Evolution of a biomass-fermenting bacterium to resist lignin phenolics

Tristan Cerisy1-4*, Tiffany Souterre1-4*, Ismael Torres-Romero1-4, Magali Boutard1,2, Ivan Dubois1,2, Julien Patrouix1,2, Karine Labadie1,2, Wahiba Berrabah1,2, Marcel Salanoubat1-4, Volker Doring1-4, Andrew Tolonen1-4†

1CEA, Genoscope, Évry France
2CNRS-UMR8030, Évry France
3Université Paris-Saclay, Évry France
4Université d’Évry, Évry France

*co-first authors
†correspondence: atolonen@genoscope.cns.fr
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 by changes in gene dosage and regulation, membrane fatty acid structure, and the surface layer. Collectively, this study demonstrates an automated framework for 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 could 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.

Introduction

Fermentation of lignocellulosic biomass by bacteria like *Clostridium phytofermentans* is central to the functioning 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.
Plant biomass is primarily composed of a macromolecular network of polysaccharides linked with lignin, a polymer of phenylpropanoid subunits with aromatic rings of varying degrees of methoxylation. Thus, when microorganisms hydrolyze lignocellulose to access sugars, they also liberate 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, but hydrolysates generally contain inhibitors at concentrations that impede growth of microorganisms by damaging the cell membrane, metabolic enzymes, and nucleic acids. The most abundant aliphatic acids are generally acetate, particularly in acetylxylan-rich hardwoods, and formate from furan breakdown. The main furans are furfural and hydroxymethylfurfural (5-HMF) that 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.

The resistance of model, sugar-fermenting bacteria such as *E. coli* to biomass inhibitors has been well-studied for aliphatic acids, furans, and phenolics. 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, developing 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.

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, H₂, 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.

**Methods**

**Cell cultivation:** *C. phytofermentans* ISDg (ATCC 700394) was cultured anaerobically in GS2 medium (15). Growth of batch cultures containing inhibitors (Table S1) was measured in 100-well microtiter plates (Bioscreen 9502550) containing 400 μl GS2 medium with 3 g L⁻¹ glucose supplemented with a given inhibitor neutralized to pH 7. Wells were inoculated with 1:10 volume cells grown to log phase in minus-inhibitor medium. The plates were sealed in the anaerobic chamber (2% H₂, 98% N₂) by press-fitting adhesive sheets (Qiagen 1018104) (16) and incubated at 37°C in a Thermo
Scientific Bioscreen C. The cell densities (OD_{600}) were measured every 15 minutes with 30 seconds shaking before each reading. Cellulose cultures were inoculated into GS2 containing 10 g L^{-1} cellulose (0.5×5 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 μ filters by vacuum filtration and drying it overnight at 65°C (17).

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% N2 gas. A 50 ml culture was maintained at 30°C with optical density readings every 30 seconds, and was transferred between growth chambers every 12 h to clean the empty chamber with 5N sodium hydroxide.

Cells were acclimated to increased ferulate using medium-swap mode, a modified chemostat (6h generation time) with dilutions every 30 minutes of stressing medium (high ferulate) if the cell density exceeded the density threshold (measured as OD_{880} 30, which is equivalent to OD_{600} 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 the WT strain in minus-ferulate medium. In turbidostat mode, 20% of the culture volume was replaced with fresh medium each time the cell density reached OD_{880} 30. Initially, the stressing medium contained 1 g L^{-1} ferulate, the highest concentration at which a WT culture could 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 L^{-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 L\(^{-1}\) ferulate, respectively.

**RNA-seq:** Log phase cultures (OD\(_{600}\) 0.8) of WT *C. phytofermentans* ISDg were diluted with 1 volume medium either lacking ferulate (-ferulate) or containing 4 g L\(^{-1}\) ferulate (+ferulate, 2 g L\(^{-1}\) 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 afterwards. Total RNA was extracted using TRI reagent (Sigma 93289) and 20 μg RNA was treated with 4U Turbo DNase (Ambion AM2238) for 30 min at 37°C. RNA was purified by Zymoclean (Zymo Research R1015) to capture RNA>200 bp. Five μg total RNA was depleted of rRNA by Ribo-Zero (Illumina MRZMB126), yielding 200-400 ng RNA, and purified by Zymo Concentrator-5 (total capture) into 10 μl water. cDNA libraries were prepared from 100 ng RNA using the Truseq Stranded mRNA kit (Illumina 15031047) and sequenced on an Illumina HiSeq2000 sequencer with paired-end 150 bp reads. Reads were aligned to the *C. phytofermentans* ISDg genome (NCBI NC_010001.1) using Bowtie 2 (18). Gene expression was calculated as reads per kb of gene per million reads (RPKM) using the easyRNASEq Bioconductor package (19). Differential expression was defined as a greater than 4-fold change in expression and a DESeq (20) p-value<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-B), CFY2 (2 clones CFY2C-D), and CFY3 (4 clones CFY3E-H). Genomic DNA (15-20 μg) was extracted from 3 ml cultures using the Sigma GenElute Bacterial Genomic DNA kit (NA2110). DNA (100-250 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 Immobilisation (SPRI) beads (Beckman Coulter A63880) and amplified by 12 cycles PCR using Kapa Hifi HotStart NGS library Amplification kit (Kapa Biosystems KK2611) with P5/P7 primers. PCR products were purified (0.8 volume SPRI beads), run on a 2% agarose gel, and DNA (700-800 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 NC_010001.1) using Bowtie 2 (18). Sequence variants (SNPs, indels) in the CFY strains relative to the reference genome were identified using the GATK (21) as described previously (22). Structural variations were detected using the Breseq split-read analysis tool (23). Insertion sequences (IS) were identified using ISfinder (24).

