Engineering of Lachnospiraceae fermentation for health and the bioeconomy

Andrew C. Tolonen

université PARIS-SACLAY

L’ingénierie de la fermentation chez les *Lachnospiraceae* pour la santé et la bioéconomie

Habilitation à diriger des recherches de l'Université Paris-Saclay

Unité de recherche : Université Paris-Saclay, Univ Évry, CNRS, CEA, Génomique métabolique, 91057, Évry, France

présentée et soutenue à Évry le 23 janvier 2024, par

Andrew C TOLONEN

Composition du jury

Bernard HENRISSAT Professor, Technical University of Denmark, Directeur de Recherche Emérite, CNRS	Rapporteur
Henri-Pierre FIEROBE Directeur de Recherche CNRS, Aix Marseille Univ	Rapporteur
Julien PICOT Directeur de Recherche, Genopole Evry	Rapporteur
Laurence ETTWILLER Research Director, New England Biolabs	Examinatrice
Andrea RAU Directrice de Recherche, INRAE Jouy en Josas	Examinatrice

université PARIS-SACLAY

Titre : L'ingénierie de la fermentation chez les *Lachnospiraceae* pour la santé et la bioéconomie

Mots clés : fermentation, lignocellulose, Clostridia, *Lachnospiraceae*

Résumé : Bien que la fermentation joue un rôle sociétal important depuis des millénaires, ses applications grandissent avec l'essor de la biotechnologie. En particulier, les innovations permettant l'ingénierie métabolique de nouveaux micro-organismes contribueront à réaliser le potentiel de la fermentation dans la bioéconomie. Mes recherches portent sur l'optimisation de la fermentation des *Lachnospiraceae*, une famille de bactéries anaérobies de la classe des Clostridia. Certaines espèces de *Lachnospiraceae* métabolisent des matières premières abondantes et peu coûteuses telles que la lignocellulose et le dioxyde de carbone en produits chimiques à valeur ajoutée. D'autres font partie des espèces dominantes du côlon humain et du rumen animal, où elles favorisent la santé intestinale et la fonction immunitaire. En utilisant principalement *Clostridium phytofermentans* comme modèle, j'applique le criblage à haut débit, l'édition du génome, et la bioinformatique pour créer des souches aux phénotypes améliorés et faire progresser notre compréhension de la biologie des *Lachnospiraceae*. À terme, mon objectif est d'exploiter ces bactéries anaérobies comme biocatalyseurs industriels et thérapeutiques.

Title : Engineering *Lachnospiraceae* fermentation for health and the bioeconomy

Keywords : fermentation, lignocellulose, Clostridia, *Lachnospiraceae*

Abstract : While fermentation has played an important societal role for millenia, its applications are expanding with the rise of biotechnology. In particular, innovations for metabolic engineering of novel microorganisms will help realize the potential of fermentation in the bioeconomy. My research focuses on optimizing fermentation by *Lachnospiraceae,* an understudied family of anaerobic bacteria in the class Clostridia. Some species of *Lachnospiraceae* metabolize abundant, low-cost feedstocks such as lignocellulose and carbon dioxide into value-added chemicals. Others are among the dominant species of the human colon and animal rumen, where they promote healthy gut and immune function. Primarily using *Clostridium phytofermentans* as a model, I am applying high-throughput screening, genome editing, and bioinformatics to create strains with improved phenotypes and to advance our understanding of the biology of *Lachnospiraceae*. Ultimately, my goal is to harness these anaerobes as industrial biocatalysts and therapeutics.

Table of Contents

COMMISSION DE LA RECHERCHE EN FORMATION RESTREINTE
DU 02 NOVEMBRE 2020

DEMANDE D'AUTORISATION A L'INSCRIPTION AU DIPLOME D'HABILITATION A DIRIGER DES RECHERCHES : COLLEGES A - B

La Commission de la Recherche se prononce à l'appui des rapports établis par les membres de la CR, des rapporteurs extérieurs et après examen des dossiers des candidats.

Nom – Prénom	Diplôme (année)	Laboratoire / Dr(ce) Laboratoire	Dr(ce) Travaux	Section CNU	Rapporteur Ext.	Rapporteur CR	Avis Fav / Déf
Andrew C. TOLONEN	Doctorat en 2005 Spécialité : Biologie moléculaire	Génomique Métabolique Dir : Patrick WINCKER	-	Sciences : Groupe 10 Section 64 Biochimie et Biologie Moléculaire	Jean-Luc PERNODET [Institut de Biologie Intégrative de la Cellule]	Florence GONNET [LAMBE – Evry-Université Paris-Saclay]	Fav

La recevabilité de la demande d'inscription au diplôme d'HDR est examinée par le Président de l'UEVE qui statue sur proposition de la Commission de la Recherche siégeant en formation restreinte sur personnalités habilitées à diriger des recherches et après avis du Directeur de travaux si le candidat en a un. L'autorisation d'inscription est ensuite notifiée au candidat.

Dans le cas d'un avis favorable, l'autorisation d'inscription au diplôme d'HDR est délivrée pour une durée de trois années, à partir de l'avis de la Commission de la Recherche, en date du 16 mars 2015.

Dans l'hypothèse d'un refus d'autorisation de l'inscription à l'HDR, un avis argumenté devra être établi par la Vice-présidente de la Commission de la Recherche et communiqué au candidat.

1ère Vice-présidente
de la Commission de la Recherche

Florence GONNET

COMMISSION DE LA RECHERCHE EN FORMATION RESTREINTE
LUNDI 2 NOVEMBRE 2020

Rapport concernant une demande d'autorisation au diplôme d'HDR

Laboratoire : Génomique Métabolique

Candidat à l'HDR :

- Nom : TOLONEN
- Prénom : Andrew
- Diplôme [année] : Doctorat en Biologie Moléculaire obtenu en Août 2005

Avis argumenté :

1. *Activité scientifique* :
   Andrew Tolonen a soutenu sa thèse en août 2005 au MIT à Cambridge en biologie moléculaire. Après deux post-doctorats (1 an à Paris et 4 ans à Boston), il est MCF depuis 2011 à l'UEVE, et chef de groupe au Genoscope. Il est en disponibilité depuis 2017, car en parallèle de sa recherche au CEA, il a été directeur associé du groupe « microbiomes et immunity au MIT et Harvard, et est directeur de l'innovation chez Kaleido Biosciences, à Lexington.
   Depuis son arrivée au CEA, sa recherche s'est focalisée sur la biologie moléculaire des usines de fermentation de clostridia, un groupe de bactéries anaérobie ayant un rôle important dans les microbiomes. 3 de ses projets en cours sont d'étudier, de concevoir et d'optimiser Clostridia par ingénierie génomique et biologie de synthèse, évolution dirigée in vivo et reconstruction des réseaux de régulation génique.

2. *Projet(s) de recherche :*
   Son futur projet de recherche est basé sur le fait que les progrès du criblage à haut débit des biomolécules et de l'ingénierie du génome promettent de révolutionner la médecine, l'agriculture et la production d'énergie renouvelable. Son programme se concentrera donc sur le développement de ces technologies pour étudier et concevoir des bactéries qui fermentent les plantes avec des rôles importants dans le cycle du carbone et la santé intestinale, pour finalement appliquer ces connaissances en médecine et en énergie renouvelable.

3. *Publications :*
   Andrew TOLONEN a à son actif 25 publications.
   Il a 3 articles dans Science, 1 dans Nature, 1 dans Nature Communications, et 1 dans Nature Microbiology, de très bons journaux.

4. *Encadrement :*
   Andrew Tolonen a déjà co-encadré deux doctorants, 5 M2 et 4 M1, et un post-doctorant.
   Dans ces 25 publications, 7 ont été rédigées avec un doctorant qu'il a encadré.

Avis du rapporteur : ☒ Favorable

Nom : Florence Gonnet
le 30/10/2020

COMMISSION DE LA RECHERCHE EN FORMATION RESTREINTE
LUNDI 2 NOVEMBRE 2020

Rapport concernant une demande d'autorisation au diplôme d'HDR

Laboratoire : Génomique Métabolique

Candidat à l'HDR :

- Nom : TOLONEN
- Prénom : Andrew
- Diplôme [année] : Doctorat en Biologie Moléculaire obtenu en Août 2005

Avis argumenté :

1. *Activité scientifique* :

Andrew TOLONEN est un microbiologiste qui a été formé dans d'excellents laboratoires où il a utilisé des approches de biologie et génétique moléculaires pour étudier divers microorganismes. Depuis 2011 il s'est intéressé principalement aux *Clostridia*, des bactéries anaérobies qui dégradent des polymères végétaux et jouent de ce fait un rôle important dans le sol comme dans le tractus digestif. Pour mener à bien ses études, il a développé de nouvelles approches d'ingénierie génomique et de biologie synthétique spécifiquement adaptées aux *Clostridia*. Il a aussi montré que les techniques d'évolution dirigée permettaient de sélectionner des souches de *Clostridia* possédant le phénotype désiré. Il a enfin utilisé une approche de biologie des systèmes pour identifier les multiples gènes codant les enzymes impliquées dans la dégradation des produits végétaux et comprendre les régulations complexes mises en jeu pour permettre à *Clostridium phytofermentans* de dégrader efficacement la biomasse végétale.

La qualité de l'activité scientifique d'A. TOLONEN est attestée par l'obtention d'une chaire d'excellence (2011-2016) et de financements (IFP, ANR) pour ses projets.

Andrew TOLONEN a pleinement démontré qu'il pouvait diriger un groupe de recherche, développer et mettre en œuvre des techniques novatrices pour effectuer des travaux originaux. Les résultats qu'il a obtenus sont susceptibles d'avoir des applications biotechnologiques, pour l'utilisation de la biomasse végétale, mais aussi en santé humaine, du fait du rôle joué par les *Clostridia* dans l'intestin.

2. *Projet(s) de recherche :*

Les projets d'A. TOLONEN se situent dans le prolongement des travaux déjà effectués sur *Clostridia*. En utilisant comme modèle *C. phytofermentans*, il va poursuivre une approche de biologie des systèmes pour mieux comprendre comment cette bactérie dégrade les polymères végétaux et interagit avec son environnement. Il va également modifier cette bactérie, par ingénierie génétique ou sélection dirigée, pour obtenir des souches avec des propriétés intéressantes pour des applications biotechnologiques.
Un deuxième aspect des projets concerne l'ingénierie du microbiote intestinal. Des glycanes particuliers peuvent être utilisés comme outils pour modifier la composition du microbiote intestinal et sa capacité à métaboliser différents substrats. Un des objectifs d'A. TOLONEN sera de développer de nouveaux glycanes pour promouvoir la croissance dans le microbiote intestinal des microorganismes bénéfiques à la santé. A long terme la production de ces glycanes par des *Clostridia* modifées est envisagée.

3. *Publications :*

A. TOLONEN a un excellent dossier de publications. Depuis 2002 il a publié 25 articles dans des revues internationales avec comité de lecture. Il est le premier auteur de 10 et le dernier auteur de six de ces articles. Ces articles ont été publiés dans de bons journaux voire dans des journaux de tout premier plan (par ex. 3 articles dans Science, 1 dans Nature, 1 dans Nature Communications, 1 dans Nature Microbiology)

4. *Encadrement :*

A. TOLONEN a déjà une bonne expérience de l'encadrement, ayant encadré quatre étudiants de M1, cinq étudiants de M2, deux doctorants (une thèse soutenue en 2018, une en cours) et un post-doctorant. La qualité de l'encadrement est attestée par le fait que la thèse soutenue en 2018 a conduit à 7 publications signées par le doctorant et A. TOLONEN, dont 3 avec le doctorant en premier auteur et A. TOLONEN en dernier auteur. Cinq des étudiants de master sont aussi auteurs dans des articles publiés par A. TOLONEN. Ceci démontre qu'il possède toutes les qualités pour diriger avec succès les recherches d'étudiants et de doctorants.

Avis du rapporteur extérieur

☒ Favorable - ☐ Assez favorable - ☐ Défavorable

Nom : Jean-Luc PERNODET Signature :

A Orsay le 26/10/2020

Andrew C Tolonen, PhD

http://www.andrewtolonen.com
https://www.linkedin.com/in/andrew-tolonen-b027301/

EXPERIENCED SYSTEMS BIOLOGIST

- Leader of interdisciplinary, international scientific teams
- Mentor of scientists and students to meet ambitious research goals
- Author of successful research proposals and publications in high-impact journals
- Innovator of technologies in DNA/RNA sequencing, small molecule/enzyme screening, metabolomics, computational biology, directed evolution
- Genome engineer in non-model organisms

PERSONAL STATEMENT

My work aims to understand the molecular mechanisms by which microorganisms interact with and modify their environment. In particular, I am interested in identifying how anaerobic microbes transform complex substrates (especially dietary fiber) into value-added chemicals and molecules that influence host immunity. To this end, I develop high-throughput screening and genome engineering technologies to study the physiology of anaerobes and create strains with improved phenotypes.

PROFESSIONAL EXPERIENCE

Alternative Energies and Atomic Energy Commission, France
Researcher 2022-present

- Lab website: https://fermlab.org
- Build a dynamic, extramurally-funded research team using systems biology-based approaches to study and engineer non-model microorganisms for applications in renewable energy and human health: https://fermlab.org/Projects
- Recruit and mentor postdocs and graduate students: https://fermlab.org/People
- Build external industry and academic collaborations

Kaleido Biosciences, Lexington MA 2019-2022
Head of immunobiology, Oncology Team Lead

- Lead a technology development team (5 scientists, 1 research associate) to develop human gut microbiome therapeutics
- Manage Immuno-oncology drug discovery program from pre-clinical testing to IND filing
- Build collaborative relationships with industry and academic opinion leaders (MD Anderson, Institut Gustave Roussy)
- Design and analyze exploratory human trials to rapidly assess efficacy of therapeutics in human studies. Identify novel biomarkers to provide early markers of clinical efficacy
- Communicate results (conferences and peer-review publication)

Broad Institute of MIT and Harvard, Cambridge MA 2016-2019
Associate Director, Microbiomes and Immunity

- Manage a large research program to study the role of microbe-host interactions in inflammatory bowel diseases, autoimmune disease, and disorders of metabolism
    - Studies with clinical samples (blood/biopsies), mouse models, high-throughput sequencing (DNA/RNA/scRNAseq), metabolomics (LC-MS), small molecule screening, microbiology (commensals and pathogens)
- Write proposals, progress reports, journal articles
- Mentor scientists, postdocs, and students in biology, chemistry, computational biology
- Direct collaborations with pharmaceutical partners

Université Paris-Saclay and Genoscope-CEA, France 2011-2016
Associate Professor and Group Leader

- Lead research to develop microbial biocatalysts for production of value-added biochemicals
- Mentor postdocs and graduate students (PhD and Masters level)
- Teach classes in Systems and Synthetic Biology (Masters level), Microbial Genomics (Masters level), Ecology and Evolution of Microbes (Undergraduate level)
- Co-found and Supervise the Évry- l'Université Paris-Saclay team for the International Genetically Engineered Machines Competition (iGEM)

EDUCATIONAL AND POST-DOCTORAL EXPERIENCE

Harvard Medical School, Boston MA 2007-2011
Research Fellow, Dr George Church Laboratory

- Molecular systems biology of Clostridia and large-scale recoding of the *E. coli* genome

Institut Pasteur, Paris, France 2006
Chateaubriand Fellow, Dr Nicole Tandeau de Marsac Laboratory

- Mechanisms of microcystin toxin production by *Microcystis aeruginosa*

Massachusetts Institute of Technology, Cambridge MA 2000-2005
Ph.D. in Biology, Dr Sallie Chisholm Laboratory
Thesis Title: *Prochlorococcus* genetic transformation and the genomics of nitrogen metabolism

Dartmouth College, Hanover NH 1994-1998
Bachelors of Arts Degree with High Honors in Biology

AWARDS AND SUPPORT

Full list of research support: https://orcid.org/0000-0001-5907-4504

PUBLICATIONS

Full lists
Google Scholar: https://scholar.google.com/citations?hl=en&user=fvS99CoAAAAJ
Pubmed: https://pubmed.ncbi.nlm.nih.gov/?term=andrew+tolonen&sort=date

Representative Publications

Tolonen AC, Beauchemin N, Bayne C, Li L, Tan J, Lee J, Meehan BM, Meisner J, Millet Y, LeBlanc G, Kottler R, Rapp E, Murphy C, Turnbaugh PJ, von Maltzahn G, Liu CM, van Hylckama Vlieg JET. (2022). Synthetic glycans control gut microbiome structure and mitigate colitis in mice. *Nat Commun* 10;13(1):1244. doi:10.1038/s41467-022-28856-x

Rostain W, Zaplana T, Boutard M, Baum C, Tabuteau S, Sanitha M, Ramya M, Guss A, Ettwiller L, Tolonen AC. (2022). Tuning of Gene Expression in *Clostridium phytofermentans* Using Synthetic Promoters and CRISPRi *ACS Synth Biol* doi:10.1021/acssynbio.2c00385

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

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* 16;7:13783. doi:10.1038/ncomms13783

Boutard M, Cerisy T, Nogue PY, Alberti A, Weissenbach J, Salanoubat M, Tolonen AC. (2014). Functional diversity of carbohydrate-active enzymes enabling a bacterium to ferment plant biomass. *PLoS Genet* 13;10(11):e1004773. doi:10.1371/journal.pgen.1004773

Career perspective: from plant genomics to microbial engineering for the bioeconomy

Introduction

As I reflect on my professional journey, I am grateful for the rich tapestry of experiences that have shaped my career studying the genetics of microorganisms with application towards environmental carbon cycling, the bioeconomy, and intestinal health. My path, which began with undergraduate research at Dartmouth College, has taken me through diverse landscapes of scientific exploration at MIT, Harvard Medical School, the Genscope-CEA, and Kaleido Biosciences. This personal statement aims to illuminate how each of these stages in my career has not only contributed to my current work, but also deepened my commitment to advancing bio-based, sustainable solutions for a more eco-conscious world.

Chapter 1: Dartmouth - The Seeds of Curiosity

My journey into scientific research began at Dartmouth College, where I embarked on an exploration of plant genomics. Here, I was introduced to the world of *Arabidopsis thaliana*, a small flowering plant that has become a model organism in plant biology. At Dartmouth, I developed my interests in molecular biology to unravel the mysteries of plant development and adaptation. My research focused on deciphering the genetic mechanisms underlying the response of *Arabidopsis* to environmental stressors, shedding light on its remarkable ability to assimilate metals, including iron, copper, and zinc. This experience instilled in me the importance of proper experimental design, data analysis, and the pursuit of knowledge with unwavering dedication. The knowledge gained during my time at Dartmouth became the bedrock of my scientific curiosity. Understanding the fundamental genetic principles governing plant biology provided me with a solid foundation on which to build my future endeavors in the bioeconomy and bioprocessing.

Chapter 2: MIT - PhD studies on photosynthetic bacteria that dominate the world's oceans

Eager to develop my expertise in biological research, I ventured to the Massachusetts Institute of Technology (MIT) for a PhD, where I transitioned from the world of plants to the intriguing realm of photosynthetic microbes, the cyanobacteria. My research at MIT was primarily centered around *Prochlorococcus*, an oceanic cyanobacterium that plays a pivotal role in the Earth's carbon and oxygen cycles. These tiny cyanobacteria (< 1 μM cell diameter) inhabit the vast expanses of the world's oceans, where they contribute significantly to global carbon fixation.

At MIT, I delved into the fascinating intricacies of marine molecular microbiology, studying how *Prochlorocccus* survives in the extremely low nitrogen concentrations in the central oceans without the ability to fix nitrogen. This shift from plant genomics to marine microbiology was a significant pivot in my academic journey, challenging me to adapt to a new discipline and gain a fresh perspective on biological systems. My time at MIT deepened my appreciation for interdisciplinary research and the importance of collaborative efforts to tackle complex environmental issues. This experience not only broadened my scientific horizons but also honed my ability to communicate and collaborate effectively, skills that have proven invaluable in my current research.

Chapter 3: Harvard - postdoctoral research on microbial genome engineering and Clostridia

My academic journey continued with a move to Harvard Medical School for my postdoctoral research, where I focused on Clostridia, a group of anaerobic bacteria with immense potential in bioprocessing and intestinal health. At Harvard, I delved into the mechanisms by which Clostridia ferment plant biomass, seeking to unlock the untapped potential of these bacteria for sustainable bioproducts. This phase of my career was characterized by a mix of experimentation at the bench and computational analysis, where I explored the metabolic pathways of these bacteria and optimized fermentation processes to enhance the yield of valuable bioproducts. The experience at Harvard not only expanded my knowledge of microbial physiology but also instilled in me a profound

appreciation for the potential of microorganisms in addressing global challenges. It was during this time that I recognized the immense potential of microbial fermentation in the emerging bioeconomy, a realization that would shape the trajectory of my future research endeavors.

Chapter 4: Genoscope-CEA - establishing a program to study microbial fermentation

To further broaden my horizons and gain a global perspective on scientific research, I moved to the Genoscope-CEA for a 5 year CNRS Chaire d'Excellence (2011-16). This international experience was instrumental in shaping my current work and provided a unique opportunity to collaborate with researchers from diverse backgrounds. At the Genoscope-CEA, I immersed myself in a vibrant scientific community that was at the forefront of environmental genomics and biotechnology. The multicultural environment of Paris-Saclay enriched my perspective on global challenges and the urgent need for sustainable solutions. I also grew as a mentor, advising two PhD students and numerous Masters and undergraduates, and teaching undergraduate and graduate level courses in microbiology and systems/synthetic biology. It also reinforced the importance of cross-cultural collaboration in addressing complex issues related to the bioeconomy, as I worked alongside colleagues from around the world who shared a common passion for advancing sustainable biotechnology.

Chapter 5 - Kaleido Biosciences - from the lab to the clinic, developing microbiome-based drugs

In my next step, I moved to the private sector at Kaleido Biosciences where I worked to develop drugs that act through the gut microbiome. My research at Kalido focused on how fermentation of complex glycans by the gut microbiome could be leveraged to modulate the composition and metabolic output of the intestinal microbiome. This research illuminated the intricate relationship between the human microbiota and host metabolism, providing insight into the vast potential for

microbiome-based interventions for drug development. Understanding how these microbial communities interact with therapeutic agents has far-reaching implications, from optimizing drug design to personalizing treatment regimens. This experience has underscored the significance of interdisciplinary research in driving innovation within the pharmaceutical industry, ultimately contributing to the development of more effective and tailored medical treatments.

Chapter 6: Genoscope-CEA: a research group focused on the bioeconomy and fermentation of bioproducts

My journey through the diverse landscapes in both public and private research labs has culminated in my current work at the Genscope-CEA, which is centered on the bioeconomy and the fermentation of valuable bioproducts. The knowledge and experiences I gained at each of these institutions have played a pivotal role in informing and shaping my approach to this field. From Dartmouth and MIT, I carry with me a deep appreciation for the importance of fundamental research and the need to understand the intricacies of biological systems at the molecular level. My time at Harvard Medical School and Kaleido Biosciences gave me the tools to collaborate with experts in various fields to develop innovative bioprocessing techniques and harness the potential of microorganisms for sustainable bioproducts.

The knowledge, skills, and perspectives I gained along this journey have equipped me to build and direct a research group to tackle the pressing challenges of our time, from sustainable bioprocessing to the development of environmentally friendly bioproducts. I am deeply grateful for the opportunities I've had to learn, grow, and collaborate with brilliant minds from around the world. As I continue on this path, I am excited about the possibilities that lie ahead. The bioeconomy is a field ripe with potential, and my diverse experiences have prepared me to be a proactive and innovative contributor to its growth. I am committed to directing a research group that pushes the boundaries of what is possible in sustainable biotechnology, driven by the belief that science and collaboration can pave the way to a more sustainable and prosperous future for our planet and its inhabitants.

Research objectives: *Lachnospiraceae* as biocatalysts for a renewable bioeconomy and human health

Abstract

While fermentation has played an important societal role for millenia, its applications are expanding with the rise of biotechnology. In particular, innovations for metabolic engineering of novel microorganisms will help realize the potential of fermentation in the bioeconomy. My research focuses on optimizing fermentation by *Lachnospiraceae,* an understudied family of anaerobic bacteria in the class Clostridia. Some species of *Lachnospiraceae* metabolize abundant, low-cost feedstocks such as lignocellulose and carbon dioxide into value-added chemicals. Others are among the dominant species of the human colon and animal rumen, where they promote healthy gut and immune function. Primarily using *Clostridium phytofermentans* as a model, I am applying high-throughput screening, genome editing, and bioinformatics to create strains with improved phenotypes and to advance our understanding of the biology of *Lachnospiraceae*. Ultimately, my goal is to harness these anaerobes as industrial biocatalysts and therapeutics.

Species of *Lachnospiraceae* inhabit diverse environments including the human intestine [1], human gingival crevice [2], rumen [3], soil [4] and deep sea hydrothermal vents [5]. As such, *Lachnospiraceae* vary widely in their metabolic specializations including carbon dioxide fixation [6], degradation of lignocellulosic polysaccharides [7] or alkanes [5], and demethylation of lignin aromatics [8]. The fermentation products of *Lachnospiraceae* differ greatly across strains, even from the same species, and include alcohols, acids, ketones, phenols, farnesol, pyrazine, and hydrogen [9]. Due to their abilities to transform abundant, low cost feedstocks such as lignocellulose and carbon dioxide into value-added biochemicals, some *Lachnospiraceae* are strong candidates to be developed into industrial biocatalysts.

Lachnospiraceae have also recently attracted much attention because of their high abundance in the human intestine, and metagenomic studies have examined associations between changes in their relative abundances and health. Administration of *Lachnospiraceae* in animal models improved survivorship following full-body radiation [10], ameliorated obesity and type II diabetes [11], protected against food allergy [12], and reduced TNF levels and the incidence of lymphoma [13]. The proposed mechanism by which these *Lachnospiraceae* promote health are by metabolizing dietary fiber to produce short-chain fatty acids and other metabolites

with anti-inflammatory, immuno-regulatory, and homeostatic effects [14]. In the following sections, I will describe my work on *Lachnospiraceae* fermentation to study the physiology of these remarkable bacteria and, ultimately, develop them for applications in biotechnology.

Targeted and evolutionary approaches to engineer fermentation

While *Lachnospiraceae* have many potential applications in industry and medicine, a lack of methods for genetic modification of *Lachnospiraceae* has hindered their use in biotechnology. *Clostridium phytofermentans* (also called *Lachnoclostridium phytofermentans*) is a model *Lachnospiraceae* isolated from forest soil that is easily cultivated and is most closely related to the human intestinal commensals [4]. *C. phytofermentans* is distinguished by its ability to ferment lignocellulose mainly to ethanol, acetate, and hydrogen [4]. This bacterium encodes 184 carbohydrate-active enzymes (CAZymes) [15] to depolymerize diverse plant polysaccharides (cellulose, hemicelluloses, and pectin) into hexoses and pentoses [16], which are taken into the cell and fermented. Due to its oxygen tolerance, simple medium requirements, and mesophilic growth, *C. phytofermentans* is straightforward to culture and represents a model system to study molecular adaptations in *Lachnospiraceae*.

Engineering fermentation pathways by targeted genome editing

To enable targeted genome engineering of *Lachnospiraceae* we established methods for DNA transfer by conjugation [18] and electroporation [19] and applied them for targeted gene inactivation to dissect the genetic basis of cellulose degradation [18], fungal interactions [20], and sugar uptake [21] in *C. phytofermentans*. By combining group II introns and Cre-lox recombination, we made a system for large-scale genomic excision and insertion of multi-gene pathways [22], which we are currently using to import alternative fermentation pathways and study the role of phage defense systems in fitness and interspecies competition. We recently built a system for CRISPR interference (CRISPRi) using a *Lachnospiraceae*-derived Cas12a protein (dLbCas12a) (Fig 2) [19]. We are applying multiplexed CRISPRi in *C. phytofermentans* to tailor fermentation products by

simultaneously targeting genes for metabolism and redox balance (NAD(P)H/NAD(P)+).

This year, we expanded our CRISPR capabilities in *C. phytofermentans* by developing methods for C-to-T base editing using a cytidine deaminase fused to either dCas9 or dLbCas12a and validated this system by introducing a premature stop coding to inactivate the *pyrF* gene. Our data supports that the efficiency of dCas-based C-to-T base editing is >50%, meaning we can isolate mutants without selection and target multiple genes in parallel. Currently, we are applying this system to inactivate genes for acetate production and synthesis of bacterial microcompartments. In the future, I will continue to develop *C. phytofermentans* as a robust, simple platform for genome engineering and apply these methods to dissect and optimize its remarkable fermentation capabilities.

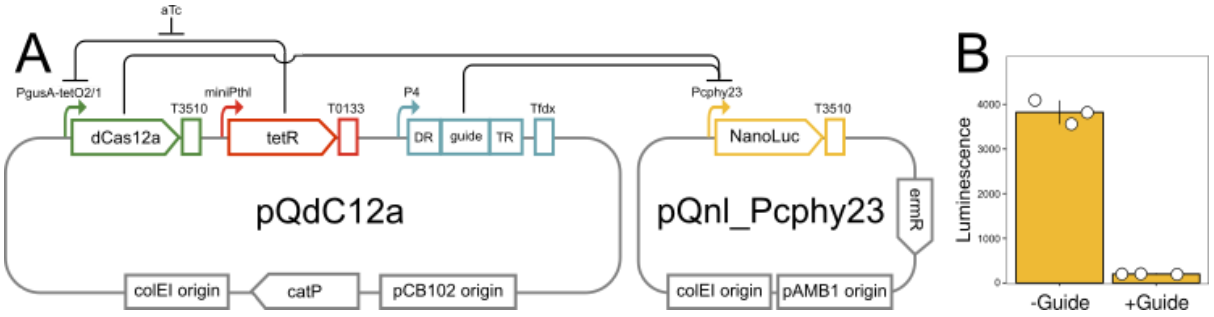


Fig 2 CRISPRi repression of gene expression in *C. phytofermentans*. (A) Two-plasmid system demonstrating dCas12a repression of the NanoLuc reporter. pQdC12a has a tet-repressible dCas12a and gRNA cassette; pQnl_Pcphy23 has the Pcphy23 promoter driving NanoLuc expression. (B) Normalized luminescence of *C. phytofermentans* expressing NanoLuc and dCas12a targeting Pcphy23 with guide or no-guide control. Luminescence was measured ±100 ng mL–1 aTc and normalized to OD600. Data points are individual cultures, and bars are means ± SD of triplicate cultures. Adapted from [23].

In vivo directed evolution of strains with improved fermentation

In vivo directed evolution is a powerful approach to study microbial adaptation and create complex, multigenic phenotypes, even when the responsible genes are unknown. To evolve microorganisms by long-term growth selection, we have fully-automated cultivation devices called GM3 automats at the Genoscope-CEA. We have applied our anaerobic GM3 to evolve *C. phytofermentans* strains with novel phenotypes. Mutant cells with growth advantages are selected in the GM3 using either of two growth regimes: turbidostat or medium-swap. Our cultivation strategy is to acclimate cells to a stressing medium (i.e. increased inhibitor concentration) using

the medium-swap regime (Fig 1A). Once oscillations are replaced by stable growth in the stressing medium, the regime is changed to a turbidostat to select for faster growth (Fig 1B). When the growth rate stabilizes in the stressing medium in the turbidostat, the selection is increased and cultures are again acclimated by medium-swap.

We applied *in vivo* directed evolution to develop inhibitor-resistant strains of *C. phytofermentans* and to study the genetic basis of inhibitor resistance [17].

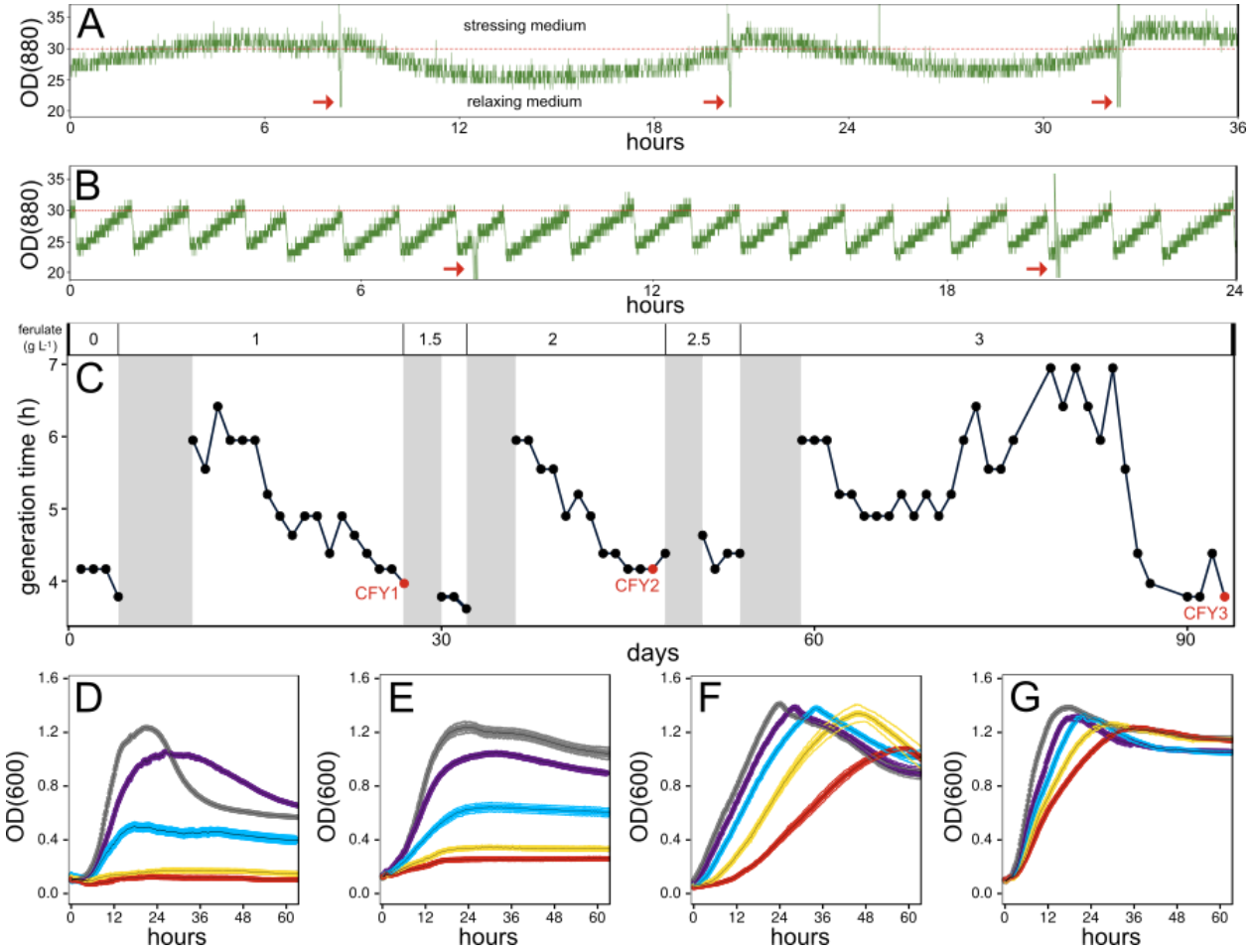


Fig 1 Growth improvement of C. phytofermentans GM3 strains in ferulate medium. (A) Cells were acclimated to increased ferulate using medium-swap mode. (B) Growth rate was improved using turbidostat mode. (C) Growth rate in medium with increasing ferulate concentrations. Shaded areas are periods of medium-swap with fixed 6h generation time. Red points are sample times for physiology and genome sequences. (D to G) Batch culture growth (OD600) of wild-type (D) and increasingly resistant clones growing in medium containing either 0 (gray), 1 (purple), 2 (blue), 4 (yellow), or 6 (red) g/L ferulate. Adapted from [17].

When microorganisms degrade lignocellulose to access sugars, they also release phenolic and acidic inhibitors that affect microbial ecology and impede industrial transformation of biomass. In this study, we initially defined the effects of a panel of

biomass inhibitors, including phenolics, furans, and aliphatic acids, on *C. phytofermentans* growth. We then focused on ferulic acid, a guaiacyl lignin precursor that is one of the most abundant phenolic inhibitors in woods, grasses, and agriculturally important crops. We applied long-term, anaerobic growth selection in a GM3 device to isolate a series of increasingly ferulate-resistant strains (Fig 1C). We isolated clones along the course of evolution (Fig 1D-G) and sequenced their genomes to identify mutations that underlie the resistance phenotype [17].

This study demonstrated an automated framework for *in vivo* evolution of anaerobes and advanced our understanding of the genetic basis of how bacteria evolve to resist chemical inhibitors. We have since evolved *C. phytofermentans* strains with other resistance phenotypes, and applied this approach to other anaerobes. In the future, I will continue to apply *in vivo* directed evolution to study the dynamics of microbial evolution and create strains with novel phenotypes (novel substrate utilization, inhibitor resistances, and enzyme activities). My long-term goal is to develop a generalizable framework whereby we can evolve microbial strains using the GM3 platform, isolate clones with novel phenotypes, and sequence and analyze their genomes to identify responsible mutations.

Engineering fermentation in gut communities

C. phytofermentans is an easily-cultured model system for development and testing of genetic methods for *Lachnospiraceae* that was isolated from soil. One of my future objectives is to port the genetic methods we developed in *C. phytofermentans* to the closely-related, commensal *Lachnospiraceae* from the human gut. For example, *Roseburia intestinalis,* an abundant health-associated commensal in the human intestine [24]. *R. intestinalis* has been linked to differentiation of regulatory T cells, activation of type 3 innate lymphoid cells, and suppression of inflammation through TLR5 [25], making it an excellent platform for engineering of anti-inflammatory probiotics.

My research at Kaleido Biosciences (2019-2021) has given me the expertise to cultivate human gut *Lachnospiraceae*, and to manipulate the composition and metabolic output of the gut microbiome. In a recent study [26], we explored the therapeutic potential of modifying fermentation potential of gut *Lachnospiraceae* using Synthetic Glycans (SGs). We developed a pre-clinical pipeline of hundreds of

synthetic glycans (SGs) that favor specific *Lachnospiraceae* that are specialized to metabolize complex polysaccharides (Fig 3).

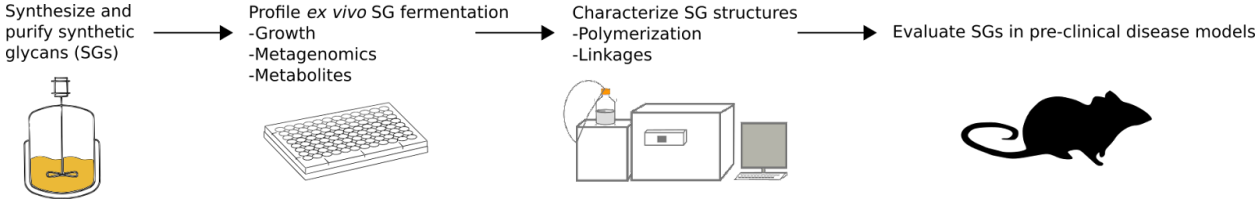


Fig 3 Schematic representation of the preclinical pipeline to evaluate the therapeutic potential of Synthetic Glycans (SGs) using an *ex vivo* platform for highly multiplexed measurements of fecal community fermentation, glycan structural analyses, and mouse models of disease. Adapted from [26].

We examined the effects of SGs on fecal communities using an *ex vivo* platform, leading us to select two SGs that elevated abundances of species of immuno-modulatory *Lachnospiraceae*. We evaluated the abilities of these two SGs to alleviate colitis in mouse models of intestinal damage and disease. In a mouse model of dextran sodium sulfate (DSS)-induced colitis, SG treatment reduced weight loss and improved diarrhea, histology, and endoscopy scores relative to treatment with DSS alone or DSS with a reference glycan. Similarly, SG treatment reduced weight loss, increased survivorship, and improved clinical scores in a mouse model of *C. difficile* infection. Ultimately, my goal is to integrate these synthetic glycans with my approaches for genetic engineering of commensal *Lachnospiraceae*. For example, engineering the CAZyme and sugar metabolism of a gut *Lachnospiraceae* such *R. intenstinalis* that is optimized to metabolize a recalcitrant, synthetic glycan would provide a means of selecting for its growth in a complex community of many species.

Conclusions

Optimized microorganisms that produce value-added chemicals from inexpensive, renewable feedstocks are a vital component of a robust bio-economy. Gut microbiome-based therapeutics will require ways to modify microbiome structure and, ultimately, to control fermentation by its dominant members. Both these societal needs can be addressed by engineering *Lachnospiraceae* species that ferment

low-cost feedstocks and that mitigate intestinal inflammation in humans. By exploiting and expanding carbon utilization and tuning fermentation products in *Lachnospiraceae*, my research will develop these bacteria as advanced biocatalysts for biotechnology.

References

1. Gosalbes MJ, Durbán A, Pignatelli M, Abellan JJ, Jiménez-Hernández N, Pérez-Cobas AE, et al. Metatranscriptomic Approach to Analyze the Functional Human Gut Microbiota. PLOS ONE. 2011;6: e17447. doi:10.1371/journal.pone.0017447
2. Antezack A, Boxberger M, La Scola B, Monnet-Corti V. Isolation and Description of Catonella massiliensis sp. nov., a Novel Catonella Species, Isolated from a Stable Periodontitis Subject. Pathog Basel Switz. 2021;10: 367. doi:10.3390/pathogens10030367
3. Seshadri R, Leahy SC, Attwood GT, Teh KH, Lambie SC, Cookson AL, et al. Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. Nat Biotechnol. 2018;36: 359–367. doi:10.1038/nbt.4110
4. Warnick TA, Methé BA, Leschine SB. Clostridium phytofermentans sp. nov., a cellulolytic mesophile from forest soil. Int J Syst Evol Microbiol. 2002;52: 1155–1160. doi:10.1099/ijs.0.02125-0
5. Schouw A, Leiknes Eide T, Stokke R, Pedersen RB, Steen IH, Bødtker G. Abyssivirga alkaniphila gen. nov., sp. nov., an alkane-degrading, anaerobic bacterium from a deep-sea hydrothermal vent system, and emended descriptions of Natranaerovirga pectinivora and Natranaerovirga hydrolytica. Int J Syst Evol Microbiol. 2016;66: 1724–1734. doi:10.1099/ijsem.0.000934
6. Greening RC, Leedle JA. Enrichment and isolation of Acetitomaculum ruminis, gen. nov., sp. nov.: acetogenic bacteria from the bovine rumen. Arch Microbiol. 1989;151: 399–406. doi:10.1007/BF00416597
7. Boutard M, Cerisy T, Nogue P-Y, Alberti A, Weissenbach J, Salanoubat M, et al. Functional diversity of carbohydrate-active enzymes enabling a bacterium to ferment plant biomass. PLoS Genet. 2014;10: e1004773. doi:10.1371/journal.pgen.1004773
8. Doré J, Bryant MP. Metabolism of One-Carbon Compounds by the Ruminal Acetogen Syntrophococcus sucromutans. Appl Environ Microbiol. 1990;56: 984–989. doi:10.1128/aem.56.4.984-989.1990
9. Abdugheni R, Wang W-Z, Wang Y-J, Du M-X, Liu F-L, Zhou N, et al. Metabolite profiling of human-originated Lachnospiraceae at the strain level. iMeta. n/a: e58. doi:10.1002/imt2.58
10. Guo H, Chou W-C, Lai Y, Liang K, Tam JW, Brickey WJ, et al. Multi-omics analyses of radiation survivors identify radioprotective microbes and metabolites. Science. 2020;370. doi:10.1126/science.aay9097
11. Hosomi K, Saito M, Park J, Murakami H, Shibata N, Ando M, et al. Oral administration of Blautia wexlerae ameliorates obesity and type 2 diabetes via metabolic remodeling of the gut microbiota. Nat Commun. 2022;13: 4477. doi:10.1038/s41467-022-32015-7
12. Feehley T, Plunkett CH, Bao R, Choi Hong SM, Culleen E, Belda-Ferre P, et al. Healthy infants harbor intestinal bacteria that protect against food allergy. Nat Med. 2019;25: 448–453. doi:10.1038/s41591-018-0324-z
13. Lu H, Xu X, Fu D, Gu Y, Fan R, Yi H, et al. Butyrate-producing Eubacterium rectale suppresses lymphomagenesis by alleviating the TNF-induced TLR4/MyD88/NF-κB axis. Cell Host Microbe. 2022;30: 1139-1150.e7. doi:10.1016/j.chom.2022.07.003
14. Reichardt N, Duncan SH, Young P, Belenguer A, McWilliam Leitch C, Scott KP, et al.

Phylogenetic distribution of three pathways for propionate production within the human gut microbiota. ISME J. 2014;8: 1323–1335. doi:10.1038/ismej.2014.14
15. Drula E, Garron M-L, Dogan S, Lombard V, Henrissat B, Terrapon N. The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Res. 2022;50: D571–D577. doi:10.1093/nar/gkab1045
16. Tolonen AC, Haas W, Chilaka AC, Aach J, Gygi SP, Church GM. Proteome-wide systems analysis of a cellulosic biofuel-producing microbe. Mol Syst Biol. 2011;7: 461. doi:10.1038/msb.2010.116
17. Cerisy T, Souterre T, Torres-Romero I, Boutard M, Dubois I, Patrouix J, et al. Evolution of a biomass-fermenting bacterium to resist lignin phenolics. Appl Environ Microbiol. 2017. doi:10.1128/AEM.00289-17
18. Tolonen AC, Chilaka AC, Church GM. Targeted gene inactivation in *Clostridium phytofermentans* shows that cellulose degradation requires the family 9 hydrolase Cphy3367. Mol Microbiol. 2009;74: 1300–1313. doi:10.1111/j.1365-2958.2009.06890.x
19. Rostain W, Zaplana T, Boutard M, Baum C, Tabuteau S, Sanitha M, et al. Tuning of Gene Expression in Clostridium phytofermentans Using Synthetic Promoters and CRISPRi. ACS Synth Biol. 2022. doi:10.1021/acssynbio.2c00385
20. Tolonen AC, Cerisy T, El-Sayyed H, Boutard M, Salanoubat M, Church GM. Fungal lysis by a soil bacterium fermenting cellulose. Environ Microbiol. 2015;17: 2618–2627. doi:10.1111/1462-2920.12495
21. Cerisy T, Iglesias A, Rostain W, Boutard M, Pelle C, Perret A, et al. ABC Transporters Required for Hexose Uptake by Clostridium phytofermentans. J Bacteriol. 2019;201. doi:10.1128/JB.00241-19
22. Cerisy T, Rostain W, Chhun A, Boutard M, Salanoubat M, Tolonen AC. A Targetron-Recombinase System for Large-Scale Genome Engineering of Clostridia. mSphere. 2019;4. doi:10.1128/mSphere.00710-19
23. Rostain W, Zaplana T, Boutard M, Baum C, Tabuteau S, Sanitha M, et al. Tuning of Gene Expression in Clostridium phytofermentans Using Synthetic Promoters and CRISPRi. ACS Synth Biol. 2022;11: 4077–4088. doi:10.1021/acssynbio.2c00385
24. Nie K, Ma K, Luo W, Shen Z, Yang Z, Xiao M, et al. Roseburia intestinalis: A Beneficial Gut Organism From the Discoveries in Genus and Species. Front Cell Infect Microbiol. 2021;11: 757718. doi:10.3389/fcimb.2021.757718
25. Shen Z, Luo W, Tan B, Nie K, Deng M, Wu S, et al. Roseburia intestinalis stimulates TLR5-dependent intestinal immunity against Crohn's disease. eBioMedicine. 2022;85. doi:10.1016/j.ebiom.2022.104285
26. Tolonen AC, Beauchemin N, Bayne C, Li L, Tan J, Lee J, et al. Synthetic glycans control gut microbiome structure and mitigate colitis in mice. Nat Commun. 2022;13: 1244. doi:10.1038/s41467-022-28856-x

frontiers | Frontiers in Bioengineering and Biotechnology

TYPE Review
PUBLISHED 04 January 2024
DOI 10.3389/fbioe.2023.1324396

OPEN ACCESS

EDITED BY
Donatella Cimini,
University of Campania Luigi Vanvitelli, Italy

REVIEWED BY
Zhengming Zhu,
Nanjing Tech University, China
Zhenshang Xu,
Qilu University of Technology, China

*CORRESPONDENCE
Andrew C. Tolonen,
atolonen@genoscope.cns.fr

RECEIVED 19 October 2023
ACCEPTED 05 December 2023
PUBLISHED 04 January 2024

CITATION
Zaplana T, Miele S and Tolonen AC (2024), *Lachnospiraceae* are emerging industrial biocatalysts and biotherapeutics.
Front. Bioeng. Biotechnol. 11:1324396.
doi: 10.3389/fbioe.2023.1324396

Lachnospiraceae are emerging industrial biocatalysts and biotherapeutics

Tom Zaplana, Solange Miele and Andrew C. Tolonen*

Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, University of Evry, Université Paris-Saclay, Evry, France

The *Lachnospiraceae* is a family of anaerobic bacteria in the class Clostridia with potential to advance the bio-economy and intestinal therapeutics. Some species of *Lachnospiraceae* metabolize abundant, low-cost feedstocks such as lignocellulose and carbon dioxide into value-added chemicals. Others are among the dominant species of the human colon and animal rumen, where they ferment dietary fiber to promote healthy gut and immune function. Here, we summarize recent studies of the physiology, cultivation, and genetics of *Lachnospiraceae*, highlighting their wide substrate utilization and metabolic products with industrial applications. We examine studies of these bacteria as Live Biotherapeutic Products (LBPs), focusing on *in vivo* disease models and clinical studies using them to treat infection, inflammation, metabolic syndrome, and cancer. We discuss key research areas including elucidation of intra-specific diversity and genetic modification of candidate strains that will facilitate the exploitation of *Lachnospiraceae* in industry and medicine.

KEYWORDS

***Lachnospiraceae*, microbial biotechnology, microbiome, biocatalyst, live biotherapeutics, Clostridia**

Introduction

The *Lachnospiraceae*, from the ancient Greek "lachnos" (wooly hair) and "spira" (coil, twist), is a family of anaerobic, mesophilic bacteria with Gram-positive ultrastructure. *Lachnospiraceae*, which generally correspond to Clostridia cluster XIVa (Whitman, 2009), inhabit diverse ecosystems. Host-associated species are found in the gastrointestinal tracts of humans (Gosalbes et al., 2011), mice (Meehan and Beiko, 2014), insects (Vera-Ponce de Leon et al., 2022), and ruminants (Seshadri et al., 2018) as well as the human gingival crevice (Antezack et al., 2021). Metagenomic studies have shown *Lachnospiraceae* account for 10%–45% of the total bacteria in feces of healthy adults (Liu C. et al., 2021) and have a life-long association with humans; they colonize the guts of infants and are enriched in the fecal microbiomes of long-living (>90 years old) individuals (Kong et al., 2016). Other *Lachnospiraceae* live in anaerobic soil (Hengstmann et al., 1999) where they recycle plant matter and mediate biological soil disinfestation, a pesticide-free method to control soil-borne pathogens (Huang et al., 2019). *Lachnospiraceae* also inhabit aquatic sediments (Lomans et al., 2001; Dai et al., 2016), Antarctic green snow (Smirnova et al., 2021), wastewater (McLellan et al., 2013), and deep sea hydrothermal vents (Schouw et al., 2016).

Recent interest in the important roles of *Lachnospiraceae* in gut and environmental ecosystems has led to advances in the genomics, cultivation, and genetic manipulation of these bacteria. Here, we examine these advances and explore how they have set the stage for applying *Lachnospiraceae* in industrial and medical biotechnology. Many metagenomic

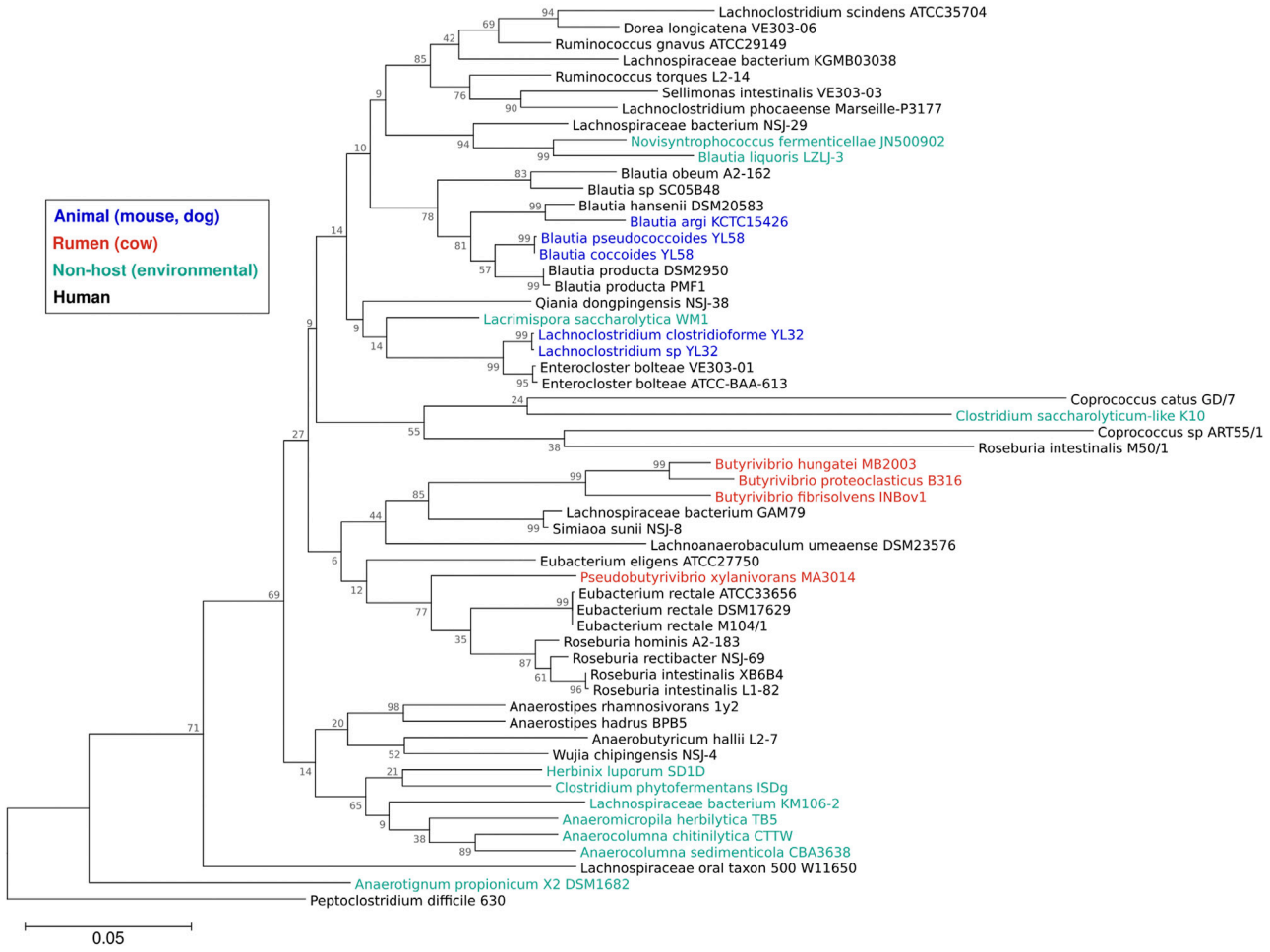


FIGURE 1
Phylogeny of the family *Lachnospiraceae*. Sequences of the 16S rRNA gene from the 57 *Lachnospiraceae* finished genomes in the IMG database were aligned using Muscle (Edgar, 2004) and a maximum likelihood tree (1,000 bootstraps) was constructed using Mega11 software (Tamura et al., 2021) with *Clostridioides (Peptoclostridium) difficile* 630 as the outgroup. Colors show strain isolation sites: animal (dog, mouse) feces (blue), cow rumen (red), human feces or oral cavity (black), or non-host environmental (aqua). Branch numbers are bootstrap percentages, branch lengths are substitution per site.

studies comparing the fecal communities of healthy and diseased subjects have drawn associations between *Lachnospiraceae* and human health. However, associations from fecal metagenomics are sometimes difficult to interpret or conflicting, likely due to differences in study design and strain-specific differences, and have been reviewed previously (Vacca et al., 2020; Liu X. et al., 2021). Thus, we focus on insights gained from preclinical, rodent models and human clinical studies involving administration of live *Lachnospiraceae*. We discuss opportunities and current needs to develop *Lachnospiraceae* to produce value-added biochemicals from low-cost feedstocks and as live biotherapeutic products (LBPs).

Phylogeny and genomes

Our knowledge of *Lachnospiraceae* diversity and genomics has greatly expanded over the past decade. In 2014, the NCBI taxonomy of *Lachnospiraceae* included 24 genera (Meehan and Beiko, 2014), which by 2023 had increased to 118 genera with 1,941 species (NCBI Genome Datasets, 2023). The number of *Lachnospiraceae* genomes has similarly increased. In 2014, NCBI included 30 *Lachnospiraceae* genomes (Meehan and Beiko, 2014), which reached 201 genomes in 2023 (NCBI Genome Datasets, 2023). The Integrated Microbial Genomes & Microbiomes (IMG) portal (Chen I.-M. A. et al., 2023) includes 1,292 *Lachnospiraceae* genomic sequences, of which 57 are finished genomes. Most of the finished *Lachnospiraceae* genomes are of strains isolated from human feces or oral cavity, with strains from the cow rumen, animal feces, and environmental isolates forming clusters in the *Lachnospiraceae* phylogeny (Figure 1).

The GC content of *Lachnospiraceae* genomes varies between 35%–50%, with members of the same genus typically sharing similar GC levels (Sorbara et al., 2020). The distribution of *Lachnospiraceae* genome sizes (mean 4.06 Mb, median 3.59 Mb) is similar across habitats (Figure 2A). *Herbinix luporum* SD1D, a cellulose-degrading strain isolated from a biogas reactor, has the smallest genome of 2.6 Mb encoding 2,632 genes (Koeck et al., 2016). *Lachnoclostridium clostridioforme* YL32, from the mouse gut, has the largest genome of 7.2 Mb with 7,735 genes (Garzetti et al., 2017). Among these finished

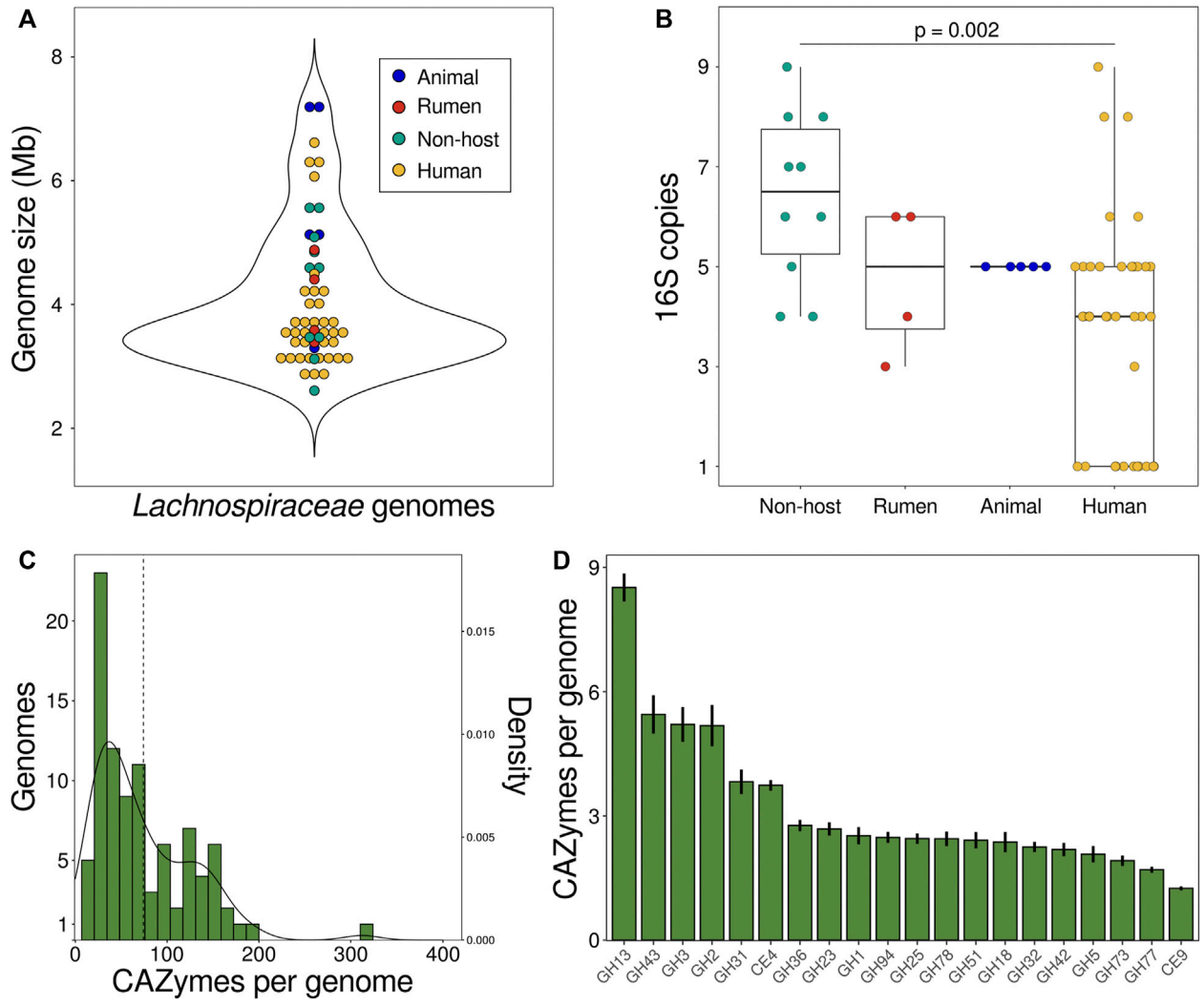


FIGURE 2
Characteristics of *Lachnospiraceae* genomes. **(A)** Genomes sizes and **(B)** number of copies of the 16S gene in 57 finished *Lachnospiraceae* genomes in the IMG database. Each data point is a genome and colors show strain isolation sites: animal (dog, mouse) feces (blue), cow rumen (red), human feces or oral cavity (yellow), or non-host environmental (aqua). **(C)** Number of CAZyme genes encoded by *Lachnospiraceae* genomes. The dashed line is the mean number of CAZymes per genome and the curve shows the kernel density estimation. **(D)** Number of CAZyme genes per genome for the 20 most abundant CAZyme families summed across *Lachnospiraceae* genomes. Bars show the mean ± SEM. Data in **(C,D)** is for degradative enzymes (glycoside hydrolases, carbohydrate esterases, pectin lyases) for 98 species in the CAZy database defined as *Lachnospiraceae* based on NCBI taxonomy. Mb DNA megabases. SEM standard error of the mean.

genomes, non-host species have a higher number of 16S genes than species isolated from humans (Figure 2B). Higher ribosomal gene copy number is linked to increased maximum growth rate and ability to respond to resource changes (Stevenson and Schmidt, 2004), suggesting these traits are under greater selection in non-host species.

