. O'mullan, "Modulation of fermentation products through vitamin supplementation." Google Patents, Oct-2013.
[228] D. Brat, E. Boles, M. Keller, and B. Wiedemann, "Prokaryotic xylose isomerase for the construction of xylose fermenting yeasts." Google Patents, 24-Mar-2015.
[229] M. E. Pyne, M. Bruder, M. Moo-Young, D. A. Chung, and C. P. Chou, "Technical guide for genetic advancement of underdeveloped and intractable Clostridium," *Biotechnol. Adv.*, vol. 32, no. 3, pp. 623–641, 2014.
[230] C. Xue, J. Zhao, L. Chen, S.-T. Yang, and F. Bai, "Recent advances and state-of-the-art strategies in strain and process engineering for biobutanol production by Clostridium acetobutylicum," *Biotechnol. Adv.*, vol. 35, no. 2, pp. 310–322, Mar. 2017.

[231] M. B. Elowitz and S. Leibler, "A synthetic oscillatory network of transcriptional regulators," pp. 335–338, 1999.
[232] K. C. B. Jennert, C. Tardif, D. I. Young, and M. Young, "Gene transfer to Clostridium cellulolyticum ATCC 35319," *Microbiology*, vol. 146, no. 12, pp. 3071–3080, 2000.
[233] P. Soucaille, R. Figge, and C. Croux, "Process for chromosomal integration and dna sequence replacement in clostridia." Google Patents, 2008.
[234] H. Dong, Y. Zhang, Z. Dai, and Y. Li, "Engineering Clostridium strain to accept unmethylated DNA," *PLoS One*, vol. 5, no. 2, 2010.
[235] Y. L. Lin and H. P. Blaschek, "Transformation of Heat-Treated Clostridium acetobutylicum Protoplasts with pUB110 Plasmid DNA.," *Appl. Environ. Microbiol.*, vol. 48, no. 4, pp. 737–742, Oct. 1984.
[236] G. Reysset and M. Sebald, "Conjugal transfer of plasmid-mediated antibiotic resistance from streptococci to Clostridium acetobutylicum.," *Ann. Inst. Pasteur. Microbiol.*, vol. 136B, no. 3, pp. 275–282, 1985.
[237] D. Lyras and J. I. Rood, "Conjugative transfer of RP4-oriT shuttle vectors from Escherichia coli to Clostridium perfringens.," *Plasmid*, vol. 39, no. 2, pp. 160–164, 1998.
[238] Y. K. Ng, M. Ehsaan, S. Philip, M. M. Collery, C. Janoir, A. Collignon, S. T. Cartman, and N. P. Minton, "Expanding the repertoire of gene tools for precise manipulation of the Clostridium difficile genome: allelic exchange using pyrE alleles.," *PLoS One*, vol. 8, no. 2, p. e56051, 2013.
[239] M. G. Group and P. Down, "Introduction of plasmids into whole cells of Clostridium acetobutylicum by electroporation," vol. 56, pp. 83–88, 1988.
[240] S. P. Allen and H. P. Blaschek, "Electroporation-induced transformation of intact cells of Clostridium perfringens.," *Appl. Environ. Microbiol.*, vol. 54, no. 9, pp. 2322–2324, 1988.
[241] M. E. Pyne, M. Moo-Young, D. a Chung, and C. P. Chou, "Development of an electrotransformation protocol for genetic manipulation of Clostridium pasteurianum.," *Biotechnol. Biofuels*, vol. 6, no. 1, p. 50, 2013.
[242] P. Trieu-Cuot, C. Carlier, C. Poyart-Salmeron, and P. Courvalin, "Shuttle vectors containing a multiple cloning site and a lacZ alpha gene for conjugal transfer of DNA from Escherichia coli to gram-positive bacteria.," *Gene*, vol. 102, no. 1, pp. 99–104, Jun. 1991.
[243] M. E. Collins, J. D. Oultram, and M. Young, "Identification of restriction fragments from two cryptic Clostridium butyricum plasmids that promote the establishment of a replication-defective plasmid in Bacillus subtilis.," *J. Gen. Microbiol.*, vol. 131, no. 8, pp. 2097–2105, Aug. 1985.
[244] D. Purdy, T. A. T. O'Keeffe, M. Elmore, M. Herbert, A. McLeod, M. Bokori-Brown, A. Ostrowski, and N. P. Minton, "Conjugative transfer of clostridial shuttle vectors from Escherichia coli to Clostridium difficile through circumvention of the restriction barrier," *Mol. Microbiol.*, vol. 46, no. 2, pp. 439–452, 2002.
[245] J. T. Heap, O. J. Pennington, S. T. Cartman, and N. P. Minton, "A modular system for Clostridium shuttle plasmids.," *J. Microbiol. Methods*, vol. 78, no. 1, pp. 79–85, Jul. 2009.
[246] J. Groom, D. Chung, D. G. Olson, L. R. Lynd, A. M. Guss, and J. Westpheling, "Promiscuous plasmid replication in thermophiles: Use of a novel hyperthermophilic replicon for genetic manipulation of Clostridium thermocellum at its optimum growth temperature," *Metab. Eng. Commun.*, vol. 3, pp. 30–38, 2016.
[247] T. McKenzie, T. Hoshino, T. Tanaka, and N. Sueoka, "The nucleotide sequence of pUB110: Some salient features in relation to replication and its regulation," *Plasmid*, vol. 15, no. 2, pp. 93–103, 1986.
[248] D. G. Olson and L. R. Lynd, "Computational design and characterization of a temperature-sensitive plasmid replicon for gram positive thermophiles.," *J. Biol. Eng.*, vol. 6, no. 1, p. 5, 2012.
[249] S.-H. Lee, M.-A. Kwon, S. Choi, S. Kim, J. Kim, Y.-A. Shin, and K. H. Kim, "A New Shuttle Plasmid That Stably Replicates in Clostridium acetobutylicum.," *J. Microbiol. Biotechnol.*, vol. 25, no. 10, pp. 1702–1708, Oct. 2015.
[250] B. Molitor, K. Kirchner, A. W. Henrich, S. Schmitz, and M. A. Rosenbaum, "Expanding the molecular toolkit for the homoacetogen Clostridium ljungdahlii," *Sci. Rep.*, vol. 6, p. 31518, 2016.
[251] D. G. Olson and L. R. Lynd, "Computational design and characterization of a temperature-sensitive plasmid replicon for gram positive thermophiles.," *J. Biol. Eng.*, vol. 6, no. 1, p. 5, 2012.
[252] G. Z. Cui, J. Zhang, W. Hong, C. Xu, Y. Feng, Q. Cui, and Y. J. Liu, "Improvement of ClosTron for successive gene disruption in Clostridium cellulolyticum using a pyrF-based screening system," *Appl. Microbiol. Biotechnol.*, vol. 98, no. 1, pp. 313–323, 2014.
[253] M. A. Al-Hinai, A. G. Fast, and E. T. Papoutsakis, "Novel system for efficient isolation of clostridium double-crossover allelic exchange mutants enabling markerless chromosomal gene deletions and DNA integration," *Appl. Environ. Microbiol.*, vol. 78, no. 22, pp. 8112–8121, 2012.
[254] M. Ehsaan, W. Kuit, Y. Zhang, S. T. Cartman, J. T. Heap, K. Winzer, and N. P. Minton, "Mutant generation by allelic exchange and genome resequencing of the biobutanol organism Clostridium acetobutylicum ATCC 824," *Biotechnol. Biofuels*, vol. 9, no. 1, p. 4, 2016.
[255] S. A. Tripathi, D. G. Olson, D. A. Argyros, B. B. Miller, T. F. Barrett, D. M. Murphy, J. D. McCool, A. K. Warner, V. B. Rajgarhia, L. R. Lynd, D. A. Hogsett, and N. C. Caiazza, "Development of pyrF-Based Genetic System for Targeted Gene Deletion in Clostridium thermocellum and Creation of a pta Mutant," *Appl. Environ. Microbiol.*, vol. 76, no. 19, pp. 6591–6599, 2010.
[256] J.-C. Blouzard, O. Valette, C. Tardif, and P. de Philip, "Random mutagenesis of Clostridium cellulolyticum by using a Tn1545 derivative.," *Appl. Environ. Microbiol.*, vol. 76, no. 13, pp. 4546–9, 2010.
[257] X. Han, X. D. Du, L. Southey, D. M. Bulach, T. Seemann, X. X. Yan, T. L. Bannam, and J. I. Rood, "Functional analysis of a bacitracin resistance determinant located on ICECp1, a novel tn916-like element from a conjugative plasmid in clostridium perfringens," *Antimicrob. Agents Chemother.*, vol. 59, no. 11, pp. 6855–6865, 2015.
[258] D. G. Olson, S. a Tripathi, R. J. Giannone, J. Lo, N. C. Caiazza, D. a Hogsett, R. L. Hettich, A. M. Guss, G. Dubrovsky, and L. R. Lynd, "Deletion of the Cel48S cellulase from Clostridium thermocellum.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, no. 41, pp. 17727–17732, 2010.
[259] H. Celik, J.-C. Blouzard, B. Voigt, D. Becher, V. Trotter, H.-P. Fierobe, C. Tardif, S. Pagès, and P. de Philip, "A Two-Component System (XydS/R) Controls the Expression of Genes Encoding CBM6-Containing Proteins in Response to Straw in Clostridium cellulolyticum," *PLoS One*, vol. 8, no. 2, p. e56063, 2013.
[260] A. Charlebois, L.-A. Jalbert, J. Harel, L. Masson, and M. Archambault, "Characterization of Genes Encoding for Acquired Bacitracin Resistance in Clostridium perfringens," *PLoS One*, vol. 7, no. 9, p. e44449, Sep. 2012.
[261] S. T. Cartman, M. L. Kelly, D. Heeg, J. T. Heap, and N. P. Minton, "Precise Manipulation of the Clostridium difficile Chromosome Reveals a Lack of Association between the tcdC Genotype and Toxin Production," *Appl. Environ. Microbiol.*, vol. 78, no. 13, pp. 4683–4690, 2012.

[262] D. A. Argyros, S. A. Tripathi, T. F. Barrett, S. R. Rogers, L. F. Feinberg, D. G. Olson, J. M. Foden, B. B. Miller, L. R. Lynd, D. A. Hogsett, and N. C. Caiazza, "High ethanol titers from cellulose by using metabolically engineered thermophilic, anaerobic microbes," *Appl. Environ. Microbiol*, vol. 77, 2011.
[263] W. Higashide, Y. Li, Y. Yang, and J. C. Liao, "Metabolic engineering of Clostridium cellulolyticum for production of isobutanol from cellulose," *Appl. Environ. Microbiol.*, vol. 77, no. 8, pp. 2727–2733, 2011.
[264] L. Girbal, I. Mortier-barrière, C. Rouanet, C. Croux, I. Mortier-barrie, and P. Soucaille, "Development of a Sensitive Gene Expression Reporter System and an Inducible Promoter-Repressor System for Clostridium acetobutylicum Development of a Sensitive Gene Expression Reporter System and an Inducible Promoter-Repressor System for Clostridium acet," vol. 69, no. 8, pp. 4985–4988, 2003.
[265] L. Feustel, S. Nakotte, P. Dürre, and P. Du, "Characterization and Development of Two Reporter Gene Systems for Clostridium acetobutylicum Characterization and Development of Two Reporter Gene Systems for Clostridium acetobutylicum," *Appl. Environ. Microbiol.*, vol. 70, no. 2, pp. 798–803, 2004.
[266] S. B. Tummala, N. E. Welker, and E. T. Papoutsakis, "Development and characterization of a gene expression reporter system for Clostridium acetobutylicum ATCC 824," *Appl. Environ. Microbiol.*, vol. 65, no. 9, pp. 3793–3799, 1999.
[267] S. M. Gaida, A. Liedtke, A. H. W. Jentges, B. Engels, and S. Jennewein, "Metabolic engineering of Clostridium cellulolyticum for the production of n-butanol from crystalline cellulose," *Microb. Cell Fact.*, vol. 15, no. 1, p. 6, 2016.
[268] D. G. Olson, M. Maloney, A. A. Lanahan, S. Hon, L. J. Hauser, and L. R. Lynd, "Identifying promoters for gene expression in Clostridium thermocellum," *Metab. Eng. Commun.*, vol. 2, pp. 23–29, 2015.
[269] S. A. Tripathi, D. G. Olson, D. A. Argyros, B. B. Miller, T. F. Barrett, D. M. Murphy, J. D. McCool, A. K. Warner, V. B. Rajgarhia, L. R. Lynd, D. A. Hogsett, and N. C. Caiazza, "Development of pyrF-Based Genetic System for Targeted Gene Deletion in Clostridium thermocellum and Creation of a pta Mutant ," *Appl. Environ. Microbiol.*, vol. 76, no. 19, pp. 6591–6599, Oct. 2010.
[270] S. B. Melville, R. Labbe, and A. L. Sonenshein, "Expression from the Clostridium perfringens cpe promoter in C. perfringens and Bacillus subtilis," *Infect. Immun.*, vol. 62, no. 12, pp. 5550–5558, 1994.
[271] N. Mani, D. Lyras, L. Barroso, P. Howarth, T. Wilkins, J. I. Rood, A. L. Sonenshein, and B. Dupuy, "Environmental response and autoregulation of Clostridium difficile TxeR, a sigma factor for toxin gene expression.," *J. Bacteriol.*, vol. 184, no. 21, pp. 5971–8, Nov. 2002.
[272] C. Perret S, "Production of heterologous and chimeric scaffoldins by Clostridium acetobutylicum ATCC 824," *J. Bacteriol.*, vol. 186, no. 1, pp. 253–257, 2004.
[273] M. Leibig, B. Krismer, M. Kolb, A. Friede, F. Götz, and R. Bertram, "Marker removal in staphylococci via Cre recombinase and different lox sites," *Appl. Environ. Microbiol.*, vol. 74, no. 5, pp. 1316–1323, 2008.
[274] J. T. Heap, O. J. Pennington, S. T. Cartman, G. P. Carter, and N. P. Minton, "The ClosTron: a universal gene knock-out system for the genus Clostridium," *J Microbiol Methods*, vol. 70, 2007.
[275] A. H. Hartman, H. Liu, and S. B. Melville, "Construction and characterization of a lactose-inducible promoter system for controlled gene expression in Clostridium perfringens.," *Appl. Environ. Microbiol.*, vol. 77, no. 2, pp. 471–8, 2011.
[276] A. Banerjee, C. Leang, T. Ueki, K. P. Nevin, and D. R. Lovley, "Lactose-inducible system for metabolic engineering of clostridium ljungdahlii," *Appl. Environ. Microbiol.*, vol. 80, no. 8, pp. 2410–2416, 2014.
[277] E. Mearls, D. Olson, C. Herring, and L. Lynd, "Development of a regulatable plasmid-based gene expression system for Clostridium thermocellum," *Appl. Microbiol. Biotechnol.*, pp. 1–11, 2015.
[278] S. M. McBride and A. L. Sonenshein, "Identification of a genetic locus responsible for antimicrobial peptide resistance in Clostridium difficile.," *Infect. Immun.*, vol. 79, no. 1, pp. 167–76, Jan. 2011.
[279] A. N. Edwards, R. A. Pascual, K. O. Childress, K. L. Nawrocki, E. C. Woods, and S. M. McBride, "An alkaline phosphatase reporter for use in Clostridium difficile," *Anaerobe*, vol. 32, pp. 98–104, 2015.
[280] H. Nariya, S. Miyata, T. Kuwahara, and A. Okabe, "Development and characterization of a xylose-inducible gene expression system for Clostridium perfringens," *Appl. Environ. Microbiol.*, vol. 77, no. 23, pp. 8439–8441, 2011.
[281] J. Zhang, Y.-J. Liu, G.-Z. Cui, and Q. Cui, "A novel arabinose-inducible genetic operation system developed for Clostridium cellulolyticum.," *Biotechnol. Biofuels*, vol. 8, p. 36, 2015.
[282] S. Nuyts, L. Van Mellaert, J. Theys, W. Landuyt, P. Lambin, and J. Anne, "The use of radiation-induced bacterial promoters in anaerobic conditions: A means to control gene expression in Clostridium-mediated therapy for cancer," *Radiat. Res.*, vol. 155, no. 5, pp. 716–723, 2001.
[283] H. Dong, W. Tao, Y. Zhang, and Y. Li, "Development of an anhydrotetracycline-inducible gene expression system for solvent-producing *Clostridium acetobutylicum*: A useful tool for strain engineering.," *Metab. Eng.*, vol. 14, no. 1, pp. 59–67, 2012.
[284] S. Nagaraju, N. K. Davies, D. J. F. Walker, M. Köpke, and S. D. Simpson, "Genome editing of Clostridium autoethanogenum using CRISPR/Cas9," *Biotechnol. Biofuels*, vol. 9, no. 1, p. 219, Dec. 2016.
[285] J. Wang, X. Ai, H. Mei, Y. Fu, B. Chen, Z. Yu, and J. He, "High-Throughput Identification of Promoters and Screening of Highly Active Promoter-5′-UTR DNA Region with Different Characteristics from Bacillus thuringiensis," *PLoS One*, vol. 8, no. 5, pp. 1–11, 2013.
[286] M. De Mey, J. Maertens, G. J. Lequeux, W. K. Soetaert, and E. J. Vandamme, "Construction and model-based analysis of a promoter library for E. coli: an indispensable tool for metabolic engineering," *BMC Biotechnology*, vol. 7. London, p. 34, 2007.
[287] C. Matsushita, O. Matsushita, M. Koyama, and a Okabe, "A Clostridium perfringens vector for the selection of promoters.," *Plasmid*, vol. 31. pp. 317–319, 1994.
[288] W. V Shaw, "Chloramphenicol acetyltransferase from chloramphenicol-resistant bacteria.," *Methods Enzymol.*, vol. 43, pp. 737–755, 1975.
[289] "Zhang Y, Grosse-Honebrink A, Minton NP. A universal mariner transposon system for forward genetic studies in the genus Clostridium. PLoS One. 2015;2;10(4):e0122411." .
[290] M. Kaji, O. Matsushita, E. Tamai, S. Miyata, Y. Taniguchi, S. Shimamoto, S. Katayama, S. Morita, and A. Okabe, "A novel type of DNA curvature present in a Clostridium perfringens ferredoxin gene: characterization and role in gene expression," *Microbiology*, vol. 149, no. 11, pp. 3083–3091, Nov. 2003.
[291] L. Abdou, C. Boileau, P. De Philip, S. Pagès, H. P. Fiérobe, and C. Tardif, "Transcriptional regulation of the Clostridium cellulolyticum cip-cel operon: A complex mechanism involving a catabolite-responsive element," *J. Bacteriol.*, vol. 190, no. 5, pp. 1499–1506, 2008.
[292] B. W. Wren, P. Mullany, C. Clayton, and S. Tabaqchali, "Molecular Cloning and Genetic Analysis of a Chloramphenicol Acetyltransferase Determinant from Clostridium dijficile," *Antimicrob. Agents Chemother.*, vol. 32, no. 8, pp. 1213–1217,

1988.
[293] P. P. Lin, K. S. Rabe, J. L. Takasumi, M. Kadisch, F. H. Arnold, and J. C. Liao, "Isobutanol production at elevated temperatures in thermophilic Geobacillus thermoglucosidasius," *Metab. Eng.*, vol. 24, pp. 1–8, 2014.
[294] N. Mani, D. Lyras, L. Barroso, P. Howarth, T. Wilkins, J. I. Rood, A. L. Sonenshein, and B. Dupuy, "Environmental response and autoregulation of Clostridium difficile TxeR, a sigma factor for toxin gene expression.," *J. Bacteriol.*, vol. 184, no. 21, pp. 5971–8, Nov. 2002.
[295] A. Ravagnani, K. C. B. Jennert, E. Steiner, R. Grünberg, J. R. Jefferies, S. R. Wilkinson, D. I. Young, E. C. Tidswell, D. P. Brown, P. Youngman, J. Gareth Morris, and M. Young, "Spo0A directly controls the switch from acid to solvent production in solvent-forming clostridia," *Mol. Microbiol.*, vol. 37, no. 5, pp. 1172–1185, 2000.
[296] A. H. Hartman, H. Liu, and S. B. Melville, "Construction and characterization of a lactose-inducible promoter system for controlled gene expression in Clostridium perfringens.," *Appl. Environ. Microbiol.*, vol. 77, no. 2, pp. 471–8, 2011.
[297] A. N. Edwards, R. A. Pascual, K. O. Childress, K. L. Nawrocki, E. C. Woods, and S. M. McBride, "An alkaline phosphatase reporter for use in Clostridium difficile," *Anaerobe*, vol. 32, pp. 98–104, 2015.
[298] B. R. Anjuwon-Foster, R. Tamayo, H. Browne, W. Leung, J. Frank, and A. Schmitz, "A genetic switch controls the production of flagella and toxins in Clostridium difficile," *PLOS Genet.*, vol. 13, no. 3, p. e1006701, Mar. 2017.
[299] E. M. Ransom, C. D. Ellermeier, and D. S. Weiss, "Use of mCherry red fluorescent protein for studies of protein localization and gene expression in Clostridium difficile," *Appl. Environ. Microbiol.*, vol. 81, no. 5, pp. 1652–1660, 2015.
[300] E. M. Ransom, K. B. Williams, D. S. Weiss, and C. D. Ellermeier, "Identification and characterization of a gene cluster required for proper rod shape, cell division, and pathogenesis in Clostridium difficile.," *J. Bacteriol.*, vol. 196, no. 12, pp. 2290–300, Jun. 2014.
[301] T. Drepper, T. Eggert, F. Circolone, A. Heck, U. Krauß, J.-K. Guterl, M. Wendorff, A. Losi, W. Gärtner, and K.-E. Jaeger, "Reporter proteins for in vivo fluorescence without oxygen," *Nat. Biotechnol.*, vol. 25, no. 4, pp. 443–445, 2007.
[302] A. M. Buckley, C. Jukes, D. Candlish, J. J. Irvine, J. Spencer, R. P. Fagan, A. J. Roe, J. M. Christie, N. F. Fairweather, and G. R. Douce, "Lighting Up Clostridium Difficile: Reporting Gene Expression Using Fluorescent Lov Domains," *Sci. Rep.*, vol. 6, no. 1, p. 23463, Sep. 2016.
[303] B. Molitor, K. Kirchner, A. W. Henrich, S. Schmitz, and M. A. Rosenbaum, "Expanding the molecular toolkit for the homoacetogen Clostridium ljungdahlii," *Sci. Rep.*, vol. 6, p. 31518, 2016.
[304] C. Cho and S. Y. Lee, "Efficient gene knockdown in *Clostridium acetobutylicum* by synthetic small regulatory RNAs," *Biotechnol. Bioeng.*, vol. 114, no. 2, pp. 374–383, Feb. 2017.
[305] M. K. Phillips-Jones, F. E.F., M. T., S. N., H. H., R. J.I., and S. T., "Use of a *lux* reporter system for monitoring rapid changes in Î±-toxin gene expression in *Clostridium perfringens* during growth," *FEMS Microbiol. Lett.*, vol. 188, no. 1, pp. 29–33, Jul. 2000.
[306] C. P. Cassona, F. Pereira, M. Serrano, and A. O. Henriques, "A Fluorescent Reporter for Single Cell Analysis of Gene Expression in Clostridium difficile," 2016, pp. 69–90.
[307] L. Saujet, F. C. Pereira, M. Serrano, O. Soutourina, M. Monot, P. V Shelyakin, M. S. Gelfand, B. Dupuy, a O. Henriques, and I. Martin-Verstraete, "Genome-wide analysis of cell type-specific gene transcription during spore formation in Clostridium difficile," *PLoS Genet*, vol. 9, no. 10, p. e1003756, 2013.
[308] X. C. Wang, S. C. Wilson, M. C. Hammond, R. D.J., D. M.J., W. C.M., L. V.T., I. A.T., C. H.K., and H. Y.-F., "Next-generation RNA-based fluorescent biosensors enable anaerobic detection of cyclic di-GMP," *Nucleic Acids Res.*, vol. 44, no. 17, pp. e139–e139, Sep. 2016.
[309] E. M. Green and G. N. Bennett, "Inactivation of an aldehyde/alcohol dehydrogenase gene from Clostridium acetobutylicum ATCC 824," *Appl Biochem Biotechnol*, vol. 57, 1996.
[310] E. M. Green, Z. L. Boynton, L. M. Harris, F. B. Rudolph, E. T. Papoutsakis, and G. N. Bennett, "Genetic manipulation of acid formation pathways by gene inactivation in Clostridium acetobutylicum ATCC 824," *Microbiol*, vol. 142, 1996.
[311] M. A. Al-Hinai, A. G. Fast, and E. T. Papoutsakis, "Novel system for efficient isolation of Clostridium double-crossover allelic exchange mutants enabling markerless chromosomal gene deletions and DNA integration," *Appl Env. Microbiol*, vol. 78, 2014.
[312] Y. K. Ng, M. Ehsaan, S. Philip, M. M. Collery, C. Janoir, A. Collignon, S. T. Cartman, and N. P. Minton, "Expanding the repertoire of gene tools for precise manipulation of the Clostridium difficile genome: allelic exchange using pyrE alleles," *PLoS One*, vol. 8, 2013.
[313] S. A. Tripathi, D. G. Olson, D. A. Argyros, B. B. Miller, T. F. Barrett, D. M. Murphy, J. D. McCool, A. K. Warner, V. B. Rajgarhia, L. R. Lynd, D. A. Hogsett, and N. C. Caiazza, "Development of pyrF-Based Genetic System for Targeted Gene Deletion in Clostridium thermocellum and Creation of a pta Mutant ," *Appl. Environ. Microbiol.*, vol. 76, no. 19, pp. 6591–6599, Oct. 2010.
[314] C. Croux, N.-P.-T. Nguyen, J. Lee, C. Raynaud, F. Saint-Prix, M. Gonzalez-Pajuelo, I. Meynial-Salles, and P. Soucaille, "Construction of a restriction-less, marker-less mutant useful for functional genomic and metabolic engineering of the biofuel producer Clostridium acetobutylicum," *Biotechnol. Biofuels*, vol. 9, no. 1, p. 23, Dec. 2016.
[315] D. A. Argyros, S. A. Tripathi, T. F. Barrett, S. R. Rogers, L. F. Feinberg, D. G. Olson, J. M. Foden, B. B. Miller, L. R. Lynd, D. A. Hogsett, and N. C. Caiazza, "High ethanol titers from cellulose by using metabolically engineered thermophilic, anaerobic microbes," *Appl. Environ. Microbiol*, vol. 77, 2011.
[316] J. T. Heap, M. Ehsaan, C. M. Cooksley, Y.-K. Ng, S. T. Cartman, K. Winzer, and N. P. Minton, "Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker.," *Nucleic Acids Res.*, vol. 40, no. 8, p. e59, 2012.
[317] M. A. Al-Hinai, A. G. Fast, and E. T. Papoutsakis, "Novel system for efficient isolation of Clostridium double-crossover allelic exchange mutants enabling markerless chromosomal gene deletions and DNA integration," *Appl Env. Microbiol*, vol. 78, 2014.
[318] C. Croux, N.-P.-T. Nguyen, J. Lee, C. Raynaud, F. Saint-Prix, M. Gonzalez-Pajuelo, I. Meynial-Salles, and P. Soucaille, "Construction of a restriction-less, marker-less mutant useful for functional genomic and metabolic engineering of the biofuel producer Clostridium acetobutylicum," *Biotechnol. Biofuels*, vol. 9, no. 1, p. 23, Dec. 2016.
[319] N. Zhang, L. Shao, Y. Jiang, Y. Gu, Q. Li, J. Liu, W. Jiang, and S. Yang, "I-SceI-mediated scarless gene modification via allelic exchange in Clostridium," *J. Microbiol. Methods*, vol. 108, pp. 49–60, 2015.
[320] N. P. Minton, M. Ehsaan, C. M. Humphreys, G. T. Little, J. Baker, A. M. Henstra, F. Liew, M. L. Kelly, L. Sheng, K. Schwarz, and Y. Zhang, "A roadmap for gene system development in Clostridium.," *Anaerobe*, vol. 41, pp. 104–112, Oct. 2016.
[321] L. J. Abraham and J. I. Rood, "Identification of Tn4451 and Tn4452, chloramphenicol resistance transposons from Clostridium perfringens.," *J. Bacteriol.*, vol. 169, no. 4, pp. 1579–84, Apr. 1987.

[322] H. Hächler, F. H. Kayser, and B. Berger-Bächi, "Homology of a transferable tetracycline resistance determinant of Clostridium difficile with Streptococcus (Enterococcus) faecalis transposon Tn916.," *Antimicrob. Agents Chemother.*, vol. 31, no. 7, pp. 1033–8, Jul. 1987.
[323] S. T. Cartman and N. P. Minton, "A mariner-based transposon system for in vivo random mutagenesis of Clostridium difficile.," *Appl. Environ. Microbiol.*, vol. 76, no. 4, pp. 1103–9, Feb. 2010.
[324] Y. Zhang, S. Xu, C. Chai, S. Yang, W. Jiang, N. P. Minton, and Y. Gu, "Development of an inducible transposon system for efficient random mutagenesis in *Clostridium acetobutylicum*," *FEMS Microbiol. Lett.*, vol. 363, no. 8, p. fnw065, Apr. 2016.
[325] Y. Chen, B. A. McClane, D. J. Fisher, J. I. Rood, and P. Gupta, "Construction of an alpha toxin gene knockout mutant of Clostridium perfringens type A by use of a mobile group II intron.," *Appl. Environ. Microbiol.*, vol. 71, no. 11, pp. 7542–7, Nov. 2005.
[326] L. Shao, S. Hu, Y. Yang, Y. Gu, J. Chen, Y. Yang, W. Jiang, and S. Yang, "Targeted gene disruption by use of a group II intron (targetron) vector in Clostridium acetobutylicum," *Cell Res.*, vol. 17, no. 11, pp. 963–965, Nov. 2007.
[327] J. T. Heap, O. J. Pennington, S. T. Cartman, G. P. Carter, and N. P. Minton, "The ClosTron: A universal gene knock-out system for the genus Clostridium," *J. Microbiol. Methods*, vol. 70, no. 3, pp. 452–464, 2007.
[328] A. M. Lambowitz and S. Zimmerly, "Mobile group II introns.," *Annu. Rev. Genet.*, vol. 38, pp. 1–35, 2004.
[329] J. Perutka, W. Wang, D. Goerlitz, and A. M. Lambowitz, "Use of Computer-designed Group II Introns to Disrupt Escherichia coli DExH/D-box Protein and DNA Helicase Genes," *J. Mol. Biol.*, vol. 336, no. 2, pp. 421–439, 2004.
[330] P. J. Enyeart, S. M. Chirieleison, M. N. Dao, J. Perutka, E. M. Quandt, J. Yao, J. T. Whitt, A. T. Keatinge-Clay, A. M. Lambowitz, and A. D. Ellington, "Generalized bacterial genome editing using mobile group II introns and Cre-lox," *Mol. Syst. Biol.*, vol. 9, no. 685, p. 685, 2013.
[331] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier, "A Programmable Dual-RNA – Guided DNA Endonuclease in Adaptice Bacterial Immunity," *Science*, vol. 337, no. August, pp. 816–822, 2012.
[332] Y. Wang, Z.-T. Zhang, S.-O. Seo, K. Choi, T. Lu, Y.-S. Jin, and H. P. Blaschek, "Markerless chromosomal gene deletion in Clostridium beijerinckii using CRISPR/Cas9 system," *J. Biotechnol.*, vol. 200, pp. 1–5, Apr. 2015.
[333] T. Xu, Y. Li, Z. Shi, C. L. Hemme, Y. Li, Y. Zhu, J. D. Van Nostrand, Z. He, and J. Zhou, "Efficient genome editing in clostridium cellulolyticum via CRISPR-Cas9 nickase," *Appl. Environ. Microbiol.*, vol. 81, no. 13, pp. 4423–4431, 2015.
[334] Q. Li, J. Chen, N. P. Minton, Y. Zhang, Z. Wen, J. Liu, H. Yang, Z. Zeng, X. Ren, J. Yang, Y. Gu, W. Jiang, Y. Jiang, and S. Yang, "CRISPR-based genome editing and expression control systems in *Clostridium acetobutylicum* and *Clostridium beijerinckii*," *Biotechnol. J.*, vol. 11, no. 7, pp. 961–972, Jul. 2016.
[335] Y. Wang, Z.-T. Zhang, S.-O. Seo, P. Lynn, T. Lu, Y.-S. Jin, and H. P. Blaschek, "Bacterial Genome Editing with CRISPR-Cas9: Deletion, Integration, Single Nucleotide Modification, and Desirable 'Clean' Mutant Selection in *Clostridium beijerinckii* as an Example," *ACS Synth. Biol.*, vol. 5, no. 7, pp. 721–732, Jul. 2016.
[336] Y. Wang, Z. T. Zhang, S. O. Seo, P. Lynn, T. Lu, Y. S. Jin, and H. P. Blaschek, "Gene transcription repression in Clostridium beijerinckii using CRISPR-dCas9," *Biotechnol. Bioeng.*, vol. 113, no. 12, pp. 2739–2743, 2016.
[337] M. R. Bruder, M. E. Pyne, M. Moo-Young, D. A. Chung, and C. P. Chou, "Extending CRISPR-Cas9 technology from genome editing to transcriptional engineering in the genus Clostridium," *Appl. Environ. Microbiol.*, vol. 82, no. 20, pp. 6109–6119, 2016.
[338] H. Huang, C. Chai, N. Li, P. Rowe, N. P. Minton, S. Yang, W. Jiang, and Y. Gu, "CRISPR/Cas9-based efficient genome editing in Clostridium ljungdahlii, an autotrophic gas-fermenting bacterium," *ACS Synth. Biol.*, p. acssynbio.6b00044, 2016.
[339] S. Wang, S. Dong, P. Wang, Y. Tao, and Y. Wang, "Genome Editing in Clostridium saccharoperbutylacetonicum N1-4 with the CRISPR-Cas9 System," *Appl. Environ. Microbiol.*, vol. 83, no. 10, pp. e00233-17, May 2017.
[340] C. Xue, J. Zhao, L. Chen, S.-T. Yang, and F. Bai, "Recent advances and state-of-the-art strategies in strain and process engineering for biobutanol production by Clostridium acetobutylicum," *Biotechnol. Adv.*, vol. 35, no. 2, pp. 310–322, Mar. 2017.
[341] A. A. Herrero and R. F. Gomez, "Development of ethanol tolerance in Clostridium thermocellum: effect of growth temperature.," *Appl. Environ. Microbiol.*, vol. 40, no. 3, pp. 571–7, Sep. 1980.
[342] Y. L. Lin and H. P. Blaschek, "Butanol Production by a Butanol-Tolerant Strain of Clostridium acetobutylicum in Extruded Corn Broth.," *Appl. Environ. Microbiol.*, vol. 45, no. 3, pp. 966–73, Mar. 1983.
[343] S. H. Baer, H. P. Blaschek, and T. L. Smith, "Effect of Butanol Challenge and Temperature on Lipid Composition and Membrane Fluidity of Butanol-Tolerant Clostridium acetobutylicum," *Appl. Environ. Microbiol.*, vol. 53, no. 12, pp. 2854–2861, 1987.
[344] S. N. Bowring and J. G. Morris, "Mutagenesis of Clostridium acetobutylicum.," *J. Appl. Bacteriol.*, vol. 58, no. 6, pp. 577–584, Jun. 1985.
[345] J. Formanek, R. Mackie, and H. P. Blaschek, "Enhanced Butanol Production by Clostridium beijerinckii BA101 Grown in Semidefined P2 Medium Containing 6 Percent Maltodextrin or Glucose," *Appl. Environ. Microbiol.*, vol. 63, no. 6, pp. 2306–2310, 1997.
[346] S. Mao, Y. Luo, T. Zhang, J. Li, G. Bao, Y. Zhu, Z. Chen, Y. Zhang, Y. Li, and Y. Ma, "Proteome Reference Map and Comparative Proteomic Analysis between a Wild Type Clostridium acetobutylicum DSM 1731 and its Mutant with Enhanced Butanol Tolerance and Butanol Yield."
[347] X. Gao, H. Zhao, G. Zhang, K. He, and Y. Jin, "Genome Shuffling of Clostridium acetobutylicum CICC 8012 for Improved Production of Acetone–Butanol–Ethanol (ABE)," *Curr. Microbiol.*, vol. 65, no. 2, pp. 128–132, Aug. 2012.
[348] X. B. Liu, Q. Y. Gu, X. Bin Yu, and W. Luo, "Enhancement of butanol tolerance and butanol yield in Clostridium acetobutylicum mutant NT642 obtained by nitrogen ion beam implantation," *J. Microbiol.*, vol. 50, no. 6, pp. 1024–1028, 2012.
[349] Y. S. Jang, A. Malaviya, and S. Y. Lee, "Acetone-butanol-ethanol production with high productivity using Clostridium acetobutylicum BKM19," *Biotechnol Bioeng*, vol. 110, no. 6, pp. 1646–1653, 2013.
[350] G. Luan, Z. Cai, F. Gong, H. Dong, Z. Lin, Y. Zhang, and Y. Li, "Developing controllable hypermutable Clostridium cells through manipulating its methyl-directed mismatch repair system," *Protein Cell*, vol. 4, no. 11, pp. 854–862, Nov. 2013.
[351] H. Li, W. Luo, Q. Wang, and X. Yu, "Direct Fermentation of Gelatinized Cassava Starch to Acetone, Butanol, and Ethanol Using Clostridium acetobutylicum Mutant Obtained by Atmospheric and Room Temperature Plasma," *Appl. Biochem. Biotechnol.*, vol. 172, no. 7, pp. 3330–3341, Apr. 2014.
[352] R. P. Desai and E. T. Papoutsakis, "Antisense RNA strategies for metabolic engineering of Clostridium acetobutylicum.," *Appl. Environ. Microbiol.*, vol. 65, no. 3, pp. 936–45, Mar. 1999.
[353] S. B. Tummala, N. E. Welker, and E. T. Papoutsakis, "Design of antisense RNA constructs for downregulation of the acetone