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 minutes. DNA was treated using IrysPrep® Reagent Kit (BioNano Genomics) to prepare NLRS (Nicked, Labeled, Repaired and Stained) 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-labelled Alexa546-dUTP and Taq Polymerase (NEB M0273). Nicks were ligated using Taq ligase (NEB M0208) with 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 ≥1000-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 a N50 of 172 kb. The NCBI assembly (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 2,689,393-3,023,191 joined by an ISL3-2 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 (25). 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 qPCR (KAPA KK4621) with primers in Table S2. Expression values are means of triplicate
measurements of duplicate cultures calculated as 2^{Ct} (26), 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 directly taken 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 relative to the 2^{Ct} of 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 growth in GS2 medium containing 6 g L^{-1} ferulate by LC/ESI-MS and 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 HESI source. HPLC separation was performed on a 5 µm, 4.6 × 150 mm Sequant ZICpHILIC 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 ± ferulate WT and CFY3E cultures and -
ferulate CFY1B, CFY2C, CFY2D, and CFY3F cultures. Late log phase cells were
collected by centrifugation from cultures grown in medium either with 2 g L⁻¹ ferulate or
lacking ferulate. Fatty acid methyl esters (FAME) were obtained from 100 mg cells by
saponification, methylation, and extraction (27) 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 x 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 (v/v/v) 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
(v/v/v). Polar lipids were resolved by 2D silica gel thin layer chromatography: dimension
1 was chloroform:methanol:water (65:25:4 v/v/v), dimension 2 was
chloroform:methanol:acetic acid:water (80:12:15:4 v/v/v/v). Total lipids were detected
using molybdatophosphoric acid and specific functional groups identified using spray
reagents specific for defined functional groups (28).

Results
Native *C. phytofermentans* inhibitor resistance: We measured growth of *C. phytofermentans* in various concentrations of 12 lignocellulosic inhibitors (Table S1) to gain a general understanding of the relative effects of aliphatic acids, furans, and phenolics (Fig 1, S1). 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 1 C-D), similar to other bacteria that reduce and detoxify furans (29), (30). Growth lags are proposed to be due to alcohol dehydrogenase (ADH) reducing furan, causing NADH depletion and acetaldehyde accumulation (31). Supportingly, the ADH protein Cphy1179 shares 28% amino acid identity with a furfural-reducing, Zn-dependent ADH (32). However, if *C. phytofermentans* detoxifies furans, this mechanism is abruptly overwhelmed at concentrations above 2 g L\(^{-1}\) 5-HMF and 1 g L\(^{-1}\) 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-F) are more toxic than hydroxybenzoic acids (Fig 1G,H), supporting 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-K) and catechol (Fig 1L). For 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 (33).

**Genome-wide mRNA expression during ferulate stress:** We quantified genome-wide mRNA expression changes at two timepoints (t=0.5 h, t=4 h) following supplementation of mid-log cultures with 2 g L⁻¹ ferulate, which reduced growth (Fig 2A) similar to the initial growth screen (Fig 1F). Three to five million read pairs were aligned to the genome for each culture (Table S3A) to calculate gene expression levels (Table S3B). The number of differentially-expressed genes (Table S3C-E) increased from 0 genes before ferulate addition (Fig 2B) to 78 genes after 30 minutes (Fig 2C), then declined to 47 genes after 4 hours (Fig 2D). The most abundant functional categories of differentially-expressed genes at t=0.5 h relate to 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 (34). Lipid genes include the *fab* gene cluster (cphy0516-23) for fatty acid biosynthesis, which was strongly repressed at t=0.5 h (Fig 2E) and recovered by t=4 h. While cultures continued active growth after the sampling, many of the differences between ± ferulate cultures at t=4 h indicate the minus-ferulate treatment had depleted nutrients in the medium, triggering expression of genes to assimilate alternative carbohydrates (Fig 2F).