Substrate utilization

Species of *Lachnospiraceae* collectively metabolize plant biomass through the assimilation of polysaccharides, peptides, and aromatics as well as subsequent transformation of the fermentation products by acetogens and cross-feeding species (Figure 3). As primary degraders of plant biomass, many *Lachnospiraceae* ferment a variety of complex polysaccharides including glucans, mannans, xylans, galactans, pectins, and arabinans (Boutard et al., 2014). Biomass-fermenting species sometimes grow faster on polysaccharides than on the constituent monosaccharides (Boutard et al., 2016). While typically able to ferment multiple polysaccharides, species of *Lachnospiraceae* are ecologically differentiated by specializing on certain substrates. As such, addition of different glycans results in compound-specific changes to the relative abundances of *Lachnospiraceae* in mixed communities (Tolonen et al., 2022). For example, *Roseburia intestinalis* efficiently metabolizes β-mannans and xylan (Leth et al., 2018; La Rosa et al., 2019), while *Roseburia faecis* ferments arabinogalactan (Sheridan et al., 2016). *Lachnospira* include pectin specialists, called pectinophiles (Cornick et al., 1994), that are stimulated by pectin addition (Bang et al., 2018). *Lachnospiraceae*

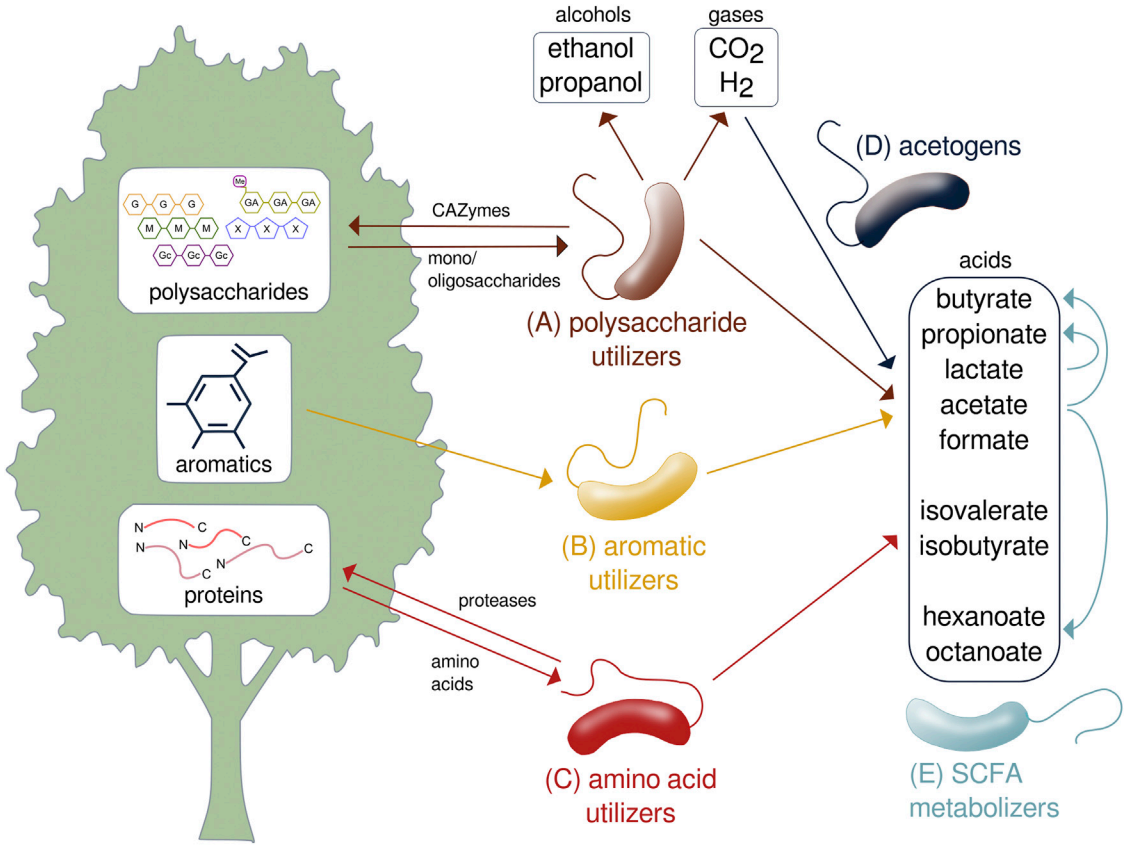


FIGURE 3
Metabolic transformation of lignocellulose performed by species of *Lachnospiraceae*. Specialized species metabolize the **(A)** polysaccharide, **(B)** aromatic, and **(C)** protein components of lignocellulose to produce gases, acids, and alcohols. **(D)** Acetogens transform carbon dioxide and hydrogen to SCFAs. **(E)** SCFAs are further transformed and assimilated by other species. SCFA, short-chain fatty acid; CO_2, carbon dioxide; H_2, hydrogen; CAZyme, carbohydrate-active enzyme.

specialized to metabolize cellulose are found in soil (Warnick et al., 2002; Wolin et al., 2003; Dai et al., 2016) and the rumen (Cai and Dong, 2010).

Genomes of plant-fermenting *Lachnospiraceae* encode numerous carbohydrate-active enzymes (CAZymes), each of which cleave a specific glycosidic linkage to depolymerize complex glycans. *Lachnospiraceae* genomes encode an average of 73 degradative CAZymes (glycoside hydrolases, pectin lyases, carbohydrate esterases) with less than 10% of genomes encoding more than 150 degradative CAZymes (Figure 2C). *Lachnospiraceae* bacterium CE91-St56, closely related to *Eisenbergiella massiliensis*, encodes over 100 more CAZymes than any other strain with 313 CAZymes including 294 glycoside hydrolases (Figure 2C). CE91-St56 was isolated from a supercentenarian (>110 years old) (Sato et al., 2021), highlighting the presence of *Lachnospiraceae* with extensive polysaccharide utilization capabilities in long-lived individuals. *Lachnospiraceae* genomes encode many families of CAZymes, the most abundant of which is GH13 to depolymerize α-glucans including starch and pullulan (Figure 2D). Other abundant CAZyme families enable degradation of substrates such as β-glucans (GH3,5,51,94), β-galactans (GH42), xylan (GH43), pectin (GH78), and chitin (GH18) (Figure 2D).

Some *Lachnospiraceae* assimilate C1 and C2 carbon sources that are produced during the initial stages of plant biomass fermentation. *Blautia hydrogenotrophica* is a human gut acetogen that grows autotrophically using hydrogen to fix carbon dioxide into acetyl-CoA by the Wood–Ljungdahl pathway (Bernalier et al., 1996) and other *Lachnospiraceae* acetogens consume H2 and CO_2 in the cow rumen (Greening and Leedle, 1989). Species that assimilate other C1 and C2 molecules include *Sporobacterium*, isolated from an olive mill, that grows on methanol (Mechichi et al., 1999). Various *Lachnospiraceae* such as *Anaerobutyricum soehngenii* (previously *Eubacterium hallii*) metabolize acetate to butyrate (Udayappan et al., 2016; Zhang et al., 2019), which represents an important cross-feeding interaction in the gut (Flint et al., 2012).

Other *Lachnospiraceae* are specialized to metabolize peptides, alkanes, and aromatics. *Falcatimonas natans*, isolated from a methanogenic reactor, ferments peptides but not carbohydrates (Watanabe et al., 2016). *Abyssivirga alkaniphila*, a hydrothermal vent species, can grow on straight and branched alkanes (C5 to C25) using thiosulfate as an external electron acceptor (Schouw et al., 2016). *Syntrophococcus sucromutans* can oxidize sugars to acetate using formate or methoxymonobenzenoids as electron acceptors (Krumholz and Bryant, 1986). *Parasporobacterium* and *Sporobacterium* are soil bacteria that grow on methoxylated

TABLE 1 Fermentation products of Lachnospiraceae.

Product	Input	Enzymatic steps (EC numbers)	Representative Species	References
Formic acid	Pyruvate	2.3.1.54	*Roseburia intestinalis*	Hillman et al. (2020)
Acetic acid	Acetyl-CoA	2.3.1.8, 2.7.2.1	*Fusicatenibacter saccharivorans*	Takada et al. (2013)
Propionic acid	Lactate (acrylate pathway)	2.8.3.1, 4.2.1.54, 1.3.1.84	*Coprococcus cactus*	Sheridan et al. (2022)
	L-fucose (propanediol pathway)	5.3.1.25, 2.7.1.51, 4.1.2.17, 1.1.1.77, 4.2.1.28, 1.2.1.87, 2.8.3.1	*Roseburia inulinivorans*	Scott et al. (2006)
	L-rhamnose (propanediol pathway)	5.3.1.14, 2.7.1.15, 4.1.2.19, 1.1.1.77, 4.2.1.28, 1.2.1.87, 2.8.3.1	*Lachnoclostridium phytofermentans*	Petit et al. (2013)
	3-oxopropionate (myo-inositol pathway)	2.8.3.5, 1.1.1.35, 4.2.1.116, 1.3.1.95, 2.8.3.1	*Anaerostipes rhamnosivorans*	Bui et al. (2021)
L-lactic acid	Pyruvate	1.1.1.27	*Lachnoclostridium phytofermentans*	Tolonen et al. (2011)
Butyric acid	Acetyl-CoA, (BCoAT)	2.3.1.9, 1.1.1.157, 4.2.1.55, 1.3.1.86, 2.8.3.8	*Roseburia intestinalis*	Vital et al. (2014)
	Acetyl-CoA (Butyrate kinase)	2.3.1.9, 1.1.1.157, 4.2.1.55, 1.3.1.86, 2.3.1.17, 2.7.2.7	*Eubacterium ventriosum*	Vital et al. (2014)
	Lysine	5.4.3.2, 5.4.3.3, 1.4.1.11, 2.3.1.247, 4.3.1.14, 1.3.1.86, 2.8.3.8	*Lachnospiraceae sp F0167*	Vital et al. (2014)
	Glutarate	1.1.99.2, 2.8.3.12, 4.2.1.167, 7.2.4.5, 1.3.1.86, 2.8.3.8	*Clostridiales* sp SS3/4	Vital et al. (2014)
	4-aminobutyrate	1.1.1.61, 2.8.3-, 4.2.1.120, 1.3.1.86, 2.8.3.8	*Anaerostipes caccae*	Vital et al. (2014)
Isobutyric acid	Valine	1.4.1.23/2.6.1.66, 4.1.1.72, 1.2.1.3	*Falcatimonas natans*	Watanabe et al. (2016)
Succinate	Pyruvate	2.7.1.40, 4.1.1.49, 1.1.1.37, 4.2.1.2, 1.3.5.1	*Blautia wexlerae*	Hosomi et al. (2022)
Isovaleric acid	Leucine	1.4.1.9/2.6.1.6, 4.1.1.72, 1.2.13	*Falcatimonas natans*	Watanabe et al. (2016)
5-aminovaleric acid	Proline	1.21.4.1	*Dorea longicatena*	Lopez et al. (2020)
Indoleacetic acid	Tryptophan	2.6.1.57, 4.1.1.74, 1.2.3.7	*Anaerobutyricum soehngenii*	Russell et al. (2013)
Hexanoic, octanoic acid	Acetyl-CoA	2.3.1.16, 1.1.1.35, 4.2.1.17, 1.3.8.7, 3.1.2.20	*Candidatus Weimeria bifida*	Scarborough et al. (2020)
Ethanol	Acetyl-CoA	1.2.1.10, 1.1.1.1	*Lachnoclostridium phytofermentans*	Tolonen et al. (2015b)
1-Propanol	L-fucose	5.3.1.25, 2.7.1.51, 4.1.2.17, 1.1.1.77, 4.2.1.28, 1.1.1.1	*Roseburia inulinivorans*	Scott et al. (2006)
	L-rhamnose	5.3.1.14, 2.7.1.15, 4.1.2.19, 1.1.1.77, 4.2.1.28, 1.1.1.1	*Lachnoclostridium phytofermentans*	Petit et al. (2013)
Hydrogen	Ferredoxin, H+	1.12.7.2	*Roseburia intestinalis*	Dostal et al. (2015)

Shown are the metabolic product, metabolite entering the pathway, enzymatic steps comprising the pathway, and a representative species containing the pathway with a supporting reference. EC number, Enzyme Commission number.

aromatic compounds as sole carbon and energy sources by catabolizing them to short-chain fatty acids (SCFAs) (Lomans et al., 2001). These aromatics are degraded by transforming the side chains and cleaving the aromatic ring using the phloroglucinol pathway (Whitman, 2009), as has recently been shown for *Clostridium scatologenes* (Zhou et al., 2023).

Metabolic products

Lachnospiraceae metabolism yields alcohols, gases, and acids with importance in industry and human health (Table 1). Species producing industrially-relevant alcohols include *Lachnoclostridium phytofermentans*, which ferments cellulose to ethanol at 68% of the maximum theoretical yield (Tolonen et al., 2011) and, similar to *Roseburia inulinivorans*, metabolizes fucose and rhamnose to 1-propanol through a 1,2-propanediol intermediate in polyhedral microcompartments (Scott et al., 2006; Petit et al., 2013). *Lachnospiraceae* produce hydrogen by co-expressing monomeric and multimeric, bifurcating [FeFe] hydrogenases along with energy conserving [NiFe] hydrogenases (Calusinska et al., 2010) to enable yields reaching 2–3 moles of hydrogen per mole glucose (Harvey et al., 2008).

Fermentation of dietary fiber by gut *Lachnospiraceae* yields three main SCFA (acetate, propionate, and butyrate) (Table 1), all of which have been shown to benefit health. Acetate, which is synthesized in two steps from acetyl-CoA, reduces adipose accumulation and improves glucose tolerance (Yamashita et al., 2007). *Lachnospiraceae*

use the acrylate, propanediol, and myo-inositol pathways to synthesize propionate (Reichardt et al., 2014; Bui et al., 2021), which is absorbed from the gut into the bloodstream to regulate cholesterol (Berggren et al., 1996) and reduce visceral and liver fat accumulation (Chambers et al., 2015). Propionate also enhances satiety (Arora et al., 2011), making it a potential way to reduce obesity. Along with *Ruminococcaceae*, *Lachnospiraceae* are the dominant butyrate producers in the human gut (Vital et al., 2014). While butyrate is most often synthesized from acetyl-CoA using butyryl-CoA transferase or butyrate kinase, some *Lachnospiraceae* can produce butyrate from lysine, glutarate, and 4-aminobutyrate (Vital et al., 2014), highlighting their abilities produce butyrate from different nutritional sources. Butyrate is the preferred energy source of colonocytes (Litvak et al., 2018), suppresses pathogens (Walker et al., 2021), and stimulates differentiation of anti-inflammatory T-regulatory cells (Furusawa et al., 2013).

Other organic acids produced by *Lachnospiraceae* include the branched-chain fatty acids isobutyrate and isovalerate that are synthesized from valine and leucine, respectively (Watanabe et al., 2016). *Candidatus* Weimeria bifida metabolizes pentoses to medium-chain fatty acids (hexanoate and octanoate) using the reverse β-oxidation cycle for SCFA chain elongation (Scarborough et al., 2020). Species of *Blautia* metabolize pyruvate to succinate (Hosomi et al., 2022) and, depending on the species, *Lachnospiraceae* can either produce L-lactate (Tolonen et al., 2011) or metabolize DL-lactate to acetate or butyrate using stereospecific lactate dehydrogenases (Sheridan et al., 2022). Gut *Lachnospiraceae* ferment proline to 5-aminovalerate in competition with *Clostridioides difficile*, representing a potential means to prevent colonization by this pathogen (Lopez et al., 2020). *Lachnospiraceae* ferment aromatic amino acids (tryptophan, tyrosine, phenylalanine) to phenolic and indolic acids (Russell et al., 2013), including compounds such as indoleacetic acid that promote intestinal homeostasis by signaling through the aryl hydrocarbon receptor (Roager and Licht, 2018).

Lachnospiraceae also produce other metabolites with health and industrial applications. *Lachnospiraceae* can modify and cleave the heterocyclic C-ring of flavonoids (Braune and Blaut, 2016) yielding molecules such as equol, which is linked to prevention of colorectal cancer (Sugiyama et al., 2013) and aging-related disorders (Mayo et al., 2019). Gut *Lachnospiraceae* generate reactive sulfur species that protect the host from oxidative stress-induced liver injury (Uchiyama et al., 2022). *Lachnospiraceae* produce farnesol (Abdugheni et al., 2022), an isoprene-derived molecule with anti-inflammatory and neuroprotective activities (Sell et al., 2022). Farnesol also has industrial applications as a fragrance ingredient (Lapczynski et al., 2008) and diesel fuel precursor (Rude and Schirmer, 2009). *Butyrivibrio fibrisolvens* synthesizes an exopolysaccharide with potential industrial applications that is rheologically similar to xanthan gum, but is composed of rare sugars including L-altrose and L-iduronic acid (Wachenheim and Patterson, 1992; Ferreira et al., 1997).

Lachnospiraceae are a potential source of other antimicrobial and immunomodulatory compounds. *Blautia obeum* synthesizes a lantibiotic, nisin O, that inhibits pathogens, including *C. difficile* and *Clostridium perfringens* (Hatziioanou et al., 2017). *Lachnospiraceae* produce pyrazines, which are being developed as antimicrobial and anti-fungal drugs (Hassan et al., 2020). Gut *Lachnospiraceae* convert primary bile acids by 7α-dehydroxylation to secondary bile acids that inhibit enteric pathogens and regulate mucosal immunity (Jin et al., 2022). Nonribosomal peptide synthetases (NRPS) of human gut *Lachnospiraceae* produce immunomodulatory secondary metabolites such as di-peptide aldehydes that act as cell-permeable cathepsin inhibitors, which could act as immunosuppressors by blocking antigen processing (Guo et al., 2017).

Cultivation and engineering

Lachnospiraceae are mesophiles that grow at 30°C–45°C and can be cultivated under standard anaerobic conditions using jars with anaerobic sachets or a glove box. A human gut *Lachnospiraceae* biobank (hLchsp) was established as part of the China General Microorganism Culture Collection by isolating strains from healthy adult fecal samples, yielding a collection of 77 species across 33 genera with *Lachnospira*, *Blautia*, and *Roseburia* being the most abundant genera (Abdugheni et al., 2022). Among the seven growth media used for isolations, Yeast Casitone Fatty Acid (YCFA) medium supported the greatest number of *Lachnospiraceae* species, which comprised 19.6% of all bacterial isolates from the fecal samples. Another project to cultivate human gut *Lachnospiraceae* using rich, non-selective media yielded 273 isolates, including all the *Lachnospiraceae* genera that were detected by metagenomic sequences of the donor feces (Sorbara et al., 2020). In contrast to gut species, more oligotrophic, GS2 medium is used to cultivate environmental heterotrophic *Lachnospiraceae* such as *Herbinix hemicellulosilytica* and *L. phytofermentans* (Warnick et al., 2002; Koeck et al., 2015). The acetogen *B. hydrogenotrophica* can be grown on DMSZ 114 or general-acetogen (GA) medium, either as an autotroph using H_2:CO_2 or as a heterotroph by adding a carbon source (Groher and Weuster-Botz, 2016).

Methods for the genetic modification of *Lachnospiraceae* are being developed to study their molecular biology and to engineer optimized strains. Early studies found that conjugative transposons bearing antibiotic resistance genes transfer DNA between *Lachnospiraceae* (Barbosa et al., 1999). Experimental methods have been developed to transfer plasmid DNA into species of *Lachnoclostridium*, *Roseburia*, *Eubacterium*, *Enterocloster*, *Lacrimispora*, and *Blautia* by conjugation with *Escherichia coli* (Tolonen et al., 2009; Cuív et al., 2015; Sheridan et al., 2019; Jin et al., 2022) and by electroporation into species of *Lachnoclostridium* and *Butyrivibrio* (Beard et al., 1995; Rostain et al., 2022) (Figure 4A).

Native plasmids from *Lachnospiraceae* have been used as expression vectors (Hefford et al., 1997) and the pMTL plasmid system (Heap et al., 2009) can be applied to identify plasmid origins and resistance markers that function in strains of interest (Jin et al., 2022; Rostain et al., 2022). Plasmid origins that have been shown to replicate in *Lachnospiraceae* include pCB102 from *Clostridium butyricum*, pBP1 from *Clostridium botulinum*, pAMβ1 from *Enterococcus faecalis*, pWV101 from *Lactococcus lactis*, pIM13 from *Bacillus subtilis*, and pCD6 from *C. difficile*, and these plasmids can be

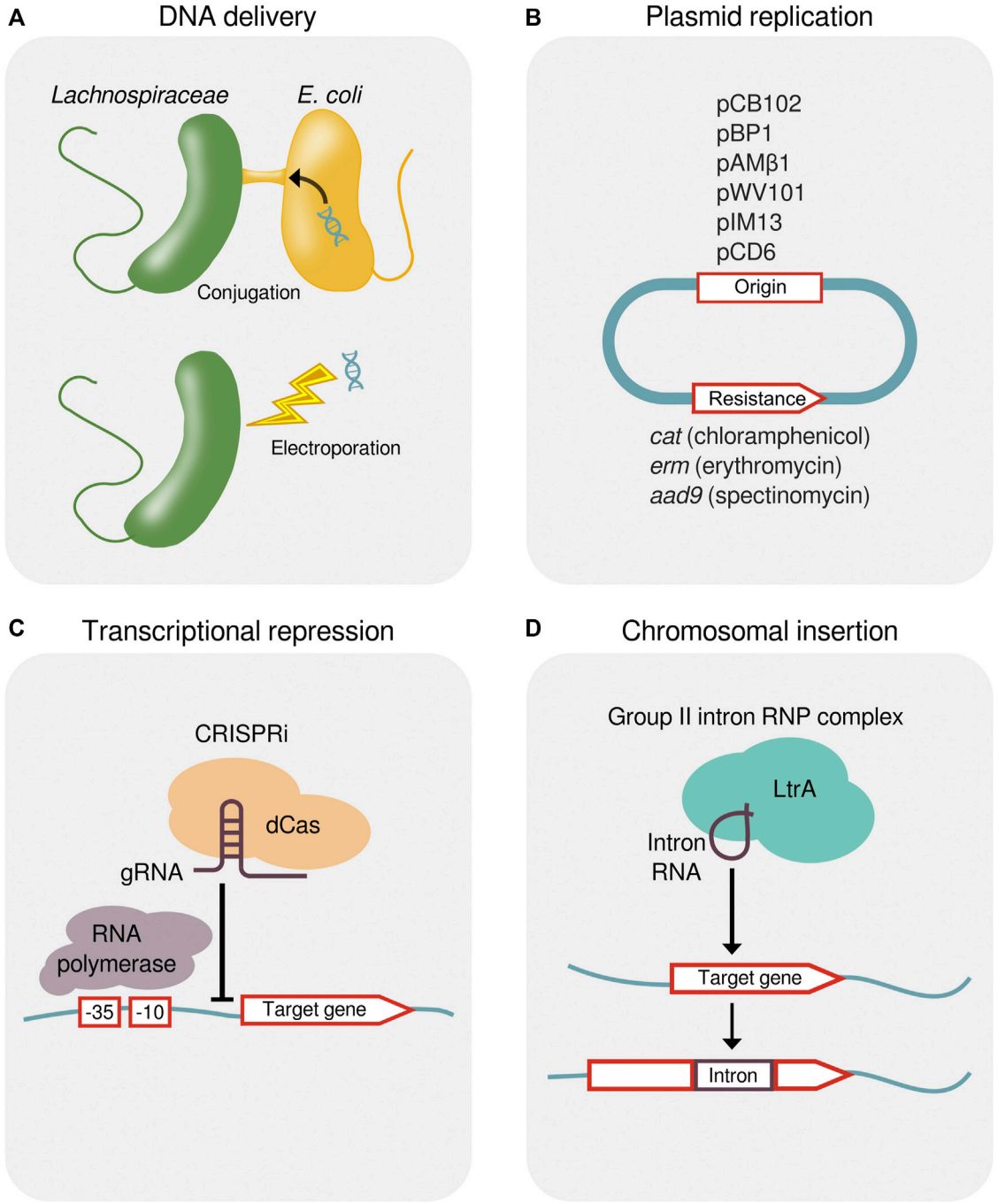


FIGURE 4
Methods for the genetic manipulation of *Lachnospiraceae*. **(A)** Delivery of foreign DNA into *Lachnospiraceae* by either conjugal transfer from *E. coli* or electroporation. **(B)** Plasmid origins that replicate in *Lachnospiraceae* and antibiotic resistance genes for plasmid selection. **(C)** Transcriptional repression by CRISPRi showing a dCas protein targeted by a gRNA to bind upstream of a gene, thereby blocking progression of RNA polymerase. **(D)** Targeted chromosomal insertion using a designed group II intron (targetron). Genomic insertion of the group II intron RNA containing a 13–16 bp target recognition sequence is facilitated by the endonuclease and reverse transcriptase activity of the LtrA protein. CRISPRi, CRISPR interference; dCas, dead CRISPR-associated protein; gRNA, guide RNA; RNP, ribonucleoprotein.

maintained using antibiotic resistance genes *catP* (chloramphenicol), *ermB* (erythromycin), or *aad9* (spectinomycin) (Sheridan et al., 2019; Jin et al., 2022; Rostain et al., 2022) (Figure 4B).

Shuttle vectors containing Gram-negative and Gram-positive origins of replication have permitted the heterologous expression of a β-(1,3–1,4)-glucanase in *Eubacterium rectale* and *R. inulinivorans* (Sheridan et al., 2019) and of a synthetic ethanol formation pathway in *L. phytofermentans* (Tolonen et al., 2015b). Plasmid-based expression of a NanoLuc reporter has identified a library of promoters of varying strength, and reporter gene expression can

be regulated by anhydrotetracycline using a promoter flanked tet repressor sites (Rostain et al., 2022). Clustered Regularly Interspaced Short Palindromic Repeats interference (CRISPRi) using dCas12a has been applied to repress the transcription of chromosomal genes for fermentation, thereby reducing production of butyrate in *Blautia luti* and *Enterocloster boltae* (Jin et al., 2022) and acetate in *L. phytofermentans* (Rostain et al., 2022) (Figure 4C).

Similar to other Clostridia, low rates of DNA transfer and homologous recombination in *Lachnospiraceae* have led to the use of other recombination systems to make targeted chromosomal changes. Designed group II intron called targetrons enabled gene inactivation by targeted chromosome insertion in various *Lachnospiraceae* with efficiencies ranging from 12.5%–100% (Tolonen et al., 2009; Tolonen et al., 2015a; Cerisy et al., 2019a; Jin et al., 2022) (Figure 4D). Multi-gene fragments can be excised and inserted by modifying targetrons to deliver *lox* sites into the genome that act as anchor points for Cre-mediated recombination, which has been applied to delete a 39 kb prophage in *L. phytofermentans* (Cerisy et al., 2019b).

Lachnospiraceae proteins that have been applied as molecular tools in other organisms are strong candidates to advance genome engineering in *Lachnospiraceae*. For example, LbCas12a, which was isolated from *Lachnospiraceae* bacterium ND 2006 (Tak et al., 2017), has been applied for genome editing in eukaryotes including fungi (Chen T. et al., 2023), plants (Kim et al., 2021), flies (Port et al., 2020), and human cells (Zhang et al., 2023). Another Cas12a variant, Lb2Cas12a, derived from *Lachnospiraceae* bacterium MA 2020, has been developed to have enhanced editing activity and broadened protospacer adjacent motif (PAM) recognition in human cells (Tran et al., 2021). A protein with anti-CRISPR function has been identified from *Lachnospiraceae* phage (Forsberg et al., 2019) and a cytosine deaminase from *R. intestinalis* has been fused to CRISPR and Transcription Activator-Like Effector (TALE) proteins to make targeted C-to-T transitions in the genomes of cultured cells and mouse embryos (Guo et al., 2023).

Bioeconomy applications

The capabilities of *Lachnospiraceae* to metabolize the polysaccharides, aromatics, and proteins that compose lignocellulose make them candidates to transform low-cost, sustainable, lignocellulosic feedstocks (i.e., forestry, agricultural, and municipal wastes) into value-added biochemicals (Table 1). For example, *L. phytofermentans*, a species with wide polysaccharide utilization capabilities, ferments corn stover to ethanol with efficiencies similar to commercial enzymes and xylose-fermenting yeast (Jin et al., 2011). Synthesis of longer chain acids from lignocellulose residues by *Ca.* Weimeria bifida (Scarborough et al., 2020) holds potential to develop 'drop in' fuels that are compatible with the current petroleum infrastructure. *F. natans* metabolizes the protein fraction of organic matter to produce isobutyrate and n-butyrate, which are used for artificial fibers, plastics and herbicides as well as isovalerate used for flavoring and perfumes (Agnihotri et al., 2022).

Lachnospiraceae have been identified as biocatalysts for conversion of biomass to hydrogen (Bu et al., 2021), and strains have been isolated that produce 2–3 moles hydrogen per mole of glucose equivalent (Harvey et al., 2008). In addition to its use as a fuel, hydrogen produced from lignocellulosic fermentation can be used as a reductant to fix CO and CO_2 by acetogens through gas fermentation (Figure 3), a process that has potential to convert industrial carbon emissions to useful biochemicals (Liew et al., 2016). *Lachnospiraceae* acetogens compete with methanogens in the cow rumen, representing a means to reduce bovine methane production (Yang et al., 2015).

Alternatively, hydrogen from lignocellulosic fermentation can be used to fix carbon dioxide by methanogens, and *Lachnospiraceae* have been shown to actuate the transformation of cellulose to methane by methanogenic consortia (Dai et al., 2016). Combination of *Lachnospiraceae* with other microorganisms to form synthetic consortia has been generally used to enhanced product formation rates (Zuroff et al., 2013; El Hage et al., 2019; Park et al., 2020), and engineered consortia including biomass-fermenting *Lachnospiraceae* and methanogens or acetogens have potential for the direct conversion of lignocellulose to methane and other value-added biochemicals.

Therapeutic applications: preclinical models

Preclinical *in vivo* models are providing evidence that supplementation with live *Lachnospiraceae* improves gut health and prevents pathogen colonization. For example, in a rat model of irritable bowel syndrome, addition of *Roseburia hominis* increased cecal butyrate content, reduced visceral hypersensitivity, and prevented the decreased expression of occludin (Zhang et al., 2019). Administration of *Blautia producta* directly inhibited growth of vancomycin-resistant *Enterococcus* (Caballero et al., 2017) and a murine *Lachnospiraceae* inhibits *C. difficile* colonization in mice (Reeves et al., 2012).

Lachnospiraceae can alleviate inflammatory and allergic diseases by modulating the immune system through production of antigens presented by innate immune cells and immunomodulatory metabolites. *R. intestinalis* promoted differentiation of regulatory T cells, activation of type 3 innate lymphoid cells, and suppression of inflammation through TLR5 (Shen et al., 2022). *E. rectale* supplementation regulated dendritic cell activation by reducing the frequency of $CD83^+$ cells and improved symptoms in a mouse model of Behçet's disease, a systemic inflammatory condition (Islam et al., 2021). Inoculation of germ-free mice with a 17 strain consortium including 8 *Lachnospiraceae* increased anti-inflammatory, regulatory T-cells ($CD4^+$, FoxP3+) in the colonic lamina propria of multiple mouse lines and alleviated colitis in experimental models (Atarashi et al., 2013). In addition, administration of this 17 strain consortium to mice in an ovalbumin (OVA)-induced allergic diarrhea mouse model reduced diarrhea and OVA-specific serum IgE levels (Atarashi et al., 2013). Additional evidence that *Lachnospiraceae* can mitigate food allergies was demonstrated by colonization of germ-free mice with *Anaerostipes caccae*, which protected against an anaphylactic response upon challenge with β-Lactoglobulin (BLG), a cow's milk allergen, and reduced BLG-specific IgE levels (Feehley et al., 2019).

Lachnospiraceae have provided benefits in preclinical models of metabolic syndrome and diabetes. Administration of *Blautia wexlerae* to male C57BL/6 mice on a high fat diet reduced body weight and multiple diabetes indicators, which was linked to *B. wexlerae* metabolites such as S-adenosylmethionine, acetylcholine, and L-ornithine conferring anti-adipogenesis and anti-inflammatory

TABLE 2 Clinical studies of Lachnospiraceae as live biotherapeutic products.

Principal Investigator; Sponsor	Trial	Intervention	Patient population	Outcome	Trial number, reference
Dr. Elaine Petrof, Queen's University	Open label trial	Synthetic community	rCDI	Both subjects clinically cured at 6 months	NCT01372943 (Petrof et al., 2013)
Dr. Lisa von Moltke, Seres Therapeutics	Phase 3, double-blind, placebo-controlled	SER-109	rCDI	Reduced rCDI at week 8 ($p < 0.001$)	NCT03183128 (Feuerstadt et al., 2022)
Dr. Michele Trucksis, Seres Therapeutics	Phase 2 double-blind, placebo-controlled	SER-109	rCDI	Reduced rCDI (44.1% vs. 53.3% with placebo), but not significant. Engraftment associated with non-recurrence ($p < 0.05$) and increased secondary bile acid concentrations ($p < 0.0001$)	NCT02437487 (McGovern et al., 2022)
Seres Therapeutics	Phase 1b Safety Study	SER-287	Mild to moderate UC	Increased clinical remission at week 8 ($p = 0.024$)	NCT02618187 (Henn et al., 2021)
Dr Eamonn Quigley, Houston Methodist; 4DPharma	Phase 2 double-blind, placebo-controlled	MRx1234 (Blautix)	IBS	Improved bowel habits ($p = 0.007$) and trend to increased overall response ($p = 0.06$)	NCT03721107 (Quigley et al., 2023)
Dr. Darrell Pardi, Mayo Clinic; Vedanta Biosciences	Phase 2 double-blind, placebo-controlled	VE303	rCDI	Reduced rCDI at week 8 ($p = 0.006$)	NCT03788434 (Louie et al., 2023)
Vedanta Biosciences	Phase 1 safety study	VE303	Healthy adults	Well tolerated, engraftment	NCT04236778 (Dsouza et al., 2022)
Dr. Patricia Bloom, University of Michigan; Vedanta Biosciences	Double-blind, placebo-controlled	VE303	Hepatic encephalopathy	Trial in progress	NCT04899115
Vedanta Biosciences	Phase 2, double-blind, placebo-controlled	VE202	Mild to moderate UC	Trial in progress	NCT05370885
Vedanta Biosciences	Phase 1 safety study	VE202	Healthy adults	Well tolerated, strain engraftment	Silber et al. (2022)
Dr. Erik Stroes, University of Amsterdam; Caelus Pharmaceuticals	Phase 2, double-blind placebo controlled	*Anaerobutyricum soehngenii* L2-7	Metabolic Syndrome	Elevated plasma GLP-1 ($p = 0.02$), increased fecal butyrate ($p = 0.06$), reduced glucose variability ($p = 0.05$)	NTR-NL6630, (Koopen et al., 2022)
Dr. James Ryan Atlantia Food Clinical Trials; Caelus Pharmaceuticals	Phase 1/2 dose finding study	*Anaerobutyricum soehngenii* L2-7	Metabolic Syndrome	Abundance of *A. soehngenii* correlated with peripheral insulin sensitivity ($p = 0.05$)	NCT04529473 (Gilijamse et al., 2020)
Oluf B Pedersen, University of Copenhagen	Placebo-controlled crossover study	*Ruminococcus torques* strain ATCC 27756	Overweight adults	Trial in progress	NCT05448274

The synthetic community in the Queen's University study consisted of 33 strains, including 10 *Lachnospiraceae*. SER-109 is purified fecal spores, 36% of genera are *Lachnospiraceae*. SER-287 is purified fecal spores, 44% of genera are *Lachnospiraceae*. MRx1234 (Blautix) is lyophilised *Blautia hydrogenotrophica*. VE303 is a consortium of 8 strains including 5 *Lachnospiraceae*. VE202 is a consortium of 16 *Clostridia* XIVa, IV, XVIII. rCDI, recurrent *Clostridioides difficile* infection; UC, ulcerative colitis; IBS, irritable bowel syndrome.

properties to adipocytes (Hosomi et al., 2022). *B. producta* suppressed lipid accumulation in HepG2 cells and gavage of *B. producta* alleviated hyperlipidemia in mice through production of 12-methylmyristic acid (Wu et al., 2021). Treatment of obese, diabetic *db/db* mice with *A. soehngenii* reduced plasma glucose, epididymal fat, and liver triglycerides and improved peripheral insulin sensitivity (Udayappan et al., 2016).

Mouse studies have also shown beneficial roles of *Lachnospiraceae* in cancer treatment. C57BL/6 mice bearing B16-F10 melanoma or CT-26 colorectal tumors showed reduced tumor growth when administered *Blautia massiliensis* (Goodman et al., 2019). Anti-PD1 mediated tumor control and survivorship of C57BL/6 mice bearing B16-F10 melanoma tumors was improved by administration of *E. rectale*, which was proposed to result from it consuming L-serine, leading to NK cell activation and tumor infiltration (Liu et al., 2023). *E. rectale* treatment also reduced the incidence of lymphoma and reduced TNF levels in sensitized Eμ-Myc mice (Lu et al., 2022). Survivorship and clinical scores of C57BL/6 mice following full body irradiation were increased by prior inoculation with a mix of 23 *Lachnospiraceae*, which resulted from increased hematopoiesis and reduced intestinal epithelial injury (Guo et al., 2020).

Therapeutic applications: clinical trials

Lachnospiraceae are being tested as LBPs in clinical studies for a number of diseases (Table 2), the most advanced of which is the treatment of recurrent *C. difficile* infections (rCDI). In an exploratory study of two rCDI patients at Queen's University, treatment with a

synthetic community of 33 strains, including 10 *Lachnospiraceae*, resulted in full remission in both patients (Petrof et al., 2013). Subsequently, microbial consortia containing *Lachnospiraceae* have been developed to combat rCDI. SER-109, consisting of spores purified from fecal samples, is composed of 77 genera of which 36% are *Lachnospiraceae* (Feuerstadt et al., 2022). In a phase 3 study, SER-109 reduced rCDI at week 8 (Feuerstadt et al., 2022). In 2023, the US Food and Drug Administration approved SER-109 under the commercial name Vowst™ to treat recurrent *C. difficile* infection (Commission of the US FDA, 2023). VE303, a consortium of 8 strains including 5 *Lachnospiraceae*, engrafted into the microbiomes of healthy volunteers to boost production of SCFA and secondary bile acids without any serious adverse events (Dsouza et al., 2022). Phase 2 results in patients supported that rCDI at 8 weeks was reduced by VE303 (Louie et al., 2023). VE303 is being tested in a phase 3 study to treat rCDI starting in 2023, and as an experimental treatment for hepatic encephalopathy (Vedanta Biosciences Inc, 2023).

LBPs that include *Lachnospiraceae* are also being developed for metabolic syndrome and inflammatory bowel disease (Table 2). For example, *A. soehngenii* L2-7 is being tested as a probiotic to improve insulin sensitivity in metabolic syndrome patients. A phase 1/2 study of 24 metabolic syndrome patients correlated *A. soehngenii* engraftment with improved peripheral insulin sensitivity (Gilijamse et al., 2020). A phase 2 study of 12 metabolic syndrome patients showed duodenal infusion of *A. soehngenii* reduced glucose variability and elevated GLP-1, secondary bile acids in plasma, duodenal REGB1 expression, and fecal SCFAs (Koopen et al., 2022). As *B. hydrogenotrophica* consumes intestinal gas (H_2, CO_2), it can treat irritable bowel syndrome (IBS) by reducing intestinal bloating. A phase 2 study of MRx1234 (Blautix), consisting of lyophilized *B. hydrogenotrophica*, improved bowel habits in IBS patients (Quigley et al., 2023).

Looking forward: challenges and opportunities

Lachnospiraceae have numerous potential applications due to their native fermentation of low-cost substrates and importance for intestinal health, but key challenges remain to harness them as industrial biocatalysts and LBPs. In particular, work is needed to isolate and characterize additional *Lachnospiraceae* strains, uncover the genetic basis of their physiological traits, and, ultimately, apply these learning to engineer optimized strains. Recently established *Lachnospiraceae* collections demonstrated methods with which many species can be cultivated (Seshadri et al., 2018; Sorbara et al., 2020; Abdugheni et al., 2022). Additional cultivation efforts could build *Lachnospiraceae* collections from environments such as soil and human patient populations. Further, *Lachnospiraceae* culture collections revealed that isolates from the same species can differ in traits such as substrate utilization and production of antimicrobials (Sorbara et al., 2020). Intraspecific comparative genomics of isolates sharing similar genomes, but with specific physiological differences, is an opportunity to define genotype-phenotype relationships in *Lachnospiraceae*.

Development of microorganisms for biotechnology often requires rewiring native metabolism to improve product yields, highlighting the importance of future research on genetic manipulation of *Lachnospiraceae*. As described above, genetic tools have been ported to *Lachnospiraceae* from other well-studied mesophiles including *B. subtilis* and *L. lactis*. Species of *Lachnospiraceae* have been established as genetically tractable hosts with methods for genetic transformation, plasmid replication and selection, transcriptional repression, and chromosomal insertions (Figure 4); a recent study demonstrates that these methods can be generally applied to many species (Jin et al., 2022). Existing molecular tools derived from *Lachnospiraceae* are promising candidates to advance genome engineering in these bacteria, including Cas proteins with broadened PAM recognition (Tran et al., 2021), anti-CRISPR proteins (Forsberg et al., 2019), and cytosine deaminases (Guo et al., 2023). Moreover, adaptive laboratory evolution can be used to generate *Lachnospiraceae* strains with complex, multigenic traits such as inhibitor tolerance that are intractable by rational genome engineering (Cerisy et al., 2017).

Engineering of *Lachnospiraceae* for industrial production of biochemicals from low-cost feedstocks will need to focus on substrate assimilation and product formation. The rate of lignocellulose solubilization remains a primary obstacle to its utilization as a feedstock (Preethi et al., 2021), which could be addressed by engineering strains with modified expression of CAZymes and associated ABC transporters to accelerate solubilization and uptake of target lignocellulosic substrates. As *Lachnospiraceae* typically produce a mixture of fermentation products, redistribution of metabolic flux by repression or inactivation of genes for undesired products is a valuable approach to streamline and increase fermentation yields.

Genetic manipulation of *Lachnospiraceae* will also be important to define the mechanisms by which they promote intestinal health. Significant advances have been made to elucidate innate and adaptive immune responses modulated by *Lachnospiraceae* (Islam et al., 2021, 1) (Atarashi et al., 2013; Feehley et al., 2019; Liu et al., 2023). It is generally believed that *Lachnospiraceae* modulate host immunity producing SCFAs. Recently, it was shown that differential recognition of *Lachnospiraceae* flagellins by TLR5 contributes to immune tolerance (Clasen et al., 2023). Genetic studies with *Lachnospiraceae* mutants will be useful to further define the mechanisms that underlie host interactions and build strains with customized immunomodulatory properties.

Industrial-scale cultivation of *Lachnospiraceae* for bioproduction and LBPs will necessitate a greater understanding of the genetics and ecology of the phage that infect them. Phage infection is recognized as a persistent threat in industrial microbiology where scale-up of bacterial populations in large bioreactors favors phage outbreaks. For example, phage infection is the main cause of fermentation failures in the dairy industry, leading to intense study of the phage of lactic acid bacteria (Fernández et al., 2017). Prophages are common in the genomes of gut *Lachnospiraceae* (Dikareva et al., 2023). Although hypervirulent phage can drive temporal variation of *R. intestinalis* abundances in the intestine, no phage infecting *Lachnospiraceae* have been deposited in public databases (Cornuault et al., 2020). Thus, isolation and characterization of *Lachnospiraceae* phage are important subjects for future research.

Over the past decade, our understanding of *Lachnospiraceae* molecular biology and physiology has greatly increased. In 2023, Vowst™ became the first FDA-approved oral medication containing live *Lachnospiraceae* (Commission of the US FDA, 2023) and clinical studies are evaluating additional therapeutic

benefits of *Lachnospiraceae* (Table 2). Development of optimized strains by genome engineering will enable us to realize the potential of *Lachnospiraceae* in biotechnology. In the next few years, we expect to see further development of *Lachnospiraceae* to produce useful chemicals and to manage disease through targeted changes to the composition and metabolites produced by the gut microbiome.

Author contributions

TZ: Writing–original draft, Writing–review and editing. SM: Writing–original draft, Writing–review and editing. AT: Conceptualization, Writing–original draft, Writing–review and editing.

Funding

The authors declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by Genoscope-CEA and the Évry Genopole under the "Action Thématique Incitative Genopole" Grant (funding ATIGE 2021 No. 2698).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abdugheni, R., Wang, W.-Z., Wang, Y.-J., Du, M.-X., Liu, F.-L., Zhou, N., et al. (2022). Metabolite profiling of human-originated Lachnospiraceae at the strain level. *iMeta* 1, e58. doi:10.1002/imt2.58

Agnihotri, S., Yin, D.-M., Mahboubi, A., Sapmaz, T., Varjani, S., Qiao, W., et al. (2022). A glimpse of the world of volatile fatty acids production and application: a review. *Bioengineered* 13, 1249–1275. doi:10.1080/21655979.2021.1996044

Antezack, A., Boxberger, M., La Scola, B., and Monnet-Corti, V. (2021). Isolation and description of catonella massiliensis sp. nov., a novel catonella species, isolated from a stable periodontitis subject. *Pathogens* 10, 367. doi:10.3390/pathogens10030367

Arora, T., Sharma, R., and Frost, G. (2011). Propionate. Anti-obesity and satiety enhancing factor? *Appetite* 56, 511–515. doi:10.1016/j.appet.2011.01.016

Atarashi, K., Tanoue, T., Oshima, K., Suda, W., Nagano, Y., Nishikawa, H., et al. (2013). Treg induction by a rationally selected mixture of Clostridia strains from the human microbiota. *Nature* 500, 232–236. doi:10.1038/nature12331

Bang, S.-J., Kim, G., Lim, M. Y., Song, E.-J., Jung, D.-H., Kum, J.-S., et al. (2018). The influence of *in vitro* pectin fermentation on the human fecal microbiome. *Amb. Express* 8, 98. doi:10.1186/s13568-018-0629-9

Barbosa, T. M., Scott, K. P., and Flint, H. J. (1999). Evidence for recent intergeneric transfer of a new tetracycline resistance gene, tet(W), isolated from Butyrivibrio fibrisolvens, and the occurrence of tet(O) in ruminal bacteria. *Environ. Microbiol.* 1, 53–64. doi:10.1046/j.1462-2920.1999.00004.x

Beard, C. E., Hefford, M. A., Forster, R. J., Sontakke, S., Teather, R. M., and Gregg, K. (1995). A stable and efficient transformation system for Butyrivibrio fibrisolvens OB156. *Curr. Microbiol.* 30, 105–109. doi:10.1007/BF00294191

Berggren, A. M., Nyman, E. M., Lundquist, I., and Björck, I. M. (1996). Influence of orally and rectally administered propionate on cholesterol and glucose metabolism in obese rats. *Br. J. Nutr.* 76, 287–294. doi:10.1079/bjn19960032

Bernalier, A., Willems, A., Leclerc, M., Rochet, V., and Collins, M. D. (1996). Ruminococcus hydrogenotrophicus sp. nov., a new H2/CO2-utilizing acetogenic bacterium isolated from human feces. *Arch. Microbiol.* 166, 176–183. doi:10.1007/s002030050373

Boutard, M., Cerisy, T., Nogue, P.-Y., Alberti, A., Weissenbach, J., Salanoubat, M., et al. (2014). Functional diversity of carbohydrate-active enzymes enabling a bacterium to ferment plant biomass. *PLoS Genet.* 10, e1004773. doi:10.1371/journal.pgen.1004773

Boutard, M., Ettwiller, L., Cerisy, T., Alberti, A., Labadie, K., Salanoubat, M., et al. (2016). Global repositioning of transcription start sites in a plant-fermenting bacterium. *Nat. Commun.* 7, 13783. doi:10.1038/ncomms13783

Braune, A., and Blaut, M. (2016). Bacterial species involved in the conversion of dietary flavonoids in the human gut. *Gut Microbes* 7, 216–234. doi:10.1080/19490976.2016.1158395

Bu, J., Wei, H.-L., Wang, Y.-T., Cheng, J.-R., and Zhu, M.-J. (2021). Biochar boosts dark fermentative H2 production from sugarcane bagasse by selective enrichment/colonization of functional bacteria and enhancing extracellular electron transfer. *Water Res.* 202, 117440. doi:10.1016/j.watres.2021.117440

Bui, T. P. N., Mannerås-Holm, L., Puschmann, R., Wu, H., Troise, A. D., Nijsse, B., et al. (2021). Conversion of dietary inositol into propionate and acetate by commensal Anaerostipes associates with host health. *Nat. Commun.* 12, 4798. doi:10.1038/s41467-021-25081-w

Caballero, S., Kim, S., Carter, R. A., Leiner, I. M., Sušac, B., Miller, L., et al. (2017). Cooperating commensals restore colonization resistance to vancomycin-resistant Enterococcus faecium. *Cell Host Microbe* 21, 592–602.e4. doi:10.1016/j.chom.2017.04.002

Cai, S., and Dong, X. (2010). Cellulosilyticum ruminicola gen. nov., sp. nov., isolated from the rumen of yak, and reclassification of Clostridium lentocellum as Cellulosilyticum lentocellum comb. nov. *Int. J. Syst. Evol. Microbiol.* 60, 845–849. doi:10.1099/ijs.0.014712-0

Calusinska, M., Happe, T., Joris, B., and Wilmotte, A. (2010). The surprising diversity of clostridial hydrogenases: a comparative genomic perspective. *Microbiol. Read.* 156, 1575–1588. doi:10.1099/mic.0.032771-0

Cerisy, T., Iglesias, A., Rostain, W., Boutard, M., Pelle, C., Perret, A., et al. (2019a). ABC transporters required for hexose uptake by Clostridium phytofermentans. *J. Bacteriol.* 201, e00241-19. doi:10.1128/JB.00241-19

Cerisy, T., Rostain, W., Chhun, A., Boutard, M., Salanoubat, M., and Tolonen, A. C. (2019b). A targetron-recombinase system for large-scale genome engineering of clostridia. *mSphere* 4, e00710-19. doi:10.1128/mSphere.00710-19

Cerisy, T., Souterre, T., Torres-Romero, I., Boutard, M., Dubois, I., Patrouix, J., et al. (2017). Evolution of a biomass-fermenting bacterium to resist lignin phenolics. *Appl. Environ. Microbiol.* 83, e00289-17. doi:10.1128/AEM.00289-17

Chambers, E. S., Viardot, A., Psichas, A., Morrison, D. J., Murphy, K. G., Zac-Varghese, S. E. K., et al. (2015). Effects of targeted delivery of propionate to the human colon on appetite regulation, body weight maintenance and adiposity in overweight adults. *Gut* 64, 1744–1754. doi:10.1136/gutjnl-2014-307913

Chen, I.-M. A., Chu, K., Palaniappan, K., Ratner, A., Huang, J., Huntemann, M., et al. (2023a). The IMG/M data management and analysis system v.7: content updates and new features. *Nucleic Acids Res.* 51, D723–D732. doi:10.1093/nar/gkac976

Chen, T., Chen, Z., Zhang, H., Li, Y., Yao, L., Zeng, B., et al. (2023b). Development of a CRISPR/Cpf1 system for multiplex gene editing in Aspergillus oryzae. *Folia Microbiol. (Praha).* doi:10.1007/s12223-023-01081-9

Clasen, S. J., Bell, M. E. W., Borbón, A., Lee, D.-H., Henseler, Z. M., de la Cuesta-Zuluaga, J., et al. (2023). Silent recognition of flagellins from human gut commensal bacteria by Toll-like receptor 5. *Sci. Immunol.* 8, eabq7001. doi:10.1126/sciimmunol.abq7001

Commission of the US Food and Drug Administration (2023). FDA approves first orally administered fecal microbiota product for the prevention of recurrence of Clostridioides difficile infection. https://www.fda.gov/news-events/press-announcements/fda-approves-first-orally-administered-fecal-microbiota-product-prevention-recurrence-clostridioides.

Cornick, N. A., Jensen, N. S., Stahl, D. A., Hartman, P. A., and Allison, M. J. (1994). Lachnospira pectinoschiza sp. nov., an anaerobic pectinophile from the pig intestine. *Int. J. Syst. Bacteriol.* 44, 87–93. doi:10.1099/00207713-44-1-87

Cornuault, J. K., Moncaut, E., Loux, V., Mathieu, A., Sokol, H., Petit, M.-A., et al. (2020). The enemy from within: a prophage of Roseburia intestinalis systematically turns lytic in the mouse gut, driving bacterial adaptation by CRISPR spacer acquisition. *ISME J.* 14, 771–787. doi:10.1038/s41396-019-0566-x

Cuív, P. Ó., Smith, W. J., Pottenger, S., Burman, S., Shanahan, E. R., and Morrison, M. (2015). Isolation of genetically tractable most-wanted bacteria by metaparental mating. *Sci. Rep.* 5, 13282. doi:10.1038/srep13282

Dai, Y., Yan, Z., Jia, L., Zhang, S., Gao, L., Wei, X., et al. (2016). The composition, localization and function of low-temperature-adapted microbial communities involved in methanogenic degradations of cellulose and chitin from Qinghai-Tibetan Plateau wetland soils. *J. Appl. Microbiol.* 121, 163–176. doi:10.1111/jam.13164

Dikareva, E., Matharu, D., Lahtinen, E., Kolho, K.-L., De Vos, W. M., Salonen, A., et al. (2023). An extended catalog of integrated prophages in the infant and adult fecal microbiome shows high prevalence of lysogeny. *Front. Microbiol.* 14, 1254535. doi:10.3389/fmicb.2023.1254535

Dostal, A., Lacroix, C., Bircher, L., Pham, V. T., Follador, R., Zimmermann, M. B., et al. (2015). Iron modulates butyrate production by a child gut microbiota *in vitro*. *mBio* 6, e01453–e01415. doi:10.1128/mBio.01453-15

Dsouza, M., Menon, R., Crossette, E., Bhattarai, S. K., Schneider, J., Kim, Y.-G., et al. (2022). Colonization of the live biotherapeutic product VE303 and modulation of the microbiota and metabolites in healthy volunteers. *Cell Host Microbe* 30, 583–598.e8. doi:10.1016/j.chom.2022.03.016

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinforma.* 5, 113. doi:10.1186/1471-2105-5-113

El Hage, R., Hernandez-Sanabria, E., Calatayud Arroyo, M., Props, R., and Van de Wiele, T. (2019). Propionate-producing consortium restores antibiotic-induced dysbiosis in a dynamic *in vitro* model of the human intestinal microbial ecosystem. *Front. Microbiol.* 10, 1206. doi:10.3389/fmicb.2019.01206

Feehley, T., Plunkett, C. H., Bao, R., Choi Hong, S. M., Culleen, E., Belda-Ferre, P., et al. (2019). Healthy infants harbor intestinal bacteria that protect against food allergy. *Nat. Med.* 25, 448–453. doi:10.1038/s41591-018-0324-z

Fernández, L., Escobedo, S., Gutiérrez, D., Portilla, S., Martínez, B., García, P., et al. (2017). Bacteriophages in the dairy environment: from enemies to allies. *Antibiotics* 6, 27. doi:10.3390/antibiotics6040027

Ferreira, F., Kenne, L., Cotta, M. A., and Stack, R. J. (1997). Structural studies of the extracellular polysaccharide from Butyrivibrio fibrisolvens strain CF3. *Carbohydr. Res.* 301, 193–203. doi:10.1016/s0008-6215(97)00097-9

Feuerstadt, P., Louie, T. J., Lashner, B., Wang, E. E. L., Diao, L., Bryant, J. A., et al. (2022). SER-109, an oral microbiome therapy for recurrent Clostridioides difficile infection. *N. Engl. J. Med.* 386, 220–229. doi:10.1056/NEJMoa2106516

Flint, H. J., Scott, K. P., Duncan, S. H., Louis, P., and Forano, E. (2012). Microbial degradation of complex carbohydrates in the gut. *Gut Microbes* 3, 289–306. doi:10.4161/gmic.19897

Forsberg, K. J., Bhatt, I. V., Schmidtke, D. T., Javanmardi, K., Dillard, K. E., Stoddard, B. L., et al. (2019). Functional metagenomics-guided discovery of potent Cas9 inhibitors in the human microbiome. *eLife* 8, e46540. doi:10.7554/eLife.46540

Furusawa, Y., Obata, Y., Fukuda, S., Endo, T. A., Nakato, G., Takahashi, D., et al. (2013). Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature* 504, 446–450. doi:10.1038/nature12721

Garzetti, D., Brugiroux, S., Bunk, B., Pukall, R., McCoy, K. D., Macpherson, A. J., et al. (2017). High-quality whole-genome sequences of the oligo-mouse-microbiota bacterial community. *Genome Announc.* 5, 007588-17. doi:10.1128/genomeA.00758-17

Gilijamse, P. W., Hartstra, A. V., Levin, E., Wortelboer, K., Serlie, M. J., Ackermans, M. T., et al. (2020). Treatment with Anaerobutyricum soehngenii: a pilot study of safety and dose-response effects on glucose metabolism in human subjects with metabolic syndrome. *NPJ Biofilms Microbiomes* 6, 16. doi:10.1038/s41522-020-0127-0

Goodman, B., Sandy, P., Papkoff, J., Gardner, H., and Ponichtera, H. (2019). *Treating cancer using a blautia strain*. US Patent application 2019/0240267 A1.

Gosalbes, M. J., Durbán, A., Pignatelli, M., Abellan, J. J., Jiménez-Hernández, N., Pérez-Cobas, A. E., et al. (2011). Metatranscriptomic approach to analyze the functional human gut microbiota. *PLOS ONE* 6, e17447. doi:10.1371/journal.pone.0017447

Greening, R. C., and Leedle, J. A. Z. (1989). Enrichment and isolation of Acetitomaculum ruminis, gen. nov., sp. nov.: acetogenic bacteria from the bovine rumen. *Arch. Microbiol.* 151, 399–406. doi:10.1007/BF00416597

Groher, A., and Weuster-Botz, D. (2016). General medium for the autotrophic cultivation of acetogens. *Bioprocess Biosyst. Eng.* 39, 1645–1650. doi:10.1007/s00449-016-1634-5

Guo, C.-J., Chang, F.-Y., Wyche, T. P., Backus, K. M., Acker, T. M., Funabashi, M., et al. (2017). Discovery of reactive microbiota-derived metabolites that inhibit host proteases. *Cell* 168, 517–526.e18. doi:10.1016/j.cell.2016.12.021

Guo, H., Chou, W.-C., Lai, Y., Liang, K., Tam, J. W., Brickey, W. J., et al. (2020). Multi-omics analyses of radiation survivors identify radioprotective microbes and metabolites. *Science* 370, eaay9097. doi:10.1126/science.aay9097

Guo, J., Yu, W., Li, M., Chen, H., Liu, J., Xue, X., et al. (2023). A DddA ortholog-based and transactivator-assisted nuclear and mitochondrial cytosine base editors with expanded target compatibility. *Mol. Cell* 83, 1710–1724.e7. doi:10.1016/j.molcel.2023.04.012

Harvey, S., Chambers, A., and Zhang, P. (2008). Overproduction of hydrogen from an anaerobic bacterium. https://apps.dtic.mil/sti/citations/ADA505872.