formation pathway of Clostridium acetobutylicum.," *J. Bacteriol.*, vol. 185, no. 6, pp. 1923–34, Mar. 2003.
[354] R. Sillers, M. A. Al-Hinai, and E. T. Papoutsakis, "Aldehyde-alcohol dehydrogenase and/or thiolase overexpression coupled with CoA transferase downregulation lead to higher alcohol titers and selectivity in *Clostridium acetobutylicum* fermentations," *Biotechnol. Bioeng.*, vol. 102, no. 1, pp. 38–49, Jan. 2009.
[355] S. Perret, H. Maamar, J.-P. Bélaich, and C. Tardif, "Use of antisense RNA to modify the composition of cellulosomes produced by Clostridium cellulolyticum," *Mol. Microbiol.*, vol. 51, no. 2, pp. 599–607, Feb. 2004.
[356] D. Raju, P. Setlow, and M. R. Sarker, "Antisense-RNA-mediated decreased synthesis of small, acid-soluble spore proteins leads to decreased resistance of clostridium perfringens spores to moist heat and UV radiation.," *Appl. Environ. Microbiol.*, vol. 73, no. 7, pp. 2048–53, Apr. 2007.
[357] S. Nakayama, T. Kosaka, H. Hirakawa, K. Matsuura, S. Yoshino, and K. Furukawa, "Metabolic engineering for solvent productivity by downregulation of the hydrogenase gene cluster hupCBA in Clostridium saccharoperbutylacetonicum strain N1-4," *Appl. Microbiol. Biotechnol.*, vol. 78, no. 3, pp. 483–493, Mar. 2008.
[358] P. Boudry, E. Semenova, M. Monot, K. A. Datsenko, A. Lopatina, O. Sekulovic, M. Ospina-Bedoya, L.-C. Fortier, K. Severinov, B. Dupuy, and O. Soutourina, "Function of the CRISPR-Cas System of the Human Pathogen Clostridium difficile.," *MBio*, vol. 6, no. 5, pp. e01112-15, Sep. 2015.
[359] D. T. Jones and D. R. Woods, "Acetone-butanol fermentation revisited," *Microbiol Rev*, vol. 50, 1986.
[360] D. T. Jones, M. Shirley, X. Wu, and S. Keis, "Bacteriophage infections in the industrial acetone butanol (AB) fermentation process.," *J. Mol. Microbiol. Biotechnol.*, vol. 2, no. 1, pp. 21–6, Jan. 2000.
[361] T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner, and H. Mori, "Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection.," *Mol. Syst. Biol.*, vol. 2, p. 2006.0008, 2006.
[362] R. S. Senger and E. T. Papoutsakis, "Genome-scale model for Clostridium acetobutylicum: Part I. Metabolic network resolution and analysis," *Biotechnol. Bioeng.*, vol. 101, no. 5, pp. 1036–1052, 2008.
[363] R. A. Thompson, S. Dahal, S. Garcia, I. Nookaew, and C. T. Trinh, "Exploring complex cellular phenotypes and model-guided strain design with a novel genome-scale metabolic model of Clostridium thermocellum DSM 1313 implementing an adjustable cellulosome.," *Biotechnol. Biofuels*, vol. 9, no. 1, p. 194, 2016.
[364] S. Dash, C. Y. Ng, and C. D. Maranas, "Metabolic modeling of clostridia: current developments and applications," *FEMS Microbiol. Lett.*, vol. 363, no. 4, p. fnw004, Feb. 2016.
[365] Enerdata, *Energy Statistical Yearbook*. 2016.
[366] U. Nations, "United Nations Framework Convention, Kyoto," Kyoto, 1997.
[367] K. Renew, "s t a t i s t i c s Key Renew able s Trends," 2016.
[368] Ren21, "The First Decade: 2004-2014, 10 Years of Renewable Energy Progress," pp. 2004–2014, 2014.
[369] Sustainable Energy for All, "Strategic Framework for results," 2016.
[370] C. Prentice, "White House denies Trump order on biofuels program pending, markets roiled," *Reuters*, 2017.
[371] European Parliament, "Directive 2009/28/EC of the European Parliament and of the Council of 23 April 2009," *Off. J. Eur. Union*, vol. 140, no. 16, pp. 16–62, 2009.
[372] E. Parliament, "Directive 2003/71/EC," *Euratom*, vol. 2001, no. May, pp. 20–30, 2004.
[373] R. Sims, M. Taylor, S. Jack, and W. Mabee, "From 1st to 2nd Generation Bio Fuel Technologies: An overview of current industry and RD&D activities," *IEA Bioenergy*, no. November, pp. 1–124, 2008.
[374] P. Rosendal, B. H. Nielsen, and N. K. Lange, "Stability of bacterial α-amylase in the starch liquefaction process," *Starch - Stärke*, vol. 31, no. 11, pp. 368–372, 1979.
[375] R. L. Antrim, "Liquefaction of granular-starch slurries using alpha-amylase in the presence of carbonate ion," 1992.
[376] G. E. Erickson, T. J. Klopfenstein, and a K. Watson, *Tilization of Feed Co-Products From Wet or Dry Milling for Beef Cattle*. 2012.
[377] P. R. Lennartsson, P. Erlandsson, and M. J. Taherzadeh, "Integration of the first and second generation bioethanol processes and the importance of by-products," *Bioresour. Technol.*, vol. 165, no. C, pp. 3–8, 2014.
[378] European Academies Science Advisory Council, *The current status of biofuels in the European Union, their environmental impacts and future prospects; EASAC policy report 19*, no. December. 2012.
[379] J.-C. Viguie, N. Ullrich, P. Porot, L. Bournay, M. Hecquet, and J. Rousseau, "BioTfueL Project: Targeting the Development of Second-Generation Biodiesel and Biojet Fuels," *Oil Gas Sci. Technol. D Ifp Energies Nouv.*, vol. 68, no. 5, pp. 935–946, 2013.
[380] G. E. Erickson, T. J. Klopfenstein, D. C. Adams, and R. J. Rasby, "Corn Processing Co-products manual: A review of current research on distillers grains and corn gluten," *Univ. Nebraska-Lincoln Nebraska Corn Board*, p. 32, 2012.
[381] M. Pauly, M. Pauly, K. Keegstra, and K. Keegstra, "Cell-wall carbohydrates and their modi cation as a resource for biofuels," *Plant J.*, vol. 54, no. 4, pp. 559–68, 2008.
[382] H. V. Scheller and P. Ulvskov, "Hemicelluloses," *Annu. Rev. Plant Biol.*, vol. 61, pp. 263–289, 2010.
[383] H.-D. Yoo, D. Kim, and S.-H. Paek, "Plant cell wall polysaccharides as potential resources for the development of novel prebiotics.," *Biomol. Ther. (Seoul).*, vol. 20, no. 4, pp. 371–9, 2012.
[384] R. M. Brown, "Cellulose Structure and Biosynthesis: What is in Store for the 21st Century?," in *Journal of Polymer Science, Part A: Polymer Chemistry*, 2004, vol. 42, no. 3, pp. 487–495.
[385] S. Brethauer and M. H. Studer, "Biochemical Conversion Processes of Lignocellulosic Biomass to Fuels and Chemicals - A Review.," *Chimia (Aarau).*, vol. 69, no. 10, pp. 572–81, 2015.
[386] M. Pauly, M. Pauly, K. Keegstra, and K. Keegstra, "Cell-wall carbohydrates and their modi cation as a resource for biofuels," *Plant J.*, vol. 54, no. 4, pp. 559–68, 2008.
[387] R. W. Whetten, J. J. MacKay, and R. R. Sederoff, "Recent Advances in Understanding Lignin Biosynthesis," *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, vol. 49, no. 1, pp. 585–609, 1998.
[388] M. S. Singhvi, S. Chaudhari, and D. V. Gokhale, "Lignocellulose processing: a current challenge," *RSC Adv.*, vol. 4, p. 8271, 2014.
[389] M. S. Singhvi, S. Chaudhari, and D. V. Gokhale, "Lignocellulose processing: a current challenge," *RSC Adv.*, vol. 4, p. 8271, 2014.
[390] A. Brandt, J. Gräsvik, J. P. Hallett, and T. Welton, "Deconstruction of lignocellulosic biomass with ionic liquids," *Green Chem.*, vol. 15, no. 3, pp. 550–583, 2013.
[391] F. H. Isikgor and C. Remzi Becer, "Lignocellulosic Biomass: a sustainable platform for production of bio-based chemicals and polymers," *Polym. Chem.*, vol. 6, pp. 4497–4559, 2015.
[392] P. Alvira, E. Tom??s-Pej??, M. Ballesteros, and M. J. Negro, "Pretreatment technologies for an efficient bioethanol production

process based on enzymatic hydrolysis: A review," *Bioresour. Technol.*, vol. 101, no. 13, pp. 4851–4861, 2010.
[393] R. P. Chandra, R. Bura, W. E. Mabee, A. Berlin, X. Pan, and J. N. Saddler, "Substrate pretreatment: The key to effective enzymatic hydrolysis of lignocellulosics?," *Adv. Biochem. Eng. Biotechnol.*, vol. 108, no. May, pp. 67–93, 2007.
[394] A. Margeot, B. Hahn-Hagerdal, M. Edlund, R. Slade, and F. Monot, "New improvements for lignocellulosic ethanol," *Curr. Opin. Biotechnol.*, vol. 20, no. 3, pp. 372–380, 2009.
[395] D. Dahnum, S. O. Tasum, E. Triwahyuni, M. Nurdin, and H. Abimanyu, "Comparison of SHF and SSF processes using enzyme and dry yeast for optimization of bioethanol production from empty fruit bunch," *Energy Procedia*, vol. 68, pp. 107–116, 2015.
[396] E. Tomás-Pejó, J. M. Oliva, M. Ballesteros, and L. Olsson, "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.*, vol. 100, no. 6, pp. 1122–1131, 2008.
[397] P. M. Bondesson and M. Galbe, "Biotechnology for Biofuels Process design of SSCF for ethanol production from steam‐pretreated , acetic‐acid‐impregnated wheat straw," *Biotechnol. Biofuels*, pp. 1–12, 2016.
[398] R. Su, Y. Ma, W. Qi, M. Zhang, F. Wang, R. Du, J. Yang, M. Zhang, and Z. He, "Ethanol Production from High-Solid SSCF of Alkaline-Pretreated Corncob Using Recombinant Zymomonas mobilis CP4," *Bioenergy Res.*, vol. 6, no. 1, pp. 292–299, 2013.
[399] S. Brethauer and M. H. Studer, "Biochemical Conversion Processes of Lignocellulosic Biomass to Fuels and Chemicals - A Review.," *Chimia (Aarau).*, vol. 69, no. 10, pp. 572–81, 2015.
[400] D. G. Olson, J. E. McBride, A. Joe Shaw, and L. R. Lynd, "Recent progress in consolidated bioprocessing," *Curr. Opin. Biotechnol.*, vol. 23, no. 3, pp. 396–405, 2012.
[401] a C. Tolonen, E. Petit, J. L. Blanchard, T. Warnick, and S. B. Leschine, "Technologies to study plant biomass fermentation using the model bacterium *Clostridium phytofermentans*," *Biol. Convers. Biomass Fuels Chem.*, no. 7, pp. 114–139, 2014.
[402] L. R. Lynd, W. H. Van Zyl, J. E. McBride, and M. Laser, "Consolidated bioprocessing of cellulosic biomass: An update," *Curr. Opin. Biotechnol.*, vol. 16, no. 5, pp. 577–583, 2005.
[403] Q. Xu, A. Singh, and M. E. Himmel, "Perspectives and new directions for the production of bioethanol using consolidated bioprocessing of lignocellulose," *Curr. Opin. Biotechnol.*, vol. 20, no. 3, pp. 364–371, 2009.
[404] R. den Haan, E. van Rensburg, S. H. Rose, J. F. Görgens, and W. H. van Zyl, "Progress and challenges in the engineering of non-cellulolytic microorganisms for consolidated bioprocessing," *Curr. Opin. Biotechnol.*, vol. 33, pp. 15–22, 2015.
[405] J.-J. Chang, F.-J. Ho, C.-Y. Ho, Y.-C. Wu, Y.-H. Hou, C.-C. Huang, M.-C. Shih, and W.-H. Li, "Assembling a cellulase cocktail and a cellodextrin transporter into a yeast host for CBP ethanol production.," *Biotechnol. Biofuels*, vol. 6, no. 1, p. 19, 2013.
[406] W. H. Van Zyl, L. R. Lynd, R. Den Haan, and J. E. McBride, "Consolidated bioprocessing for bioethanol production using saccharomyces cerevisiae," *Adv. Biochem. Eng. Biotechnol.*, vol. 108, no. April, pp. 205–235, 2007.
[407] J. L. Cotter, M. S. Chinn, and A. M. Grunden, "Ethanol and acetate production by Clostridium ljungdahlii and Clostridium autoethanogenum using resting cells.," *Bioprocess Biosyst. Eng.*, vol. 32, no. 3, pp. 369–380, Apr. 2009.
[408] D. G. Olson, J. E. McBride, A. Joe Shaw, and L. R. Lynd, "Recent progress in consolidated bioprocessing," *Curr. Opin. Biotechnol.*, vol. 23, no. 3, pp. 396–405, 2012.
[409] Nrel, "A look back at the U. S. Department of Energy's aquatic species program: biodiesel from algae," *Report*, vol. 328, p. 291 p, 1998.
[410] S. Behera, R. Singh, R. Arora, N. K. Sharma, M. Shukla, and S. Kumar, "Scope of algae as third generation biofuels.," *Front. Bioeng. Biotechnol.*, vol. 2, no. February, p. 90, 2014.
[411] Biofuel.org.uk, "Third Generation Biofuels."
[412] K. Ullah, M. Ahmad, Sofia, V. K. Sharma, P. Lu, A. Harvey, M. Zafar, S. Sultana, and C. N. Anyanwu, "Algal biomass as a global source of transport fuels: Overview and development perspectives," *Prog. Nat. Sci. Mater. Int.*, vol. 24, no. 4, pp. 329–339, 2014.
[413] C. Y. Chen, X. Q. Zhao, H. W. Yen, S. H. Ho, C. L. Cheng, D. J. Lee, F. W. Bai, and J. S. Chang, "Microalgae-based carbohydrates for biofuel production," *Biochem. Eng. J.*, vol. 78, no. January 2017, pp. 1–10, 2013.
[414] E. M. Aro, "From first generation biofuels to advanced solar biofuels," *Ambio*, vol. 45, no. 1, pp. 24–31, 2016.
[415] M. E. Martin, H. Richter, S. Saha, and L. T. Angenent, "Traits of selected Clostridium strains for syngas fermentation to ethanol," *Biotechnol. Bioeng.*, vol. 113, no. 3, pp. 531–539, 2016.
[416] E. Marcellin, J. B. Behrendorff, S. Nagaraju, S. DeTissera, S. Segovia, R. Palfreyman, J. Daniell, C. Licona-Cassani, L. Quek, R. Speight, M. P. Hodson, S. D. Simpson, W. P. Mitchell, M. Köpke, and L. K. Nielsen, "Low carbon fuels and commodity chemicals from waste gases – Systematic approach to understand energy metabolism in a model acetogen," *Green Chem.*, pp. 3020–3028, 2016.
[417] M. (Lanzatech) Koepke, "Carbon recycling for fuel and chemical production," in *BASF Science Symposium*, 2015.
[418] G. M. Church, M. B. Elowitz, C. D. Smolke, C. a Voigt, and R. Weiss, "Realizing the potential of synthetic biology.," *Nat. Rev. Mol. Cell Biol.*, vol. 15, no. 4, pp. 289–294, 2014.
[419] M. E. E. Essalem and W. J. Mitchell, "Identification of a glucose-mannose phosphotransferase system in Clostridium beijerinckii," pp. 1–18, 2016.
[420] W. J. Mitchell, "The Phosphotransferase System in Solventogenic Clostridia," *J. Mol. Microbiol. Biotechnol.*, vol. 25, no. 2–3, pp. 129–142, 2015.
[421] Y. Nataf, S. Yaron, F. Stahl, R. Lamed, E. A. Bayer, T. H. Scheper, A. L. Sonenshein, and Y. Shoham, "Cellodextrin and laminaribiose ABC transporters in Clostridium thermocellum," *J. Bacteriol.*, vol. 91, no. 1, pp. 203–209, 2009.
[422] J. Ravachol, P. de Philip, R. Borne, P. Mansuelle, M. J. Maté, S. Perret, and H.-P. Fierobe, "Mechanisms involved in xyloglucan catabolism by the cellulosome-producing bacterium Ruminiclostridium cellulolyticum," *Sci. Rep.*, vol. 6, no. February, p. 22770, 2016.
[423] R. D. Finn, P. Coggill, R. Y. Eberhardt, S. R. Eddy, J. Mistry, A. L. Mitchell, S. C. Potter, M. Punta, M. Qureshi, A. Sangrador-Vegas, G. A. Salazar, J. Tate, and A. Bateman, "The Pfam protein families database: towards a more sustainable future," *Nucleic Acids Res.*, vol. 44, no. D1, pp. D279-285, Jan. 2016.
[424] L. J. Deakin, S. Clare, R. P. Fagan, L. F. Dawson, D. J. Pickard, M. R. West, B. W. Wren, N. F. Fairweather, G. Dougan, and T. D. Lawley, "The Clostridium difficile spo0A gene is a persistence and transmission factor.," *Infect. Immun.*, vol. 80, no. 8, pp. 2704–11, Aug. 2012.
[425] V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho, and B. Henrissat, "The carbohydrate-active enzymes database (CAZy) in 2013," *Nucleic Acids Res*, vol. 42, 2014.
[426] N. Sternberg, "Bacteriophage P1 site-specific recombination," *J. Mol. Biol.*, vol. 150, no. 4, pp. 603–608, 1981.
[427] J. M. Lambert, R. S. Bongers, and M. Kleerebezem, "Cre-lox-based system for multiple gene deletions and selectable-marker

removal in Lactobacillus plantarum.," *Appl. Environ. Microbiol.*, vol. 73, no. 4, pp. 1126–35, Feb. 2007.
[428] B. Sauer, "Functional Expression of the cre-lox Site-Specific Recombination System in the Yeast Saccharomyces cerevisiae," *Mol. Cell. Biol.*, vol. 7, no. 6, pp. 2087–2096, 1987.
[429] B. Sauer and N. Henderson, "Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1," *Genetics*, vol. 85, pp. 5166–5170, 1988.
[430] E. A. Johnson, A. Madia, and A. L. Demain, "Chemically defined minimal medium for growth of the anaerobic cellulolytic thermophile Clostridium thermocellum," *Applied and Environmental Microbiology*, vol. 41, no. 4. pp. 1060–1062, 1981.
[431] G. Mercy, J. Mozziconacci, V. F. Scolari, K. Yang, G. Zhao, A. Thierry, Y. Luo, L. A. Mitchell, M. Shen, Y. Shen, R. Walker, W. Zhang, Y. Wu, Z. Xie, Z. Luo, Y. Cai, J. Dai, H. Yang, Y.-J. Yuan, J. D. Boeke, J. S. Bader, H. Muller, and R. Koszul, "3D organization of synthetic and scrambled chromosomes," *Science (80-.).*, vol. 355, no. 6329, 2017.
[432] N. R. Baral and A. Shah, "Microbial inhibitors: formation and effects on acetone-butanol-ethanol fermentation of lignocellulosic biomass," *Appl. Microbiol. Biotechnol.*, vol. 98, no. 22, pp. 9151–9172, Nov. 2014.
[433] P. Marlière, J. Patrouix, V. Döring, P. Herdewijn, S. Tricot, S. Cruveiller, M. Bouzon, and R. Mutzel, "Chemical Evolution of a Bacterium's Genome," *Angew. Chemie Int. Ed.*, vol. 50, no. 31, pp. 7109–7114, Jul. 2011.
[434] D. Baum, "Reading a Phylogenetic Tree : The Meaning of Monophyletic Groups," *Nat. Educ.*, vol. 1, no. 1, pp. 1–5, 2009.

APPENDICES

Appendix 1

M. Boutard, **T. Cerisy**, P.-Y. Nogue, A. Alberti, J. Weissenbach, M. Salanoubat, and A. C. Tolonen, "Functional Diversity of Carbohydrate-Active Enzymes Enabling a Bacterium to Ferment Plant Biomass.," *PLoS Genet.*, vol. 10, no. 11, p. e1004773, 2014.

OPEN ACCESS Freely available online

PLOS | GENETICS

Functional Diversity of Carbohydrate-Active Enzymes Enabling a Bacterium to Ferment Plant Biomass

Magali Boutard[1,2,3], Tristan Cerisy[1,2,3], Pierre-Yves Nogue[1,2,3], Adriana Alberti[1], Jean Weissenbach[1,2], Marcel Salanoubat[1,2,3], Andrew C. Tolonen[1,2,3]*

1 Genoscope, CEA, DSV, IG, Évry, France, **2** CNRS-UMR8030, Évry, France, **3** Department of Biology, Université d'Évry Val d'Essonne, Évry, France

Abstract

Microbial metabolism of plant polysaccharides is an important part of environmental carbon cycling, human nutrition, and industrial processes based on cellulosic bioconversion. Here we demonstrate a broadly applicable method to analyze how microbes catabolize plant polysaccharides that integrates carbohydrate-active enzyme (CAZyme) assays, RNA sequencing (RNA-seq), and anaerobic growth screening. We apply this method to study how the bacterium *Clostridium phytofermentans* ferments plant biomass components including glucans, mannans, xylans, galactans, pectins, and arabinans. These polysaccharides are fermented with variable efficiencies, and diauxies prioritize metabolism of preferred substrates. Strand-specific RNA-seq reveals how this bacterium responds to polysaccharides by up-regulating specific groups of CAZymes, transporters, and enzymes to metabolize the constituent sugars. Fifty-six up-regulated CAZymes were purified, and their activities show most polysaccharides are degraded by multiple enzymes, often from the same family, but with divergent rates, specificities, and cellular localizations. CAZymes were then tested in combination to identify synergies between enzymes acting on the same substrate with different catalytic mechanisms. We discuss how these results advance our understanding of how microbes degrade and metabolize plant biomass.

Citation: Boutard M, Cerisy T, Nogue P-Y, Alberti A, Weissenbach J, et al. (2014) Functional Diversity of Carbohydrate-Active Enzymes Enabling a Bacterium to Ferment Plant Biomass. PLoS Genet 10(11): e1004773. doi:10.1371/journal.pgen.1004773

Editor: Paul M. Richardson, MicroTrek Incorporated, United States of America

Received July 16, 2014; **Accepted** September 24, 2014; **Published** November 13, 2014

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. RNA sequencing files in FASTQ format have been deposited in the European Nucleotide Archive under accessions ERP006991-ERP007002.

Funding: This work was funded by the Genoscope-CEA and a CNRS chaire d'excellence to ACT. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: atolonen@genoscope.cns.fr

Introduction

Plants annually produce 200 billion tons of lignocellulosic biomass [1], which is metabolized by specialized microbes in diverse environments. For instance, recycling of plant biomass by soil [2] and marine [3] microbes is a key part of the global carbon cycle and intestinal bacteria ferment indigestible plant fiber to short chain fatty acids that constitute 60–85% of calories in ruminants and 5–10% in humans [4]. Further, as only 2% of cellulosic biomass is currently used by humans [5], it is a vast potential feedstock that industrial microbes could convert into energy and commodities. Elucidating how microbes depolymerize and metabolize plant biomass is thus important to understand carbon flow in the environment, to promote healthy human nutrition and prevent disease [6], and to develop industrial processes based on cellulosic bioconversion.

Most of plant biomass is in the cell wall, a macromolecular network of phenolic lignin and three types of polysaccharides (cellulose, hemicelluloses, and pectins) whose relative abundances vary widely among species and tissues (Table S1). The load bearing structure of the cell wall consists of cellulose fibrils tethered by various types of hemicellulose. Hemicellulose is enriched in xylan [7] and xyloglucan [8] in dicots, arabinoxylan in monocots [9], and galacto- and glucomannans in gymnosperms [10]. Outside the cell wall, mannans also act as storage polysaccharides in seeds [11], similar to starch. Pectins are cross-linked galacturonic acid-based polysaccharides that act in cellular adhesion and primary wall extension. More than 60% of pectin is often homogalacturonan (HG) [12], which is esterified with methanol to various degrees. Rhamnogalacturonan I (RGI) [13], the second most abundant pectin, can have galactan and arabinan side chains on the rhamnose residues [14]. Because plant tissues are composed of such heterogeneous polysaccharides, plant-degrading microbes express a myriad of carbohydrate-active enzymes (CAZymes) [15], each of which modifies or cleaves a specific type of sugar linkage.

Here we demonstrate a strategy for systematic analysis of the enzymatic machinery used by microbes to degrade and metabolize plant polysaccharides. Among these microbes, the plant-fermenting clostridia are of particular interest for being a dominant group in the human gut microbiome [6] and top candidates to transform cellulosic biomass into fuels and commodities [16], [17]. We studied *Clostridium phytofermentans* [18], a soil bacterium with 171 CAZyme-encoding genes (Table S2) including 116 glycoside hydrolases in 44 different families. We first quantified growth on comprehensive panel of plant polysaccharides and sugars (Table S3). Strand-specific RNA sequencing revealed all genes whose expression changed on the various substrates. In particular, we focused on up-regulated CAZyme genes and determined how they are organized into regulons that respond to specific

Author Summary

Plant-fermenting bacteria are important for the global carbon cycle, human nutrition, and industrial production of renewable fuels and commodities from cellulosic biomass. Plants are primarily composed of heterogeneous polysaccharides, requiring plant-degrading microbes to encode many carbohydrate-active enzymes (CAZymes) to cleave different sugar linkages. Here we develop a broadly applicable method to study how microbes catabolize plant biomass by determining the combination of CAZymes that depolymerize each polysaccharide into sugars, how the cell alters global mRNA expression, and the efficiency with which each polysaccharide is metabolized. We apply this method to investigate how *Clostridium phytofermentans*, a bacterium encoding 171 CAZymes, ferments polysaccharides. We assimilate our results into a genetic model of how this bacterium metabolizes plant biomass and discuss how these results further our understanding of microbial plant fermentation.

polysaccharides. A set of 56 up-regulated CAZymes were cloned, purified, and an ''each enzyme versus each substrate'' screen quantified their abilities to bind and cleave plant polysaccharides. These enzymes were then tested in combination to identify synergies for polysaccharide degradation. We discuss how the results can be integrated to further our knowledge of how microbes metabolize plant biomass.

Results/Discussion

Growth on polysaccharides and sugars

We developed a high resolution, microtiter anaerobic growth assay that shows *C. phytofermentans* ferments diverse plant polysaccharides (Fig. 1) and their constituent monosaccharides (Fig. S1), but with widely varying cell yields and growth rates (Table S4). It also forms colonies on solid medium containing each polysaccharide except arabinogalactan II (AGII) (Fig. S2). Growth was fastest on HG (Fig. 1A, generation time 0.70h), similar to rumen microbes that digest pectin more rapidly than cellulose and hemicellulose [19]. Although *C. phytofermentans* ferments both galacturonic acid (Fig. S1F) and rhamnose (Fig. S1H), cell yield was low on RGI (Fig. 1B). *C. phytofermentans* grows well on galactan (Fig. 1C), xylans (Fig. 1F–G), mannans (Fig. 1H–I), xyloglucan (Fig. 1J), and starch (Fig. 1L). Limited growth on AGII (Fig. 1E) relative to galactan supports that *C. phytofermentans* cleaves β-1,4 galactan, but not the β-1,3 and β-1,6-galactose bonds in AGII. Poor growth on arabinan (Fig. 1D) is similar to arabinose (Fig. S1G), suggesting this sugar is transported or metabolized inefficiently. *C. phytofermentans* grows well on cellulose plates (Fig. S2) and solubilizes cellulosic substrates such as filter paper and raw corn stover (Fig. S3), but weak growth on carboxymethylcellulose (CMC) might result from either lack of a suitable endoglucanase or carboxymethyl side groups inhibiting its metabolism.

C. phytofermentans shows diauxic growth on the mixed sugar polysaccharides galactomannan (Fig. 1I) and xyloglucan (Fig. 1J). For each of these substrates, one of the component sugars (galactose or glucose) supports faster growth than the other (mannose or xylose) (Fig. S1, Table S4). Growth on various mixtures of galactose/mannose (Fig. S4) and of glucose/xylose (Fig. S5) shows rapid metabolism of the preferred sugar followed by slower growth on the other one. However, in both cases when the favored sugar reached 75% of the total, the other sugar does not appear to be metabolized.

Similar to some ruminal [19] and human gut microbes [20], *C. phytofermentans* often grows faster on polysaccharides than the constituent sugars (Table S4). When presented with mixtures of xylan and xylose, this bacterium shows diauxic growth with preferential metabolism of xylan (Fig. S6), which is surprising because xylan must be cleaved to xylose to be metabolized. Growth on polysaccharides could be energetically favorable if significant ATP is saved by simultaneous transport of multiple sugar units in a single oligosaccharide [21] or by intracellular phosphorolysis of oligosaccharides [22]. *C. phytofermentans* encodes at least a dozen phosphorylases [23] [24], which cleave oligosaccharides without using ATP. Although the mechanisms regulating sugar metabolism in *C. phytofermentans* are unknown, diauxic growth supports carbon catabolite repression prioritizes growth on preferred sugars and polysaccharides.

Gene expression

We quantified mRNA expression by strand-specific RNA sequencing during log-phase growth on 8 polysaccharides, 3 monosaccharides, and raw corn stover as a complex biomass substrate. An average of 17.3 million mRNA reads were mapped per sample (Table S5), yielding expression (RPKM) values (Table S6) that were highly correlated ($r^2 = 0.96$–0.99) between duplicate cultures for all conditions (Fig. S7). The reads were also highly strand-specific (Fig. S8), which will facilitate their future use for *de novo* transcriptome assembly, gene annotation and detection of antisense transcription. The fraction of reads mapping to CAZymes during growth on glucose was 2.0%, but this increased greatly on polysaccharides, especially cellulose (11.9%) and stover (31.0%). We assessed which genes were significantly differentially expressed on each polysaccharide relative to glucose using DESeq [25] (Table S7). Expression of CAZyme genes on polysaccharides relative to glucose (Fig. 2) shows that between 15 (cellobiose) and 40 (stover) CAZymes were significantly up-regulated per treatment (Table S8) with a total of 92 CAZymes up-regulated on at least one polysaccharide.

The differentially expressed CAZymes are putatively classified by the CAZy database as 67 glycoside hydrolases, 6 carbohydrate esterases, 4 polysaccharide lyases, 14 glycosyl transferases, and 2 CBM proteins. We analyzed the specificity of the CAZyme transcriptional response by K-means clustering the expression profiles of these genes (Fig. 3, Table S9). Cluster A consists six genes that were highly up-regulated on multiple substrates: the GH26 *cphy1071*, the GH11 *cphy2105*, two GH18 chitinases *cphy1799* and *cphy1800* [26], the GH9 cellulase *cphy3367* [27] [28] and the GH48 cellulase *cphy3368* [29]. Clusters B–F respond to specific polysaccharides such as homogalacturonan (clusters B,C), starch (cluster D), xylan (cluster E), and cellulose/arabinan (cluster F). *C. phytofermentans* thus perceives signals from individual polysaccharides and responds by up-regulating specific transcriptional regulons that enable it to tailor its complement of CAZymes to the polysaccharide substrate.

CAZyme activities

A set of 56 CAZymes up-regulated on polysaccharides were His-tagged, overexpressed, purified, and their abilities to bind and cleave polysaccharides were quantified. The CAZy database classifies these enzymes putatively as 47 glycoside hydrolases, 4 polysaccharide lyases, and 4 carbohydrate esterases (Table S2); putative glycosyltransferases were not examined as they are not involved in polysaccharide catabolism [6]. Thirty-two enzymes have significant cleavage or binding activities (Fig. 4, Table S11).

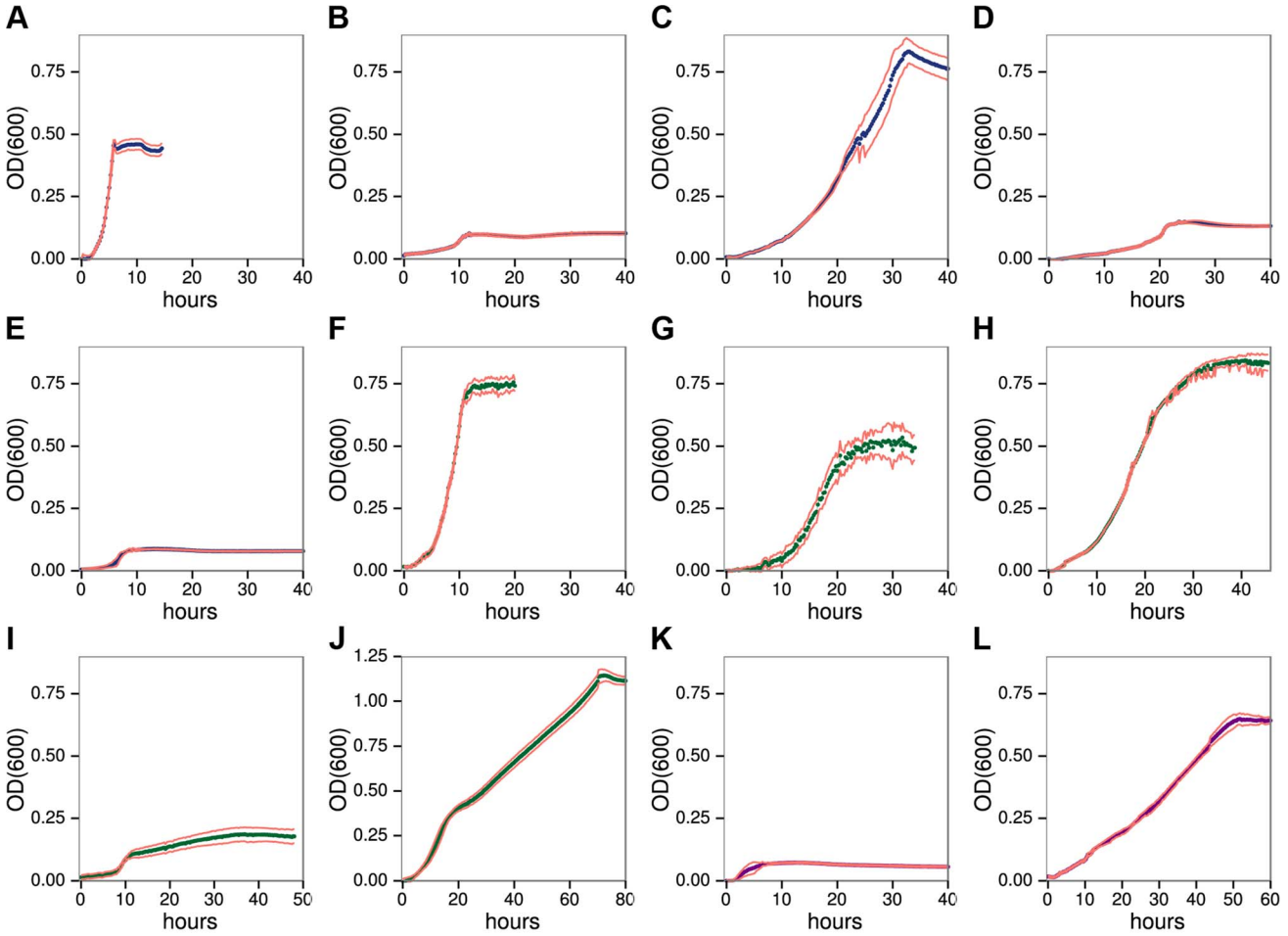

Figure 1. ***C. phytofermentans*** **growth on pectic A–E, hemicellulosic F–J, and glucan K–L.** Polysaccharides: homogalacturonan **A**, rhamnogalacturonan I **B**, galactan **C**, arabinan **D**, arabinogalactan II **E**, xylan **F**, arabinoxylan **G**, glucomannan **H**, galactomannan **I**, xyloglucan **J**, carboxymethylcellulose **K**, starch **L**. Growth was measured as OD_{600} every 15 minutes. Each point is the mean of six cultures; red lines show one standard deviation.
doi:10.1371/journal.pgen.1004773.g001

Some substrates such as β-1,4-galactan appear to be cut by a single, highly active enzyme, while multiple CAZymes from the same family degrade other substrates such as xylan (GH10), mannan (GH26), starch (GH13), and HG (PL9). CAZymes from multiple families together depolymerize substrates such as xyloglucan (GH2,5,12,31), glucomannan (GH5,GH9,GH26) and galactomannan (GH5,GH26).

We found 15 CAZymes that bind insoluble polysaccharides, most commonly cellulose and mannan (Fig. 4B). Unexpectedly, no CAZyme bound corn stover, suggesting that partial digestion of raw biomass is needed to facilitate enzyme binding. Nine enzymes that bound substrates have carbohydrate binding modules (CBM), but some enzymes such as the cellulase Cphy1163 can bind their substrate without one. While CBM are known to discriminate between polysaccharides such as cellulose and mannan [30], we observed overlap with cellulase CBMs binding mannan and vice versa. Further, CBM from xylanases can bind cellulose and mannose, but with lower affinity, showing that CBM often bind a range of polysaccharides. Consistent with their cleavage activities, GH13 were the only enzymes to bind starch. Enzymes with CBM usually also have catalytic modules, but Cphy1713, a CAZyme with a CBM32 and no catalytic module, binds galactomannan. CBM32 are known to bind galactose and this protein may function similar to one in Yersinia that is proposed to bind oligosaccharides to prevent them from leaking out of the cell [31].

Thirty-two CAZy families have multiple members, which often have divergent cleavage activities and cellular localizations. Cphy1510 has the highest activity among the four GH10 active on xylan (Fig. 5A). Cphy3010, the GH10 with lowest activity, is the only one lacking a secretion signal, supporting it acts intracellularly on xylo-oligosaccharides while the other GH10 are extracellular. Members of the GH5 family act on a wide range of polysaccharides [32]. *C. phytofermentans* encodes 3 GH5 enzymes, among which one is active on galactomannan and two on xyloglucan (Fig. 5B). The GH5 Cphy1163 has no activity on either of these substrates, but is the most active on cellulose and glucomannan. The 3 GH26 also vary in substrate specificities (Fig. 5C); all the GH26 are similarly active on β-mannan, but only Cphy1071 has cellulase activity and it has lower activity on gluco- and galactomannan. Sequenced-based families are thus useful to make general substrate predictions for CAZymes, but experiments are needed to determine substrate range and catalytic efficiency.