Gene expression at t=0.5 h shows abrupt ferulate stress induces expression of genes encoding the efflux pump *cphy1055-6*, which is similar to *E. coli mdlAB* conferring resistance to organic solvents (35). Many of the genes up-regulated at t=0.5 h are co-located in two genomic regions. The first region encodes *tad* (tight adherence)
cphy0029-40 genes for Flp-type type IV pili assembly. Type IV pili are widespread in clostridia (36) for adhesion to solid substrates to form protective biofilms (37), reflecting how ferulate represses motility genes in C. beijerinckii (38). The other cluster cphy1838-45 includes genes for the flavin mononucleotide (FMN) binding proteins WrbA (39) and two NADPH:FMN reductases. NADPH:FMN reductase inactivation confers ferulate resistance in C. beijerinckii by an unknown mechanism (40). While this appears in opposition to our data that NADPH:FMN reductases are up-regulated 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 the metal coordination with E. coli YhhW, which cleaves the plant phenolic quercetin (41). C. phytofermentans may thus up-regulate genes to transform or detoxify plant phenolics, similar to some ruminal clostridia (42).

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 2-5 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, growth rate at
the higher ferulate concentration was improved in turbidostat mode (Fig 3B). We initiated the growth selection with stressing medium containing 1 g L\(^{-1}\) ferulate, the highest concentration at which we could establish a stable 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 3 g L\(^{-1}\) ferulate medium as WT in the absence of ferulate (Fig 3C). Clones isolated along the growth selection are progressively more ferulate resistant in batch culture (Fig 3D-G); while no growth was observed above 2 g L\(^{-1}\) ferulate in the WT strain (Fig 3D), CFY3 clones grow robustly at the ferulate solubility limit (6 g L\(^{-1}\)) (Fig 3G). We assessed the ferulate resistance of 2 clones from each of the CFY1 (CFY1A-B) and CFY2 (CFY2C-D) time points and 4 clones from the CFY3 time point (CFY3E-H). The duplicate CFY1 and CFY2 clones showed similar ferulate resistance, but CFY3H is much less ferulate-resistant than the 3 other clones (Fig S2), showing 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 WT (Fig 4A) and show accelerated cellulose degradation in medium supplemented with ferulate (Fig 4B), supporting 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 (Fig S3), 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...
the ferulate concentration was unaltered with no products corresponding to reduced, demethoxylated, or decarboxylated ferulate (Fig S4). Thus, even though C. phytofermentans up-regulates potential phenol-degrading enzymes in response to ferulate, the cell adapted to ferulate by reinforcing the cell or excluding this molecule, rather than detoxifying it.

As toxicity of aromatic molecules is often associated with disruption of the cell membrane, we profiled fatty acids (FA) to determine if ferulate resistance is associated with altered membrane phospholipids (Table S4). 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 minus-ferulate medium was 185-fold higher than WT. Related clostridia similarly increase plasmalogens in response to aliphatic alcohol stress (43) (44), likely to fine tune membrane fluidity and protect from redox-mediated damage (45). The distribution of FA chain lengths in WT cells (Fig 4D) is similar to other clostridia, but with fewer unsaturated FA and more cyclopropanes (46), both of which reduce membrane fluidity to protect from solvent stress (47). While the addition of ferulate had little immediate effect on the FA chains of WT cells (Fig S5A-B), the CFY strains showed altered FA relative to WT in the absence of ferulate (Fig 4D). The CFY1B FA profile was the most perturbed with increased hydroxylated C16 and unsaturated fatty acids, largely C18:1, which is associated with increased ethanol tolerance in E. coli (48). CFY3F shifted to branched FA (especially C15) and longer chain lengths (C18, C20), which increases membrane rigidity (10) to potentially combat the membrane fluidizing effects of ferulate.
C. phytofermentans fatty acids are decorated with a diversity of phospo-, 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 CFY1-3 clones giving between 106 and 705-fold coverage (Table S5A) to identify DNA variants relative to wild-type (Table S5B). Seven single-nucleotide variations (SNV) and short insertions/deletions (indels) are present in all the CFY genomes (Fig 5A), which likely fixed in the population during an early selective sweep. These variants caused non-synonymous changes in 5 proteins including a homolog of Cap5F (Cphy3503), a protein for biosynthesis of capsular polysaccharides (49) that is associated with biofilm formation (50) and stress resistance (51). Strains subsequently accrued strain-specific mutations consistent with the population exploring alternative mutational pathways to improve ferulate resistance, particularly by modifying sensor kinases that could 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 (52), the fatty acid dehydratase FabZ (Cphy0520), and the reductase FabV (Cphy1286) for the final step in fatty acid elongation. The genomes of CFY3E-G (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 (53). The S-layer is a protein lattice that
provides mechanical stabilization, sites for extracellular protein attachment, and a selective barrier for molecules (54).