Hassan, N. W., Saudi, M. N., Abdel-Ghany, Y. S., Ismail, A., Elzahhar, P. A., Sriram, D., et al. (2020). Novel pyrazine based anti-tubercular agents: design, synthesis, biological evaluation and *in silico* studies. *Bioorg Chem.* 96, 103610. doi:10.1016/j.bioorg.2020.103610

Hatziioanou, D., Gherghisan-Filip, C., Saalbach, G., Horn, N., Wegmann, U., Duncan, S. H., et al. (2017). Discovery of a novel lantibiotic nisin O from Blautia obeum A2-162, isolated from the human gastrointestinal tract. *Microbiol. Read.* 163, 1292–1305. doi:10.1099/mic.0.000515

Heap, J. T., Pennington, O. J., Cartman, S. T., and Minton, N. P. (2009). A modular system for Clostridium shuttle plasmids. *J. Microbiol. Methods* 78, 79–85. doi:10.1016/j.mimet.2009.05.004

Hefford, M. A., Kobayashi, Y., Allard, S. E., Forster, R. J., and Teather, R. M. (1997). Sequence analysis and characterization of pOM1, a small cryptic plasmid from Butyrivibrio fibrisolvens, and its use in construction of a new family of cloning vectors for Butyrivibrios. *Appl. Environ. Microbiol.* 63, 1701–1711. doi:10.1128/aem.63.5.1701-1711.1997

Hengstmann, U., Chin, K. J., Janssen, P. H., and Liesack, W. (1999). Comparative phylogenetic assignment of environmental sequences of genes encoding 16S rRNA and numerically abundant culturable bacteria from an anoxic rice paddy soil. *Appl. Environ. Microbiol.* 65, 5050–5058. doi:10.1128/AEM.65.11.5050-5058.1999

Henn, M. R., O'Brien, E. J., Diao, L., Feagan, B. G., Sandborn, W. J., Huttenhower, C., et al. (2021). A phase 1b safety study of SER-287, a spore-based microbiome therapeutic, for active mild to moderate ulcerative colitis. *Gastroenterology* 160, 115–127.e30. doi:10.1053/j.gastro.2020.07.048

Hillman, E. T., Kozik, A. J., Hooker, C. A., Burnett, J. L., Heo, Y., Kiesel, V. A., et al. (2020). Comparative genomics of the genus Roseburia reveals divergent biosynthetic pathways that may influence colonic competition among species. *Microb. Genom* 6, mgen000399. doi:10.1099/mgen.0.000399

Hosomi, K., Saito, M., Park, J., Murakami, H., Shibata, N., Ando, M., et al. (2022). Oral administration of Blautia wexlerae ameliorates obesity and type 2 diabetes via metabolic remodeling of the gut microbiota. *Nat. Commun.* 13, 4477. doi:10.1038/s41467-022-32015-7

Huang, X., Liu, L., Zhao, J., Zhang, J., and Cai, Z. (2019). The families Ruminococcaceae, Lachnospiraceae, and Clostridiaceae are the dominant bacterial groups during reductive soil disinfestation with incorporated plant residues. *Appl. Soil Ecol.* 135, 65–72. doi:10.1016/j.apsoil.2018.11.011

Islam, S. M. S., Ryu, H.-M., Sayeed, H. M., Byun, H.-O., Jung, J.-Y., Kim, H.-A., et al. (2021). Eubacterium rectale attenuates HSV-1 induced systemic inflammation in mice by inhibiting CD83. *Front. Immunol.* 12, 712312. doi:10.3389/fimmu.2021.712312

Jin, M., Balan, V., Gunawan, C., and Dale, B. E. (2011). Consolidated bioprocessing (CBP) performance of Clostridium phytofermentans on AFEX-treated corn stover for ethanol production. *Biotechnol. Bioeng.* 108, 1290–1297. doi:10.1002/bit.23059

Jin, W.-B., Li, T.-T., Huo, D., Qu, S., Li, X. V., Arifuzzaman, M., et al. (2022). Genetic manipulation of gut microbes enables single-gene interrogation in a complex microbiome. *Cell* 185, 547–562.e22. doi:10.1016/j.cell.2021.12.035

Kim, D., Hager, M., Brant, E., and Budak, H. (2021). Efficient genome editing in wheat using Cas9 and Cpf1 (AsCpf1 and LbCpf1) nucleases. *Funct. Integr. Genomics* 21, 355–366. doi:10.1007/s10142-021-00782-z

Koeck, D. E., Ludwig, W., Wanner, G., Zverlov, V. V., Liebl, W., and Schwarz, W. H. (2015). Herbinix hemicellulosilytica gen. nov., sp. nov., a thermophilic cellulose-degrading bacterium isolated from a thermophilic biogas reactor. *Int. J. Syst. Evol. Microbiol.* 65, 2365–2371. doi:10.1099/ijs.0.000264

Koeck, D. E., Maus, I., Wibberg, D., Winkler, A., Zverlov, V. V., Liebl, W., et al. (2016). Complete genome sequence of Herbinix luporum SD1D, a new cellulose-degrading bacterium isolated from a thermophilic biogas reactor. *Genome Announc.* 4, 006877-16. doi:10.1128/genomeA.00687-16

Kong, F., Hua, Y., Zeng, B., Ning, R., Li, Y., and Zhao, J. (2016). Gut microbiota signatures of longevity. *Curr. Biol.* 26, R832–R833. doi:10.1016/j.cub.2016.08.015

Koopen, A., Witjes, J., Wortelboer, K., Majait, S., Prodan, A., Levin, E., et al. (2022). Duodenal Anaerobutyricum soehngenii infusion stimulates GLP-1 production, ameliorates glycaemic control and beneficially shapes the duodenal transcriptome in metabolic syndrome subjects: a randomised double-blind placebo-controlled cross-over study. *Gut* 71, 1577–1587. doi:10.1136/gutjnl-2020-323297

Krumholz, L. R., and Bryant, M. P. (1986). Syntrophococcus sucromutans sp. nov. gen. nov. uses carbohydrates as electron donors and formate, methoxymonobenzenoids or Methanobrevibacter as electron acceptor systems. *Arch. Microbiol.* 143, 313–318. doi:10.1007/BF00412795

Lapczynski, A., Bhatia, S. P., Letizia, C. S., and Api, A. M. (2008). Fragrance material review on farnesol. *Food Chem. Toxicol.* 46 (Suppl. 11), S149–S156. doi:10.1016/j.fct.2008.06.046

La Rosa, S. L., Leth, M. L., Michalak, L., Hansen, M. E., Pudlo, N. A., Glowacki, R., et al. (2019). The human gut Firmicute Roseburia intestinalis is a primary degrader of dietary β-mannans. *Nat. Commun.* 10, 905. doi:10.1038/s41467-019-08812-y

Leth, M. L., Ejby, M., Workman, C., Ewald, D. A., Pedersen, S. S., Sternberg, C., et al. (2018). Differential bacterial capture and transport preferences facilitate co-growth on dietary xylan in the human gut. *Nat. Microbiol.* 3, 570–580. doi:10.1038/s41564-018-0132-8

Liew, F., Martin, M. E., Tappel, R. C., Heijstra, B. D., Mihalcea, C., and Köpke, M. (2016). Gas fermentation—a flexible platform for commercial scale production of low-carbon-fuels and chemicals from waste and renewable feedstocks. *Front. Microbiol.* 7, 694. doi:10.3389/fmicb.2016.00694

Litvak, Y., Byndloss, M. X., and Bäumler, A. J. (2018). Colonocyte metabolism shapes the gut microbiota. *Science* 362, eaat9076. doi:10.1126/science.aat9076

Liu, C., Du, M.-X., Abuduaini, R., Yu, H.-Y., Li, D.-H., Wang, Y.-J., et al. (2021a). Enlightening the taxonomy darkness of human gut microbiomes with a cultured biobank. *Microbiome* 9, 119. doi:10.1186/s40168-021-01064-3

Liu, N., Chen, L., Yan, M., Tao, Q., Wu, J., Chen, J., et al. (2023). Eubacterium rectale improves the efficacy of anti-PD1 immunotherapy in melanoma via l-serine-mediated NK cell activation. *Research* 6, 0127. doi:10.34133/research.0127

Liu, X., Mao, B., Gu, J., Wu, J., Cui, S., Wang, G., et al. (2021b). Blautia-a new functional genus with potential probiotic properties? *Gut Microbes* 13, 1–21. doi:10.1080/19490976.2021.1875796

Lomans, B. P., Leijdekkers, P., Wesselink, J. J., Bakkes, P., Pol, A., van der Drift, C., et al. (2001). Obligate sulfide-dependent degradation of methoxylated aromatic compounds and formation of methanethiol and dimethyl sulfide by a freshwater sediment isolate, Parasporobacterium paucivorans gen. nov., sp. nov. *Appl. Environ. Microbiol.* 67, 4017–4023. doi:10.1128/AEM.67.9.4017-4023.2001

Lopez, C. A., McNeely, T. P., Nurmakova, K., Beavers, W. N., and Skaar, E. P. (2020). Clostridioides difficile proline fermentation in response to commensal clostridia. *Anaerobe* 63, 102210. doi:10.1016/j.anaerobe.2020.102210

Louie, T., Golan, Y., Khanna, S., Bobilev, D., Erpelding, N., Fratazzi, C., et al. (2023). VE303, a defined bacterial consortium, for prevention of recurrent Clostridioides difficile infection: a randomized clinical trial. *Jama* 329, 1356–1366. doi:10.1001/jama.2023.4314

Lu, H., Xu, X., Fu, D., Gu, Y., Fan, R., Yi, H., et al. (2022). Butyrate-producing Eubacterium rectale suppresses lymphomagenesis by alleviating the TNF-induced TLR4/MyD88/NF-κB axis. *Cell Host Microbe* 30, 1139–1150.e7. doi:10.1016/j.chom.2022.07.003

Mayo, B., Vázquez, L., and Flórez, A. B. (2019). Equol: a bacterial metabolite from the daidzein isoflavone and its presumed beneficial health effects. *Nutrients* 11, 2231. doi:10.3390/nu11092231

McGovern, B. H., Ford, C. B., Henn, M. R., Pardi, D. S., Khanna, S., Hohmann, E. L., et al. (2021). SER-109, an investigational microbiome drug to reduce recurrence after Clostridioides difficile infection: lessons learned from a phase 2 trial. *Clin. Infect. Dis.* 72, 2132–2140. doi:10.1093/cid/ciaa387

McLellan, S. L., Newton, R. J., Vandewalle, J. L., Shanks, O. C., Huse, S. M., Eren, A. M., et al. (2013). Sewage reflects the distribution of human faecal *Lachnospiraceae*. *Environ. Microbiol.* 15, 2213–2227. doi:10.1111/1462-2920.12092

Mechichi, T., Labat, M., Garcia, J. L., Thomas, P., and Patel, B. K. (1999). Sporobacterium olearium gen. nov., sp. nov., a new methanethiol-producing bacterium that degrades aromatic compounds, isolated from an olive mill wastewater treatment digester. *Int. J. Syst. Bacteriol.* 49 (Pt 4), 1741–1748. doi:10.1099/00207713-49-4-1741

Meehan, C. J., and Beiko, R. G. (2014). A phylogenomic view of ecological specialization in the Lachnospiraceae, a family of digestive tract-associated bacteria. *Genome Biol. Evol.* 6, 703–713. doi:10.1093/gbe/evu050

NCBI Genome Datasets (Lachnospiraceae) (2023). Genome, https://www.ncbi.nlm.nih.gov/datasets/genome/?taxon=186803&reference_only=true.

O Sheridan, P., Martin, J. C., Lawley, T. D., Browne, H. P., Harris, H. M. B., Bernalier-Donadille, A., et al. (2016). Polysaccharide utilization loci and nutritional specialization in a dominant group of butyrate-producing human colonic Firmicutes. *Microb. Genom* 2, e000043. doi:10.1099/mgen.0.000043

Park, H., Patel, A., Hunt, K. A., Henson, M. A., and Carlson, R. P. (2020). Artificial consortium demonstrates emergent properties of enhanced cellulosic-sugar degradation and biofuel synthesis. *NPJ Biofilms Microbiomes* 6, 59. doi:10.1038/s41522-020-00170-8

Petit, E., LaTouf, W. G., Coppi, M. V., Warnick, T. A., Currie, D., Romashko, I., et al. (2013). Involvement of a bacterial microcompartment in the metabolism of fucose and rhamnose by Clostridium phytofermentans. *PLoS ONE* 8, e54337. doi:10.1371/journal.pone.0054337

Petrof, E. O., Gloor, G. B., Vanner, S. J., Weese, S. J., Carter, D., Daigneault, M. C., et al. (2013). Stool substitute transplant therapy for the eradication of *Clostridium difficile* infection: 'RePOOPulating' the gut. *Microbiome* 1, 3. doi:10.1186/2049-2618-1-3

Port, F., Starostecka, M., and Boutros, M. (2020). Multiplexed conditional genome editing with Cas12a in Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* 117, 22890–22899. doi:10.1073/pnas.2004655117

Preethi, M., Kumar, G., Karthikeyan, O. P., and Varjani, S. (2021). Lignocellulosic biomass as an optimistic feedstock for the production of biofuels as valuable energy source: techno-economic analysis, Environmental Impact Analysis, Breakthrough and Perspectives. *Environ. Technol. Innovation* 24, 102080. doi:10.1016/j.eti.2021.102080

Quigley, E. M. M., Markinson, L., Stevenson, A., Treasure, F. P., and Lacy, B. E. (2023). Randomised clinical trial: efficacy and safety of the live biotherapeutic product MRx1234 in patients with irritable bowel syndrome. *Aliment. Pharmacol. Ther.* 57, 81–93. doi:10.1111/apt.17310

Reeves, A. E., Koenigsknecht, M. J., Bergin, I. L., and Young, V. B. (2012). Suppression of *Clostridium difficile* in the gastrointestinal tracts of germfree mice inoculated with a murine isolate from the family Lachnospiraceae. *Infect. Immun.* 80, 3786–3794. doi:10.1128/IAI.00647-12

Reichardt, N., Duncan, S. H., Young, P., Belenguer, A., McWilliam Leitch, C., Scott, K. P., et al. (2014). Phylogenetic distribution of three pathways for propionate production within the human gut microbiota. *ISME J.* 8, 1323–1335. doi:10.1038/ismej.2014.14

Roager, H. M., and Licht, T. R. (2018). Microbial tryptophan catabolites in health and disease. *Nat. Commun.* 9, 3294. doi:10.1038/s41467-018-05470-4

Rostain, W., Zaplana, T., Boutard, M., Baum, C., Tabuteau, S., Sanitha, M., et al. (2022). Tuning of gene expression in Clostridium phytofermentans using synthetic promoters and CRISPRi. *ACS Synth. Biol.* 11, 4077–4088. doi:10.1021/acssynbio.2c00385

Rude, M. A., and Schirmer, A. (2009). New microbial fuels: a biotech perspective. *Curr. Opin. Microbiol.* 12, 274–281. doi:10.1016/j.mib.2009.04.004

Russell, W. R., Duncan, S. H., Scobbie, L., Duncan, G., Cantlay, L., Calder, A. G., et al. (2013). Major phenylpropanoid-derived metabolites in the human gut can arise from microbial fermentation of protein. *Mol. Nutr. Food Res.* 57, 523–535. doi:10.1002/mnfr.201200594

Sato, Y., Atarashi, K., Plichta, D. R., Arai, Y., Sasajima, S., Kearney, S. M., et al. (2021). Novel bile acid biosynthetic pathways are enriched in the microbiome of centenarians. *Nature* 599, 458–464. doi:10.1038/s41586-021-03832-5

Scarborough, M. J., Myers, K. S., Donohue, T. J., and Noguera, D. R. (2020). Medium-chain fatty acid synthesis by "Candidatus Weimeria bifida" gen. Nov., sp. nov., and "Candidatus pseudoramibacter fermentans" sp. nov. *Appl. Environ. Microbiol.* 86, 022422-19. doi:10.1128/AEM.02242-19

Schouw, A., Leiknes Eide, T., Stokke, R., Pedersen, R. B., Steen, I. H., and Bødtker, G. (2016). Abyssivirga alkaniphila gen. nov., sp. nov., an alkane-degrading, anaerobic bacterium from a deep-sea hydrothermal vent system, and emended descriptions of Natranaerovirga pectinivora and Natranaerovirga hydrolytica. *Int. J. Syst. Evol. Microbiol.* 66, 1724–1734. doi:10.1099/ijsem.0.000934

Scott, K. P., Martin, J. C., Campbell, G., Mayer, C.-D., and Flint, H. J. (2006). Whole-genome transcription profiling reveals genes up-regulated by growth on fucose in the human gut bacterium "Roseburia inulinivorans.". *J. Bacteriol.* 188, 4340–4349. doi:10.1128/JB.00137-06

Sell, L. B., Ramelow, C. C., Kohl, H. M., Hoffman, K., Bains, J. K., Doyle, W. J., et al. (2022). Farnesol induces protection against murine CNS inflammatory demyelination and modifies gut microbiome. *Clin. Immunol.* 235, 108766. doi:10.1016/j.clim.2021.108766

Seshadri, R., Leahy, S. C., Attwood, G. T., Teh, K. H., Lambie, S. C., Cookson, A. L., et al. (2018). Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. *Nat. Biotechnol.* 36, 359–367. doi:10.1038/nbt.4110

Shen, Z., Luo, W., Tan, B., Nie, K., Deng, M., Wu, S., et al. (2022). Roseburia intestinalis stimulates TLR5-dependent intestinal immunity against Crohn's disease. *eBioMedicine* 85, 104285. doi:10.1016/j.ebiom.2022.104285

Sheridan, P. O., Louis, P., Tsompanidou, E., Shaw, S., Harmsen, H. J., Duncan, S. H., et al. (2022). Distribution, organization and expression of genes concerned with anaerobic lactate utilization in human intestinal bacteria. *Microb. Genom* 8, 000739. doi:10.1099/mgen.0.000739

Sheridan, P. O., Martin, J. C., Minton, N. P., Flint, H. J., O'Toole, P. W., and Scott, K. P. (2019). Heterologous gene expression in the human gut bacteria Eubacterium rectale and Roseburia inulinivorans by means of conjugative plasmids. *Anaerobe* 59, 131–140. doi:10.1016/j.anaerobe.2019.06.008

Silber, J., Norman, J., Kanno, T., Crossette, E., Szabady, R., Menon, R., et al. (2022). Randomized, double-blind, placebo (pbo)-controlled, single- and multiple-dose phase 1 study of ve202, a defined bacterial consortium for treatment of ibd: safety and colonization dynamics of a novel live biotherapeutic product (lbp) in healthy adults. *Inflamm. Bowel Dis.* 28, S65–S66. doi:10.1093/ibd/izac015.106

Smirnova, M., Miamin, U., Kohler, A., Valentovich, L., Akhremchuk, A., Sidarenka, A., et al. (2021). Isolation and characterization of fast-growing green snow bacteria from coastal East Antarctica. *MicrobiologyOpen* 10, e1152. doi:10.1002/mbo3.1152

Sorbara, M. T., Littmann, E. R., Fontana, E., Moody, T. U., Kohout, C. E., Gjonbalaj, M., et al. (2020). Functional and genomic variation between human-derived isolates of Lachnospiraceae reveals inter- and intra-species diversity. *Cell Host Microbe* 28, 134–146.e4. doi:10.1016/j.chom.2020.05.005

Stevenson, B. S., and Schmidt, T. M. (2004). Life history implications of rRNA gene copy number in *Escherichia coli*. *Appl. Environ. Microbiol.* 70, 6670–6677. doi:10.1128/AEM.70.11.6670-6677.2004

Sugiyama, Y., Masumori, N., Fukuta, F., Yoneta, A., Hida, T., Yamashita, T., et al. (2013). Influence of isoflavone intake and equol-producing intestinal flora on prostate cancer risk. *Asian Pac J. Cancer Prev.* 14, 1–4. doi:10.7314/apjcp.2013.14.1.1

Tak, Y. E., Kleinstiver, B. P., Nuñez, J. K., Hsu, J. Y., Horng, J. E., Gong, J., et al. (2017). Inducible and multiplex gene regulation using CRISPR-Cpf1-based transcription factors. *Nat. Methods* 14, 1163–1166. doi:10.1038/nmeth.4483

Takada, T., Kurakawa, T., Tsuji, H., and Nomoto, K. (2013). Fusicatenibacter saccharivorans gen. nov., sp. nov., isolated from human faeces. *Int. J. Syst. Evol. Microbiol.* 63, 3691–3696. doi:10.1099/ijs.0.045823-0

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38, 3022–3027. doi:10.1093/molbev/msab120

Tolonen, A. C., Beauchemin, N., Bayne, C., Li, L., Tan, J., Lee, J., et al. (2022). Synthetic glycans control gut microbiome structure and mitigate colitis in mice. *Nat. Commun.* 13, 1244. doi:10.1038/s41467-022-28856-x

Tolonen, A. C., Cerisy, T., El-Sayyed, H., Boutard, M., Salanoubat, M., and Church, G. M. (2015a). Fungal lysis by a soil bacterium fermenting cellulose. *Environ. Microbiol.* 17, 2618–2627. doi:10.1111/1462-2920.12495

Tolonen, A. C., Chilaka, A. C., and Church, G. M. (2009). Targeted gene inactivation in Clostridium phytofermentans shows that cellulose degradation requires the family 9 hydrolase Cphy3367. *Mol. Microbiol.* 74, 1300–1313. doi:10.1111/j.1365-2958.2009.06890.x

Tolonen, A. C., Haas, W., Chilaka, A. C., Aach, J., Gygi, S. P., and Church, G. M. (2011). Proteome-wide systems analysis of a cellulosic biofuel-producing microbe. *Mol. Syst. Biol.* 7, 461. doi:10.1038/msb.2010.116

Tolonen, A. C., Zuroff, T. R., Ramya, M., Boutard, M., Cerisy, T., and Curtis, W. R. (2015b). Physiology, genomics, and pathway engineering of an ethanol-tolerant strain of Clostridium phytofermentans. *Appl. Environ. Microbiol.* 81, 5440–5448. doi:10.1128/AEM.00619-15

Tran, M. H., Park, H., Nobles, C. L., Karunadharma, P., Pan, L., Zhong, G., et al. (2021). A more efficient CRISPR-Cas12a variant derived from Lachnospiraceae bacterium MA2020. *Mol. Ther. Nucleic Acids* 24, 40–53. doi:10.1016/j.omtn.2021.02.012

Uchiyama, J., Akiyama, M., Hase, K., Kumagai, Y., and Kim, Y.-G. (2022). Gut microbiota reinforce host antioxidant capacity via the generation of reactive sulfur species. *Cell Rep.* 38, 110479. doi:10.1016/j.celrep.2022.110479

Udayappan, S., Manneras-Holm, L., Chaplin-Scott, A., Belzer, C., Herrema, H., Dallinga-Thie, G. M., et al. (2016). Oral treatment with Eubacterium hallii improves insulin sensitivity in db/db mice. *NPJ Biofilms Microbiomes* 2, 16009. doi:10.1038/npjbiofilms.2016.9

Vacca, M., Celano, G., Calabrese, F. M., Portincasa, P., Gobbetti, M., and De Angelis, M. (2020). The controversial role of human gut Lachnospiraceae. *Microorganisms* 8, E573. doi:10.3390/microorganisms8040573

Vedanta Biosciences, Inc (2023). Vedanta Biosciences VE303 program. https://www.vedantabio.com/pipeline/ve303.

Vera-Ponce de Leon, A., Schneider, M. G., Jahnes, B. C., Sadowski, V., Camuy-Vélez, L. A., Duan, J., et al. (2022). Genetic drift and host-adaptive features likely underlie cladogenesis of insect-associated Lachnospiraceae. *Genome Biol. Evol.* 14, evac086. doi:10.1093/gbe/evac086

Vital, M., Howe, A. C., and Tiedje, J. M. (2014). Revealing the bacterial butyrate synthesis pathways by analyzing (meta)genomic data. *MBio* 5, e00889. doi:10.1128/mBio.00889-14

Wachenheim, D. E., and Patterson, J. A. (1992). Anaerobic production of extracellular polysaccharide by Butyrivibrio fibrisolvens nyx. *Appl. Environ. Microbiol.* 58, 385–391. doi:10.1128/aem.58.1.385-391.1992

Walker, A. C., Bhargava, R., Vaziriyan-Sani, A. S., Pourciau, C., Donahue, E. T., Dove, A. S., et al. (2021). Colonization of the *Caenorhabditis elegans* gut with human enteric bacterial pathogens leads to proteostasis disruption that is rescued by butyrate. *PLoS Pathog.* 17, e1009510. doi:10.1371/journal.ppat.1009510

Warnick, T. A., Methé, B. A., and Leschine, S. B. (2002). Clostridium phytofermentans sp. nov., a cellulolytic mesophile from forest soil. *Int. J. Syst. Evol. Microbiol.* 52, 1155–1160. doi:10.1099/ijs.0.02125-0

Watanabe, M., Kaku, N., Ueki, K., and Ueki, A. (2016). Falcatimonas natans gen. nov., sp. nov., a strictly anaerobic, amino-acid-decomposing bacterium isolated from a methanogenic reactor of cattle waste. *Int. J. Syst. Evol. Microbiol.* 66, 4639–4644. doi:10.1099/ijsem.0.001403

Whitman, W. B. (2009). *Systematic bacteriology*. New York, NY, USA: Springer. doi:10.1007/978-0-387-68489-5

Wolin, M. J., Miller, T. L., Collins, M. D., and Lawson, P. A. (2003). Formate-dependent growth and homoacetogenic fermentation by a bacterium from human feces: description of Bryantella formatexigens gen. nov., sp. nov. *Appl. Environ. Microbiol.* 69, 6321–6326. doi:10.1128/AEM.69.10.6321-6326.2003

Wu, C., Xu, W., Yu, W., Li, Z., Zhang, Y., and Zhang, F. (2021). Strain-level screening of human gut microbes identifies Blautia producta as a novel anti-hyperlipidemic probiotic via the production of 12-methylmyristic acid. https://www.researchsquare.com/article/rs-989302/v1.

Yamashita, H., Fujisawa, K., Ito, E., Idei, S., Kawaguchi, N., Kimoto, M., et al. (2007). Improvement of obesity and glucose tolerance by acetate in Type 2 diabetic Otsuka Long-Evans Tokushima Fatty (OLETF) rats. *Biosci. Biotechnol. Biochem.* 71, 1236–1243. doi:10.1271/bbb.60668

Yang, C., Guan, L., Liu, J., and Wang, J. (2015). Rumen fermentation and acetogen population changes in response to an exogenous acetogen TWA4 strain and *Saccharomyces cerevisiae* fermentation product. *J. Zhejiang Univ. Sci. B* 16, 709–719. doi:10.1631/jzus.B1500013

Zhang, J., Song, L., Wang, Y., Liu, C., Zhang, L., Zhu, S., et al. (2019). Beneficial effect of butyrate-producing Lachnospiraceae on stress-induced visceral hypersensitivity in rats. *J. Gastroenterol. Hepatol.* 34, 1368–1376. doi:10.1111/jgh.14536

Zhang, L., Li, G., Zhang, Y., Cheng, Y., Roberts, N., Glenn, S. E., et al. (2023). Boosting genome editing efficiency in human cells and plants with novel LbCas12a variants. *Genome Biol.* 24, 102. doi:10.1186/s13059-023-02929-6

Zhou, Y., Wei, Y., Jiang, L., Jiao, X., and Zhang, Y. (2023). Anaerobic phloroglucinol degradation by Clostridium scatologenes. *mBio* 14, e0109923–23. doi:10.1128/mbio.01099-23

Zuroff, T. R., Barri Xiques, S., and Curtis, W. R. (2013). Consortia-mediated bioprocessing of cellulose to ethanol with a symbiotic Clostridium phytofermentans/yeast co-culture. *Biotechnol. Biofuels* 6, 59. doi:10.1186/1754-6834-6-59

CHAPTER 7

Technologies to Study Plant Biomass Fermentation Using the Model Bacterium Clostridium Phytofermentans

ANDREW C. TOLONEN,*[a,b,c] ELSA PETIT,[d] JEFFREY L. BLANCHARD,[e] TOM WARNICK[f] AND SUSAN B. LESCHINE[f]

[a] CEA, DSV, IG, Genoscope, Evry, France; [b] CNRS-UMR8030, Evry, France; [c] Université d'Evry Val d'Essonne, Evry, France; [d] Amherst College, Amherst MA, USA; [e] Department of Biology, University of Massachusetts, Amherst MA, USA; [f] Department of Veterinary and Animal Sciences, University of Massachusetts, Amherst MA, USA
*Email: atolonen@genoscope.cns.fr

7.1 Introduction

Dwindling oil reserves and climate change have made the development of renewable, carbon-balanced energy resources a global necessity. As cellulosic biomass is the world's most abundant biological energy source,[1] fermentation of inedible, cellulosic feedstocks such as prairie grasses, corn stover, wood chips and agricultural waste represents one of the most promising avenues for replacing petroleum-based fuels and products. Indeed, over 1.3 billion metric tons of cellulosic biomass could be used annually to produce

RSC Energy and Environment Series No. 10
Biological Conversion of Biomass for Fuels and Chemicals: Explorations from Natural Utilization Systems
Edited by Jianzhong Sun, Shi-You Ding and Joy Doran-Peterson

Published by the Royal Society of Chemistry, www.rsc.org

fuels and chemicals in North America without affecting food supplies.[2] Fermentation on this scale could provide enough ethanol for 65% of the USA's ground transportation fuel at current levels,[3] which is attractive for USA energy security because it imports 60% of it's petroleum.[4] Encouraged by this opportunity, the US Congress passed the Energy Independence and Security Act (EISA) in 2007 mandating an annual production of 36 billion gallons of biofuels by 2022, of which 16 billion gallons should come from sources such as cellulosic feedstocks.

Presently, the main barrier to the expansion of cellulosic-based fuels and bioproducts is the recalcitrance of biomass.[4] While plant biomass is composed primarily of pentose and hexose sugars, they are linked within high-molecular-weight polysaccharides to form a complex network that resists biological and chemical degradation. Furthermore, no single crop can be cultivated in all geographic areas and candidate feedstocks vary widely in their polysaccharide composition: cellulose (24–45%), hemicellulose (18–29%), and pectin (1–24%).[5] Due to the complexity and variability of plant biomass, the cost of breaking it into sugars using purified enzymes and chemical pretreatment currently doubles the price per unit of carbohydrate, which erases the cost savings of biomass relative to corn.[6]

Consolidated bioprocessing (CBP)[7] is a promising strategy to overcome biomass recalcitrance in a cost-effective manner by using a single microbial culture to combine biomass deconstruction and sugar fermentation into one step. The projected costs for converting biomass to ethanol using CBP are four-fold lower than a two-step process based on simultaneous saccharification and co-fermentation (SSCF) using purified cellulases.[8] However, CBP requires a single microbe or stable microbial consortia with five properties: (1) enzyme production to degrade diverse plant polysaccharides; (2) fermentation of 5- and 6-carbon sugars; (3) minimal production of inhibitory side-products; (4) growth in conditions compatible with industrial bioreactors; and (5) the ability to produce a high concentration of fuel product. *Clostridium phytofermentans* is remarkable in possessing all five of these properties, enabling it to ferment biomass with saccharification rates and ethanol yields comparable to SSCF.[9]

C. Phytofermentans, a mesophilic anaerobe isolated from forest soil,[10] which grows on both the soluble and insoluble components of plant biomass by first cleaving plant polysaccharides (cellulose, hemicellulose, starch, and pectin) and then fermenting the resulting pentose and hexose sugars. According to the CAZy database,[11] the *C. phytofermentans* genome encodes 169 carbohydrate-active enzymes (CAZys) to cut diverse plant substrates into sugars, which is the largest number of CAZys among sequenced clostridia. This chapter details how *C. phytofermentans* is unique among the characterized microbes in its broad substrate range and efficient ethanol production, making it an excellent model system for cellulosic biofuels research. Specifically, we will focus on three areas of *C. phytofermentans* research: isolation and physiological characterization; systems biology; and genetic engineering.

7.2 Isolation and Physiology

7.2.1 Isolation

This chapter focuses on the *C. phytofermentans* strain ISDgT ATCC 700394, which was isolated in August 1991 from damp silt in a stream bed up-stream of a mill pond dam near Quabbin Reservoir in central Massachusetts, USA.[10] The strain name "ISDg" refers to the isolation site: "Intermittent Stream behind Dam". Samples were recorded alphabetically, so "g" is the seventh isolate from the ISD location. The "T" means strain ISDg is the "type strain" used for taxonomy.

To isolate strain ISDgT, a spatula of 0.2–0.5 g soil was inoculated into GS-2 medium[12] with 6 g L^{-1} ball-milled cellulose as a sole carbon source (GS-2C medium) and incubated anaerobically for 1 month at 30 °C. A 0.1 mL aliquot was twice transferred to 4 mL GS-2C medium and incubated for 2 weeks. The culture was then diluted into tubes containing 4 mL melted GS-2C soft-agar medium and poured onto plates with a basal layer of GS-2 medium with no added carbon. Colonies that produced clear zones in cellulose overlays were twice streaked on GS-2 cellobiose plates and colonies were picked and grown in GS-2 cellobiose liquid medium. Cultures were transferred to GS-2C medium to confirm they were cellulolytic and stocks were frozen at –80 °C in 15% glycerol.

At least 16 different cellulolytic anaerobes were isolated from nine different soil samples during this sampling survey. Four of the eight isolates from the ISD site were cellulolytic and a preliminary characterization of all 16 isolates showed that ISDg produced the highest ethanol titers. Ribosomal sequencing showed that *C. phytofermentans* belongs to clostridia group 14 A, which is phylogenetically distant from well-studied cellulolytic clostridia such as *C. thermocellum* and *C. cellulovorans* (Figure 7.1(a)). *C. phytofermentans* ISDg was thus chosen for further study based on its ability to efficiently ferment cellulose to ethanol and its phylogenetic distance from previously characterized clostridia.

7.2.2 Growth Conditions

C. phytofermentans ISDgT cells are straight rods 0.5–0.8 μm by 3–15 μm in size. Cells growing on soluble substrates are motile, usually with one or two flagella (Figure 7.1(b)). When growing on insoluble plant matter such as cellulose, cells are shorter, non-flagellated, and adhere to the substrate (Figure 7.1(c)). Adhesion to biomass is an adaptation used by some cellulolytic microbes to increase enzyme concentrations near the substrate and to exclude competitors from the liberated sugars.[13] *C. phytofermentans* can bind both cellulose and hemicellulose,[14] even though it lacks the cellulosomes that enable attachment by other clostridia. Cells growing on hemicellulose sometime bear surface nodules, suggesting that hemicellulose particles bind to the cell surface. The mechanism by which *C. phytofermentans* adheres to plant substrates is unknown, but may include cell-surface-binding proteins, secretion of an

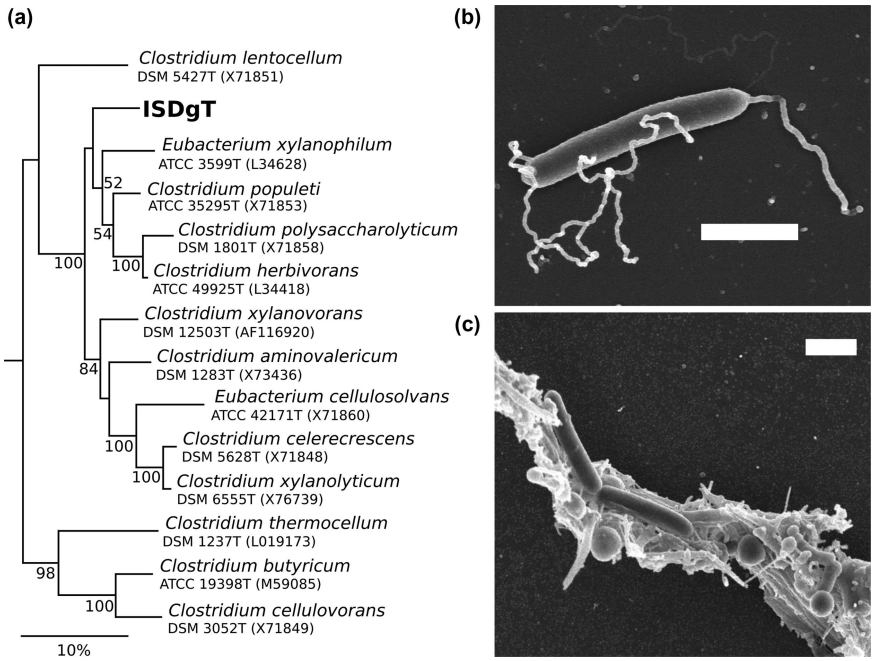


Figure 7.1 (a) Phylogenetic tree based on the maximum likelihood analysis of *C. phytofermentans* and related clostridia 16S rRNA gene sequences. Clostridia are members of group 14a, except *Clostridium cellulovorans* (cluster 1), *Clostridium thermocellum* (cluster 3) and *Clostridium lentocellum* (cluster 14b) which are included as cluster representatives. Bootstrap values greater than 50%, expressed as percentages of 100 replications, are shown at the branching points. The bar represents 10 substitutions per 100 base pairs. *E. Coli* was used as the outgroup to root the tree (image adapted from ref. 10). (b) Scanning electron microscopy shows a log-phase cell with two flagella. (c) Scanning electron micrograph of cells growing on cellulose fibers. Cells are adhered to cellulose and spores are also visible. White scale bars are 1 μm.

extra-polysaccharide glycocalyx, or cell surface pili that facilitate binding as in *Ruminococcus albus*.[15]

C. phytofermentans grows fastest at 37 °C. It will grow slowly at temperatures as low as 15 °C and as high as 42 °C. Cells grow well at pH 6.0–9.0, but poorly at pH 9.5 and not at all at pH 5.5. The maximum growth rate occurs when the initial pH is 8.0, whereas the highest growth yield is achieved when the initial pH is 8.5. Strain ISDgT ferments numerous carbon sources: pentoses (arabinose, ribose, xylose), hexoses (glucose, glucuronic acid, galactose, galacturonic acid, fructose, mannose), disaccharides (cellobiose, gentiobiose, lactose, maltose), and polysaccharides (cellulose, xylan, pectin, starch, laminarin). ISDgT does not grow on inulin, mannan, glycerol, pyruvate, sucrose, trehalose or tryptone.[10]

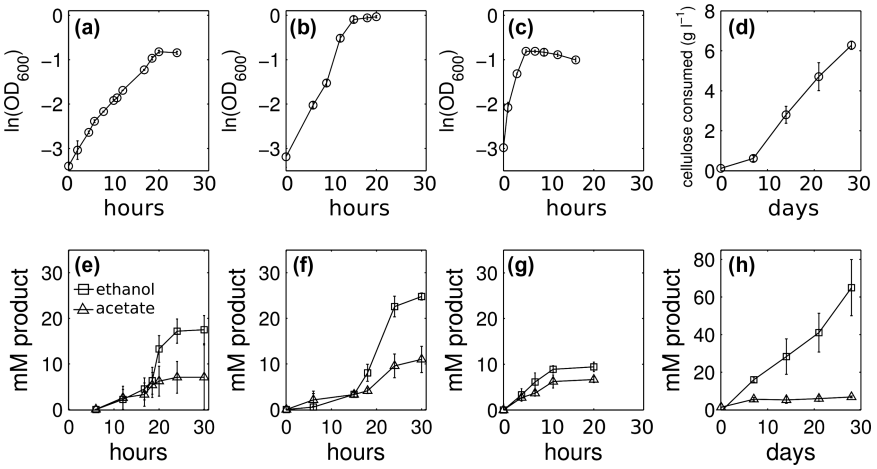


Figure 7.2 (a)–(d) *C. phytofermentans* growth and (e)–(h) accumulation of fermentation products on (a) and (e) 0.3 g L^{-1} xylose, (b) and (f) glucose, (c) and (g) hemicellulose and (d) and (h) 1.2 g L^{-1} cellulose. Data points are means of triplicate cultures. Error bars show one standard deviation and are smaller than the symbols where not apparent. Growth on xylose (a), glucose (b), and hemicellulose (c) was quantified as OD_{600}. Growth on cellulose (d) was measured as dry mass of cellulose in culture. Production of ethanol and acetate, the two most abundant fermentation products, was measured by HPLC as described in Tolonen *et al.*[14]

C. phytofermentans sometimes grows better on polysaccharides than on their constituent sugars. For example, growth on xylose (Figure 7.2(a)) is slower than on hemicellulose xylan (Figure 7.2(c)), even though hemicellulose xylan is a β-1,4-D-xylopyranose polymer that must be cleaved to xylose and isomerized before glycolysis. Faster growth on polysaccharides may result from *C. phytofermentans* having higher affinity transporters for oligosaccharides than for sugar monomers. Also, because ATP for sugar transport is used on a per-molecule basis, uptake of oligosaccharides saves energy relative to sugar monomers.[16] Phosphorolysis of xylan by a yet unidentified phosphorylase might also save ATP relative to hydrolysis by a mechanism similar to cellodextrin phosphorylases.[17] However, when xylan is hydrolyzed, the resulting xylose is isomerized to xylulose by xylose isomerase (Cphy0200, Cphy1219) before being phosphorylated to xylulose 5-phosphate by xylulokinase (Cphy3419). This suggests that the terminal xylose residues on xylosaccharides would have to be isomerized before phosphorolysis or the phosphorylase would need to isomerize xylose in parallel.

Although standard GS-2 media contains ammonium chloride or urea as a nitrogen source, there is evidence that *C. phytofermentans* can fix nitrogen if grown in a defined medium without nitrogen.[18] Nitrogen fixation is common among clostridia, indeed *C. pasteurianum* was the first free-living, nitrogen-fixing organism described almost 120 years ago.[19] The genomes of nitrogen-fixing clostridia contain *nif* genes: *nifH*, *nifD*, and *nifK* encode nitrogenase

proteins and *nifE*, *nifB-N* and *nifV* construct the iron–molybdenum nitrogenase co-factor. *C. phytofermentans* has potential orthologs to *nifH* (*cphy2693*, *cphy3935*), *nifB* (*cphy2391*, *cphy1480*), *nifV* (*cphy3171*), and two genes to provide [Fe–S] clusters for synthesis of the Fe–Mo co-factor, *nifS* (*cphy2907*) and *nifU* (*cphy2907*, *cphy3260*). However, *nifD*, *nifK*, and *nifE* are missing and the putative *nif* genes in *C. phytofermentans* are not co-localized in the genome as they are in *C. pasteurianum*, *C. acetobutylicum*, and *C. beijerinckii*.[20] Further work is needed to characterize nitrogen fixation by *C. phytofermentans*, but it is an interesting opportunity to ferment biomass without nitrogen supplementation and to produce additional hydrogen.[18]

A recent advance in the culturing of *C. phytofermentans* is the formulation of chemically defined media that do not contain the yeast extract present in GS-2 medium. These defined media consist of a carbon source, an inorganic phosphate, inorganic nitrogen, salts ($MgCl_2$, $CaCl_2$, $FeSO_4$), cysteine HCl as a reducing agent, sodium citrate, and seven B-vitamins: thiamine (B1), riboflavin (B2), nicotinate (B3), pantethine (B5), pyridoxal (B6), biotin (B7), and folinic acid (B9).[21] B-Vitamin auxotrophies are common in clostridia. *C. thermocellum* is auxotrophic for pyridoxal, biotin, folate, and cobalamin (B12).[12] Similarly, *C. cellulolyticum* cannot synthesize biotin, folate, nicotinate, riboflavin, panthoteate, thiamine, and cobalamin.[22] An unanticipated benefit of using chemically defined media, at least in *C. cellulolyticum*, is that specific ethanol production rates are 10-fold higher than in complex media.[22] The authors reasoned that in complex media, *C. cellulolyticum* accumulates pyruvate leading to NADH-induced growth arrest, whereas cells in defined media are better able to regulate NADH-to-NAD^+ ratios.

Because B-vitamin addition is expensive for large-scale industrial reactors, we want to understand the genetic basis of *C. phytofermentans* B-vitamin auxotrophies (Figure 7.3) to enable the development of prototrophic strains. Four of the seven B-vitamin auxotrophies (thiamine, riboflavin, pantothenate, and biotin) result from *C. phytofermentans* missing entire multi-gene pathways that would be challenging to transfer to the genome. Although *C. phytofermentans* lacks most or all of the genes to make these vitamins, they are each co-factors for many required enzymes. For example, the pyrophosphate derivative (TPP) of thiamine (B1) is a co-enzyme of highly expressed decarboxylase enzymes such as pyruvate ferredoxin oxidoreductase (Cphy3558), oxoglutarate ferredoxin oxidoreductase (Cphy3122-3), and transketolase (Cphy0014). Rather than making thiamine, *C. phytofermentans* scavenges it using a thiamine transporter (Cphy0729). The riboflavin (B2) derivatives flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD) are co-factors that mediate redox reactions; 10 proteins are annotated as FMN-binding and 42 bind FAD. Riboflavin appears to be taken into the cell with a transporter (Cphy0344) that has a high similarity to the *B. subtilis* riboflavin transporter YpaA. Pantothenate (B5) is a building block of co-enzyme A, which is required to oxidize pyruvate and to build fatty acids. Rather than building pantothenate, *C. phytofermentans* uptakes pantethine, splits it into pantetheine, and converts it to CoA. Biotin (B7) is a co-factor of at least six carboxylase

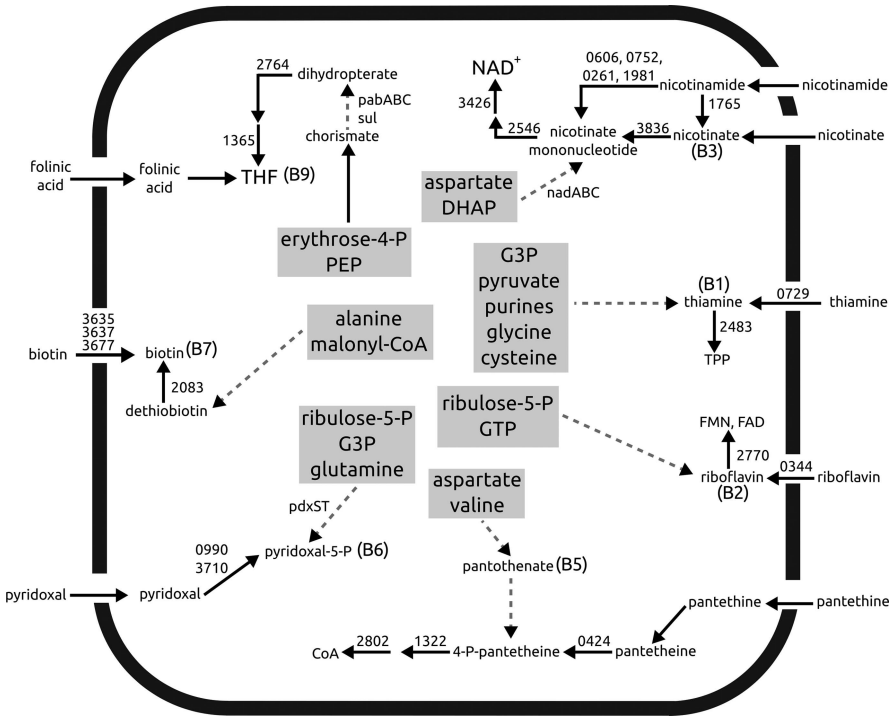


Figure 7.3 Map of the genetic basis of *C. phytofermentans* B-vitamin auxotrophies. The metabolites used for *de novo* synthesis of each vitamin are shown. The arrows represent one or more enzymatic steps: solid arrows are reactions catalyzed in *C. phytofermentans* and dashed arrows are missing reactions. The numbers denote NCBI protein annotations. For vitamins B3, B6, and B9, the *B. subtilis* genes that might complement *C. phytofermentans* auxotrophies are shown. Abbreviations: glyceraldehyde-3-phosphate (G3P), guanosine triphosphate (GTP), phosphoenolpyruvate (PEP), tetrahydrofolate (THF), flavin mononucleotide (FMN), flavin adenine dinucleotide (FAD), thiamine pyrophosphate (TPP).

enzymes such as acetyl-CoA carboxylase, (Cphy0519, Cphy0521-3) which adds a carboxyl group to acetyl-CoA to form malonyl-CoA for fatty acid synthesis. *C. Phytofermentans* likely uptakes biotin with a tripartite biotin transporter.[23]

Auxotrophies for three other B-vitamins, nicotinate, pyridoxal, and folate, result from *C. phytofermentans* missing only one or a few enzymes, suggesting these auxotrophies could be readily overcome by gene transfer. *C. Phytofermentans* requires NAD^+ for various electron-transfer reactions, but needs nicotinate (B3) or nicotinamide for growth because it lacks the *nadABC* genes for *de novo* synthesis of nicotinate mononucleotide from aspartate and dihydroxyacetone phosphate. *C. phytofermentans* can, however, salvage NAD^+ from nicotinate in three steps: (1) nicotinate and D-ribose-5-phosphate are combined to form nicotinate mononucleotide by Cphy3836; (2) nicotinate adenine dinucleotide is formed from nicotinate mononucleotide and ATP by

Cphy2546; and (3) finally Cphy3426 adds NH_2 to form NAD^+, which can be converted to $NADP^+$ by Cphy2509. According to KEGG, *C. phytofermentans* also has a second pathway to convert nicotinate and nicotinamide to the mononucleotide forms using Cphy0752, Cphy0261, and Cphy0606, which were all up-regulated on cellulose relative to glucose.[14] Although proteins for the other NAD^+ synthesis pathway were not similarly up-regulated, *C. phytofermentans* may have an increased need of NAD^+ during growth on cellulose. Pyridoxal-5-phosphate (B6), PLP, is a putative co-factor for 15 *C. phytofermentans* enzymes including transaminases, amino acid racemases and decarboxylases, and sulfhydrylases. PLP synthesis in *B. subtilis* requires the PdxST complex,[24] also called "YaaDE", which makes PLP from either ribose-5-phosphate or ribulose 5-phosphate, glyceraldehyde 3-phosphate or dihydroxyacetone phosphate, and glutamine.[25] Because *C. phytofermentans* makes all the substrates of the PdxST complex, it is feasible that transformation of *C. phyofermentans* with *B. subtilis pdxS–pdxT* could rescue the PLP auxotrophy.

Tetrahydrofolate (THF), the active form of folic acid (B9), is required for many single-carbon-transfer reactions, but the pathway to make THF is incomplete in *C. phytofermentans*. It uses erythrose-4-phosphate and PEP to build shikimate which can, in turn, be used to make chorismate. However, the *pabABC* and *sul* genes to convert chorismate to dihydropteroate are missing. Following these three steps, *C. phytofermentans* can use dihydropteroate make dihydrofolate (DHF) using Cphy2764 and can reduce DHF to THF with dihydrofolate reductase (Cphy1365). Experimental evidence supports that the THF auxotrophy can only be subsidized with folinic acid, even though it is not a cofactor in any known reactions.[26] It is unclear if this preference for folinic acid is due to its increased stability,[27] if *C. phytofermentans* has a specific folinic acid requirement, or if the putative folate transporter is more efficient at transporting folinic acid.

A benefit of the alternative and missing enzymes for THF metabolism is that *C. phytofermentans* is resistant to anti-folate antibiotics, which can thus be used to purge contaminant bacteria from mixed cultures. For example, dihydrofolate reductase (Cphy1365) is similar to other clostridial orthologs, but not to those of *E. coli* or *B. subtilis*. As trimethoprim acts by inhibiting dihydrofolate reductase, the variant enzyme likely underlies the natural resistance of *C. phytofermentans* to trimethoprim such that this antibiotic can be used to remove *E. coli* following conjugation.[28] Also, the lack of dihydropteroate synthase to make dihydropteroate from 4-Aminobenzoic acid (PABA) should make *C. phytofermentans* resistant to sulfonamide antibiotics.

7.2.3 Fermentation

Ethanol is the generally the primary fermentation product for *C. phytofermentans* and ethanol yields can approach the theoretical maximum. To calculate ethanol yields, we assume that each glucose molecule entering glycolysis forms two pyruvate molecules that each can be converted to one ethanol molecule, giving a maximum yield of two ethanols formed per glucose

consumed. As the initial glucose concentration in cultures was 16 mM (3 g^{-1}), ethanol reaching 24.81 mM in 30 h (Figure 7.2(f)) corresponds to a yield of 77%. After 48 h, ethanol concentrations in glucose cultures reach >95% of the maximum theoretical yield.[14] In cellulose cultures, stable cell densities (10^7–10^8 colony forming units (CFU) mL^{-1}) result in a constant rate of cellulose degradation (Figure 7.2(d)) and ethanol formation (Figure 7.2(h)). Assuming cellulose is composed of polymerized glucose, the conversion of cellulose to ethanol is 68% of the maximum theoretical yield. Furthermore, the ethanol-to-acetate ratio of 9.54 in the cellulose cultures is among the highest yields reported for clostridia.[7]

One xylose can be fermented to 5/3 ethanol equivalents; xylose is converted to xylulose-5-phosphate and fed into the non-oxidative pentose phosphate pathway where three xylulose-5-phosphate molecules are converted into two fructose-6-phosphates (four ethanols) and one glyceraldehyde-3-phosphate (one ethanol). Conversion of 20 mM (3 g^{-1}) xylose to 17.1 mM ethanol in 24 h (Figure 7.2(e)) thus represents a 52% ethanol yield. Assuming that xylan (Sigma X0502) is made entirely of xylose, the initial xylose concentration in the hemicellulose cultures is 22.8 mM and the final ethanol concentration of 10.35 mM (Figure 7.2(g)) gives an ethanol yield of 27.2%. Rapid growth on hemicellulose (Figure 7.2(c)) is thus accompanied by a reduced ethanol-to-acetate ratio (Figure 7.2(g)).

A shift to produce lower ethanol-to-acetate ratios during growth on hemicellulose is reinforced by recent results that acetate can be the most abundant fermentation product when *C. phytofermentans* grows on AFEX-treated corn stover.[9] This study showed that acetate was mostly produced during early growth on AFEX-treated stover when hemicellulose was preferentially consumed; afterwards the fermentation products shifted to ethanol when glucans were metabolized. Acetate could be viewed as the preferred fermentation product because its synthesis makes ATP, whereas ethanol formation wastes potential energy by oxidizing NADH to maintain redox balance. Efficient carbon assimilation during growth on hemicellulose may allow the cell to produce more acetate while still making enough ethanol to maintain redox balance. During growth on cellulose, cells still need to make ethanol to maintain redox balance, but are unable to also produce as much acetate for ATP because it is more expensive to harvest carbon.

7.3 Systems Biology

A comprehensive, molecular-level understanding of how *C. phytofermentans* ferments biomass will enhance our ability to optimize strains for transformation of plant matter into useful biochemicals. Because biomass fermentation by *C. phytofermentans* is such a complex process involving a myriad of enzymes, systems biology is proving to be a vital tool for rapid, high-throughput analysis of changes in cell state under different environmental conditions. Our systems biology studies seek an improved understanding of three areas: genome annotation (defining of genes, regulatory elements, and regulons), assembling

proteins into metabolic and regulatory networks, and identifying the function of key enzymes for cellulosic breakdown and fermentation. Initially, we used approaches including genome sequencing and proteomics. These systems-level analyses show that *C. phytofermentans* has a treasure trove of novel, cellulolytic and fermentative enzymes that provide a rational basis to engineer strains for improved biomass fermentation.

7.3.1 Genome Sequence

The *C. phytofermentans* ISDg ATCC 700394 complete genome has been sequenced by the US Department of Energy Joint Genome Institute. The genome has a 35.4% G + C content and consists of a single, circular 4 847 594-bp chromosome encoding 3902 protein coding sequences (CDS). Presently, 2481 (62.2%) of the CDS have a predicted function according to the JGI annotation. The genome also contains 89 non-coding RNAs including eight rRNA operons (16 S, 5 S, 23 S), a 6 S RNA, a tmRNA, and 61 tRNAs. Assuming G–U wobble in the codon–anti-codon pairing, the 61 tRNA can recognize all 61 possible codons. The genome also encodes 33 putative riboswitches that control processes such as amino acid and cobalamin biosynthesis, sodium balance, and perhaps biofilm formation.

The *C. phytofermentans* genome sequence is a basis for understanding the machinery used by this bacterium to ferment biomass.[29] The genome has 169 CAZys including 31 glycosyl transferases, 13 carbohydrate esterases, 9 polysaccharide lyases, and 115 glycoside hydrolases that are spread across 39 sequence families.[11] This abundance and diversity of CAZys enables this bacterium to grow on so many plant substrates. For example, *C. thermocellum* grows on cellulose, but not hemicellulose or pectin, and its genome encodes only 70 glycoside hydrolases in 23 families. Because *C. thermocellum* specializes in cellulose degradation, its CAZys include 16 glycoside hydrolase family 9 (GH9) cellulases. In contrast, *C. phytofermentans* has only one CAZy in GH9, which is required for cellulose degradation.[28] *C. Phytofermentans* CAZys also differ from other cellulolytic clostridia by lacking dockerin domains and the genome does not have a scaffoldin protein to assemble a cellulosome. Cellulolytic enzymes in *C. phytofermentans* are thus either freely secreted or anchored to the cell in a cellulosome-independent manner. The absence of a cellulosome will likely make *C. phytofermentans* cellulolytic enzymes particularly well-suited for over-expression and heterologous expression because they can function without binding a cellulosomal scaffold.

The ability of *C. phytofermentans* cells to uptake numerous substrates is highlighted by the genome encoding 137 genes for ATP-binding cassette (ABC-type) transporters. Along with the transporters, are 53 genes for extra-cellular-solute-binding proteins (ESBs), which are membrane-attached lipoproteins that capture substrates and pass them to the membrane transporter. Forty-seven of the 53 *C. phytofermentans* ESBs are involved in sugar transport and can be very highly expressed. The oligosaccharide-binding ESB Cphy2466 was the third most highly expressed protein in the cell during growth on cellulose.[14]

7.3.2 Gene Expression

We recently completed a proteome-wide analysis of *C. phytofermentans* to identify the hydrolytic and metabolic enzymes used to ferment different cellulosic substrates.[14] Protein quantification by mass spectrometry (LC-MS/MS) remains a challenging research area and we showed that a labeling method we call ''ReDi proteomics'' gives comprehensive, accurate, low-cost quantification of a microbial proteome and can discern extra-cellular proteins. ReDi labeling is advantageous to alternative approaches for stable isotope incorporation such as ^{15}N labeling and Stable Isotope Labeling by Amino acids in Cell culture (SILAC) in not requiring strains with specific amino acid auxotrophies or optimization of growth in synthetic media. Also, because ReDi isotope labels are added after peptides are isolated, supernatant and cellular proteins can be differentially labeled to identify extra-cellular proteins. Specifically, we applied ReDi proteomics to quantify differences in each protein in cellulosic cultures (hemicellulose and cellulose) relative to protein levels in glucose cultures. We also identified the extra-cellular proteome using ReDi protemics to measure the concentration differences of each protein between the supernatant and culture lysates. To understand the physiological consequences of changes to the proteome, we integrated the proteomics with detailed measurements of growth, fermentation, enzyme activities, and electron microscopy.

Changes in a total of 2567 proteins (65% of the predicted coding sequences) were quantified, including 357 previously hypothetical proteins. Growth on cellulosic substrates entailed numerous proteome changes. Comparing protein expression levels from glucose and hemicellulose cultures showed greater than two-fold differences for 20% of proteins; 49% proteins were expressed at two-fold different levels between glucose and cellulose cultures. As a control, fewer than 4% of proteins differed by two-fold levels when comparing two glucose cultures. Expression changes on hemicellulose were largely confined to pentose transport and assimilation, the pentose phosphate pathway, and CAZys, whereas expression changes on cellulose extended throughout metabolism. Many proteins were up-regulated on cellulose (CAZys, transporters, glycolytic proteins, NADH synthesis); others were repressed (flagella, fatty acid synthesis, housekeeping functions, DNA replication, transcription, translation). Also, the overall fraction of the proteome comprised of hypothetical proteins was much higher on cellulose, suggesting that many proteins of unknown function relate to cellulose metabolism.

Defining the set of secreted proteins ''the secretome'' was a high priority because enzymes to degrade insoluble biomass must be extra-cellular. We combined ReDi proteomics with localization prediction by PsortB v2.0[30] and SignalP3.0[31] to identify extra-cellular proteins and the cellular mechanisms used for export. The *C. phytofermentans* secretome consists of more than 300 proteins that function mostly in carbohydrate and protein degradation, cell surface and flagellar assembly, and transport. Furthermore, assigning CAZys to intra-cellular and extra-cellular proteins showed which steps in polysaccharide degradation occur inside the cell. Both substrates are initially

cleaved by multiple extra-cellular enzymes and transported into the cell as oligosaccharides before being catabolized to sugars, an adaptation that likely conserves energy and prevents sugars from being available to competing microbes.

C. phytofermentans expressed more than 100 CAZys and altered their stoichiometries to each substrate. The first step to metabolize hemicellulose is cleavage of the xylan backbone by extra-cellular GH10 and GH11 endo-xylanases. Cphy2108, the most highly expressed extra-cellular xylanase, is predicted to attach to the peptidoglycan wall with an LPXTG motif.[32] Cleavage of hemicellulose by a surface-bound xylanase may help to retain xylosaccharides in close proximity to the cell. In parallel with cleavage of the xylan backbone, additional extra- and intra-cellular enzymes remove the glucuronic and acetic acid side-chains, which are common in hemicellulose.[33] *C. phytofermentans* likely transport xylosaccharides into the cell before cleavage by two main intra-cellular exo-xylosidases: Cphy3009 cleaves xylosides from the non-reducing end of xylosaccharides and Cphy3207 acts on the reducing end.