CAZymes mixtures can degrade polysaccharides more efficiently than individual enzymes. We assessed pairwise interactions between each CAZyme and a second enzyme on cellulose (Cphy3367), xylan (Cphy2105), glucomannan (Cphy1071), and

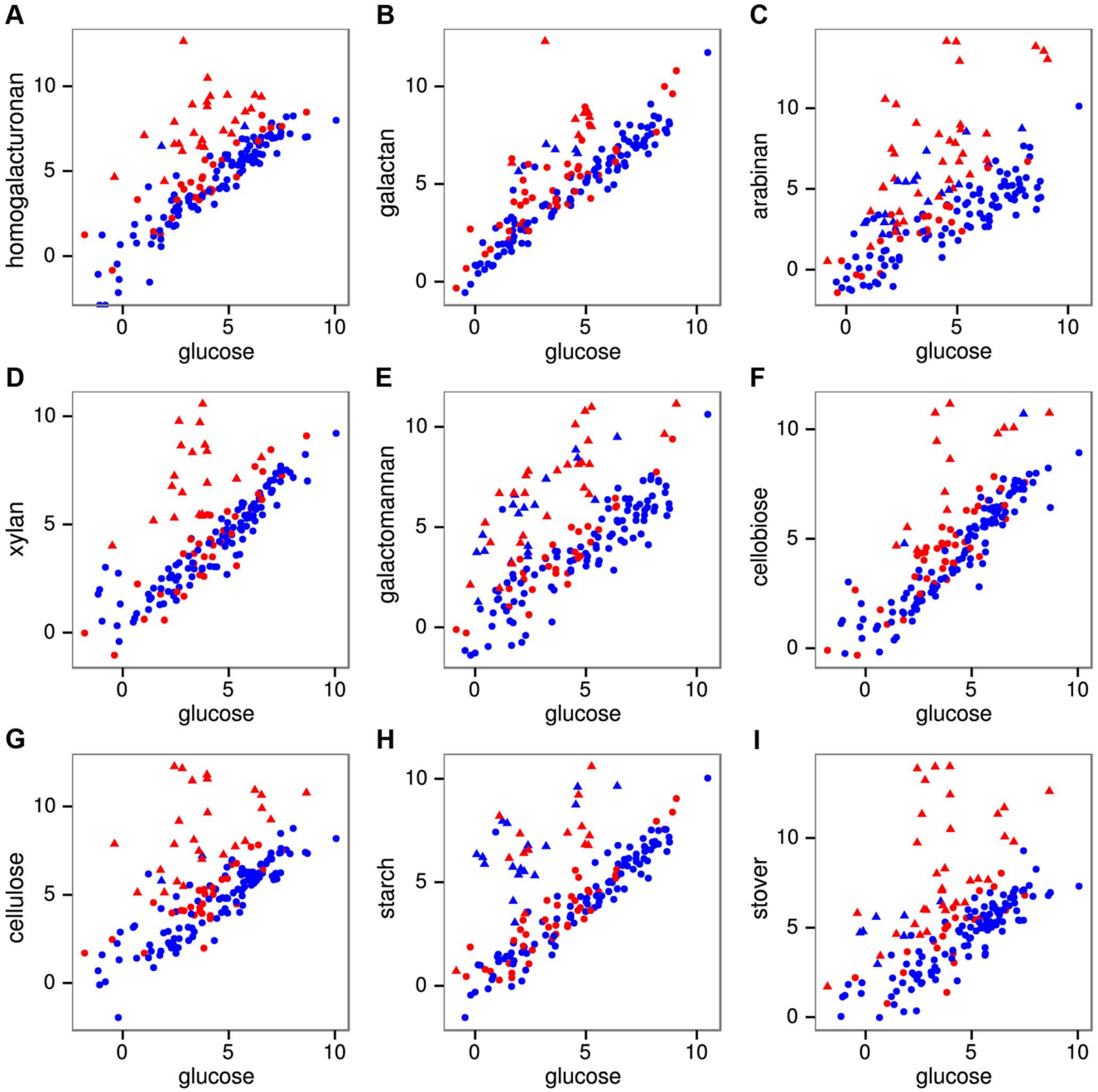

Figure 2. mRNA expression of all 171 CAZymes during growth on pectins A–C, hemicelluloses D–E, glucans F–H, and raw corn stover I relative to expression on glucose. Expression was quantified as $\log_2$(RPKM) with significantly differentially expressed genes on a given polysaccharide shown as triangles and unchanged genes as circles. The 56 purified CAZymes are red and others are blue.
doi:10.1371/journal.pgen.1004773.g002

homogalacturonan (Cphy1687) (Fig. 5D–G). Similar to results showing synergy between the GH9 Cphy3367 and a *B. subtilis* GH5 [33], we found that a mix of Cphy3367 and the GH5 Cphy1163 has higher activity on cellulose than either enzyme alone (Fig. 5D), supporting they have complementary roles in cellulolysis. CAZymes can also potentiate other enzymes that have no activity by themselves. For example, the xylanase Cphy2105 activates the putative xylosidases Cphy3009 and Cphy3207 on xylan (Fig. 5E). Similarly, Cphy1071 activates the putative mannosidase Cphy1719 on glucomannan (Fig. 5F). Activities of the GH28 Cphy2567 and Cphy3310 are enhanced by the carbohydrate esterase Cphy1687 (Fig. 5G), supporting this enzyme demethylesterifies homogalacturonan to facilitate its degradation. This carbohydrate esterase did not, however, increase cleavage by the PL9 enzymes that were the most active on homogalacturonan.

Global correlations between CAZyme mRNA expression and cleavage activities were weak for all polysaccharides (Fig. S12), mostly because many CAZyme genes are up-regulated on substrates upon which they have no activity. CAZymes up-regulated on multiple substrates (Fig. 3, cluster A) may act as 'carbon scouts' [34] that degrade complex substrates into inducing

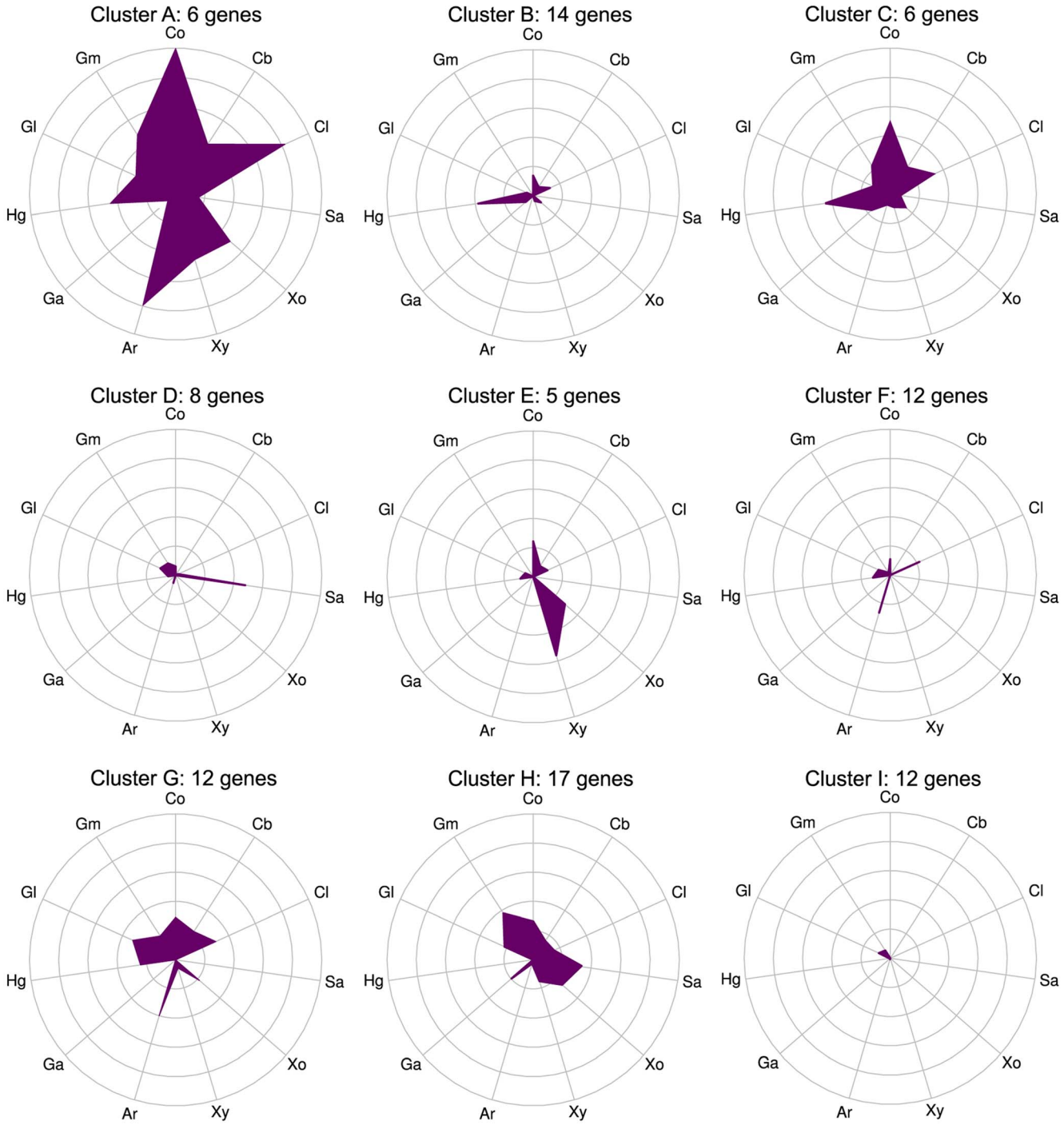

Figure 3. CAZymes clustered based on gene expression patterns (clusters A–I) show that some genes respond to multiple carbon sources while others are substrate-specific. mRNA expression changes (log_2 expression ratios relative to glucose) for all 92 CAZyme genes differentially expressed on at least 1 polysaccharide relative to glucose were separated into nine clusters using K-means. Plot centers are expression on glucose and concentric rings show log_2 up-regulation on the following carbon sources: cellobiose (Cb), filter paper cellulose (Cl), starch (Sa), xylose (Xo), xylan (Xy), arabinan (Ar), galacturonic acid (Ga), homogalacturonan (Hg), galactan (Gl), galactomannan (Gm), raw corn stover (Co). Gene membership of clusters is shown in Table S9.
doi:10.1371/journal.pgen.1004773.g003

molecules used to fine-tune the expression of hydrolytic enzymes. As described above, some CAZymes such as xylosidases (Fig. 5D) are inactive on intact xylan, but are potentiated by other xylanases. The GH18 Cphy1799 and Cphy1800 are the most highly-upregulated CAZymes on cellulose (Fig. S12), but are chitinases with no activity on cellulose or other plant substrates

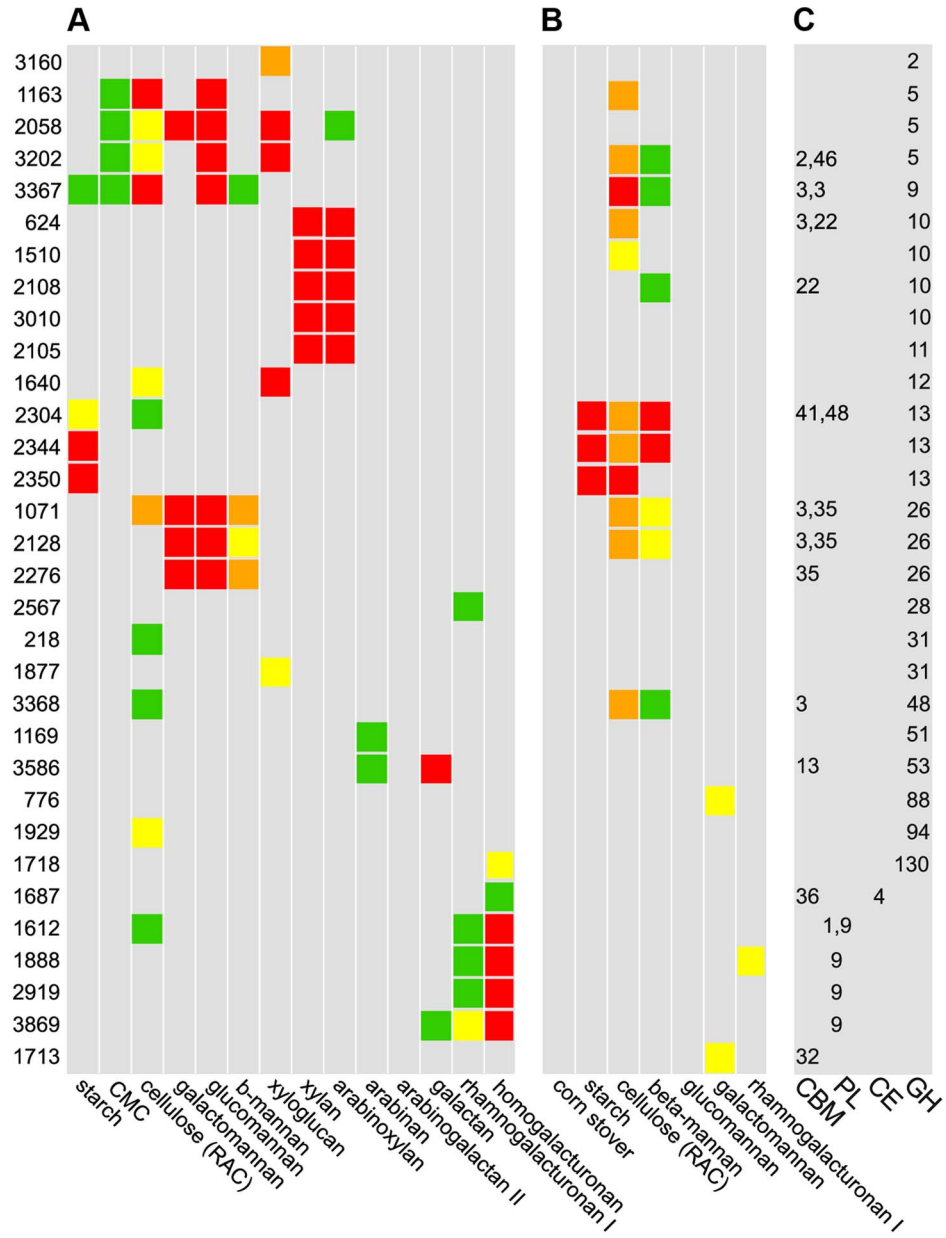

Figure 4. Cleavage A, binding B, and CAZy database classification C of purified enzymes. A Polysaccharide cleavage was quantified as nmol reducing sugar released per milligram enzyme per minute: >160 red, 80–160 orange, 40–80 yellow, 20–40 green, <20 gray. **B** Binding to insoluble polysaccharides was quantified as percentage enzyme bound to substrate: >30% red, 20–30% orange, 15–20% yellow, 10–15% green, < 10% gray. **C** CAZy database classifications: glycoside hydrolases (GH), carbohydrate esterases (CE), polysaccharide lyases (PL), and carbohydrate binding domains (CBM). Among 56 purified CAZymes, only the 32 enzymes for which activities were found are shown.
doi:10.1371/journal.pgen.1004773.g004

[26]. As such, the set of up-regulated CAZymes is useful to identify active enzymes, but strong up-regulation does not necessarily indicate activity on a given substrate.

Conclusions

We assimilated our results into a model of *C. phytofermentans* polysaccharide catabolism that shows degradation by active CAZymes and uses mRNA expression profiles to predict how these substrates are transported and metabolized (Fig. 6). Unlike other clostridia that transport sugars with numerous phosphotransferase systems (PTS) [35] [36], *C. phytofermentans* encodes a single, lowly expressed PTS and also lacks the symporters to transport xylose and arabinose [37]. Instead, *C. phytofermentans* responds to carbon sources by up-regulating between two (galacturonic acid) and twenty-two (arabinan) ABC transporters (Fig. 6). Expression changes support that oligosaccharides and monosaccharides are uptaken by distinct transporters. For example, different ABC transporters are up-regulated on xylose and xylan. Similarly, different transporters respond to glucose, cellobiose, and cellulose. Intracellular cellodextrins are cleaved by at least one cellodextin phosphorylase (GH94); hexoses are phosphorylated, likely by a ROK hexokinase (Cphy0329) and a putative galactokinase (Cphy2237), and fed into glycolysis. While hexokinases may have wide substrate activity [38], poor growth on

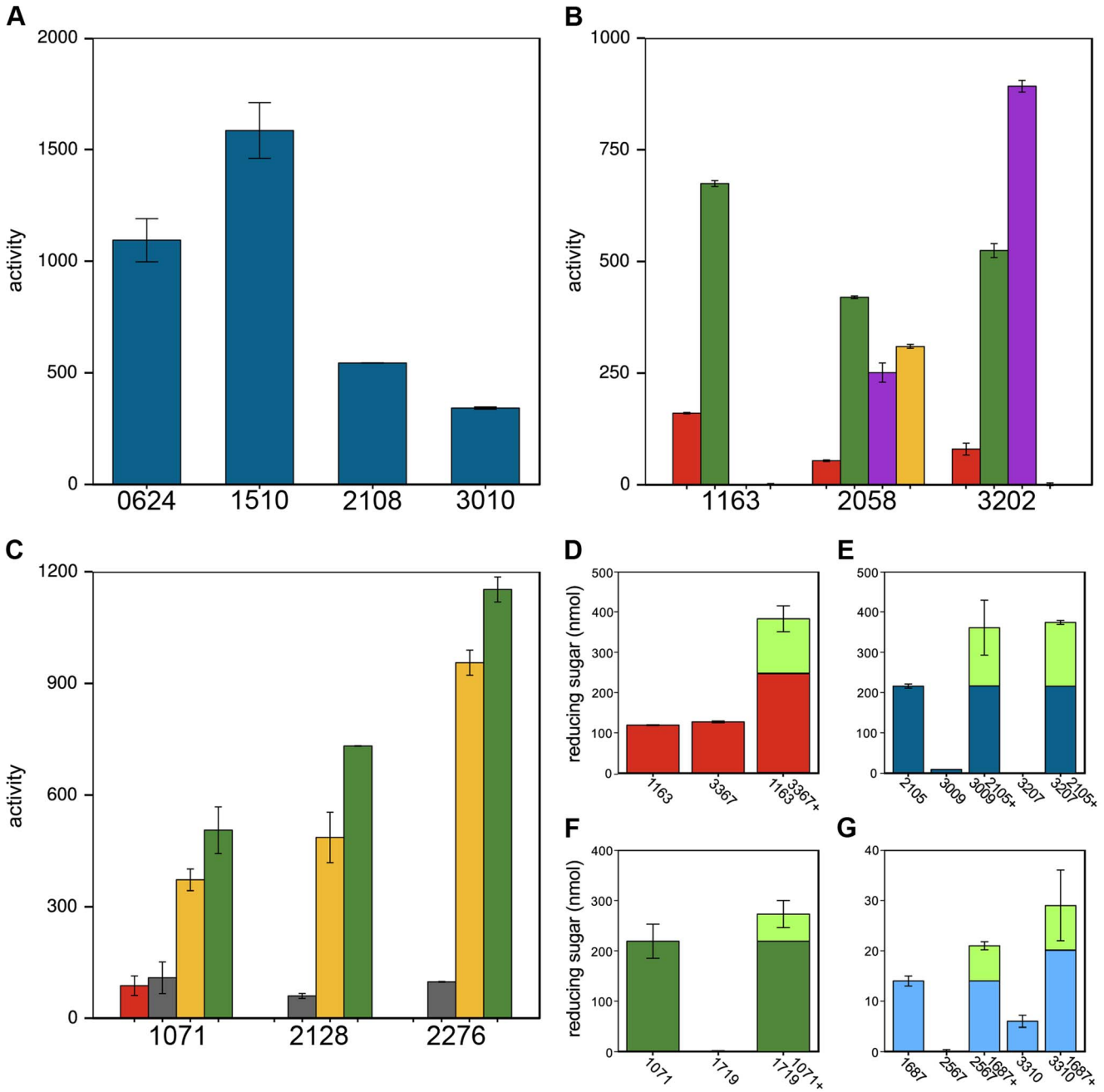

Figure 5. Members of the same CAZy family vary in polysaccharide cleavage activities and CAZymes can by potentiated by other enzymes. A Variation in cleavage activities of GH10 enzymes on xylan. **B** GH5 and **C** GH26 family members differ in their activities and substrate specificities on amorphous cellulose (red), glucomannan (green), xyloglucan (violet), galactomannan (yellow), mannan (gray). Enzyme activities in **A**–**C** are nmol reducing sugar released per milligram enzyme per minute. **D**–**G** CAZyme mixtures have higher activities than the individual enzymes. **D** Cphy1163 and Cphy3367 alone and together on amorphous cellulose. **E** Cphy2105, Cphy3009, and Cphy3207 alone and the latter two enzymes plus Cphy2105 on xylan. **F** Cphy1719 and Cphy1071 alone and together on glucomannan. **G** Cphy1687, Cphy2567, and Cphy3310 alone and the latter two enzymes plus Cphy1687 on homogalacturonan. In **D**–**G**, enzyme activities are shown as reducing sugar (nmol) produced by individual and combined enzymes. The fraction of the reducing sugar produced by the mixed enzymes that exceeds the sum of the individual enzymes is shown in green.
doi:10.1371/journal.pgen.1004773.g005

mannose could be due to inefficient mannose phosphorylation. The pentoses xylose and arabinose are isomerized and metabolized by the pentose phosphate pathway (PPP). Weak growth on arabinose could be due to inefficient transport or the lack of the phosphoketolase in the PPP enabling rapid L-arabinose metabolism by *C. acetobutylicum* [39].

Plant degrading microbes differ widely in their abilities to depolymerize and metabolize polysaccharides, likely reflecting niche differentiation to alleviate resource competition. Among soil clostridia, *C. thermocellum* ferments cellulose, but not xylan [40]. *C. cellulolyticum* grows faster on xylose than xylan and faster on cellobiose than glucose [41], both of which differ from *C.*

phytofermentans. Similar specialization exists in the human gut microbiome where microbes catabolize different glycans in dietary fiber [20]. The strategy presented here of high-resolution anaerobic growth measurements, RNA sequencing, and CAZyme assays complements other methods such as proteomics [42] and metagenomics. Elucidating how microbes metabolize polysaccharides is key to understanding the function of plant-degrading microbial communities and to develop improved enzyme mixtures and recombinant microbes for industrial processing of plant biomass.

Materials and Methods

Growth measurements

C. phytofermentans was cultured anaerobically at 30°C in GS2 medium [43]. Growth was quantified in medium containing 3 g l^{-1} mono-, di-, or polysaccharides (Table S3 for product numbers) in 100-well microtiter plates (Bioscreen 9502550) that were sealed by press-fitting adhesive sheets (Qiagen 1018104) under the lids. Growth was measured every 15 minutes as OD_{600} at 30°C using a Thermo Scientific Bioscreen C with 30 seconds shaking before each reading. Growth was not measured for ß-mannan and amorphous cellulose cultures because these low solubility substrates occluded optical density measurements. Growth on insoluble substrates (15 g l^{-1} of 0.5×5 cm strips of filter paper cellulose or raw corn stover) was measured in 10 ml cultures as substrate consumption by collection on 11 µm filters by vacuum filtration, drying overnight at 65°C, and weighing. Growth on solid GS2 medium with 3 g l^{-1} polysaccharide and 15 g l^{-1} agar was tested by incubating plates anaerobically for 10 days at 30°C (Fig. S2). Regenerated amorphous cellulose (RAC) for GS2 agar plates and enzyme assays was prepared from Avicel PH-101 by phosphoric acid treatment [44].

RNA-seq

Cultures containing soluble substrates were sampled for RNA-seq in mid-log phase. Filter paper cellulose cultures were sampled under the same conditions as in [45]; corn stover cultures were sampled after 3 days. Cells were arrested with RNA stabilization buffer, collected by centrifugation (5 kg, 5 min, 4°C), and RNA was isolated with an Ambion Ribopure Bacteria Kit (AM1925). Twenty µg total RNA was treated with Turbo DNase (Invitrogen AM1907), phenol-chloroform extracted, ethanol precipitated, and resuspended in 15 µl 10 mM Tris-Cl, pH 8.5. Ribosomal RNA was depleted from 10 µg RNA using a MICROBExpress kit (Invitrogen AM1905), giving a typical yield of 1–2 µg RNA. 500 ng of rRNA-depleted RNA was fragmented with magnesium buffer (NEB E6101) for 2.5 minutes at 94°C, ethanol precipitated, and resuspended in 10 µl water. RNA was resolved on an Agilent Bioanalyzer 2100 to confirm it was 200–800 bp.

Single stranded cDNA was made from 500 ng fragmented RNA by Superscript III first strand synthesis (Invitrogen 18080-051) including 200 ng/µl actinomycin D (Invitrogen 11805017) and 120 ng/µl random hexamers (Invitrogen 48190-011). Single stranded DNA was twice phenol-chloroform extracted, ethanol precipitated, and resuspended in 52 µl water. The second cDNA strand was synthesized by the Invitrogen double stranded cDNA

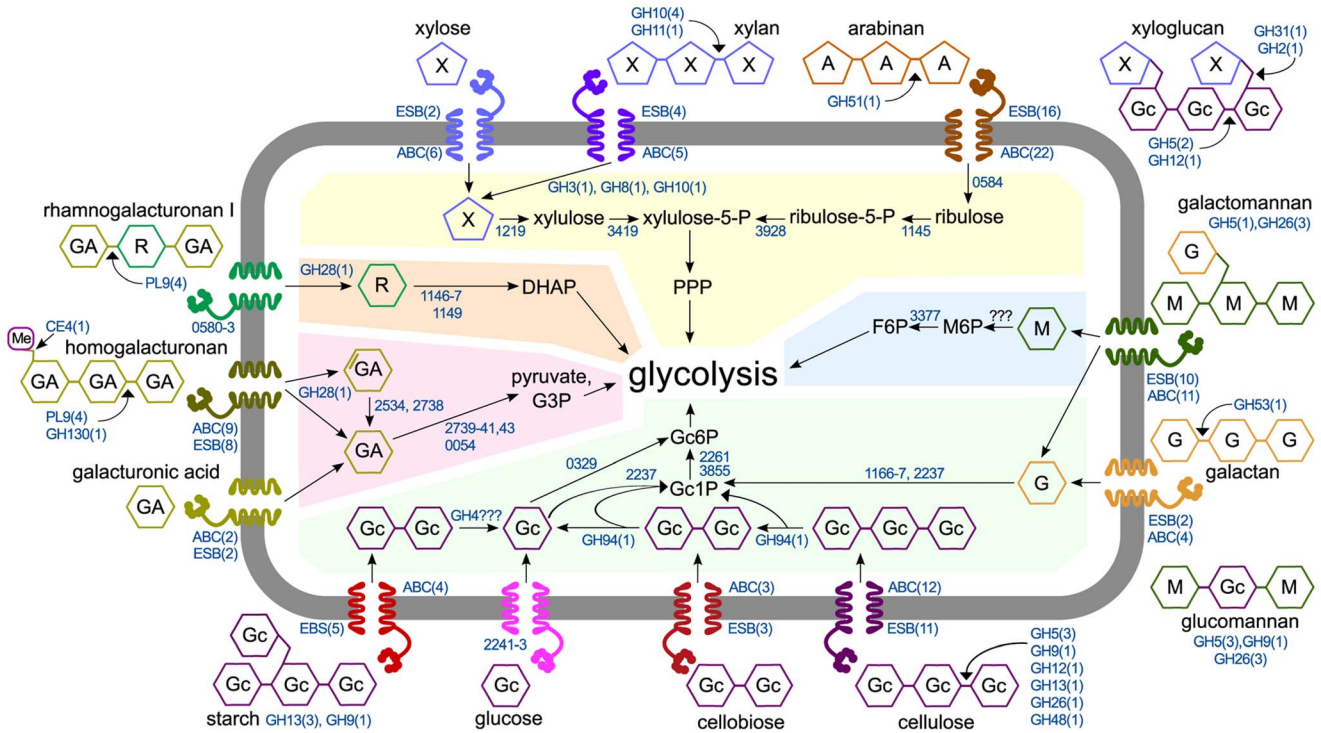

Figure 6. Model of polysaccharide degradation and metabolism by *C. phytofermentans*. CAZymes (shown as the number of enzymes in CAZy families) are based on purified activities and are intra- or extracellular based on putative secretion signals. Metabolic enzymes are shown as NCBI numbers and are proposed based on mRNA expression. Rhamnose transport and assimilation is based on pathway from [55]. Abbreviations are D-galacturonic acid (GA), L-rhamnose (R), D-mannose (M), D-glucose (Gc), D-galactose (G), D-xylose (X), L-arabinose (A), fructose (F), phosphate (P), pentose phosphate pathway (PPP), dihydroxyacetone-phosphate (DHAP), glyceraldehyde-3-phosphate (G3P). For each substrate, the number of significantly up-regulated extracellular solute binding proteins (ESB) and ABC transporters (ABC) are shown. Shaded regions show metabolism of glucose (green), mannose (blue), xylose and arabinose (yellow), rhamnose (orange), and galacturonic acid (red).
doi:10.1371/journal.pgen.1004773.g006

synthesis kit (Invitrogen 11917-010) using 250 μM dNTP (dATP, dCTP, dGTP, dUTP). Double stranded cDNA was purified using 1.8 volumes of Solid Phase Reversible Immobilisation (SPRI) beads [46] prepared from carboxylate-modified microparticles [47] (ThermoScientific 6515-2105-050250) and resuspended in 50 μl 10 mM Tris-Cl, pH 8.5. DNA was resolved on an Agilent Bioanalyzer to confirm cDNA was 100–500 bp.

Sequencing libraries were prepared from ~250 ng cDNA using the Illumina Truseq DNA Kit (Illumina FC-121-2001). DNA was purified with 1.8 volumes of SPRI beads after end repair. Following adapter ligation, 200–500 bp DNA was selected by Pippin Prep gel extraction (Sage Science). The second cDNA strand (contains dUTP) was then degraded by incubating with 1 unit USER enzyme (NEB NEB M5505S) at 37°C for 15 minutes, followed by 5 minutes at 95°C. Single stranded DNA was PCR amplified and size selected with 1.8 volumes of SPRI beads. Multiplexed libraries were normalized to 2 ng μl^{-1}, pooled, and resolved on an Agilent Bioanalyzer to confirm the peak was at 200–500 bp. Libraries were sequenced on a Illumina HiSeq2000 sequencer. RNA-seq reads were mapped to the *C. phytofermentans* genome (NCBI NC_010001.1) using Bowtie [48] to report the single, best alignment (see Table S5 for parameters). Gene expression was measured as RPKM (sum of nucleotides in gene per million mapped nucleotides per gene kb) using RSEQtools [49] (Table S6), excluding rRNA reads from RPKM calculations. Differentially expressed genes (Tables S7, S8) were identified by DESeq [25] using the statistical thresholds defined in [50] of >4-fold expression and a p-value$<10^{-5}$, which corresponds to a p-value<0.05 adjusted for multiple testing of the 3902 genes in *C. phytofermentans* genome by Bonferroni correction.

Enzyme purification and activities

A set of 56 CAZymes (Table S2) that were up-regulated on plant polysaccharides were His-tagged, overexpressed, and purified. Primers (Table S10) were designed to clone CAZyme genes by Ligation-Independent Cloning [51] into pET-22B(+). Genes were cloned with C-terminal His-tags except for *cphy1687* and *cphy2105*, which were cloned with N-terminal tags to improve yields. The 26 enzymes predicted to be secreted by SignalP [52] were cloned as their mature forms, omitting the N-terminal secretion signals. Forward and reverse sequences of the 56 genes cloned in pET-22B(+) were confirmed by sequencing (Fig. S9). Plasmids were transformed into *E. coli* BL21(DE3) (Novagen 70235), grown in 50 ml TB medium to OD_{600} = 1.0, and induced by adding 500 μM IPTG and incubating overnight at 20°C. Cells were pelleted, resuspended in lysis buffer: 50 mM phosphate buffer pH = 8, 0.5M NaCl, 10 mM imidazole, 15% glycerol, 1 mM pefabloc (Sigma 76307). Cells were lysed by sonication (Cole-Parmer Vibracell CV33) in the presence of lysozyme (Novagen 71230). His-tagged proteins were purified from 50 ml culture on Ni-NTA spin columns (Qiagen 31014) and quantified by Bradford assay, giving an average yield of 3 mg protein. Purified proteins were visualized on 12% SDS-PAGE gels (Novex 12% Bis-Tris Gel NP0342BOX) (Fig. S10).

Enzymatic polysaccharide cleavage was quantified by 3,5-dinitrosalicylic acid (DNS) assay [53] in 96 well clear, flat bottom plates (Molecular Devices X6011). Fifty μg enzyme was incubated with 0.25% polysaccharide in 25 mM potassium phosphate buffer pH = 7 for 30 minutes at 37°C, 1 volume DNS reagent was added, incubated at 95°C for 15 minutes, and read at OD_{600}. Enzyme activities (nmol reducing sugar per mg enzyme per min) were calculated from DNS readings by subtracting blanks (+polysaccharide,-enzyme) and calculating sugar produced using standard curves (Fig. S11). Polysaccharide cleavage by enzyme pairs was tested on amorphous cellulose, xylan, glucomannan, and homogalacturonan as the reducing sugar (mM) produced by 25 μg of each enzyme alone and combined with a second enzyme (25 μg Cphy3367 for cellulose, 5 μg Cphy2105 for xylan, 5 μg Cphy1071 for glucomannan, 25 μg Cphy1687 for homogalacturonan). Reducing sugar was measured by DNS assay after 30 minutes incubation as described above. Enzyme binding to insoluble polysaccharides was quantified using a method adapted from [54]. Fifty μg enzyme was suspended in 200 μl 0.1 M phosphate buffer pH = 7 with 6 mg polysaccharide and incubated with rotation for 5 h at 4°C. The polysaccharide was collected by centrifuging at 12 kg for 5 min and the enzyme concentration in the supernatant was calculated before and after incubation by Bradford assay, yielding the percentage of enzyme bound to the polysaccharide.

Supporting Information

Figure S1 *C. phytofermentans* growth curves on 3 g l^{-1} di- and monosaccharides: A D-cellobiose, B D-glucose, C D-galactose, D D-xylose, E D-mannose, F D-galacturonic acid, G D-arabinose, H L-rhamnose. Blue curve is mean density (OD_{600}) of 6 cultures; red curves show one standard deviation.
(PNG)

Figure S2 *C. phytofermentans* growth on solid GS2 medium containing 3 g l^{-1} polysaccharides. Plates were incubated anaerobically at 30°C for 10 days. Colonies were observed on all substrates except arabinogalactan II.
(PNG)

Figure S3 *C. phytofermentans* growth on 15 g l^{-1} A raw corn stover and B filter paper cellulose as a sole carbon source in GS2 medium. Growth was measured as consumption of insoluble substrate. Data are means of triplicate cultures; error bars are one standard deviation. Gray bars show when samples were taken for RNA-seq. The cellulose sample for RNA-seq was taken under the same conditions as those for proteomics in [45].
(PNG)

Figure S4 *C. phytofermentans* growth on mixtures of galactose and mannose (3 g l^{-1} total for all treatments): A galactomannan (80% mannose, 20% galactose), B D-galactose, C D-mannose, D 80% D-mannose and 20% D-galactose, E 50% D-mannose and 50% D-galactose, F 20% D-mannose and 80% D-galactose. Blue line shows mean OD_{600} of 6 cultures, red lines show range of one standard deviation. Growth on individual sugars shows that *C.phytofermentans* grows faster and to higher density on D-galactose than D-mannose. Growth is diauxic on galactomannan and sugar mixtures A, D, E, and F supporting that galactose is metabolized preferentially before mannose.
(PNG)

Figure S5 *C. phytofermentans* growth on mixtures of glucose and xylose (3 g l^{-1} total for all treatments): A xyloglucan (45% glucose, 35% xylose), B D-glucose, C D-xylose, D 75% D-glucose and 25% D-xylose, E 50% D-glucose and 50% D-xylose, F 25% D-glucose and 75% D-xylose. Blue line shows mean OD_{600} of 6 cultures, red lines show range of one standard deviation. Growth on individual sugars shows that *C.phytofermentans* grows faster and to higher density on D-glucose than on D-xylose. Growth is diauxic on xyloglucan and mixtures of glucose and xylose A, D, E, F supporting that D-glucose is metabolized preferentially before D-xylose.
(PNG)

Figure S6 *C. phytofermentans* growth on mixtures of xylan and xylose (3 g l^{-1} total): A xylan, B D-xylose, C 25% D-xylose and

75% xylan, D 50% D-xylose and 50% xylan, E 75% D-xylose and 25% xylan. Black curve is the mean OD_{600} of 6 cultures; the red curves show the range of 1 standard deviation. Growth curves show that xylan is metabolized before its constituent monomer, D-xylose.
(PNG)

Figure S7 Expression values (log_2(RPKM)) for RNA sequencing of replicate cultures of all carbon source treatments are highly correlated. A–H are single-end reads and I–M are paired-end reads. Spearman correlation coefficients are shown on each panel.
(PNG)

Figure S8 Reads from dUTP sequencing are highly strand-specific with an average of 33,715 times more reads mapping to the expected strand for each of the 8 rRNA operons (16S-5S-23S) in the glucose samples. Operons 1–3 are transcribed in the positive direction; Operons 4–8 are in the reverse direction. Reads mapping to the expected strand are in magenta and the opposing strand are green. Note, dUTP sequencing reads map to the opposite strand. Thus, genes transcribed in the positive orientation are sequenced with reads on the reverse strand.
(PNG)

Figure S9 Forward and reverse sequence alignments of 56 CAZyme genes cloned into pET-22B(+) confirm that genes had the correct sequences.
(PDF)

Figure S10 Purified CAZyme visualized on 12% SDS-PAGE gels (Nupage bis-Tris novex gel IM-8042). Masses (kDa) of full proteins are shown for each CAZyme; masses with asterisks are secreted proteins for which the N-terminal secretion signal was not cloned, resulting in an expected size slightly smaller than the full protein mass shown. Mass ladders with blue borders are from images that were cropped by omitting intermediate lanes.
(PNG)

Figure S11 Sugar standard curves from DNS assays used to convert polysaccharide DNS readings to sugar concentrations.
(PNG)

Figure S12 Comparison of mRNA expression versus enzyme activity for purified CAZymes on A homogalacturonan, B galactan, C arabinan, D xylan, E galactomannan, F cellulose, G starch. mRNA expression is expressed as log_2(RPKM) on the polysaccharide relative to glucose. Enzyme activity is log_2(nmol sugar per mg enzyme per minute).
(PNG)

Table S1 Plant polysaccharides used in this study: natural abundances, plant source of tested compound, chemical structure, and sugar composition. References: [56] [57] [58] [59] [60] [61] [62] [63].
(XLS)

Table S2 *C. phytofermentans* CAZymes in the CAZy database [15] including gene name, NCBI accession, gene (bp) and protein (amino acid) length, predicted protein molecular weight (kDa), N-terminal secretion signal cleavage position (SignalP [52] Y-max position), CAZy classification, and annotation. The 56 CAZymes purified in this study are shown in blue.
(XLS)

Table S3 Vendor product information for mono-, di-, and polysaccharides used in this study.
(XLS)

Table S4 *C. phytofermentans* growth rates (generation time in hours) and cell yields (maximum OD_{600}) growing on polysaccharides and di-/monosaccharides. Growth rate on each substrate was calculated by log-transforming the mean growth curve (Fig. 1,S1) and calculating the slop of a linear regression fit to the portion of the curve describing exponential growth. Regressions were fit to each of the two growth phases on galactomannan and xyloglucan.
(XLS)

Table S5 Mapping of RNA-seq reads to the *C. phytofermentans* genome, excluding reads mapping to rRNA operons. Reads were mapped using Bowtie [48] to report the single, best alignment between RNA-seq reads in.fastq files and the *C. phytofermentans* NCBI genome file (NC_010001.gbk). Reads were mapped using the following command: bowtie –best -k 1 [bowtie database of NC_010001.fna] –un unmappedReads.txt -q [.fastq file] mappedReads.txt.
(XLS)

Table S6 mRNA expression (RPKM) of all *C. phytofermentans* genes (page 1) and of only CAZymes (page 2) during steady state growth on sugars and polysaccharides. RPKM is defined as the sum of nucleotides from the reads that overlap with a given annotation entry normalized per million mapped nucleotides and the length of the annotation item. RPKM were calculated using RSEQtools [49] from mapped reads using MRFquant files. rRNA reads were removed from mrf files before running mrfQuant so that expression levels are normalized to mRNA reads. Samples are labeled whether they were sequenced using single-end (S) or paired-end (P) reads. The 56 CAZymes purified in this study are shown in blue on page 2.
(XLS)

Table S7 Differential expression of all *C. phytofermentans* genes on nine polysaccharides relative to glucose. Differentially expressed genes (DEseq p-value$<10^{-5}$ and a >4-fold differential expression) are shown in blue. Sheets show differential expression relative to glucose of the following carbon sources: homogalacturonan, galactan, arabinan, xylan, galactomannan, cellulose, cellobiose, starch, stover. Paired-end sequenced polysaccharide samples are compared to the paired-end glucose sample; Single-end sequenced polysaccharide samples are compared to the single-end glucose sample (see Table S6).
(XLS)

Table S8 Differentially expressed CAZyme genes on nine polysaccharides relative to glucose. Page 1 shows all 92 differentially expressed CAZymes and the polysaccharides upon which they were up-regulated. Subsequent pages show the expression values of CAZymes up-regulated on specific polysaccharides. Purified CAZymes are shown in blue. Proportion of CAZyme genes with significant mRNA expression changes that were purified: homogalacturonan (25/27), galactan (8/13), arabinan (37/54), xylan (19/19), galactomannan (28/45), cellobiose (13/15), cellulose (26/27), starch (14/32), stover (33/40).
(XLS)

Table S9 Membership of CAZyme genes in K-means clusters. The 92 differentially-expressed CAZymes were clustered based on their expression (log_2 expression ratios relative to glucose) as shown in Fig. 3.
(XLS)

Table S10 Primers for LIC cloning of 56 CAZymes. LIC sequences that overlap with the pET-22B(+) cloning plasmid are shown in red. His-tag sequences are in green. Genes of putatively secreted proteins were cloned without their predicted N-terminal secretion signals (Table S1). Genes were cloned with C-terminal

His-tags, except for Cphy1687 and Cphy2105, which were re-cloned with N-terminal His-tags to improve yield.
(XLS)

Table S11 Polysaccharide cleavage activities (nmol reducing sugar per mg enzyme per minute) of 56 purified CAZymes on 14 polysaccharides. Enzyme (50 μg) was incubated with 0.25% polysaccharide in 25 mM potassium phosphate buffer pH = 7 for 30 minutes at 37°C. Activities: >160 red, >80 orange, >40 yellow, >20 green.
(XLS)

References

1. Shekhar C (2011) Future fuel: could biomass be the new petroleum? Chem Biol 18: 1199–1200. doi:10.1016/j.chembiol.2011.10.010.
2. Schwarz WH (2001) The cellulosome and cellulose degradation by anaerobic bacteria. Appl Microbiol Biotechnol 56: 634–649. doi:10.1007/s002530100710.
3. Edwards JL, Smith DL, Connolly J, McDonald JE, Cox MJ, et al. (2010) Identification of carbohydrate metabolism genes in the metagenome of a marine biofilm community shown to be dominated by gammaproteobacteria and bacteroidetes. Genes 1: 371–384. doi:10.3390/genes1030371.
4. McNeil NI (1984) The contribution of the large intestine to energy supplies in man. Am J Clin Nutr 39: 338–342.
5. Pauly M, Keegstra K (2008) Cell-wall carbohydrates and their modification as a resource for biofuels. Plant J Cell Mol Biol 54: 559–568. doi:10.1111/j.1365-313X.2008.03463.x.
6. El Kaoutari A, Armougom F, Gordon JI, Raoult D, Henrissat B (2013) The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. Nat Rev Microbiol 11: 497–504. doi:10.1038/nrmicro3050.
7. Deutschmann R, Dekker RFH (2012) From plant biomass to bio-based chemicals: latest developments in xylan research. Biotechnol Adv 30: 1627–1640. doi:10.1016/j.biotechadv.2012.07.001.
8. Bauer WD, Talmadge KW, Keegstra K, Albersheim P (1973) The Structure of Plant Cell Walls: II. The Hemicellulose of the Walls of Suspension-cultured Sycamore Cells. Plant Physiol 51: 174–187. doi:http://dx.doi.org/10.1104/pp.51.1.174.
9. Burke D, Kaufman P, McNeil M, Albersheim P (1974) The Structure of Plant Cell Walls: VI. A Survey of the Walls of Suspension-cultured Monocots. Plant Physiol 54: 109–115. doi:http://dx.doi.org/10.1104/pp.54.1.109.
10. Capek P, Alföldi J, Lisková D (2002) An acetylated galactoglucomannan from *Picea abies* L. Karst. Carbohydr Res 337: 1033–1037. doi:10.1016/S0008-6215(02)00090-3.
11. Buckeridge MS (2010) Seed cell wall storage polysaccharides: models to understand cell wall biosynthesis and degradation. Plant Physiol 154: 1017–1023. doi:10.1104/pp.110.158642.
12. Caffall KH, Mohnen D (2009) The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. Carbohydr Res 344: 1879–1900. doi:10.1016/j.carres.2009.05.021.
13. Ridley BL, O'Neill MA, Mohnen D (2001) Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. Phytochemistry 57: 929–967. doi:10.1016/S0031-9422(01)00113-3.
14. Mohnen D (2008) Pectin structure and biosynthesis. Curr Opin Plant Biol 11: 266–277. doi:10.1016/j.pbi.2008.03.006.
15. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42: D490–495. doi:10.1093/nar/gkt1178.
16. Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS (2002) Microbial cellulose utilization: fundamentals and biotechnology. Microbiol Mol Biol Rev MMBR 66: 506–577, table of contents. doi:10.1128/MMBR.66.3.506-577.2002.
17. Tolonen AC, Petit E, Blanchard JL, Warnick T, Leschine SB (2013) Technologies to Study Plant Biomass Fermentation Using the Model Bacterium *Clostridium phytofermentans*. In: J. Sun, S.Y. Ding, J.D. Peterson (Eds.), Biological conversion of biomass for fuels and chemicals, Royal Society of Chemistry, Cambridge, pp. 114–139. doi:10.1039/9781849734738-00114
18. Warnick TA, Methé BA, Leschine SB (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.
19. Hatfield RD, Weimer PJ (1995) Degradation characteristics of isolated and in situ cell wall lucerne pectic polysaccharides by mixed ruminal microbes. J Sci Food Agric 69: 185–196. doi:10.1002/jsfa.2740690208.
20. Martens EC, Lowe EC, Chiang H, Pudlo NA, Wu M, et al. (2011) Recognition and Degradation of Plant Cell Wall Polysaccharides by Two Human Gut Symbionts. PLoS Biol 9: e1001221. doi:10.1371/journal.pbio.1001221.
21. Muir M, Williams L, Ferenci T (1985) Influence of transport energization on the growth yield of *Escherichia coli*. J Bacteriol 163: 1237–1242.
22. Zhang Y-HP, Lynd LR (2005) Cellulose utilization by *Clostridium thermocellum*: bioenergetics and hydrolysis product assimilation. Proc Natl Acad Sci U S A 102: 7321–7325. doi:10.1073/pnas.0408734102.
23. Nihira T, Nakai H, Chiku K, Kitaoka M (2012) Discovery of nigerose phosphorylase from *Clostridium phytofermentans*. Appl Microbiol Biotechnol 93: 1513–1522. doi:10.1007/s00253-011-3515-9.
24. Nakajima M, Nishimoto M, Kitaoka M (2009) Characterization of three beta-galactoside phosphorylases from *Clostridium phytofermentans*: discovery of d-galactosyl-beta1->4-l-rhamnose phosphorylase. J Biol Chem 284: 19220–19227. doi:10.1074/jbc.M109.007666.
25. Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11: R106. doi:10.1186/gb-2010-11-10-r106.
26. Tolonen AC, Cerisy T, El-Sayyed H, Boutard M, Salanoubat M, et al. (2014) Fungal lysis by a soil bacterium fermenting cellulose. Environ Microbiol. doi:10.1111/1462-2920.12495.
27. Tolonen AC, Chilaka AC, Church GM (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.
28. Zhang X-Z, Sathitsuksanoh N, Zhang Y-HP (2010) Glycoside hydrolase family 9 processive endoglucanase from *Clostridium phytofermentans*: heterologous expression, characterization, and synergy with family 48 cellobiohydrolase. Bioresour Technol 101: 5534–5538. doi:10.1016/j.biortech.2010.01.152.
29. Zhang X-Z, Zhang Z, Zhu Z, Sathitsuksanoh N, Yang Y, et al. (2010) The noncellulosomal family 48 cellobiohydrolase from *Clostridium phytofermentans* ISDg: heterologous expression, characterization, and processivity. Appl Microbiol Biotechnol 86: 525–533. doi:10.1007/s00253-009-2231-1.
30. Knox JP (2008) Revealing the structural and functional diversity of plant cell walls. Curr Opin Plant Biol 11: 308–313. doi:10.1016/j.pbi.2008.03.001.
31. Abbott DW, Hrynuik S, Boraston AB (2007) Identification and characterization of a novel periplasmic polygalacturonic acid binding protein from *Yersinia enterolitica*. J Mol Biol 367: 1023–1033. doi:10.1016/j.jmb.2007.01.030.
32. Aspeborg H, Coutinho PM, Wang Y, Brumer H, Henrissat B (2012) Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). BMC Evol Biol 12: 186. doi:10.1186/1471-2148-12-186.
33. Liao H, Zhang X-Z, Rollin JA, Zhang Y-HP (2011) A minimal set of bacterial cellulases for consolidated bioprocessing of lignocellulose. Biotechnol J 6: 1409–1418. doi:10.1002/biot.201100157.
34. Benz JP, Chau BH, Zheng D, Bauer S, Glass NL, et al. (2014) A comparative systems analysis of polysaccharide-elicited responses in *Neurospora crassa* reveals carbon source-specific cellular adaptations. Mol Microbiol 91: 275–299. doi:10.1111/mmi.12459.
35. Nölling J, Breton G, Omelchenko MV, Makarova KS, Zeng Q, et al. (2001) Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. J Bacteriol 183: 4823–4838. doi:10.1128/JB.183.16.4823-4838.2001.
36. Al Makishah NH, Mitchell WJ (2013) Dual substrate specificity of an N-acetylglucosamine phosphotransferase system in *Clostridium beijerinckii*. Appl Environ Microbiol 79: 6712–6718. doi:10.1128/AEM.01866-13.
37. Servinsky MD, Kiel JT, Dupuy NF, Sund CJ (2010) Transcriptional analysis of differential carbohydrate utilization by *Clostridium acetobutylicum*. Microbiol Read Engl 156: 3478–3491. doi:10.1099/mic.0.037085-0.
38. Chen M, Chen L, Zou Y, Xue M, Liang M, et al. (2011) Wide sugar substrate specificity of galactokinase from *Streptococcus pneumoniae* TIGR4. Carbohydr Res 346: 2421–2425. doi:10.1016/j.carres.2011.08.014.
39. Servinsky MD, Germane KL, Liu S, Kiel JT, Clark AM, et al. (2012) Arabinose is metabolized via a phosphoketolase pathway in *Clostridium acetobutylicum* ATCC 824. J Ind Microbiol Biotechnol 39: 1859–1867. doi:10.1007/s10295-012-1186-x.
40. Ng TK, Ben-Bassat A, Zeikus JG (1981) Ethanol Production by Thermophilic Bacteria: Fermentation of Cellulosic Substrates by Cocultures of *Clostridium thermocellum* and *Clostridium thermohydrosulfuricum*. Appl Environ Microbiol 41: 1337–1343.
41. Xu C, Huang R, Teng L, Wang D, Hemme CL, et al. (2013) Structure and regulation of the cellulose degradome in *Clostridium cellulolyticum*. Biotechnol Biofuels 6: 73. doi:10.1186/1754-6834-6-73.
42. Tolonen AC, Haas W (2014) Quantitative Proteomics Using Reductive Dimethylation for Stable Isotope Labeling. J Vis Exp (89), e51416, doi:10.3791/51416.

Acknowledgments

We are grateful to the following people for their technical assistance: S Leschine, J Blanchard and T Warnick for microbiology, J Levin and K Labadie for RNA sequencing, and S Gas and CEH Tolonen for informatics. We thank the Biopolymers Facility at Harvard Medical School and the Genoscope-CEA for RNA sequencing.

Author Contributions

Conceived and designed the experiments: ACT JW MS. Performed the experiments: MB TC PYN AA ACT. Analyzed the data: MB TC PYN AA ACT. Wrote the paper: ACT.

43. Johnson EA, Madia A, Demain AL (1981) Chemically Defined Minimal Medium for Growth of the Anaerobic Cellulolytic Thermophile *Clostridium thermocellum*. Appl Environ Microbiol 41: 1060–1062.
44. Hong J, Ye X, Wang Y, Zhang Y-HP (2008) Bioseparation of recombinant cellulose-binding module-proteins by affinity adsorption on an ultra-high-capacity cellulosic adsorbent. Anal Chim Acta 621: 193–199. doi:10.1016/j.aca.2008.05.041.
45. Tolonen AC, Haas W, Chilaka AC, Aach J, Gygi SP, et al. (2011) Proteome-wide systems analysis of a cellulosic biofuel-producing microbe. Mol Syst Biol 7: 461. doi:10.1038/msb.2010.116.
46. DeAngelis MM, Wang DG, Hawkins TL (1995) Solid-phase reversible immobilization for the isolation of PCR products. Nucleic Acids Res 23: 4742–4743. doi:10.1093/nar/23.22.4742.
47. Rohland N, Reich D (2012) Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Res 22: 939–946. doi:10.1101/gr.128124.111.
48. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10: R25. doi:10.1186/gb-2009-10-3-r25.
49. Habegger L, Sboner A, Gianoulis TA, Rozowsky J, Agarwal A, et al. (2011) RSEQtools: a modular framework to analyze RNA-Seq data using compact, anonymized data summaries. Bioinforma Oxf Engl 27: 281–283. doi:10.1093/bioinformatics/btq643.
50. Mandlik A, Livny J, Robins WP, Ritchie JM, Mekalanos JJ, et al. (2011) RNA-Seq-based monitoring of infection-linked changes in *Vibrio cholerae* gene expression. Cell Host Microbe 10: 165–174. doi:10.1016/j.chom.2011.07.007.
51. Aslanidis C, de Jong PJ (1990) Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res 18: 6069–6074. http://dx.doi.org/10.1093/nar/18.20.6069.
52. Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8: 785–786. doi:10.1038/nmeth.1701.
53. Miller GL (1959) Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. Anal Chem 31: 426–428. doi:10.1021/ac60147a030.
54. Watanabe T, Ito Y, Yamada T, Hashimoto M, Sekine S, et al. (1994) The roles of the C-terminal domain and type III domains of chitinase A1 from *Bacillus circulans* WL-12 in chitin degradation. J Bacteriol 176: 4465–4472.
55. Petit E, LaTouf WG, Coppi MV, Warnick TA, Currie D, 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.
56. Zablackis E, Huang J, Müller B, Darvill AG, Albersheim P (1995) Characterization of the cell-wall polysaccharides of *Arabidopsis thaliana* leaves. Plant Physiol 107: 1129–1138. doi:http://dx.doi.org/10.1104/pp.107.4.1129.
57. Prasad S, Singh A, Joshi HC (2007) Ethanol as an alternative fuel from agricultural, industrial and urban residues. Resour Conserv Recycl 50: 1–39. doi:10.1016/j.resconrec.2006.05.007.
58. Chow PS, Landhäusser SM (2004) A method for routine measurements of total sugar and starch content in woody plant tissues. Tree Physiol 24: 1129–1136. doi:10.1093/treephys/24.10.1129.
59. Scheller HV, Ulvskov P (2010) Hemicelluloses. Annu Rev Plant Biol 61: 263–289. doi:10.1146/annurev-arplant-042809-112315.
60. Dahal P, Nevins DJ, Bradford KJ (1997) Relationship of Endo-[beta]-D-Mannanase Activity and Cell Wall Hydrolysis in Tomato Endosperm to Germination Rates. Plant Physiol 113: 1243–1252. doi:http://dx.doi.org/10.1104/pp.113.4.1243.
61. Pettolino FA, Walsh C, Fincher GB, Bacic A (2012) Determining the polysaccharide composition of plant cell walls. Nat Protoc 7: 1590–1607. doi:10.1038/nprot.2012.081.
62. Oxenboll Sørensen S, Pauly M, Bush M, Skjøt M, McCann MC, et al. (2000) Pectin engineering: modification of potato pectin by in vivo expression of an endo-1,4-beta-D-galactanase. Proc Natl Acad Sci U S A 97: 7639–7644. doi:10.1073/pnas.130568297.
63. Moore JP, Nguema-Ona E, Chevalier L, Lindsey GG, Brandt WF, et al. (2006) Response of the leaf cell wall to desiccation in the resurrection plant *Myrothamnus flabellifolius*. Plant Physiol 141: 651–662. doi:10.1104/pp.106.077701.

Appendix 2

M. Boutard, L. Ettwiller, **T. Cerisy**, A. Alberti, K. Labadie, M. Salanoubat, I. Schildkraut, and A. C. Tolonen, "Global repositioning of transcription start sites in a plant-fermenting bacterium," *Nat. Commun.*, vol. 7, p. 13783, 2016.

nature COMMUNICATIONS

ARTICLE

Received 9 Aug 2016 | Accepted 1 Nov 2016 | Published 16 Dec 2016

DOI: 10.1038/ncomms13783 OPEN

Global repositioning of transcription start sites in a plant-fermenting bacterium

Magali Boutard[1,2], Laurence Ettwiller[3], Tristan Cerisy[1,2,4,5], Adriana Alberti[1], Karine Labadie[1], Marcel Salanoubat[1,2,4,5], Ira Schildkraut[3] & Andrew C. Tolonen[1,2,4,5]

Bacteria respond to their environment by regulating mRNA synthesis, often by altering the genomic sites at which RNA polymerase initiates transcription. Here, we investigate genome-wide changes in transcription start site (TSS) usage by *Clostridium phytofermentans*, a model bacterium for fermentation of lignocellulosic biomass. We quantify expression of nearly 10,000 TSS at single base resolution by Capp-Switch sequencing, which combines capture of synthetically capped 5′ mRNA fragments with template-switching reverse transcription. We find the locations and expression levels of TSS for hundreds of genes change during metabolism of different plant substrates. We show that TSS reveals riboswitches, non-coding RNA and novel transcription units. We identify sequence motifs associated with carbon source-specific TSS and use them for regulon discovery, implicating a LacI/GalR protein in control of pectin metabolism. We discuss how the high resolution and specificity of Capp-Switch enables study of condition-specific changes in transcription initiation in bacteria.

[1] 1CEA, DRF, IG, Genoscope, Évry 91000, France. [2] CNRS-UMR8030, Évry 91000, France. [3] New England Biolabs, Inc., Ipswich, Massachusetts 01938, USA. [4] Université Paris-Saclay, Évry 91000, France. [5] Université d'Évry, Évry 91000, France. Correspondence and requests for materials should be addressed to A.C.T. (email: atolonen@genoscope.cns.fr).

Bacteria translate environmental signals into cellular responses using a network of regulatory RNA and proteins that control genome-wide transcription patterns. Many of these regulators affect where RNA polymerase initiates messenger RNA (mRNA) synthesis at transcription start sites (TSS). As such, locating and quantifying changes in TSS usage is an important step to understand bacterial gene regulation. Here, we investigate TSS architecture in *Clostridium phytofermentans* ISDg, a soil bacterium that ferments plant biomass into ethanol, H_2 and acetate[1], and belongs to the *Lachnospiraceae* family that includes gut commensals with important roles in host nutrition[2,3]. This anaerobic mesophile metabolizes diverse plant components including cellulose, hemicellulose and pectin by tailoring expression of many carbohydrate-active enzymes (CAZymes) and other metabolic enzymes to the available substrate[4,5]. *C. phytofermentans* has a 4.8 Mb genome with 3,926 predicted protein-encoding genes[3], and its ability to alter gene expression in response to carbon sources and other environmental cues is mediated by over 300 transcription regulator proteins[6] and numerous non-coding RNA including metabolite-sensing riboswitches[7].

We investigate genome-wide patterns of *C. phytofermentans* transcription initiation on heterogeneous plant substrates by demonstrating an approach called Capp-Switch sequencing. The initiating nucleotide of nascent mRNA is distinguished by a 5′ triphosphate (5′-PPP), which has been exploited for genome-wide TSS identification with dRNA-seq[8] by depleting rRNA and other monophosphorylated transcripts using terminal exonuclease (TEX). dRNA-seq has been applied to diverse bacteria[9–13], but incomplete and non-specific degradation of processed RNA requires TSS identification to be based on statistical comparison of read coverage in +TEX and −TEX samples. Capp-Switch avoids these problems by capturing and purifying 5′ mRNA fragments, which are reverse transcribed with template-switching to tagged cDNA for high-throughput sequencing (Fig. 1). The 5′-PPP of mRNA are modified by vaccinia capping enzyme (VCE) to bear a biotinylated guanosine cap that facilitates their capture and purification using streptavidin magnetic beads. Recently, TSS were identified by Cappable-Seq[14] using VCE to add a desthiobiotin cap for bead-based capture of 5′ mRNA, which were then eluted from the beads and de-capped to ligate adapters for reverse transcription to tagged cDNA. Capp-Switch streamlines this approach by reverse transcribing the 5′ mRNA fragments using template-switching by Moloney murine leukemia virus reverse (MMLV) transcriptase[15]. Template-switching avoids adapter ligation and enables synthesis of 5′-tagged cDNA without releasing RNA from the beads, permitting use of an irreversible, biotinylated cap to increase RNA capture affinity. In all, we show Capp-Switch is a robust method that yields a genome-wide, strand-specific, quantitative map of TSS at single nucleotide resolution.

We apply Capp-Switch sequencing to define a genome-wide map of 9,457 TSS during *C. phytofermentans* growth on raw biomass, heterogeneous polysaccharides (cellulose, hemicellulose and pectin) and their constituent sugars. We use this TSS map to investigate features controlling gene regulation, such as RNA polymerase binding sites, 5′ untranslated region (UTR) structure, alternative promoters, operons and non-standard (leaderless and antisense) transcription. We identify sequence motifs associated with groups of TSS that are differentially expressed on specific carbon sources and show these motifs can be used to reconstruct transcription factor regulons. By integrating Capp-Switch data with an updated genome annotation, RNA-seq and proteomics, we discover novel transcriptional units (TU) and protein-encoding genes. Finally, we discuss how Capp-Switch sequencing can be applied as a general approach to explore transcription regulation in prokaryotes.

Results

General transcriptome features. Capp-Switch sequencing quantified TSS with high reproducibility between duplicate model substrate (Fig. 2a) and raw biomass (Fig. 2b) cultures. We identified 9,457 TSS across treatments (Supplementary Data 1), one-third of which were expressed in both sugar and polysaccharide cultures (Fig. 2c). Most reads (74%) contribute to InterS TSS (Fig. 2d), which we observed upstream of 898 genes. Among these, 687 genes (77%) are predicted to start operons[16] (Supplementary Data 2), supporting these operon predictions and the existence of many sub-operons. The 5′ UTR, spanning from the primary TSS to the start codon, is less than 100 bp for most genes, but there is no correlation between 5′ UTR length and TSS

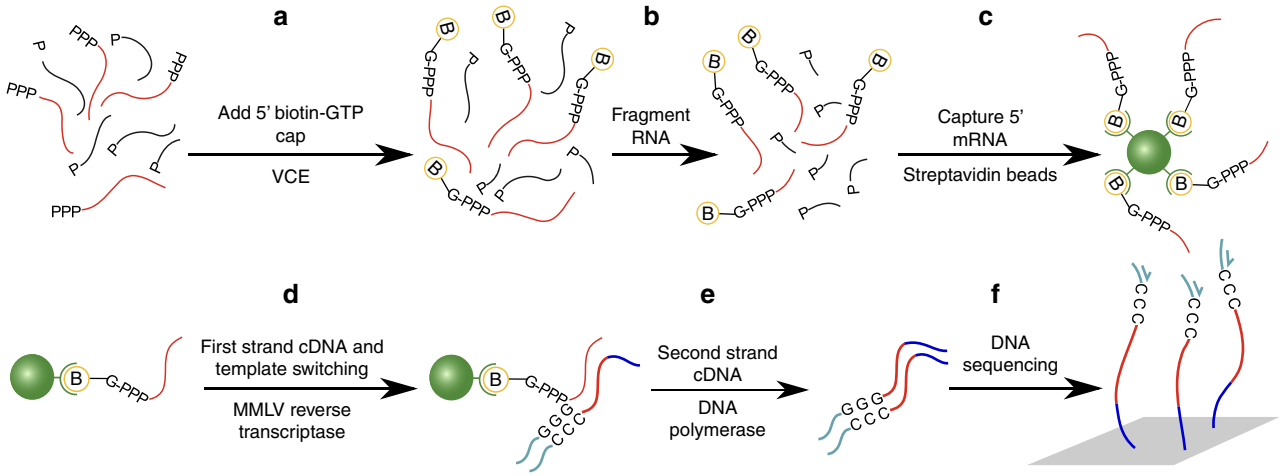

Figure 1 | Overview of the Capp-Switch sequencing approach. Capp-Switch includes (**a**–**c**) capture of 5′ mRNA fragments and (**d**–**f**) cDNA synthesis and sequencing. (**a**) The mRNA 5′ triphosphate is capped with biotin-GTP by VCE. (**b**) RNA is fragmented and (**c**) the capped 5′ mRNA fragments are captured on streptavidin magnetic beads and separated from other RNA. (**d**) The 5′ mRNA fragments are reverse transcribed to single-stranded cDNA using MMLV reverse transcriptase. An oligonucleotide hybridizes to the 3′ overhang and the complementary sequence is synthesized by the MMLV template-switching activity. (**e**) Double-stranded cDNA is synthesized using primers that hybridize to the single-stranded cDNA termini. (**f**) The cDNA is sequenced on a high-throughput platform.

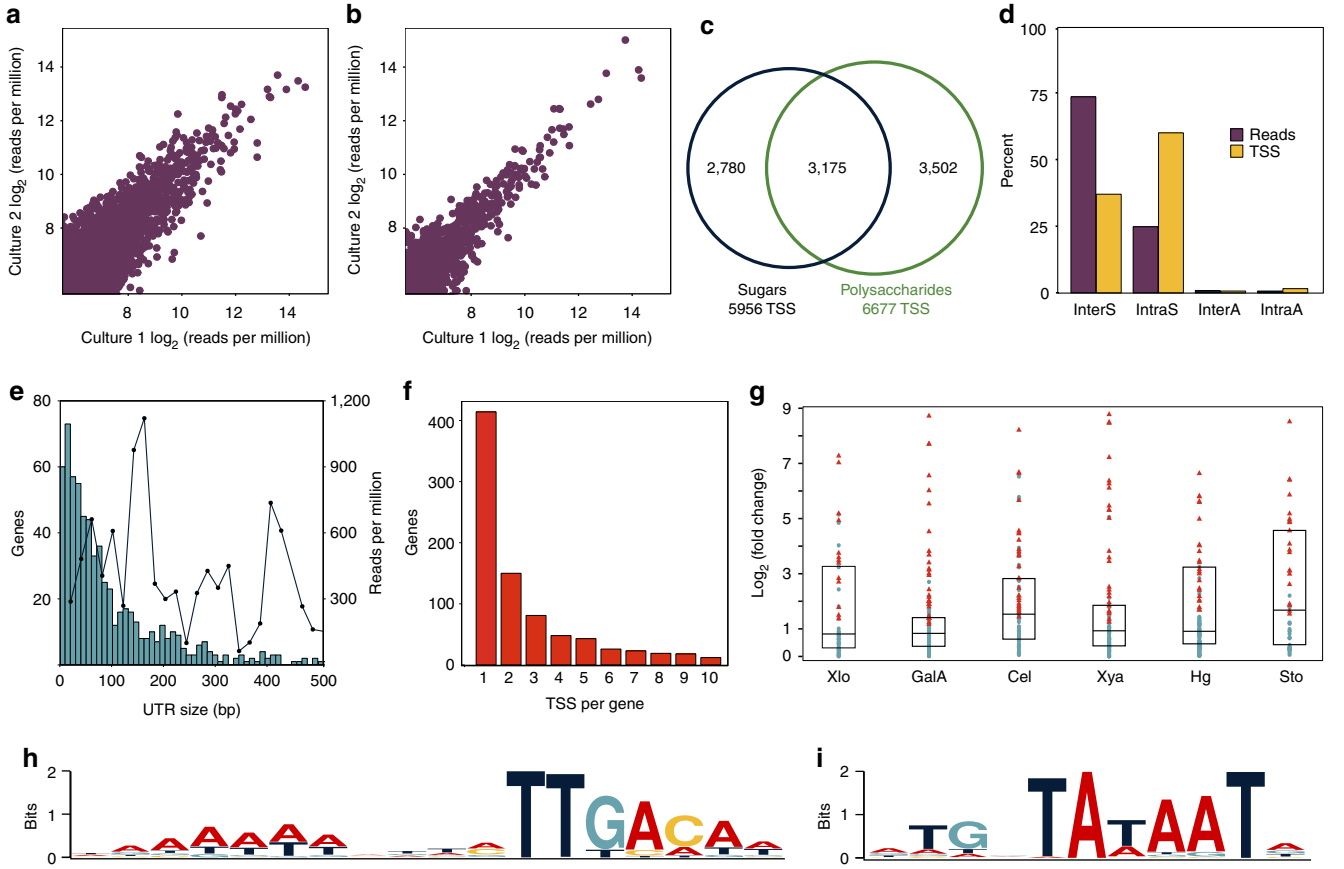

Figure 2 | General features of TSS identification by Capp-Switch sequencing. Capp-Switch reproducibly quantifies TSS usage in duplicate (**a**) glucose (4,399 TSS; $R^2 = 0.96$) and (**b**) stover (1,532 TSS; $R^2 = 0.99$) cultures. (**c**) Venn diagram showing overlap of TSS identified in at least one monosaccharide and one polysaccharide or biomass treatment. (**d**) Percentage of reads (purple) and TSS (yellow) classified as InterS, IntraS, InterA or IntraA summed across treatments. (**e**) The length of most 5′ UTR (primary TSS to start codon) is <100 bp (blue bars with left *Y* axis), but UTR length does not correlate with expression strength (black line with right *Y* axis). TSS strength is the average reads per million for all TSS in a 20 bp 5′ UTR size interval. Results show glucose data. (**f**) Distribution of the number of InterS TSS per gene for data summed across treatments. (**g**) Genes with substrate-specific TSS are often differentially expressed. The *Y* axis is the absolute value of $\log_2$ (RPKM substrate/RPKM glucose) from RNA-seq for all genes with InterS TSS specific to that substrate. Substrates are xylose (Xlo $n = 50$ genes), galacturonic acid (GalA $n = 146$ genes), cellulose (Cel $n = 94$ genes), xylan (Xya $n = 91$ genes), pectin (Hg $n = 119$ genes) and stover (Sto $n = 48$ genes). Symbols: red triangles are differentially expressed genes, blue circles unchanged genes, box shows median and interquartile range. Promoter regions upstream of TSS expressed on three sugars and polysaccharides show consensus (**h**) −35 and (**i**) −10 motifs recognized by RNA polymerase.

strength (Fig. 2e). Studies in other bacteria report many leaderless mRNA without 5′ UTR and ribosome binding sites (RBS)[11]. Four per cent of InterS TSS are potentially leaderless in *C. phytofermentans*, but these genes generally have another upstream TSS and retain a typical RBS similar to highly expressed *C. phytofermentans* genes (Supplementary Fig. 1).

Most genes were expressed from a single, primary TSS on all substrates (Fig. 2f), but 191 (21%) genes altered their primary TSS in response to carbon source. Further, genes with substrate-specific InterS TSS are often differentially expressed on that carbon source (χ^2 test, $P < 0.01$ for all substrates relative to glucose) (Fig. 2g), supporting that changing TSS is a widespread means of transcription regulation. In total, more than a thousand TSS are specific to each polysaccharide (Supplementary Fig. 2A). Xylan-specific (Supplementary Fig. 2B) and pectin-specific (Supplementary Fig. 2C) TSS are primarily associated with carbohydrate metabolism genes, while the most abundant functional category of cellulose-specific TSS is prophage genes (Supplementary Fig. 2D). The *C. phytofermentans* genome includes a large prophage island that is not predicted to encode a viable phage[3], but whose transcription is up-regulated on cellulose and biomass (Supplementary Fig. 3). This burst of transcriptional initiation at viral genes could indicate prophage excision was triggered on cellulosic substrates, that is, by low carbon stress, or that viral proteins contribute to bacterial fitness[17].

Sequences upstream of primary TSS generally contain the sigma-A-type consensus −35 and −10 hexamers recognized by RNA polymerase (RNAP) and associated elements that likely contribute to promoter function in this organism. An A-rich region upstream of the -35 hexamer (TTGACA) (Fig. 2h) resembles the 'UP element' that stimulates transcription initiation by interacting with the RNAP alpha subunit[18]. Also, the Pribnow hexamer (TATAAT) has an upstream TG di-nucleotide (Fig. 2i), which enhances transcription in certain other bacteria[19–21] by interacting with the RNAP sigma-A subunit[22]. In contrast, searching upstream of IntraS TSS identified an AT-rich stretch ~10 bp upstream of the TSS lacking RNAP binding sites (Supplementary Fig. 4A), suggesting IntraS TSS often result from promiscuous initiation at AT-rich sequences. We observed IntraS TSS comprised that more than 50% of TSS (Fig. 2d), albeit with fewer reads per site than InterS TSS. dRNA-seq studies have rationalized similarly abundant intragenic TSS as resulting from incomplete TEX degradation[12], but our data support these TSS

bear 5′-PPP indicative of transcription initiation. IntraS TSS are preferentially found in the 5′ end of genes (Supplementary Fig. 4B), supporting they are under selective pressure and may have roles including expression of alternative protein isoforms or as mimicry molecules to sequester other RNA and ribonucleases from their mRNA targets[9].

Capp-Switch reads (Fig. 3a–d) start at specific positions with respect to known genes showing TSS at single base resolution, whereas RNA-seq reads begin throughout genes (Fig. 3e–h). We observed four common TSS situations: genes with a single upstream TSS, genes with both upstream and intragenic TSS, genes with multiple TSS on a single substrate and genes with substrate-specific TSS. For example, the glyceraldehyde 3-phosphate dehydrogenase (*gadph*) gene is constitutively transcribed from a single TSS (Fig. 3a). The pyruvate ferredoxin oxidoreductase (*pfor*) gene is transcribed from a single, upstream TSS and another, weaker TSS in the coding sequence (Fig. 3b). The *cel5A* cellulase gene[23] is simultaneously transcribed from multiple TSS on cellulose (Fig. 3c), as are other cellulases (Supplementary Fig. 5). CAZyme expression in *C. phytofermentans* is controlled by carbon source[24,25] and our data supports their regulation involves multiple promoters. The *cphy1510* gene encoding the most active xylanase[5] is transcribed from three TSS on xylan and a different, upstream TSS on pectin (Fig. 3d). Similarly, genes for other CAZymes including three cellulases, one other xylanase, four pectinases and two glycosyl transferases changed their primary TSS as a function of carbon source. We confirmed the positions of the primary TSS identified by Capp-Switch for *gadph*, *pfor* (IntraS and primary TSS), *cphy2243* and *cphy1510* (xylan and pectin) using 5′ RACE (Supplementary Fig. 6).