Cellulose fibers are depolymerized by the concerted action of endo- and exo-cellulases. Endo-cellulases randomly cleave internal bonds in cellulose chains to disrupt the crystalline structure and expose individual polysaccharide chains. The most highly secreted endo-cellulase was Cphy3367, which is the only GH9 enzyme in *C. phytofermentans*. Cphy3367 is required for cellulose degradation[28] and solubilizes cellulose *in vitro*.[34] *C. phytofermentans* also highly up-regulated 3 GH5 endo-cellulases, which likely have accessory roles in cellulose cleavage. Exo-cellulases channel the freed cellulose chains through an active site tunnel to cleave two to four units from the ends of the exposed chains.[35] Cphy3368 is an exo-cellulase[36] that is likely expressed from the same mRNA as Cphy3367 and these two enzymes synergize to convert cellulose to cellodextrins and glucose.[34]

C. phytofermentans transports cellodextrins into the cell before splitting them into sugar monomers. Cellodextrin cleavage occurs primarily by three cellodextrin phosphorylases (Cphy3854, Cphy0430, Cphy1929) that remove terminal glucosides using phosphate as an attacking group. Phosphorolysis avoids ATP hydrolysis for phosphorylation of glucose during import, which is required for glucose transport by the *E. coli* phosphotransferase system. While the ATP savings correlate with the length of the transported cellodextrin, uptake and phosphorolytic cleavage of glucans can reduce cellular ATP requirements by 25%.[37]

Sugars are catabolized by the Embden–Meyerhof–Parnas glycolysis pathway using several non-standard enzymes that, similar to the cellulodextrin phosphorylases described above, highlight how *C. phytofermentans* is pressured to conserve ATP because of the reduced efficiency of fermentation relative to oxidative respiration. For example, ATP is likely conserved during glycolysis by using reversible, pyrophosphate (PPi)-dependent glycolytic enzymes such as PPi-dependent phosphofructokinase and pyruvate phosphate dikinase. *C. phytofermentans* lacks the Ppa phosphatase used by *E. coli* to cleave PPi and

instead uses it as a phosphate donor and source of a high-energy bond. PPi-Dependent glycolytic enzymes can increase glycolytic yield from 2 to 5 ATP38. In general, clostridia have low pyrophosphatase activities and high intracellular PPi levels,[39] supporting the general use of PPi as an energy-saving adaptation.

C. phytofermentans likely also conserves energy by using ferredoxin-dependent enzymes in the place of enzymes that use NADH. Pyruvate from glycolysis is converted to acetyl-CoA using a highly expressed pyruvate ferredoxin oxidoreductase (PFOR, Cphy3558) that oxidizes pyruvate and transfers electrons to ferredoxin. In contrast, *E. coli* and *B. subtilis* convert pyruvate to acetyl-CoA using NAD^+-reducing pyruvate dehydrogenase. *C. phytofermentans* also converts 2-oxoglutarate to succinyl-CoA using a ferredoxin-dependent 2-oxoglutarate oxidoreductase (Cphy3122-3) instead of NADH-dependent 2-oxoglutarate dehydrogenase. Reduced ferredoxin ($E'_0 \leq -420\,\text{mV}$) has a more negative redox potential than NADH ($E'_0 = -320\,\text{mV}$) and this 100 mV energy difference could be exploited for energy generation. Similar to *C. ljungdahlii*,[40] *C. phytofermentans* highly expresses an Rnf-type NADH ferredoxin oxidoreductase[29] that could use electrons from reduced ferredoxin to pump either sodium or protons across the cell membrane and subsequently deposit the electrons on NAD^+. The resulting electrochemical gradient could then be harnessed to form ATP by a membrane-bound ATPase.

Reducing equivalents in the form of NADH are consumed by reducing acetyl-CoA to ethanol with multiple, highly expressed alcohol dehydrogenase (ADH) enzymes: five Fe-dependent ADHs and one Zn-dependent ADH. The two Fe-dependent ADHs, Cphy3925 and Cphy1029, are among the most highly expressed proteins under all conditions. Cphy3925 shares 54% amino acid identity with *E. coli* AdhE, a bifunctional aldehyde/alcohol dehydrogenase that converts acetyl-CoA directly to ethanol. The other highly expressed ADH, Cphy1029, shares 35% amino acid identity with *Z. mobilis* AdhB, the primary ADH in this organism. This combination of ADH enzymes rapidly consumes excess reducing equivalents and facilitates the high ethanol yields. Another adaptation to consume excess reducing equivalents is the high expression of both monomeric and bifurcating [FeFe]–hydrogenases that re-oxidize ferredoxin and NADH. Monomeric hydrogenases oxidize only ferredoxin, but the bifurcating hydrogenases oxidize both ferredoxin and NADH in a 1 : 1 ratio,[41] thereby saving ferredoxin for use by the Rnf complex. All subunits for two multi-meric, bifurcating hydrogenases were highly expressed on all carbon sources, whereas a monomeric ferredoxin-dependent hydrogenase was expressed at a 10-fold lower level.

In addition to identifying the principle enzymes used to deconstruct and ferment cellulosic substrates (Figure 7.4), other metabolic adaptations likely indirectly enhance cellulosic fermentation. For example, tryptophan biosynthesis enzymes are up-regulated during growth on cellulose. While the tryptophan precursors leading up to chorismate are used for various compounds, the up-regulated enzymes make anthranilate (Cphy3847-8) and are

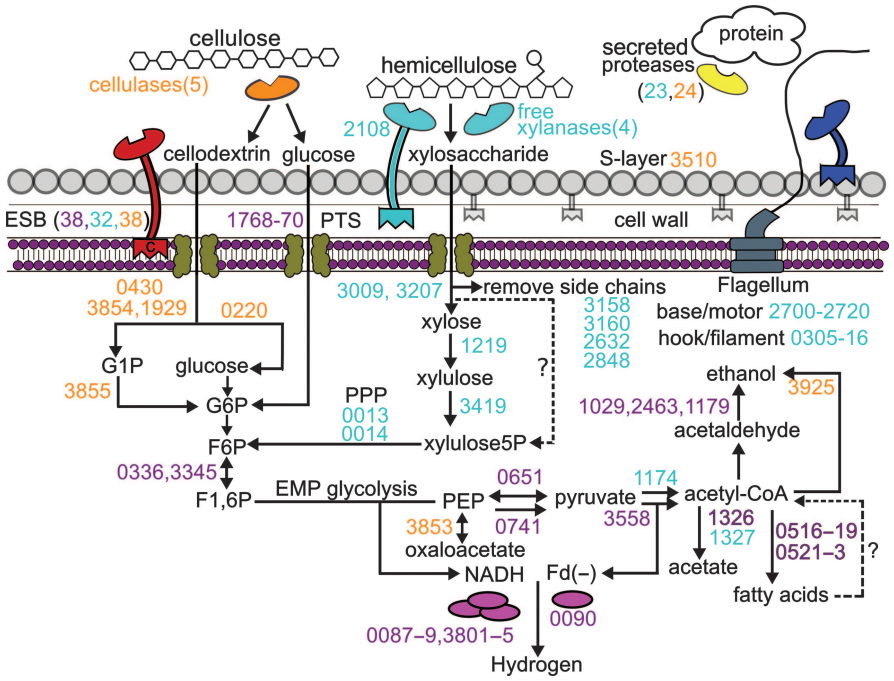


Figure 7.4 Model of the key secreted and intracellular enzymes for the degradation and fermentation of plant biomass. The numbers are NCBI protein annotations and are colored by the highest protein expression on glucose (purple), hemicellulose (turquoise), and cellulose (orange). The numbers in parentheses show the number of proteins of related function. Dashed arrows represent unknown reactions.
Image adapted from ref. 14.

involved in the final steps of tryptophan synthesis (Cphy3842-3, Cphy3845-6), suggesting the expression changes show a direct need for tryptophan. Increased production of tryptophan could increase ethanol tolerance by an unknown mechanism similar to yeast[42,43] and enable translation of CAZys with tryptophan-rich carbohydrate-binding modules (CBMs). CBMs in carbohydratase proteins enhance catalysis by keeping the hydrolytic domain in close association with its substrate[44] using a hydrophobic platform of tryptophan residues[45] that form van der Waals interactions with sugar rings.[46] Thirteen glycoside hydrolases with CBMs were up-regulated on cellulose and the concomitant high expression of tryptophan-synthesis proteins suggests that tryptophan may be an important requirement for cellulolysis.

Fatty acid synthesis was repressed during growth on cellulose relative to glucose, including enzymes for synthesis of malonyl-CoA (Cphy0519, Cphy0521-3) and malonyl-ACP (Cphy 0516), and for fatty acid elongation (Cphy0517. The greater expense required to harvest carbon from cellulose likely caused a metabolic shift to store less carbon and energy as fatty acids, which could complicate efforts to use microbes to produce fatty-acid-based

biofuels from cellulosic biomass.[47] Along with repression of fatty acid synthesis, we expected to see increased expression of β-oxidation enzymes to break down fatty acids into acetyl. However, KEGG and homology searches support that *C. phytofermentans* and other clostridia including *C. cellulolyticum* and *C. thermocellum* do not have any of the enzymes for fatty acid β-oxidation, making it unclear if or how these clostridia degrade fatty acids.

A novel means to express cellulolytic enzymes on the cell surface is suggested by the most highly expressed protein in the proteome, Cphy3510, a secreted protein that shares significant sequence similarity with the *Bacillus anthracis* S-layer protein, Sap.[48] The S-layer is a protective outer coat in some bacteria that is generally composed of a single protein that spontaneously assembles into a two-dimensional lattice.[49] Because S-layers of $\sim 5\times10^5$ protein molecules are required to cover a microbial cell,[50] the S-layer protein is often the most abundant protein in the proteome. Transmission electron micrographs of *C. phytofermentans* shows a surface layer exterior to the cell wall,[10,14] leading us to propose that *C. phytofermentans* is covered by a Cphy3510-based S-layer. The S-layer has been shown to anchor CAZys in *B. stearothermophilis*[51] and *Thermoanaerobacterium thermosulfurigenes*[52] using C-terminal repeats of about 50 amino acids. Thus, the identification of the S-layer protein in *C. phytofermentans* provides insight into how CAZys may be anchored to the cell surface and suggests a new strategy for engineering polypeptides for cell surface display by binding them to Cphy3510.

In addition to the genomic and proteomic data described above, we are analyzing metabolomic and transcriptomic data with the ultimate goal of integrating these datasets into a comprehensive molecular model of biomass fermentation. These datasets will together reveal genetic regulatory elements, the organization of protein networks, and the dynamics of metabolite pools during cellulosic fermentation. This molecular systems-level understanding of cell state will ultimately enable improved predictions about how *C. phytofermentans* should be genetically engineered for industrial biomass fermentation.

7.4 Genetic Engineering

Although cellulolytic clostridia have been studied for decades, a lack of methods for their genetic manipulation has hindered our ability to learn how they ferment biomass and to create robust strains for industrial bioconversion. Here we discuss progress and future opportunities to develop *C. phytofermentans* as a platform microbe for genetic engineering by focusing on three major prerequisites: (1) delivery of foreign DNA; (2) plasmid origins and resistance markers; and (3) chromosomal integration.

7.4.1 DNA Delivery

Transfer of foreign DNA to clostridia has been achieved by conjugation, electroporation, and natural transformation. We have shown that DNA can be

reliably transferred to *C. phytofermentans* by conjugal transfer from *E. coli*,[28] a method that also works in cellulolytic *C. cellulolyticum*[53] and in non-cellulolytic clostridia including *C. acetobutylicum*,[54] *C. perfringens*,[55] and *C. difficile*.[56] Specifically, we used the broad range RP4 conjugal apparatus encoded by pRK24,[57] which can transfer any plasmid that has the 760-bp RP4 conjugal origin of transfer (oriT) sequence.[58]

Following conjugal transfer, it is necessary to eliminate both *E. coli* donors and *C. phytofermentans* cells that did not receive plasmid DNA. As *C. phytofermentans* is naturally resistant to nalidixic acid and trimethoprim, these antibiotics can be used to kill *E. coli* donors. Alternatively, if plates are supplemented with X-gal and IPTG, *lacZ+ E. coli* strains form blue colonies whereas *C. phytofermentans* colonies are white. The formation of white *C. phytofermentans* colonies on X-gal/IPTG plates is surprising because it grows on lactose and has multiple putative β-galactosidases. Perhaps X-gal is either inefficiently transported into the cell or is not a substrate of *C. phytofermentans* enzymes. A third option to selectively remove *E. coli* from post-conjugation liquid cultures, is inoculation with an *E. coli* phage such as T7.[59] After selection, the absence of residual *E. coli* can be confirmed by plating an aliquot on solid LB medium and incubating aerobically at 37 °C overnight. Selection for *C. phytofermentans* transconjugants that received a resistance marker (see the following section) supports the conjugal delivery of DNA to *C. phytofermentans* with an efficiency of $\sim$1 transconjugant per 10^6 cells.

Although DNA delivery by electroporation has not yet been shown in *C. phytofermentans*, this method works in various clostridia including *C. acetobutylicum*,[60] *C. tyrobutyricum*,[61] *C. ljungdahlii*,[40] *C. paraputrificum*,[62] and the cellulolytic strains *C. cellulolyticum*[53] and *C. thermocellum*.[63] Electroporation is often inhibited in gram-positive bacteria such as clostridia by their thick peptidoglycan wall, but efficiencies can be improved by weakening the cell wall with muralytic enzymes,[64] glycine,[65] or isoniacin.[66] *C. phytofermentans* is an excellent candidate for electroporation because the genome does not appear to encode any DNA-restriction enzymes, which have plagued electroporation efforts in other clostridia such as *C. acetobutylicum*[67] and *C. perfringens*.[68]

An exciting possibility that could accelerate large-scale genome engineering in *C. phytofermentans* would be to find methods to transfer DNA into the cell by natural transformation. Natural transformation is based on genetic competence, the ability of a cell to uptake naked DNA from the environment.[69] In the first report of natural competence in clostridia, strains of *Thermoanaerobacterium* and *Thermoanaerobacter* were shown to be naturally competent with efficiencies from 1×10^{-3} to 1.9×10^{-6} transformants per CFU.[70] These thermophiles, which ferment xylan and certain sugars to a range of fermentation products, had previously been transformed by electroporation,[71] but now several strains are known to be competent in the early exponential phase in a process that depends upon a TP4 locus, *comEA*, *comEC*, and a *cinA recA* locus.

The genes for natural competence in these thermophilic clostridia all have homologs in *C. phytofermentans*, suggesting it may also be transformable if the

conditions for competence were found. *C. phytofermentans* has several genes for assembly of DNA-binding type IV pili and structures resembling these pili are visible in electron micrographs.[14] *C. phytofermentans* has gene homologs of both parts of the ComE DNA transporter: the membrane-bound dsDNA receptor ComEA (*cphy1953*) and the ComEC transmembrane channel (*cphy1957*). Furthermore, the *comEC* homologs are separated in the genome by a putative two-component system (c*phy1954*, *cphy1955*), suggesting competence may be regulated by a sensor kinase/response regulator pair similar to *B. subtilis*.[69] After the transforming DNA has been internalized, it must be delivered to the genome for recombination. In *B. subtilis*, there is a specifically positioned ssDNA binding apparatus that receives ssDNA from the DNA-uptake machinery and processes the DNA for recombination with chromosomal DNA.[72] *C. phytofermentans* has gene homologs of the *B. subtilis* proteins that mediate this process including: DprA, (*cphy2723*), SsbB (*cphy3773*, *cphy2987*, *cphy3483*), *CoiA* (*cphy0885*), RecA (*cphy2439*), and RecN (*cphy2499*). However, even if *C. phytofermentans* has a complete competence machinery, conditions for natural transformation may be abstruse. *Bacillus megaterium* has a *comE* locus that can complement the *B. subtilis* ComE transporter, but *B. megaterium* has not been found to be naturally competent.[73]

7.4.2 Antibiotic Resistance and Plasmid Origins

In parallel with methods for DNA delivery, we are working to identify antibiotic resistance genes and plasmid origins for clostridial shuttle and suicide vectors. The two most commonly used antibiotics used for clostridia are erythromycin and chloramphenicol. For selections in *C. phytofermentans*, we use a $200\,\mu g\,mL^{-1}$ erythromycin to select for the *erm* gene from *S. pneumoniae* Tn1545 that functions in both gram-negative and gram-positive bacteria.[74] Although we have not had problems with spontaneous resistance to erythromycin, a 1 : 1 combination of erythromycin and lincomycin could potentially give a stronger selection, as with *C. thermocellum*.[63] These antibiotics can be combined because lincomycin and erythromycin resistance are conferred by the same *mls* gene (macrolide–lincosamide–streptogramin).[75] Because the mutations that confer resistance to each antibiotic are independent, the probability that both mutations arise simultaneously is much lower. *C. phytofermentans* is also sensitive to $10\,\mu g\,mL^{-1}$ chloramphenicol and the chloramphenicol acetyltransferase gene *catP* functions as a selectable marker. If desired, thiamphenicol, a methyl-sulfonyl analog of chloramphenicol, can be used instead for selections with *catP*.[76] *C. phytofermentans* is, however, naturally resistant to kanamycin and streptomycin,[10] so these antibiotics should not be used for selection.

Genetic methods using both replicative and non-replicative plasmid origins have been used in clostridia. Plasmids bearing the *Enterococcus* pAMß1 origin replicate in *C. phytofermentans* during antibiotic selection, but are rapidly lost in the absence of selection.[28] Specifically, we found that 80% of cells can be cured of pAMß1 plasmids by diluting cultures 1 : 100 with a medium lacking

antibiotic and growing to late log phase through five serial transfers. The cultures that have been cured of pAMß1 plasmids regain erythromycin sensitivity, permitting further manipulations using the same plasmid and resistance marker. While conditionally replicating plasmids are ideal for many applications, a high-copy-number plasmid that replicates indefinitely without selection would be useful in large, industrial reactors where antibiotic addition is prohibitive. We are currently investigating plasmid origins that may allow stable, high-copy replication of plasmids in *C. phytofementans* without selection, including the pBP1 origin from *C. botulinum*, the pCB101 origin from *C. butyricum*[77] that replicates at high copy in *C. acetobutylicum*,[54] the pIM13 origin from *B. subtilis*,[78] and the pCD6 origin from *C. difficile*.[56]

7.4.3 Chromosomal Insertion

Targeted chromosomal changes in clostridia have traditionally been made by either single or double cross-over homologous recombination. Single cross-over homologous recombination has been used to insert a plasmid into the *pta* genes of *C. acetobutylicum*[79] and of *C. tyrobutyricum*.[61] Similarly, a double cross-over was used to delete the *pta*[80] and *cel48s*[81] genes from *C. thermocellum*. Because the frequencies of chromosomal insertion by homologous recombination initially appeared to be prohibitively low in *C. phytofermentans*, we developed a general system for targeted, chromosomal gene inactivation in *C. phytofermentans* based on group II introns, catalytic RNAs that self-splice into genomic DNA in a site-specific manner.[82]

The group II intron we developed to make targeted chromosomal insertions in *C. phytofermentans* is based on the *Lactococcus lactis* Ll.LtrB group II intron,[83] which is available from Sigma–Aldrich as the ''Targetron System''. It consists of a 0.9 kb Ll.LtrB-deltaORF intron flanked by short exon sequences and a downstream *ltrA* gene encoding a protein with endo-nuclease and reverse transcriptase activity.[84] Because the DNA-binding specificity of the group II intron is conferred by a short 13–16-bp sequence, the intron can be customized to integrate into a desired DNA target site by a simple two-step cross-over PCR. Group II introns can, in theory, function in any bacterial taxa into which plasmid DNA can be delivered because intron insertion does not require host-supplied factors. Furthermore, the insertion frequencies of group II introns into the genomic DNA of diverse bacteria are quite high: 0.1–22% in *E. coli*,[83] 37–100% in gram-positive *Staphylococcus aureus*,[85] and 2.5–100% in clostridia.[86]

When a strong *C. phytofermentans* promoter is used to drive expression of the group II intron, it inserts into the genome with such a high efficiency (often near 100%) that mutants can be easily isolated without selecting for integration. Because the intron insertions are stable in the absence of selection, the intron need not contain a resistance gene. Once the plasmid carrying the intron and *ltrA* gene is cured from *C. phytofermentans*, no antibiotic resistance genes remain. Multiple intron insertions can thus be made in the same strain without the need for independent resistance markers, which is particularly helpful because so few resistance markers are known to function in clostridia.

The elements to make our general system for targeted chromosomal insertions in *C. phytofermentans* are combined in the plasmid, pQint (Figure 7.5(a)). This plasmid has an RP4 conjugal origin of transfer, an

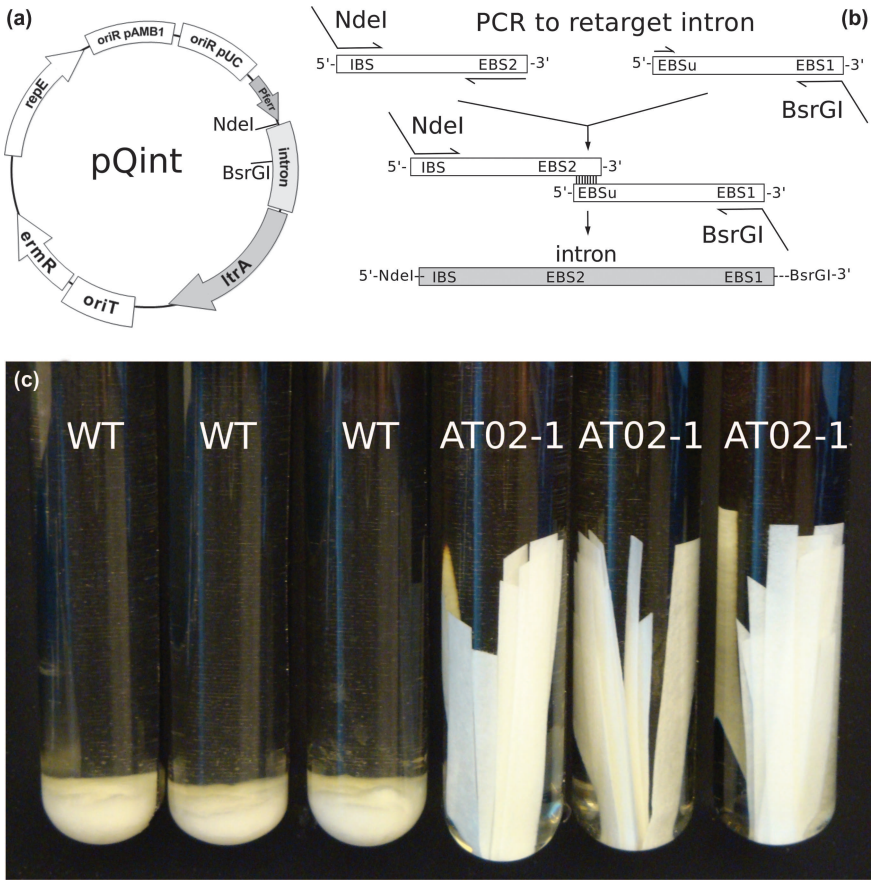


Figure 7.5 The genetic system to make targeted, chromosomal insertions in the *C. phytofermentans* chromsome. (a) Plasmid map of pQint. This plasmid has two origins of replication: a *Enterococcus* pAMß1 origin (for *C. phytofermentans*) and a pUC origin (for *E. coli*). The *C. phytofermentans* pyruvate ferrodoxin oxidoreductase (*cphy3558*) promoter (Pferr) drives the strong expression of the intron cassette and down-stream *ltrA* gene. The oriT facilitates conjugal transfer by an RP4-type conjugal apparatus. The *erm* gene from *S. pneumoniae* Tn1545 functions in *E. coli* and *C. phytofermentans*. (b) The intron cassette can be targeted to insert in nearly any desired DNA sequence by two-step, cross-over PCR. After PCR, the targeted intron is cloned into the NdeI and BsrGI sites of pQint. (c) Insertion of an intron into the gene encoding the sole family 9 hydrolase Cphy3367 results in a strain (AT02-1) that is unable to degrade cellulose. Cellulose degradation was measured as dry mass of cellulose remaining in culture. After four weeks, the cellulose strips in the wild-type tubes had broken down, while the strips in the AT02-1 cultures appeared unchanged. Images adapted from ref. 28.

erythromycin-resistance gene that functions in both gram-negative and gram-positive bacteria, origins of replication for *E. coli* (pUC) and for gram-positive bacteria (*Enterococcus* pAMß1), and a group II intron down-stream of the strong *cphy3558* promoter. This intron can be targeted to insert anywhere in the *C. phytofermentans* chromosome by customizing the targeting sequence using PCR (Figure 7.5(b)). In an initial study, we inserted a group II intron into *cphy3367*, the only family 9 glycoside hydrolase in *C. phytofermentans*.[28] Inactivation of *cphy3367* resulted in a strain (AT02-1) that grew normally on glucose, cellobiose and hemicellulose, but had lost the ability to degrade cellulose (Figure 7.5(c)). These findings reveal a central role played by Cphy3367 in cellulose degradation and show for the first time that a single gene can be required for cellulolysis.[87] We are currently applying this system to dissect the set of cellulolytic enzymes in *C. phytofermentans* in order to identify key genes for the degradation of cellulosic biomass and to develop strains that ferment biomass to specific fermentation products.

7.5 Future Directions

This chapter has highlighted how *C. phytofermentans* is an excellent model system for cellulosic biofuel research because of its unique ability to directly ferment diverse plant polysaccharides to ethanol and its ease of cultivation in the laboratory. Our recent work on *C. phytofermentans* has focused on optimizing growth conditions, high-throughput genomics and proteomics to characterize metabolic networks, and genetic engineering methods. These studies both provide insight into how microbes ferment biomass and reveal unmet challenges that need to be addressed.

The results described here show promising targets for metabolic engineering to make *C. phytofermentans* strains that are more cost-effective for industrial fermentation. For example, *C. phytofermentans* is similar to other cellulolytic clostridia in lacking the ability to make several B-vitamins (Figure 7.3), which are expensive to add to large industrial reactors. Genomic analysis suggests that some of these auxotrophies could be overcome by addition of a few genes, either by putting the genes on a replicating plasmid or by integrating them into the genome using a group II intron. Specifically, these B-vitamin auxotophies might be overcome by adding the following genes: *nadABC* for nicotinamide (B3), *B. subtilis pdxST* for pyridoxal-5-phosphate (B6), and *B subtilis pabABC* and *sul* to convert chorismate to dihydropteroate for folate (B9).

Proteomic identification of highly expressed and secreted enzymes allows us to reduce the 169 *C. phytofermentans* CAZys into a set of high-priority targets to engineer for improved biomass degradation. For example, Cphy3367 is the second-most highly expressed CAZy on cellulose and it is required for cellulose degradation; supporting it would be an ideal gene both to over-express in *C. phytofermentans* and to express in other microbes to improve cellulolysis. Similarly, Cphy2105 and Cphy2108 appear to be the main secreted hemicellulases, suggesting they would be good candidates to manipulate for the degradation of hemicellulose-rich substrates. Along with the secreted enzymes

to degrade polysaccharides *in situ*, the principle intra-cellular cellodextrin phosphorylases (Cphy3854, Cphy0430, Cphy1929) and xylosidases (Cphy3207, Cphy3009) should also be over-expressed to cleave oligosaccharides into sugars that can be readily metabolized.

Even though ethanol fermentation by *C. phytofermentans* is quite efficient, metabolic engineering to streamline fermentation could further improve ethanol yields. We observed that the ethanol-to-acetate ratio is much lower on hemicellulose (Figure 7.2(g)) than on other substrates, especially cellulose (Figure 7.2(h)). The proteomics of hemicellulose cultures showed elevated expression of acetate kinase (Cphy1327), supporting that inactivation of this enzyme would improve ethanol yields similar to other clostridia.[79,61,80] The growth rate on hemicellulose (doubling time 1.16 h, Figure 7.2(c)) is also the fastest among the substrates we have tested. It is unknown if high acetate production, which yields ATP by substrate-level phosphorylation, directly enables the rapid growth rate on hemicellulose, but it would be interesting to test if inactivating acetate kinase cripples the growth on hemicellulose.

In addition to altering the expression of native enzymes, methods such as directed evolution and domain shuffling can be used to improve the activities of these enzymes under industrially relevant conditions. For example, the *C. phytofermentans* hydrolase Cphy3202 was amplified by error-prone PCR to create a library of random 20 000 variants, which was then screened to identify a mutant with a two-fold increased half-life at 60 °C relative to the wild-type.[88] Domain shuffling has also been applied to three fungal class II cellobiohydrolases to create themostable chimeric proteins.[89] As an alternative to enzyme screening, a technique called ''chemical complementation'' allows the selection of highly active cellulases *in vivo*.[90] This method is a reverse yeast three-hybrid assay in which cleavage of a tetrasaccharide substrate decreases expression of a toxic URA3 reporter, thereby selecting for cells with high-activity cellulases. While this assay requires a synthetic saccharide substrate with methotrexate and dexamethasone handles that bears little resemblance to plant biomass, it can analyze more enzyme variants ($\sim 10^8$) than is feasible by *in vivo* enzyme screening and was applied to domain-shuffled cellulases to find a mutant with a six-fold increase in K_{cat}/K_m over the parent enzyme.[90] Chemical complementation could be applied to *C. phytofermentans* CAZys to develop variants with enhanced activities.

Cost-effective biofuel production requires high product titers, which is challenging because ethanol and other prospective biofuels are toxic to microbes at high concentrations. *C. phytofermetans* grows normally in ethanol concentrations up to 4% (40 g l^{-1}). Even though this ethanol tolerance is higher than other clostridia such as *C. thermocellum* (1–2% ethanol tolerance),[91] industrial biofuel production will likely require strains that tolerate even higher ethanol concentrations. Numerous studies have shown that ethanol tolerance can be improved by random mutagenesis[92] or serial transfer.[91,93,94] Two promising recombinant DNA methods have also recently been used to express genes conferring increased biofuel tolerance in clostridia. First, transforming *C. acetobutylicum* with a random genomic library of its own DNA was used to

identify genes that improve growth during butanol stress when at increased copy number.[95] Second, heterologous expression in *E. coli* of 43 efflux pumps from various sequenced bacterial genomes was shown to reduce toxicity by exporting the biofuels from the cell.[96] Both these methods are based on heterologous expression of plasmid DNA, which is now tractable in *C. phytofermentans* using the genetic methods described in this chapter.

Plant biomass is an abundant, low-cost feedstock that could enable large-scale production of fuels and biochemicals. In order for cellulosic biofuels to replace a significant fraction of our current petroleum needs, improvements are needed in many areas such as the genetic engineering of bioenergy crops, improved cultivation methods for these plants, and microbes to efficiently convert plant matter into useful chemicals. This chapter has described recent advances in developing *C. phytofermentans* as a model microbe to study biomass fermentation. We hope that these studies on *C. phytofermentans* will help to enable future advances to produce biofuels that both reduce our need for fossil fuels and facilitate sustainable economic development.

References

1. S. B. Leschine, *Annu. Rev. Microbiol.*, 1995, **49**, 399.
2. D. Perlack, L. L. Wright, A. F. Turhollow, R. L. Graham, B. L. Stokes and D. C. Erbach, *A Joint Study by the US Department of Energy and the US Department of Agriculture*, 2005, U.S. Department of Energy (Oak Ridge TN, USA).
3. C. Somerville, *Science*, 2006, **312**, 1277.
4. J. Houghton, S. Weathervax and J. Ferrell, *US Department of Energy Roadmap*, Department of Energy, 2006, U.S. Department of Energy (Oak Ridge TN, USA).
5. J. Doran-Peterson, D. M. Cook and S. K. Brandon, *Plant J.*, 2008, **54**, 582.
6. L. R. Lynd, M. S. Laser, D. Bransby, B. E. Dale, B. Davison, R. Hamilton, M. Himmel, M. Keller, J. D. McMillan, J. Sheehan and C. E. Wyman, *Nat. Biotechnol.*, 2008, **26**, 169.
7. L. R. Lynd, P. J. Weimer, W. H. van Zyl and I. S. Pretorius, *Microbiol. Mol. Biol. Rev.*, 2002, **66**, 506.
8. L. R. Lynd, W. H. van Zyl, J. E. McBride and M. Laser, *Curr. Opin. Biotechnol.*, 2005, **16**, 577.
9. M. Jin, V. Balan, C. Gunawan and B. E. Dale, *Biotechnol. Bioeng.*, 2011, **108**, 1290.
10. T. A. Warnick, B. A. Methé and S. B. Leschine, *Int. J. Syst. Evol. Microbiol.*, 2002, **52**, 1155.
11. B. L. Cantarel, P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard and B. Henrissat, *Nucleic Acids Res.*, 2009, **37**, D233.
12. E. A. Johnson, A. Madia and A. L. Demain, *Appl. Environ. Microbiol.*, 1981, **41**, 1060.
13. Y. Lu, Y.-H. P. Zhang and L. R. Lynd, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 16165.

14. A. C. Tolonen, W. Haas, A. C. Chilaka, J. Aach, S. P. Gygi and G. M. Church, *Mol. Syst. Biol.*, 2011, **7**, 461.
15. M. Morrison and J. Miron, *FEMS Microbiol. Lett.*, 2000, **185**, 109.
16. M. Muir, L. Williams and T. Ferenci, *J. Bacteriol.*, 1985, **163**, 1237.
17. Y.-H. P. Zhang and L. R. Lynd, *J. Bacteriol.*, 2005, **187**, 99.
18. S. Harvey, A. Chambers and P. Zhang, presented at the 26th Army Science Conference, Orlando, 2008.
19. S. Winogradsky, *Arch. Sci. Biol.*, 1895, **3**, 297.
20. J. S. Chen, J. Toth and M. Kasap, *J. Ind. Microbiol. Biotechnol.*, 2001, **27**, 281.
21. S. Leschine, Susan and T. Warnick, *US Pat.*, 7682811 (B2), 2010.
22. E. Guedon, M. Desvaux, S. Payot and H. Petitdemange, *Microbiology*, 1999, **145**, 1831.
23. P. Hebbeln, D. A. Rodionov, A. Alfandega and T. Eitinger, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 2909.
24. B. R. Belitsky, *J. Bacteriol.*, 2004, **186**, 1191.
25. M. Strohmeier, T. Raschle, J. Mazurkiewicz, K. Rippe, I. Sinning, T. B. Fitzpatrick and I. Tews, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 19284.
26. P. Stover and V. Schirch, *Trends Biochem. Sci.*, 1993, **18**, 102.
27. S. Ogwang, H. T. Nguyen, M. Sherman, S. Bajaksouzian, M. R. Jacobs, W. H. Boom, G.-F. Zhang and L. Nguyen, *J. Biol. Chem.*, 2011, **286**, 15377.
28. A. C. Tolonen, A. C. Chilaka and G. M. Church, *Mol. Microbiol.*, 2009, **74**, 1300.
29. Petit *et al.*, PLoS One (2013). Accepted.
30. J. L. Gardy, M. R. Laird, F. Chen, S. Rey, C. J. Walsh, M. Ester and F. S. L. Brinkman, *Bioinformatics*, 2005, **21**, 617.
31. J. D. Bendtsen, H. Nielsen, G. von Heijne and S. Brunak, *J. Mol. Biol.*, 2004, **340**, 783.
32. M. Zhou, J. Boekhorst, C. Francke and R. J. Siezen, *BMC Bioinf.*, 2008, **9**, 173.
33. D. Shallom and Y. Shoham, *Curr. Opin. Microbiol.*, 2003, **6**, 219.
34. X.-Z. Zhang, N. Sathitsuksanoh and Y.-H. P. Zhang, *Bioresour. Technol.*, 2010, **101**, 5534.
35. C. Divne, J. Ståhlberg, T. Reinikainen, L. Ruohonen, G. Pettersson, J. K. Knowles, T. T. Teeri and T. A. Jones, *Science*, 1994, **265**, 524.
36. X.-Z. Zhang, Z. Zhang, Z. Zhu, N. Sathitsuksanoh, Y. Yang and Y.-H. P. Zhang, *Appl. Microbiol. Biotechnol.*, 2010, **86**, 525.
37. Y.-H. P. Zhang and L. R. Lynd, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 7321.
38. C. H. Slamovits and P. J. Keeling, *Eukaryotic Cell*, 2006, **5**, 148.
39. J. K. Heinonen and H. L. Drake, *FEMS Microbiol. Lett.*, 1988, **52**, 205.
40. M. Köpke, C. Held, S. Hujer, H. Liesegang, A. Wiezer, A. Wollherr, A. Ehrenreich, W. Liebl, G. Gottschalk and P. Dürre, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 13087.

41. G. J. Schut and M. W. W. Adams, *J. Bacteriol.*, 2009, **191**, 4451.
42. T. Hirasawa, K. Yoshikawa, Y. Nakakura, K. Nagahisa, C. Furusawa, Y. Katakura, H. Shimizu and S. Shioya, *J. Biotechnol.*, 2007, **131**, 34.
43. X. Q. Zhao and F. W. Bai, *J. Biotechnol.*, 2009, **144**, 23.
44. P. Tomme, H. Van Tilbeurgh, G. Pettersson, J. Van Damme, J. Vandekerckhove, J. Knowles, T. Teeri and M. Claeyssens, *Eur. J. Biochem.*, 1988, **170**, 575.
45. T. Ponyi, L. Szabó, T. Nagy, L. Orosz, P. J. Simpson, M. P. Williamson and H. J. Gilbert, *Biochemistry*, 2000, **39**, 985.
46. J. Lehtiö, J. Sugiyama, M. Gustavsson, L. Fransson, M. Linder and T. T. Teeri, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 484.
47. E. J. Steen, Y. Kang, G. Bokinsky, Z. Hu, A. Schirmer, A. McClure, S. B. Del Cardayre and J. D. Keasling, *Nature*, 2010, **463**, 559.
48. I. Etienne-Toumelin, J. C. Sirard, E. Duflot, M. Mock and A. Fouet, *J. Bacteriol.*, 1995, **177**, 614.
49. U. B. Sleytr and P. Messner, *Annu. Rev. Microbiol.*, 1983, **37**, 311.
50. U. B. Sleytr and P. Messner, *J. Bacteriol.*, 1988, **170**, 2891.
51. E. Egelseer, I. Schocher, M. Sára and U. B. Sleytr, *J. Bacteriol.*, 1995, **177**, 1444.
52. M. Matuschek, G. Burchhardt, K. Sahm and H. Bahl, *J. Bacteriol.*, 1994, **176**, 3295.
53. K. C. Jennert, C. Tardif, D. I. Young and M. Young, *Microbiology*, 2000, **146**, 3071.
54. D. R. Williams, D. I. Young and M. Young, *J. Gen. Microbiol.*, 1990, **136**, 819.
55. D. Lyras and J. I. Rood, *Plasmid*, 1998, **39**, 160.
56. D. Purdy, T. A. T. O'Keeffe, M. Elmore, M. Herbert, A. McLeod, M. Bokori-Brown, A. Ostrowski and N. P. Minton, *Mol. Microbiol.*, 2002, **46**, 439.
57. R. Meyer, D. Figurski and D. R. Helinski, *Mol. Gen. Genet.*, 1977, **152**, 129.
58. D. G. Guiney and E. Yakobson, *Proc. Natl. Acad. Sci. U. S. A.*, 1983, **80**, 3595.
59. A. C. Tolonen, G. B. Liszt and W. R. Hess, *Appl. Environ. Microbiol.*, 2006, **72**, 7607.
60. S. Nakotte, S. Schaffer, M. Böhringer and P. Dürre, *Appl. Microbiol. Biotechnol.*, 1998, **50**, 564.
61. Y. Zhu, X. Liu and S.-T. Yang, *Biotechnol. Bioeng.*, 2005, **90**, 154.
62. K. Sakka, M. Kawase, D. Baba, K. Morimoto, S. Karita, T. Kimura and K. Ohmiya, *J. Biosci. Bioeng.*, 2003, **96**, 304.
63. M. V. Tyurin, S. G. Desai and L. R. Lynd, *Appl. Environ. Microbiol.*, 2004, **70**, 883.
64. P. T. Scott and J. I. Rood, *Gene*, 1989, **82**, 327.
65. H. Holo and I. F. Nes, *Appl. Environ. Microbiol.*, 1989, **55**, 3119.
66. M. V. Tyurin, C. R. Sullivan and L. R. Lynd, *Appl. Environ. Microbiol.*, 2005, **71**, 8069.

67. L. D. Mermelstein and E. T. Papoutsakis, *Appl. Environ. Microbiol.*, 1993, **59**, 1077.
68. C. K. Chen, C. M. Boucle and H. P. Blaschek, *FEMS Microbiol. Lett.*, 1996, **140**, 185.
69. D. Dubnau, *Microbiol. Rev.*, 1991, **55**, 395.
70. A. J. Shaw, D. A. Hogsett and L. R. Lynd, *Appl. Environ. Microbiol.*, 2010, **76**, 4713.
71. V. Mai and J. Wiegel, *Appl. Environ. Microbiol.*, 2000, **66**, 4817.
72. D. Kidane and P. L. Graumann, *Cell*, 2005, **122**, 73.
73. M. Lammers, H. Nahrstedt and F. Meinhardt, *J. Basic Microbiol.*, 2004, **44**, 451.
74. P. Trieu-Cuot, C. Carlier, C. Poyart-Salmeron and P. Courvalin, *Gene*, 1991, **102**, 99.
75. M. Monod, C. Denoya and D. Dubnau, *J. Bacteriol.*, 1986, **167**, 138.
76. J. R. O'Connor, D. Lyras, K. A. Farrow, V. Adams, D. R. Powell, J. Hinds, J. K. Cheung and J. I. Rood, *Mol. Microbiol.*, 2006, **61**, 1335.
77. J. G. Morris and N. P. Minton, *J. Gen. Microbiol.*, 1981, **127**, 325.
78. H. Azeddoug, J. Hubert and G. Reysset, *J. Gen. Microbiol.*, 1992, **138**, 1371.
79. E. M. Green, Z. L. Boynton, L. M. Harris, F. B. Rudolph, E. T. Papoutsakis and G. N. Bennett, *Microbiology*, 1996, **142**, 2079.
80. 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, *Appl. Environ. Microbiol.*, 2010, **76**, 6591.
81. 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, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 17727.
82. A. M. Lambowitz and S. Zimmerly, *Annu. Rev. Genet.*, 2004, **38**, 1.
83. M. Karberg, H. Guo, J. Zhong, R. Coon, J. Perutka and A. M. Lambowitz, *Nat. Biotechnol.*, 2001, **19**, 1162.
84. H. Guo, M. Karberg, M. Long, J. P. Jones 3rd, B. Sullenger and A. M. Lambowitz, *Science*, 2000, **289**, 452.
85. J. Yao, J. Zhong, Y. Fang, E. Geisinger, R. P. Novick and A. M. Lambowitz, *RNA*, 2006, **12**, 1271.
86. J. T. Heap, O. J. Pennington, S. T. Cartman, G. P. Carter and N. P. Minton, *J. Microbiol. Methods*, 2007, **70**, 452.
87. D. B. Wilson, *Mol. Microbiol.*, 2009, **74**, 1287.
88. W. Liu, X.-Z. Zhang, Z. Zhang and Y.-H. P. Zhang, *Appl. Environ. Microbiol.*, 2010, **76**, 4914.
89. P. Heinzelman, C. D. Snow, I. Wu, C. Nguyen, A. Villalobos, S. Govindarajan, J. Minshull and F. H. Arnold, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 5610.
90. P. Peralta-Yahya, B. T. Carter, H. Lin, H. Tao and V. W. Cornish, *J. Am. Chem. Soc.*, 2008, **130**, 17446.
91. T. I. Williams, J. C. Combs, B. C. Lynn and H. J. Strobel, *Appl. Microbiol. Biotechnol.*, 2007, **74**, 422.

92. P. Tailliez, H. Girard, J. Millet and P. Beguin, *Appl. Environ. Microbiol.*, 1989, **55**, 207.
93. A. A. Herrero and R. F. Gomez, *Appl. Environ. Microbiol.*, 1980, **40**, 571.
94. X. Shao, B. Raman, M. Zhu, J. R. Mielenz, S. D. Brown, A. M. Guss and L. R. Lynd, *Appl. Microbiol. Biotechnol.*, 2011, **92**, 641.
95. J. R. Borden and E. T. Papoutsakis, *Appl. Environ. Microbiol.*, 2007, **73**, 3061.
96. M. J. Dunlop, Z. Y. Dossani, H. L. Szmidt, H. C. Chu, T. S. Lee, J. D. Keasling, M. Z. Hadi and A. Mukhopadhyay, *Mol. Syst. Biol.*, 2011, **7**, 487.

Aim 1: Rational and evolutionary genome engineering to optimize microbial fermentation

Genome engineering is a powerful approach to construct bacterial strains with improved phenotypes. In particular, we are interested in engineering the genomes of Clostridia, a group of bacteria composed of species with important roles in soil and intestinal ecosystems that have numerous applications in industrial microbiology. For example, efficient production of biochemicals from lignocellulose by Clostridia will require creation of strains with enhanced substrate assimilation, streamlined carbon flux, and increased fermentation product titres. To this end, one of our research aims is to develop methods for genome engineering of Clostridia and apply them to optimize their metabolisms. The two approaches we are using can broadly be classified as rational and evolutionary genome engineering. Below is a summary of our current research interests in these areas, followed by two published studies highlighting our progress to date.

Rational genome engineering involves the deliberate modification of the genome, based on a deep understanding of the biology and genetics of the organism in question. This approach involves using detailed knowledge of the genome's structure and function to make targeted, predictable changes. We have developed a suite of methods for the rational genome engineering of Clostridia using *C. phytofermentans* as a model (Table 1). These methods enable efficient introduction of foreign DNA into the cell, expression of heterologous genes, and repression or inactivation of chromosomal genes. We have applied these methods to make fundamental insights into *C. phytofermentans* biology, including how it degrades complex polymers, uptakes sugars, and produces fermentation products. Below, is a description of two focus areas of our rational genome engineering efforts: base editing using dCas9 and dCas12a and construction and testing of pathways for production of non-natural molecules of interest.

Table 1 Progress to date on the rational genome engineering of *C. phytofermentans.*

Method	Application	References
Delivery of plasmid DNA by electroporation and by conjugal transfer from *E. coli*	Basis for genetic transformation	[1] [2]
Reporter gene expression based on fluorescence using a flavin-based fluorescent protein (FbFP) or luminescence (NanoLuc)	Identification of a library of promoters spanning 100-fold expression levels	[3] [2]
A set of plasmid replicons and resistance markers for heterologous gene expression	Expression of a synthetic ethanol formation pathway	[4]
Targeted gene inactivation using designed group II introns (Targetrons)	Identification of genes for polysaccharide degradation and transport of liberated sugars	[1] [5] [6]
Insertion and deletion of multi-gene fragments using Targetrons plus Cre-*lox*	Deletion of a 39 kb genomic prophage	[3]
Repression of fermentation genes by CRISPRi using LbCas12a	Reduction of acetate formation	[2]
Assembly and expression of multi-gene pathways	Synthetic production of butanol and pigments	In progress
C-to-A base editing using dCas9 and dLbCas12a for scarless, targeted gene inactivation	The role of defense systems in regulating plasmid transformability	In progress
Multiplex C-to-A based editing using dCas9	Inactivation of genes for unwanted fermentation products	In progress

Evolutionary genome engineering, sometimes called adaptive laboratory evolution, is an approach that enables creation of strains with improved phenotypes even when the underlying genetic elements are not known. This approach is particularly well-suited to improving complex, multigenic traits and can be applied to species for which methods for rational genetic engineering have not yet been developed. Moreover, evolutionary engineering can be applied in cases where Genetically Modified Organisms (GMOs) are not permitted due to regulations or or consumer preferences. The main limitation of evolutionary approaches is that the desired trait for improvement must be linked to fitness (i.e. growth rate). As such, traits linked to substrate utilization or resistance can easily be improved, whereas those linked to metabolite production require creative solutions to link them to fitness [7].

A platform at the Genoscope-CEA, called the Genemat system, facilitates evolutionary genome engineering of microorganisms by long-term growth selection [8]. Cells with growth advantages are selected in the Genemat using either of two growth regimes: turbidostat or medium-swap. In turbidostat mode, cells with increased growth rate are selected using a single medium. In medium swap mode, cultures grow in a chemostat regime with dilutions of stressing medium if density exceeds a threshold and with relaxing medium otherwise. Our cultivation strategy is to acclimate cells to a stressing medium (i.e. increased inhibitor concentration) using the medium-swap regime. Once oscillations are replaced by stable growth in the stressing medium, the regime is changed to a turbidostat to select for faster growth. When the growth rate stabilizes in the stressing medium in the turbidostat, the selection is increased and cultures are again acclimated by medium-swap.

The Genemat system can be applied for trait improvement of any microbial species, provided a few practical considerations are met. The species needs to be cultivable with a reasonable growth rate in liquid medium. All elements of the medium must be soluble such that they do not clog the tubing. Finally, as long-term growth selections require many liters of medium, all elements of the medium must be relatively inexpensive. We have previously evolved strains of C*. phytofermentans* that resist lignin phenolics using the Genemat system [9]. Below, is a description of two focus areas of our evolutionary genome engineering efforts: butanol resistance in *C. phytofermentans* and *C. cellulolyticum* and acetate utilization in *Yarrowia lipolytica*.

Aim 1a: Multiplex gene inactivation by base editing

We recently demonstrated a C-to-T base editing method in *C. phytofermentans* that consists of a a Cas9D10A nickase, a cytidine deaminase (rAPOBEC1), and uracil DNA glycosylase inhibitor (UGI). We applied this system to inactivate the *pyrE* gene in *C. phytofermentans* by introducing a premature stop codon, resulting in a strain that is resistant to the toxic 5-Fluoroorotic acid (5-FOA). We also developed an algorithm for *in silico* analysis of the base editing capabilities of base editors with Cas9 (NGG) and Cas12a (TTTV, TNTN) protospacer adjacent motif (PAM) preferences, revealing that a

toolbox including both dCas9 and dCas12a base editors would significantly expand the targeting scope for gene inactivation in the genome. Currently, we are establishing the baseline efficiency of dLbCas12a-mediated base editing in *C. phytofermentans* with the goal of enabling both systems.

As a first step to applying base editing-mediated gene inactivation, we will study the role of bacterial defense systems in regulating transformability of *C. phytofermentans*. Defense systems are mechanisms that bacteria use to protect themselves from various threats such as viral infections and attacks from other microorganisms. *C. phytofermentans* encodes a putative defense system which could have an anti-plasmid function. These genes (*cphy3888-90*) are homologous to MksBEF, which acts as a complex molecular machine to regulate plasmid transmission in *C. glutamicum* [10]. We will apply base editing-mediated gene editing to inactivate *cphy2888-90* and examine their roles in regulating horizontal gene transfer, as measured by plasmid transfer efficiency of each mutant strain relative to WT.

Aim 1b: Pathway engineering for novel fermentation products

To facilitate the delivery and testing of synthetic metabolic pathways in *C. phytofermentans*, we modified the set of pQmod plasmids routinely used to transform *C. phytofermentans* to be compatible with the MoClo (Modular Cloning) system for combinatorial pathway assembly [11]. MoClo is for efficient cloning and re-ordering of genetic parts by Golden Gate assembly, enabling the optimization of pathway design by modulating the promoters, ribosome binding sites, and gene orders. Using this MoClo-based cloning system, we can assemble multiple variants of pathways for the biosynthesis pathways of interest. In particular, we will apply the MoClo system to study the potential of *C. phytofermentans* to be developed as a platform for microbial production of pigments.

Pigments have a huge variety of chemical structures and are used as food colorants, dyes in the textile or cosmetic industry, or for specific biological activities such as antioxidants, antibacterials, or anticancer agents. In particular, we are focusing on

violacein, which has applications as dyes and also has antifungal, nematocidal, anti-protozon, anticancer and immunomodulatory activities [12]. To permit biosynthesis of violacein, a five gene pathway consisting of *vioA-E* from *C. vaccinii* for production of violacein from two molecules of L-tryptophan will be PCR amplified from *C. vaccinii*. The *vioA-E* pathway will be assembled as a single operon using 5 different gene orders and fused to *C. phytofermentans* promoters of three different strengths, yielding 15 variants. The pathway variants will be cloned, transformed into *C. phytofermentans,* and pigment production will be measured colorimetrically. Once a gene order and promoter combination has been identified that maximizes pigment production in standard glucose-based growth medium, the pathway will be transferred into the *C. phytofermentans* genome using group II introns and Cre-lox, and pigment production will be measured by HPLC.

Aim 1c. Evolution of inhibitor resistance in Clostridia

C. phytofermentans and *C. cellulolyticum* are two Clostridial species that efficiently metabolize polysaccharide components of lignocellulose to acetate and ethanol. One of our research goals is to engineer strains of these species that produce other fermentation products such as n-butanol, which is an attractive fuel and solvent that can catalytically convert butanol to jet and diesel fuel at high yield [13] Unfortunately, butanol is highly toxic to cells [14] and inhibits growth at less than 1% (v/v) for most microbial species, including native butanol producers [15]. Butanol toxicity is a complex phenotype involving multiple cellular processes. It disrupts membrane stability by fluidizing the phospholipid bilayer, resulting in ion leakage and reduced electrochemical potential [16]. To date, we have evolved cultures of *C. phytofermentans* and *C. cellulolyticum* cultures to resist elevated butanol concentrations in the Genemat platform (Fig 1). We have sequenced the genomes of clones isolated along the evolutionary trajectory and are using comparative sequencing-based approaches to identify candidate mutations responsible for the butanol resistance phenotypes.

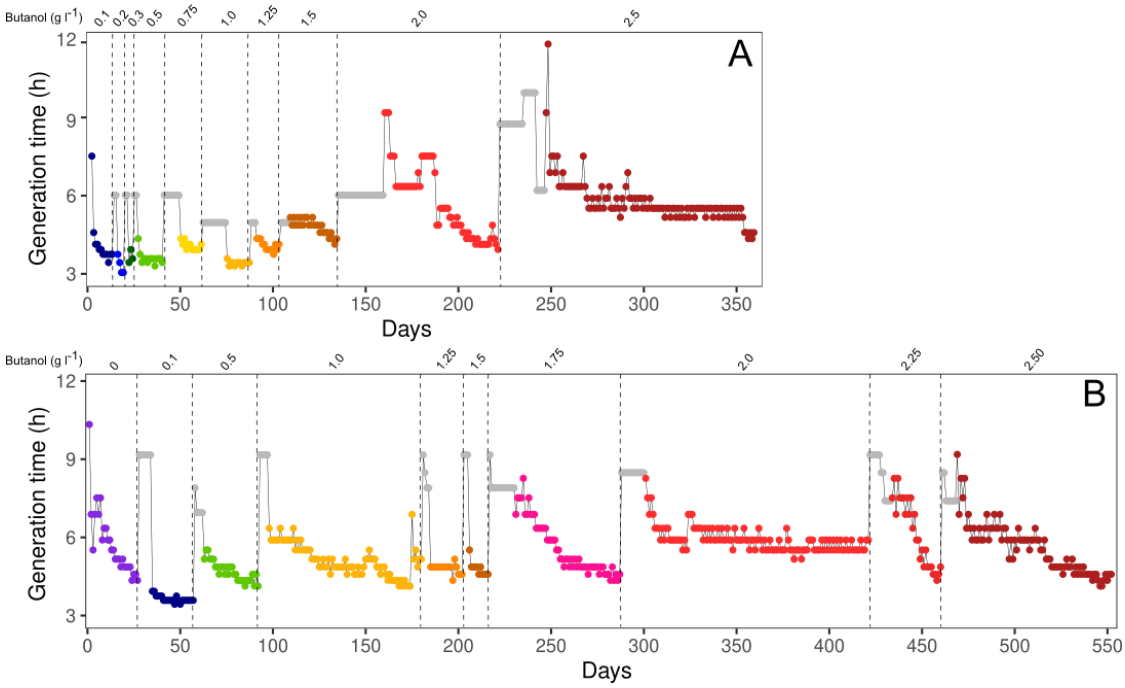


Fig 1 Adaptation of (A) *C. phytofermentans* or (B) *C. cellulolyticum* to growth with increasing levels of butanol present in the medium. Plots show generation time (hours) of cultures growing either in medium-swap mode with fixed generation time (gray) or in turbostat mode (other colors). Butanol concentrations in g l^{-1} are shown above plots.

Aim 1d. Evolution of *Yarrowia lipolytica* acetate utilization

Y*arrowia lipolytica*, an oleaginous yeast, is a microbial chassis for the production of alkanes [17], terpenes [18], and organic acids [19]. Most studies in *Y. lipolytica* use glucose or glycerol as a carbon source. However, acetate is a less expensive molecule that is abundant in industrial waste streams. *Y. lipolytica* has enzymes to assimilate acetate to fatty acids through the sequential activities of acetyl-CoA synthase, acetyl-CoA carboxylase, and fatty acid synthase, but growth is poor on acetate relative to glucose. Previously, growth of *Y. lipolytica* on acetate has been improved by over-expression of these enzymes [20]. However, acetate toxicity remains problematic.

In this project, we are adapting *Y. lipolytica* to growth in medium containing increasing acetate concentrations using the Genemat platform. To date, we have evolved two *Y. lipolytica* cultures to grow on 50 g l^{-1} acetate at rates similar to the WT strain growing on glucose (Fig 2). We will continue to increase the acetate

concentrations in these cultures until stable growth becomes infeasible. We will then select clones for genome sequencing to identify mutations associated with enhanced acetate utilization. The outcome of this project will be to isolate and characterize *Y. lipolytica* clones with enhanced growth on acetate and to identify mutations that can be re-engineered into other *Y. lipolytica* strains to enhance acetate assimilation.

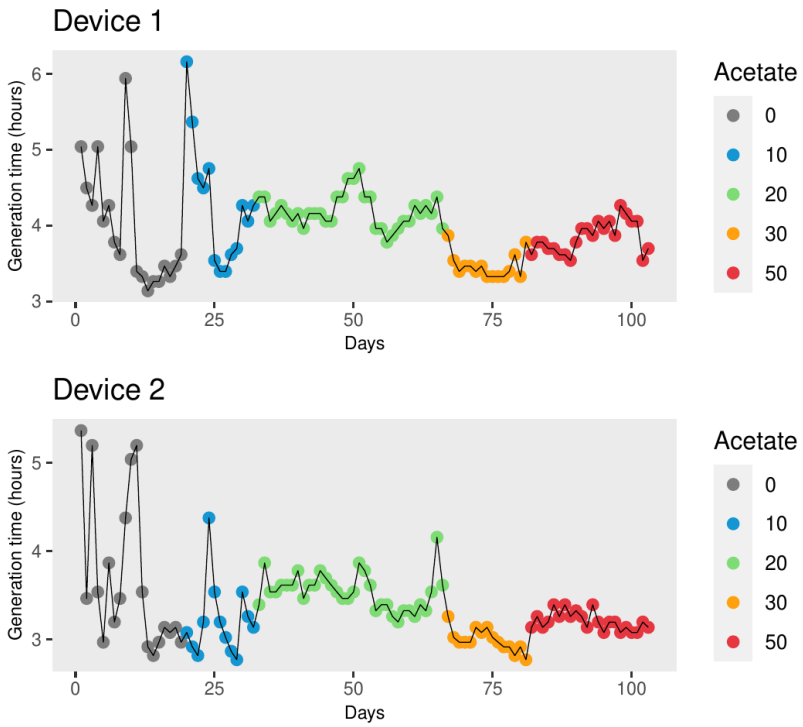


Fig 2 Generation times of *Y. lipolytica* W29 cultures in the GM3 with increasing acetate concentrations. Plot shows culture in Genemat 19, device 1 (top) and Genemat 19, device 2 (bottom). Colors show acetate concentrations in g l^{-1}. Data points with zero acetate (gray) correspond to the initial acclimatization period with 10 g l^{-1} glucose.

References

1. Tolonen AC, Chilaka AC, Church GM. Targeted gene inactivation in *Clostridium phytofermentans* shows that cellulose degradation requires the family 9 hydrolase Cphy3367. Molecular Microbiology. 2009;74: 1300–1313. doi:10.1111/j.1365-2958.2009.06890.x
2. Rostain W, Zaplana T, Boutard M, Baum C, Tabuteau S, Sanitha M, et al. Tuning of Gene Expression in Clostridium phytofermentans Using Synthetic Promoters and CRISPRi. ACS Synth Biol. 2022;11: 4077–4088. doi:10.1021/acssynbio.2c00385
3. Cerisy T, Rostain W, Chhun A, Boutard M, Salanoubat M, Tolonen AC. A Targetron-Recombinase System for Large-Scale Genome Engineering of Clostridia. mSphere. 2019;4. doi:10.1128/mSphere.00710-19
4. Tolonen AC, Zuroff TR, Ramya M, Boutard M, Cerisy T, Curtis WR. Physiology, Genomics, and Pathway Engineering of an Ethanol-Tolerant Strain of Clostridium phytofermentans.

Appl Environ Microbiol. 2015;81: 5440–5448. doi:10.1128/AEM.00619-15
5. Tolonen AC, Cerisy T, El-Sayyed H, Boutard M, Salanoubat M, Church GM. Fungal lysis by a soil bacterium fermenting cellulose. Environ Microbiol. 2015;17: 2618–2627. doi:10.1111/1462-2920.12495
6. Cerisy T, Iglesias A, Rostain W, Boutard M, Pelle C, Perret A, et al. ABC Transporters Required for Hexose Uptake by Clostridium phytofermentans. J Bacteriol. 2019;201. doi:10.1128/JB.00241-19
7. Konstantinidis D, Pereira F, Geissen E-M, Grkovska K, Kafkia E, Jouhten P, et al. Adaptive laboratory evolution of microbial co-cultures for improved metabolite secretion. Molecular Systems Biology. 2021;17: e10189. doi:10.15252/msb.202010189
8. Mutzel R, Marliere P. Method and Device for Selecting Accelerated Proliferation of Living Cells in Suspension. WO/2000/034433, 2000. Available: https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2000034433
9. Cerisy T, Souterre T, Torres-Romero I, Boutard M, Dubois I, Patrouix J, et al. Evolution of a biomass-fermenting bacterium to resist lignin phenolics. Appl Environ Microbiol. 2017. doi:10.1128/AEM.00289-17
10. Böhm K, Giacomelli G, Schmidt A, Imhof A, Koszul R, Marbouty M, et al. Chromosome organization by a conserved condensin-ParB system in the actinobacterium Corynebacterium glutamicum. Nat Commun. 2020;11: 1485. doi:10.1038/s41467-020-15238-4
11. Iverson SV, Haddock TL, Beal J, Densmore DM. CIDAR MoClo: Improved MoClo Assembly Standard and New E. coli Part Library Enable Rapid Combinatorial Design for Synthetic and Traditional Biology. ACS Synth Biol. 2016;5: 99–103. doi:10.1021/acssynbio.5b00124
12. Choi SY, Lim S, Yoon K, Lee JI, Mitchell RJ. Biotechnological Activities and Applications of Bacterial Pigments Violacein and Prodigiosin. J Biol Eng. 2021;15: 10. doi:10.1186/s13036-021-00262-9
13. Anbarasan P, Baer ZC, Sreekumar S, Gross E, Binder JB, Blanch HW, et al. Integration of chemical catalysis with extractive fermentation to produce fuels. Nature. 2012;491: 235–239. doi:10.1038/nature11594
14. Huffer S, Clark ME, Ning JC, Blanch HW, Clark DS. Role of alcohols in growth, lipid composition, and membrane fluidity of yeasts, bacteria, and archaea. Appl Environ Microbiol. 2011;77: 6400–6408. doi:10.1128/AEM.00694-11
15. Jones DT, Woods DR. Acetone-butanol fermentation revisited. Microbiol Rev. 1986;50: 484–524.
16. Baer SH, Blaschek HP, Smith TL. Effect of Butanol Challenge and Temperature on Lipid Composition and Membrane Fluidity of Butanol-Tolerant Clostridium acetobutylicum. Appl Environ Microbiol. 1987;53: 2854–2861.
17. Bruder S, Moldenhauer EJ, Lemke RD, Ledesma-Amaro R, Kabisch J. Drop-in biofuel production using fatty acid photodecarboxylase from Chlorella variabilis in the oleaginous yeast Yarrowia lipolytica. Biotechnol Biofuels. 2019;12: 202. doi:10.1186/s13068-019-1542-4
18. Liu L, Markham K, Blazeck J, Zhou N, Leon D, Otoupal P, et al. Surveying the lipogenesis landscape in Yarrowia lipolytica through understanding the function of a Mga2p regulatory protein mutant. Metab Eng. 2015;31: 102–111. doi:10.1016/j.ymben.2015.07.004
19. Rzechonek DA, Dobrowolski A, Rymowicz W, Mirończuk AM. Aseptic production of citric and isocitric acid from crude glycerol by genetically modified Yarrowia lipolytica. Bioresour Technol. 2019;271: 340–344. doi:10.1016/j.biortech.2018.09.118
20. Increased Lipid Production in Yarrowia lipolytica from Acetate through Metabolic Engineering and Cosubstrate Fermentation | ACS Synthetic Biology. [cited 3 Jul 2023]. Available: https://pubs.acs.org/doi/10.1021/acssynbio.1c00405

ACS Synthetic Biology

pubs.acs.org/synthbio Research Article

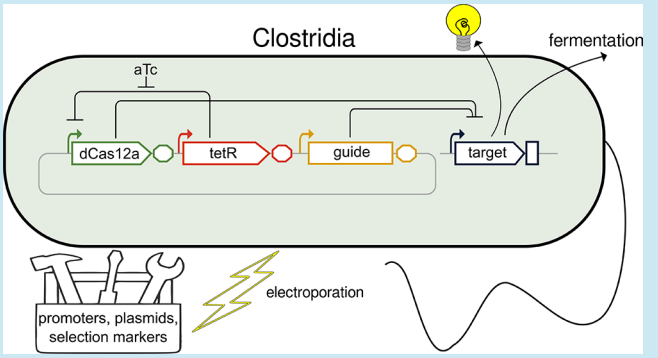
Tuning of Gene Expression in *Clostridium phytofermentans* Using Synthetic Promoters and CRISPRi

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

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

ACCESS | Metrics & More | Article Recommendations | Supporting Information

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



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

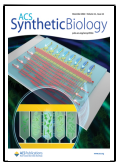
INTRODUCTION

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

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

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

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



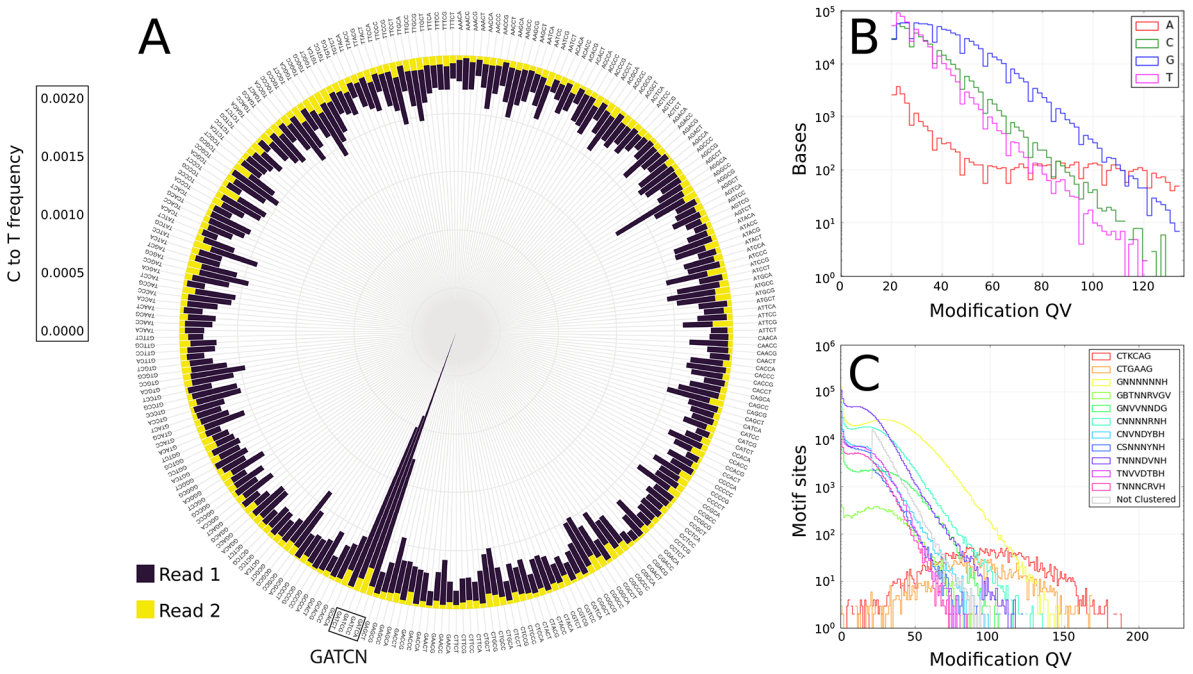


Figure 1. Base modification detection in the *C. phytofermentans* genome by (A) RIMS-seq sequencing and (B, C) SMRT sequencing. (A) RIMS-seq data analysis showing frequency of C-to-T mutations on Read 1 (blue bars) and Read 2 (yellow bars), indicating m5C methylation at the 5′-GATC-3′ motif ($p = 1.23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

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

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

RESULTS

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

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

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

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

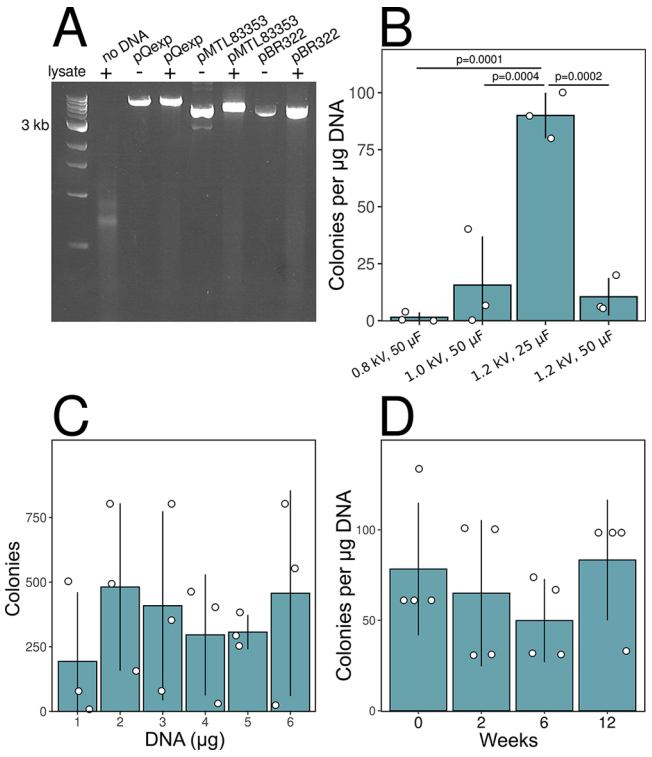


Figure 2. Factors affecting efficacy of *C. phytofermentans* electroporation. (A) Treatment of linearized plasmid DNA with *C. phytofermentans* culture lysate. Three plasmids (pQexp, pMTL83353, and pBR322) isolated from methylation-deficient *E. coli* ER2796 were linearized with *Nhe*I, incubated with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

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

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

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

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

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

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

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

Table 1. Plasmids and Strains Used in This Study

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

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

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

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

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

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

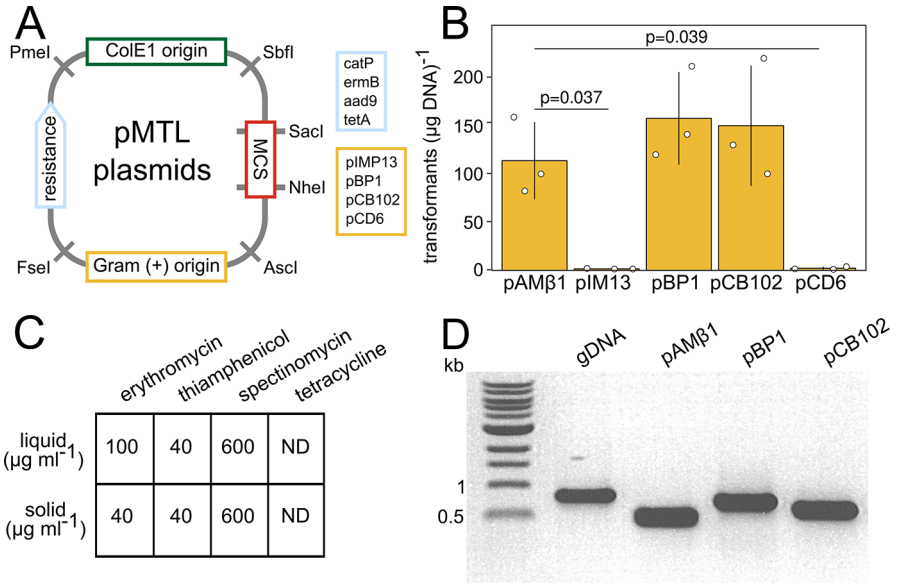


Figure 3. Plasmid replicons and antibiotic resistance genes that function in *C. phytofermentans*. (A) Structure of modular pMTL plasmids showing the MCS, Gram-positive replicon, antibiotic resistance marker, and Gram-negative ColE1 replicon. Restriction enzyme sites between modular units are shown along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

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

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

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

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

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

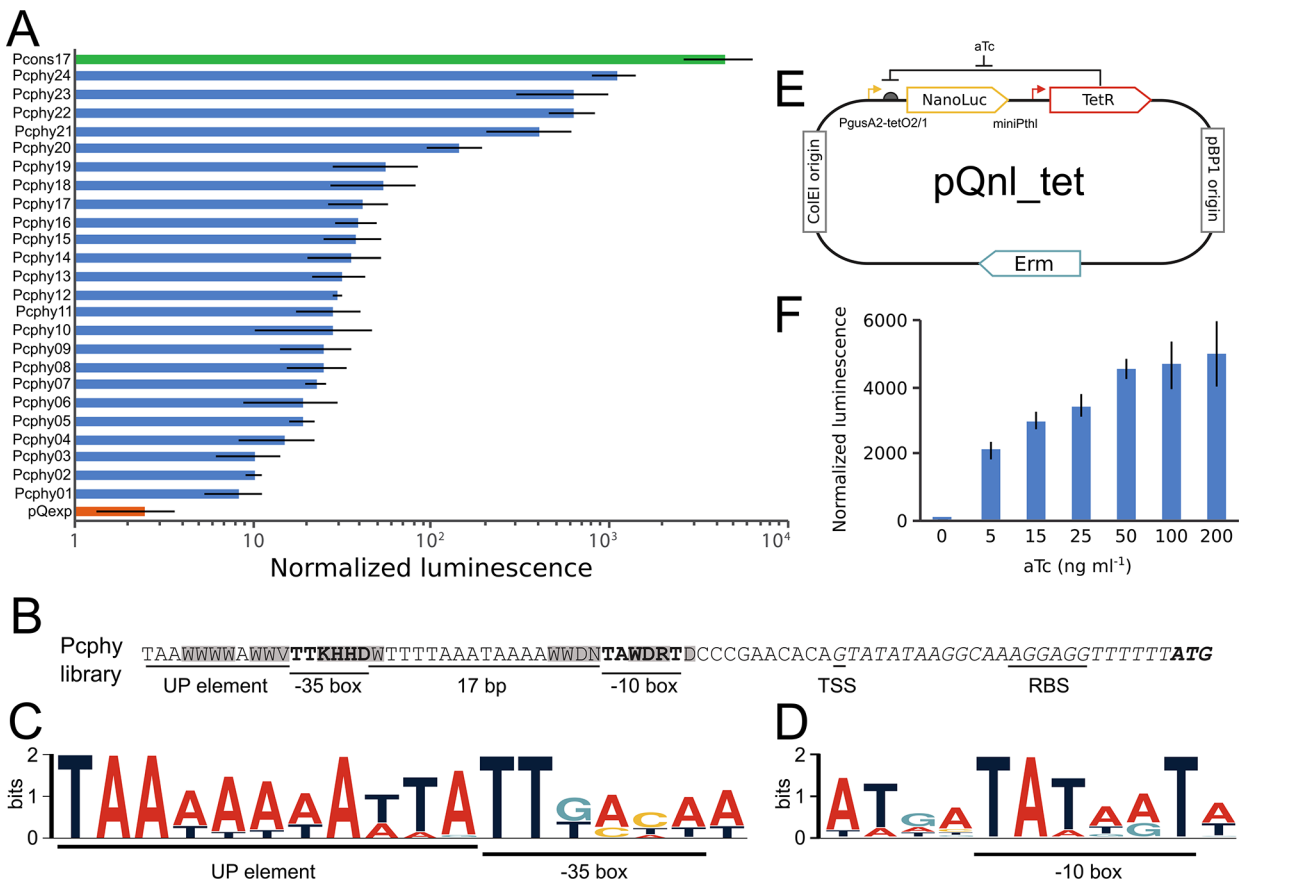


Figure 4. Modulation of gene expression in *C. phytofermentans* using constitutive and inducible promoters. (A) Promoter strengths of Pcons17, Pcphy1−24, and minus-NanoLuc control (pQexp) measured as normalized luminescence. (B) Nucleotide sequence showing degenerate positions used to build the Pcphy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

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

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

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

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

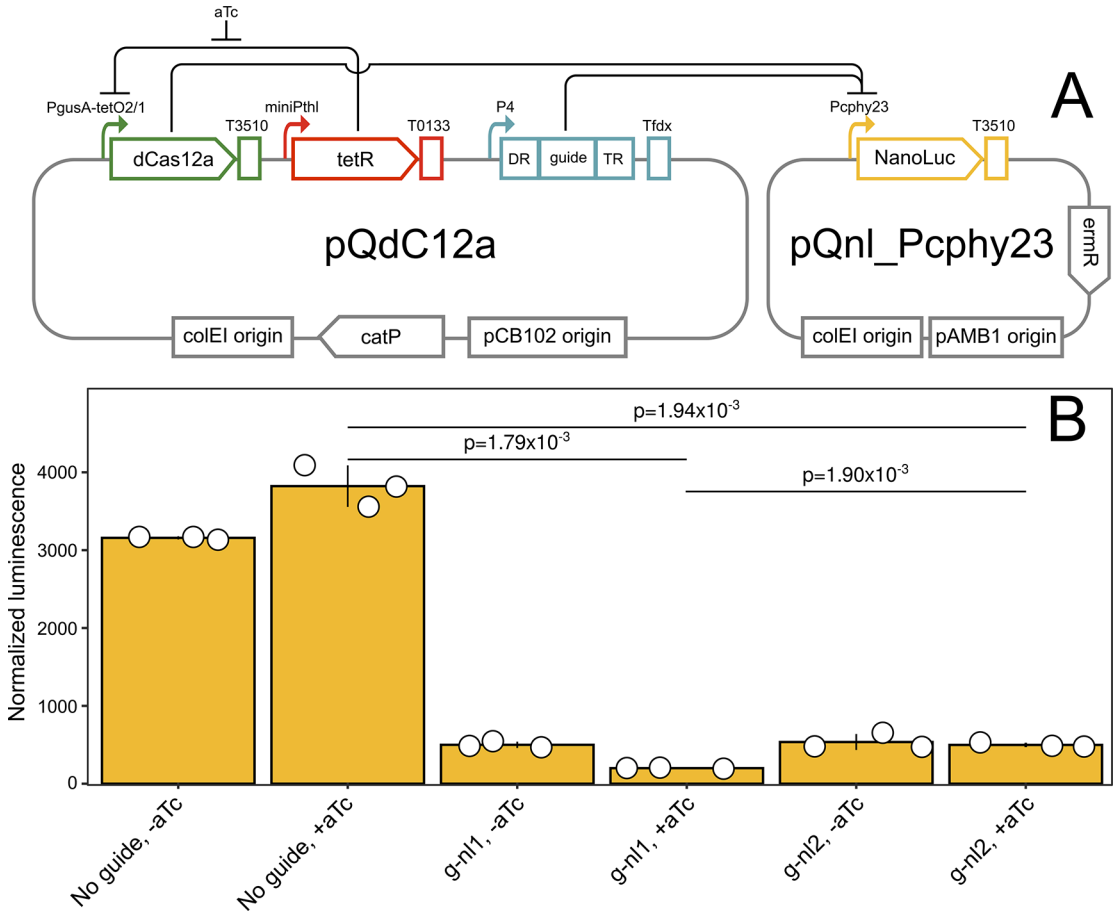


Figure 5. CRISPRi repression of reporter gene expression in *C. phytofermentans*. (A) Two-plasmid system demonstrating dCas12a repression of the NanoLuc reporter. pQdC12a has a tet-repressible dCas12a and gRNA cassette; pQnl_Pcphy23 has the Pcphy23 promoter driving NanoLuc expression. (B) Normalized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

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

DISCUSSION

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

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

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

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

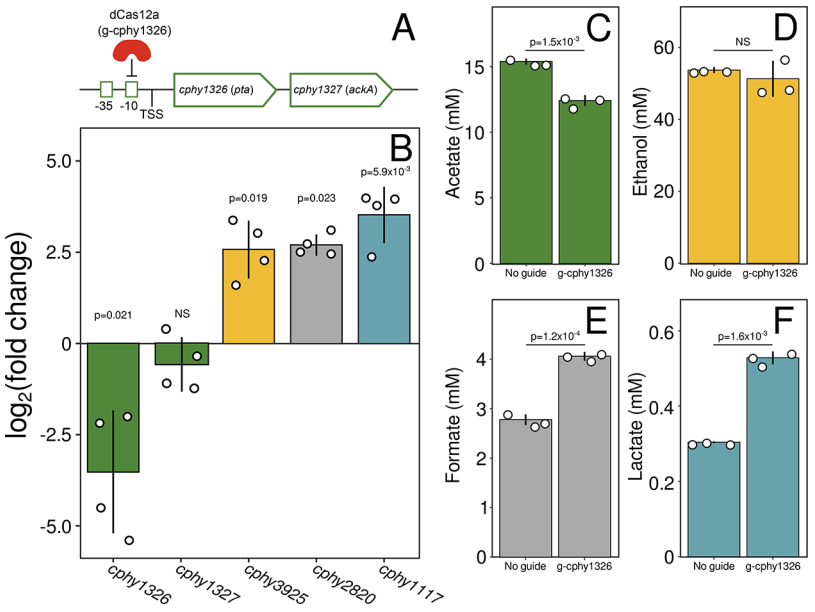


Figure 6. CRISPRi repression of fermentation gene expression in *C. phytofermentans*. (A) dCas12a was targeted to repress transcription of the acetate biosynthesis operon by binding the promoter −10 box upstream of the *cphy1326* (*pta*) gene using guide g-cphy1326. (B) Transcription of fermentation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

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

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

METHODS AND MATERIALS

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

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

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

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

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

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

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

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

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

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

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

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

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

growth to late log phase before measurement of luciferase expression.

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

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

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

■ ASSOCIATED CONTENT

Supporting Information

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

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

■ AUTHOR INFORMATION

Corresponding Author

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

Authors

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

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

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

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

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

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

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

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

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

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

Author Contributions

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

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

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

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

REFERENCES

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

(23) Baum, C.; Lin, Y.-C.; Fomenkov, A.; Anton, B. P.; Chen, L.; Yan, B.; Evans, T. C.; Roberts, R. J.; Tolonen, A. C.; Ettwiller, L. Rapid identification of methylase specificity (RIMS-seq) jointly identifies methylated motifs and generates shotgun sequencing of bacterial genomes. *Nucleic Acids Res.* **2021**, *49*, No. e113.

(24) Flusberg, B. A.; Webster, D. R.; Lee, J. H.; Travers, K. J.; Olivares, E. C.; Clark, T. A.; Korlach, J.; Turner, S. W. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* **2010**, *7*, 461−465.

(25) Li, T.; Zhang, X.; Luo, F.; Wu, F.-X.; Wang, J. MultiMotif-Maker: A Multi-Thread Tool for Identifying DNA Methylation Motifs from Pacbio Reads. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **2020**, *17*, 220−225.

(26) Anton, B. P.; Mongodin, E. F.; Agrawal, S.; Fomenkov, A.; Byrd, D. R.; Roberts, R. J.; Raleigh, E. A. Complete Genome Sequence of ER2796, a DNA Methyltransferase-Deficient Strain of *Escherichia coli* K-12. *PLoS One* **2015**, *10*, No. e0127446.

(27) Tolonen, A. C.; Chilaka, A. C.; Church, G. M. Targeted gene inactivation in *Clostridium phytofermentans* shows that cellulose degradation requires the family 9 hydrolase Cphy3367. *Mol. Microbiol.* **2009**, *74*, 1300−1313.

(28) Tolonen, A. C.; Cerisy, T.; El-Sayyed, H.; Boutard, M.; Salanoubat, M.; Church, G. M. Fungal lysis by a soil bacterium fermenting cellulose. *Environ. Microbiol.* **2015**, *17*, 2618−2627.

(29) Tolonen, A. C.; Zuroff, T. R.; Ramya, M.; Boutard, M.; Cerisy, T.; Curtis, W. R. Physiology, Genomics, and Pathway Engineering of an Ethanol-Tolerant Strain of *Clostridium phytofermentans*. *Appl. Environ. Microbiol.* **2015**, *81*, 5440−5448.

(30) Cerisy, T.; Iglesias, A.; Rostain, W.; Boutard, M.; Pelle, C.; Perret, A.; Salanoubat, M.; Fierobe, H.-P.; Tolonen, A. C. ABC Transporters Required for Hexose Uptake by *Clostridium phytofermentans*. *J. Bacteriol.* **2019**, *201* (15), No. e00241-19.

(31) Cerisy, T.; Rostain, W.; Chhun, A.; Boutard, M.; Salanoubat, M.; Tolonen, A. C. A Targetron-Recombinase System for Large-Scale Genome Engineering of Clostridia. *mSphere* **2019**, *4* (6), e00710-19.

(32) Bhattacharjee, D.; Sorg, J. A. Factors and Conditions That Impact Electroporation of *Clostridioides difficile* Strains. *mSphere* **2020**, *5* (2), No. e00941-19.

(33) Hammes, W.; Schleifer, K. H.; Kandler, O. Mode of action of glycine on the biosynthesis of peptidoglycan. *J. Bacteriol.* **1973**, *116*, 1029−1053.

(34) Sloan, J.; McMurry, L. M.; Lyras, D.; Levy, S. B.; Rood, J. I. The Clostridium perfringens Tet P determinant comprises two overlapping genes: tetA(P), which mediates active tetracycline efflux, and tetB(P), which is related to the ribosomal protection family of tetracycline-resistance determinants. *Mol. Microbiol.* **1994**, *11*, 403−415.

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

(35) Potapov, V.; Ong, J. L.; Kucera, R. B.; Langhorst, B. W.; Bilotti, K.; Pryor, J. M.; Cantor, E. J.; Canton, B.; Knight, T. F.; Evans, T. C.; Lohman, G. J. S. Comprehensive Profiling of Four Base Overhang Ligation Fidelity by T4 DNA Ligase and Application to DNA Assembly. *ACS Synth. Biol.* **2018**, *7*, 2665−2674.
(36) Boutard, M.; Ettwiller, L.; Cerisy, T.; Alberti, A.; Labadie, K.; Salanoubat, M.; Schildkraut, I.; Tolonen, A. C. Global repositioning of transcription start sites in a plant-fermenting bacterium. *Nat. Commun.* **2016**, *7*, 13783.
(37) de Jong, A.; Pietersma, H.; Cordes, M.; Kuipers, O. P.; Kok, J. PePPER: a webserver for prediction of prokaryote promoter elements and regulons. *BMC Genomics* **2012**, *13*, 299.
(38) Xu, T.; Li, Y.; Shi, Z.; Hemme, C. L.; Li, Y.; Zhu, Y.; Van Nostrand, J. D.; He, Z.; Zhou, J. Efficient Genome Editing in *Clostridium cellulolyticum* via CRISPR-Cas9 Nickase. *Appl. Environ. Microbiol.* **2015**, *81*, 4423−4431.
(39) Salis, H. M.; Mirsky, E. A.; Voigt, C. A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **2009**, *27*, 946−950.
(40) Mazo-Vargas, A.; Park, H.; Aydin, M.; Buchler, N. E. Measuring fast gene dynamics in single cells with time-lapse luminescence microscopy. *Mol. Biol. Cell* **2014**, *25*, 3699−3708.
(41) Ross, W.; Gosink, K. K.; Salomon, J.; Igarashi, K.; Zou, C.; Ishihama, A.; Severinov, K.; Gourse, R. L. A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science* **1993**, *262*, 1407−1413.
(42) Graves, M. C.; Rabinowitz, J. C. *In vivo* and *in vitro* transcription of the *Clostridium pasteurianum* ferredoxin gene. Evidence for "extended" promoter elements in gram-positive organisms. *J. Biol. Chem.* **1986**, *261*, 11409−11415.
(43) Gao, L.; Cox, D. B. T.; Yan, W. X.; Manteiga, J. C.; Schneider, M. W.; Yamano, T.; Nishimasu, H.; Nureki, O.; Crosetto, N.; Zhang, F. Engineered Cpf1 variants with altered PAM specificities. *Nat. Biotechnol.* **2017**, *35*, 789−792.
(44) Wasels, F.; Jean-Marie, J.; Collas, F.; López-Contreras, A. M.; Lopes Ferreira, N. A two-plasmid inducible CRISPR/Cas9 genome editing tool for *Clostridium acetobutylicum*. *J. Microbiol Methods* **2017**, *140*, 5−11.
(45) Hong, W.; Zhang, J.; Cui, G.; Wang, L.; Wang, Y. Multiplexed CRISPR-Cpf1-Mediated Genome Editing in *Clostridium difficile* toward the Understanding of Pathogenesis of C. difficile Infection. *ACS Synth. Biol.* **2018**, *7*, 1588−1600.
(46) Liao, C.; Slotkowski, R. A.; Achmedov, T.; Beisel, C. L. The *Francisella novicida* Cas12a is sensitive to the structure downstream of the terminal repeat in CRISPR arrays. *RNA Biol.* **2019**, *16*, 404−412.
(47) Zhang, X.; Wang, J.; Cheng, Q.; Zheng, X.; Zhao, G.; Wang, J. Multiplex gene regulation by CRISPR-ddCpf1. *Cell Discovery* **2017**, *3*, 17018.
(48) Boutard, M.; Cerisy, T.; Nogue, P.-Y.; Alberti, A.; Weissenbach, J.; Salanoubat, M.; Tolonen, A. C. Functional diversity of carbohydrate-active enzymes enabling a bacterium to ferment plant biomass. *PLoS Genet.* **2014**, *10*, No. e1004773.
(49) Mermelstein, L. D.; Welker, N. E.; Bennett, G. N.; Papoutsakis, E. T. Expression of cloned homologous fermentative genes in *Clostridium acetobutylicum* ATCC 824. *Bio/Technology* **1992**, *10*, 190−195.
(50) Ao, X.; Yao, Y.; Li, T.; Yang, T.-T.; Dong, X.; Zheng, Z.-T.; Chen, G.-Q.; Wu, Q.; Guo, Y. A Multiplex Genome Editing Method for *Escherichia coli* Based on CRISPR-Cas12a. *Front. Microbiol.* **2018**, *9*, 2307.
(51) Cerisy, T.; Souterre, T.; Torres-Romero, I.; Boutard, M.; Dubois, I.; Patrouix, J.; Labadie, K.; Berrabah, W.; Salanoubat, M.; Doring, V.; Tolonen, A. Evolution of a biomass-fermenting bacterium to resist lignin phenolics. *Appl. Environ. Microbiol.* **2017**, *83* (11), No. e00289-17.
(52) Cavedon, K.; Leschine, S. B.; Canale-Parola, E. Cellulase system of a free-living, mesophilic clostridium (strain C7). *J. Bacteriol.* **1990**, *172*, 4222−4230.
(53) Eid, J.; Fehr, A.; Gray, J.; Luong, K.; Lyle, J.; Otto, G.; Peluso, P.; Rank, D.; Baybayan, P.; Bettman, B.; Bibillo, A.; Bjornson, K.; Chaudhuri, B.; Christians, F.; Cicero, R.; Clark, S.; Dalal, R.; Dewinter, A.; Dixon, J.; Foquet, M.; Gaertner, A.; Hardenbol, P.; Heiner, C.; Hester, K.; Holden, D.; Kearns, G.; Kong, X.; Kuse, R.; Lacroix, Y.; Lin, S.; Lundquist, P.; Ma, C.; Marks, P.; Maxham, M.; Murphy, D.; Park, I.; Pham, T.; Phillips, M.; Roy, J.; Sebra, R.; Shen, G.; Sorenson, J.; Tomaney, A.; Travers, K.; Trulson, M.; Vieceli, J.; Wegener, J.; Wu, D.; Yang, A.; Zaccarin, D.; Zhao, P.; Zhong, F.; Korlach, J.; Turner, S. Real-time DNA sequencing from single polymerase molecules. *Science* **2009**, *323*, 133−138.
(54) Crooks, G. E.; Hon, G.; Chandonia, J.-M.; Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **2004**, *14*, 1188−1190.

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

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

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.

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-

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

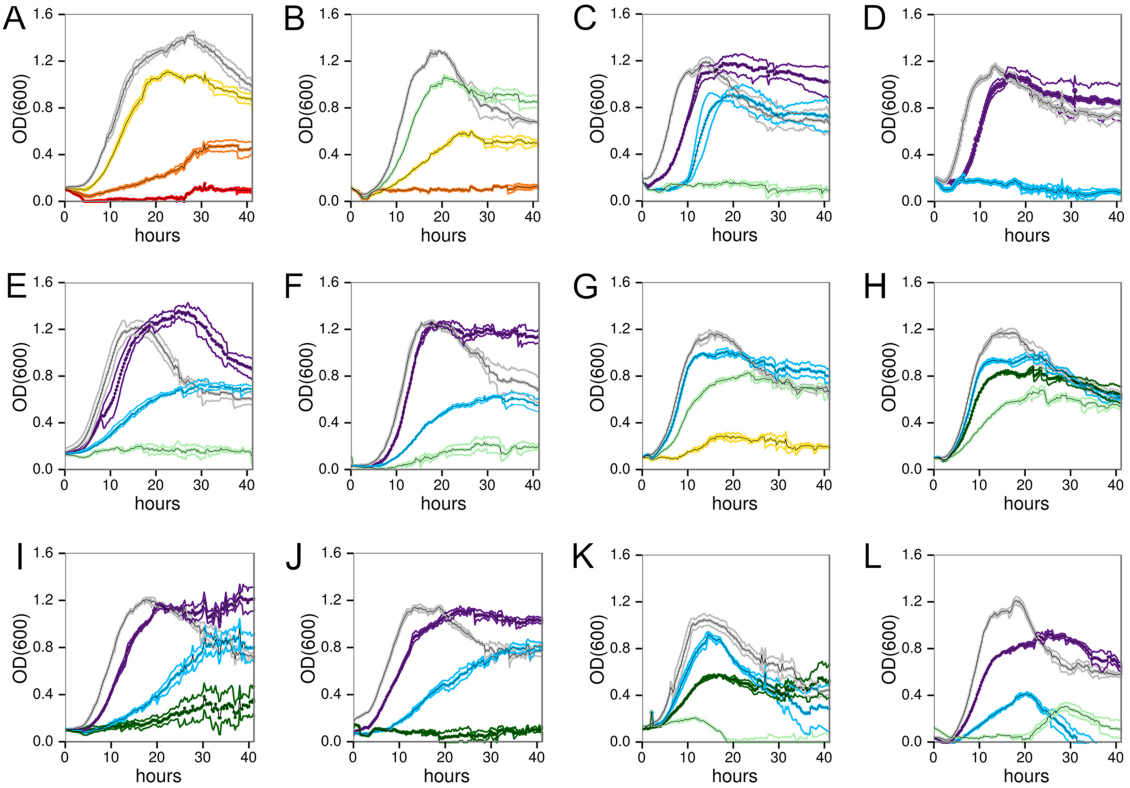


FIG 1 *C. phytofermentans* growth in GS2 glucose medium containing different concentrations of acetate (A), formate (B), 5-hydroxymethylfurfural (C), furfural (D), coumarate (E), ferulate (F), vanillate (G), 4-hydroxybenzoic acid (H), vanillin (I), benzaldehyde (J), syringaldehyde (K), and catechol (L). 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

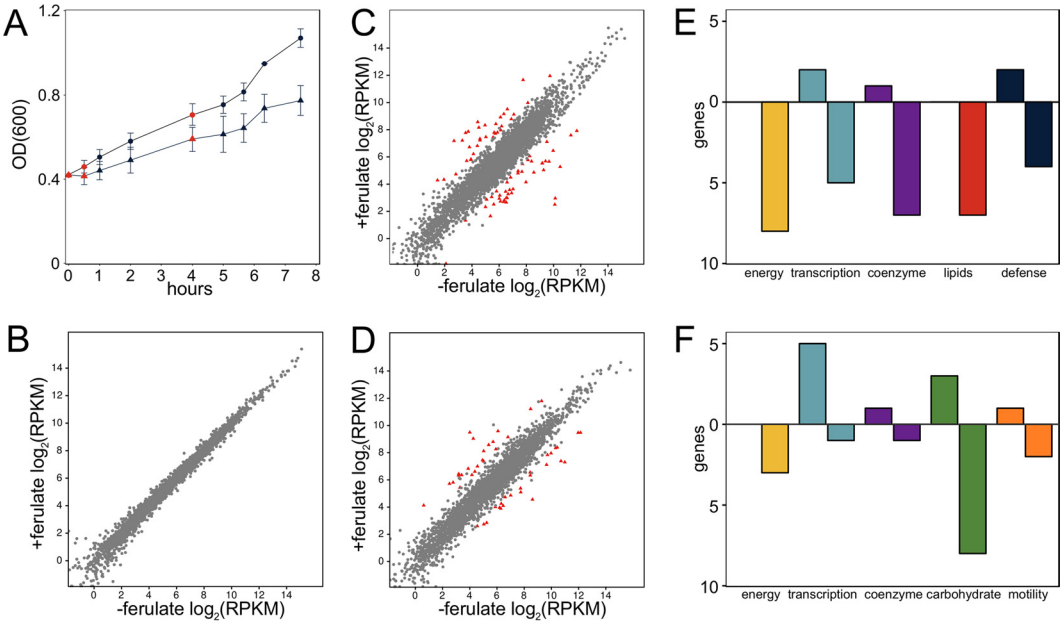


FIG 2 *C. phytofermentans* growth and gene expression during ferulate stress. (A) Growth in log-phase cultures in medium either lacking ferulate (−ferulate, circles) or containing 2 g liter^{-1} ferulate (+ferulate, triangles). Points are mean cell density (OD_{600}) of duplicate cultures ± SD with red 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

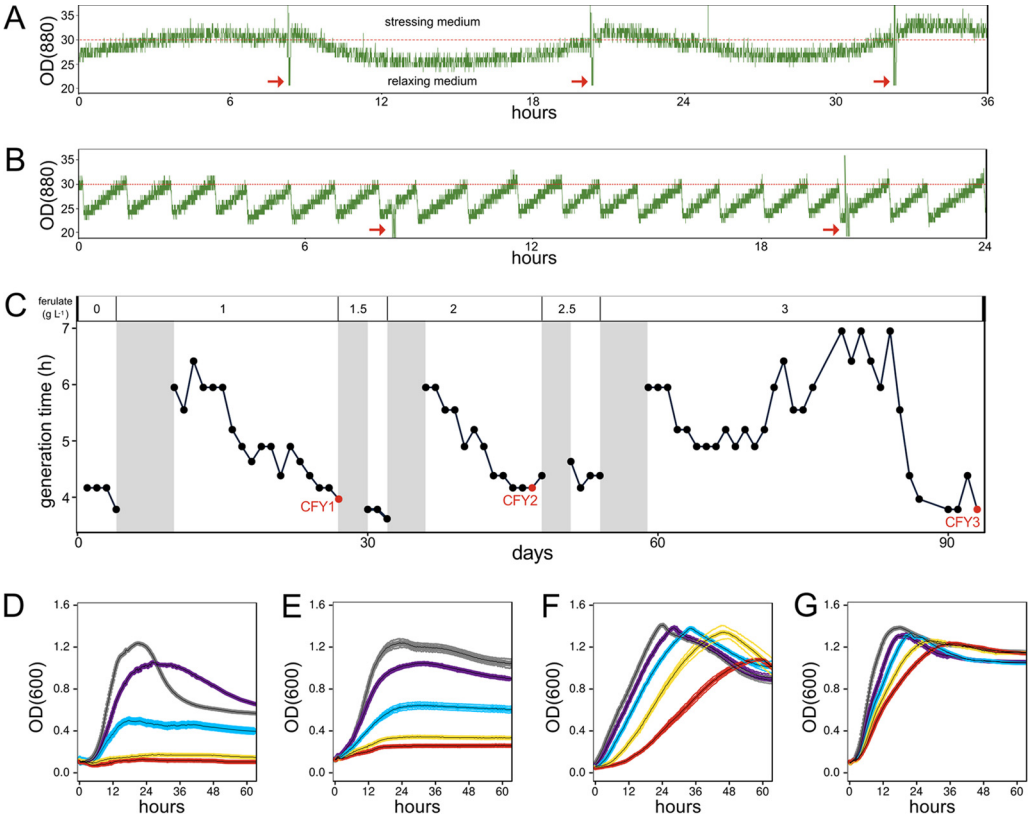


FIG 3 Growth improvement of *C. phytofermentans* GM3 strains in ferulate medium. (A) Cells were acclimated to increased ferulate using medium-swap mode, a chemostat with dilutions of stressing medium if density exceeds a threshold and with relaxing medium otherwise. (B) Growth rate was improved using 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,

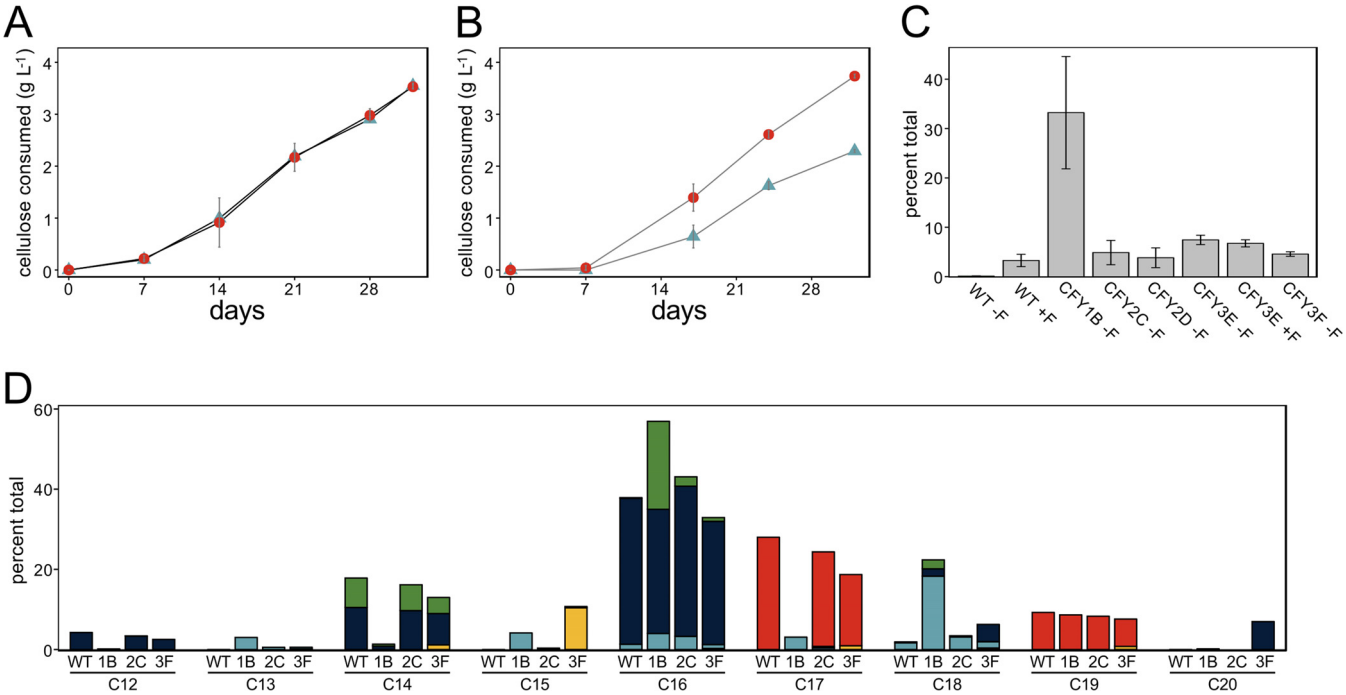


FIG 4 Growth physiology and membrane fatty acid composition of *C. phytofermentans* WT and GM3 strains. (A and B) Rate of cellulose degradation by CFY3E (red circles) and WT (blue triangles) in medium lacking ferulate (A) and medium supplemented with 2 g liter^{-1} ferulate (B). (C) Plasmalogen content 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

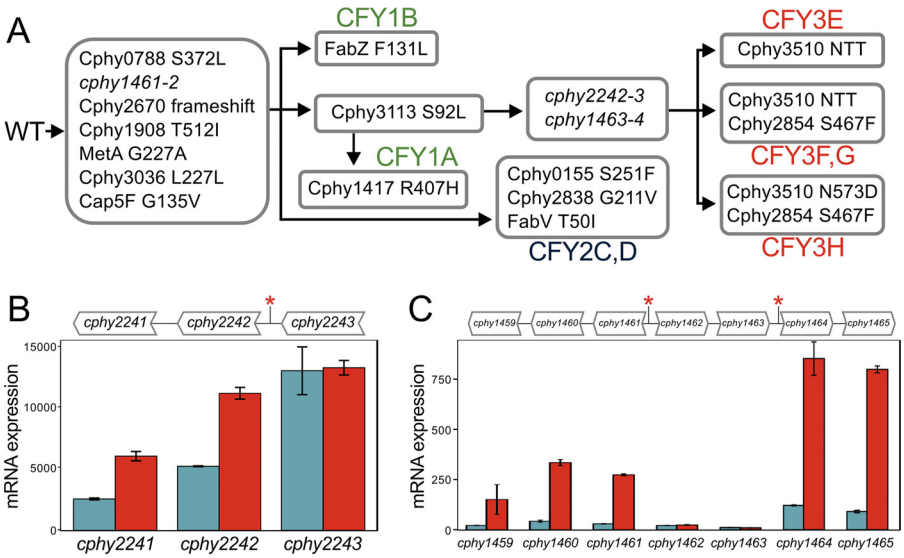


FIG 5 Small-scale genome differences in *C. phytofermentans* GM3 strains. (A) Accumulation of single-nucleotide variants and small indels in the genomes of clones isolated from the CFY1 to CFY3 time points. (B and C) mRNA expression of the *cphy2241-cphy2243* operon (B) and the *cphy1459-cphy1465* genes (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

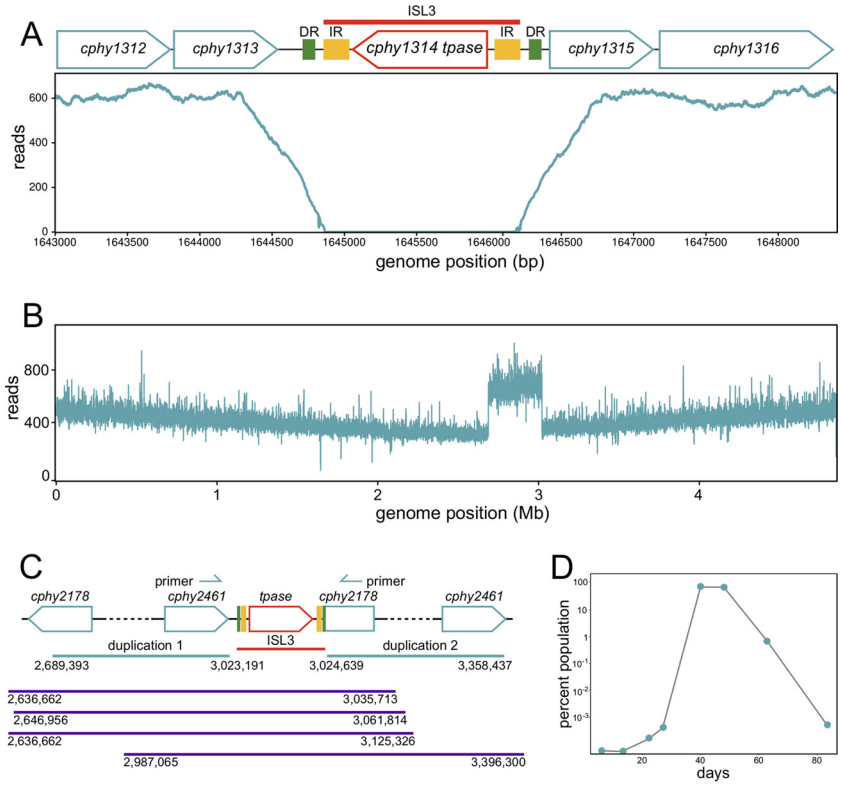


FIG 6 ISL*3*-associated genome changes in GM3 strains. (A) Read coverage showing deletion of the ISL*3-1* element including the *cphy1314* transposase gene (CFY1A shown). (B) Read coverage in CFY2C and CFY2D (shown) reveals the duplication of a 333-kb region from *cphy2178-cphy2461* (genome position 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.

Aim 2: Genome-wide studies of cellular adaptations in bacteria

High-throughput methods including DNA and RNA sequencing, proteomics, and metabolomics have emerged as transformative tools to enable systems biology-based studies in microbiology. One of our aims is to develop and apply these approaches to study molecular adaptations in bacteria, with a particular focus on Clostridia. To this end, we are studying mechanisms regulating gene expression such as DNA methylation and transcription initiation. To link how gene expression changes affect phenotypes, we examine genome-wide changes in RNA expression, quantify intracellular and secreted proteins, and measure metabolites in culture using untargeted metabolomics. Below is a summary of our research interests in these areas, followed by two published studies highlighting our progress to date.

Aim 2a: DNA methylation and transcription initiation provides insights into bacterial gene regulation

Bacterial gene regulation is a complex and finely tuned process that allows microorganisms to adapt to their environment, respond to changing conditions, and ensure their survival. DNA and RNA sequencing technologies have revolutionized our ability to dissect bacterial gene regulation by providing insights into aspects of gene regulation such as identification of regulatory elements and association of genes into operons. We have developed methods to interrogate two mechanisms that are critical to the regulation of gene expression: DNA methylation and initiation of transcription.

The methylation status of the genome plays a crucial role in the regulation of gene expression in bacteria. DNA methylation has been mainly described as part of the sequence-specific restriction modification systems to resist invasion of foreign DNA. Approximately 90% of bacterial genomes contain at least one of the three common forms of DNA methylation: 5-methylcytosine (m5C), N4-methylcytosine (m4C) and N6-methyladenine (m6A). These sites are defined by the methylase specificity and, in the case of RM systems, tend to be fully methylated to avoid cuts by the cognate

restriction enzyme. We developed an approach called RIMS-seq [1] to simultaneously sequence bacterial genomes and globally profile m5C methylase specificity using a protocol that closely resembles the standard Illumina DNA-seq with a single, additional step. In RIMS-seq, m5C is identified based on elevated C-to-T mutations on Read 1 of Illumina sequencing.

We profiled the *C. phytofermentans* genome using RIMS-seq, revealing m5C modification at 5′-GATC-3′ [2]. Even though the *C. phytofermentans* genome is methylated, *in vitro* treatment of unmethylated plasmid DNA with *C. phytofermentans* lysate did not result in DNA cleavage. Further, the electroporation efficiencies of *C. phytofermentans* were similar whether the plasmid was isolated from *C. phytofermentans* (and thus pre-methylated), *E. coli* DH5α (DAM, DCM methylation), or *E. coli* ER2796 (no methylation). We thus concluded that although the *C. phytofermentans* genome is methylated, the associated restriction enzymes are likely inactive and do not impede electrotransformation. While DNA methylation does not appear to influence transformation efficiency in *C. phytofermentans*, we expect that profiling m5C could be critical to identify sites that need to be pre-methylated in order to avoid restriction of foreign DNA in other bacteria.

While DNA sequencing provides a largely static view of the genetic blueprint, RNA sequencing offers dynamic insights into gene regulation by quantifying the transcriptional activity of genes. By comparing the transcriptomes of bacteria under various conditions (e.g., different growth phases, stress conditions, or in response to specific signals), we can identify differentially expressed genes that enable the cell to respond to environmental changes. A primary mechanism by which bacteria regulate mRNA synthesis in response to environmental changes is by altering the genomic sites at which RNA polymerase initiates transcription, called Transcription Start Sites (TSSs). As such, locating and quantifying changes in TSS usage is an important step to understand bacterial gene regulation. We developed a method called Capp-Switch sequencing to investigate TSS architecture based on capturing and purifying 5′ mRNA fragments, which are reverse transcribed with template-switching to tagged cDNA for high-throughput sequencing [3].

We applied Capp-Switch sequencing to define a genome-wide map of TSSs during *C. phytofermentans* growth on raw biomass, heterogeneous polysaccharides, and their constituent sugars. This TSS map provided a means to investigate features controlling gene regulation including RNA polymerase binding sites, 5′ untranslated region structure, alternative promoters, operons and non-standard (leaderless and antisense) transcription. We also identified sequence motifs associated with groups of TSSs 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 discovered novel transcriptional units and protein-encoding genes. Capp-Switch sequencing thus represents a general approach that can be applied to explore transcription regulation in prokaryotes.

Aim 2b: Proteomics identifies cellular adaptations for growth on lignocellulose

Proteomics captures information on protein abundance, localization (including secreted proteins), and activity, which is often more reflective of the actual cellular state than transcript levels. We developed a method for quantitative proteomics based on isotopic labeling of peptides by reductive dimethylation (ReDi labeling), permitting accurate quantification of protein expression in complex protein mixtures using mass spectrometry [4]. ReDi labeling is performed using either regular (light) or deuterated (heavy) forms of formaldehyde and sodium cyanoborohydride to add two methyl groups to each free amine. ReDi labeling involves reacting peptides with formaldehyde to form a Schiff base, which is then reduced by cyanoborohydride. This reaction dimethylates free amino groups on N-termini and lysine side chains and monomethylates N-terminal prolines. Protein samples for comparison are digested into peptides, labeled to carry either light or heavy methyl tags, mixed, and co-analyzed by LC-MS/MS. Relative protein abundances are quantified by comparing the ion chromatogram peak areas of heavy and light labeled versions of the constituent peptide extracted from the full MS spectra.

We applied ReDi proteomics to perform a systems-level analysis including 2500 proteins of how *C. phytofermentans* ferments different cellulosic substrates [5]. We

found that fermentation of cellulosic substrates by C. phytofermentans involves secretion of numerous CAZymes as well as proteins for binding of extracellular solutes, proteolysis, and motility. The most highly expressed protein in the proteome is a secreted protein that appears to compose a surface layer to support the cell and anchor cell surface proteins, including some enzymes for plant degradation. Once the secreted CAZymes cleave insoluble plant polysaccharides into oligosaccharides, they are taken into the cell to be further degraded by intracellular CAZymes, enabling more efficient sugar transport, conserving energy by phosphorolytic cleavage, and ensuring the sugar monomers were not available to competing microbes. Sugars are catabolized by EMP glycolysis incorporating reversible, PPi-dependent glycolytic enzymes, and pyruvate ferredoxin oxidoreductase. Together, these methods formed a general, proteome-based strategy to evaluate the efficiency of cellulosic bioconversion and to identify enzyme targets to engineer for improving this process.

Aim 2c: Repressors of colibactin production revealed by culture-based metabolomics

Mass spectrometry-based metabolomics enables the simultaneous quantification of thousands of metabolite features, offering the possibility to identify those that undergo changes in abundance under varying environmental conditions. We recently used untargeted, mass spectrometry-based metabolomics to compare concentrations of >1000 extracellular metabolites to identify candidate repressors of colibactin in *pks+ E.coli.* Colibactin is a small-molecule produced by various strains of *Enterobacteriaceae* that damages DNA by inducing interstrand crosslinks and double strand breaks. Colibactin-producing (*pks*+) *E. coli* promote tumorigenesis in mouse models of colorectal cancer and are elevated in abundance in CRC patient biopsies, making it important to identify the regulatory systems governing colibactin production [6].

In our study, we applied a systems biology approach to explore metabolite repression of colibactin production in *pks+ E. coli* [7]. We performed a time-series of untargeted metabolomics of *pks+ E. coli* growing in conditions with different ClbP

activities, revealing tryptophan as a potential repressor. We examine how L-tryptophan supplementation in *E. coli* ATCC 25922 cultures affects genome-wide transcription, ClbP activity, and colibactin genotoxicity. Together, our experiments identify how L-tryptophan is linked to copper export and reduced ClbP activity, representing a potential strategy to reduce colibactin genotoxicity in the intestine.

References

1. Baum C, Lin Y-C, Fomenkov A, Anton BP, Chen L, Yan B, et al. Rapid identification of methylase specificity (RIMS-seq) jointly identifies methylated motifs and generates shotgun sequencing of bacterial genomes. Nucleic Acids Res. 2021;49: e113. doi:10.1093/nar/gkab705
2. Rostain W, Zaplana T, Boutard M, Baum C, Tabuteau S, Sanitha M, et al. Tuning of Gene Expression in Clostridium phytofermentans Using Synthetic Promoters and CRISPRi. ACS Synth Biol. 2022;11: 4077–4088. doi:10.1021/acssynbio.2c00385
3. Boutard M, Ettwiller L, Cerisy T, Alberti A, Labadie K, Salanoubat M, et al. Global repositioning of transcription start sites in a plant-fermenting bacterium. Nat Commun. 2016;7: 13783. doi:10.1038/ncomms13783
4. Tolonen AC, Haas W. Quantitative Proteomics Using Reductive Dimethylation for Stable Isotope Labeling. Journal of Visualized Experiments. 2014 [cited 2 Sep 2014]. doi:10.3791/51416
5. Tolonen AC, Haas W, Chilaka AC, Aach J, Gygi SP, Church GM. Proteome-wide systems analysis of a cellulosic biofuel-producing microbe. Mol Syst Biol. 2011;7: 461. doi:10.1038/msb.2010.116
6. Buc E, Dubois D, Sauvanet P, Raisch J, Delmas J, Darfeuille-Michaud A, et al. High Prevalence of Mucosa-Associated E. coli Producing Cyclomodulin and Genotoxin in Colon Cancer. PLOS ONE. 2013;8: e56964. doi:10.1371/journal.pone.0056964
7. Bayne C, Boutard M, Zaplana T, Tolonen A. Reduced colibactin genotoxicity by L-tryptophan is linked to copper-mediated ClbP inhibition in pks+ Escherichia coli. In review.

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

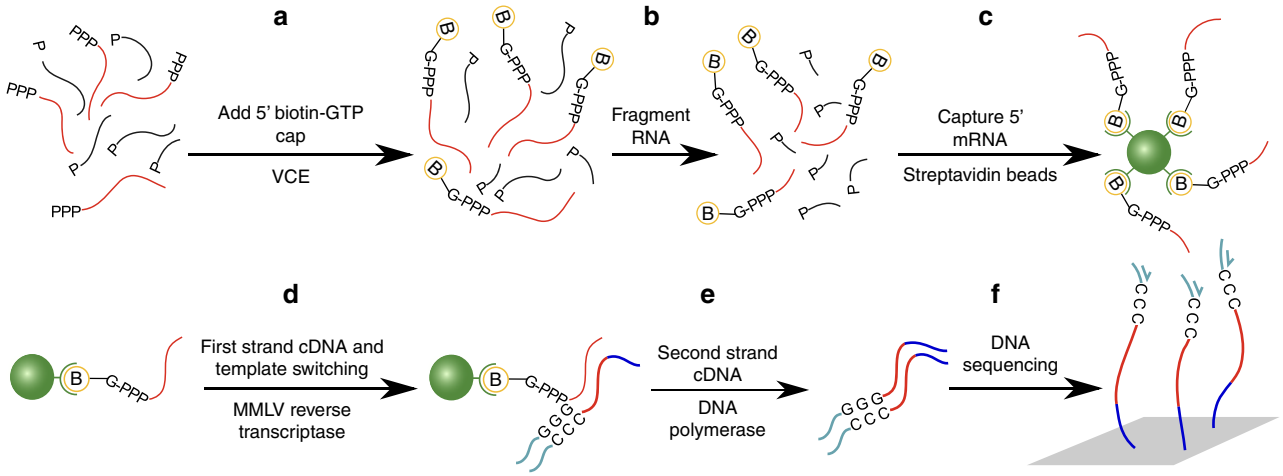


Figure 1 | Overview of the Capp-Switch sequencing approach. Capp-Switch includes (**a**–**c**) capture of 5′ mRNA fragments and (**d**–**f**) cDNA synthesis and sequencing. (**a**) The mRNA 5′ triphosphate is capped with biotin-GTP by VCE. (**b**) RNA is fragmented and (**c**) the capped 5′ mRNA fragments 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.

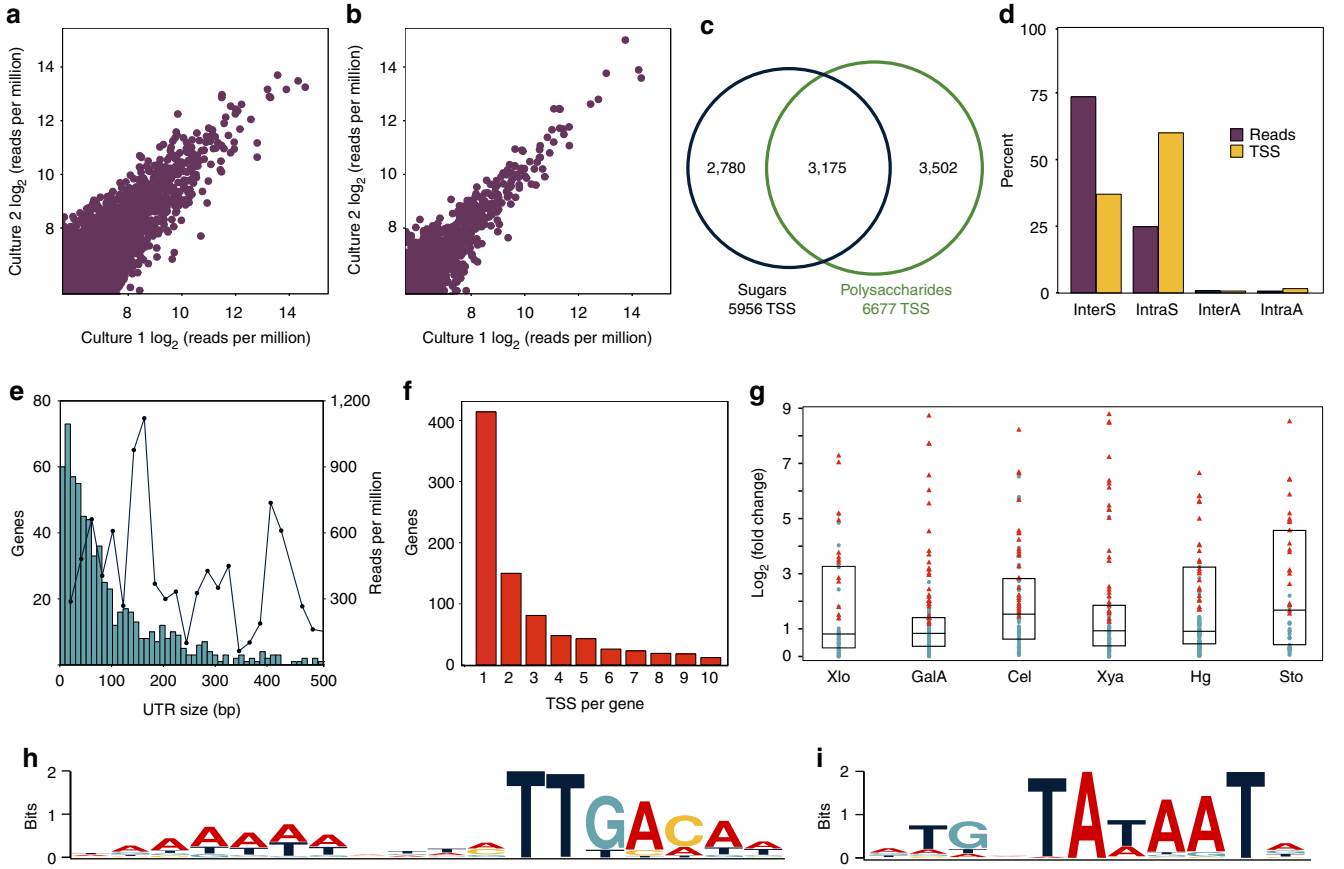


Figure 2 | General features of TSS identification by Capp-Switch sequencing. Capp-Switch reproducibly quantifies TSS usage in duplicate (**a**) glucose (4,399 TSS; $R^2 = 0.96$) and (**b**) stover (1,532 TSS; $R^2 = 0.99$) cultures. (**c**) Venn diagram showing overlap of TSS identified in at least one 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.