Motifs associated with TSS clusters. We clustered TSS based on expression across carbon sources and searched sequences surrounding TSS for overrepresented motifs (Supplementary Fig. 7; Supplementary Data 3), revealing TSS clusters that share motifs with potential regulatory functions (Fig. 4). For example, the TSS cluster up-regulated on galacturonic acid and homogalacturonan (HG) (Fig. 4c) has a palindromic motif resembling the *cre* operator (TGAAAGCGCTTTCA) bound by *B. subtilis* CcpA[26,27], a LacI/GalR regulator of numerous carbon metabolism genes. LacI/GalR genes often have upstream copies of their operators to auto-repress transcription[28], and we found three copies of the galacturonic acid cluster motif in the 5′ UTR of *cphy2742*, a LacI/GalR gene specifically up-regulated on galacturonic acid (Fig. 5a). Further, three of the six LacI/GalR genes with detected primary TSS have upstream variants of the *cre* operator that are conserved in their orthologs from related species (Fig. 5b–d), leading us to propose *C. phytofermentans* LacI/GalR regulators recognize related, but distinct, operators to control separate regulons. Supportingly, the putative Cphy2742 operator (Fig. 5b) is upstream of 22 genes in the *C. phytofermentans* genome (Supplementary Table 1) including 3 CAZymes (PL9 pectin lyases) that degrade HG to galacturonic acid[5] and transcription units containing all genes needed to assimilate galacturonic acid[29] (Supplementary Fig. 8).

The putative Cphy2742 operator sites are co-located with or downstream of TSS for HG degradation and galacturonic acid metabolism genes (Fig. 5e), supporting Cphy2742 binds these sites to block transcription. Transcription of the *pl9* genes *cphy2919* and *cphy3869* switches to upstream primary TSS on galacturonic acid relative to HG, but all TSS are close enough to be potentially regulated by Cphy2742 operators. The *pta-ackA* (*cphy1326-7*) acetate synthesis operon also has a Cphy2742 operator and both *pta-ackA* expression and acetate formation are elevated on galacturonic acid (Supplementary Fig. 9). While *B. subtilis* CcpA represses most of its targets, it activates *pta* and *ackA* transcription[30,31] by binding upstream of their promoters[32]. The Cphy2742 operator is also upstream of the *pta* gene TSS, suggesting Cphy2742 may similarly activate transcription of the *pta-ackA* operon as well as the glycolytic gene *ppdK* and the hydrolase gene *cphy0367*. Collectively, we propose Cphy2742 represses a comprehensive set of pectin fermentation genes by binding a conserved palindrome at or downstream of their TSS to block transcription. In response to a galacturonic acid-based signal, Cphy2742 de-represses itself and its targets, and may activate transcription of acetate synthesis and other aspects of carbon metabolism by binding upstream of TSS.

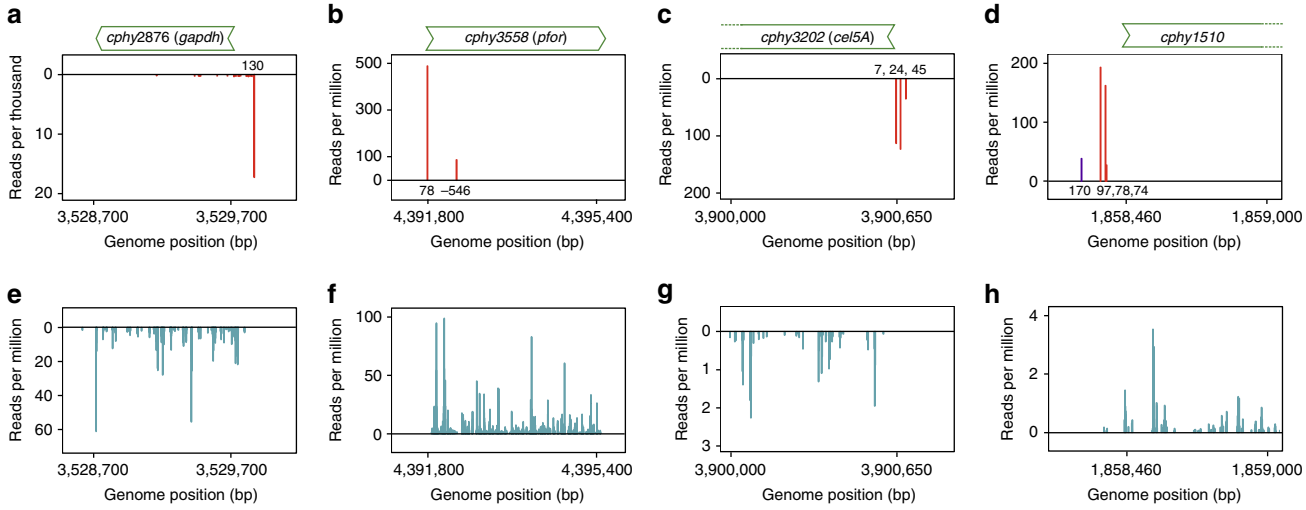

Figure 3 | Capp-Switch reads start at specific genome positions corresponding to putative TSS. The number of reads starting at each genome position are shown for Capp-Switch (**a**–**d**) and RNA-seq (**e**–**h**). The *cphy2876 gapdh* gene (**a**,**e**) has a single TSS (glucose data shown). The *cphy3558 pfor* gene (**b**,**f**) has an upstream TSS and an intragenic sense TSS (glucose data shown). The *cphy3202 cel5A* cellulase gene (**c**,**g**) has three TSS during growth on cellulose. The *cphy1510* xylanase gene (**d**,**h**) is expressed from three TSS on xylan (red bars) and a single, upstream TSS on pectin (purple). Plots show the number of reads starting at each genome position with forward strand reads on the positive *Y* axis and reverse strand reads on the negative *Y* axis. Distance to the start codon is shown at the base of TSS peaks.

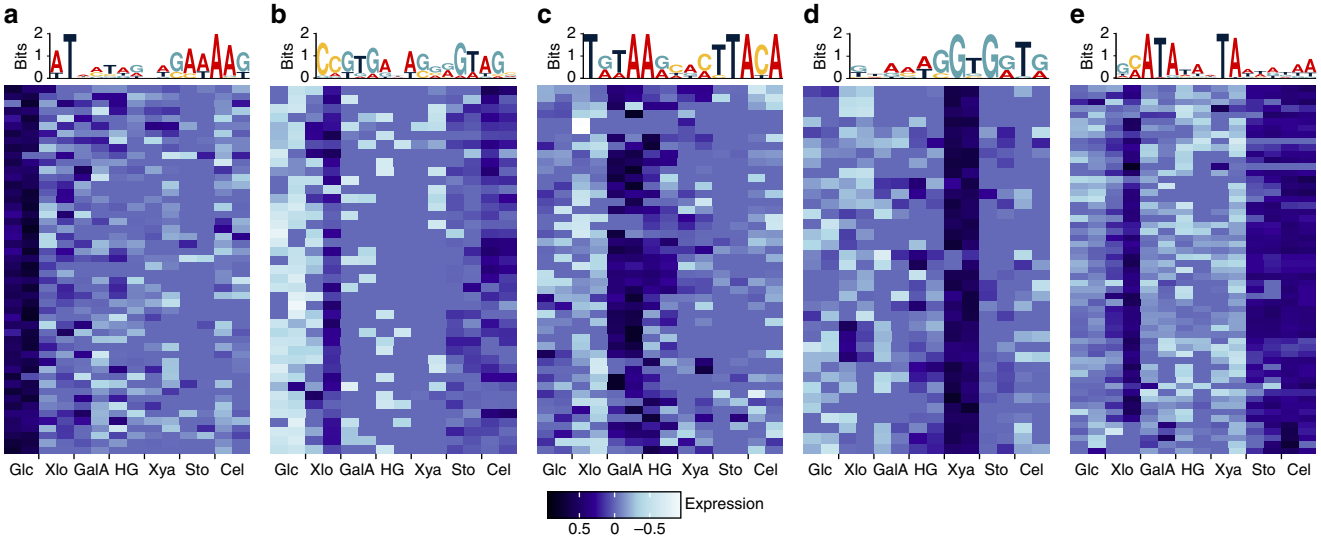

Figure 4 | TSS in carbon source-specific clusters share DNA sequence motifs. TSS clusters differentially expressed on (**a**,**b**) glucose, (**c**) galacturonic acid and HG, (**d**) xylan and (**e**) stover and cellulose are shown along with their associated sequence motifs. Rows are expression of a TSS cluster member and columns are duplicate glucose (Glc), xylose (Xlo), galacturonic acid (GalA), homogalacturonan (HG), xylan (Xya), stover (Sto) and cellulose (Cel) cultures. Colours show TSS expression as $\log_2$-transformed read counts scaled to a median of zero for each TSS.

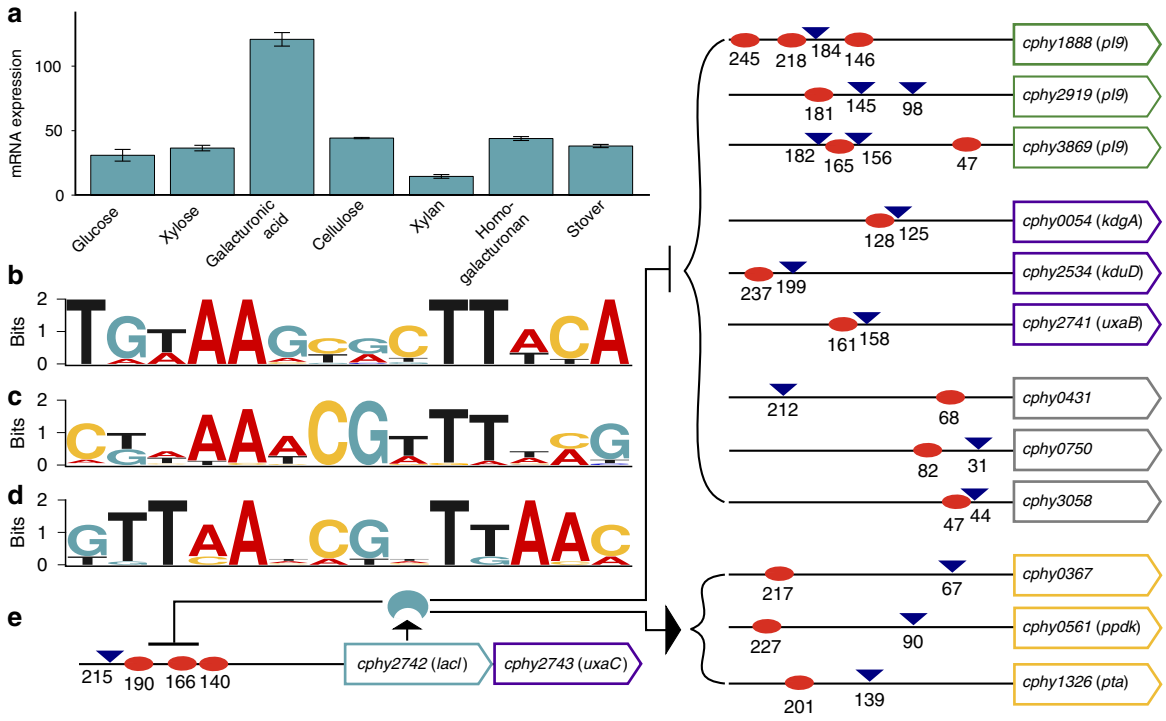

Figure 5 | The role of the LacI/GalR regulator Cphy2742 in galacturonic acid and pectin metabolism. (**a**) Transcription of the LacI/GalR gene *cphy2742* is up-regulated on galacturonic acid relative to other carbon sources. Bars shows average RNA-seq RPKM of duplicate cultures; error bars are one s.d. (**b**–**d**) Upstream palindromes resembling *cre* operator sites found upstream of *C. phytofermentans* LacI/GalR genes and their orthologs from related genomes (**b**) *cphy2742* (motif $e = 1.1 \times 10^{-8}$), (**c**) *cphy2467* (motif $e = 2.4 \times 10^{-8}$) and (**d**) *cphy1883* (motif $e = 8.9 \times 10^{-2}$). (**e**) Twelve genes have both TSS (blue triangles) and putative Cphy2742 operators (red ovals) including genes for pectin lyases (green), galacturonic acid metabolism (purple), general carbon metabolism (yellow) and other or unknown (grey). The distance from the translation start is shown for each site.

Antisense and novel transcripts. Recent studies found 30–40% of TSS are antisense in other bacteria[8,9,13]. However, antisense transcription appears rare in *C. phytofermentans:* <1% of TSS were antisense either between (InterA) or within genes (IntraA) (Fig. 2d). To further investigate whether diffuse antisense transcription was underestimated by our TSS thresholds, we classified all mapped read starts, including those not meeting TSS thresholds. Even then, InterA and IntraA classes together comprise <4% reads. This dearth of antisense transcription may relate to the early evolutionary divergence of the Clostridiales[33]. Alternatively, we would not detect antisense transcripts that were processed to remove 5′-PPP or that are below the 200 bp size threshold of our cDNA libraries, but studies in other bacteria using larger size thresholds found antisense TSS

in ~35% of genes[10]. While comparatively rare, antisense transcription appears to have important cellular functions. For example, we observed an antisense TSS in the 5′ UTR of the sporulation regulator *spoOA* (*cphy2497*) that also opposes transcription of the *spoIVB* peptidase (*cphy2498*) (Fig. 6a). This TSS was expressed on all sugars, but not polysaccharides, supporting antisense transcription has a role in repressing sporulation during log growth in sugar-replete conditions.

TSS reveal novel transcriptional features such as a TU downstream of the glycoside hydrolase *cphy2658* that is up-regulated to have the strongest initiation site in the genome on cellulose and corn stover (Fig. 6b). This region contains a hypothetical open-reading frame (ORF) in the MaGe annotation (*clops3132*) that has no similar sequences in Genbank, but the ORF lacks an ribosome binding site (RBS), and we did not detect any expressed peptides from this region by mass spectrometry, suggesting it is a non-coding RNA. The most highly expressed ABC transporter on glucose is a putative operon (*cphy2241-3*) with a single TSS (Supplementary Fig. 5C,F). On all other carbon sources, we observed repression of *cphy2241-3* along with appearance of an upstream, antisense TU (Fig. 6c) that has no mapped peptides or predicted ORF. Non-coding RNA are often associated with ABC transporters in clostridia[34], and they may also regulate ABC transport in this organism.

The *C. phytofermentans* genome may encode significantly more genes than in the NCBI Genbank annotation. Classifying TSS using the MaGe annotation showed 735 (7%) TSS map to MaGe-specific *clops* genes of unknown function (Supplementary Data 4), including 64 *clops* genes with InterS TSS. We examined which of these novel TU encode proteins by mapping *C. phytofermentans* MS/MS peptide spectra to the genome translated in all frames, identifying peptides outside the predicted proteome in 21 InterS, 13 IntraS, 5 InterA and 25 IntraA regions (Supplementary Data 5). The combination of TSS and expressed peptides supports ORFs with N-terminal extensions such as *cphy0891* (Supplementary Fig. 10A) and the existence of novel ORFs. For example, *clops3461*, which overlaps with *cphy2929* on the opposite strand (Fig. 6d), and an antisense overlapping ORF in *cphy1953* encoding the ComEA competence protein (Supplementary Fig. 10B).

TSS also show mechanisms of RNA-mediated gene regulation. Comparative genomics with other clostridia detected a putative T-box upstream of the *C. phytofermentans trp* operon[34]. In low tryptophan conditions, the T-box promotes antitermination of the *trp* operon by base pairing with uncharged $tRNA^{trp}$ (ref. 35). We observed transcription halted abruptly in the 5′ UTR of the *trp* operon in glucose cultures (Fig. 6e), consistent with T-box-mediated repression. In cellulose cultures, antitermination in the T-box enabled *trp* operon mRNA expression, potentially enabling translation of the trytophan-rich carbohydrate binding modules in cellulases and other CAZymes. TSS also support riboswitches associated with genes for metabolism of flavin mononucleotide (FMN), cobalamin, thiamine pyrophosphate (TPP) and lysine (Supplementary Data 6). For example, *C. phytofermentans* is auxotrophic for thiamine, which it uptakes by a thiamine transporter, Cphy0729 (ref. 36). The *cphy0729* gene has a single, constitutive TSS with an extended 5′ UTR containing a putative TPP-sensing riboswitch (Fig. 6f) that could regulate transporter expression in response to intracellular TPP levels[37].

Discussion

The strategy presented here to quantify condition-specific changes in transcription initiation by Capp-Switch sequencing could be generally applied to dissect the regulation of complex bacterial phenotypes. In this study, we explored the transcriptional programme enabling *C. phytofermentans* to ferment the cellulosic, hemicellulosic and pectic components of plant biomass.

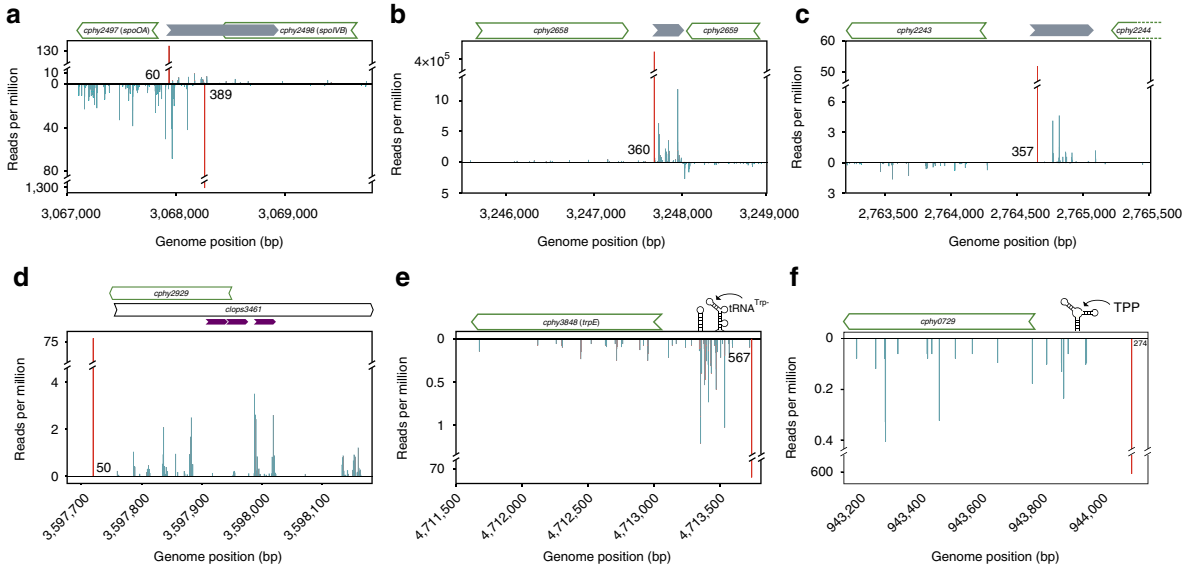

Figure 6 | TSS show genome features. (**a**) The *cphy2497 spoOA* gene has both a primary TSS and an antisense TSS in the 5′ UTR (grey arrow) that were observed on all sugars (glucose data shown). (**b**) A novel transcription unit (grey arrow) is up-regulated to be the most highly expressed TSS on biomass. (**c**) Induction of a transcription unit (grey arrow) upstream of the ABC transporter *cphy2243* is associated with repression of the transporter. This TSS was observed on all substrates except glucose (cellulose data shown). (**d**) A primary TSS, RNA-seq reads, and three in-frame peptides expressed on cellulose support the MaGe-predicted *clops3461* gene rather than the annotated *cphy2929* gene. Positions of peptides detected by mass spectrometry (purple) are shown. (**e**) The *trpE* (*cphy3848*) gene has an upstream T-box that terminates transcription in the 5′ UTR during log-phase growth on glucose. (**f**) The thiamine transporter (*cphy0729*) has an extended 5′ UTR containing a TPP-binding riboswitch. All plots show the number of reads starting at each genome position for RNA-seq (blue) and Capp-Switch (red). Numbers at base of TSS peaks are distances to start codons of (**a**) *cphy2497*, (**c**) *cphy2243*, (**d**) *clops3461*, (**e**) *cphy3848*, (**f**) *cphy0729* and (**b**) the *cphy2659* stop codon.

We found that growth on these different carbon sources entailed widespread TSS changes, including use of substrate-specific TSS for genes encoding biomass-degrading enzymes such as cellulases, xylanases and pectinases. Substrate-specific TSS could enable tuning of expression by changing promoters or the regulatory properties (that is, binding sites or secondary structure) of the 5′ UTR. We observed that genes encoding cellulases and other enzymes are simultaneously expressed from more than one TSS. Multiple regulators may control transcription of these genes, reflecting the numerous transcription factors encoded by this organism (Supplementary Data 7). Genes for biomass-degrading enzymes in other Clostridiales are regulated by various transcription factors including a two-component system for hemicellulases[38], a LacI/GalR protein for β-1-3 glucanases[39] and alternative sigma factors for cellulases[40]. We defined TSS clusters that were differentially expressed on specific carbon sources and used them to guide the discovery of sequence motifs with potential regulatory function, leading us to identify the LacI/GalR Cphy2742 as a putative regulator of pectin metabolism. Combining TSS mapping with motif searching could be broadly applied to LacI/GalR regulators and other types of transcription factors. For example, each of the 4 TetR regulators for which we detected TSS also have conserved, TSS-associated palindromes that resemble operator sites (Supplementary Fig. 11).

We also gained insight into regulatory mechanisms such as antisense transcription, leaderless transcription and non-coding RNA. We observed that antisense and leaderless transcription are much rarer than reported in other bacteria and it will be interesting to see if they are similarly uncommon in closely-related bacteria. We also show that integration of Capp-Switch TSS mapping with RNA-seq and proteomics enables discovery of novel transcription units and protein-encoding genes. Transcription initiation is a complex and important component of gene regulation for which most of the underlying mechanisms in *C. phytofermentans* are yet unknown. Further, these results illustrate how little we know about gene regulation in plant-fermenting clostridia, a group of bacteria with important roles in soil and gut microbiomes that have significant potential to serve as biocatalysts for industrial transformation of plant biomass.

Methods

Bacterial cultivation. *C. phytofermentans* ISDg (ATCC 700394) was cultured anaerobically at 30 °C in GS2 medium[41] containing 5 g l^{-1} of either D-(+)-glucose (Sigma G5767), D-(+)-xylose (Sigma X3877), D-galacturonic acid sodium salt (Sigma 73960), regenerated amorphous cellulose (RAC) from Avicel PH-101 (Sigma 11365), birchwood xylan (Sigma X0502), apple pectin (HG) (Sigma P8471) or raw corn stover (Qteros Inc) cut in 0.5 × 3.0 cm strips. RAC was prepared by phosphoric acid treatment[42]. Duplicate cultures were sampled in mid-log phase or after 2 days (RAC) or 3 days (stover). Fermentation products were quantified by HPLC[43].

Capp-Switch library preparation. Total RNA was extracted from duplicate cultures for each treatment using TRI reagent (Sigma 93289) and treated with Turbo DNase (Ambion AM2238) at 0.2 U μg^{-1} RNA for 30 min at 37 °C. RNA was purified by Zymo Concentrator-5 (Zymo Research R1015) (>200 bp capture) into 15 μl water. RNA was 5′ capped using VCE (NEB M2080) at 3 U μg^{-1} RNA with 0.1 mM SAM and 0.5 mM 3′ biotin-GTP (NEB N0760) for 30 min at 37 °C and purified by Zymo Concentrator-5 (>200 bp capture) with two additional washes into 45 μl water. RNA was fragmented for 30 s at 94 °C using NEBNext Magnesium-based RNA fragmentation buffer (NEB E6101) and purified by Zymo Concentrator-5 (total RNA capture) into 100 μl water. Streptavidin magnetic beads (NEB S1421S) were pre-washed twice with low-salt buffer (10 mM Tris, 50 mM NaCl, 1 mM EDTA), twice with binding buffer (10 mM Tris, 500 mM NaCl, 1 mM EDTA) and resuspended at 4 mg ml^{-1} beads in binding buffer. Capped RNA fragments were bound to streptavidin beads for 20 min at room temperature and magnetically separated from other RNA by washing twice with binding buffer and twice with low-salt buffer to elute non-bound RNA. Beads were washed once with 1 mM Tris–HCl pH 7.5 and resuspended in 1 mM Tris–HCl pH 7.5.

RNA was converted to single-strand cDNA by SMARTscribe MMLV reverse transcriptase (Clontech 634836) at 10 U μl^{-1} with 2.5 mM DTT, 1 mM dNTP, 1.2 μM SMARTer stranded oligo and 0.6 μM SMART stranded N6 primer (Clontech 634836) by incubating 90 min at 42 °C and 10 min at 70 °C. Beads were collected and the supernatant was combined with the liquid fraction after the beads were washed with 30 μl 1 mM Tris pH 7.5. The cDNA was twice purified using 1 volume of solid phase reversible immobilization (SPRI) beads (Beckman Coulter A63880). cDNA was left on beads after the second purification and double-stranded cDNA was synthesized by 18 cycles PCR using SeqAmp DNA polymerase (Clontech 638504) with 0.25 μM primers (Universal Forward PCR primer and indexed Reverse PCR primer) and then SPRI purified with 1 volume of beads. DNA was sequenced on Illumina MiSeq with 150 bp paired-end reads chemistry.

TSS identification and classification. Sequencing reads were quality filtered[44] and the 3 bp MMLV reverse transcriptase 3′ non-template extension was removed from the 5′ end of forward (R1) reads. Reads were mapped to the *C. phytofermentans* ISDg genome (NCBI NC_010001.1) using Bowtie 2 (version 2.2.4)[45]. Alignments showed 87–98% of reads mapped to unique positions in the *C. phytofermentans* genome, yielding between 0.4 million (corn stover) and 3.4 million (glucose) reads per culture (Supplementary Table 2). TSS were identified using R1 reads by calculating the number of reads starting at each genomic position, clustering read counts within a 5 bp sliding window, and retaining the position with the greatest number of reads. TSS were defined as genome positions with greater than 10 read starts per million reads in both duplicate cultures. Capp-switch TSS were confirmed by 5′ RACE (Sigma 03353621001) using primers in Supplementary Table 3 to amplify PCR products, which were resolved by electrophoresis, excised and sequenced.

Genes in the NCBI and MicroScope (MaGe) annotations[46] were used to divide TSS into four categories: InterS (intergenic TSS with downstream gene in same orientation), InterA (intergenic TSS with downstream gene opposite orientation), IntraS (intragenic TSS in gene with same orientation) or IntraA (intragenic TSS in gene with opposite orientation). The InterS TSS with the most reads for each gene was defined as the primary TSS. Capp-Switch results were compared with strand-specific (dUTP) RNA-seq of *C. phytofermentans* grown in the same culture conditions[5]. RNA-seq gene expression was calculated as RPKM using the Bioconductor[47] package 'easyRNASeq' and differential expression was defined as a DESeq[48] (version 1.22.1) *P*-value <0.05 adjusted for multiple testing of the 3,902 genes in *C. phytofermentans* genome by Bonferroni correction. Peptides corresponding to novel ORFs were identified by mapping peptide MS/MS spectra from glucose, xylan and cellulose cultures[4] to the genome translated in all six frames. Peptides were identified from spectra using SEQUEST and filtered to a 5% false discovery rate using a target-decoy approach[49,50] including a target database and a decoy of the reversed sequences.

Motif analysis. Sequence motifs were identified using MEME[51] with a background model of di-nucleotide frequencies in the *C. phytofermentans* genome. Searches for RNA polymerase binding site motifs included positions 25–50 bp (−35 motif) and 5–20 bp (−10 motif) upstream of all primary TSS expressed on the three sugars and polysaccharides. The top palindromic motifs associated with LacI/GalR and TetR regulators were found by searching sequences from −250 (upstream) to +50 bp (downstream) relative to the start codon of *C. phytofermentans* genes and their putative orthologs from related genomes identified by top reciprocal BLAST searches (Supplementary Table 4). These motifs were used for genome-wide scans from −250 to +50 bp within all *C. phytofermentans* genes using MAST[52]. To cluster TSS by expression, the 1,188 TSS with at least a 30-fold change in read counts between two conditions were $\log_2$-transformed and each TSS was normalized to have a median value of 0 across conditions and scaled so the sum of the squared expression levels is 1. TSS were separated into 24 clusters by *K*-means using the city-block similarity metric. Significant motifs ($e<0.001$) associated with individual *K*-means clusters were identified by searching −100 to +10 bp with respect to each TSS.

Data availability. The authors confirm that all data underlying the findings are fully available without restriction. RNA sequencing files in FASTQ format are available in the European Nucleotide Archive under study accession PRJEB13063.

References

1. Warnick, T. A., Methé, B. A. & Leschine, S. B. *Clostridium phytofermentans* sp. nov., a cellulolytic mesophile from forest soil. *Int. J. Syst. Evol. Microbiol.* **52,** 1155–1160 (2002).
2. Meehan, C. J. & Beiko, R. G. A phylogenomic view of ecological specialization in the Lachnospiraceae, a family of digestive tract-associated bacteria. *Genome Biol. Evol.* **6,** 703–713 (2014).
3. Petit, E. *et al.* Genome and transcriptome of *Clostridium phytofermentans*, catalyst for the direct conversion of plant feedstocks to fuels. *PLoS ONE* **10,** e0118285 (2015).
4. Tolonen, A. C. *et al.* Proteome-wide systems analysis of a cellulosic biofuel-producing microbe. *Mol. Syst. Biol.* **7,** 461 (2011).
5. Boutard, M. *et al.* Functional diversity of carbohydrate-active enzymes enabling a bacterium to ferment plant biomass. *PLoS Genet.* **10,** e1004773 (2014).

6. Hunter, S. *et al.* InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res.* **40,** D306–D312 (2012).
7. Nawrocki, E. P. *et al.* Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res.* **43,** D130–D137 (2015).
8. Sharma, C. M. *et al.* The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* **464,** 250–255 (2010).
9. Mitschke, J. *et al.* An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp. PCC6803. *Proc. Natl Acad. Sci. USA* **108,** 2124–2129 (2011).
10. Schlüter, J.-P. *et al.* Global mapping of transcription start sites and promoter motifs in the symbiotic α-proteobacterium *Sinorhizobium meliloti* 1021. *BMC Genomics* **14,** 156 (2013).
11. Cortes, T. *et al.* Genome-wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in *Mycobacterium tuberculosis*. *Cell Rep.* **5,** 1121–1131 (2013).
12. Shao, W., Price, M. N., Deutschbauer, A. M., Romine, M. F. & Arkin, A. P. Conservation of transcription start sites within genes across a bacterial genus. *MBio.* **5,** e01398-14 (2014).
13. Thomason, M. K. *et al.* Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in *Escherichia coli*. *J. Bacteriol.* **197,** 18–28 (2015).
14. Ettwiller, L., Buswell, J., Yigit, E. & Schildkraut, I. A novel enrichment strategy reveals unprecedented number of novel transcription start sites at single base resolution in a model prokaryote and the gut microbiome. *BMC Genomics* **17,** 199 (2016).
15. Zhu, Y. Y., Machleder, E. M., Chenchik, A., Li, R. & Siebert, P. D. Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *BioTechniques* **30,** 892–897 (2001).
16. Dehal, P. S. *et al.* MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic Acids Res.* **38,** D396–D400 (2010).
17. Bondy-Denomy, J. & Davidson, A. R. When a virus is not a parasite: the beneficial effects of prophages on bacterial fitness. *J. Microbiol.* **52,** 235–242 (2014).
18. Ross, W. *et al.* A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science* **262,** 1407–1413 (1993).
19. 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.* **261,** 11409–11415 (1986).
20. Helmann, J. D. Compilation and analysis of *Bacillus subtilis* sigma A-dependent promoter sequences: evidence for extended contact between RNA polymerase and upstream promoter DNA. *Nucleic Acids Res.* **23,** 2351–2360 (1995).
21. Burns, H. D., Ishihama, A. & Minchin, S. D. Open complex formation during transcription initiation at the *Escherichia coli* galP1 promoter: the role of the RNA polymerase alpha subunit at promoters lacking an UP-element. *Nucleic Acids Res.* **27,** 2051–2056 (1999).
22. Barne, K. A., Bown, J. A., Busby, S. J. & Minchin, S. D. Region 2.5 of the *Escherichia coli* RNA polymerase sigma70 subunit is responsible for the recognition of the 'extended-10' motif at promoters. *EMBO J.* **16,** 4034–4040 (1997).
23. Liu, W., Zhang, X.-Z., Zhang, Z. & Zhang, Y.-H. P. Engineering of *Clostridium phytofermentans* Endoglucanase Cel5A for improved thermostability. *Appl. Environ. Microbiol.* **76,** 4914–4917 (2010).
24. 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.* **74,** 1300–1313 (2009).
25. Tolonen, A. C. *et al.* Fungal lysis by a soil bacterium fermenting cellulose. *Environ. Microbiol.* **17,** 2618–2627 (2015).
26. Weickert, M. J. & Chambliss, G. H. Site-directed mutagenesis of a catabolite repression operator sequence in *Bacillus subtilis*. *Proc. Natl Acad. Sci. USA*. **87,** 6238–6242 (1990).
27. Marciniak, B. C. *et al.* High- and low-affinity cre boxes for CcpA binding in *Bacillus subtilis* revealed by genome-wide analysis. *BMC Genomics* **13,** 401 (2012).
28. Francke, C., Kerkhoven, R., Wels, M. & Siezen, R. J. A generic approach to identify transcription factor-specific operator motifs; Inferences for LacI-family mediated regulation in *Lactobacillus plantarum* WCFS1. *BMC Genomics* **9,** 145 (2008).
29. Richard, P. & Hilditch, S. D-galacturonic acid catabolism in microorganisms and its biotechnological relevance. *Appl. Microbiol. Biotechnol.* **82,** 597–604 (2009).
30. Grundy, F. J., Waters, D. A., Allen, S. H. & Henkin, T. M. Regulation of the *Bacillus subtilis* acetate kinase gene by CcpA. *J. Bacteriol.* **175,** 7348–7355 (1993).
31. Presecan-Siedel, E. *et al.* Catabolite regulation of the *pta* gene as part of carbon flow pathways in *Bacillus subtilis*. *J. Bacteriol.* **181,** 6889–6897 (1999).
32. Fujita, Y. Carbon catabolite control of the metabolic network in *Bacillus subtilis*. *Biosci. Biotechnol. Biochem.* **73,** 245–259 (2009).
33. Paredes, C. J., Alsaker, K. V. & Papoutsakis, E. T. A comparative genomic view of clostridial sporulation and physiology. *Nat. Rev. Microbiol.* **3,** 969–978 (2005).
34. Chen, Y., Indurthi, D. C., Jones, S. W. & Papoutsakis, E. T. Small RNAs in the genus *Clostridium*. *MBio.* **2,** e00340-10 (2011).
35. Merino, E. & Yanofsky, C. Transcription attenuation: a highly conserved regulatory strategy used by bacteria. *Trends Genet.* **21,** 260–264 (2005).
36. Tolonen, A. C., Petit, E., Blanchard, J. L., Warnick, T. & Leschine, S. B. in *Biological Conversion of Biomass for Fuels and Chemicals* (eds Sun, J. *et al.*) 114–139 (Royal Society of Chemistry, 2013).
37. Winkler, W., Nahvi, A. & Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419,** 952–956 (2002).
38. Celik, H. *et al.* A two-component system (XydS/R) controls the expression of genes encoding CBM6-containing proteins in response to straw in *Clostridium cellulolyticum*. *PLoS ONE* **8,** e56063 (2013).
39. Newcomb, M., Chen, C.-Y. & Wu, J. H. D. Induction of the celC operon of *Clostridium thermocellum* by laminaribiose. *Proc. Natl Acad. Sci. USA* **104,** 3747–3752 (2007).
40. Nataf, Y. *et al. Clostridium thermocellum* cellulosomal genes are regulated by extracytoplasmic polysaccharides via alternative sigma factors. *Proc. Natl Acad. Sci. USA* **107,** 18646–18651 (2010).
41. Cavedon, K., Leschine, S. B. & Canale-Parola, E. Cellulase system of a free-living, mesophilic clostridium (strain C7). *J. Bacteriol.* **172,** 4222–4230 (1990).
42. Hong, J., Ye, X., Wang, Y. & Zhang, Y.-H. P. Bioseparation of recombinant cellulose-binding module-proteins by affinity adsorption on an ultra-high-capacity cellulosic adsorbent. *Anal. Chim. Acta* **621,** 193–199 (2008).
43. Tolonen, A. C. *et al.* Physiology, genomics, and pathway engineering of an ethanol-tolerant strain of *Clostridium phytofermentans*. *Appl. Environ. Microbiol.* **81,** 5440–5448 (2015).
44. Alberti, A. *et al.* Comparison of library preparation methods reveals their impact on interpretation of metatranscriptomic data. *BMC Genomics* **15,** 912 (2014).
45. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–359 (2012).
46. Vallenet, D. *et al.* MicroScope—an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. *Nucleic Acids Res.* **41,** D636–D647 (2013).
47. Delhomme, N., Padioleau, I., Furlong, E. E. & Steinmetz, L. M. easyRNASeq: a bioconductor package for processing RNA-Seq data. *Bioinformatics* **28,** 2532–2533 (2012).
48. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11,** R106 (2010).
49. Elias, J. E. & Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **4,** 207–214 (2007).
50. Tolonen, A. C. & Haas, W. Quantitative proteomics using reductive dimethylation for stable isotope labeling. *J Vis. Exp.* **89,** e51416 (2014).
51. Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2,** 28–36 (1994).
52. Bailey, T. L. & Gribskov, M. Combining evidence using *p*-values: application to sequence homology searches. *Bioinformatics* **14,** 48–54 (1998).

Acknowledgements

This work was funded by a CNRS Chaire d'Excellence to A.C.T. and the Genoscope-CEA. We thank NEB for providing reagents (biotin-GTP, vaccinia capping enzyme and streptavidin beads), the Genoscope-CEA sequencing platform for RNA sequencing and the LABGeM group for supporting the MicroScope (MaGe) annotation resource.