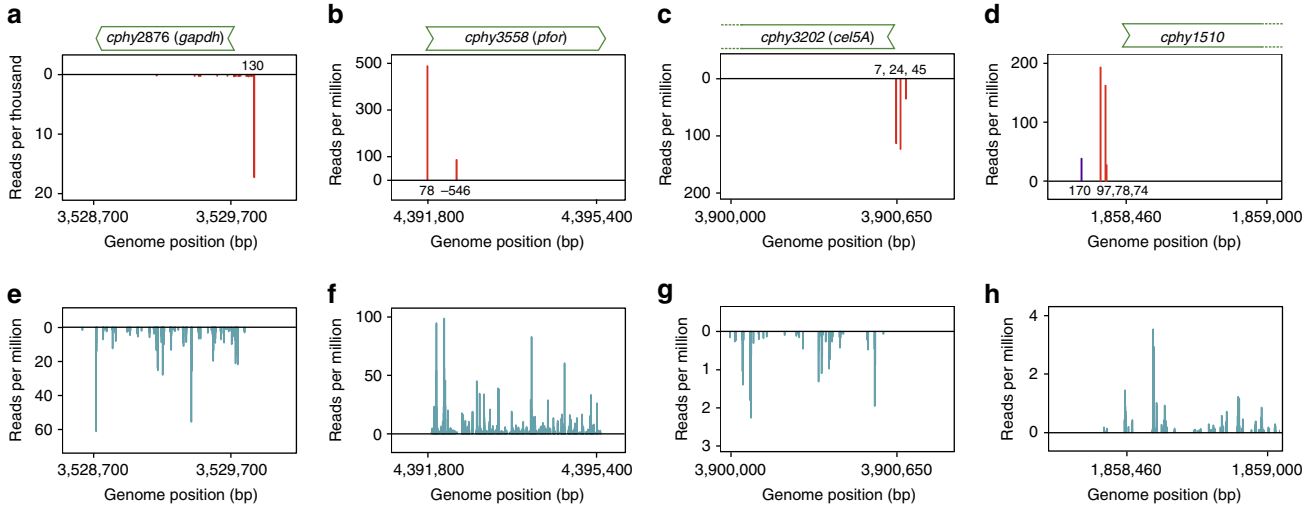


Figure 3 | Capp-Switch reads start at specific genome positions corresponding to putative TSS. The number of reads starting at each genome position are shown for Capp-Switch (**a**–**d**) and RNA-seq (**e**–**h**). The *cphy2876 gapdh* gene (**a**,**e**) has a single TSS (glucose data shown). The *cphy3558 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.

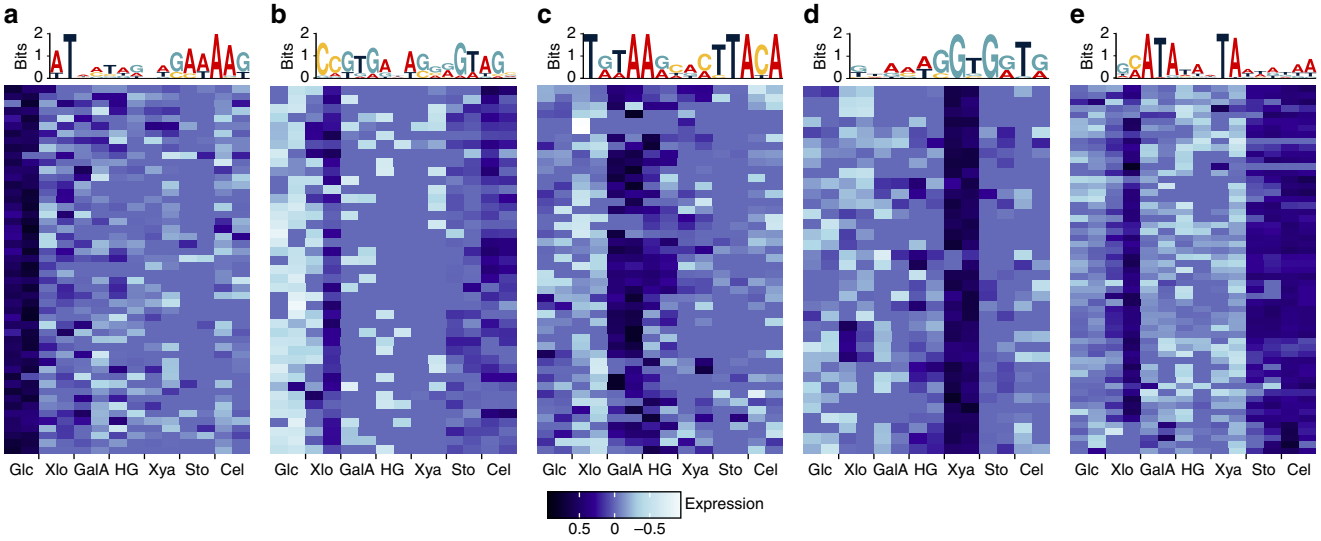


Figure 4 | TSS in carbon source-specific clusters share DNA sequence motifs. TSS clusters differentially expressed on (**a**,**b**) glucose, (**c**) galacturonic acid and HG, (**d**) xylan and (**e**) stover and cellulose are shown along with their associated sequence motifs. Rows are expression of a 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.

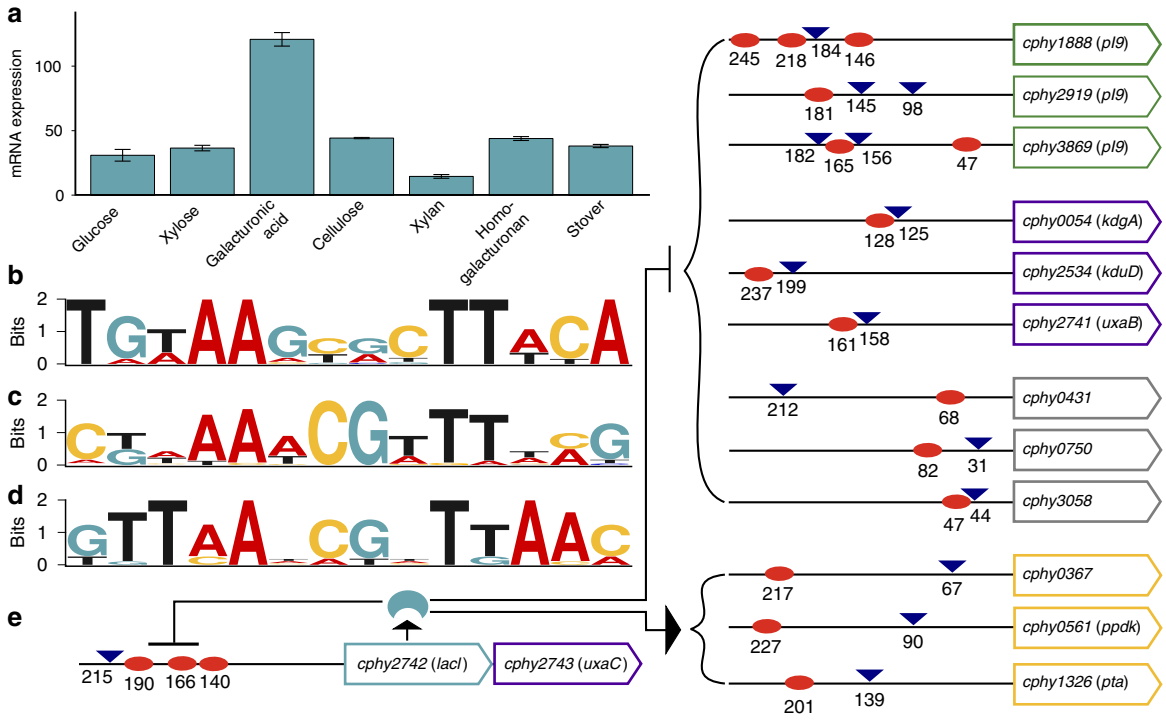


Figure 5 | The role of the LacI/GalR regulator Cphy2742 in galacturonic acid and pectin metabolism. (**a**) Transcription of the LacI/GalR gene *cphy2742* is up-regulated on galacturonic acid relative to other carbon sources. Bars shows average RNA-seq RPKM of duplicate cultures; error bars are one s.d. (**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.

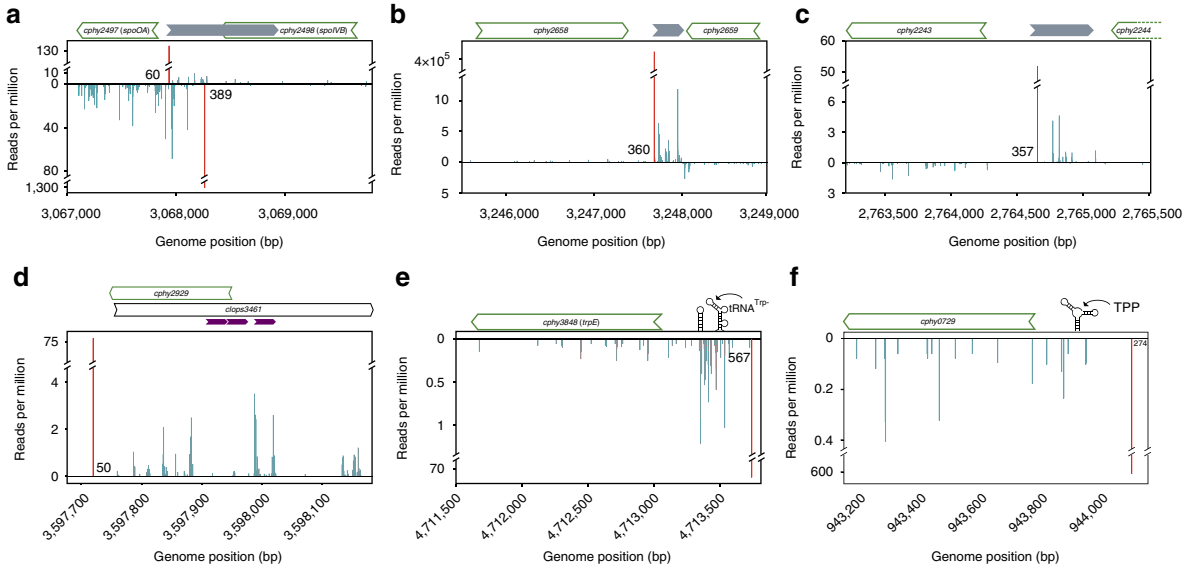


Figure 6 | TSS show genome features. (**a**) The *cphy2497 spoOA* gene has both a primary TSS and an antisense TSS in the 5′ UTR (grey arrow) that were observed on all sugars (glucose data shown). (**b**) A novel transcription unit (grey arrow) is up-regulated to be the most highly expressed TSS on biomass. (**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.

Molecular Systems Biology 7; Article number 461; doi:10.1038/MSB.2010.116
Citation: ***Molecular Systems Biology* 7:461**

www.molecularsystemsbiology.com

molecular systems biology

Proteome-wide systems analysis of a cellulosic biofuel-producing microbe

Andrew C Tolonen[1,*], Wilhelm Haas[2,*], Amanda C Chilaka[3], John Aach[1], Steven P Gygi[2] and George M Church[1]

[1] Department of Genetics, Harvard Medical School, Boston, MA, USA, [2] Department of Cell Biology, Harvard Medical School, Boston, MA, USA and [3] Department of Biology, Northeastern University, Boston, MA, USA
* Corresponding authors. AC Tolonen, Department of Genetics, Harvard Medical School, NRB 238, 77 Avenue Louis Pasteur, Boston, MA 02115, USA. Tel.: +1 617 432 6510; Fax: +1 617 432 6510; E-mail: tolonen@alum.mit.edu or W Haas, Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA. Tel.: +1 617 432 3155; Fax: +1 617 432 1144; E-mail: wilhelm_haas@hms.harvard.edu

Received 9.9.10; accepted 30.11.10

Fermentation of plant biomass by microbes like *Clostridium phytofermentans* recycles carbon globally and can make biofuels from inedible feedstocks. We analyzed *C. phytofermentans* fermenting cellulosic substrates by integrating quantitative mass spectrometry of more than 2500 proteins with measurements of growth, enzyme activities, fermentation products, and electron microscopy. Absolute protein concentrations were estimated using Absolute Protein EXpression (APEX); relative changes between treatments were quantified with chemical stable isotope labeling by reductive dimethylation (ReDi). We identified the different combinations of carbohydratases used to degrade cellulose and hemicellulose, many of which were secreted based on quantification of supernatant proteins, as well as the repertoires of glycolytic enzymes and alcohol dehydrogenases (ADHs) enabling ethanol production at near maximal yields. Growth on cellulose also resulted in diverse changes such as increased expression of tryptophan synthesis proteins and repression of proteins for fatty acid metabolism and cell motility. This study gives a systems-level understanding of how this microbe ferments biomass and provides a rational, empirical basis to identify engineering targets for industrial cellulosic fermentation.
Molecular Systems Biology **7**: 461; published online 18 January 2011; doi:10.1038/MSB.2010.116
Subject Categories: proteomics; cellular metabolism
Keywords: bioenergy; clostridium; proteomics

Introduction

Cellulosic biomass is the world's most abundant biological energy source (Leschine, 1995). Recycling this vast carbon sink by cellulolytic microbes is one of the largest material flows in the global carbon cycle (Falkowski *et al*, 2000). Microbes could industrially convert over 1.3 billion metric tons of cellulosic biomass to fuels and chemicals per year in North America (Perlack *et al*, 2005), which could sustainably provide enough ethanol for 65% of US ground transportation fuel at current levels (Somerville, 2006). However, plant biomass is composed primarily of high-molecular weight polysaccharides in a quasicrystalline structure, making the deconstruction of biomass a key challenge to developing cellulosic biofuels (Houghton *et al*, 2006). Consolidated bioprocessing (Lynd *et al*, 2002) is a promising strategy to overcome biomass recalcitrance by using microbes such as *Clostridium phytofermentans* that secrete enzymes to both depolymerize biomass and then ferment the resulting hexose and pentose sugars to a biofuel such as ethanol.

C. phytofermentans is a mesophile from forest soil that ferments both of the main components of plant biomass, cellulose and hemicellulose, to ethanol and hydrogen (Warnick *et al*, 2002). As a group 14 clostridium, this microbe is phylogenetically distant from well-studied cellulolytic clostridia. The *C. phytofermentans* genome encodes 161 carbohydrate-active enzymes (CAZy) including 108 glycoside hydrolases spread across 39 families (Cantarel *et al*, 2009), highlighting the elaborate set of enzymes needed to break-down different biomass types. Hydrolases in most clostridia have dockerin domains to bind a scaffolding protein on the cell exterior forming a multienzyme cellulosome. *C. phytofermentans* lacks scaffolding and dockerin domains, suggesting that cellulolytic enzymes are either freely secreted or are anchored to the cell in a novel, cellulosome-independent manner.

Faced with the complexity of metabolizing biomass, systems-level strategies are needed to identify hydrolases and metabolic enzymes to engineer microbes for improved cellulosic bioconversion. We demonstrate such a strategy (Figure 1) in *C. phytofermentans* by integrating analyses of

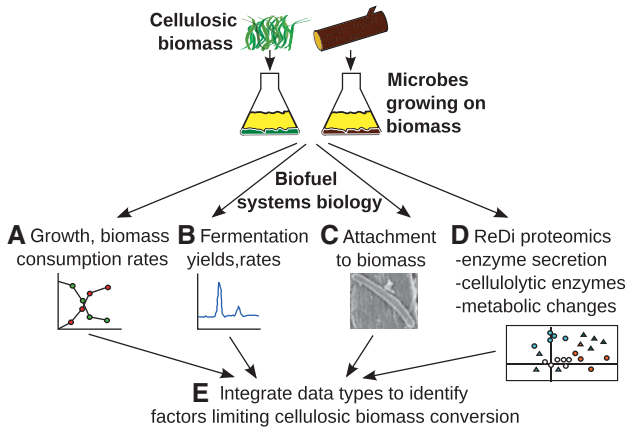


Figure 1 Integrated systems biology strategy to study cellulosic bioconversion. Cultures metabolizing different biomass substrates were examined for (**A**) growth and biomass consumption rates (Figure 2A–C), (**B**) fermentation production rates and yields (Figure 2D–F), and (**C**) ability of the microbe to adhere to cellulosic substrates (Figure 2G–I). (**D**) Supernatant and cellular protein samples were taken for reductive dimethylation (ReDi) proteomics and analyzed for enzyme secretion (Figure 4), abundances of cellulolytic enzymes (Figure 5), and proteome-wide changes (Figure 6). (**E**) These data were integrated to identify key enzymes for each step in biomass deconstruction and fermentation (Table I, Figure 7).

growth, fermentation, enzyme activities, and electron microscopy with quantitative mass spectrometry-based proteomics of more than 2500 proteins. Protein concentrations were estimated by machine learning-supported spectral counting (Absolute Protein EXpression, APEX) (Lu *et al*, 2007). Protein levels on hemicellulose and cellulose relative to glucose were determined using reductive methylation (Hsu *et al*, 2003; Boersema *et al*, 2009), here called reductive dimethylation (ReDi) labeling, to chemically incorporate hydrogen or deuterium isotopes at lysines and N-terminal amines of tryptic peptides. We show that ReDi labeling gives accurate, low-cost quantification of a microbial proteome and can be used to discern extracellular proteins. *C. phytofermentans* expressed more than 100 CAZy and adapted their stoichiometries to each cellulosic substrate. Cellulosic fermentation entailed additional changes such as increased tryptophan and nicotinamide synthesis, use of alternative glycolytic enzymes, and adhesion to the substrate. We describe how these data provide a blueprint showing promising genetic targets to engineer microbes for more efficient conversion of biomass to fuels and biomaterials.

Results

Growth, fermentation, and cell adhesion

Wild-type *C. phytofermentans* ATCC 700394 is well suited for cellulosic biofuels as cultures were actively growing on glucose, hemicellulose, and cellulose and converting these substrates primarily to ethanol when samples were taken for proteomics (Figure 2A–F). Growth was faster on hemicellulose (Figure 2B) than on glucose (Figure 2A) or xylose (Supplementary Figure S1), which is unexpected because hemicellulose is a beta-1,4-D-xylopyranose polymer that must be cleaved to xylose and isomerized before glycolysis. Ethanol titers reached 77% of the maximum theoretical yield in the glucose cultures (30 h, Figure 2D) and 27% in the hemicellulose cultures (24 h, Figure 2E) during the sampling period (see Supplementary Figs S2-5 for growth and ethanol yield calculations). Final ethanol concentrations in glucose cultures were >95% of the maximum theoretical yield after 48 h (Supplementary Figure S6). Stable cell densities (10^7–10^8 CFU ml^{-1}) in the cellulose cultures resulted in linear rates of cellulose degradation (Figure 2C) and ethanol formation (Figure 2F) that correspond to a direct conversion of cellulose to ethanol at 68% of the maximum theoretical yield. The cellulose cultures produced an ethanol/acetate ratio (9.54) similar to the highest yields reported for clostridia (Lynd *et al*, 2002).

Adhesion to plant substrates is an important adaptation in some cellulolytic bacteria to enhance cellulolysis (Lu *et al*, 2006) by increasing enzyme concentrations near the substrate and excluding competitors from the liberated sugars. *C. phytofermentans* was adhered to both cellulosic substrates when samples were taken for proteomics, even though it lacks cellulosomes that enable adhesion in other clostridia. Cells growing on hemicellulose (Figure 2H, Supplementary Figure S7) were sometimes laden with surface nodules, suggesting that hemicellulose particles were bound to the cell surface. Cells in cellulose cultures were shorter, non-flagellated, and adhered to cellulose (Figure 2I, Supplementary Figure S8). The mechanism by which *C. phytofermentans* adheres to plant substrates is unknown, but may include cell surface-binding proteins, secretion of an extrapolysaccharide glycocalyx, or cell surface pili visible by transmission electron microscopy (Supplementary Figure S9) that may facilitate binding as in *Ruminococcus albus* (Morrison and Miron, 2000).

Protein quantification

Proteins from glucose, hemicellulose, and cellulose culture lysates and the supernatants of hemicellulose and cellulose cultures were identified by liquid chromatography tandem mass spectrometry (LC-MS/MS). We quantified 2567 proteins across all five treatments (Figure 3A), representing 65% of the proteome (Supplementary Table SI). The fraction of proteins identified was comparable across all 23 Clusters of Orthologous Genes (COG) categories (Figure 3B), showing that the protein expression profiles cover diverse cellular functions. Although 357 hypothetical proteins were detected and thereby validated as real proteins, hypotheticals had the lowest percent identification among COGs, suggesting many are either expressed under specific conditions or are not expressed.

Absolute protein abundances measured by APEX were highly correlated between duplicate cultures (Supplementary Figure S10). Comparing total protein synthesis allocated to each COG (summed APEX for all proteins in COG) in the glucose and cellulose cultures gives a general picture of cellular changes on these substrates (Figure 3C). These COG expression differences include processes directly related to cellulose degradation (i.e., increased carbohydrate metabolism) and indirect changes such as a reduced transcription and translation on cellulose, which likely resulted from slower growth. Hypothetical proteins were higher on cellulose

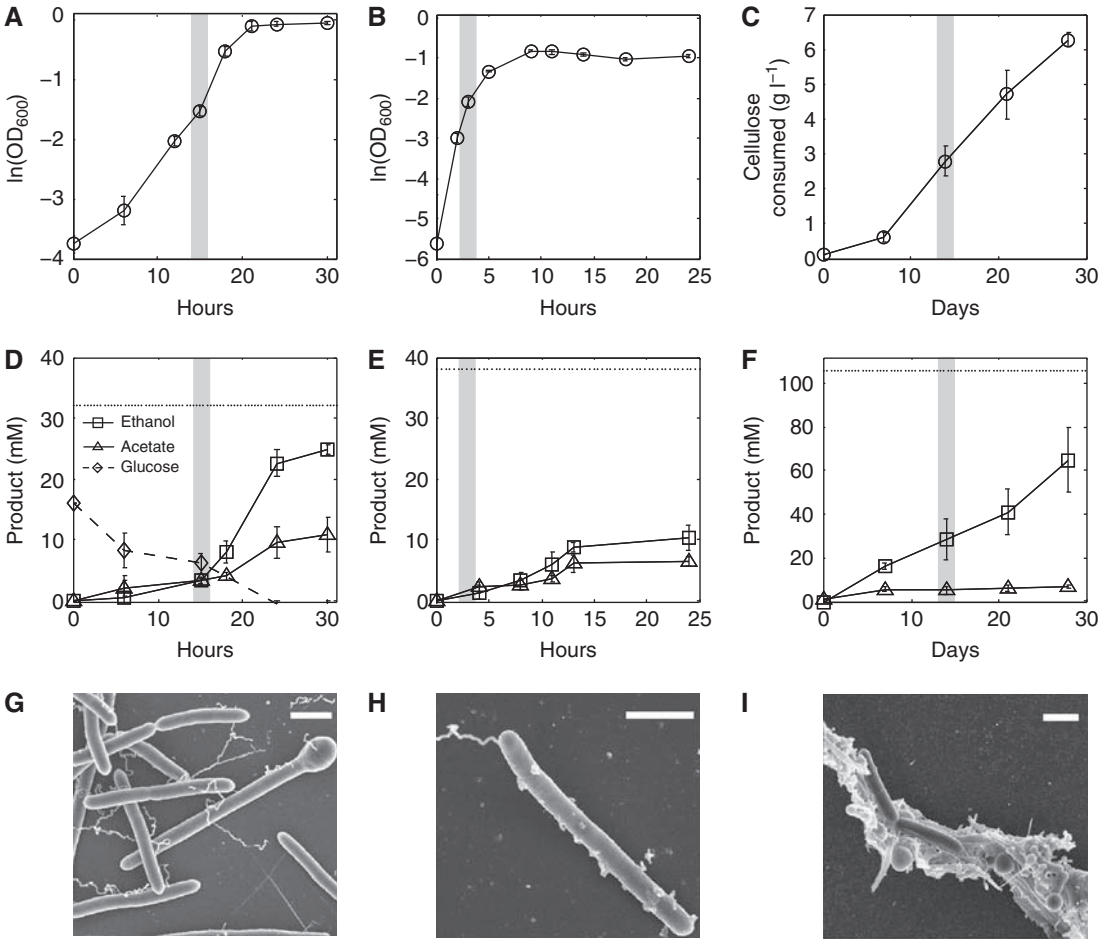


Figure 2 Growth (**A**–**C**), fermentation (**D**–**F**), and cell morphology (**G**–**I**) of *C. phytofermentans* on different carbon sources. Data points are means of triplicate cultures; error bars show one s.d. and are smaller than the symbols where not apparent. Gray bars show when samples were taken for mass spectrometry. Growth on glucose (A) and hemicellulose (B) was quantified as OD_{600}. Growth on cellulose (C) was measured as dry mass of cellulose in culture. Production of ethanol and acetate, the two most abundant fermentation products, and glucose consumption in the glucose treatment was measured by HPLC. Dotted lines show maximum theoretical yield of ethanol. Scanning electron microscopy shows cells growing on glucose (G), hemicellulose (H), and cellulose (I). White scale bar is 1 μm.

(Figure 3C), suggesting many proteins of unknown function have roles in cellulose metabolism.

ReDi labeling modified the primary amines of all tryptic peptides with methyl groups bearing either deuterium (heavy label) or hydrogen (light label). The MS1 peak area ratio (MPA ratio), based on ion intensity peaks of the same peptide with different isotopic labels, was used to quantify proteome-wide expression changes between treatments. As an initial validation of the accuracy and reproducibility of ReDi labeling, we compared replicate experiments in which differentially labeled samples were mixed at several ratios. ReDi quantifications correspond to the ratios at which the samples were mixed (Supplementary Table SII) and replicate experiments were highly correlated across all ratios (r^2=0.96) (Supplementary Figure S11).

ReDi labeling was applied to quantify proteome-wide expression changes on cellulosic substrates (heavy labeled) versus glucose (light labeled) (Supplementary Table SIII). We observed that 94% of proteins were expressed within twofold levels for differentially labeled duplicate glucose cultures, whereas 80% of proteins were within twofold when glucose was compared with hemicellulose and 49% when glucose was compared with cellulose (Figure 3D). Growth on cellulosic substrates thus results in widespread changes to the proteome relative to growth on glucose. Global expression changes on cellulose versus glucose were well correlated (r^2=0.82) in replicate pairs of cultures (Figure 3E). Expression of CAZy proteins were also correlated with their mRNA expression changes on hemicellulose (r^2=0.71) and cellulose (r^2=0.77) relative to glucose (Figure 3F).

Secretome

Protein secretion mechanisms in *C. phytofermentans* are of particular interest because extracellular enzymes are needed to degrade insoluble, cellulosic polymers. The secretome was defined for hemicellulose (Supplementary Table SVI) and cellulose (Supplementary Table SVII) cultures using ReDi labeling to measure concentration differences of each protein between the supernatant (heavy labeled) and the lysate (light labeled) of the same culture (Supplementary Figure S12). Although the genome encodes several protein secretion

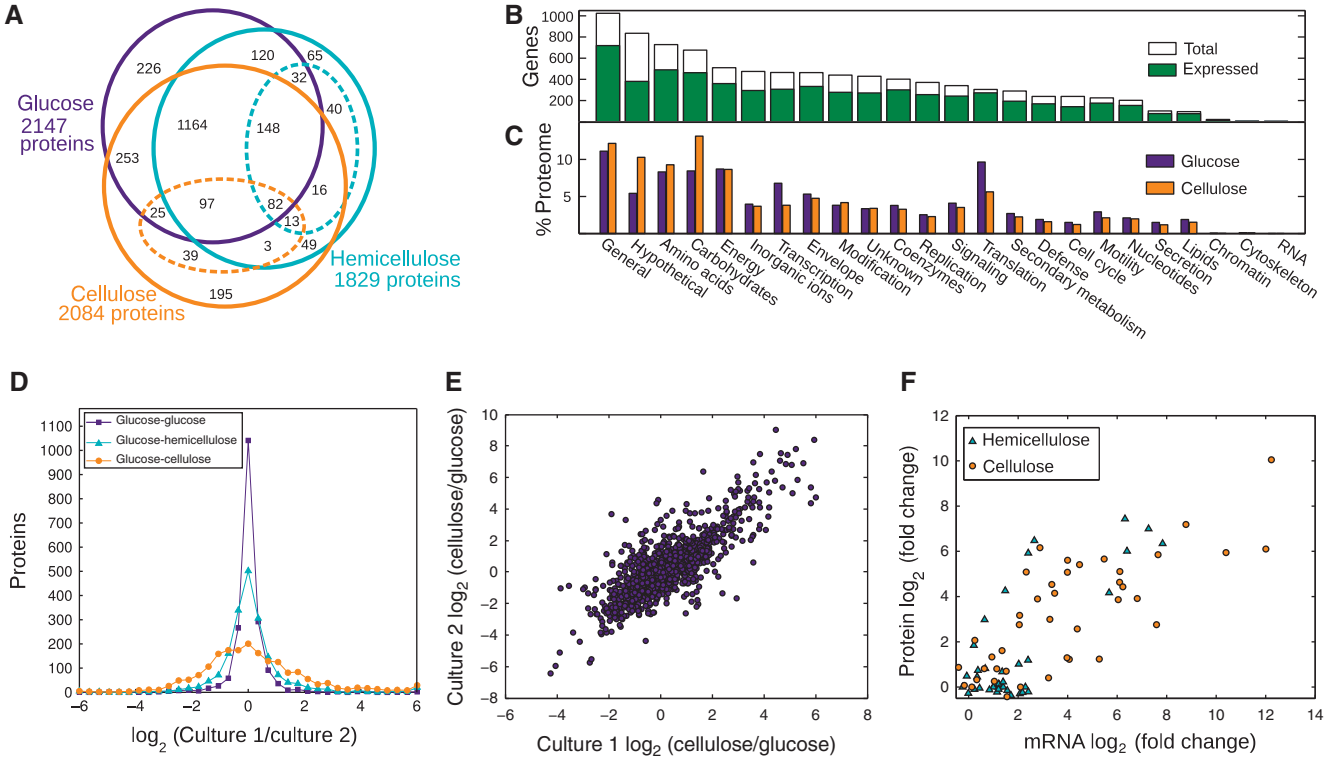


Figure 3 Protein identification (**A**–**C**) and quantification (**D**–**F**) by mass spectrometry. (A) Venn diagram of proteins identified in each treatment. Protein subsets in the hemicellulose and cellulose culture supernatants are shown with dashed ellipses. In total, 2567 of 3926 (65%) putative proteins were detected. (B) The 65% overall protein identification rate is conserved across Clusters of Orthologous Genes (COG) functional categories. (C) The percent of the proteome shown as summed Absolute Protein EXpression (APEX) values in each COG category for cells growing on glucose and cellulose. (D) Relative protein expression in different cultures quantified by ReDi labeling. The fraction of proteins expressed within twofold levels for the glucose treatment compared with a duplicate glucose culture (94%), hemicellulose (80%), and cellulose (49%) cultures. (E) Fold change in protein expression (MS1 peak area ratio, MPA ratio) for cellulose versus glucose duplicate cultures is highly correlated (r^2=0.82). (F) Scatter plot of mRNA versus protein expression of 40 carbohydrate-active enzymes on cellulose (orange circles, r^2=0.77) and hemicellulose (turquoise triangles, r^2=0.71) versus glucose. The mRNA fold change was measured by qRT–PCR (−ΔΔCt).

pathways, the Sec system appears to be the primary means of secretion. N-terminal Sec-dependent secretion signals scored by SignalP3.0 (Bendtsen *et al*, 2004) (Supplementary Table SVIII) correlate with the probability of a protein being in the supernatant (Figure 4A) and many highly expressed CAZy have putative Sec secretion signals (Table I). Reduced N-terminal MS/MS peptide coverage of supernatant proteins with putative signal peptides supports the expectation that the signals were removed from mature proteins (Supplementary Figure S13). Defining the *C. phytofermentans* secretion signals will help to optimize secretion of heterologous proteins in this organism and to refine signal peptide predictions in related microbes. Type I signal peptides can be separated into a positively charged N-region, a hydrophobic H region and an alanine-rich cleavage site (Figure 4B). Type II signals for membrane-anchored lipoproteins have a shorter N-region and an lipobox anchor domain similar to *C. acetobutylicum* (Desvaux *et al*, 2005) containing the conserved cysteine for lipification (Figure 4C).

Proteins in culture supernatants act primarily in carbohydrate and protein degradation, cell surface and flagellar assembly, and transport (Figure 4D). Flagellar proteins were more abundant in the hemicellulose supernatant, consistent with SEM images (Figure 2H,I) showing that *C. phytofermentans* has a motile, planktonic phase on hemicellulose and a sessile phase when adhered to cellulose. Culture supernatants also contained numerous lipoproteins tethered to the cell surface. Lipoproteins are secreted by the Sec system, but attach to the membrane by the lipified N-terminal cysteine (von Heijne, 1989) in the lipobox. The most abundant lipoproteins were 43 expressed extracellular solute-binding proteins (ESBs) that capture substrates and pass them to a membrane transporter. In all, 47 of the 53 *C. phytofermentans* ESBs are involved in sugar transport and can be very highly expressed; the oligosaccharide-binding ESB Cphy2466 was the third most highly expressed protein on cellulose. In addition to nutrient-binding proteins, several CAZy are surface-attached lipoproteins.

The most highly expressed protein in the proteome, Cphy3510, is an unannotated, secreted protein with sequence similarity to the *Bacillus anthracis* S-layer protein, Sap (Etienne-Toumelin *et al*, 1995). The S-layer is a protein coat covering some bacteria that serves to attach extracellular proteins, gives mechanical stabilization, and acts as a molecular sieve (Sára and Sleytr, 1987). S-layers are generally composed of a single protein subunit which, on export by the Sec-system, spontaneously assembles into a symmetrical lattice (Sleytr and Messner, 1983). S-layers require $\sim 5 \times 10^5$ protein molecules (Sleytr and Messner, 1988), often making the S-layer protein the most abundant of the cell. Transmission

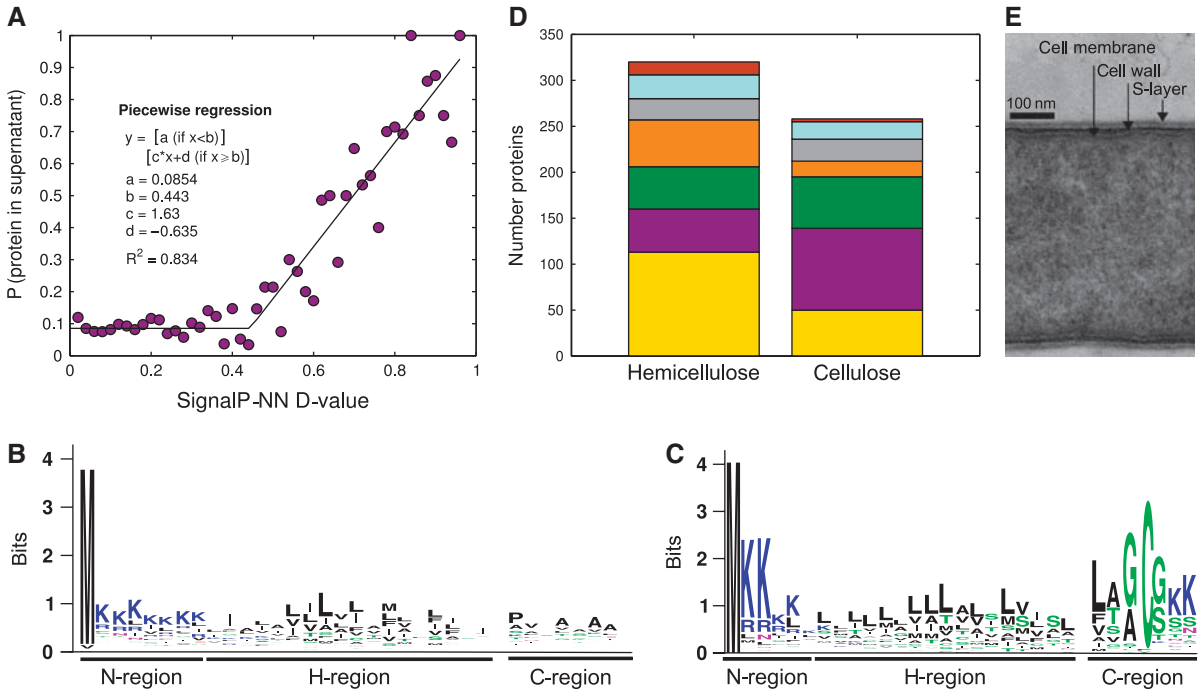


Figure 4 The C. *phytofermentans* secretome. (**A**) Proteins with high-scoring N-terminal signal peptides have a greater probability of being in culture supernatants. Fraction of proteins in proteome at each SignalP-NN D value observed in the supernatants of hemicellulose or cellulose cultures. Data were fit to a piecewise linear regression with the leftmost regression to a horizontal line. Consensus sequences of (**B**) type I and (**C**) type II lipoprotein N-terminal signal peptides for proteins found in the supernatant of cellulose cultures. (**D**) Functional categories of proteins in culture supernatants: rust, flagellum; turquoise, cell wall/surface; gray, proteases; orange, transport; green, CAZy; purple, other; gold, unknown. (**E**) Transmission electron micrograph of a *C. phytofermentans* cell cross section showing the cell membrane, cell wall, and surface layer.

electron micrographs of *C. phytofermentans* show a surface layer exterior to the cell wall (Figure 4E; Warnick *et al*, 2002). On the basis of the high expression of Cphy3510 and its sequence similarity to Sap, we propose that the cell is covered by an Cphy3510-based S-layer that supports the cell and anchors cell surface proteins, including plant degradation enzymes.

Carbohydratases

Cells have elevated secreted and cellular cellulolytic enzyme activities on cellulosic substrates (Figure 5A). Supernatant proteins have higher activity against hemicellulose than cellulose even in cellulose cultures, showing the high activity of hemicellulases in this organism. Although most enzyme activity is found in culture supernatants, cellular enzymes also contribute to degrading these polymers (Figure 5A). Supportingly, growth on plant substrates involves the upregulation of numerous extra and intracellular CAZy, which are divided into distinct subsets that responded to each substrate (Figure 5B). CAZy were expressed at similar, basal levels on glucose (Figure 5C), whereas CAZy changed stoichiometries and represent a greater fraction of the proteome on hemicellulose and cellulose (Figure 5D and E). Combining CAZy expression with localization predicted by PsortB v2.0 (Gardy *et al*, 2005) and SignalP3.0 (Bendtsen *et al*, 2004) allows putative identification of the main enzymes to catalyze each step in the deconstruction of cellulosic substrates (Table I).

The first step of hemicellulose breakdown is cleavage of the xylan backbone by several extracellular endoxylanases (Table I). The most highly expressed endoxylanase, Cphy2108, is predicted by LocateP (Zhou *et al*, 2008) to have an LPXTG motif for covalent attachment to the peptidoglycan wall. Adhesion of a surface-bound xylanase to hemicellulose could account for the cell surface particles seen by electron microscopy (Figure 2H). Xylosaccharides are hydrolyzed to xylose by two main cellular exoxylosidases: Cphy3009 cleaves xylosides from the non-reducing end of xylosaccharides and Cphy3207 acts on the reducing end. Hardwood xylan such as the birch wood used in this study often has side chains of glucuronic and acetic acid (Shallom and Shoham, 2003), which are removed by additional enzymes (Table I).

Depolymerization of cellulose fibers is catalyzed by endocellulases, which cut random sites on the cellulose surface to generate free chain ends. Among the many CAZy upregulated on cellulose (Figure 5B), the most highly secreted putative endocellulase Cphy3367 is required for cellulose degradation (Tolonen *et al*, 2009) and solubilizes cellulose *in vitro* (Zhang *et al*, 2010a). Exocellulases channel the freed cellulose chains through an active site tunnel to iteratively cleave saccharides (Divne *et al*, 1994). Cphy3368 is an exocellulase (Zhang *et al*, 2010b) that is expressed similar to Cphy3367 (Figure 5) and these two enzymes synergize to accelerate cellulolysis *in vitro* (Zhang *et al*, 2010a). Expression data supports the resulting cellodextrins are transported for intracellular catabolism by hydrolysis with beta-glucosidases and phosphorolytic cleavage with cellodextrin phosphorylases. Cphy0220 was the most highly expressed putative beta-glucosidase (Figure 5D). Three cellodextrin phosphorylases (Table I) cleave terminal glucosides from cellodextrins using

Table I Highly expressed enzymes for each step in the degradation of hemicellulose and cellulose

Protein	Function	CAZy	Hemicellulose	Cellulose	Psort	Signalp-NN
Hemicellulose						
Xylan backbone to xylosaccharides						
Cphy2108	Cell surface 1,4-β-D-xylanase	GH10, CBM22	5737.9(S)	563.8(S)	Cell wall	0.9
Cphy2105	1,4-β-D-Xylanase	GH11	3120.7(S)	6007.9	Extracellular	0.97
Cphy1510	1,4-β-D-Xylanase	GH10	2074.5(S)	29559.8(S)	Extracellular	0.9
Cphy0624	1,4-β-D-Xylanase	GH10, CBM22	1055.6(S)	4767.9(S)	Extracellular	0.65
Cphy3010	1,4-β-D-Xylanase	GH10	933.1	0.0	Extracellular	—
Xylosaccharides to xylose						
Cphy3009	Non-reducing end 1,4-β-xylosidase	GH3	5982.0	542.6	Cytoplasmic	—
Cphy3207	Reducing end 1,4-β-xylosidase	GH8	2473.2	0.0	Unknown	—
Remove hemicellulose side groups						
Cphy3158	α-1,2-Glucuronosidase	GH67	3508.9	82.4	Cytoplasmic	—
Cphy3160	β-Glucuronidase, galactosidase	GH2	813.3	1701.1	Cytoplasmic	—
Cphy2632	Arabinase	GH43	703.4	0.0	Cytoplasmic	—
Cphy2848	α-Glucuronidase, galactosidase	GH4	701.1	199.9(S)	Cytoplasmic	—
Cphy3862	Xylanase, carboxylesterase	GH10, CE15	466.8(S)	1511.7(S)	Extracellular	0.9
Cphy2730	Acetyl xylan esterase	CE4	155.4	225.6	Cytoplasmic	—
Cellulose						
Hydrolysis of cellulose to cellodextrins						
Cphy3368	Exocellulase	GH48, CBM3	1911.7(S)	9730.3(S)	Extracellular	0.65
Cphy3367	Bifunctional endo, exocellulase	GH9, CBM3	2181.4(S)	9277.7(S)	Extracellular	0.87
Cphy3202	Cellulase	GH5, CBM2, CBM46	1550.3(S)	3485.3(S)	Unknown	0.81
Cphy2058	Cellulase	GH5	424.5(S)	1199.0(S)	Unknown	0.62
Cphy1163	Cellulase	GH5	136.9(S)	525.1	Unknown	0.82
Hydrolysis of cellodextrins						
Cphy0220	Non-reducing end β-glucosidase	GH3	688.9	3469.6	Cytoplasmic	—
Cphy1169	Endo-1,4-β-D-glucanase	GH51	112.0	615.3	Cytoplasmic	—
Phosphorolytic cleavage of cellodextrins to glucose-1-phosphate						
Cphy3854	Cellodextrin, cellobiose phosphorylase	GH94	136.4	7602.4	Membrane	—
Cphy0430	Cellodextrin, cellobiose phosphorylase	GH94	166.2	4421.4	Membrane	—
Cphy1929	Cellodextrin, cellobiose phosphorylase	GH94	133.6	2743.0	Membrane	—
Other highly expressed CAZy						
Cphy1799	Chitinase	GH18, CBM12	175.3(S)	43042.0	Unknown	0.77
Cphy1800	Chitinase	GH18, CBM12	101.9(S)	31498.1(S)	Extracellular	0.8
Cphy0218	α-Glucosidase	GH31	184.5	3366.0	Cytoplasmic	—
Cphy1687	Polysaccharide deacetylase	CE4, CBM36	0.0	2551.3(S)	Extracellular	0.858
Cphy1888	Pectin lyase	PL9	0.0	2334.6	Extracellular	0.72
Cphy1071	β-Mannanase	GH26, CBM3, CBM35	764.5(S)	1664.3(S)	Extracellular	0.68
Cphy1652	Chitinase	GH18, CBM50	0.0	1329.0(S)	Cytoplasmic	—
Cphy2128	β-Mannanase	GH26, CBM3, CBM35	153.7(S)	956.0(S)	Extracellular	0.78
Cphy1943	Chitinase	GH19	625.6(S)	875.4	Unknown	0.77

Protein ID, putative function, CAZy category, APEX expression, localization (PsortB v2.0), and significant signal peptide predictions (SignalP3.0-NN D-value > 0.45) are shown. APEX values followed by (S) are supernatant proteins.

phosphate as an attacking group. Phosphorolysis avoids ATP hydrolysis for direct phosphorylation of glucose during import, such as with the *Escherichia coli* phosphotransferase system. Although the ATP savings correlates with the length of the transported cellodextrin, uptake and phosphorolytic cleavage of glucans can reduce cellular ATP requirements by 25% (Zhang and Lynd, 2005).

Carbon metabolism

Because of the low ATP yield of anaerobic metabolism, *C. phytofermentans* needs to optimize energy output from carbon metabolism. In addition to cellodextrin phosphorylases, ATP is conserved by highly expressing reversible, pyrophosphate (PPi)-dependent glycolytic enzymes such as PPi-dependent phosphofructokinase and pyruvate phosphate dikinase. *C. phytofermentans* lacks the Ppa phosphatase used by *E. coli* to cleave PPi and instead uses it as a phosphate donor and source of a high-energy bond. PPi-dependent glycolytic enzymes can boost yield of glycolysis from two to five ATP (Slamovits and Keeling, 2006). Clostridia have low pyrophosphatase activities and high intracellular PPi levels (Heinonen and Drake, 1988), supporting a key role of PPi in their metabolism.

To maximize glycolytic turnover, reducing equivalents from glycolysis are consumed by reducing acetyl-CoA to ethanol with five Fe-dependent and one Zn-dependent alcohol dehydrogenase (ADH) proteins. Two Fe-dependent ADH, Cphy3925 and Cphy1029, were among the several most highly expressed proteins in all treatments. Cphy3925 shares 54% amino acid identity with *E. coli* AdhE, a bifunctional aldehyde/ADH that converts acetyl-CoA directly to ethanol.

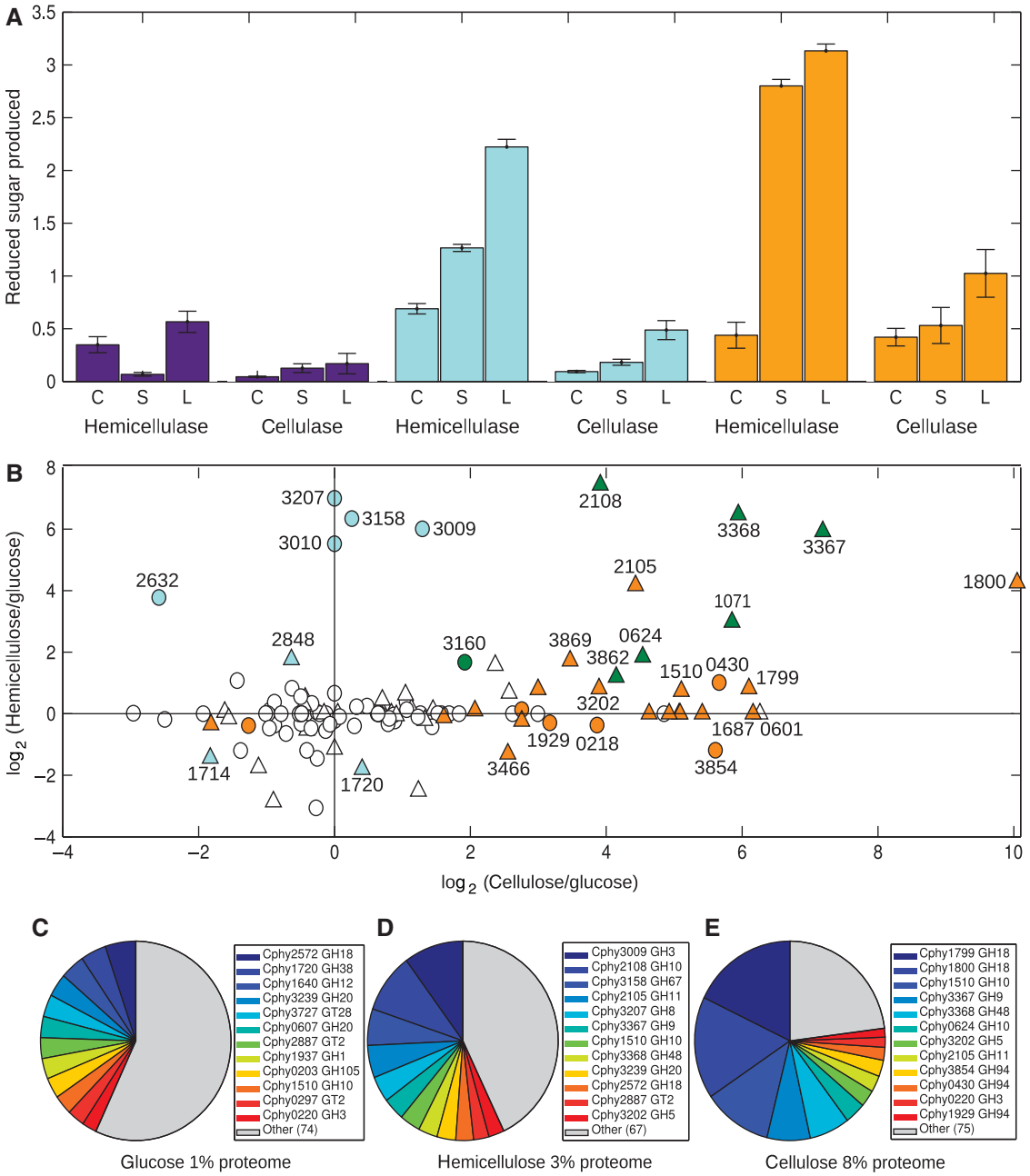


Figure 5 Carbohydrate-active enzyme (CAZy) expression and activities in glucose, hemicellulose, and cellulose cultures. (**A**) Secreted and cellular cellulolytic enzyme activities. Protein lysates from cultures grown on glucose (purple), hemicellulose (turquoise), or cellulose (orange) prepared from the cellular fraction (C), supernatant (S), or whole-culture lysates (L). Proteins were incubated with hemicellulose (hemicellulase assay) or carboxymethylcellulose substrate (cellulase assay), reducing sugars were assayed using dinitrosalicyclic acid, and were normalized to protein concentration. (**B**) CAZy expression changes (MS1 peak area ratio, MPA ratio) on hemicellulose and cellulose versus glucose showing differentially expressed proteins ($P<0.01$) on hemicellulose (turquoise), cellulose (orange), or both (green). Symbols show cellular proteins (circles) and supernatant proteins (triangles). (**C**–**E**) Shifts in the relative abundances of CAZy proteins in glucose (C) hemicellulose (D), and cellulose (E) treatments by Absolute Protein EXpression (APEX) show acclimation to different carbon sources. Fraction of proteome comprised of CAZy proteins in each treatment is shown.

The other highly expressed ADH, Cphy1029, shares 35% amino acid identity with *Zymomonas mobilis* AdhB, the primary ADH in this organism. *C. phytofermentans* also has a Zn-dependent ADH Cphy1179 that is homologous to the other *Z. mobilis* ADH, AdhA. This combination of ADHs facilitates the high ethanol yields (Figure 2). Another adaptation to increase glycolytic throughput is the high expression of two types of [FeFe]-hydrogenases that produce hydrogen by oxidizing reduced ferredoxin and NADH from glycolysis. Monomeric hydrogenases oxidize only ferredoxin, whereas the bifurcating hydrogenases oxidize both ferredoxin and NADH in a 1:1 ratio (Schut and Adams, 2009). All subunits for two multimeric, bifurcating hydrogenases were highly expressed on all carbon sources; a monomeric ferredoxin-dependent hydrogenase was expressed at a 10-fold lower level.

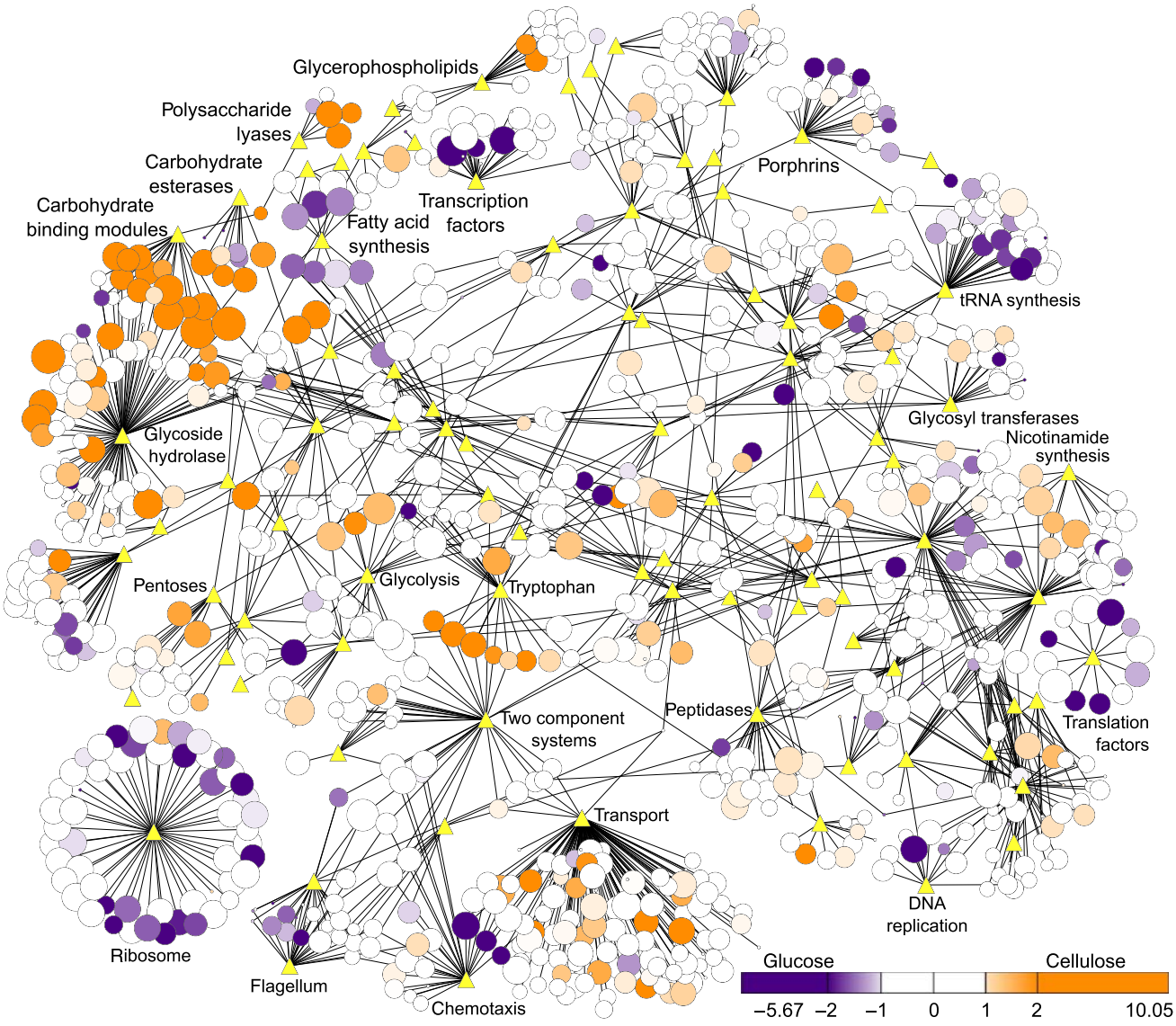


Figure 6 Proteome-wide expression changes on cellulose versus glucose visualized as a Cytoscape interaction network (Shannon *et al*, 2003). Nodes are proteins (circles) or KEGG/carbohydrate-active enzyme (CAZy) categories (yellow diamonds); edges are protein interactions defined by KEGG or CAZy databases. Protein node sizes show expression on cellulose as $\log_2$ (Absolute Protein EXpression, APEX). Node colors are expression changes as cellulose/glucose $\log_2$ protein ratios (MS1 peak area ratio, MPA ratio). Proteins less than twofold changed are white, higher on cellulose are graded orange, and higher on glucose are graded purple (see legend).

Proteome-wide interaction networks of proteins (circles) connected to CAZy and KEGG categories (triangles) show that expression changes on cellulose extend to diverse aspects of metabolism (Figure 6), while changes on hemicellulose (Supplementary Figure S14) occur mainly in pentose assimilation, the pentose phosphate pathway, CAZy, and transporters. Growth on cellulose results in repression of flagella, fatty acid synthesis, and many housekeeping functions (DNA replication, transcription, and translation). In addition to numerous transporters and carbohydratases, proteins upregulated on cellulose are involved in glycolysis, assimilation of alternative sugars, and nicotinamide synthesis to increase glycolytic flux.

The tryptophan biosynthesis proteins are upregulated on cellulose, which could increase ethanol tolerance similar to yeast (Hirasawa *et al*, 2007; Zhao and Bai, 2009) and enable production of hydrolases with tryptophan-rich carbohydrate-binding modules (CBM). CBM enhance catalysis by keeping carbohydratases in close association with their substrates (Tomme *et al*, 1988). Thirteen glycoside hydrolases with CBM were upregulated on cellulose. CBM proteins have elevated tryptophan contents relative to the proteome (Supplementary Figure S15) because CBM contact carbohydrates using a hydrophobic platform of tryptophan residues (Ponyi *et al*, 2000) that form van der Waals interactions with sugar rings (Lehtiö *et al*, 2003). The high expression of tryptophan synthesis proteins suggests that tryptophan may be an important factor limiting conversion of cellulose to ethanol.

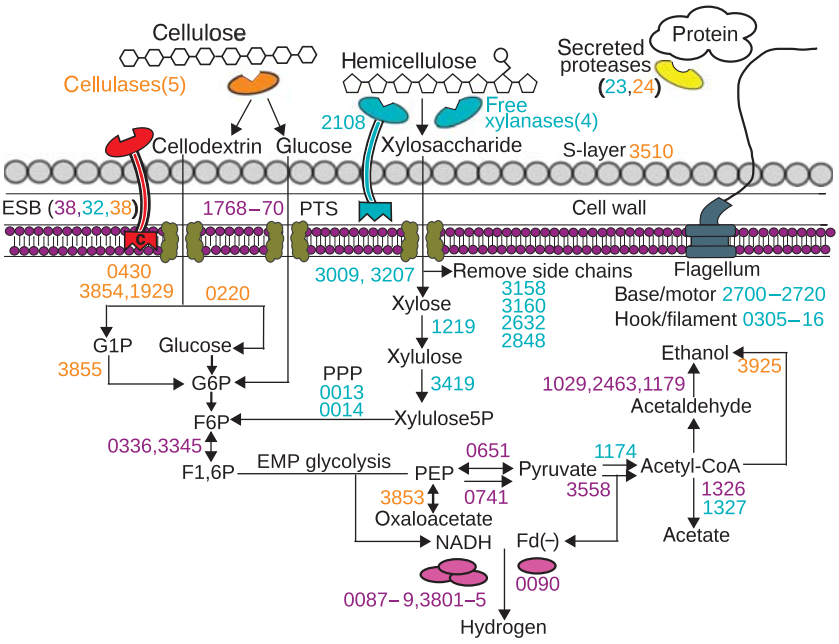


Figure 7 Model of the key secreted and intracellular proteins for the degradation and fermentation of plant biomass. Protein ID numbers are colored by highest Absolute Protein EXpression (APEX) expression on glucose (purple), hemicellulose (turquoise), cellulose (orange). Number in parentheses show the number of proteins of related function.