Author contributions

L.E., A.A., M.S., I.S. and A.C.T. conceived the project. M.B., T.C. and K.L. collected data. M.B., L.E., I.S. and A.C.T. analysed the results. A.C.T. wrote the paper.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Boutard, M. *et al.* Global repositioning of transcription start sites in a plant-fermenting bacterium. *Nat. Commun.* **7**, 13783 doi: 10.1038/ncomms13783 (2016).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Appendix 3

A. C. Tolonen, **T. Cerisy**, H. El-Sayyed, M. Boutard, M. Salanoubat, and G. M. Church, "Fungal lysis by a soil bacterium fermenting cellulose," Environ. Microbiol., vol. 17, no. 8, pp. 2618–2627, 2015.

environmental microbiology

Environmental Microbiology (2015) **17**(8), 2618–2627 doi:10.1111/1462-2920.12495

Fungal lysis by a soil bacterium fermenting cellulose

Andrew C. Tolonen,[1,2,3]* Tristan Cerisy,[1,2,3] Hafez El-Sayyed,[1,2,3] Magali Boutard,[1] Marcel Salanoubat[1,2,3] and George M. Church[4]

[1]*CEA, DSV, IG, Genoscope, Évry, France.*

[2]*CNRS-UMR8030, Évry, France.*

[3]*Université d'Évry Val d'Essonne, Évry, France.*

[4]*Department of Genetics, Harvard Medical School, Boston, MA, USA.*

Summary

Recycling of plant biomass by a community of bacteria and fungi is fundamental to carbon flow in terrestrial ecosystems. Here we report how the plant fermenting, soil bacterium *Clostridium phytofermentans* enhances growth on cellulose by simultaneously lysing and consuming model fungi from soil. We investigate the mechanism of fungal lysis to show that among the dozens of different glycoside hydrolases *C. phytofermentans* secretes on cellulose, the most highly expressed enzymes degrade fungi rather than plant substrates. These enzymes, the GH18 Cphy1799 and Cphy1800, synergize to hydrolyse chitin, a main component of the fungal cell wall. Purified enzymes inhibit fungal growth and mutants lacking either GH18 grow normally on cellulose and other plant substrates, but have a reduced ability to hydrolyse chitinous substrates and fungal hyphae. Thus, *C. phytofermentans* boosts growth on cellulose by lysing fungi with its most highly expressed hydrolases, highlighting the importance of fungal interactions to the ecology of cellulolytic bacteria.

Introduction

Plant polysaccharides are the largest pool of reduced carbon in the terrestrial biosphere, making their recycling a major part of the global carbon cycle. While abundant, these polysaccharides form a recalcitrant matrix in the plant cell wall that can only be degraded by a specialized community of fungi and bacteria that secrete numerous carbohydrate-active enzymes (CAZy). Among these microbes, the cellulolytic clostridia are a group of Gram-positive bacteria that ferment insoluble plant polysaccharides to a variety of products. Due to the high abundance and diversity of cellulolytic clostridia, an understanding of how they recycle plant biomass is critical to soil microbial ecology. Cellulolytic clostridia are also among the best candidates for consolidated bioprocessing (Lynd *et al.*, 2002), an industrial process using a single microbe to both hydrolyse plant biomass and ferment the sugars directly to fuels and chemicals.

Clostridium phytofermentans is a forest soil mesophile that is unique among described bacteria in its ability to degrade diverse plant polysaccharides (cellulose, hemicellulose, starch, pectin) and ferment the resulting sugars primarily to ethanol and hydrogen (Warnick *et al.*, 2002). The *C. phytofermentans* genome encodes 170 CAZy, more than 100 of which are expressed on either cellulose or hemicellulose with stoichiometries tailored to each substrate (Tolonen *et al.*, 2011). Degradation of insoluble plant polysaccharides is an energy-intensive process that initially occurs outside the cell, meaning *C. phytofermentans* likely needs adaptations to scavenge nutrients and to exclude other microbes from accessing plant-derived sugars before they are imported. As fungi are often the dominant members of the plant biomass-degrading microbial community (Schneider *et al.*, 2012b), we studied how *C. phytofermentans* interacts with three phylogenetically diverse model fungi from forest soil: *Aspergillus niger*, *Trichoderma harzianum* and *Cladosporium cladosporioides*.

In this study, we show that *C. phytofermentans* rapidly lyses fungal biomass during growth on cellulose, which stimulates growth and accelerates cellulose degradation. We investigate the mechanism of fungal lysis to show that it involves the enzymes Cphy1799 and Cphy1800. These are the most highly expressed CAZy during growth on cellulose both at the mRNA (Tolonen *et al.*, 2009) and protein levels (Tolonen *et al.*, 2011), exceeding expression of the cellulases. They belong to glycoside hydrolase family 18 (GH18), which consists primarily of chitinases (Cantarel *et al.*, 2009). Chitin is a 1,4-beta-N-acetyl-D-glucosamine (GlcNAc) polymer that is the primary structural component of the fungal cell wall (Ruiz-Herrera and Xoconostle-Cazares, 1995), but it is not present in plants. We purify these GH18 enzymes and characterize their hydrolytic activities on polysaccharides and their effect on fungal growth. We also inactivate each gene in

Received 11 January, 2014; accepted 26 April, 2014. *For correspondence. E-mail atolonen@genoscope.cns.fr; Tel. (+33) 0160 878477; Fax (+33) 0160 872514.

C. phytofermentans to directly examine its role during growth on plant and fungal biomass. Based on our results, we discuss the importance of fungal lysis to the ecology of cellulolytic bacteria such as *C. phytofermentans* that decompose plant biomass in forest soils.

Results

We observed that *C. phytofermentans* lysed and consumed soil fungi in cellulose cultures, thereby increasing cell densities and cellulose degradation (Fig. 1). Specifically, *A. niger* and *T. harzianum* were rapidly lysed while *C. cladosporioides* largely resisted degradation (Fig. 1A). Fungal cell mass remained stable or increased slightly in control cultures lacking *C. phytofermentans* and no cellulose degradation was observed, likely because of the low oxygen and nutrient concentrations in GS2 medium. Lysis of *A. niger* and *T. harzianum* stimulated *C. phytofermentans* cell densities twofold to fivefold relative to cultures without fungi (Fig. 1B). Higher cell densities in cultures containing fungi accelerated cellulose degradation (Fig. 1C). Enzymes secreted by *C. phytofermentans* during cellulose fermentation thus hydrolyse fungal hyphae and this bacterium obtains nutrition from lysed fungal biomass. Further, addition of fungal biomass to *C. phytofermentans* cultures represents a novel means to improve cellulolysis by boosting cell densities without repressing cellulase expression, as would occur by adding sugars.

We investigated the role in fungal lysis of the GH18 enzymes Cphy1799 and Cphy1800, the most highly expressed CAZy during cellulosic fermentation (Tolonen *et al*., 2011). These proteins both contain GH18 domains, N-terminal signals for Sec-dependent secretion (SP), two fibronectin type III (Fn3) domains (Watanabe *et al*., 1994) and a family 12 carbohydrate-binding domain (CBM12) that could bind chitin (Watanabe *et al*., 1994) and cellulose (Morimoto *et al*., 2007). The seven *C. phytofermentans* GH18 domains align with other clostridial GH18 (Supporting Information Fig. S4) to form several clusters (Fig. 2B). Cluster A contains Cphy1800 and Cphy2572 along with clostridial endochitinases active on crystalline chitin including *C. paraputrificum* ChiC (Morimoto *et al*., 2007) and ChiA of the *C. thermocellum* cellulosome (Zverlov *et al*., 2002). Cphy1799 is the only clostridial member of cluster B, a group of enzymes common in Actinobacteria such as *Streptomyces coelicolor* Chi18bA, which is active on soluble chitin substrates (Kawase *et al*., 2006). The four other *C. phytofermentans* GH18 are common in clostridia and are similar to the bacillus spore cortex lytic enzymes YaaH and YdhD, suggesting they are involved in sporulation.

We examined the regulation of *cphy1799* and *cphy1800* transcription by quantifying mRNA levels on different carbon sources (Fig. 2C). Both genes are similarly upregulated on cellulose and chitin and repressed by glucose, supporting they are subject to glucose-mediated carbon catabolite repression (CCR). CCR could be mediated by a CcpA repressor as in other clostridia (Abdou *et al*., 2008; Antunes *et al*., 2012); *C. phytofermentans* encodes 20 LacI/GalR repressors similar to CcpA. Although inducer exclusion is considered less important in Firmicutes (Deutscher, 2008), it could also have a role in repressing these enzymes by CCR. GH18 expression continuously increased from 1 h to 8 h after glucose cultures were transferred to minus-carbon medium (Supporting Information Fig. S5), but did not reach levels seen in cellulose and chitin cultures (Fig. 2C). While additional mechanisms may curtail gene expression during carbon

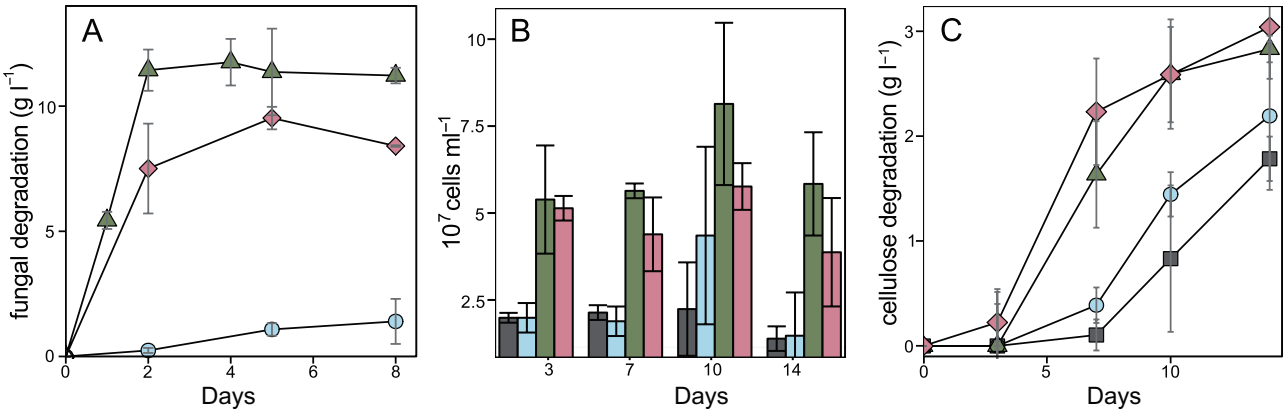

Fig. 1. Fungal lysis by *C. phytofermentans* in cellulose cultures promotes growth and accelerates cellulose degradation. (A) Rates of fungal lysis by *C. phytofermentans* in cellulose cultures. *C. phytofermentans* (B) cell densities and (C) cellulose degradation rates in cellulose cultures with and without fungi. Treatments: no fungi (gray, squares), *C. cladosporioides* (blue, circles), *T. harzianum* (pink, diamonds) and *A. niger* (green, triangles). Error bars show one standard deviation.

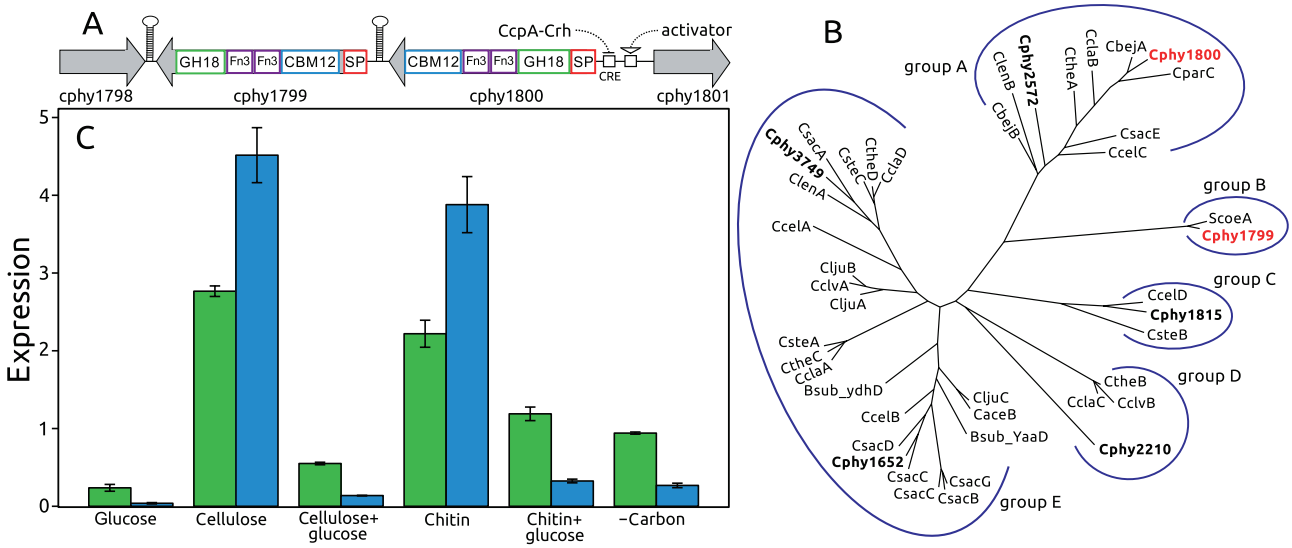

Fig. 2. A. Cphy1799 and Cphy1800 genome organization showing protein domains, transcriptional terminators and proposed dual regulation by a repressor (CcpA-Crh) and a transcription activator.
B. Maximum likelihood tree of clostridial GH18 protein sequences shows that sequences similar to Cphy1800 are found in other clostridia. Cphy1799 clusters with *Streptomyces* and is not found in other clostridia (see supplementary information for protein sequences and alignment).
C. mRNA expression of *cphy1799* (green) and *cphy1800* (blue) is upregulated on cellulose and chitin, repressed by glucose, and intermediate in absence of carbon. mRNA expression was measured by qRT-PCR and quantified as $2^{-\Delta Ct}$ normalized to 16S rRNA expression; error bars show one standard deviation.

limitation, the modest GH18 expression in minus-carbon cultures suggests that maximal transcription requires an inducer (likely a chito- or cellodextrin), similar to CAZy in other clostridia (Lin and Wilson, 1987; Newcomb *et al.*, 2007). Much work is needed to understand CAZy regulation in *C. phytofermentans*, but we propose that CcpA repression and activation by a polysaccharide-derived inducer could both play roles in the regulation of these GH18 genes (Fig. 2A).

The activities of purified Cphy1799 and Cphy1800 on 14 polysaccharides show that hydrolysis is specific to chitin (Fig. 3A). Efficient chitinolysis occurs by endochitinases randomly cleaving chitin into short chains, from which exochitinases progressively release chitobiose. Quantification of the exo- and endohydrolytic activities using 4-methylumbelliferyl (4MU)-conjugated chitin substrates revealed that the primary activity of Cphy1799 is exochitinolytic, Cphy1800 is endochitinolytic, and that a mix of the two enzymes is required to efficiently cleave both substrates (Fig. 3B). We also tested the ability of purified Cphy1799 and Cphy1800 to inhibit fungal growth. Treatment with Cphy1799 and Cphy1800 reduced growth of both *C. cladosporioides* (Fig. 3C and D) and *T. harzianum* (Fig. 3E and F). Thus, *C. cladosporioides* is inhibited by these enzymes on solid medium even though it resists lysis in liquid culture. Colony sizes of *A. niger* were similar to no enzyme controls, but enzyme-treated plates had reduced growth apart from the initial colony (Supporting Information Fig. S6), suggesting that GH18 reduced spore propagation or germination.

To directly examine the roles of these enzymes in growth and fungal degradation, we made *C. phytofermentans* gene inactivation strains by inserting group II introns into the 5′ end of the GH18 catalytic domains of *cphy1799* (strain AT03) and *cphy1800* (strain AT04) (Fig. 4A). As this is only the second report of targeted genetics in *C. phytofermentans*, we took care to ensure that the group II intron insertions were both accurate and specific and we examined their effects on gene expression. Southern blots of AT03 (Fig. 4B) and AT04 (Fig. 4C) probed with intron fragments show the mutant strains have single intron insertions. Polymerase chain reaction (PCR) amplification (Fig. 4D) and sequencing of intron-genome junction fragments (Supporting Information Fig. S4) confirmed the introns inserted into the genome at the desired locations. In the wildtype strain, *cphy1799* and *cphy1800* are coexpressed as a 4.9 kb transcript that is not expressed on glucose, lowly expressed on xylan and highly expressed on cellulose (Fig. 4E and F). Expression of the *cphy1799-cphy1800* transcript was not perturbed by intron insertion, but a 2.2 kb *cphy1800* transcript was abolished (Fig. 4F). Similar to the *C. cellulolyticum* cip-cel hydrolase cluster (Maamar *et al.*, 2006), *C. phytofermentans* likely expresses these GH18 genes on alternative transcripts with different stabilities to fine-tune their expression.

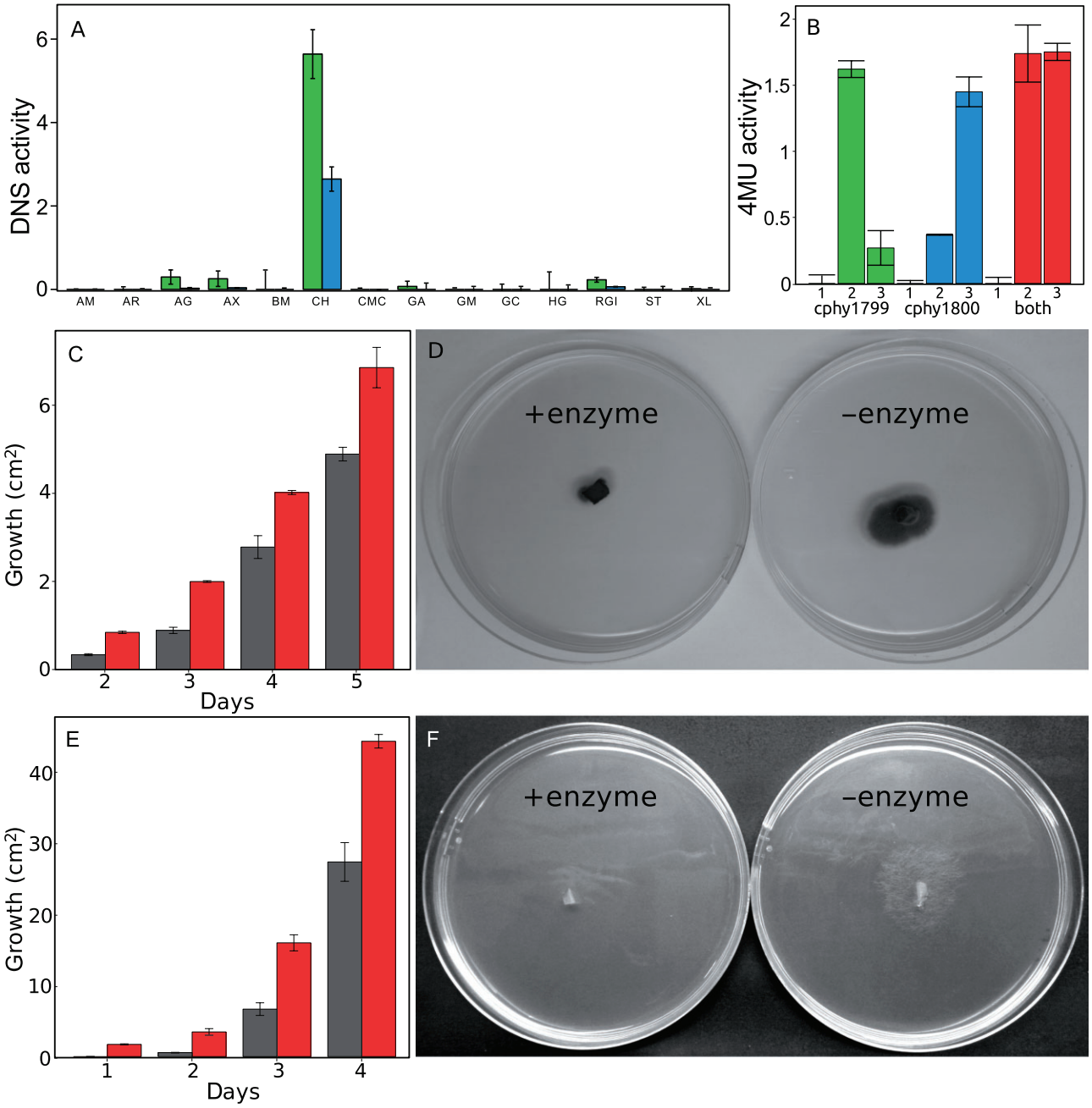

Fig. 3. Cphy1799 and Cphy1800 hydrolyse chitin and inhibit fungal growth. (A) Purified Cphy1799 (green) and Cphy1800 (blue) hydrolyse chitin (CH), but not other polysaccharides: alpha-mannan (AM), arabinan (AR), arabinogalactan (AG), arabinoxylan (AX), beta-mannan (BM), carboxymethylcellulose (CMC), galactan (GA), galactomannan (GM), glucomannan (GC), homogalacturonan (HG), rhamnogalacturonan I (RGI), starch (ST) and xylan (XL). (B) Activities of Cphy1799 (green), Cphy1800 (blue) and a 1:1 mix of both enzymes (red) on glucosamidine (1), chitobiose (2), and chitotriose (3) 4MU substrates. Fungi *C. cladosporioides* (C,D) and *T. harzianum* (E,F) were incubated with Cphy1799 and Cphy1800 (red bars) or no enzyme (gray bars), plated, and fungal colony area was quantified daily. Bars are means of triplicate plates. 4MU activity is μg 4MU released mg enzyme^{-1} second^{-1} and 3,5-dinitrosalicylic acid (DNS) activity is mM reducing sugar produced mg enzyme^{-1} minute^{-1}. All error bars are one standard deviation.

The GH18-inactivated strains grow similar to wildtype in medium with glucose (Fig. 5A), xylan (Fig. 5B), cellulose (Fig. 5C) and corn stover (Fig. 5D) as the sole carbon source. Thus, even though these are the two most highly expressed hydrolases during growth on cellulose, they do not contribute to hydrolysis of cellulose or these other plant substrates. However, both mutant strains have lost the ability of wildtype to solubilize chitin (Fig. 5E and F). These GH18 are thus the principal chitinases in the genome and both are required for chitinolysis. Although

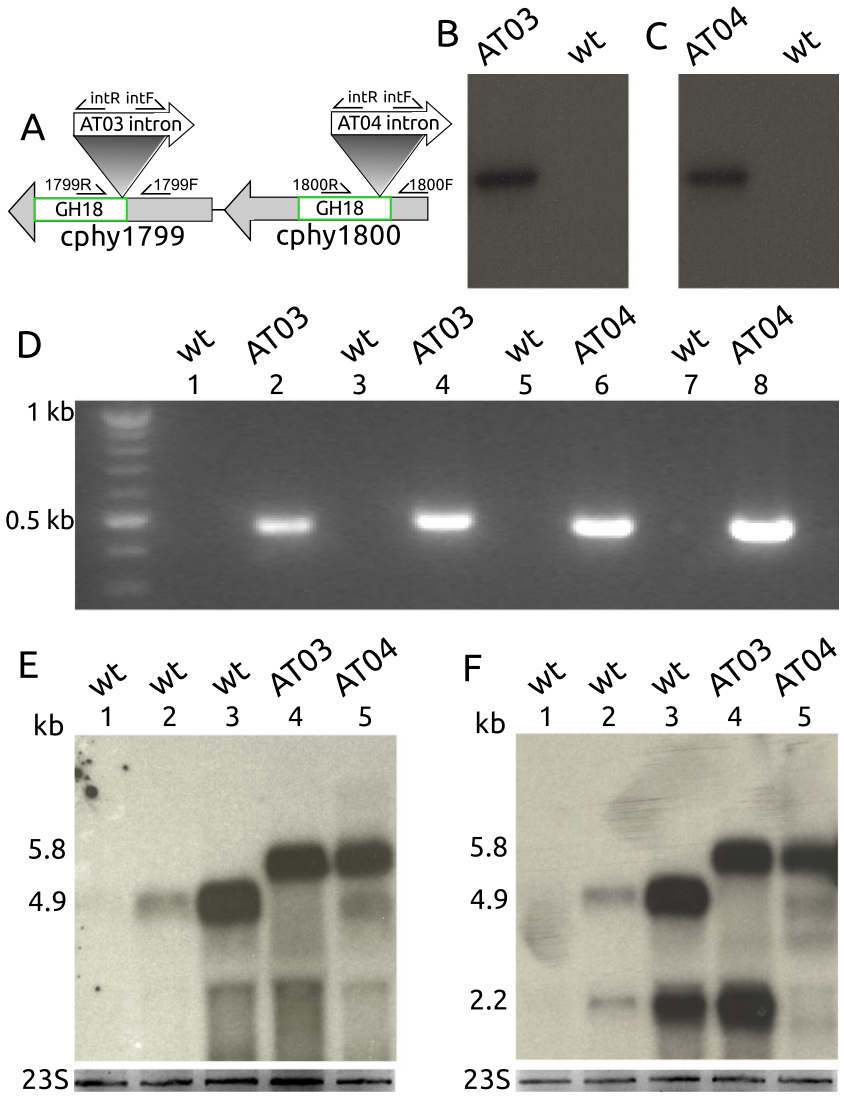

Fig. 4. Accurate, specific inactivation of *cphy1799* and *cphy1800* by group II intron insertions and their effects on mRNA expression. (A) The *cphy1799-cphy1800* operon showing intron insertion sites and the location of primers used to PCR amplify genome-intron junctions. Southern blots using an intron probe show single insertions in (B) AT03 and (C) AT04, but not in wildtype. (D) PCR of genome-intron junctions localize intron insertions AT03 and AT04 strains. Lanes 1 and 2: 5′ *cphy1799*-intron junction. Lanes 3 and 4: 3′ *cphy1799*-intron junction. Lanes 5 and 6: 5′ *cphy1800*-intron junction. Lanes 7 and 8: 3′ *cphy1800*-intron junction. Northern blot of (E) *cphy1799* and (F) *cphy1800* mRNA expression of wildtype growing on glucose (lane 1), xylan (lane 2) and cellulose (lane 3) and AT03 (lane 4) and AT04 (lane 5) strains growing on cellulose. Ethidium-stained 23S rRNA shows equal RNA loading in each lane.

C. phytofermentans can solubilize chitin flakes, it grows poorly on colloidal chitin and on the chitin monomer, GlcNAc (Fig. 5G) as carbon sources. GlcNAc is an amino sugar that can also serve as a microbial nitrogen source in N-limited soils (Reguera and Leschine, 2001). However, we found that replacing ammonium with GlcNAc resulted in reduced growth similar to minus-ammonium controls (Fig. 5H), supporting that *C. phytofermentans* cannot assimilate nitrogen from GlcNAc. Fungal lysis may stimulate *C. phytofermentans* growth by providing access to other components of the fungal cell wall such as glucans, mannans and glycoproteins (Bowman and Free, 2006).

Both GH18 mutants are less able than wildtype to hydrolyse *T. harzianum* and *A. niger* (Fig. 6) in liquid cellulose cultures. More than half of the *T. harzianum* biomass was solubilized by wildtype after 2 days, whereas this took 9 days in the GH18 mutant cultures (Fig. 6A). Most of the *A. niger* biomass was hydrolysed in 1 day by wildtype, but not until 3–4 days by the GH18 mutants (Fig. 6B), leading to visible differences in *A. niger* remaining in culture after 2 days (Fig. 6C). Although GH18 mutants hydrolyse fungi more slowly than wildtype, the fungi were ultimately lysed in these cultures. Fungal lysis by *C. phytofermentans* thus involves additional enzymes.

Discussion

Plant litter is degraded in forest soil by a community of cellulolytic microbes that is often dominated by fungi. Here we show that the plant-fermenting soil bacterium *C. phytofermentans* rapidly lyses and consumes fungi during growth on cellulose, stimulating growth and

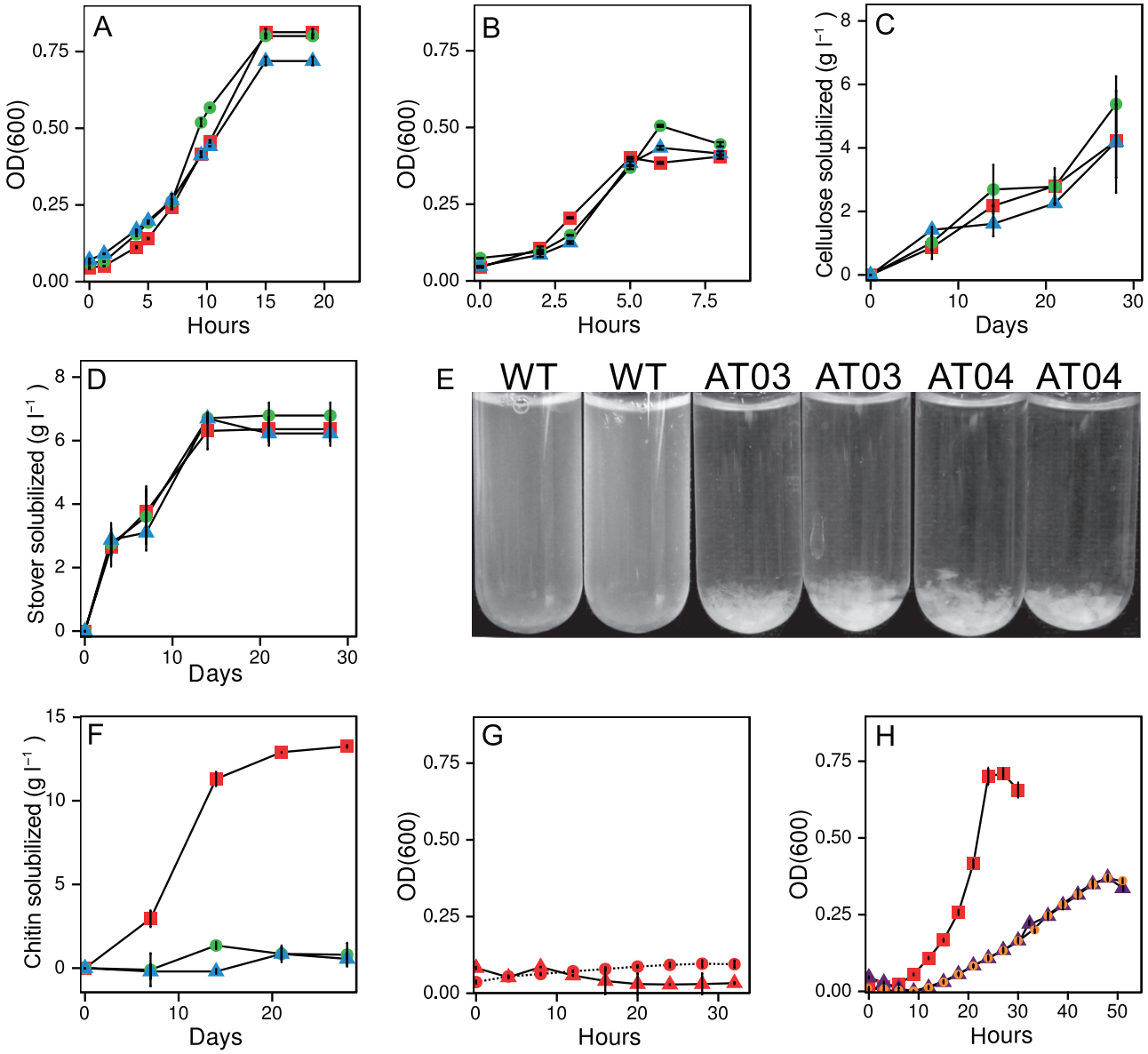

Fig. 5. GH18-inactivated strains AT03 (green circles) and AT04 (blue triangles) grow similar to wildtype (WT, red squares) on (A) glucose, (B) xylan, (C) cellulose and (D) corn stover, but have lost the ability of the WT strain to solubilize (E,F) chitin. (E) Chitin flakes are solubilized in wildtype cultures after 2 weeks, but remain unchanged in AT03 and AT04 cultures. (G) WT grows poorly on either colloidal chitin (dashed line) or the chitin monomer GlcNAc (solid line) as a carbon source. (H) WT cannot assimilate GlcNAc as an nitrogen source as shown by growth in MQM5.1 glucose medium with ammonium (red squares), minus ammonium (orange circles) and minus ammonium, plus GlcNAc (purple triangles). The limited growth in minus ammonium medium is likely due to the amino acids in MQM5.1. Growth curves are means of duplicate cultures; error bars are one standard deviation.

accelerating cellulose degradation (Fig. 1). *Clostridium phytofermentans* expresses at least 87 CAZy on cellulose (Tolonen *et al.*, 2011) and here we implicate the two most highly expressed CAZy, Cphy1799 and Cphy1800, in fungal lysis. Expression of these genes increases greatly on cellulose (Figs 2C and 4E–F), but purified enzymes are inactive on plant polysaccharides (Fig. 3A) and mutants lacking either enzyme grow normally on cellulose and other plant substrates (Fig. 5B–D). These enzymes perform complementary roles in chitinolysis (Fig. 3B) and both GH18 mutants have lost the ability to solubilize chitin (Fig. 5E–F). The high chitinolytic activities of these enzymes enables them to inhibit fungal growth (Fig. 3C–F) and GH18 mutants are less able to lyse fungi compared with wildtype (Fig. 6). We thus conclude that these GH18 enzymes are highly expressed on cellulose to lyse chitin in the fungal cell wall rather than plant substrates. Fungal lysis could increase

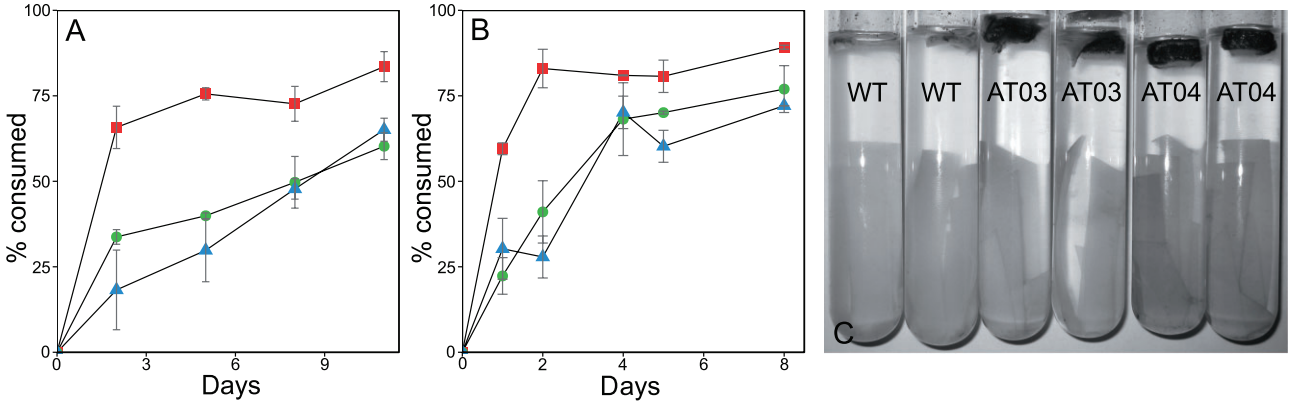

Fig. 6. GH18-inactivated *C. phytofermentans* strains AT03 (green circles) and AT04 (blue triangles) are less able than wildtype (WT, red squares) to lyse the fungi (A) *T. harzianum* and (B,C) *A. niger* in GS2 cellulose cultures. Percent remaining fungal biomass was determined by collecting fungal mass on filters and comparing to the initial mass. (C) *C. phytofermentans-A. niger* co-cultures after 2 days show increased fungal lysis in WT cultures relative to GH18 mutants. Error bars show one standard deviation.

C. phytofermentans fitness during cellulolysis by killing other cells, removing cell mass (i.e. fungal hyphae) that block access to cellulose and liberating nutrients.

Similar to Cphy1799 and Cphy1800, other bacterial GH18 have been shown to inhibit fungal spore germination, germ tube elongation, and to burst spore and hyphal tips. For example, *Bacillus circulans* WL-12 ChiA, which is similar to Cphy1800, inhibits spore germination by *Botrytis elliptica* (Chen *et al.*, 2004). While GH18 mutants lysed fungi more slowly than wildtype (Fig. 6), the mutants did ultimately degrade the fungi, supporting that other, less highly expressed enzymes are also involved. For example, Cphy1943 (GH19) is similar to the *S. coelicolor* Chi19F, which hydrolyses colloidal chitin and inhibits fungal growth, likely by lysing fungal hyphal tips where chitin is not yet crystallized (Kawase *et al.*, 2006). GH18 such as Cphy1799 and Cphy1800 that hydrolyse crystalline chitin (Fig. 5E and F) could complement GH19 by degrading mature hyphal filaments. *Clostridium phytofermentans* expresses other fungal-specific CAZy such as the GH65 Cphy1874 that hydrolyses nigeran, a glucan found in hyphal wall of *A. niger* (Nihira *et al.*, 2012).

We examined the effect of purified GH18 enzymes and *C. phytofermentans* on three, cosmopolitan, model fungi from forest soil. Purified GH18 inhibit growth of these fungi during robust, aerobic growth on YPD medium (Fig. 3C–F, Supporting Information Fig. S6). Further, *C. phytofermentans* rapidly lysed and consumed *A. niger* and *T. harzianum* (Fig. 1) in GS2 medium under anaerobic conditions (< 1% oxygen). While *A. niger* grows slowly at low oxygen concentration (0.5%) (Hall and Denning, 1994) and *T. harzianum* has been isolated from anaerobic soil (Kurakov *et al.*, 2008), their limited growth in controls lacking *C. phytofermentans* show that these fungi prefer different soil habitats than *C. phytofermentans*. We thus propose a few ecological scenarios by which *C. phytofermentans* may interact with these and other fungi in soil. *Clostridium phytofermentans* may degrade aerobic and facultatively anaerobic fungi at the oxygen interface or when these fungi infiltrate anaerobic soil. Soil is a temporally dynamic environment with fluctuating oxygen conditions; aerobic fungi may proliferate when oxygen rises whereas *C. phytofermentans* grows and degrades these fungi when oxygen is depleted. Finally, *C. phytofermentans* may interact with anaerobic fungi. To the best of our knowledge, an anaerobic fungus that grows in the same conditions as *C. phytofermentans* has not yet been isolated. However, anaerobic fungi exist in the cow rumen (Lowe *et al.*, 1987) where they are inhibited by chitinolytic bacteria (Kopecný *et al.*, 1996) and facultatively anaerobic cellulolytic fungi have been reported in soil (Durrant *et al.*, 1995).