Discussion

Cellulolytic microbes such as *C. phytofermentans* reveal novel strategies for developing microbes to overcome the recalcitrance of cellulosic feedstocks, currently the main barrier to cellulosic biofuels (Houghton *et al*, 2006). We analyzed *C. phytofermentans* growth and fermentation of different cellulosic polymers (Figure 2), comprehensively quantified proteome changes that enabled fermentation of each substrate, and distilled the data into a model of cellulosic fermentation (Figure 7) showing key enzymes that can be engineered in this bacteria and other hosts to potentially optimize cellulosic biofuels. Growth on biomass involves secretion of numerous CAZy as well as proteins for motility, binding of extracellular solutes, proteolysis, and formation of a proteinaceous surface layer (Figure 4D). Oligosaccharides were uptaken before breakdown by intracellular CAZy, enabling more efficient sugar transport, conserving energy by phosphorolytic cleavage, and ensuring the sugar monomers were not available to competing microbes. Sugars were catabolized using an EMP glycolysis incorporating reversible, PPi-dependent glycolytic enzymes and fermented using pyruvate ferredoxin oxidoreductase and multiple ADHs. Growth on cellulose also resulted in broad metabolic changes such as increased tryptophan and nicotinamide synthesis and repression of fatty acid synthesis (Figure 6), which suggest ways to optimize cellulosic fermentation by supplementing growth media or genetic modification.

Cellulosic bioconversion can be accelerated by expression of cellulolytic enzymes in their native hosts as well as in model organisms such as yeast (Den Haan *et al*, 2007; Tsai *et al*, 2009) and *E. coli* (Steen *et al*, 2010). *C. phytofermentans* cellulolytic enzymes may be particularly well suited to enable cellulosic bioconversion by heterologous expression. Unlike other clostridia that package cellulolytic enzymes into a cellulosome, *C. phytofermentans* freely secretes these enzymes such that they can be functionally expressed without relying on a cellulosomal scaffold. However, considering the *C. phytofermentans* genome encodes 161 CAZy, it would take 12 880 experiments to find the most effective enzyme pair for any particular substrate. High throughput methods such as quantitative proteomics are thus needed to systematically quantify the enzymes used by cellulolytic microbes to metabolize different biomass types. In this study, we build upon proteomic studies of cellulosomes in other clostridia (Gold and Martin, 2007; Raman *et al*, 2009; Blouzard *et al*, 2010), to show how *C. phytofermentans* alters the stoichiometries of more than 100 CAZy (Figure 5) as well as diverse metabolic processes (Figure 6) when fermenting cellulosic substrates. By identifying the most highly upregulated and the secreted enzymes, we can prioritize targets for more efficient bioconversion of different types of biomass (Ito *et al*, 2010). This approach is supported in that Cphy3367, which is the second most highly upregulated CAZy on cellulose and was secreted into the supernatant (Figure 5B), has recently been shown to be essential for cellulolysis (Tolonen *et al*, 2009) and to solubilize cellulose *in vitro* (Zhang *et al*, 2010a). In this study, we focused on chemically defined substrates. In the future, we will apply these methods to untreated plant biomass composed of diverse polysaccharides and non-fermentable, soluble components.

In addition to cellulolytic enzymes, cellulosic biofuels can be optimized by more efficient fermentation. *C. phytofermentans* ferments cellulose to ethanol with high specificity (Figure 2F), supporting that accelerating cellulose deconstruction will give larger gains than streamlining end products.

This high fermentation efficiency also suggests heterologous expression of the alternative glycolytic enzymes and multiple ADHs from *C. phytofermentans* (Figure 7) could improve ethanol yields in other microbes. *C. phytofermentans* grows rapidly on hemicellulose (Figure 2B), but produces more acetate relative to ethanol (Figure 2E). Faster growth on hemicellulose than on glucose or xylose may result from faster uptake of xylosaccharides by high-affinity transporters, energy savings because of transport of oligosaccharides relative to sugar monomers (Muir *et al*, 1985), or the presence of an unidentified xylosaccharide phosphorylase. Higher relative acetate produced on hemicellulose than on cellulose is also supported by higher expression of acetate kinase, which could be inactivated to improve ethanol yields on hemicellulose-rich substrates. Additionally, acetate accumulation in hemicellulose cultures may partially result from cleavage of acetyl side chains from xylose subunits, which are common to hardwood xylans such as the birchwood used here (Shallom and Shoham, 2003).

Proteome-wide expression changes by ReDi proteomics are consistent with mRNA measurements by qRT–PCR (Figure 3F), but are advantageous in directly quantifying proteins and allowing discrimination of secreted and cellular enzymes. ReDi labeling is advantageous to stable isotope labeling by amino acids in cell culture (SILAC) (Ong and Mann, 2006), a widely used alternative approach for stable isotope incorporation, in not requiring strains with specific amino acid auxotrophies or optimization of growth on synthetic medium. As ReDi uses inexpensive, highly quantitative chemistry and a small mass tag, it is easily incorporated into routine proteomics protocols (Boersema *et al*, 2009). ReDi proteomics will be a key part of research to enable cellulosic biofuels and can be applied to many other research areas as well.

Materials and methods

Growth and HPLC

C. phytofermentans ISDg ATCC 700394 was cultured anaerobically at 32°C in GS2 medium (Johnson *et al*, 1981) with either 3 g l^{-1} glucose, 3 g l^{-1} birch wood hemicellulose xylan (Sigma X0502, >90% poly-beta-1,4-D-xylopyranose), or 10 g l^{-1} cellulose (0.5 × 5 cm strips of Whatman #1 filter paper cat 1001-090, cellulose content >98%). Degradation of cellulose was quantified as the dry mass of cellulose remaining in culture. Cellulose was collected on 11 μ filters by vacuum filtration, dried overnight at 65°C, and weighed. Growth in glucose and hemicellulose cultures was quantified as OD_{600} (OD_{600}=1.1 is 10^9 cells ml^{-1}). Fermentation products were quantified by HPLC using a Biorad aminex HPX-87H 300 × 7.8 mm column maintained at 60°C with a Waters 2414 refractive index detector (see Supplementary Information for growth and fermentation calculations).

Electron microscopy

Scanning electron microscopy samples were bound to glass cover slips with 0.1% poly-L-lysine and incubated for 15 min in primary fixative (2.5% glutaraldehyde, 2.5% formaldehyde, and 0.1 M sodium cacodylate). Samples were rinsed for 30 min in 0.1 M sodium cacodylate and immersed for 15 min in 1% osmium tetroxide, and 0.1 M sodium cacodylate (pH 7.2). Cells were dehydrated with a 30–100% ethanol graded series, critical point dried, and sputter coated. The samples were examined on a Hitachi S-4800 FE-SEM. Cells for transmission electron microscopy were pelleted and suspended in primary fixative for 30 min at room temperature. The fixed cells were washed in water, embedded in 2% agarose, rinsed with 0.1 M sodium cacodylate, and post-fixed (1% osmium tetroxide, and 0.1 M cacodylate, pH 7.4) for 2 h. Samples were dehydrated with a 30–100% ethanol graded series and polymerized in Spurrs/Quetol at 55C for 24 h. Agarose was sliced (<90 nm) and visualized on a JEOL JEM-1010 TEM.

Quantitative proteomics

Culture protein lysates were prepared by French press of glucose and hemicellulose cultures in mid-log phase and from cellulose cultures after 2 weeks of growth. Supernatant proteins were isolated by three rounds of centrifugation (5k, 15 min) with transfer of the supernatant to a fresh tube, followed by 0.1 μ filtration. Proteins were precipitated with $\frac{1}{4}$ volume of 100% (w/w) trichloroacetic acid (TCA) and incubated for 60 min at 37°C in 1% SDS, 0.2 M NaOH, and 10 mM DTT. Cysteines were alkylated with 30 mM iodoacetamide at room temperature in the dark for 60 min. Proteins were again TCA precipitated and resuspended in 50 mM Tris–HCl (pH 7.6), 1 M urea and digested overnight at 37°C with sequencing grade trypsin (Promega, WI, USA) in a ~50:1 substrate-to-enzyme ratio.

Peptide solutions were acidified with trifluoroacetic acid (TFA) to a final concentration of 0.5% and 500 μg of peptides were subjected to C18 reversed-phase extraction using a 100 mg Sep-Pak cartridge (Waters, MA). Following sample loading and a 4 ml wash step with 0.1% TFA, the cartridge was washed with 2 ml of 0.2 M citric acid/sodium phosphate buffer (pH 5.5) and peptides were labeled through ReDi with 8 ml of 0.8% formaldehyde (D_2 labeled or in normal isotopic distribution, Sigma, MO) and 0.12 M sodium cyanoborohydride (Sigma) or sodium cyanoborodeuteride (CDN Isotopes, Canada) for 8 min. After washing with 0.1% TFA and 0.5% acetic acid AcOH, labeled peptides were eluted with 40% acetonitrile (ACN), 0.5% AcOH and 80% ACN, 0.5% AcOH. Light and heavy peptides were mixed and separated into 16 fractions by immobilized pH gradient isoelectric focusing (IPG-IEF) on a 18 cm pH 3–10 IPG strip using a IPGphor apparatus (Chick *et al*, 2008) (GE Healthcare, WI). Peptides were desalted using C_{18}-StageTips (Rappsilber *et al*, 2003) and analyzed by microcapillary lLC-MS/MS using a hybrid dual pressure linear ion trap/Orbitrap mass spectrometer (LTQ Orbitrap Velos, Thermo Scientific, Bremen, Germany) essentially as described (Haas *et al*, 2006). Peptides were dissolved in 5% formic acid (FA) and 5% ACN and ~1 μg was separated on a 100 μm × 20 cm C_{18}-reversed phase (Magic C18AQ, Michrom Bioresources) column applying a gradient from 10% ACN to 32% ACN in 0.125% FA over a 75 or 100 min gradient at a flow rate of ~300 nl min^{-1}. The mass spectrometer was operated in a data-dependent mode with a full MS scan (resolution, 60 000) acquired in the Orbitrap analyzer followed by linear ion trap MS/MS spectra on the 20 most abundant ions detected in the full MS spectrum. Automatic gain control (AGC) targets were 3×10^6 for the full MS and 2000 for MS/MS, maximum ion accumulation times were set to 1000 ms (MS) and 150 ms (MS/MS). Fragmented peptide precursor ions were excluded from further selection for MS/MS for 20–60 s (see RAW data). The RAW data files have been deposited in Proteome Commons (http://www.proteomecommons.org) and can be downloaded using the following hash key:

3n++Ey/FMTRCCMvwxPAdNuSTL2VVGaWyODpYlSLqyk24mYz
CVhaLcFsvcdRvtDi83gNOKYUhPPn84zLjRdBjWA2kUIgAAAAAAAB
FUw==.

Peptides were identified from MS/MS spectra using SEQUEST and were filtered to a 1% MS/MS spectra assignment false discovery rate using a target-decoy sequence database (Elias and Gygi, 2010) of the 3926 *C. phytofermentans* (NCBI NC_010001.faa) proteins, common contaminants such as trypsin and human keratins, and a decoy component of the reversed sequences of these proteins. Absolute protein abundances within each treatment were estimated from MS/MS spectral counts using APEX, a machine learning-supported analysis based on spectral counting (see APEX section of Supplementary Information). Extracted ion current peak areas (MS1 peaks) of light and heavy versions of the same peptide were compared to determine their relative concentration differences using VISTA (Bakalarski *et al*, 2008). Protein expression changes were calculated

as the median peptide concentration change, here called the median MPA ratio. Significant differential expression was assessed using aggregate Z-scores for all peptides in a protein (see ReDi section of Supp Info). Peptide pairs were quantified only when their average signal-to-noise ratio was above five and proteins for which exclusively light or heavy peptides were detected are reported only if their identification was based on at least two unique peptides. After this additional filtering, the protein false discovery rate for each experiment was lower than 5%. Expression measurements are from individual cultures (APEX) or culture comparisons (ReDi). Replicate experiments were run to establish high reproducibility between biological replicates for APEX (Supplementary Figure S10) and for ReDi labeling of biological duplicates (Figure 3D), treatment comparisons (Figure 3E), and differentially labeled samples mixed at various ratios (Supplementary Figure S11).

DNS reduced sugar assay

Enzymatic activities of protein lysates against cellulosic substrates were quantified by 3,5-dinitrosalicylic acid (DNS) assay (Miller, 1959). Proteins from glucose, hemicellulose, and cellulose cultures were assayed as French press lysates of cellular pellets resuspended in fresh medium, whole-culture lysates, or supernatant fractions. Proteins were incubated with either 0.1% hemicellulose xylan or carboxymethylcellulose for 15 min at 50°C before addition of DNS (Sigma D0550) and boiling for 15 min. Reduced sugars were quantified colorimetrically as OD_{600}. Reduced sugar equivalents were calibrated using standard curves of glucose (for carboxymethylcellulose) or xylose (for hemicellulose). Enzyme activities were normalized to protein concentrations by BCA assay (Pierce product 23227).

qRT–PCR

Cells were collected by centrifugation for mRNA analysis from cultures at the same time points as protein samples. A Ribopure Bacteria Kit (Ambion AM1925) was used to extract RNA; remaining DNA was removed using Dnase I (Ambion AM2222). One microgram RNA was reverse transcribed to single-stranded DNA using the Superscript First Strand cDNA Synthesis Kit (Invitrogen 11904018). Real-time PCR amplification was conducted using a MJ Research DNA Engine Opticon II machine by monitoring incorporation of SYBR green I (Invitrogen S7563). PCR primers (Supplementary Table SIV) amplified 100 bp products from each gene. Relative gene expression was quantified using the comparative C_T method (Schmittgen and Livak, 2008) with the 16S ribosomal sequence serving as the internal control gene (Supplementary Table SV). Expression values are the means of duplicate measurements from duplicate cultures.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

We thank B Fowle and W Curtis for technical expertize and S Leschine, T Warnick and J Blanchard for help and sharing *C. phytofermentans* cultures. This work was funded by DOE GtL grant DE-FG02-02ER-63445 to GMC and an SRA with Qteros Inc.

Author Contributions: ACT, WH, JA, SPG, and GMC designed experiments. ACT, WH, and ACC conducted experiments. ACT, WH, and JA analyzed the data. ACT, WH, and JA wrote the paper.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Bakalarski CE, Elias JE, Villén J, Haas W, Gerber SA, Everley PA, Gygi SP (2008) The impact of peptide abundance and dynamic range on stable-isotope-based quantitative proteomic analyses. *J Proteome Res* **7:** 4756–4765

Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* **340:** 783–795

Blouzard JC, Coutinho PM, Fierobe HP, Henrissat B, Lignon S, Tardif C, Pagès S, de Philip P (2010) Modulation of cellulosome composition in *Clostridium cellulolyticum*: adaptation to the polysaccharide environment revealed by proteomic and carbohydrate-active enzyme analyses. *Proteomics* **10:** 541–554

Boersema PJ, Raijmakers R, Lemeer S, Mohammed S, Heck AJR (2009) Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat Protoc* **4:** 484–494

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

Chick JM, Haynes PA, Molloy MP, Bjellqvist B, Baker MS, Len AC (2008) Characterization of the rat liver membrane proteome using peptide immobilized pH gradient isoelectric focusing. *J Proteome Res* **7:** 1036–1045

Den Haan R, Rose SH, Lynd LR, van Zyl WH (2007) Hydrolysis and fermentation of amorphous cellulose by recombinant *Saccharomyces cerevisiae*. *Metab Eng* **9:** 87–94

Desvaux M, Khan A, Scott-Tucker A, Chaudhuri RR, Pallen MJ, Henderson IR (2005) Genomic analysis of the protein secretion systems in *Clostridium acetobutylicum* ATCC 824. *Biochim Biophys Acta* **1745:** 223–253

Divne C, Ståhlberg J, Reinikainen T, Ruohonen L, Pettersson G, Knowles JK, Teeri TT, Jones TA (1994) The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. *Science* **265:** 524–528

Elias JE, Gygi SP (2010) Target-decoy search strategy for mass spectrometry-based proteomics. *Methods Mol Biol* **604:** 55–71

Etienne-Toumelin I, Sirard JC, Duflot E, Mock M, Fouet A (1995) Characterization of the *Bacillus anthracis* S-layer: cloning and sequencing of the structural gene. *J Bacteriol* **177:** 614–620

Falkowski P, Scholes RJ, Boyle E, Canadell J, Canfield D, Elser J, Gruber N, Hibbard K, Högberg P, Linder S, Mackenzie FT, Moore III B, Pedersen T, Rosenthal Y, Seitzinger S, Smetacek V, Steffen W (2000) The global carbon cycle: a test of our knowledge of earth as a system. *Science* **290:** 291–296

Gardy JL, Laird MR, Chen F, Rey S, Walsh CJ, Ester M, Brinkman FS (2005) PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics* **21:** 617–623

Gold ND, Martin VJJ (2007) Global view of the *Clostridium thermocellum* cellulosome revealed by quantitative proteomic analysis. *J Bacteriol* **189:** 6787–6795

Haas W, Faherty BK, Gerber SA, Elias JE, Beausoleil SA, Bakalarski CE, Li X, Villén J, Gygi SP (2006) Optimization and use of peptide mass measurement accuracy in shotgun proteomics. *Mol Cell Proteomics* **5:** 1326–1337

Heinonen JK, Drake HL (1988) Comparative assessment of inorganic pyrophosphate and pyrophosphatase levels of *Escherichia coli, Clostridium pasteurianum*, and *Clostridium thermoaceticum*. *FEMS Microbiol Lett* **52:** 205–208

Hirasawa T, Yoshikawa K, Nakakura Y, Nagahisa K, Furusawa C, Katakura Y, Shimizu H, Shioya S (2007) Identification of target genes conferring ethanol stress tolerance to *Saccharomyces cerevisiae* based on DNA microarray analysis. *J Biotechnol* **131:** 34–44

Houghton J, Weatherwax S, Ferrell J (2006) Breaking the biological barriers to cellulosic ethanol. *US Dept of Energy Research Roadmap*. pp 1–206 (http://www.genomicscience.energy.gov/biofuels/b2bworkshop.shtml#page = news)

Hsu J, Huang S, Chow N, Chen S (2003) Stable-isotope dimethyl labeling for quantitative proteomics. *Anal Chem* **75:** 6843–6852

Ito J, Petzold CJ, Mukhopadhyay A, Heazlewood JL (2010) The role of proteomics in the development of cellulosic biofuels. *Curr Proteomics* **7:** 121–134

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

Lehtiö J, Sugiyama J, Gustavsson M, Fransson L, Linder M, Teeri TT (2003) The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules. *Proc Natl Acad Sci USA* **100:** 484–489

Leschine SB (1995) Cellulose degradation in anaerobic environments. *Annu Rev Microbiol* **49:** 399–426

Lu P, Vogel C, Wang R, Yao X, Marcotte EM (2007) Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol* **25:** 117–124

Lu Y, Zhang YH, Lynd LR (2006) Enzyme-microbe synergy during cellulose hydrolysis by *Clostridium thermocellum*. *Proc Natl Acad Sci USA* **103:** 16165–16169

Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS (2002) Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev* **66:** 506–577

Miller GL (1959) Use of dinitrosalicylic acid reagent for determination of reducing suga. *Anal Chem* **31:** 426

Morrison M, Miron J (2000) Adhesion to cellulose by *Ruminococcus albus*: a combination of cellulosomes and Pil-proteins? *FEMS Microbiol Lett* **185:** 109–115

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

Ong S, Mann M (2006) A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nat Protoc* **1:** 2650–2660

Perlack RD, Wright LL, Turhollow A, Graham RL, Stokes B, Erbach DC (2005) Biomass as a feedstock for a bioenergy and bioproducts industry: the technical feasibility of a billion-ton annual supply. *A Joint Study by the US Dept of Energy and the US Dept of Agriculture*. pp 1–78 (http://www.ornl.gov/~webworks/cppr/y2001/rpt/123021.pdf)

Ponyi T, Szabó L, Nagy T, Orosz L, Simpson PJ, Williamson MP, Gilbert HJ (2000) Trp22, Trp24, and Tyr8 play a pivotal role in the binding of the family 10 cellulose-binding module from *Pseudomonas* xylanase A to insoluble ligands. *Biochemistry* **39:** 985–991

Raman B, Pan C, Hurst GB, Rodriguez Jr M, McKeown CK, Lankford PK, Samatova NF, Mielenz JR. (2009) Impact of pretreated Switchgrass and biomass carbohydrates on *Clostridium thermocellum* ATCC 27405 cellulosome composition: a quantitative proteomic analysis. *PLoS ONE* **4:** e5271

Rappsilber J, Ishihama Y, Mann M (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* **75:** 663–670

Sára M, Sleytr UB (1987) Molecular sieving through S layers of *Bacillus stearothermophilus* strains. *J Bacteriol* **169:** 4092–4098

Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc* **3:** 1101–1108

Schut GJ, Adams MWW (2009) The iron-hydrogenase of *Thermotoga maritima* utilizes ferredoxin and NADH synergistically: a new perspective on anaerobic hydrogen production. *J Bacteriol* **191:** 4451–4457

Shallom D, Shoham Y (2003) Microbial hemicellulases. *Curr Opin Microbiol* **6:** 219–228

Slamovits CH, Keeling PJ (2006) Pyruvate-phosphate dikinase of oxymonads and parabasalia and the evolution of pyrophosphate-dependent glycolysis in anaerobic eukaryotes. *Eukaryot Cell* **5:** 148–154

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13:** 2498–2504

Sleytr UB, Messner P (1983) Crystalline surface layers on bacteria. *Annu Rev Microbiol* **37:** 311–339

Sleytr UB, Messner P (1988) Crystalline surface layers in procaryotes. *J Bacteriol* **170:** 2891–2897

Somerville C (2006) The billion-ton biofuels vision. *Science* **312:** 1277

Steen EJ, Kang Y, Bokinsky G, Hu Z, Schirmer A, McClure A, Del Cardayre SB, Keasling JD (2010) Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* **463:** 559–562

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

Tomme P, Van Tilbeurgh H, Pettersson G, Van Damme J, Vandekerckhove J, Knowles J, Teeri T, Claeyssens M. (1988) Studies of the cellulolytic system of *Trichoderma reesei* QM 9414. Analysis of domain function in two cellobiohydrolases by limited proteolysis. *Eur J Biochem* **170:** 575–581

Tsai S, Oh J, Singh S, Chen R, Chen W (2009) Functional assembly of minicellulosomes on the *Saccharomyces cerevisiae* cell surface for cellulose hydrolysis and ethanol production. *Appl Environ Microbiol* **75:** 6087–6093

von Heijne G (1989) The structure of signal peptides from bacterial lipoproteins. *Protein Eng* **2:** 531–534

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

Zhang YH, Lynd LR (2005) Cellulose utilization by *Clostridium thermocellum*: bioenergetics and hydrolysis product assimilation. *Proc Natl Acad Sci USA* **102:** 7321–7325

Zhang X, Sathitsuksanoh N, Zhang YH (2010a) Glycoside hydrolase family 9 processive endoglucanase from *Clostridium phytofermentans*: heterologous expression, characterization, and synergy with family 48 cellobiohydrolase. *Bioresour Technol* **101:** 5534–5538

Zhang XZ, Zhang Z, Zhu Z, Sathitsuksanoh N, Yang Y, Zhang YH (2010b) The noncellulosomal family 48 cellobiohydrolase from *Clostridium phytofermentans* ISDg: heterologous expression, characterization, and processivity. *Appl Microbiol Biotechnol* **86:** 525–533

Zhao XQ, Bai FW (2009) Mechanisms of yeast stress tolerance and its manipulation for efficient fuel ethanol production. *J Biotechnol* **144:** 23–30

Zhou M, Boekhorst J, Francke C, Siezen RJ (2008) LocateP: genome-scale subcellular-location predictor for bacterial proteins. *BMC Bioinformatics* **9:** 173

Aim 3: *Lachnospiraceae* dynamics in complex communities

Studies are increasingly finding that certain species of gut *Lachnospiraceae* promote healthy intestinal and immune function. In particular, administration of live *Lachnospiraceae* in pre-clinical *in vivo* models and human clinical studies support their potential to mitigate health problems including infection, inflammation, allergies, insulin insensitivity, and cancer [1] [2] [3] [4]. In order to maximize the health-promoting properties of gut *Lachnospiraceae*, we need a greater understanding of interventions that promote their growth and in the context of complex gut communities. To this end, we designed a community of species spanning the phylogenetic diversity of the human gut microbiome, which we are using to study microbiome dynamics in response to environmental perturbations (Table 1). In addition, at Kaleido Biosciences we designed synthetic glycans as chemical interventions that remodel the composition of the gut microbiome to favor *Lachnospiraceae,* and tested their therapeutic potential in mouse models of intestinal inflammation [5]. Below is a summary of our research interests in these areas, followed by a published study highlighting our progress to date.

Aim 3a: Microbiome structure and gene expression in response to antibiotic treatment

Antibiotics are critical tools in modern medicine, effectively combating bacterial infections and saving countless lives. However, the impact of antibiotics on the composition of the gut microbiome is complex. While antibiotics target harmful bacteria, they can also inadvertently harm health-promoting commensals, including *Lachnospiraceae*. We are examining the microbiome response to antibiotic treatment by quantifying changes in the transcription and composition of an artificial gut community upon treatment with ciprofloxacin. Ciprofloxacin is a broad-spectrum, fluoroquinolone antibiotic that inhibits bacterial replication by inhibiting enzymes essential in DNA replication, namely the DNA gyrase (a type II DNA topoisomerase) and DNA topoisomerase IV.

Bacteria	Phylum	NCBI taxid
Bifidobacterium adolescentis E298b (Variant c)	Actinobacteria	1680
Bifidobacterium angulatum DSM 20098	Actinobacteria	518635
Bifidobacterium catenulatum DSM 16992	Actinobacteria	566552
Bifidobacterium longum sub. Infantalis ATCC 15697	Actinobacteria	391904
Bifidobacterium pseudocatenulatum DSM 20438	Actinobacteria	547043
Collinsella aerofaciens JCM 7790	Actinobacteria	74426
Akkermansia muciniphila ATCC BAA-835	Bacteroidetes	349741
Bacteroides cellulosilyticus WH2	Bacteroidetes	1268240
Bacteroides coprophilus DSM 18228	Bacteroidetes	547042
Bacteroides finegoldii CL09T03C10	Bacteroidetes	997888
Parabacteroides merdae CL09T00C40	Bacteroidetes	999421
Prevotella copri DSM 18205	Bacteroidetes	537011
Bacteroides caccae ATCC 43185	Bacteroidetes	411901
Bacteroides thetaiotaomicron VPI-5482	Bacteroidetes	226186
Bacteroides uniformis ATCC 8492	Bacteroidetes	411479
Bacteroides vulgatus ATCC 8482	Bacteroidetes	435590
Odoribacter splanchnicus DSM 20712	Bacteroidetes	709991
Parabacteroides distasonis ATCC 8503	Bacteroidetes	435591
Clostridium symbiosum WAL-14163	Firmicutes	742740
Anaerobutyricum hallii DSM 3353	Firmicutes	411469
Blautia coccoides YL58	Firmicutes	1532
Blautia hansenii DSM 20583	Firmicutes	537007
Blautia hydrogenotrophica ATCC BAA-2371	Firmicutes	53443
Blautia obeum ATCC 29174	Firmicutes	411459
Blautia producta ATCC 27340	Firmicutes	1121114
Clostridium celatum DSM 1785	Firmicutes	545697
Clostridium scindens ATCC 35704	Firmicutes	411468
Coprococcus catus VPI C6-61 [NCTC 11835]	Firmicutes	116085
Dorea formicigenerans ATCC 27755	Firmicutes	411461
Dorea longicatena DSM 13814	Firmicutes	411462
Eubacterium eligens ATCC 27750	Firmicutes	515620
Eubacterium rectale ATCC 33656	Firmicutes	515619
Eubacterium ventriosum ATCC 27560	Firmicutes	411463
Faecalibacterium prausnitzii VPI C13-51	Firmicutes	853
Holdemanella biformis DSM 3989	Firmicutes	518637
Lachnospira multipara ATCC 19207	Firmicutes	1282887
Lactobacillus casei subsp casei ATCC 393	Firmicutes	1423732
Roseburia intestinalis DSM 14610	Firmicutes	166486
Ruminococcus gnavus AGR2154	Firmicutes	1384063
Tyzzerella nexilis DSM 1787	Firmicutes	500632
Clostridioides difficile 630	Firmicutes	272563
Enterocloster bolteae ATCC BAA-613	Firmicutes	411902
Enterococcus faecium ATCC 700221	Firmicutes	1352
Escherichia coli ATCC BAA-97	Proteobacteria	562
Succinivibrio dextrinosolvens DSM 3072	Proteobacteria	1123324
Klebsiella pneumoniae ATCC 33259	Proteobacteria	72407
Salmonella enterica ATCC 27869	Proteobacteria	108619

Table 1 Bacterial species in the synthetic gut microbiome community. All species can be cultivated independently and have sequenced genomes.

We used a multi-omic approach to analyze transcriptional responses and community composition changes in response to ciprofloxacin treatment. We profiled mRNA expression and mapping of transcription start sites (TSSs) to study the functional response as well as 16S and RIMS-seq (shotgun sequencing) to study the community-wide composition changes. Overall, the 16S and RIMS-seq data that ciprofloxacin treatment results in a decrease in Proteobacteria and Bacteroidetes and an increase in Firmicutes. Among the 47 bacteria present in the community, 15 bacteria

were significantly more abundant and 16 were significantly less abundant following ciprofloxacin treatment. The species with greatest increases following ciprofloxacin treatment include *Lachnospiraceae* such as *Ruminococcus gnavus*, *Tyzzerella nexilis, Enterocloster bolteae, Clostridium scindens, Dorea formicigenerans, and* 4 species of *Blautia* (Fig 2A). Proteobacteria (*E. coli, Salmonella, Klebsiella pneumoniae*) and Bacteroidetes *(Odoribacter splanchnicus, Parabacteroides merdae, Bacteroides finegoldii, Prevotella copri*) decreased in relative abundance after ciprofloxacin addition.

Our preliminary analysis shows that species with the greatest reduction in relative abundance after 48h also showed the greatest number of differentially expressed genes at 5 min (Fig 2B). Currently, we are analyzing these gene expression changes in detail to see if there are biomarker genes whose rapid changes in gene expression can be use to predict longer-term changes in community composition (Fig 2B).

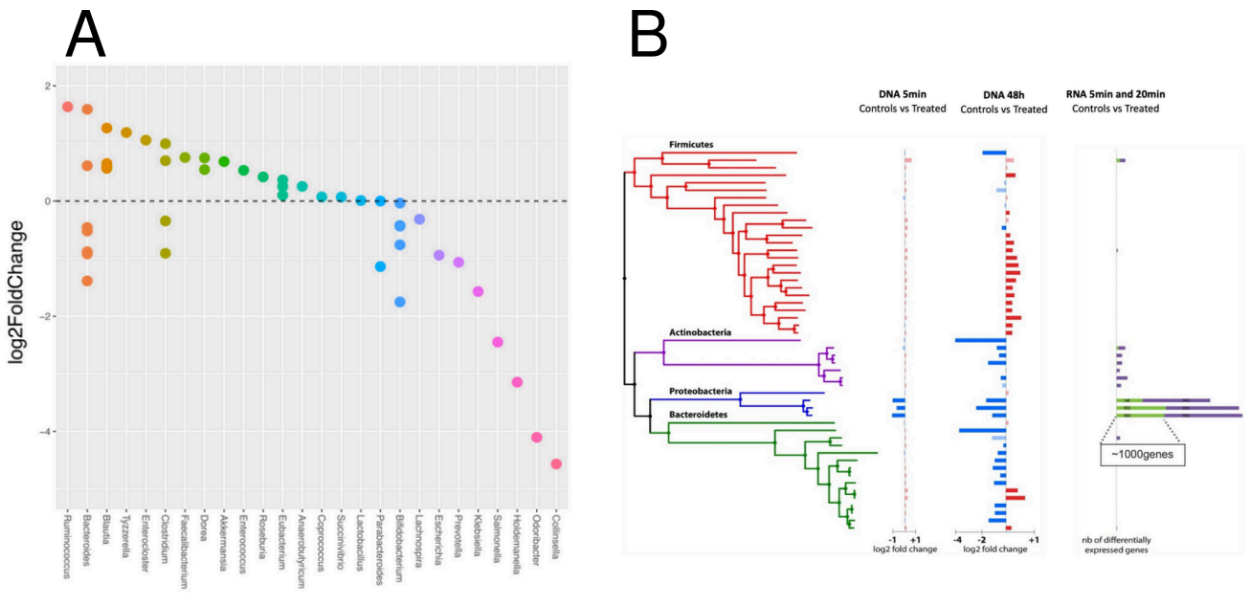


Fig 2 (A) Changes in the relative abundance (log_2 fold change) of bacterial genera 48h after ciprofloxacin treatment as measured by metagenomic sequencing. (B) Comparison of changes in relative abundance by metagenomics and number of differentially expressed genes measured by RNAseq.

Aim 3b: Synthetic glycans that promote *Lachnospiraceae* alleviate colitis

Metabolism of complex glycans by the gut microbiome plays a critical role in promoting health [6], leading to interest in developing glycan therapeutics that act by modulating the composition and output of the gut microbiome. We synthesized a library of hundreds of synthetic glycans (SGs) that span the chemical diversity of dietary fiber and we evaluated their fermentation properties by fecal communities, leading to identification of SGs that promote health-associated bacterial taxa such as *Lachnospiraceae* and production of beneficial metabolites [5]. We also showed that SGs are poor growth substrates for model enteric pathogens in comparison to reference glycans, including prebiotics. We demonstrated the *in vivo* therapeutic potential of *Lachnospiraceae*-promoting SGs in two mouse models of colitis: DSS-mediated colitis and *C. difficile* infection. SG treatment reduced DSS-mediated weight loss and improved scores of diarrhea, endoscopy, and colonic histology. In the *C. difficile* infection model, SG treatment significantly reduced weight loss, increased survival, and improved clinical scores.

Delivery of rationally-optimized, complex glycans is a promising option to drive the composition of the microbiota toward health-promoting states by leveraging taxa-specific differences in glycan metabolism, one of the dominant factors shaping gut microbiome composition [6]. Our study demonstrates SGs are a novel, glycan-based modality to manipulate the properties of human gut communities with flexibility and precision. In particular, SGs that promote the abundance of *Lachnospiraceae* show therapeutic potential in mouse models of chemical and pathogen-induced intestinal inflammation. Our future work will explore additional modalities to manage disease using prebiotic interventions that promote the growth of *Lachnospiraceae*.

References

1. Caballero S, Kim S, Carter RA, Leiner IM, Sušac B, Miller L, et al. Cooperating Commensals Restore Colonization Resistance to Vancomycin-Resistant Enterococcus faecium. Cell Host Microbe. 2017;21: 592-602.e4. doi:10.1016/j.chom.2017.04.002
2. Shen Z, Luo W, Tan B, Nie K, Deng M, Wu S, et al. Roseburia intestinalis stimulates TLR5-dependent intestinal immunity against Crohn's disease. eBioMedicine. 2022;85.

doi:10.1016/j.ebiom.2022.104285
3. Feehley T, Plunkett CH, Bao R, Choi Hong SM, Culleen E, Belda-Ferre P, et al. Healthy infants harbor intestinal bacteria that protect against food allergy. Nat Med. 2019;25: 448–453. doi:10.1038/s41591-018-0324-z
4. Liu N, Chen L, Yan M, Tao Q, Wu J, Chen J, et al. Eubacterium rectale Improves the Efficacy of Anti-PD1 Immunotherapy in Melanoma via l-Serine-Mediated NK Cell Activation. Research. 2023;6: 0127. doi:10.34133/research.0127
5. Tolonen AC, Beauchemin N, Bayne C, Li L, Tan J, Lee J, et al. Synthetic glycans control gut microbiome structure and mitigate colitis in mice. Nat Commun. 2022;13: 1244. doi:10.1038/s41467-022-28856-x
6. Patnode ML, Beller ZW, Han ND, Cheng J, Peters SL, Terrapon N, et al. Interspecies Competition Impacts Targeted Manipulation of Human Gut Bacteria by Fiber-Derived Glycans. Cell. 2019;179: 59-73.e13. doi:10.1016/j.cell.2019.08.011

nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-022-28856-x OPEN

Synthetic glycans control gut microbiome structure and mitigate colitis in mice

Andrew C. Tolonen[1✉], Nicholas Beauchemin[1,6], Charlie Bayne[1], Lingyao Li[1], Jie Tan[1], Jackson Lee[1], Brian M. Meehan[1,7], Jeffrey Meisner[1], Yves Millet[1], Gabrielle LeBlanc[1], Robert Kottler[2], Erdmann Rapp[2,3], Chris Murphy[1,8], Peter J. Turnbaugh[4], Geoffrey von Maltzahn[1,5], Christopher M. Liu[1,9] & Johan E. T. van Hylckama Vlieg[1✉]

Relative abundances of bacterial species in the gut microbiome have been linked to many diseases. Species of gut bacteria are ecologically differentiated by their abilities to metabolize different glycans, making glycan delivery a powerful way to alter the microbiome to promote health. Here, we study the properties and therapeutic potential of chemically diverse synthetic glycans (SGs). Fermentation of SGs by gut microbiome cultures results in compound-specific shifts in taxonomic and metabolite profiles not observed with reference glycans, including prebiotics. Model enteric pathogens grow poorly on most SGs, potentially increasing their safety for at-risk populations. SGs increase survival, reduce weight loss, and improve clinical scores in mouse models of colitis. Synthetic glycans are thus a promising modality to improve health through selective changes to the gut microbiome.

[1] Kaleido Biosciences, Lexington, MA 02421, USA. [2] glyXera GmbH, 39120 Magdeburg, Germany. [3] Max Planck Institute for Dynamics of Complex Technical Systems, 39106 Magdeburg, Germany. [4] Department of Microbiology and Immunology, University of California San Francisco, San Francisco, CA 94143, USA. [5] Flagship Pioneering, Cambridge, MA 02142, USA. [6] Present address: Seres Therapeutics, Cambridge, MA 02139, USA. [7] Present address: Pareto Bio, Cambridge, MA 02140, USA. [8] Present address: Bacainn Therapeutics, Inc and Morningside BioPharma Advisory, Concord, MA 01742, USA. [9] Present address: Exo Therapeutics, Watertown, MA 02472, USA. ✉email: tolonen@alum.mit.edu; johan.van-hylckama-vlieg@kaleido.com

Gut microbiome composition and metabolic output have been associated with initiation and progression of diseases including auto-immune diseases, cancer, metabolic syndrome, and liver disease[1–5]. Targeted manipulation of the microbiome is thus a promising therapeutic strategy being pursued by numerous public and private initiatives to treat disease[6]. In particular, glycan intervention is a chemical-based approach to alter the microbiome that leverages how ecological niches in the gut are occupied by species that are specialized to metabolize different carbon sources[7,8]. Alimentary interventions with complex glycans comprising dietary fiber significantly shift the composition and output of the gut microbiome within a few days and have been linked to prevention of type 2 diabetes, obesity, and cancer[9–12].

Glycans with documented health benefits currently marketed as prebiotics include fructo-oligosaccharides (FOS), galacto-oligosaccharides (GOS), xylo-oligosaccharides (XOS), pullulan, and lactulose[13–17]. These glycans are extracted from agricultural materials or enzymatically synthesized and lack the structural complexity of dietary fiber, which is composed of a matrix of β-glucans, starches, hemicelluloses, and pectins with diverse monosaccharides and glycosidic bonds. To exploit the potential of glycans to modulate the microbiome, there is a need for novel molecules spanning the chemical and structural diversity of dietary glycans that can be efficiently and consistently produced. By controlling the starting materials and reaction conditions, we synthesized a library of hundreds of synthetic glycans (SGs) with different monosaccharides, bond types, and degrees of polymerization (DPs)[18,19]. These SGs enable a wide range of targeted changes to the microbiome and potentially open new avenues for the prevention and treatment of disease.

In this study, the fermentation of hundreds of SGs polymerized from diverse, naturally occurring monosaccharides are compared to reference glycans using an ex vivo platform for highly multiplexed measurements of growth, metabolic output, and taxonomy. Based on the ex vivo assays, we select SGs for chemical structural analyses of their polymerization profiles and glycosidic linkages. As some reference glycans have been shown to alleviate intestinal inflammation and improve barrier function, as well as reduce colitis from infection, we compare the effects of SG and reference glycan treatment in mouse models of these intestinal pathologies[20,21]. Together, this experimental pipeline (Fig. 1a) evaluates the function and therapeutic potential of SGs as microbiome modulators and demonstrates additional benefits of SGs compared to reference glycans, including compounds currently marketed as prebiotics.

Results

SGs vary in composition and fermentation dynamics. We selected a set of 653 SGs and 110 commercially available, reference glycans that are compositionally diverse and sufficiently soluble for culture-based growth and metabolite assays (Supplementary Data 1, Supplementary Table 1). Both SGs and reference glycans are composed of a similar set of monosaccharides (Fig. 1b), but unlike the reference glycans, most SGs (73%) contain multiple, different monosaccharides (Fig. 1c), demonstrating how SGs can be built to include dietary sugars in novel and complex combinations. SGs were catalytically synthesized[19] and span a wide range of average molecular masses (Fig. 1d) with a median of 1.7 kDa (range 0.3 kDa-77.5 kDa), corresponding to a polymerization of approximately ten monosaccharides. We profiled these SGs and reference glycans in a panel of ex vivo assays and highlighted the performance of two SGs with different compositions, BRF (glucose) and BQM (galactose and glucose), which were subsequently selected for structural analysis and mouse models.

We investigated variation in fermentation dynamics across the 763 glycans by measuring growth and pH kinetics of anaerobic fecal cultures. Hierarchical clustering of growth and pH parameters (Fig. 1e, Supplementary Fig. 1) identified groups of glycans with distinct growth (Fig. 1f) and pH (Fig. 1g) profiles. Reference glycans were enriched in group 1 (Fisher exact test $p = 7.5 \times 10^{-26}$) and group 4 (Fisher exact test $p = 4.8 \times 10^{-14}$), which both supported rapid growth resulting in a precipitous pH drop. Group 1 included pullulan, lactulose, GOS, and FOS; group 4 included XOS. Group 5, which included BQM and BRF, was enriched in SGs ($p = 6.0 \times 10^{-17}$) that were well-fermented at controlled rates with gradual reductions in pH. Thus, SG composition affects fermentability and SGs are generally fermented more slowly than reference glycans, likely due to their compositional complexity (Fig. 1c). In addition, these fermentation parameters were well correlated between fecal samples from two different donors (Supplementary Fig. 1c–n). While more data is needed to establish that SG effects are conserved across populations, these data suggest SGs have similar fermentation dynamics between individuals.

SGs change microbiome metabolic output. Fermentation of glycans by the gut microbiome produces metabolites including short-chain fatty acids (SCFAs) that have important physiological and immunological benefits[15,22]. We developed a high-throughput matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry method to quantify yields of two SCFAs, butyrate and propionate, in fecal cultures fermenting each of the 763 glycans. Butyrate is the preferred energy source of colonocytes, leading to healthy colonocyte function and maintenance of an anaerobic gut environment[23]. Propionate has immunological and metabolic effects locally in the gut and is absorbed into the bloodstream to regulate cholesterol[24,25]. In addition, butyrate and propionate promote differentiation of naive T cells into anti-inflammatory regulatory T cells[26,27]. Butyrate and propionate production by some SGs were similar to those of reference glycans, but many SGs were distinct in producing high propionate levels (Fig. 2a), supporting these SGs favor growth of propionate-producing taxa such as *Bacteroidaceae* and *Ruminococcaceae*[25].

We investigated the extent to which SCFA production underlies pH dynamics during glycan fermentation by measuring a time series of three main SCFAs (acetate, propionate, and butyrate) by gas chromatography in cultures fermenting a subset of glycans. The correlation between pH and production of these SCFAs was weak, but the relative levels of SCFAs and ammonia, an abundant microbial product, were strongly correlated with pH (Supplementary Fig. 2). While these SCFA measurements did not include lactate because it is non-volatile and can not be quantified by standard gas chromatography[28], our data supports that changes in pH observed during glycan fermentation are associated with the balance between SCFA and ammonia production.

In addition to SCFAs, glycan fermentation produces gaseous compounds that can lead to abdominal bloating, a major symptom limiting tolerability of reference glycans in patients with gastrointestinal disorders such as irritable bowel syndrome[29]. As the volume and rate of gas produced depend on diet and microbiome composition[30], we measured gas production rates by fecal communities growing on BRF, BQM, and randomly selected glycans from each of the five fermentation groups in Fig. 1e–g (Supplementary Data 2). The rate of gas production varied widely among groups (Fig. 2b) and reflected fermentation dynamics. Glycans in fermentation groups 1 and 4, mostly reference glycans that are rapidly fermented, had the most

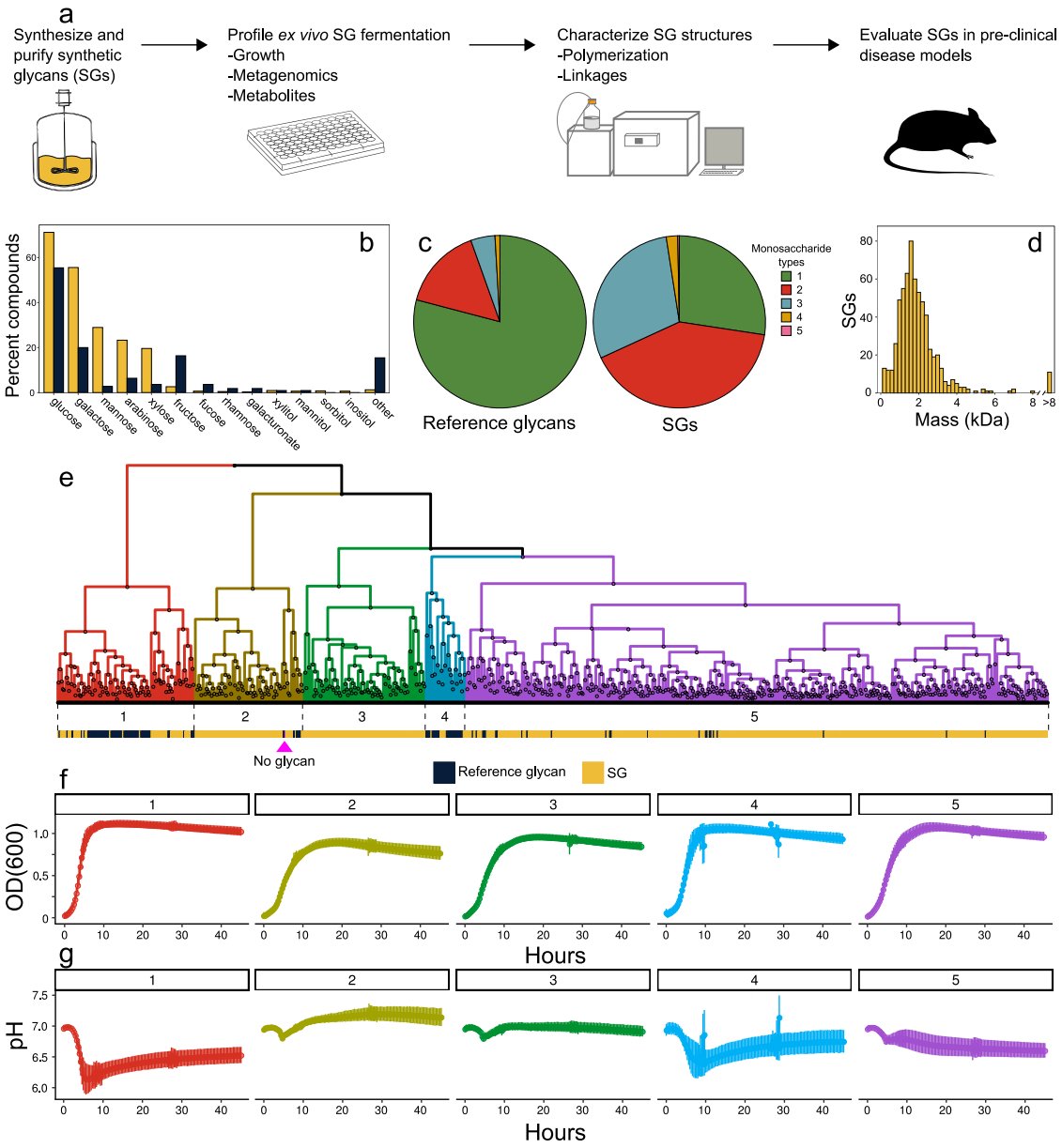


Fig. 1 Analytical pipeline and description of glycan compositions and fermentation dynamics. a Schematic representation of the analytical pipeline. **b**–**f** Monosaccharide compositions and fermentation dynamics of 653 SGs and 110 reference glycans. **b** Percentages of SGs (yellow) and reference glycans (indigo) containing various monosaccharide types. **c** Number of monosaccharide types composing each SG or reference glycan. **d** Distribution of weight average molecular weights of SGs measured by SEC. **e**–**g** Growth (OD_{600}) and pH dynamics of triplicate fecal cultures fermenting 5 g l^{-1} of a single SG or reference glycan in MM29 medium. **e** Hierarchical clustering of glycans into five fermentation groups based on twelve growth and pH parameters. Bars below the dendrogram show compound class: SG (yellow), reference glycan (indigo), or no glycan (magenta). Mean (**f**) growth and (**g**) pH curves (±SD) for each glycan fermentation group shown in **e**. Source data are provided as a Source Data file. SGs Synthetic Glycans, SEC size exclusion chromatography, OD_{600} optical density at 600 nm, SD standard deviation, kDa kilodalton.

rapid gas production. SGs in group 5, which included BRF and BQM, produced gas more moderately, potentially improving tolerability in humans.

SGs alter microbiome taxonomic composition. Glycan fermentation can result in divergent changes to microbiome composition, even differentially promoting closely related species[8]. Therefore, we applied shotgun metagenomic sequencing to examine fecal cultures fermenting a compositionally diverse set of 190 SGs and 40 reference glycans (Supplementary Data 3). While there is heterogeneity in the level of microbiome diversity resulting from fermentation of different SGs, SG fermentation generally resulted in higher taxonomic diversity (Fig. 2c) and species richness (Fig. 2d) relative to reference glycans, as shown by BQM and BRF promoting higher diversity relative to reference glycans (Fig. 2e).

Fecal community compositions following SG fermentations spanned the taxonomic space covered by reference glycan fermentations, but also reached novel compositions (Fig. 2f) that can be explored by clustering based on taxa abundances. SG-enriched taxa clusters had elevated abundances of *Lachnospiraceae* (clusters 1–3) and *Parabacteroides* (clusters 2–3) (Fig. 2g, Supplementary Fig. 3a–c). BRF is in glycan cluster 1 (Fig. 2g, Supplementary Fig. 3a) with increased abundances of members of the family *Lachnospiraceae* including *Roseburia*, a key butyrate

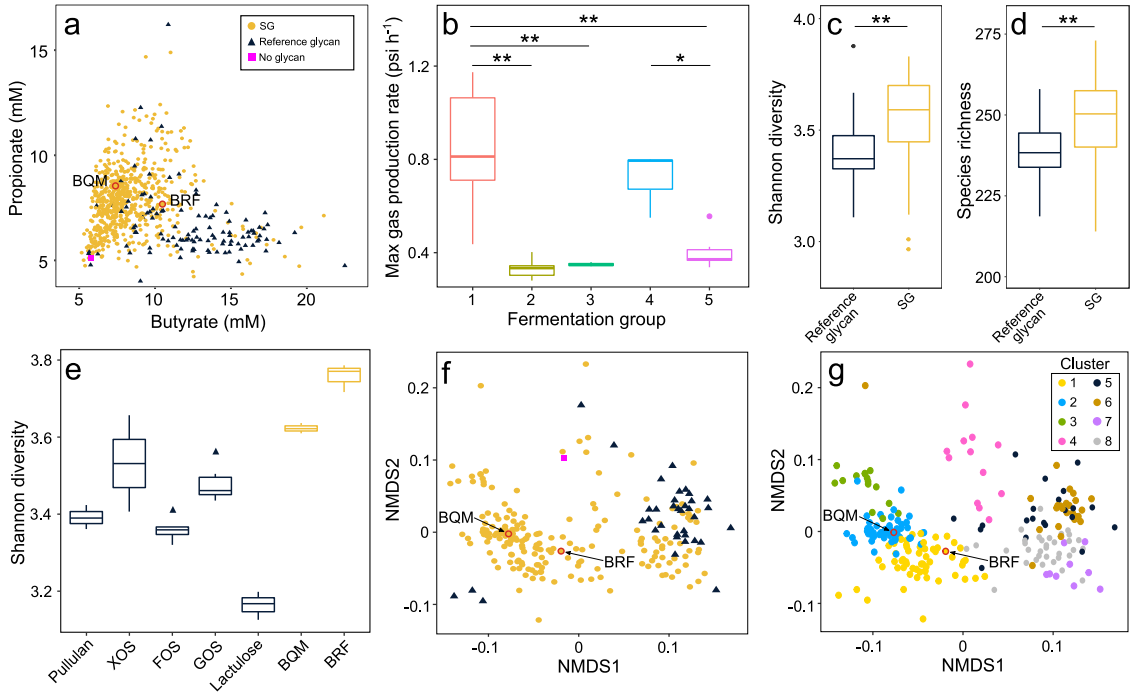


Fig. 2 Effects of glycans on fecal community metabolic output and taxonomic composition. a Yields of two SCFAs, butyrate and propionate, from fecal cultures fermenting either an SG (yellow circles, $n = 653$), reference glycan (indigo triangles, $n = 110$), or no glycan (magenta square). **b** Maximum gas production rate (psi h^{-1}) during fecal culture fermentation of glycans from each of the five fermentation groups in Fig. 1e–g. **c** Shannon diversity and **d** species richness of fecal cultures fermenting SGs (yellow, $n = 190$) versus reference glycans (indigo, $n = 40$). **e** Shannon diversity of fecal cultures fermenting BRF or BQM (yellow) is higher than reference glycans (indigo) for all comparisons except BQM versus XOS (Kruskal–Wallis followed by Dunn's comparison test, $p < 0.05$). **f** NMDS of metagenomic data calculated based on a matrix of Bray–Curtis dissimilarities using species-level mapping of sequencing reads from fecal cultures grown on either an SG (yellow circles, $n = 190$), reference glycan (indigo triangles, $n = 40$), or no glycan (magenta square). **g** NMDS as in **f** colored by differences in taxonomic composition defined by eight K-means clusters based on species-level mapping of sequencing reads. Data for each glycan is the mean of (**a**–**d**, **f**, **g**) three or (**e**) six replicate fecal cultures grown on 5 g l^{-1} of each SG or reference glycan for 45 h in MM29 medium. **a**, **f**, **g** BRF and BQM highlighted in red. **b**–**e** Box plots show median and interquartile ranges. Asterisks show significance (*$p < 0.05$, **$p < 0.01$) by **b** Tukey's test or **c**, **d** two-sided Wilcoxon rank-sum test. Source data are provided as a Source Data file. SG Synthetic Glycan, SCFA short-chain fatty acid, XOS xylo-oligosaccharides, FOS fructo-oligosaccharides, GOS galacto-oligosaccharides, NMDS non-metric multidimensional scaling.

producer that is reduced in the microbiomes of patients with Crohn's disease[31], and *Fusicatenibacter*, which is decreased in patients with active ulcerative colitis[32]. BQM is in glycan cluster 2 (Fig. 2g, Supplementary Fig. 3b) with elevated levels of other *Clostridiaceae* and *Parabacteroides*, taxa associated with remission in Crohn's disease[33]. Clusters enriched in reference glycan fermentations contained elevated levels of *Lactobacillaceae* (cluster 5) and *Bifidobacterium* (clusters 6–8) (Fig. 2g, Supplementary Fig. 3e–h), which are adapted to metabolize simple oligosaccharides such as FOS[34].

SGs are poor growth substrates for model pathogens. It is important to establish that microbiome-modulating glycans do not promote proliferation of pathogens. In particular, intestinal colonization by a multidrug-resistant (MDR) pathogen such as carbapenem-resistant *Enterobacteriaceae* (CRE) expressing extended-spectrum carbapenemases or vancomycin-resistant *Enterococcus* (VRE) is a major risk factor for bloodstream infection[35,36]. Currently, no approved therapies exist for de-colonization of these pathogens from the gut. We hypothesized that SGs will confer a competitive growth advantage to polysaccharide-utilizing commensals relative to pathogens in gut microbiomes[37]. Initially, we examined the ability of a panel of laboratory and MDR clinical *Enterobacteriaceae* and *Enterococcus* strains (Supplementary Table 2) to grow on each of a compositionally diverse set of 148 SGs including BRF and BQM and 32 reference glycans including FOS (Supplementary Data 4) as a sole carbohydrate source in defined medium. Most strains of the model enteric pathogens *Klebsiella pneumoniae* (Fig. 3a), *Escherichia coli* (Fig. 3b), and *Enterococcus faecium* (Fig. 3c) exhibited limited growth on SGs with final cell densities significantly lower than when grown on reference glycans including FOS.

To assess changes in the relative abundances of model pathogens in gut microbiomes during glycan fermentation, we spiked a healthy donor fecal community with either CRE *K. pneumoniae*, CRE *E. coli*, or VRE *E. faecium*, cultured it in a medium containing each SG or reference glycan (Supplementary Data 5), and measured changes in relative abundances of the pathogen by 16S rRNA gene sequencing. The relative abundances of *K. pneumonia* (Fig. 3d), *E. coli* (Fig. 3e), and *E. faecium* (Fig. 3f) were significantly lower after growth on SGs including BRF and BQM than on reference glycans, consistent with the SGs favoring growth of commensal taxa.

In contrast to the limited growth of pathogens on SGs relative to reference glycans, pure cultures of phylogenetically diverse, gut commensals (Firmicutes, Bacteroidetes, and Actinobacteria) grown in defined medium often reached similar cell densities with either BRF, BQM, or glucose as the sole carbohydrate source (Supplementary Fig. 4a–f). We measured a time series of BRF and BQM consumption by a healthy donor fecal community using high-performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD), which showed the fecal community fully consumed these SGs by progressively metabolizing fractions of increasing DP (Supplementary Fig. 5).

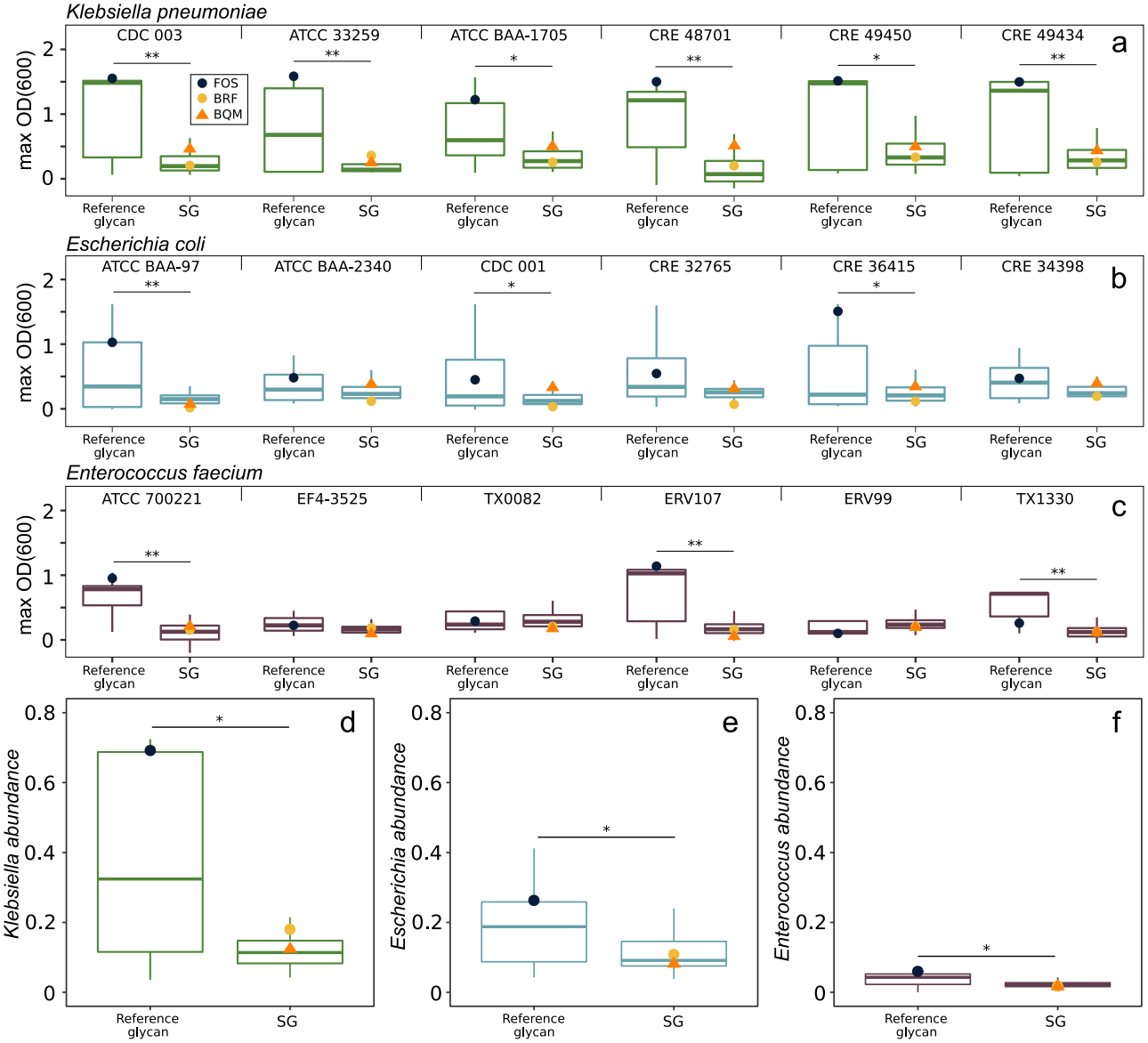


Fig. 3 Enteric pathogen growth in pure culture and relative abundances in fecal communities grown on glycans. Six strains of **a** *Klebsiella pneumoniae*, **b** *Escherichia coli*, or **c** *Enterococcus faecium* were cultured with 5 g l^{-1} of a single SG ($n = 148$) or reference glycan ($n = 32$) in CM3 medium. Data is the mean maximum growth (OD_{600}) of triplicate cultures; strain names are shown above each plot. Fecal communities from a healthy donor were OD_{600}-normalized to contain 8% of **d** *K. pneumoniae* CDC 003, **e** *E. coli* CDC 001, or **f** *E. faecium* ATCC 700221 and cultured in triplicate with 5 g l^{-1} of an SG ($n = 45$) or reference glycan ($n = 17$) for 45 h in MM29 medium. The relative abundances of the pathogens were quantified by 16S rRNA gene sequencing. Data points show FOS (indigo circle), BRF (yellow circle), or BQM (orange triangle) cultures. Box plots show median and interquartile ranges. Asterisks show significance (*$p < 0.05$, **$p < 0.01$) by two-sided Wilcoxon rank-sum test. Source data are provided as a Source Data file. SGs Synthetic Glycans, OD_{600} optical density at 600 nm, FOS fructo-oligosaccharides.

SGs reproducibly shift microbiome taxonomic and CAZyme profiles across individuals. Based upon the assimilated data from the ex vivo experiments, we focused on BRF and BQM because they have different monosaccharide compositions, are well-fermented, enrich different, anti-inflammatory taxa associated with resolution of colitis, promote high taxonomic diversity, and are poor growth substrates for pathogen-associated genera. To gain insight into how these SGs alter microbiome taxonomy and function, we assessed the effects of treatment with BRF, BQM, or lactulose on fecal communities from additional healthy donors by metagenomics (ten donors for SGs, seven donors for lactulose). Lactulose was included as a reference glycan that is hydrolyzed and metabolized by the colonic microbiota[38]. Lactulose is, to our knowledge, the only medically prescribed glycan that acts primarily through modulation of the microbiome and is commonly used to prevent and treat hepatic encephalopathy[39].

Ex vivo fecal metagenomics revealed glycan treatments resulted in taxonomic shifts that were conserved across individuals (Fig. 4a–c). Both SGs promoted *Parabacteroides* along with Clostridia such as *Ruminiclostridium* and multiple genera of *Lachnospiraceae*. There were also differences in taxa promoted by each SG. For example, *Roseburia* and *Prevotella* abundance were specifically elevated in the BRF treatment (Fig. 4a), similar to metagenomics of the BRF-containing glycan cluster 1 (Supplementary Fig. 3a). In contrast, lactulose treatment increased the relative abundance of *Bifidobacterium* and *Lactobacillus*, similar to the metagenomics of glycan cluster 5 that includes lactulose (Supplementary Fig. 3e).

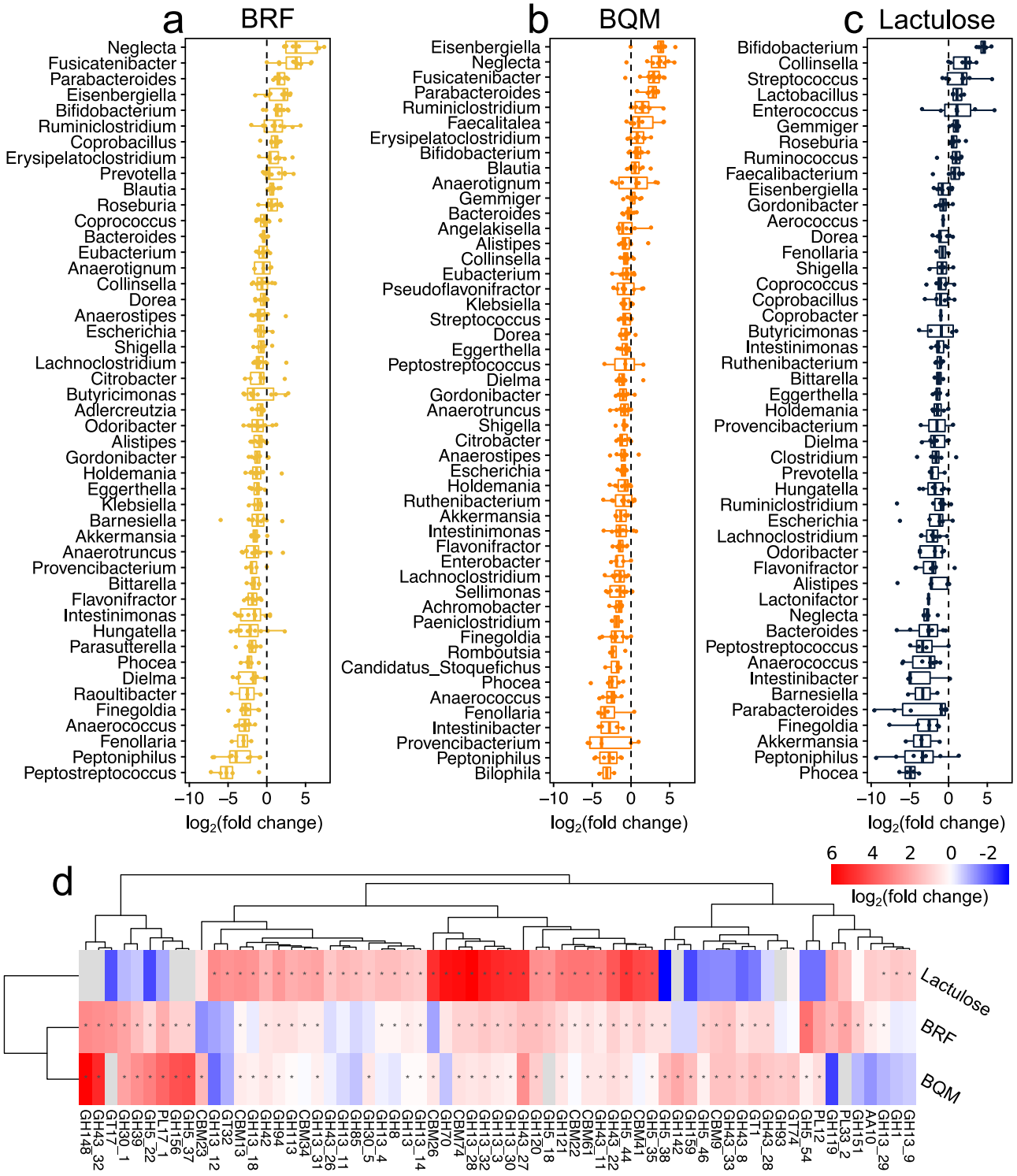


Fig. 4 Changes in fecal community composition in response to glycans across human donors. Differences in relative abundances of genera in ex vivo fecal cultures grown on (**a**) BRF, (**b**) BQM, or (**c**) lactulose relative to no-glycan controls. Data points show median log_2FC in genus abundances for each donor; box plots show median and interquartile range across donors. Genera with significant changes ($p < 0.05$ after FDR correction) are shown. **d** Differences in CAZyme gene abundances in ex vivo cultures grown on different glycans (median log_2FC versus no-glycan controls). CAZyme family/subfamilies with significant abundance changes ($p < 0.05$ after FDR correction) and $log_2FC > 1$ on at least one glycan versus no-glycan controls are shown with hierarchical clustering based on Euclidean distance. Colors are FC with gray showing families not detected in the no-glycan control. Asterisks indicate significantly elevated abundance. **a**–**d** Fecal cultures from healthy donors (ten donors for SGs, seven donors for lactulose) were grown in triplicate for 45 h in MM29 medium supplemented with 5 g l^{-1} glycan, as appropriate, and sequenced by metagenomics. Statistical significance for genus and CAZyme changes was determined by fitting a linear mixed-effect model on rank transformed genera/CAZyme abundance data with glycan treatment as fixed effect and subject as random effect. Source data are provided as a Source Data file. SGs Synthetic Glycans, FC fold change, FDR false discovery rate, CAZyme carbohydrate-active enzyme, GH glycoside hydrolase, CBM carbohydrate-binding modul, PL polysaccharide lyase, GT glycosyltransferase, AA auxiliary activity.

Glycan treatment was also associated with widespread changes in the abundances of carbohydrate-active enzyme (CAZyme) genes in fecal metagenomes (Fig. 4d). Metabolism of an oligosaccharide glycan requires specific CAZymes to cleave its glycosidic linkages[40,41], leading to promotion of the subset of taxa in the microbiome encoding relevant CAZymes[42,43]. As such, the abundances of genes for CAZymes that are active on an oligosaccharide should increase in metagenomes following treatment. The abundances of glycoside hydrolases (GHs) known to cleave lactulose such as GH42[44] were most highly elevated in the lactulose cohort, along with thirteen GH13 subfamilies (Fig. 4d). GH13 are not known to cleave lactulose, but are widespread in the genomes of Bifidobacteria[45], which increased in response to lactulose (Fig. 4c). Both SGs resulted in increased abundances of GH148 (β-1,3-glucanases) along with six subfamilies of GH5 containing predicted β-glucosidases and β-glucanases that are likely active on the glucose linkages present in BRF and BQM. In addition, the metagenomes of the BQM-treated communities were enriched in galactanase CAZyme families such as GH159 (β-D-galactofuranosidase)[46] and GH43 (7 subfamilies) that putatively cleave the galactose linkages in BQM (Fig. 4d).

SG structural features. We performed a detailed characterization of the chemical structures of BRF and BQM. The chemical catalysis methods for synthesis of BRF and BQM (see Methods) yielded an ensemble of structures that were defined as a function of monosaccharide input, catalyst, and reaction conditions[19]. Multiplexed capillary gel electrophoresis with laser-induced fluorescence detection (xCGE-LIF) showed BRF (Fig. 5a) and BQM (Fig. 5b) have low monomer content (1.1% and 2.2%, respectively). We observed a series of discrete peaks indicating multiple structures at each DP below five, but the increased structural complexities at higher DPs exceeded resolution. Size exclusion chromatography of BRF (Fig. 5c) and BQM (Fig. 5d) confirmed their low monomer contents and showed similar average DPs of 11.3 and 10.9, respectively. Both compounds have a polydispersity index of 1.9 due to the range of molecular masses comprising each SG.

We performed linkage analysis of BRF (Fig. 5e) and BQM (Fig. 5f) to examine the abundances and types of glycosidic bonds. Both glycans contained mixtures of 1,2- 1,3-, 1,4-, and 1,6-bonds with frequent branching, as >2 linkages were present in 19% of monosaccharide subunits in BRF and 14% of subunits in BQM. We further compared the glycosidic linkages of BRF and BQM relative to reference glycans by two-dimensional nuclear magnetic resonance spectroscopy (2D-NMR). The 2D-NMR profiles of BRF versus pullulan (Supplementary Fig. 6a) and BQM versus GOS (Supplementary Fig. 6b) support that these SGs contain a greater diversity of glycosidic bonds with distinct stereo- and regiochemistries. The structural complexity of SGs likely contributes to their comparatively slow fermentation by commensals (Fig. 1f, g).

SGs demonstrate therapeutic potential in mouse models. We evaluated the therapeutic potential of BRF and BQM in mouse models of intestinal damage and disease. Initially, we examined the effects of BRF or FOS in a mouse model of dextran sodium sulfate (DSS)-induced colitis (Fig. 6a). BRF treatment reduced DSS-mediated weight loss (Fig. 6a) and improved scores of diarrhea (Fig. 6b) and endoscopy (Fig. 6c). Colonic histology showed DSS induced lesions including epithelial erosion or mucosal ulceration, loss of colonic glands, and inflammation of the mucosa, which was reduced by BRF treatment relative to DSS alone (Fig. 6d). In contrast, FOS treatment failed to reduce weight loss or improve scores of stool, endoscopy, or histology relative to DSS treatment without glycan supplementation (Fig. 6a–d).

We also examined if BRF and BQM could alleviate pathologies associated with *C. difficile* infection. First, we confirmed that *C. difficile* grows poorly on BRF or BQM as a sole carbohydrate source in defined medium (Supplementary Fig. 4g). We then tested if BRF, BQM, or FOS have in vivo efficacy in a *C. difficile* murine infection model (Fig. 6e). Fecal metagenomics on day 6 after *C. difficile* infection revealed a divergence in fecal microbiome compositions across treatments (Supplementary Fig. 7a) with the abundances of different genera elevated in each glycan treatment (Supplementary Fig. 7b–e). Further, we found these microbiome changes translated into improved outcomes. Glycan treatment significantly reduced weight loss (Fig. 6e) and increased survival (Fig. 6f) relative to the no-glycan treatment. Whereas 3 of 12 mice survived in the no-glycan treatment, 11 of 12 mice survived BRF treatment and BQM treatment had 100% survival, similar to the vancomycin-treated controls (Fig. 6f). In addition, all glycans improved clinical scores relative to the no-glycan treatment (Fig. 6g). While FOS improved all metrics relative to the no-glycan treatment, BQM further reduced weight loss (Fig. 6e), increased survivorship (Fig. 6f), and improved clinical scores (Fig. 6g) relative to FOS.

Discussion

Perturbations to the taxonomic composition and metabolic output of the gut microbiome have been linked to numerous non-communicable and immune-related pathologies. Delivery of rationally optimized, complex glycans is a promising option to drive the composition of the microbiota toward health-promoting states by leveraging taxa-specific differences in glycan metabolism, one of the dominant factors shaping gut microbiome composition[7]. Our study demonstrates SGs are a novel, glycan-based modality to manipulate the properties of human gut communities with flexibility and precision. SGs are synthetic compounds with mixtures of glycosidic bonds that recapitulate features of the chemical complexity of dietary fiber, while avoiding challenges associated with natural fibers including geographic and seasonal variability of raw material sourcing. Moreover, SGs are well-metabolized by polysaccharide-utilizing commensals, but not important pathogens. While still requiring validation in human trials, SGs could represent an antibiotic-sparing approach to control infection with no known mechanism of resistance.

Both of the SGs that we tested in mouse models shifted fecal community composition to elevate potentially anti-inflammatory taxa that are negatively associated with inflammatory bowel disease (IBD) including *Lachnospiraceae* (especially *Roseburia* and *Fusicatenibacter*), *Ruminiclostridium*, and *Parabacteroides*[31–33,47]. As such, the SGs BQM and BRF could potentially alleviate altered microbiome compositions associated with mild and moderate inflammation, which is supported by how these SGs lessen pathologies associated with DSS colitis (Fig. 6a–d) and *C. difficile* infection (Fig. 6e–g) in mice. SG-mediated microbiome shifts could also be beneficial for other diseases. *Roseburia* is reduced in the microbiomes of colorectal cancer patients[48]. *Fusicatenibacter* and *Parabacteroides* are proposed to promote resistance to *C. difficile* infection[49,50] and *Parabacteroides* is associated with decreased hepatic steatosis[51].

SGs extend the benefits of reference glycans and avoid many of the limitations of microbiome interventions based on live bacteria including dosing, engraftment, and inconsistent manufacturing. SGs could also be applied to promote growth and engraftment of live bacteria therapeutics. SGs selected for clinical development can be produced using controlled manufacturing processes and

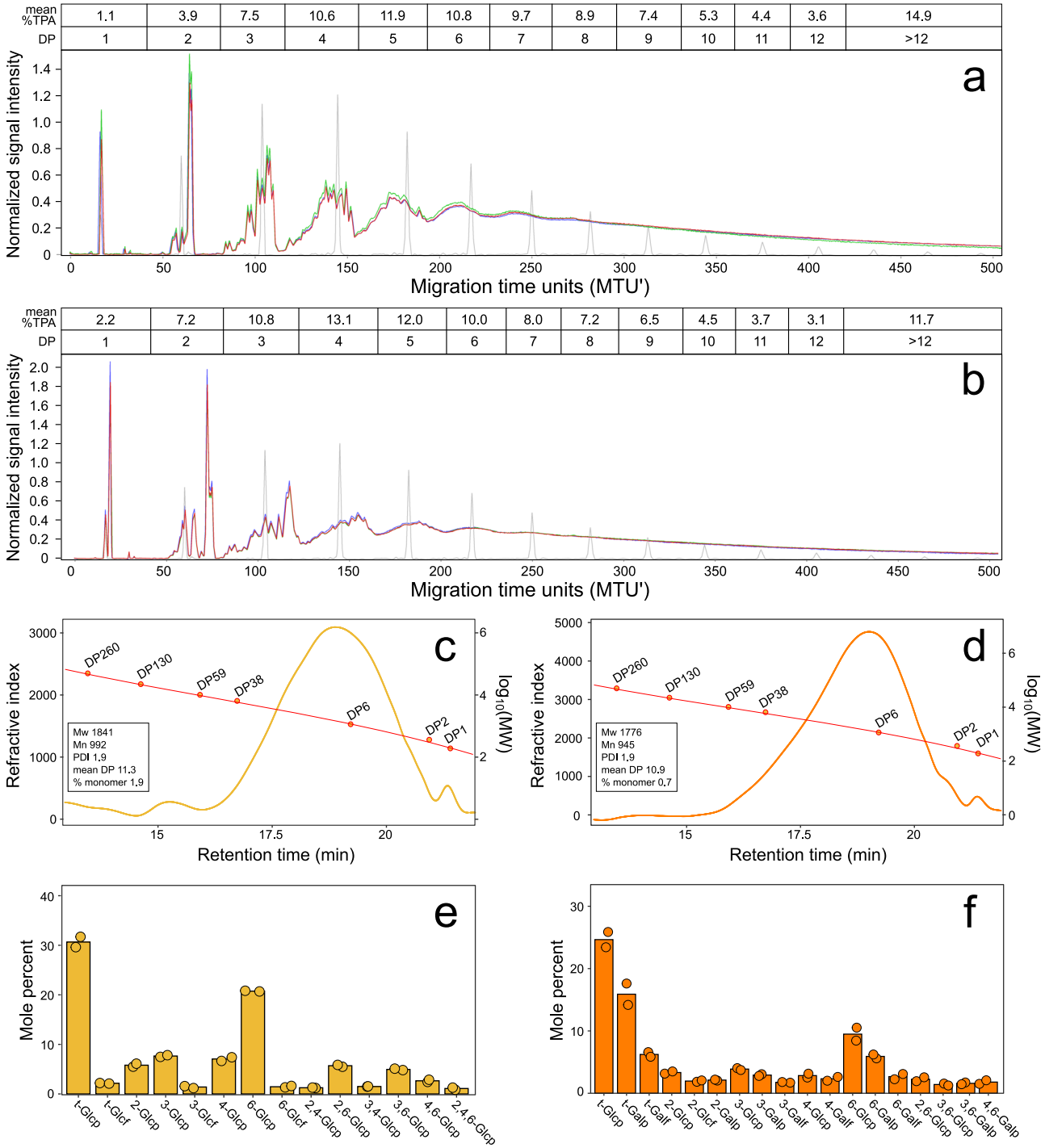


Fig. 5 Structural features of SGs. xCGE-LIF analysis of the DP of (**a**) BRF and (**b**) BQM. Triplicate measurements of each SG (red, green, purple) and oligomaltose standards (gray) are shown relative to normalized migration time units (MTU'). The mean percent TPA at each DP is above the plots. SEC chromatograms of (**c**) BRF and (**d**) BQM showing distributions relative to molecular weight (MW) standards with insets showing polymerization parameters. Refractive index in millivolts and molecular weights in Daltons. **e**–**f** Abundances (mole percent) of monosaccharides with different glycosidic linkages in (**e**) BRF and (**f**) BQM. Labels show all linkage types at >1% mole percent with residues linked only at 1- position as terminal "t-" residues. Bars show means and points show measurements for independent syntheses of each glycan. Source data are provided as a Source Data file. SGs Synthetic Glycans, xCGE-LIF multiplexed capillary gel electrophoresis with laser-induced fluorescence detection, DP degree of polymerization, TPA total peak area, SEC size exclusion chromatography, Mw weight average molecular weight, Mn number average molecular weight, PDI polydispersity index, Glcp glucopyranose, Glcf glucofuranose, Galp galactopyranose, Galf galactofuranose.

have been evaluated as Generally Regarded as Safe (GRAS) for their intended use in clinical research, allowing generation of proof of mechanism information by streamlined, human interventions. Together, our results show SGs are chemically diverse glycans that effectuate novel taxonomic and metabolic shifts to gut communities and reduce symptoms of colitis in

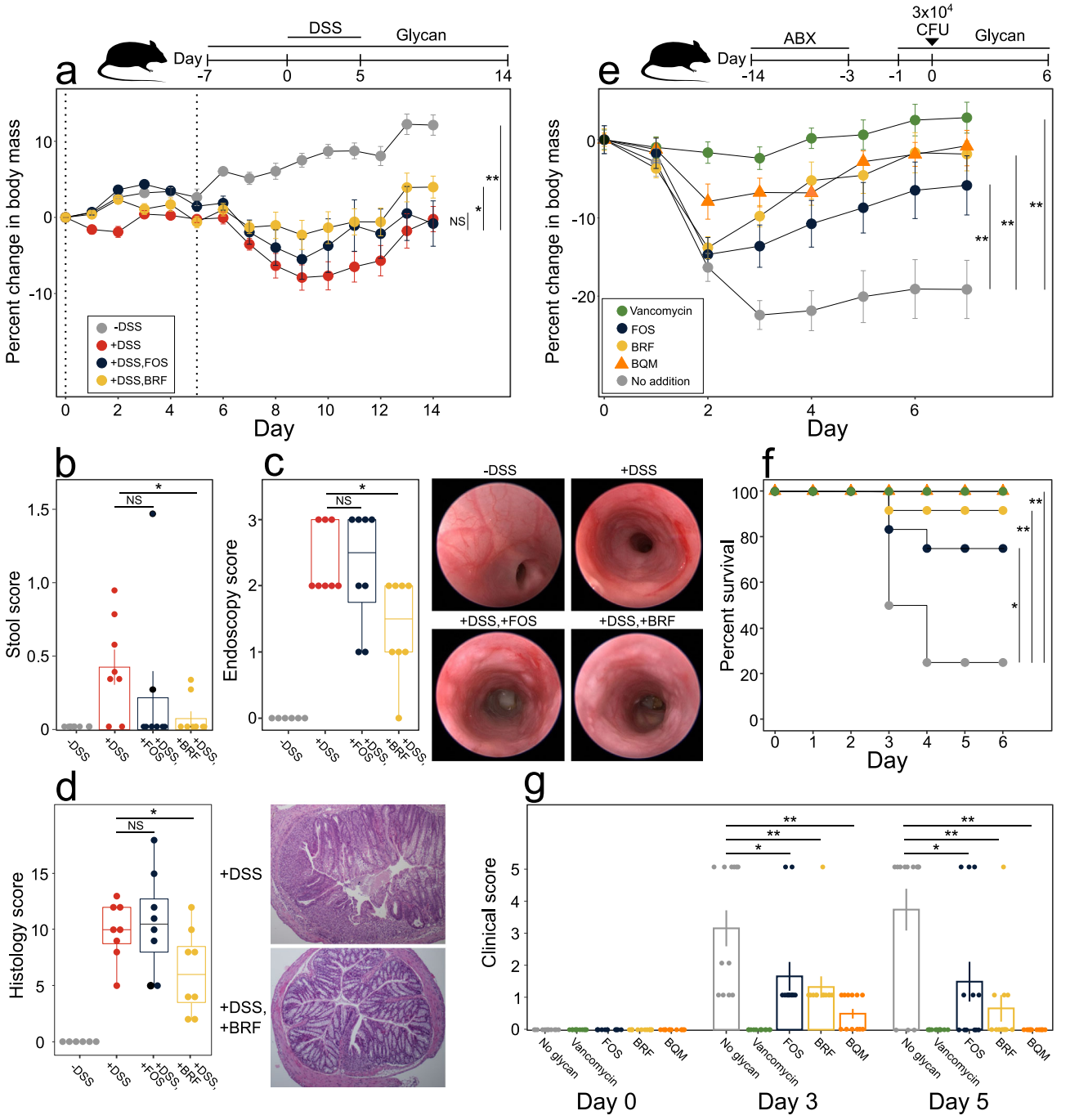


Fig. 6 Glycan effects in mouse models of DSS colitis and *C. difficile* infection. a–d Mice were treated in drinking water with 2.5% DSS (days 0–5, dashed lines) and 1% (v/v) glycans (days 7–14), as appropriate. Treatment groups (eight animals per group): −DSS (gray), +DSS (red), +DSS, FOS (indigo), +DSS, BRF (yellow). Treatment group comparisons of (**a**) body weight, (**b**) stool score averaged over days 0–14, (**c**) day 14 endoscopy scores with representative images, and (**d**) day 14 histology scores with representative 100x magnified H&E stained distal colon micrographs. **e–g** Mice were treated with antibiotics (days -14–3), infected with *C. difficile* (day 0), and treated with 50 mg kg^{-1} vancomycin daily (days 0–4) or 1% (v/v) glycans in drinking water (days 1–6), as appropriate. Treatment groups (12 animals per group): no glycan (gray), vancomycin (green), FOS (indigo), BRF (yellow circles), and BQM (orange triangles). Treatment group comparisons of (**e**) body weight, (**f**) survival, and (**g**) clinical scores. Data in (**a**, **b**, **e**, **g**) show treatment group means ± SEM. Box plots in (**c**, **d**) show median and interquartile range. Data points in (**b**–**d**, **g**) show individual mice. **a**, **e** Statistics on body mass changes are based on area under the curve for all individual mice. Asterisks show significance (*$p < 0.05$, **$p < 0.01$) by (**a**–**e**, **g**) two-sided Wilcoxon rank-sum test or (**f**) log-rank test. Source data are provided as a Source Data file. DSS dextran sodium sulfate, FOS fructo-oligosaccharides, ABX antibiotics, CFU colony forming units, SEM standard error of the mean, NS non-significant.

mouse models. Future work will further expand the potential of SGs in disease management by modulating the composition and metabolites produced by the gut microbiome.

Methods

Synthesis of BRF and BQM. BRF was synthesized by combining D-(+)-glucose (100.0 g, 555.1 mmol), Dowex® Marathon™ C (7.0 g, 5% dry weight ratio to glucose, 29% moisture content), and 30.0 mL DI water in a 1000 mL 3-necked round bottom flask equipped with an overhead stirrer, thermocouple plug, and a short-path distillation head. The mixture was stirred continuously at 100 rpm using a glass stirring shaft equipped with a Teflon half-moon paddle. The reaction mixture was run at 130 °C for 4 h. To quench the reaction, 60 mL DI water was added to the reaction mixture. The Dowex resin was removed by vacuum filtration through a fritted-glass filter. The solution was then diluted to 25.0 °Bx and purified by precipitation. The solution was slowly poured into absolute ethanol to form a cloudy solution with a final water:ethanol ratio of 1:9 (v/v). The cloudy solution was then centrifuged at 2100 × *g* for 2 h. The supernatant was removed, and the precipitate was collected and dissolved in water. The residual ethanol was removed under reduced pressure. The solution was frozen at −20 °C and lyophilized to yield the final product of white powder (60.5 g, 61% yield).

BQM was synthesized by combining D-(+)-glucose (50.0 g, 277.5 mmol), D-(+)-galactose (50.0 g, 277.5 mmol), Dowex® Marathon™ C (7.0 g, 5% dry weight ratio to monosaccharides, 29% moisture content), and 30.0 mL DI water in a 1000 mL 3-necked round bottom flask equipped with an overhead stirrer, thermocouple plug, and a short-path distillation head. The mixture was stirred continuously at 100 rpm using a glass stirring shaft equipped with a Teflon half-moon paddle. The reaction mixture was run at 130 °C for 4 h. To quench the reaction, 60 mL DI water was added to the reaction mixture. The Dowex resin was removed by vacuum filtration through a fritted-glass filter. The solution was then diluted to 25.0 °Bx and purified by precipitation. The solution was slowly poured into absolute ethanol to form a cloudy solution with a final water:ethanol ratio = 1:9 (v/v). The cloudy solution was then centrifuged at 2100 x *g* for 2 h. The supernatant was removed, and the precipitate was collected and dissolved in water. The residual ethanol was removed under reduced pressure. The solution was frozen at −20 °C and lyophilized to yield the final product of white powder (57.2 g, 57% yield).

Size exclusion chromatography (SEC). Glycan samples (300 mg) were resuspended in 10 mL water, 0.2 mm filtered, and 10 µL of glycan solution was injected to an Agilent 1100 HPLC system with refractive index (RI) detector equipped with a guard column (Agilent PL aquagel-OH, 7.5 × 50 mm, 5 µm, PL1149-1530), and two SEC columns (2X Agilent PL aquagel-OH 20,PL1120-6520) in tandem connection. The sample was run with 0.1 M $NaNO_3$ mobile phase, 28 min run time, 0.9 mL min^{-1} flow rate with the column and RI detector at 40 °C. The peaks of the sample were integrated and the weight-average molecular mass (M_w), number average molecular mass (M_n), mean degree of polymerization (mean DP), and polydispersity index (PDI) were determined using Agilent Cirrus GPC/SEC software (version 3.4.2). The calibration curve was generated from polymer standard solutions (10 mg mL^{-1}) of D-(+) Glucose Mp 180, Carbosynth Ltd Standard; Maltose Mp 342, Carbosynth Ltd Standard; Maltohexaose Mp 990, Carbosynth Ltd Standard; Nominal Mp 6100 Pullulan Standard, PSS # PPS-pul6k; Nominal Mp 9600 Pullulan Standard, PSS # PPS-pul10k; Nominal Mp 22000 Pullulan Standard, PSS # PPSpul22k; and Nominal Mp 43000 Pullulan Standard, PSS # PPSpul43k.

xCGE-LIF oligosaccharide analysis. The DP distribution of SGs were determined using a glyXboxCE™ (glyXera GmbH, Magdeburg, Germany) based on multiplexed capillary gel electrophoresis with laser-induced fluorescence detection (xCGE-LIF)[52]. SG samples were prepared using a glyXprep kit for oligosaccharide analysis (KIT-glyX-OS.P-APTS-48-01, glyXera). Briefly, samples were diluted 100-fold with ultrapure water and labeled with fluorescent dye 8-aminopyrene-1,3,6-trisulfonic acid (APTS). Two µL of each SG sample, 2 µL APTS Labeling Solution, and 2 µL ReduX Solution were mixed thoroughly and incubated for 3 h at 37 °C. The labeling reaction was stopped by adding 150 µL Stopping Solution and the excess of unreacted APTS and salt were removed using glyXbeads for hydrophilic interaction chromatography solid phase extraction (HILIC-SPE). The samples were aliquoted to wells containing 200 µL glyXbead slurry and incubated for 5 min at ambient temperature for binding, followed by washing and elution steps.

The purified APTS-labeled glycans were analyzed on a glyXboxCE™ system (glyXera) equipped with a 50 cm 4-capillary array, filled with POP-7™ polymer (4333466 and 4363929, Thermo Fisher Scientific). The sample (1 µL) was mixed with 1 µL prediluted 1st NormMiX (STD-glyX-1stN-100Rn-01, glyXera)[53] internal standard to align migration times for migration time unit (MTU') calculations. The mixture was combined with 9 µL glyXinject (C-glyXinj-1.6mL-01, glyXera) and subjected to xCGE-LIF analysis. The samples were electrokinetically injected and measured with a running voltage of 15 kV for 40 min. Data were analyzed with glyXtoolCE™ (glyXera) glycoanalysis software to perform migration time alignment, raw data smoothing, DP-range specific interval picking (summing up of the total measuring signal within the DP ranges after comparison with a reference oligo-maltose ladder), and normalization of total peak areas.

Glycan linkage analysis. Permethylation was performed as previously described with slight modification[54]. The glycan sample was purified to a monosaccharide content <1.0% using chromatography before permethylation. The glycan sample (500 µg) was dissolved in dimethyl sulfoxide (DMSO) for 30 min with gentle stirring. A freshly prepared sodium hydroxide suspension in DMSO was added, followed by a 10 min incubation. Iodomethane (100 µL) was added, followed by a 20 min incubation. A repeated round of sodium hydroxide and iodomethane treatment was performed for complete permethylation. The permethylated sample was extracted, washed with dichloromethane (DCM), and blow dried with nitrogen gas. The sample was hydrolyzed in 2 M trifluoroacetic acid (TFA) for 2 h, reduced overnight with sodium borodeuteride (10 mg mL^{-1} in 1 M ammonia), and acetylated using acetic anhydride/TFA. The derivatized material was extracted, washed with DCM, and concentrated to 200 µL. Glycosyl linkage analysis was performed on an Agilent 7890 A GC equipped with a 5975 C MSD detector (EI mode with 70 eV), using a 30-meter RESTEK RTX®-2330 capillary column. The GC temperature program: 80 °C for 2 min, a ramp of 30 °C min^{-1} to 170 °C, a ramp of 4 °C min^{-1} to 245 °C, and a final holding time of 5 min. The helium flow rate was 1 mL min^{-1}, and the sample injection was 1 µL with a split ratio of 10:1.

2D HSQC NMR. Structures of SGs were compared to reference glycans: pullulan (P4516 Sigma) or galacto-oligosaccharides (DOMO Vivinal GOS). Glycans were lyophilized and a 20 mg sample was dissolved in 200 µL of deuterium oxide (D_2O) with 0.1% (v/v) acetone as the internal standard. The solutions were placed into 3 mm NMR tubes and HSQC (heteronuclear single quantum coherence) spectra were recorded at 25 °C on a Bruker AVANCE III 500 MHz spectrometer equipped with a 5 mm BBF-H-D-05 probe with *Z*-axis gradient using the Bruker program "HSQCETGP.hires2". HSQC experiments were performed with eight scans and a 1 s recycle delay. Each spectrum was acquired from 7.5–1.5 ppm in F2 (^1H) with 1024 data points and 120–50 ppm in F1 (^{13}C) with 256 data points. The resulting spectra were analyzed using the MestReNova software (version: 12.0.0-20080) from Mestrelab Research.

Glycan consumption analysis. Consumption of SGs during fermentation by fecal cultures was measured at ProDigest (Ghent, Belgium) by high-performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD) using an ICS-3000 chromatography (Dionex, Sunnyvale, CA, USA) equipped with a CarboPacPA20 column (Dionex)[55]. Sample preparation involved initial dilution of the sample with ultrapure water followed by deproteinization with acetonitrile (1:1), centrifugation (21,380 × *g*, 10 min), and filtration (0.2 µm PTFE, 13 mm syringe filter, VWR International) prior to injection (5 µL) into the column. Qualitative glycan fingerprints were obtained by plotting the detected signal (nC) against the elution time.

Human fecal sample collection. Informed consent was obtained from all donors before fecal sample collection. To collect samples for ex vivo cultivation, donors were instructed to collect the bowel movement into a sample collection unit, which was immediately sealed and placed on ice. Within 4 h of bowel movement, sample collection units were transferred into an anaerobic chamber, unsealed, and processed into fecal slurries. To prepare fecal slurries for culturing, samples were transferred into filtered blender bags (Interscience) and diluted using 1x phosphate-buffered saline (PBS) and glycerol to result in 20% fecal slurry (w/w) containing 15% glycerol (w/w). Diluted samples were homogenized in a lab blender (Interscience 032230), flash frozen in a dry ice/ethanol bath, and stored at −80 °C.

Fecal samples from four donors were selected for ex vivo cultivations as representative samples from among a set of 32 donors by assessing the homogeneity of donors at basal conditions by metagenomic sequencing (Supplementary Fig. 8). The donor 1 sample was used for fermentation dynamics, metabolite analysis, gas production, and SG consumption analysis. The donor 2 sample was used for fermentation dynamics comparisons. The donor 3 sample was used for glycan culture metagenomics. The donor 4 sample was used for model pathogen spike-ins. In addition, the reproducibility of fecal community responses to glycan treatment across individuals was assessed by ex vivo cultivations of fecal samples from ten randomly selected healthy subjects for SGs (BRF, BQM) and seven healthy subjects for lactulose. These samples were not pre-screened by metagenomic sequencing to avoid biases in sample selection.

Microbial cultivation. Fecal communities were cultured in Mega Medium 29 (MM29) (Supplementary Table 3) and single strains were grown in Clostridial Minimal 3 (CM3) medium (Supplementary Table 4). All cultures were grown anaerobically at 37 °C containing 5 g l^{-1} of the appropriate glycan as the sole carbohydrate source. Growth was measured in 60 µl cultures in 384 well plates (3860 Corning) sealed using a Breathe-Easy® sealing membrane (Z380059 Sigma). Cell densities at an optical density of 600 nm (OD_{600}) and pH were measured using a Biotek Synergy H1 multi-mode plate reader outfitted with a Biostack 4 plate stacker. Fecal cultures were grown in MM29 medium for 45 h before DNA extraction for metagenomic or 16S rRNA gene sequencing. Gas production was measured continuously in 25 ml cultures fermenting different glycans in MM29 medium grown in 125 mL glass bottles sealed with an Ankom^{RF} gas production module (Ankom, Macedon NY, USA).

Pathogen growth in pure culture was measured for strains each of *K. pneumoniae, E. coli*, and *E. faecium* (Supplementary Table 2) cultured anaerobically in CM3 medium supplemented with 5 g l^{-1} of each glycan (Supplementary Data 4). To measure pathogen abundance in fecal cultures, fecal slurries were OD-normalized to contain 8% of each pathogen strain, grown for 45 h in MM29 medium supplemented with 5 g l^{-1} of each glycan (Supplementary Data 5), and relative abundances were quantified by 16S rRNA gene sequencing.

Culture pH was measured by supplementing the medium with 2 μM BCECF [2,7-bis-(2-carboxyethyl)-5-(and -6)-carboxyfluorescein] (C3411 Sigma) and the ratio of BCECF fluorescence at the pH-sensitive point (485 nm excitation; 540 nm emission) relative to the pH-insensitive isosbestic point (450 nm excitation; 540 nm emission) was measured[56]. The pH was calculated relative to a standard curve of media at known pH by fitting a sigmoidal curve using four parameter logistic regression. The R package, phgrofit (version 1.0.2)[57], was developed to extract physiological descriptors from the kinetic pH and OD_{600} curves (Supplementary Fig. 1). Glycans were clustered based on twelve fermentation parameters that were transformed into *Z*-scores by subtracting the mean across all glycans and dividing by the standard deviation. Hierarchical clustering analysis on these *Z*-score values was performed using the Manhattan distance metric and the complete agglomeration method; cluster number was selected based on the "elbow method" to minimize the within cluster variation (sum of squares distance).

DNA sequencing and analysis. Fecal microbiomes were characterized by 16S rRNA gene or shotgun metagenomic sequencing (Diversigen, MN USA). DNA was extracted from human and mouse fecal samples using Qiagen DNeasy PowerSoil extraction plates, quantified using the Quant-iT PicoGreen dsDNA assay (Thermo Fisher), and stored at −80 °C until analysis. 16S libraries were prepared by PCR amplification with the 515 F/806 R primer set and metagenomic libraries were prepared using the NexteraXT kit before sequencing on the Illumina platform. The average reads per sample for shotgun metagenomics was 0.5 million reads and for 16S metagenomics was 25 thousand reads.

Sequences of 16S rRNA genes were analyzed by UNOISE clustering[58] and denoising of raw sequences followed by DADA2/RDP taxonomic calling at the genus or family level[59]. Taxa count tables from shotgun metagenomic sequencing data were generated by the SHOGUN pipeline[60] using a database including the first 20 strains per species in RefSeq v87. Metagenomic reads with ambiguous species matches were excluded from taxa counts to reduce spurious matches.

To cluster SGs based on taxonomic response using metagenomics data from ex vivo cultures, the relative abundance of each species was averaged across triplicate cultures for each glycan. *K*-means clustering was performed based on species relative abundances differences relative to no-glycan controls defined by species-level mapping of sequencing reads. The number of clusters ($K = 8$) was selected using the "elbow method" to minimize the within-cluster variation (sum of squares distance) while also minimizing *K*. For ecological analysis, relative abundance was averaged across technical replicates for each species. The *vegan* R package[61] (version 2.5-6) was used to compute the Shannon diversity, Bray–Curtis dissimilarity, and multidimensional scaling measures and coordinates.

Sequencing reads corresponding to carbohydrate-active enzyme (CAZyme) genes in metagenomic data were identified based on alignment at 97% identity to a CAZyme gene sequence database, which includes all annotated CAZyme genes from all strains in the same database used for taxonomy assignment (see above). CAZyme genes in each strain were annotated using the Hidden Markov Models from dbCAN2[62]. Metagenomic reads were mapped to the CAZyme database using the BURST aligner[63] to build a table of gene length-normalized count values for each CAZyme family/subfamily in a metagenomic sample.

Short-chain fatty acid (SCFA) quantification. To quantify butyrate and propionate by matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry, fecal culture supernatants and SCFA standards in culture medium were derivatized by 3-Nitrophenylhydrazine hydrochloride (3NPH, N21804 Sigma) through *N*-(3-Dimethylaminopropyl)-*N′*-ethylcarbodiimide hydrochloride (EDC, Sigma E6383) coupling[64]. Culture media (10 μL) was mixed with 10 μL of 100 mM EDC in 50:50 acetonitrile/water (v/v) containing 7.5% pyridine and 10 μL of 100 mM 3NPH in 50:50 acetonitrile/water (v/v), incubated at 45 °C for 1 h, and diluted with 30 μL of 50:50 acetonitrile/water (v/v). The matrix internal standard (IS) solution was prepared by derivatizing 25 mM of ^{13}C-butyric acid and ^{2}D-propionic acid using 3NPH and mixing with saturated 9-Aminoacridine (9AA, Sigma 92817) in 50:50 methanol/water (v/v) at a 1:4 volume ratio.

SCFA reaction mixtures were mixed 1:1 with matrix IS solution, co-crystallized onto a MALDI HTS target plate by a TTP Mosquito liquid handling system, and dried at room temperature. A Bruker Daltonics ultrafleXtreme MALDI-TOF/TOF mass spectrometer was used to acquire mass spectra in reflectron negative ion mode using 65% laser from laser attenuator and 2200-shots accumulation for each sample. Data was collected and processed using Bruker PharmaPulse 2.1 software. Peak area of derivatized butyrate (m/z 222.2) and propionate (m/z 208.2) were monitored and normalized with the peak area of corresponding internal standards (m/z 224.2 and 210.2, respectively). The concentrations of propionate and butyrate in each sample were calculated using calibration standards. Measurements of propionate and butyrate by MALDI-TOF and by gas chromatography with flame-ionization detection (GC-FID) were confirmed to be well correlated for ex vivo fecal cultures fermenting SGs (Supplementary Fig. 9).

To quantify acetate, butyrate, and propionate in fecal microbial culture supernatants by GC-FID, 50 μL of each sample and calibration standards (2.5, 10.0, 15.0, 20.0, 30.0, and 40.0 mM of each SCFA) were mixed with 20 μL of a 400 mM 2-ethylbutyric acid internal standard solution prepared in HPLC-grade water. Immediately before injection, 30 uL of each mixed sample was acidified with 100 μL of 6% formic acid and 1 μL was injected. The GC-FID (7890 A, Agilent) was run as follows: 15 m × 0.53 mm × 0.50 μm DB-FFAP column (Agilent); carrier gas was helium at 28.819 mL min^{-1}; inlet conditions were 250 °C with a 5 mL min^{-1} purge flow at a 4:1 split ratio. The initial oven temperature of 70 °C was increased by 70 °C over 1 min, increased by 100 °C over 1 min, and held at 240 °C for 1.8 min. Chromatogram peaks of SCFA standards were used to derive a retention time window and quantification was based on plotting each standard against detector response (x) with a 1/x weighting factor.

DSS colitis mouse model. All animal studies were performed in facilities accredited by the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC). Male C57Bl/6 mice (Taconic Biosciences) at 6–8 weeks of age were randomized into cages each containing a treatment group of eight mice at Biomodels LLC (Waltham MA, USA). Mice were kept in rooms provided with filtered air at 65–75 °F with 30–70% relative humidity. Rooms were on a 12 h:12 h light:dark cycle with no twilight and a minimum of 12–15 air changes per hour. Animals were fed diet 5053 (Lab Supply, Fort Worth TX, USA) throughout the study with water provided *ad libitum*. All animals were weighed daily and assessed visually for the presence of diarrhea and/or bloody stool. Mice with >30% weight loss were euthanized. Colitis was induced by supplementing the drinking water with 2.5% dextran sodium sulfate (DSS) from days 0–5. Glycans were administered in the drinking water as 1% (v/v) filter-sterilized solutions from days 7–14.

Stool consistency for each mouse was scored daily (Supplementary Table 5). Colitis in each mouse was assessed by video endoscopy under isoflurane anesthesia on day 14. During each endoscopic procedure, images were recorded and colitis severity was scored by a blinded observer (Supplementary Table 6). Histopathological changes in tissue samples were assessed by a board-certified veterinary pathologist at Inotiv (Missouri, USA). Tissue samples were prepared by fixation in formalin and longitudinal and cross-sections of proximal and distal colon were processed, embedded in paraffin, sectioned at ~5–6 μm, mounted on glass slides, stained with hematoxylin and eosin (H&E) examined microscopically, and scored using a qualitative and semi-quantitative grading system (Supplementary Table 7).

***C. difficile* infection mouse model**. Female C57Bl/6 mice (Harlan Laboratories) weighing 16–18 g were randomly allocated into treatment groups of 12 animals (3 animals per cage) at Trans Pharm Pre-clinical Solutions (Jackson MI, USA). Animal room conditions were the same as in the DSS colitis study. Mice were fed Teklad Global Rodent Diet 2918 (Harlan Laboratoires) and water *ad libitum*. All mice received an antibiotic cocktail (0.5 mg ml^{-1} kanamycin, 0.044 mg ml^{-1} gentamicin, 1062 U ml^{-1} colistin, 0.27 mg ml^{-1} metronidazole, 0.16 mg ml^{-1} ciprofloxacin, 0.1 mg ml^{-1} ampicillin, 0.06 mg ml^{-1} vancomycin, 1% (v/v) glucose) in their drinking water from day 14–5 followed by a single oral dose of 10 mg kg^{-1} clindamycin on day 3. On day 0, mice received an oral gavage of 3×10^4 viable *C. difficile* spores (strain VPI 10463 ATCC 43255). Mice in the vancomycin control group received oral gavage of 50 mg kg^{-1} vancomycin daily from day 0–4. Glycans were administered in the drinking water as 1% (v/v) filter-sterilized solutions from day 1–6. Glycan efficacy was assessed by daily measurements of animal survival, body weight, and a clinical score of disease severity (Supplementary Table 8). Mouse feces were placed into a 96-well plate that was kept on wet ice and then frozen at −80 °C. Separate disposable forceps were used for each cage of animals to avoid cross-contamination.

Statistical analysis. Statistical analyses were performed using R software. Box plots show median and interquartile ranges. Values are presented as the mean ± SD for ex vivo fermentation dynamics, bacterial growth, and glycan linkage analyses and as the mean ± SEM in mouse models. Comparisons of ex vivo cultures fermenting reference glycans versus SGs were evaluated by two-sided Wilcoxon rank-sum test for Shannon diversity, species richness, and enteric pathogen growth and relative abundances. Differences in Shannon diversity between cultures fermenting individual glycans were evaluated by Kruskal–Wallis followed by Dunn's comparison test. Statistical significance for gas production experiments was performed by Tukey's test after confirming the data adhered to a normal distribution by Shapiro Wilks test. Statistical significance for abundance changes in genera and CAZyme families across individuals was determined by fitting a linear mixed effect model on rank transformed genera/CAZyme abundance data with glycan treatment as fixed effect and subject as random effect, and corrected to a false discovery rate < 0.05.

In mouse models, animals were allocated randomly to each treatment group and different groups were processed identically. Body mass changes are based on the area under the curve for each individual mouse. Statistical significance of comparisons between no-glycan controls and other treatments for body mass, stool

scores, endoscopy scores, and histology scores were evaluated by two-sided Wilcoxon rank-sum test. Differences in survivorship between no-glycan controls and other treatments were calculated by log-rank test.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The DNA sequencing reads for this study are available in the NCBI SRA database as project accessions PRJNA800419, PRJNA800763, PRJNA800792, and PRJNA800405. Supporting data are available in the Supplementary Information, Supplementary Data, and Source data file. All requests will be reviewed by Kaleido Biosciences to verify whether the request is subject to any intellectual property or confidentiality obligations. Source data are provided with this paper.

Code availability

The methods and code for the R package phgrofit (version 1.0.2) to extract physiological descriptors from the kinetic pH and OD_{600} data have been deposited in Github and are freely available[57].

Received: 24 August 2021; Accepted: 8 February 2022;
Published online: 10 March 2022

References

1. Vatanen, T. et al. Variation in microbiome LPS immunogenicity contributes to autoimmunity in humans. *Cell* **165**, 842–853 (2016).
2. Duscha, A. et al. Propionic acid shapes the multiple sclerosis disease course by an immunomodulatory mechanism. *Cell* **180**, 1067–1080.e16 (2020).
3. Routy, B. et al. Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. *Science* **359**, 91–97 (2018).
4. Dabke, K., Hendrick, G. & Devkota, S. The gut microbiome and metabolic syndrome. *J. Clin. Invest.* **129**, 4050–4057 (2019).
5. Bajaj, J. S. The role of microbiota in hepatic encephalopathy. *Gut Microbes* **5**, 397–403 (2014).
6. Taroncher-Oldenburg, G. et al. Translating microbiome futures. *Nat. Biotechnol.* **36**, 1037–1042 (2018).
7. Patnode, M. L. et al. Interspecies competition impacts targeted manipulation of human gut bacteria by fiber-derived glycans. *Cell* **179**, 59–73.e13 (2019).
8. Martens, E. C. et al. Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol.* **9**, e1001221 (2011).
9. Baxter, N. T. et al. Dynamics of human gut microbiota and short-chain fatty acids in response to dietary interventions with three fermentable fibers. *MBio* https://doi.org/10.1128/mBio.02566-18 (2019).
10. Zhao, L. et al. Gut bacteria selectively promoted by dietary fibers alleviate type 2 diabetes. *Science* **359**, 1151–1156 (2018).
11. Menni, C. et al. Gut microbiome diversity and high-fibre intake are related to lower long-term weight gain. *Int. J. Obes. 2005* **41**, 1099–1105 (2017).
12. O'Keefe, S. J. D. et al. Fat, fibre and cancer risk in African Americans and rural Africans. *Nat. Commun.* **6**, 6342 (2015).
13. Holscher, H. D. et al. Effects of prebiotic-containing infant formula on gastrointestinal tolerance and fecal microbiota in a randomized controlled trial. *JPEN J. Parenter. Enter. Nutr.* **36**, 95S–105S (2012).
14. Walton, G. E. et al. A randomised crossover study investigating the effects of galacto-oligosaccharides on the faecal microbiota in men and women over 50 years of age. *Br. J. Nutr.* **107**, 1466–1475 (2012).
15. Lecerf, J.-M. et al. Xylo-oligosaccharide (XOS) in combination with inulin modulates both the intestinal environment and immune status in healthy subjects, while XOS alone only shows prebiotic properties. *Br. J. Nutr.* **108**, 1847–1858 (2012).
16. Hong, L. et al. Pullulan nanoparticles as prebiotics enhance the antibacterial p[roperties of Lactobacillus plantarum through the induction of mild stress in probiotics. *Front. Microbiol.* **10**, 142 (2019).
17. Ballongue, J., Schumann, C. & Quignon, P. Effects of lactulose and lactitol on colonic microflora and enzymatic activity. *Scand. J. Gastroenterol. Suppl.* **222**, 41–44 (1997).
18. Von Maltzahn, G., Silverman, J., Yamanaka, Y., Milwid, J. & Geremia, J. Glycan therapeutics and related methods thereof. US Patent 10,314,853 (2019).
19. Geremia, J., Liu, C. & Murphy, A. Methods of producing glycan polymers. US Patent app 16/761115 (2020).
20. Hansen, C. H. F. et al. Targeting gut microbiota and barrier function with prebiotics to alleviate autoimmune manifestations in NOD mice. *Diabetologia* **62**, 1689–1700 (2019).
21. Lewis, S., Burmeister, S. & Brazier, J. Effect of the prebiotic oligofructose on relapse of *Clostridium difficile*-associated diarrhea: a randomized, controlled study. *Clin. Gastroenterol. Hepatol.***3**, 442–448 (2005).
22. Whelan, K. et al. Fructooligosaccharides and fiber partially prevent the alterations in fecal microbiota and short-chain fatty acid concentrations caused by standard enteral formula in healthy humans. *J. Nutr.* **135**, 1896–1902 (2005).
23. Litvak, Y., Byndloss, M. X. & Bäumler, A. J. Colonocyte metabolism shapes the gut microbiota. *Science* **362**, eaat9076 (2018).
24. Berggren, A. M., Nyman, E. M., Lundquist, I. & Björck, I. M. Influence of orally and rectally administered propionate on cholesterol and glucose metabolism in obese rats. *Br. J. Nutr.* **76**, 287–294 (1996).
25. El Hage, R., Hernandez-Sanabria, E., Calatayud Arroyo, M., Props, R. & Van de Wiele, T. Propionate-producing consortium restores antibiotic-induced dysbiosis in a dynamic in vitro model of the human intestinal microbial ecosystem. *Front. Microbiol.* **10**, 1206 (2019).
26. Smith, P. M. et al. The microbial metabolites, short-chain fatty acids, regulate colonic Treg cell homeostasis. *Science* **341**, 569–573 (2013).
27. Arpaia, N. et al. Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. *Nature* **504**, 451–455 (2013).
28. Darwin, W. C. & Cord-Ruwisch, R. Concurrent lactic and volatile fatty acid analysis of microbial fermentation samples by gas chromatography with heat pre-treatment. *J. Chromatogr. Sci.* **56**, 1–5 (2018).
29. Seo, A. Y., Kim, N. & Oh, D. H. Abdominal bloating: pathophysiology and treatment. *J. Neurogastroenterol. Motil.* **19**, 433–453 (2013).
30. Azpiroz, F. Intestinal gas dynamics: mechanisms and clinical relevance. *Gut* **54**, 893–895 (2005).
31. Takahashi, K. et al. Reduced abundance of butyrate-producing bacteria species in the fecal microbial community in Crohn's disease. *Digestion* **93**, 59–65 (2016).
32. Takeshita, K. et al. A single species of clostridium subcluster XIVa decreased in ulcerative colitis patients. *Inflamm. Bowel Dis.* **22**, 2802–2810 (2016).
33. Cruz, P. D. et al. Association between specific mucosa-associated microbiota in Crohn's disease at the time of resection and subsequent disease recurrence: A pilot study. *J. Gastroenterol. Hepatol.* **30**, 268–278 (2015).
34. Meyer, D. & Stasse-Wolthuis, M. The bifidogenic effect of inulin and oligofructose and its consequences for gut health. *Eur. J. Clin. Nutr.* **63**, 1277–1289 (2009).
35. Stoma, I. et al. Compositional flux within the intestinal microbiota and risk for bloodstream infection with gram-negative bacteria. *Clin. Infect. Dis.* **73**, e4627–e4635 (2020).
36. Liss, B. J. et al. Intestinal colonisation and blood stream infections due to vancomycin-resistant enterococci (VRE) and extended-spectrum beta-lactamase-producing Enterobacteriaceae (ESBLE) in patients with haematological and oncological malignancies. *Infection* **40**, 613–619 (2012).
37. Dubin, K. & Pamer, E. G. Enterococci and their interactions with the intestinal microbiome. *Microbiol. Spectr.* https://doi.org/10.1128/microbiolspec.BAD-0014-2016 (2014).
38. Hallmann, F. Toxicity of commonly used laxatives. *Med. Sci. Monit.* **6**, 618–628 (2000).
39. Morgan, M. Y. Current state of knowledge of hepatic encephalopathy (part III): non-absorbable disaccharides. *Metab. Brain Dis.* **31**, 1361–1364 (2016).
40. Boutard, M. et al. Functional diversity of carbohydrate-active enzymes enabling a bacterium to ferment plant biomass. *PLoS Genet.* **10**, e1004773 (2014).
41. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495 (2014).
42. Salyers, A. A., West, S. E., Vercellotti, J. R. & Wilkins, T. D. Fermentation of mucins and plant polysaccharides by anaerobic bacteria from the human colon. *Appl. Environ. Microbiol.* **34**, 529–533 (1977).
43. Koropatkin, N. M., Cameron, E. A. & Martens, E. C. How glycan metabolism shapes the human gut microbiota. *Nat. Rev. Microbiol.* **10**, 323–335 (2012).
44. Di Lauro, B. et al. Isolation and characterization of a new family 42 β-galactosidase from the thermoacidophilic bacterium *Alicyclobacillus acidocaldarius*: Identification of the active site residues. *Biochim. Biophys. Acta* **1784**, 292–301 (2008).
45. Milani, C. et al. Bifidobacteria exhibit social behavior through carbohydrate resource sharing in the gut. *Sci. Rep.* **5**, 15782 (2015).
46. Rietschel-Berst, M. et al. Extracellular exo-beta-galactofuranosidase from *Penicillium charlesii*: isolation, purification, and properties. *J. Biol. Chem.* **252**, 3219–3226 (1977).
47. Sokol, H. et al. Prominence of ileal mucosa-associated microbiota to predict postoperative endoscopic recurrence in Crohn's disease. *Gut* **69**, 462–472 (2020).
48. Wang, T. et al. Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. *ISME J.* **6**, 320–329 (2012).

49. Crobach, M. J. T. et al. The bacterial gut microbiota of adult patients infected, colonized or noncolonized by *Clostridioides difficile*. *Microorganisms* **8**, 677 (2020).
50. Mullish, B. H. et al. Microbial bile salt hydrolases mediate the efficacy of faecal microbiota transplant in the treatment of recurrent *Clostridioides difficile* infection. *Gut* **68**, 1791–1800 (2019).
51. Wang, K. et al. *Parabacteroides distasonis* alleviates obesity and metabolic dysfunctions via production of succinate and secondary bile acids. *Cell Rep.* **26**, 222–235.e5 (2019).
52. Kottler, R. et al. Development of a high-throughput glycoanalysis method for the characterization of oligosaccharides in human milk utilizing multiplexed capillary gel electrophoresis with laser-induced fluorescence detection. *Electrophoresis* **34**, 2323–2336 (2013).
53. Hennig, R. et al. Towards personalized diagnostics via longitudinal study of the human plasma N-glycome. *Biochim. Biophys. Acta* **1860**, 1728–1738 (2016).
54. Hakomori, S. A rapid permethylation of glycolipid, and polysaccharide catalyzed by methylsulfinyl carbanion in dimethyl sulfoxide. *J. Biochem. (Tokyo)* **55**, 205–208 (1964).
55. Rovalino-Córdova, A. M., Fogliano, V. & Capuano, E. Effect of bean structure on microbiota utilization of plant nutrients: An *in-vitro* study using the simulator of the human intestinal microbial ecosystem (SHIME®). *J. Funct. Foods* **73**, 104087 (2020).
56. Heux, S., Philippe, B. & Portais, J.-C. High-throughput workflow for monitoring and mining bioprocess data and its application to inferring the physiological response of *Escherichia coli* to perturbations. *Appl. Environ. Microbiol.* **77**, 7040–7049 (2011).
57. Bayne, C. *phgrofit R package (version 1.0.2)*. https://doi.org/10.5281/zenodo.5948925 (2022).
58. Edgar, R. C. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. *bioRxiv* https://doi.org/10.1101/081257 (2016).
59. Callahan, B. J. et al. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
60. Hillmann, B. et al. Evaluating the information content of shallow shotgun metagenomics. *MSystems* https://doi.org/10.1128/mSystems.00069-18 (2018).
61. Oksanen, J. et al. *vegan: Community Ecology Package. R package (version 2.5-6)*. https://CRAN.R-project.org/package=vegan (2020).
62. Zhang, H. et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **46**, W95–W101 (2018).
63. Al-Ghalith, G. & Knights, D. BURST enables mathematically optimal short-read alignment for big data. *bioRxiv* https://doi.org/10.1101/2020.09.08.287128 (2020).
64. Han, J., Lin, K., Sequeira, C. & Borchers, C. H. An isotope-labeled chemical derivatization method for the quantitation of short-chain fatty acids in human feces by liquid chromatography-tandem mass spectrometry. *Anal. Chim. Acta* **854**, 86–94 (2015).

Acknowledgements

We thank the Kaleido Discovery and Clinical groups for their hard work and expertise. We thank Stephan Reiling, Michael Mahowald, Tanya Yatsunenko, Adarsh Jose, Jon Lawrence, and Uli Binné for helpful discussions, AMRI for SCFA analyses, and L Jung (PrecisionScientia) for critical reading of the manuscript. This study was funded by Kaleido Biosciences.

Author contributions

A.C.T., G.v.M., C.M.L., and J.E.v.H.V. designed research. N.B., C.B., B.M., J.M., Y.M., L.L., G.L., C.M., R.K., and E.R. performed research. A.C.T., N.B., C.B., J.T., J.L, and R.K. analyzed data. A.C.T., P.J.T., and J.E.v.H.V. wrote the paper.

Competing interests

P.J.T. is on the Kaleido Biosciences Scientific Advisory Board. R.K. and E.R. are employees of GlyXera. All other authors either are or were employees of Kaleido Biosciences.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-28856-x.

Correspondence and requests for materials should be addressed to Andrew C. Tolonen or Johan E. T.van Hylckama Vlieg.

Peer review information *Nature Communications* thanks Rebeca Martin, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

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