The complex and diverse community of soil microbes that degrade plant litter likely interact to benefit and harm one another in many unknown ways. This study shows the cellulolytic bacterium *C. phytofermentans* can consume fungi to increase cell abundances when fermenting cellulose. Indeed, two antifungal enzymes are secreted more abundantly than cellulases when fermenting cellulose. In addition to playing an important role in *C. phytofermentans* ecology, these GH18 enzymes could have various industrial applications (Bhattacharya *et al.*, 2007) and could affect product yields when fermenting cellulosic substrates with bacterial-fungal consortia (Zuroff *et al.*, 2013). Cellulolytic clostridia such as *C. phytofermentans* encode hundreds of hydrolases with diverse functions, but much work is needed to understand their roles in soil ecosystems and to facilitate their use in biotechnology.

Experimental procedures

Clostridium phytofermentans ISDg ATCC 700394 was cultured anaerobically at 30°C in GS2 medium (Johnson *et al*., 1981) or defined MQM5.1 medium (Lee *et al*., 2012) with either 3 g l^{-1} glucose, xylan (Sigma X0502), colloidal chitin (Sigma C9752), GlcNAc (Sigma A3286) or 15 g l^{-1} chitin (Sigma C9752), cellulose (Whatman filter paper cat 1001-090), corn stover or fungus. Colloidal chitin was prepared as described in Usharani and Gowdaa (2011). Cellulose and corn stover were cut into 0.5 × 5 cm strips. Degradation of insoluble substrates (chitin, cellulose, corn stover and fungi) was quantified as dry mass remaining in culture after vacuum collection on 11 μm filters. To test if *C. phytofermentans* assimilates nitrogen from GlcNAc, cultures were grown in MQM5.1 glucose medium, centrifuged, washed in MQM5.1 lacking ammonium and transferred to MQM5.1 glucose medium containing either 15 mM ammonium (standard recipe), minus ammonium or minus ammonium plus 15 mM GlcNAc. *Aspergillus niger*, *T. harzianum* and *C. cladosporioides* were grown in liquid PD medium (4 g l^{-1} potato extract, 20 g l^{-1} dextrose) or on PDA plates (PD plus 15 g l^{-1} agar) at 25°C. Liquid fungal cultures were inoculated using spores suspended from PDA plates. Fungal hyphae from liquid PD cultures were weighed and transferred to *C. phytofermentans* cellulose cultures for fungal degradation experiments.

mRNA expression was quantified by quantitative reverse transcription PCR (qRT-PCR) using primers in Supporting Information Table S1 and the methods previously described (Tolonen *et al*., 2011). Expression values are means of triplicate cultures calculated as $2^{-\Delta Ct}$, normalized to 16S rRNA levels. Carbon source treatments were sampled as follows: glucose (mid-log phase), cellulose or chitin (1-week growth), cellulose + glucose and chitin + glucose (1-week growth, addition of 0.3 g l^{-1} glucose, and sampled after 8 h incubation) and minus-carbon (mid-log phase glucose cultures resuspended in GS2 with no carbon and sampled after 1 h, 3 h, 6 h and 8 h incubation).

Mature forms of *cphy1799* and *cphy1800* lacking N-terminal secretion signals were cloned with C-terminal His-tags into pET-22b(+) by Ligation Independent Cloning (Aslanidis and de Jong, 1990) (primers in Supporting Information Table S1) and inserts were sequence verified (Supporting Information Fig. S1). Plasmids were transformed into BL-21 (DE3) cells (Novagen 71012) and expression was induced by 500 μM IPTG. Cells were lysed using Lysonase Bioprocessing Reagent (Novagen 71230) and an Ultrasonic Processor (Thomas Scientific). Enzymes were purified using Ni-NTA Spin columns (Qiagen 31314) and imidazole was removed by dialysing three times in 0.1 M phosphate buffer pH7 using Amicon Centrifugal filter units (Millipore UFC801024). Purified enzymes were resolved by SDS-PAGE (Supporting Information Fig. S2) and concentrations were determined by Bradford assay.

Chitinase activities were assayed using three 4MU-labelled chitin substrates that fluoresce upon cleavage (Sigma CS1030): 4-MU-$(GlcNAc)_3$ (endochitinase substrate), 4-MU-$(GlcNAc)_2$ (exochitinase substrate) and 4-MU-GlcNAc (glucosaminidase substrate). Proteins and 4MU-chitin substrates were incubated at 37°C for 30 min and fluorescence was measured at 360 nm excitation, 450 nm emission. Fluorescence measurements were normalized to protein concentrations measured by Bradford assay. Enzyme activities on purified polysaccharides were quantified by incubating for 30 min at 37°C and then using the 3,5-dinitrosalicylic acid method as previously described (Tolonen *et al*., 2011). The effect of GH18 on fungal growth was quantified by cutting a 25 mm^2 agar square from within the fungal colony on a PDA agar plate and incubating for 6 h at 37°C in 0.1 M phosphate buffer pH7 with or without 2.5 mg ml^{-1} enzyme (equal parts Cphy1799, Cphy1800). The agar chunk was transferred to a fresh PDA plate and fungal growth was measured from plate images as colony surface areas using ImageJ (Schneider *et al*., 2012a).

Clostridium phytofermentans mutants were made by insertional inactivation of c*phy1799* (strain AT03) and *cphy1800* (AT04 strain) with the group II intron in plasmid pQint (Tolonen *et al*., 2009). The intron was targeted to insert in the antisense orientation 756 bp from the start of *cphy1799* or 466 bp from the start of *cphy1800* by two-step, cross-over PCR (primers in Supporting Information Table S1). Targeted pQint plasmids were transferred into *C. phytofermentans* by conjugal transfer from *Escherichia coli* (Tolonen *et al*., 2009). The location of intron insertions in strains AT03 and AT04 were confirmed by PCR of genome-intron junction fragments (primers in Supporting Information Table S1) and sequencing (Supporting Information Fig. S3).

RNA for Northern blots was isolated from 10 ml glucose and xylan cultures and 100 ml cellulose cultures with an Ambion Ribopure Bacteria Kit (AM1925). DNA was removed with DNase I (Ambion AM2222), RNA was ethanol precipitated and 10 μg RNA was resolved on a 1% agarose formaldehyde gel. The gel was shaken in 0.05 M NaOH for 30 min, followed by 30 min in 0.1 M Tris pH 7.5 and 20 min in 1× saline sodium citrate (SSC). RNA was transferred to a Genescreen (Perkin Elmer NEF987001PK) membrane and cross-linked in a UV stratalinker. The 500 bp probes of *cphy1799* and *cphy1800* were prepared by PCR using primers cphy1799probeF,R and cphy1800probeF,R (Supporting Information Table S1). The probe was labelled with ^{32}P dATP using Klenow and random hexamers. The probe was hybridized to the blot for 6 h at 42°C; the blot was washed in SSC, and imaged using a phosphoimager. Southern blotting with an intron probe amplified by primers intronProbeF,R (Supporting Information Table S1) used the methods previously described (Tolonen *et al*., 2009).

Acknowledgements

We thank J Dupont for fungal cultures, D Spatt for technical expertise and D LePaslier for reading the manuscript. This work was funded by Genoscope, a CNRS Chaire d'Excellence to ACT and DOE GtL Grant No. DE-FG02-02ER-63445 to GMC. The authors declare no competing commercial interests in relation to this work.

References

Abdou, L., Boileau, C., de Philip, P., Pagès, S., Fiérobe, H.-P., and Tardif, C. (2008) Transcriptional regulation of the *Clostridium cellulolyticum* cip-cel operon: a complex mechanism involving a catabolite-responsive element. *J Bacteriol* **190:** 1499–1506.

Antunes, A., Camiade, E., Monot, M., Courtois, E., Barbut, F., Sernova, N.V., *et al*. (2012) Global transcriptional control by glucose and carbon regulator CcpA in *Clostridium difficile*. *Nucleic Acids Res* **40:** 10701–10718.

Aslanidis, C., and de Jong, P.J. (1990) Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res* **18:** 6069–6074.

Bhattacharya, D., Nagpure, A., and Gupta, R.K. (2007) Bacterial chitinases: properties and potential. *Crit Rev Biotechnol* **27:** 21–28.

Bowman, S.M., and Free, S.J. (2006) The structure and synthesis of the fungal cell wall. *Bioessays* **28:** 799–808.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* **37:** D233–D238.

Chen, C.-Y., Wang, Y.-H., and Huang, C.-J. (2004) Enhancement of the antifungal activity of *Bacillus subtilis* F29-3 by the chitinase encoded by *Bacillus circulans chiA* gene. *Can J Microbiol* **50:** 451–454.

Deutscher, J. (2008) The mechanisms of carbon catabolite repression in bacteria. *Curr Opin Microbiol* **11:** 87–93.

Durrant, L.R., Canale-Parola, E., and Leschine, S.B. (1995) Facultatively anaerobic cellulolytic fungi from soil. In *The Significance and Regulation of Soil Biodiversity*. Collins, H.P., Robertson, G.P., and Klug, M.J. (eds). The Netherlands: Springer, pp. 161–167.

Hall, L.A., and Denning, D.W. (1994) Oxygen requirements of *Aspergillus* species. *J Med Microbiol* **41:** 311–315.

Johnson, E.A., Madia, A., and Demain, A.L. (1981) Chemically defined minimal medium for growth of the anaerobic cellulolytic thermophile *Clostridium thermocellum*. *Appl Environ Microbiol* **41:** 1060–1062.

Kawase, T., Yokokawa, S., Saito, A., Fujii, T., Nikaidou, N., Miyashita, K., and Watanabe, T. (2006) Comparison of enzymatic and antifungal properties between family 18 and 19 chitinases from *S. coelicolor* A3(2). *Biosci Biotechnol Biochem* **70:** 988–998.

Kopecný, J., Hodrová, B., and Stewart, C.S. (1996) The effect of rumen chitinolytic bacteria on cellulolytic anaerobic fungi. *Lett Appl Microbiol* **23:** 199–202.

Kurakov, A.V., Lavrent'ev, R.B., Nechitaĭlo, T.I., Golyshin, P.N., and Zviagintsev, D.G. (2008) Diversity of facultatively anaerobic microscopic mycelial fungi in soils. *Mikrobiologiia* **77:** 103–112.

Lee, S.J., Warnick, T.A., Pattathil, S., Alvelo-Maurosa, J.G., Serapiglia, M.J., McCormick, H., *et al*. (2012) Biological conversion assay using *Clostridium phytofermentans* to estimate plant feedstock quality. *Biotechnol Biofuels* **5:** 5. doi: 10.1186/1754-6834-5-5

Lin, E., and Wilson, D.B. (1987) Regulation of β-1,4-endoglucanase synthesis in *Thermomonospora fusca*. *Appl Environ Microbiol* **53:** 1352–1357.

Lowe, S.E., Theodorou, M.K., and Trinci, A.P. (1987) Cellulases and xylanase of an anaerobic rumen fungus grown on wheat straw, wheat straw holocellulose, cellulose, and xylan. *Appl Environ Microbiol* **53:** 1216–1223.

Lynd, L.R., Weimer, P.J., van Zyl, W.H., and Pretorius, I.S. (2002) Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev* **66:** 506–577, table of contents.

Maamar, H., Abdou, L., Boileau, C., Valette, O., and Tardif, C. (2006) Transcriptional analysis of the cip-cel gene cluster from *Clostridium cellulolyticum*. *J Bacteriol* **188:** 2614–2624.

Morimoto, K., Yoshimoto, M., Karita, S., Kimura, T., Ohmiya, K., and Sakka, K. (2007) Characterization of the third chitinase Chi18C of *Clostridium paraputrificum* M-21. *Appl Microbiol Biotechnol* **73:** 1106–1113.

Newcomb, M., Chen, C.-Y., and Wu, J.H.D. (2007) Induction of the celC operon of *Clostridium thermocellum* by laminaribiose. *Proc Natl Acad Sci USA* **104:** 3747–3752.

Nihira, T., Nakai, H., Chiku, K., and Kitaoka, M. (2012) Discovery of nigerose phosphorylase from *Clostridium phytofermentans*. *Appl Microbiol Biotechnol* **93:** 1513–1522.

Reguera, G., and Leschine, S.B. (2001) Chitin degradation by cellulolytic anaerobes and facultative aerobes from soils and sediments. *FEMS Microbiol Lett* **204:** 367–374.

Ruiz-Herrera, J., and Xoconostle-Cazares, B. (1995) Molecular and genetic control of chitin biosynthesis in fungi. *Arch Med Res* **26:** 315–321.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012a) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9:** 671–675.

Schneider, T., Keiblinger, K.M., Schmid, E., Sterflinger-Gleixner, K., Ellersdorfer, G., Roschitzki, B., *et al*. (2012b) Who is who in litter decomposition? Metaproteomics reveals major microbial players and their biogeochemical functions. *ISME J* **6:** 1749–1762.

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.

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

Usharani, T.R., and Gowdaa, T.K.S. (2011) Cloning of chitinase gene from *Bacillus thuringiensis*. *Indian J Biotechnol* **10:** 264–269.

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.

Watanabe, T., Ito, Y., Yamada, T., Hashimoto, M., Sekine, S., and Tanaka, H. (1994) The roles of the C-terminal domain and type III domains of chitinase A1 from *Bacillus circulans* WL-12 in chitin degradation. *J Bacteriol* **176:** 4465–4472.

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

Zverlov, V.V., Fuchs, K.-P., and Schwarz, W.H. (2002) Chi18A, the endochitinase in the cellulosome of the thermophilic, cellulolytic bacterium *Clostridium thermocellum*. *Appl Environ Microbiol* **68:** 3176–3179.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Sequences of Cphy1799 and Cphy1800 with C-terminal His tags cloned into pET-22b(+). Cphy1799 and Cphy1800 are both secreted proteins with N-terminal signal peptides that are removed upon export from the cell. SignalP (http://www.cbs.dtu.dk/services/SignalP/) predicts that the signal peptide of Cphy1799 is cleaved at residue 32 and Cphy1800 at residue 31. We amplified the mature forms (lacking N-terminal secretion signals) of *cphy*1799 and *cphy*1800 with C-terminal His tags using primers shown in Supporting Information Table S1. Supporting Information Fig. S3 shows four alignments between expected and observed sequences: cphy1799 forward strand, cphy1799 reverse strand, cphy1800 forward strand and cphy1800 reverse strand.

Fig. S2. SDS-PAGE gel showing purified His-tagged Cphy1799 and Cphy1800 proteins. The ladder in lane 1 shows protein sizes in kDa. The expected size of Cphy1799 is 57 kDa and Cphy1800 is 71 kDa.

Fig. S3. Sequences of genome-intron junctions in strains AT03 and AT04. Transconjugant cultures were transferred in -erm medium to induce loss of pQint and then plated on GS2 agar medium. Colonies were picked and intron insertions were confirmed by sequencing. Highlighted regions show intron-junction fragment with intron (red) and genome (cyan).

Fig. S4. Alignment of clostridial GH18 sequences (species abbreviation and accession number) used for phylogenetic tree construction. Forty-one GH18 protein sequences from non-pathogenic clostridia (including the six GH18 in *C. phytofermentans*) were downloaded from the CAZy database (www.cazy.org). *Bacillus subtilis* YaaH and YdhD and *S. coelicolor* ChiA were included as well-characterized GH18 proteins for comparison. The catalytic domains, as defined by the GenBank accession entries, were aligned using the MUSCLE alignment algorithm (http://www.drive5.com/muscle/) and then improved manually. A maximum likelihood phylogenetic tree was created from the aligned GH18 sequences using the Seaview software package (http://pbil.univ-lyon1.fr/software/seaview.html) to implement the PhyML method using these parameters:

Running 'C:phyml_3.0.1_win32.exe' -d aa -m LG -b 100 -v 0.0 -c 4 -a e -s BEST --rand_start --n_rand_starts 5 -i C:seaviewtemp_5.phy

Species abbreviations in the GH18 alignment and tree: *Clostridium phytofermentans* ISDg (Cphy), *Clostridium paraputrificum* (Cpar), *Clostridium lentocellum* DSM 5427 (Clen), *Clostridium beijerinckii* NCIMB 8052 (Cbej), *Clostridium thermocellum* ATCC 27405 (Cthe), *Clostridium clariflavum* DSM 19732 (Ccla), *Clostridium cellulolyticum* H10 ATCC 35319 (Ccel), *Clostridium saccharolyticum* WM1 (Csac), *Clostridium ljungdahlii* DSM 13528 (Clju), *Clostridium cellulovorans* 743B (Cclv), *Clostridium acetobutylicum* ATCC 824 (Cace), *Clostridium stercorarium subsp. stercorarium* DSM 8532 (Cste), *Bacillus subtilis subsp. subtilis* str 168 (Bsub), *Streptomyces coelicolor* A3 (Scoe).

Fig. S5. mRNA expression of *cphy1799* (green) and *cphy1800* (blue) continuously increased over the 8 h time course following transfer to GS2 medium lacking supplemental carbon. mRNA expression was measured by qRT-PCR and quantified as $2^{-\Delta Ct}$ normalized to 16S rRNA expression; error bars show one standard deviation.

Fig. S6. Treatment of *Aspergillus niger* with GH18 enzymes reduces secondary colony formation. 25 mm^2 agar squares of *A. niger* were incubated in 0.1 M phosphate buffer without (–enzymes) or with 2.5 mg ml^{-1} GH18 enzymes, equal parts Cphy1799 and Cphy1800 (+ enzymes) and images were taken 5 days later. Duplicate plates for each treatment show GH18 enzymes did not affect growth from the central colony, but GH18-treated plates had no additional colonies for on the plate.

Table S1. Primer sequences and plasmids used in this study.

Appendix 4

A. C. Tolonen, T. R. Zuroff, M. Ramya, M. Boutard, **T. Cerisy**, and W. R. Curtis, "Physiology, Genomics, and Pathway Engineering of an Ethanol-Tolerant Strain of Clostridium phytofermentans.," Appl. Environ. Microbiol., vol. 81, no. 16, pp. 5440–8, 2015.

AEM
Journals.ASM.org

Physiology, Genomics, and Pathway Engineering of an Ethanol-Tolerant Strain of *Clostridium phytofermentans*

Andrew C. Tolonen,[a] Trevor R. Zuroff,[b]* Mohandass Ramya,[c] Magali Boutard,[a] Tristan Cerisy,[a] Wayne R. Curtis[b]

Genoscope-CEA, CNRS-UMR8030, Université d'Évry, Évry, France[a]; Department of Chemical Engineering, The Pennsylvania State University, University Park, Pennsylvania, USA[b]; Department of Genetic Engineering, SRM University, Kattankulathur, India[c]

Novel processing strategies for hydrolysis and fermentation of lignocellulosic biomass in a single reactor offer large potential cost savings for production of biocommodities and biofuels. One critical challenge is retaining high enzyme production in the presence of elevated product titers. Toward this goal, the cellulolytic, ethanol-producing bacterium *Clostridium phytofermentans* was adapted to increased ethanol concentrations. The resulting ethanol-tolerant (ET) strain has nearly doubled ethanol tolerance relative to the wild-type level but also reduced ethanol yield and growth at low ethanol concentrations. The genome of the ET strain has coding changes in proteins involved in membrane biosynthesis, the Rnf complex, cation homeostasis, gene regulation, and ethanol production. In particular, purification of the mutant bifunctional acetaldehyde coenzyme A (CoA)/alcohol dehydrogenase showed that a G609D variant abolished its activities, including ethanol formation. Heterologous expression of *Zymomonas mobilis* pyruvate decarboxylase and alcohol dehydrogenase in the ET strain increased cellulose consumption and restored ethanol production, demonstrating how metabolic engineering can be used to overcome disadvantageous mutations incurred during adaptation to ethanol. We discuss how genetic changes in the ET strain reveal novel potential strategies for improving microbial solvent tolerance.

The conversion of lignocellulosic biomass to fuels and commodities represents a large-scale, renewable alternative to petroleum. This multistep bioconversion is traditionally performed in a series of independent processes, but consolidated bioprocessing (CBP) is an alternative paradigm with potential economic advantages (1). In CBP, enzyme production, hydrolysis, and fermentation occur in a single reactor, leading to savings in capital and operating costs as well as increased efficiencies due to system synergies (2). Here we studied *Clostridium phytofermentans*, a promising CBP candidate that ferments plant biomass primarily to ethanol (3, 4). *C. phytofermentans* hydrolyzes pretreated corn stover (both glucans and xylans) with efficiencies similar to those seen with simultaneous saccharification and cofermentation (SSCF) using commercial enzymes and xylose-fermenting yeast (*Saccharomyces cerevisiae*) (5). Fermentation of pretreated corn stover by *C. phytofermentans* reaches a titer of 7 g liter^{-1} ethanol (6), and stable cocultures of *C. phytofermentans* and *S. cerevisiae* cdt-1 ferment ~70 g liter^{-1} cellulose to 22 g liter^{-1} ethanol (7), which is an ethanol concentration that reduces *C. phytofermentans* growth. Thus, application of CBP bacteria such as *C. phytofermentans* will likely require improving their solvent tolerances without compromising enzyme production or fermentation of soluble carbohydrates to ethanol.

Considerable effort has focused on adapting clostridia to increased solvent levels and investigating the genetic and physiological changes associated with adaptation to solvents (8–13). Other studies have shown increased ethanol production in clostridia that primarily produce fermentation products other than ethanol. *C. cellulolyticum* expressing pyruvate decarboxylase and alcohol dehydrogenase (ADH) overcame pyruvate accumulation and shifted fermentation products from lactate to acetate and ethanol (14). In *C. thermocellum*, redirection of carbon flow through pyruvate kinase (15), inactivation of lactate dehydrogenase and phosphotransacetylase (16), and deletion of hydrogenases (17) all improve ethanol production. These results demonstrate that, although the genetic tools are being developed only now, engineering improved ethanol production in cellulolytic clostridia is possible. However, development of strains that are ethanol tolerant (ET) and that also produce ethanol in high titers remains a significant challenge.

Here we sought to develop a strain of *C. phytofermentans* with both improved resistance and production of ethanol, particularly from cellulose. We isolated an ethanol-tolerant (ET) *C. phytofermentans* strain by serial transfer into increasing ethanol levels and characterized its growth and fermentation properties. We sequenced the ET strain genome to reveal genomic mutations that arose during adaptation and overcame reduced ethanol yield in the ET strain by heterologous expression of an alternative ethanol formation pathway. We discuss how the findings from this study improve our understanding of how microbes adapt to elevated concentrations of solvents such as ethanol.

Received 23 February 2015 **Accepted** 27 May 2015

Accepted manuscript posted online 5 June 2015

Citation Tolonen AC, Zuroff TR, Ramya M, Boutard M, Cerisy T, Curtis WR. 2015. 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.

Editor: J. L. Schottel

Address correspondence to Andrew C. Tolonen, atolonen@genoscope.cns.fr.

* Present address: Trevor R. Zuroff, Biodomain, Shell Technology Center, Houston, Texas, USA.

A.C.T. and T.R.Z. contributed equally to this article.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.00619-15.

doi:10.1128/AEM.00619-15

TABLE 1 Bacterial strains, primers, and plasmids used in this study

Strain name	Genotype or description[a]	Source or sequence
Strains		
Clostridium phytofermentans ISDg	ATCC type strain 700394	Susan Leschine Laboratory, University of Massachusetts, Amherst, Amherst, MA, USA
E. coli Top 10	*hsdR mcrA endA1 recA1 rpsL* (Str^r) (cloning strain)	Invitrogen Corporation
E. coli S17-1	RP4-2 (Km::Tn*7* Tc::Mu-1) *recA1 endA1* (conjugal strain)	Yale *E. coli* Stock Center
Primers		
pdcAdhB_F	Forward primer for amplification of *pdc* and *adhB* from pES120	5′-TTTTTCGAATTCACCGGATCCCTGCAGTAGGAGGAATTAACC-3′
pdcAdhB_R	Reverse primer for amplification of *pdc* and *adhB* from pES120	5′-ATATTTCGATCGATTGCATGCTTAGAAAGCGCTCAGGAAGAG-3′
pQexp_F	Forward primer to confirm *pdc-adhB* insertion in pQexp	5′-AAACCTAGGTAATTGAGGAAAGTTACAATTA-3′
pQexp_R	Reverse primer to confirm *pdc-adhB* insertion in pQexp	5′-GAATGGCGCCTGATGCG-3′
cphy3925F	Forward primer to amplify Cphy3925 coding sequence	5′-AAAGAAGGAGATAGGATCATGACGAAGAAAGTGGAATTA-3′
cphy3925R	Reverse primer to amplify Cphy3925 coding sequence	5′-GTGTAATGGATAGTGATCTTAATGGTGATGGTGATGATGTTTACCG TAGTACACTTTAAGATAG-3′
Plasmids		
pES120	Source of *Z. mobilis pdc* and *adhB* genes	Jay Keasling Laboratory, University of California, Berkeley, Berkeley, CA, USA
pQexp	Replicating plasmid for *C. phytofermentans*	Andrew Tolonen Laboratory, Genoscope-CEA, Évry, France
pQexpE	pQexp with *Z. mobilis pdc* and *adhB* cloned into the unique BamHI and PvuI sites	This study

[a] Str, streptomycin, Km, kanamycin; Tc, tetracycline.

MATERIALS AND METHODS

Culturing. *C. phytofermentans* ISDg (ATCC 700394) was grown anaerobically by preparing cultures in a Coy anaerobic chamber with a 1.5% H_2/98.5% N_2 atmosphere. Cultures were incubated without shaking at 30°C in GS2 medium (18) adjusted to a pH of 7 and supplemented with carbon sources as described elsewhere in the text. Growth kinetics were monitored by optical density at 600 nm (OD_{600}) in sealed 100-well microtiter plates (Bioscreen 9502550) as previously described (19); cultures were briefly shaken to resuspend cells before each optical density measurement. Cellulose, cellobiose, and glucose cultures for substrate consumption and fermentation product analysis were grown in 100-ml serum bottles, which were sealed with butyl rubber stoppers after degassing.

The *C. phytofermentans* ethanol-tolerant (ET) strain was selected by serial transfer (1:50 dilution) into culture tubes containing 10 ml of GS2 medium supplemented with increasing ethanol concentrations. Starting with cultures in 4% (vol/vol) (31.5 g liter^{-1}) ethanol, cultures were transferred weekly to fresh medium containing the same ethanol concentration and to medium with a 1%-higher ethanol concentration. If no growth was observed at the higher concentration after 1 week, cultures were retransferred to the same ethanol concentrations. If a culture grew at a higher ethanol concentration, this culture was transferred again to that ethanol concentration and to a 1%-greater ethanol concentration. Growth was observed at 5% ethanol after 7 weekly transfers, 6% after 13 transfers, and 7% after 19 transfers. Each time the ethanol tolerance improved, cells were

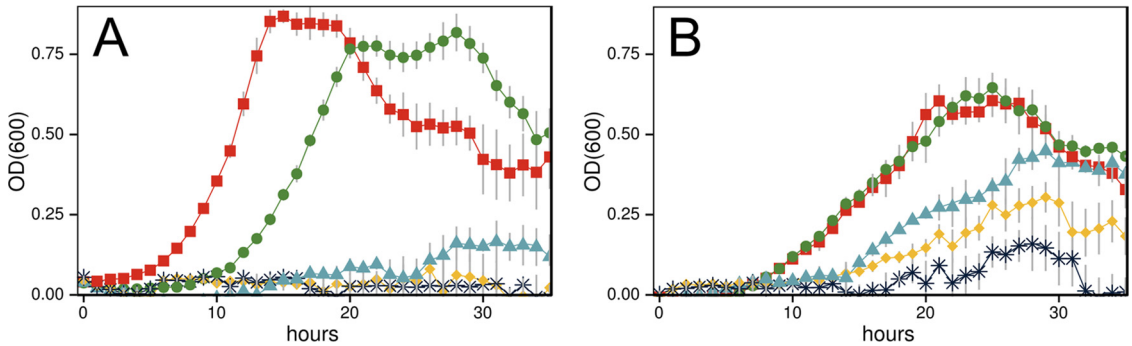

FIG 1 Growth of wild-type (WT) (A) and ethanol-tolerant (ET) (B) *C. phytofermentans* strains at 30°C in GS2 medium with 3 g liter^{-1} glucose supplemented with ethanol at the following levels (vol/vol): 0% (red squares), 2% (green circles), 4% (blue triangles), 6% (yellow diamonds), and 7% (black X's). Data points of growth (OD_{600}) represent means of results from triplicate 400-μl cultures in sealed microtiter plates; error bars represent 1 standard deviation.

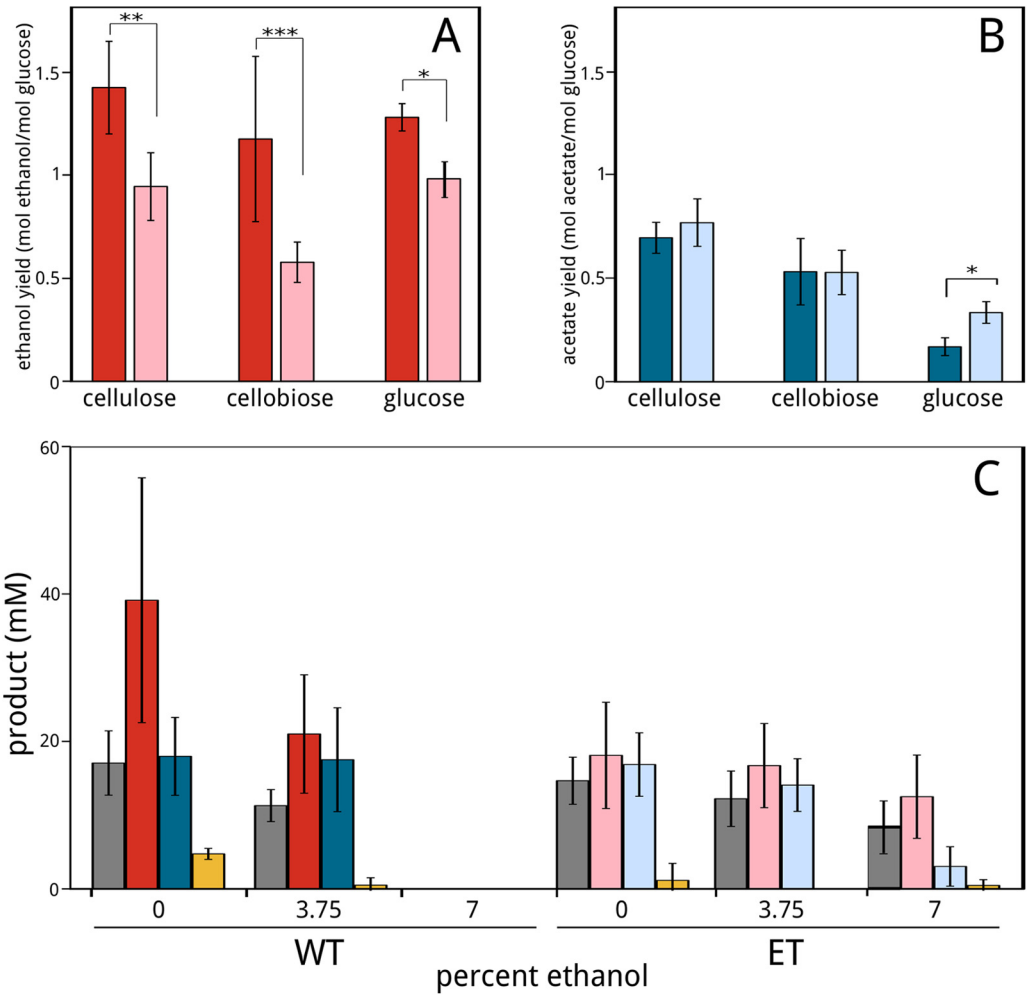

FIG 2 *C. phytofermentans* yields (mole of product per mole of glucose equivalent consumed) of ethanol from WT (dark red) and ET (light red) strains (A) and acetate from WT (dark blue) and ET (light blue) strains (B). Statistical differences between WT and ET yield averages are indicated at *P* values of <0.01 (*), <0.05 (**), and <0.1 (***) using Student's *t* test. (C) *C. phytofermentans* WT and ET cellobiose consumption and fermentation products in GS2 cellobiose medium with 0, 3.75, or 7% (vol/vol) added ethanol. Bars show concentrations (in millimoles) of cellobiose consumption (gray) and production of ethanol (red), acetate (blue), and formate (yellow) by WT (dark bars) and ET (light bars) strains. All cultures were grown in serum bottles containing GS2 medium at 30°C with 30 g liter^{-1} cellulose, 10 g liter^{-1} cellobiose, or 30 g liter^{-1} glucose and measured after 14 days. Bars represent averages of the results from 4 cultures; error bars represent 1 standard deviation.

plated, individual colonies were picked, and liquid cultures were reinoculated to ensure that selection was based on a specific strain with increased ethanol tolerance and not on a consortium of strains that collectively survived the increased ethanol concentration. The ET strain is thus a colony-purified isolate from a mother culture that grew in GS2 medium supplemented with 7% (vol/vol) ethanol. After colony purification, it was confirmed the ET strain has ethanol resistance similar to that of the mother culture.

Cellulose and fermentation analysis. The level of cellulose remaining in the culture was measured by taking a 1-ml sample from a 10-ml culture tube with a sterile syringe, placing it in a preweighed 1.7-ml microcentrifuge tube, and centrifuging at 13,000 × *g* for 10 min. The supernatant was removed, and the cellulose pellet was washed and centrifuged again at 13,000 × *g* for 10 min. The rinsed pellet was placed at 70°C to dry until a constant mass was reached. The contribution of cellular biomass to total cellulose weight was not accounted for and was assumed to be minimal due to low anaerobic biomass yields.

Fermentation product concentrations were measured in 0.22-μm-pore-size-filtered culture supernatant using an Agilent 1100 high-performance liquid chromatograph (HPLC) with a Jasco RI-1531 refractive index detector (RID) and an Aminex HPX-87H cation exchange column (Bio-Rad). The HPLC was run using a 0.01 M sulfuric acid mobile phase, 65°C column temperature, 30°C RID temperature, 25 μl sample volume, and 0.6 ml/min operating flow rate. Product formation is reported relative to the concentration in the medium at the point of inoculation. Gas phase measurements were made by removing 1 ml of headspace and injecting 100 μl into a gas chromatograph (Model 8610C multiple-gas analyzer; SRI Instruments). Argon was used as a carrier gas and was adjusted to 30 lb/in^2 gauge pressure. A stainless steel molecular sieve (13×) and silica gel-packed columns were used for sample separation, and the components were detected using a thermal conductivity detector (TCD). The column compartment temperature was held initially at 40°C for 3.5 min and then ramped to 160°C for 2 min and to 300°C for 10 min, after which the column was allowed to cool to 40°C for the remainder of the sample run.

Genome sequencing and variant analysis of the ET strain. A total of 12 μg of genomic DNA was extracted from a 4-ml ET strain culture using a Sigma GenElute bacterial genomic DNA kit (NA2110). DNA was sequenced on an Illumina MiSeq instrument with an insertion size of 795 bp

TABLE 2 Genomic DNA variants in the ET strain[a]

Protein	Length (amino acids)	Amino acid variant	Confidence (*Q*) value	Annotation
Energy and metabolism				
Cphy3925 (AdhE)	872	G609D	8,929	Fe-dependent bifunctional acetaldehyde-CoA/alcohol dehydrogenase
Cphy0215 (RnfA)	191	C26S	5,569	Ferredoxin:NAD^+ oxidoreductase (Fno); couples electron transfer from reduced ferredoxin to NAD^+ with cation transport out of the cell to create an electrochemical gradient
Cphy3255	259	20-bp insertion at Q91	15,938	FMN-dependent nitroreductase
Transport				
Cphy0543 (MgtA/MgtB)	920	K417N	8,661	P-type ATPase for Mg^{2+} uptake transporter or Ca^{2+}/Mg^{2+} antiporter
Cphy3778	258	1-bp deletion, Y239 frameshift	11,687	Na^+ efflux transporter, ABC-2 transmembrane permease (PFAM accession no. PF06182)
Membrane and cell wall				
Cphy0233 (PlsD)	237	D80N	9,246	Membrane synthesis: glycerol-3-phosphate acyltransferase; transfers a fatty acid to the 1 position of glycerol-3-phosphate
Cphy0107 (MurC)	469	G115D	8,841	Peptidoglycan synthesis: ATP-dependent ligation of L-alanine and UDP-*N*-acetylmuramic acid to form UDP-*N*-acetylmuramyl-L-alanine
Gene regulation				
Cphy3040	301	D182N	8,214	LysR transcriptional activator/repressor
Cphy3687 (PolB)	1,278	D189Y	8,824	β-Subunit of DNA-directed RNA polymerase
Noncoding changes				
Cphy0267	524	Q65Q (synonymous)	7,470	Modification methyltransferase
Intergenic G→T transversion		None	52	Transversion between 2 convergently transcribed genes, encoding Cphy1313 and Cphy1314
Cphy3036 (ApbE)	381	L244L (synonymous)	9,011	Thiamine biosynthesis: membrane-associated lipoprotein

[a] Data include the gene name, encoded protein length, amino acid (if coding) or DNA variant, confidence value (Q) of the variant call, and annotation of the mutated protein. The probability (*P*) that a variant exists in the genome is reported as a Phred-scaled probability, i.e., $Q = -10 \times \log_{10}(1 - P)$, meaning that a *Q* value of 100 indicates an error probability $(1 - P)$ of 10^{-10} (see GATK reports in the supplemental material for more information). FMN, flavin mononucleotide.

and 300-bp paired-end reads. A total of 5,077,282 reads passed quality filtering and mapping by Picard Tools (https://github.com/broadinstitute/picard), yielding approximately 250-fold genome coverage. Sequence variants (single nucleotide polymorphisms [SNPs] and indels) in the ET genome relative to the reference strain genome (NCBI accession no. NC_010001.1) were identified using the Genome Analysis Toolkit (GATK) (20) (see the supplemental material for detailed descriptions of the filtering and variant-calling methods).

ADH purification and activity measurements. The Cphy3925-encoding genes from the wild-type (WT) and ET strains were cloned by ligation-independent cloning (21) into pET-22B(+) as previously described (19). Genes were cloned with C-terminal His tags using primer pair cphy3925F/cphy3925R (Table 1) and confirmed by sequencing. Plasmids were transformed into *Escherichia coli* BL21(DE3) (Novagen 70235) and grown in 50 ml TB medium (12 g $liter^{-1}$ tryptone, 24 g $liter^{-1}$ yeast extract, 4 ml $liter^{-1}$ glycerol) to an OD_{600} of 1, and expression was induced by adding 500 μM IPTG (isopropyl-β-D-thiogalactopyranoside) and incubating overnight at 20°C. Cells were pelleted, resuspended in lysis buffer (50 mM phosphate buffer [pH 8], 0.5 M NaCl, 10 mM imidazole, 15% glycerol, 1 mM Pefabloc [Sigma 76307]), and lysed by sonication (Cole-Parmer Vibracell CV33) with lysozyme (Novagen 71230). His-tagged proteins were purified from 50 ml culture on nickel-nitrilotriacetic acid (Ni-NTA) spin columns (Qiagen 31014) and visualized on 12% SDS-PAGE gels (Novex 12% Bis-Tris gel NP0342BOX). Enzyme activities were measured as described in reference 22 in 100 μl of 100 mM Tris-HCl (pH 8) containing cofactor [0.2 mM NADH(P)H or 2 mM $NAD(P)^+$] and substrate (300 μM acetyl coenzyme A [acetyl-CoA] or CoA, 18 mM acetaldehyde, 2 M ethanol) as appropriate. NAD(P)H was measured as 340-nm absorbance (extinction coefficient, 6.22 $mM^{-1} \cdot cm^{-1}$) using a SAFAS UVmc2 spectrophotometer at room temperature.

Plasmid construction and transformations. pQexpE is derived from pQexp (23), a plasmid that replicates stably with erythromycin selection in *E. coli* (200 μg ml^{-1} erythromycin) and *C. phytofermentans* (40 μg ml^{-1}erythromycin on plates and 200 μg ml^{-1} erythromycin in liquid culture). To construct pQexpE, the *Z. mobilis pdc* and *adhB* genes were PCR amplified from pES120 (24) using primer pair pdcAdhB_F/dcAdhB_R (Table 1) and cloned into the unique BamHI and PvuI sites of pQexp. The insertion was confirmed by sequencing using primer pair pQexp_F/pQexp_R. pQexpE was conjugally transferred to *C. phytofermentans* using donor strain *E. coli* S17-1. Conjugation was performed as described in references 23 and 25, except polyethersulfone membranes were used to support 50-μl culture mixtures and only nalidixic acid without trimethoprim was used to select against *E. coli* following mating. Positive *C. phytofermentans* transconjugants containing pQexpE were confirmed by colony PCR using primer pair pQexp_F/pQexp_R to amplify the 2.9-kb *pdc-adhB* operon.

Nucleotide sequence accession number. The FASTQ-formatted DNA sequencing files for the ET genome were submitted to the European Nucleotide Archive under primary accession no. PRJEB7255.

RESULTS AND DISCUSSION

Isolation and physiology of an ethanol-tolerant *C. phytofermentans* strain. The growth of wild-type (WT) *C. phytofermen-*

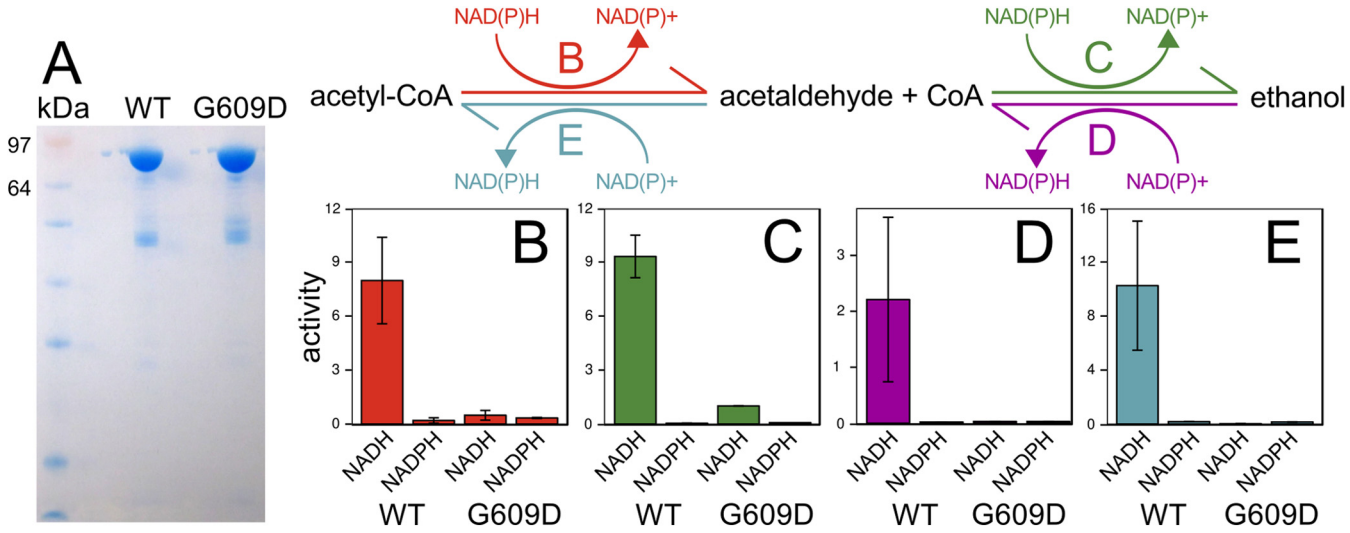

FIG 3 Comparison of activities of purified Cphy3925 AdhE from wild-type and ET strains. (A) SDS-PAGE gel of purified Cphy3925 from the wild-type (WT) and ET (G609D) strains showing single bands of the expected 95-kDa molecular mass. (B to E) Reactions for the two-step, bidirectional interconversion of acetyl-CoA acetaldehyde and ethanol: reduction of 300 μM acetyl-CoA to acetaldehyde (red) (B), reduction of 18 mM acetaldehyde to ethanol (green) (C), oxidation of 2 M ethanol to acetaldehyde (purple) (D), and oxidation of 18 mM acetaldehyde to acetyl-CoA (blue) (E). Enzyme activities are shown in millimoles of NAD(P)H per micromole of enzyme per second measured using NADH(P)H or $NAD(P)^+$ cofactors. Bar heights represent averages of duplicate activity measurements, and error bars represent 1 standard deviation.

tans was monitored in cultures supplemented with 0, 2, 4, 6, or 7% (vol/vol) ethanol (Fig. 1A), and generation times (hours) and maximum cell densities (OD_{600}) were calculated (see Table S1 in the supplemental material). The WT strain cultures grew similarly at 0% ethanol and 2% ethanol but growth was significantly inhibited at 4% ethanol, and very little growth was observed at 6 and 7% ethanol. In contrast, the ET strain grew at 4, 6, and 7% ethanol, with a maximum cell density (OD_{600}) at 7% ethanol similar to that of the WT strain at 4% ethanol (Fig. 1B; see also Table S1).

While the ET strain was more ethanol resistant, it grew more slowly than the WT strain and grew at reduced cell yields in cultures without added ethanol (Fig. 1). Glucose consumption by the ET strain was also slower than glucose consumption by the WT strain, reflecting the reduced growth rate. When grown on 30 g $liter^{-1}$ glucose, the ET strain consumed only 20 g $liter^{-1}$ substrate in 200 h, while the WT strain completely exhausted the glucose in less than 150 h (see Fig. S1 in the supplemental material). Thus, the enhanced growth of the ET strain at a high ethanol concentration is accompanied by reduced growth under standard conditions, supporting the idea that the ET strain has altered physiology, with results similar to those seen with *C. thermocellum* adapted to 5% ethanol (26, 27). However, we observed no morphological differences between ET and WT colonies grown on GS2 agar plates or cells grown in liquid GS2 medium (see Fig. S2).

The ET strain also produces less ethanol per unit of sugar consumption than the WT strain. For example, ethanol yield (moles of product per mole of glucose equivalent consumed) by the ET strain decreased 25 to 50% relative to that seen with the WT strain when growing on cellulose, cellobiose, and glucose (Fig. 2A), whereas the acetate yields of the strains were similar (Fig. 2B). We also measured cellobiose consumption and formation of the primary fermentation products ethanol, acetate, and formate in WT and ET cultures grown in cellobiose medium supplemented with 0, 3.75, or 7% ethanol (Fig. 2C). Ethanol production by the WT strain was significantly lower at 3.75% ethanol, and cellobiose consumption and fermentation ceased at 7% ethanol. In contrast, cellobiose consumption and ethanol production by the ET strain decreased only slightly with increased ethanol supplementation, demonstrating the robustness of the ethanol resistance phenotype. In all, these results highlight the phenotypic advantages of the ET strain with respect to metabolizing and producing ethanol at elevated ethanol concentrations but also indicate that those advantages occur at the expense of lower ethanol yields and slower growth at low ambient concentrations of ethanol.

Genome sequence of the ET strain. The *C. phytofermentans* ET strain genome contains 12 variants relative to the WT strain (Table 2), many fewer than a *C. thermocellum* ET genome with similar ethanol resistance that had 200 to 500 changes (9). While ethanol resistance is likely a complex, multigenic trait, the small number of changes in the ET strain genome shed light on DNA variants that could have functional roles in ethanol tolerance. Two of the 12 mutations were in genes encoding transcriptional regulators that could effectuate broad gene expression changes: PolB, the β-subunit of DNA-directed RNA polymerase, and a LysR regulator, Cphy3040. LysR-type regulators often colocalize in the genome with their targets (28) and the gene encoding Cphy3040 is adjacent to a gene encoding a NAD-dependent aldehyde dehydrogenase, suggesting that this regulator is related to alcohol formation.

Ethanol increases the permeability of the cell membrane, resulting in toxic leakage of metabolites out of the cell (29). Ethanol resistance thus often involves membrane modifications such as altered protein content (8) or longer chain fatty acids and more plasmalogen lipids (30) that increase membrane rigidity to mitigate the fluidizing effect of ethanol. The ET strain has a D80N change in the putative acyl-acceptor binding pocket (NCBI accession no. cd07989) of Cphy0233, a homolog of *C. butyricum* PlsD that transfers a fatty acyl group to the *sn*-1 position of glycerol-3-phosphate in phospholipid biosynthesis (31). The D80N

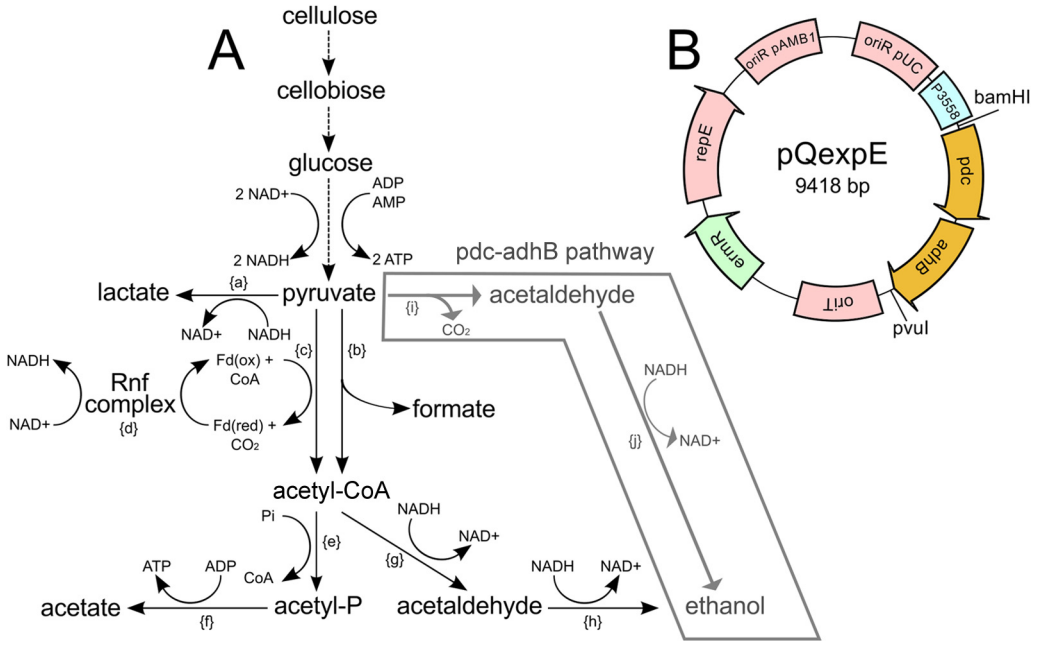

FIG 4 (A) Diagram of *C. phytofermentans* carbon metabolism showing insertion of the *Z. mobilis pdc-adhB* alternative ethanol formation pathway. Enzymatic steps: {a}, lactate dehydrogenase (Cphy1117); {b}, pyruvate formate lyase (Cphy1174); {c}, pyruvate ferredoxin oxidoreductase (Cphy3558); {d}, Rnf ferredoxin: NAD^+ oxidoreductase complex (Cphy0211 to Cphy0216); {e}, phosphate acetyltransferase (Cphy1326); {f}, acetate kinase (Cphy1327); {g}, acetaldehyde dehydrogenase (Cphy1428 or Cphy3925); {h}, alcohol dehydrogenase (Cphy3925 or Cphy1029); {i}, *Zymomonas mobilis* pyruvate decarboxylase (*pdc*); {j}, alcohol dehydrogenase (*adhB*). Dashed lines represent multienzyme reactions where all enzymes are not listed. (B) Plasmid map of pQexpE for *pdc-adhB* expression in *C. phytofermentans*. Plasmid features: Gram-negative pUC origin of replication (*oriR* pUC), *C. phytofermentans* pyruvate ferredoxin oxidoreductase promoter (P3558) to express the *Z. mobilis pdc-adhB* genes, the RP4 conjugal origin of transfer (*oriT*), the Gram-negative/Gram-positive erythromycin resistance gene from TN1545 (*ermR*), the Gram-positive pAMB1 origin (*oriR* pAMB1), and *repE*-encoded protein.

Cphy0233 mutation may thus enable synthesis of a more rigid, ethanol-resistant cell membrane by altering which fatty acids are incorporated into phospholipids.

The ET strain has a C26S mutation in the RnfA subunit of the membrane-bound Rnf complex that couples efflux of H^+ (32) or Na^+ (33) with electron transfer from reduced ferredoxin to NAD^+ (34). The resulting electrochemical gradient is harnessed by an F_0F_1 ATPase for ATP synthesis. The *C. phytofermentans* Rnf complex (Cphy0211 to Cphy0216) and F_oF_1 ATPase (Cphy3735 to Cphy3742) are highly expressed on all tested carbon sources (19) and may be important for energy conservation, similarly to *C. ljungdahlii* (32). However, Rnf generates NADH, which may not be tolerated by the ET strain that cannot reoxidize NADH by AdhE-mediated ethanol formation (see below). The C26S RnfA variant may thus cripple the Rnf complex, which sacrifices ATP production, but may benefit the ET strain by balancing cellular NADH/NAD^+ ratios.

The ET strain also has mutations in two transporters putatively involved in cation homeostasis. Cphy0543 is homologous to MgtA, a P-type ATPase upregulated at low ambient Mg^{2+} concentrations (35) to mediate Mg^{2+} uptake (36) or Ca^{2+}/Mg^{2+} antiport (37). Cphy3778, the membrane component of an ABC transporter (PFAM accession no. PF06182), appears to be cotranscribed with Cphy3780, an ABC-type Na^+ efflux protein (NCBI accession no. cd03267). In *Bacillus subtilis*, this Na^+ efflux system is induced by ethanol and is proposed to compensate for an influx of extracellular Na^+ resulting from a weakened membrane barrier (38). The variants in these cation transporters may increase their activities to alleviate cation leakage due to ethanol stress.

AdhE activities. The ET strain has a G609D variant in Cphy3925 AdhE, a putative acetaldehyde-CoA dehydrogenase and alcohol dehydrogenase (ADH). The G609D mutation is in a conserved position in the active site of the C-terminal ADH domain (NCBI accession no. cd08178). A previous study reported an ethanol-tolerant *C. thermocellum* strain with AdhE mutations (P704L and H735R) that shifted the cofactor specificity from

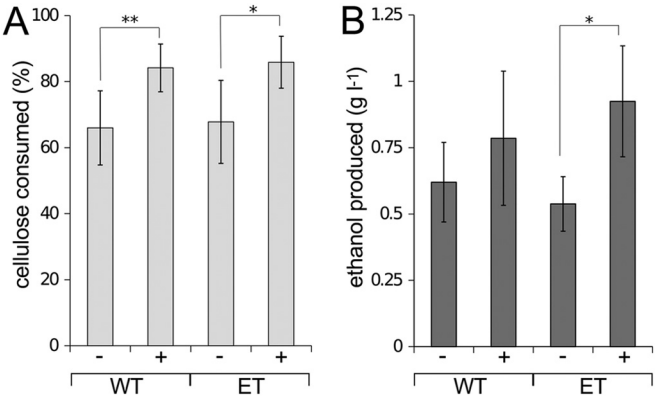

FIG 5 Cellulose consumption (A) and ethanol production (B) by non-plasmid-bearing (−) and *pdc-adhB*-containing (+) WT and ET cultures at 30°C in GS2 medium with 3 g $liter^{-1}$ α-cellulose after 30 days. Bars represent averages of the results from duplicate cultures in serum bottles from three independent experiments, and error bars represent 1 standard deviation. Statistical differences between treatments are indicated at *P* values of <0.01 (*) and <0.05 (**) using Student's *t* test.

TABLE 3 Carbon dioxide and hydrogen gas production by *C. phytofermentans* WT and ET cultures with or without the *Z. mobilis pdc-adhB* genes after 2 weeks of growth at 30°C in GS2 medium containing 100 g liter^{-1} α-cellulose[a]

Strain	Gas produced (ml)	H_2 (mM)	CO_2 (mM)	H_2/CO_2 (mol/mol)
WT	26.3 ± 0.4	4.21 ± 0.11	6.26 ± 0.08	0.67
WT (*pdc-adhB*)	40.0 ± 5.4	4.62 ± 0.13	10.02 ± 0.07	0.46
ET	25.2 ± 4.7	3.47 ± 0.12	4.46 ± 0.19	0.78
ET (*pdc-adhB*)	31.0 ± 2.3	4.41 ± 0.15	7.20 ± 0.13	0.61

[a] Data represent averages ± standard deviations of the results from four replicates.

NADH to NADPH, which was proposed to confer ethanol resistance by altering the internal redox balance (9). To determine the effect of the G609D mutation on Cphy3925 enzyme activity, we purified WT and ET versions of the enzyme (Fig. 3A) and tested their *in vitro* catalysis of the two-step, bidirectional reactions converting acetyl-CoA to ethanol using either NADH or NADPH cofactors.

The mutated Cphy3925 lost NAD(H)-dependent activities, but, unlike the mutated AdhE in *C. thermocellum*, the G609D mutation did not result in NADPH-dependent ADH activity (Fig. 3B to E). Instead, our results support the notion that the ET strain arrested AdhE-mediated interconversion of acetyl-CoA, acetaldehyde, and ethanol, which helps explain why the *C. phytofermentans* ET strain had lower ethanol yield. AdhE loss of function could mitigate ethanol stress by reducing intracellular levels of ethanol and its highly toxic precursor, acetaldehyde. *C. phytofermentans* encodes four Fe-dependent ADHs in addition to Cphy3925 as well as a Zn-dependent ADH. All 6 ADHs are expressed, and Cphy3925 and Cphy1029 are among the most highly expressed proteins on all tested carbon sources (19, 39). *C. phytofermentans* thus likely produces ethanol by the concerted action of multiple ADHs, and these other ADHs, especially Cphy1029, are responsible for ethanol produced by the ET strain.

Ethanol pathway engineering. To augment ethanol production by the ET strain, an alternative ethanol production pathway comprised of pyruvate decarboxylase (Pdc) and alcohol dehydrogenase (AdhB) from *Zymomonas mobilis* (Fig. 4A) was transferred into *C. phytofermentans* on the replicating pQexpE plasmid (Fig. 4B). Together, these enzymes couple decarboxylation of pyruvate to ethanol with the oxidation of NADH and thus represent an alternative to the AdhE ethanol formation pathway. We chose to express foreign enzymes rather than a WT copy of Cphy3925 because AdhE multimerizes (40) such that the mutant AdhE could have a dominant-negative effect in a merodiploid.

Expression of pQexpE increased cellulolysis by ~30% in both the WT and ET strains (Fig. 5A) and boosted ethanol production by 70% relative to the ET strain ($P < 0.01$), thereby restoring ethanol yields to WT levels (Fig. 5B). CO_2 production increased disproportionately relative to H_2 production in WT and ET strains expressing pQexpE (Table 3). Elevated CO_2 synthesis is likely due to increased pyruvate decarboxylation by the Pdc enzyme. Previous results showed that Pdc/AdhB expression enhanced cellulolysis and ethanol production in WT *C. cellulolyticum*, which was proposed to result from consumption of excess pyruvate that otherwise leads to metabolic arrest (14). Increased metabolism (cellulolysis and production of CO_2 and ethanol) by *C. phytofermentans* expressing Pdc/AdhB might be due to alleviation of inhibition by excess pyruvate. Alternatively, expression or activity of glycolytic enzymes might be regulated by NADH levels such that NADH reoxidation by Pdc/AdhB stimulates glycolysis, which results in increased substrate utilization.

Conclusions. In this study, we investigated the genetic basis and phenotypic consequences of microbial ethanol tolerance by isolating, characterizing, and engineering an ethanol-resistant (ET) strain of *Clostridium phytofermentans*. The ET strain grows at higher ethanol concentrations than the wild-type strain (Fig. 1) and continues to produce ethanol at a 7% ambient ethanol concentration (Fig. 2C) but has impaired growth (Fig. 1) and ethanol yield (Fig. 2A) relative to the wild type. The genome sequence of the ET strain revealed 12 mutations in genes involved in diverse aspects of metabolism (Table 2), including a G609D variant in the bifunctional acetaldehyde CoA/alcohol dehydrogenase AdhE that abolishes its activity (Fig. 3). We complemented the AdhE mutation in the ET strain by expressing pyruvate decarboxylase (Pdc) and alcohol dehydrogenase B (AdhB) from *Zymomonas mobilis* on the pQexpE plasmid (Fig. 4), which boosted substrate conversion (Fig. 5A) and restored ethanol production (Fig. 5B).

Additional work is needed to enhance *C. phytofermentans*' plant biomass degradation and ethanol formation rates and product titers. Recently, improvement of *C. phytofermentans*' growth on cellobiose, cellulose, and xylan by experimental evolution yielded strains that also produced ethanol more quickly (41). The genome sequence of the ET strain presented here suggests other novel approaches to potentially improve ethanol resistance and production. For example, our results suggest that further studies on ethanol resistance should focus on PlsD-mediated fatty acid incorporation into phospholipids, LysR-regulated gene expression patterns, overexpression of the Rnf complex to stimulate AdhE-mediated ethanol production, and prevention of cation leakage.

ACKNOWLEDGMENTS

C. phytofermentans ISDg was kindly provided by Susan Leschine (University of Massachusetts, Amherst).

A.C.T. was supported by a CNRS Chaire d'Excellence and the Genoscope-CEA. T.R.Z. was supported by a GRFP Fellowship from the National Science foundation (grant no. DGE1255832) and a John and Jeannette McWhirter Fellowship from The Pennsylvania State University. M.R. was supported by the Faculty Abroad Program, SRM University. Anaerobic growth capabilities for this work were financed by the Genoscope-CEA and a grant from the Department of Energy Advanced Research Project Agency—Energy (ARPA-e; no. DE-AR0000092) to W.R.C.

We thank Alex Rajangam for facilitating M.R.'s international exchange, Sergio Florez for useful feedback, Patrick Hillery and Sam Bjork for experimental assistance, and Nymul Khan for supporting GC analysis. We acknowledge the Huck Institute of the Life Sciences Shared Fermentation Facility for providing HPLC analytical capabilities and the Genoscope-CEA Sequencing Platform for sequencing the ET strain. Finally, we thank George Church and Noel Goddard for facilitating a sabbatical leave for W.R.C. at Harvard, where these studies were initiated.

REFERENCES

1. **Lynd LR, van Zyl WH, McBride JE, Laser M.** 2005. Consolidated bioprocessing of cellulosic biomass: an update. Curr Opin Biotechnol **16:** 577–583. http://dx.doi.org/10.1016/j.copbio.2005.08.009.
2. **Lu Y, Zhang Y-HP, Lynd LR.** 2006. Enzyme-microbe synergy during

cellulose hydrolysis by *Clostridium thermocellum*. Proc Natl Acad Sci U S A **103:**16165–16169. http://dx.doi.org/10.1073/pnas.0605381103.

3. **Warnick TA, Methé BA, Leschine SB.** 2002. *Clostridium phytofermentans* sp. nov., a cellulolytic mesophile from forest soil. Int J Syst Evol Microbiol **52:**1155–1160. http://dx.doi.org/10.1099/ijs.0.02125-0.
4. **Tolonen AC, Petit E, Blanchard JL, Warnick T, Leschine SB.** 2013. Technologies to study plant biomass fermentation using the model bacterium *Clostridium phytofermentans*, p 114–139. *In* Sun J, Ding SY, Peterson JD (ed), Biological conversion of biomass for fuels and chemicals: explorations from natural utilization systems. Royal Society of Chemistry, Cambridge, United Kingdom. http://dx.doi.org/10.1039/9781849734738-00114.
5. **Jin M, Balan V, Gunawan C, Dale BE.** 2011. Consolidated bioprocessing (CBP) performance of *Clostridium phytofermentans* on AFEX-treated corn stover for ethanol production. Biotechnol Bioeng **108:**1290–1297. http://dx.doi.org/10.1002/bit.23059.
6. **Jin M, Gunawan C, Balan V, Dale BE.** 2012. Consolidated bioprocessing (CBP) of AFEX™-pretreated corn stover for ethanol production using *Clostridium phytofermentans* at a high solids loading. Biotechnol Bioeng **109:**1929–1936. http://dx.doi.org/10.1002/bit.24458.
7. **Zuroff TR, Xiques SB, Curtis WR.** 2013. Consortia-mediated bioprocessing of cellulose to ethanol with a symbiotic *Clostridium phytofermentans*/yeast co-culture. Biotechnol Biofuels **6:**59. http://dx.doi.org/10.1186/1754-6834-6-59.
8. **Shao X, Raman B, Zhu M, Mielenz JR, Brown SD, Guss AM, Lynd LR.** 2011. Mutant selection and phenotypic and genetic characterization of ethanol-tolerant strains of *Clostridium thermocellum*. Appl Microbiol Biotechnol **92:**641–652. http://dx.doi.org/10.1007/s00253-011-3492-z.
9. **Brown SD, Guss AM, Karpinets TV, Parks JM, Smolin N, Yang S, Land ML, Klingeman DM, Bhandiwad A, Rodriguez M, Raman B, Shao X, Mielenz JR, Smith JC, Keller M, Lynd LR.** 2011. Mutant alcohol dehydrogenase leads to improved ethanol tolerance in *Clostridium thermocellum*. Proc Natl Acad Sci U S A **108:**13752–13757. http://dx.doi.org/10.1073/pnas.1102444108.
10. **Zhu X, Cui J, Feng Y, Fa Y, Zhang J, Cui Q.** 2013. Metabolic adaption of ethanol-tolerant *Clostridium thermocellum*. PLoS One **8:**e70631. http://dx.doi.org/10.1371/journal.pone.0070631.
11. **Yang S, Giannone RJ, Dice L, Yang ZK, Engle NL, Tschaplinski TJ, Hettich RL, Brown SD.** 2012. *Clostridium thermocellum* ATCC27405 transcriptomic, metabolomic and proteomic profiles after ethanol stress. BMC Genomics **13:**336. http://dx.doi.org/10.1186/1471-2164-13-336.
12. **Ezeji T, Milne C, Price ND, Blaschek HP.** 2010. Achievements and perspectives to overcome the poor solvent resistance in acetone and butanol-producing microorganisms. Appl Microbiol Biotechnol **85:**1697–1712. http://dx.doi.org/10.1007/s00253-009-2390-0.
13. **Tomas CA, Beamish J, Papoutsakis ET.** 2004. Transcriptional analysis of butanol stress and tolerance in *Clostridium acetobutylicum*. J Bacteriol **186:**2006–2018. http://dx.doi.org/10.1128/JB.186.7.2006-2018.2004.
14. **Guedon E, Desvaux M, Petitdemange H.** 2002. Improvement of cellulolytic properties of *Clostridium cellulolyticum* by metabolic engineering. Appl Environ Microbiol **68:**53–58. http://dx.doi.org/10.1128/AEM.68.1.53-58.2002.
15. **Deng Y, Olson DG, Zhou J, Herring CD, Joe Shaw A, Lynd LR.** 2013. Redirecting carbon flux through exogenous pyruvate kinase to achieve high ethanol yields in *Clostridium thermocellum*. Metab Eng **15:**151–158. http://dx.doi.org/10.1016/j.ymben.2012.11.006.
16. **Argyros DA, Tripathi SA, Barrett TF, Rogers SR, Feinberg LF, Olson DG, Foden JM, Miller BB, Lynd LR, Hogsett DA, Caiazza NC.** 2011. High ethanol titers from cellulose by using metabolically engineered thermophilic, anaerobic microbes. Appl Environ Microbiol **77:**8288–8294. http://dx.doi.org/10.1128/AEM.00646-11.
17. **Biswas R, Zheng T, Olson DG, Lynd LR, Guss AM.** 2015. Elimination of hydrogenase active site assembly blocks H_2 production and increases ethanol yield in *Clostridium thermocellum*. Biotechnol Biofuels **8:**20. http://dx.doi.org/10.1186/s13068-015-0204-4.
18. **Cavedon K, Leschine SB, Canale-Parola E.** 1990. Cellulase system of a free-living, mesophilic clostridium (strain C7). J Bacteriol **172:**4222–4230.
19. **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. http://dx.doi.org/10.1371/journal.pgen.1004773.
20. **McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA.** 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res **20:**1297–1303. http://dx.doi.org/10.1101/gr.107524.110.
21. **Aslanidis C, de Jong PJ.** 1990. Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res **18:**6069–6074. http://dx.doi.org/10.1093/nar/18.20.6069.
22. **Yan RT, Chen JS.** 1990. Coenzyme A-acylating aldehyde dehydrogenase from *Clostridium beijerinckii* NRRL B592. Appl Environ Microbiol **56:**2591–2599.
23. **Tolonen AC, Chilaka AC, Church GM.** 2009. Targeted gene inactivation in *Clostridium phytofermentans* shows that cellulose degradation requires the family 9 hydrolase Cphy3367. Mol Microbiol **74:**1300–1313. http://dx.doi.org/10.1111/j.1365-2958.2009.06890.x.
24. **Bokinsky G, Peralta-Yahya PP, George A, Holmes BM, Steen EJ, Dietrich J, Lee TS, Tullman-Ercek D, Voigt CA, Simmons BA, Keasling JD.** 2011. Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered *Escherichia coli*. Proc Natl Acad Sci U S A **108:**19949–19954. http://dx.doi.org/10.1073/pnas.1106958108.
25. **Tolonen AC, Cerisy T, El-Sayyed H, Boutard M, Salanoubat M, Church GM.** 25 May 2014, posting date. Fungal lysis by a soil bacterium fermenting cellulose. Environ Microbiol http://dx.doi.org/10.1111/1462-2920.12495.
26. **Williams TI, Combs JC, Lynn BC, Strobel HJ.** 2007. Proteomic profile changes in membranes of ethanol-tolerant *Clostridium thermocellum*. Appl Microbiol Biotechnol **74:**422–432. http://dx.doi.org/10.1007/s00253-006-0689-7.
27. **Biswas R, Prabhu S, Lynd LR, Guss AM.** 2014. Increase in ethanol yield via elimination of lactate production in an ethanol-tolerant mutant of *Clostridium thermocellum*. PLoS One **9:**e86389. http://dx.doi.org/10.1371/journal.pone.0086389.
28. **Maddocks SE, Oyston PCF.** 2008. Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. Microbiology **154**(Pt 12)**:**3609–3623. http://dx.doi.org/10.1099/mic.0.2008/022772-0.
29. **Ingram LO.** 1990. Ethanol tolerance in bacteria. Crit Rev Biotechnol **9:**305–319.
30. **Timmons MD, Knutson BL, Nokes SE, Strobel HJ, Lynn BC.** 2009. Analysis of composition and structure of *Clostridium thermocellum* membranes from wild-type and ethanol-adapted strains. Appl Microbiol Biotechnol **82:**929–939. http://dx.doi.org/10.1007/s00253-009-1891-1.
31. **Heath RJ, Goldfine H, Rock CO.** 1997. A gene (*plsD*) from *Clostridium butyricum* that functionally substitutes for the *sn*-glycerol-3-phosphate acyltransferase gene (*plsB*) of *Escherichia coli*. J Bacteriol **179:**7257–7263.
32. **Tremblay P-L, Zhang T, Dar SA, Leang C, Lovley DR.** 2012. The Rnf complex of *Clostridium ljungdahlii* is a proton-translocating ferredoxin: NAD^+ oxidoreductase essential for autotrophic growth. mBio **4:**e00406-412. http://dx.doi.org/10.1128/mBio.00406-12.
33. **Biegel E, Müller V.** 2010. Bacterial Na^+-translocating ferredoxin:NAD^+ oxidoreductase. Proc Natl Acad Sci U S A **107:**18138–18142. http://dx.doi.org/10.1073/pnas.1010318107.
34. **Biegel E, Schmidt S, Müller V.** 2009. Genetic, immunological and biochemical evidence for a Rnf complex in the acetogen *Acetobacterium woodii*. Environ Microbiol **11:**1438–1443. http://dx.doi.org/10.1111/j.1462-2920.2009.01871.x.
35. **Snavely MD, Gravina SA, Cheung TT, Miller CG, Maguire ME.** 1991. Magnesium transport in *Salmonella typhimurium*. Regulation of *mgtA* and *mgtB* expression. J Biol Chem **266:**824–829.
36. **Hmiel SP, Snavely MD, Florer JB, Maguire ME, Miller CG.** 1989. Magnesium transport in *Salmonella typhimurium*: genetic characterization and cloning of three magnesium transport loci. J Bacteriol **171:**4742–4751.
37. **Neef J, Andisi VF, Kim KS, Kuipers OP, Bijlsma JJE.** 2011. Deletion of a cation transporter promotes lysis in *Streptococcus pneumoniae*. Infect Immun **79:**2314–2323. http://dx.doi.org/10.1128/IAI.00677-10.
38. **Cheng J, Guffanti AA, Krulwich TA.** 1997. A two-gene ABC-type transport system that extrudes Na^+ in *Bacillus subtilis* is induced by ethanol or

protonophore. Mol Microbiol **23:**1107–1120. http://dx.doi.org/10.1046/j.1365-2958.1997.2951656.x.

39. **Tolonen AC, Haas W, Chilaka AC, Aach J, Gygi SP, Church GM.** 2011. Proteome-wide systems analysis of a cellulosic biofuel-producing microbe. Mol Syst Biol **7:**461. http://dx.doi.org/10.1038/msb.2010.116.
40. **Kessler D, Leibrecht I, Knappe J.** 1991. Pyruvate-formate-lyase-deactivase and acetyl-CoA reductase activities of *Escherichia coli* reside on a polymeric protein particle encoded by *adhE*. FEBS Lett **281:**59–63. http://dx.doi.org/10.1016/0014-5793(91)80358-A.
41. **Mukherjee S, Thompson LK, Godin S, Schackwitz W, Lipzen A, Martin J, Blanchard JL.** 2014. Population level analysis of evolved mutations underlying improvements in plant hemicellulose and cellulose fermentation by *Clostridium phytofermentans*. PLoS One **9:**e86731. http://dx.doi.org/10.1371/journal.pone.0086731.

THESIS/THÈSE

English

Title: Rational and evolution-based engineering of Clostridium phytofermentans

Abstract:

The research aim of this thesis project was to develop and apply new tools for targeted chromosomal changes and *in vivo* directed evolution of *Clostridium phytofermentans,* a mesophilic anaerobe that ferments lignocellulosic biomass. The introduction presents an overview of previous basic and applied research in Clostridia biology and genetics, including both pathogenic and environmental strains. A focus on the biofuel industry is reported to describe applications of *C. phytofermentans* in this competitive industry. Chapter one presents a study using functional genomics and targeted gene inactivation to identify hexose sugar transporters in *C. phytofermentans*. Chapter two describes the application of Genome Editing *via* Targetron and Recombinases (GETR) to make large-scale chromosomal deletions in this bacterium. Chapter 3 presents the *in vivo* directed evolution of *C. phytofermentans* to enhance its resistance to lignocellulosic inhibitors including analyses of genome-wide transcription patterns and genomic variants that arose in the resistant strains. Together, this thesis work highlights the advantages and limitations of both targeted and evolutional approaches to study and engineer Clostridia.

Keywords: Clostridia, *Clostridium phytofermentans*, genome editing, synthetic biology, microbiology, ABC transporter, directed-evolution.

Français

Titre : Evolution dirigée et ingénierie rationnelle chez *Clostridium phytofermentans*

Résumé :

Le but de la recherche menée dans le cadre de cette thèse est de développer et d'appliquer de nouveaux outils de modifications chromosomiques ciblées et d'évolution dirigée *in vivo* chez *Clostridium phytofermentans*, une bactérie mésophile et anaérobie qui fermente la biomasse lignocellulosique. L'introduction présente une vue d'ensemble des recherches fondamentales et appliquées dans la biologie et la génétique des Clostridia, incluant les souches pathogènes et environnementales. Une description de l'industrie des biocarburants est détaillée pour montrer les applications possibles de *C. phytofermentans* dans ce domaine très compétitif. Le premier chapitre présente l'utilisation de la génomique fonctionnelle et de l'inactivation de gènes cibles pour identifier des transporteurs d'hexoses chez *C. phytofermentans*. Le second chapitre décrit l'application de la technique *Genome Editing via Targetron and Recombinases (GETR)* pour effectuer de larges délétions chromosomiques chez cette bactérie. Le troisième chapitre présente l'évolution dirigée *in vivo* chez *C. phytofermentans* pour améliorer sa résistance aux inhibiteurs issus de la biomasse lignocellulosique, incluant l'analyse des variations transcriptomiques et génomiques des souches résistantes. Dans son ensemble, ce travail de thèse souligne les avantages et limites de l'approche ciblée ou de l'approche par évolution *in vivo*, pour étudier et modifier les Clostridia.

Mots Clés : Clostridia, *Clostridium phytofermentans*, édition de génomes, biologie de synthèse, microbiologie, Transporteur ABC, évolution dirigée.