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 (cphy2241-3) was duplicated in the CFY3 strains (Fig S6A-B). 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 (55). Duplication of this putative REP increases the mRNA secondary structure of the cphy2243-2242 intergenic region (Fig S6B), supporting it functions similar to REP that increase expression by forming stable stem-loop structures that protect mRNA from ribonucleases (56) (57). Supportingly, we found that mRNA expression of the two genes downstream of the insertion were elevated (Fig 5B), which could have increased fitness because the GM3 growth selections were done in glucose medium. The mRNA expression of genes in two co-located operons with upstream point mutations were up-regulated in the CFY strains (Fig 5C). The A-to-G transition upstream of cphy1464 created a TG di-nucleotide 2 bp upstream of the Pribnow hexamer (Fig S6A) that enhances transcription in other bacteria (58) (59) and is present in the consensus -10 promoter sequence in C. phytofermentans (60). We propose the up-regulated operons cphy1459-61 and cphy1464-5 either enable increased production of 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 (Table S6) including 12 ISL3 comprised of 2 isoforms: 8 ISL3-1 elements and 4 ISL3-2 elements. IS elements inactivate genes through their transposition and act as substrates for homologous recombination. In addition, the IR of all 12 ISL3 contain a 5′-TTGACA-3′ sequence matching an outward-facing, consensus -35 box from this organism (60) (Table S6), suggesting ISL3 could activate expression of adjacent genes (61). An ISL3-1 was precisely deleted in all CFY genomes as evidenced by reduced read coverage (Fig 6A) as well as BioNano optical mapping and Sanger sequencing (Fig S7A-B), showing ISL3 are active in C. *phytofermentans*. Further, the CFY2C,D 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 ISL3-2 insertion (Fig 6C, Fig S7C). We did not observe any extrachromosomal DNA by pulsed-field gel in the CFY2C strain, supporting the *cphy2178*-ISL3-*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 localize 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*-ISL3-*cphy2461* fragment (Fig 6D). The duplication arose between day 13 and 20 and overtook the population to comprise 68% cells by day 40, supporting it was the subject of positive selection. Subsequently, this variant declined in the population, representing 1% 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 influences 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 C16 and decrease of C17 cyclopropanes, which are synthesized from C16:1. 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 up-regulates transcription of genes for efflux pumps, biofilm formation, and flavoproteins including two NADPH:FMN reductases that are associated with ferulate stress in *C. beijerinckii* (40). 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-positives 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 (54), potentially excluding ferulate, and stabilizes the cell through non-covalently linkage to cell wall polysaccharides using threonine residues (62). A three residue (NTT) 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-G); the CFY3 strain with lower resistance (CFY3H) had a N573D variant in Cphy3510. S-layer proteins differ greatly among bacteria and neither of the Cphy3510 variants are in known domains, but our results support that modification of the S-layer could be an effective strategy to improve inhibitor resistance.

In addition to minor genomic changes (SNV 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 ISL3-2 insertion, supporting this large genome rearrangement was positively-selected during ferulate selection. Tandem duplications of large chromosomal regions have been detected in other bacteria (55) 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 recombined with each other, the intervening region was deleted to leave a single, central IS element (63). 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 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 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.

References

1. Metzger JO, Hüttermann A. 2009. Sustainable global energy supply based on
lignocellulosic biomass from afforestation of degraded areas. Naturwissenschaften

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

http://dx.doi.org/10.1007/10_2007_070


24


Acknowledgements

This work was funded by the Genoscope-CEA, a CNRS Chaire d’Excellence (to ACT), and the Agence Nationale de la Recherche Grant ANR-16-CE05-0020 (to ACT). We thank P Marlière for the GM3, E Darii for mass spectrometry, and the Genoscope-CEA sequencing platform for DNA and RNA sequencing.

Competing Interests

The authors declare no competing financial interests.

Data Availability

The authors confirm that 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 ERP018602 (RNA-seq) and ERP018603 (whole genome sequencing).

Figures

Fig 1 C. phytofermentans growth GS2 glucose medium containing different concentrations of A acetate, B formate, C 5-hydroxymethylfurfural, D furfural, E coumarate, F ferulate, G vanillate, H 4-hydroxybenzoic acid, I vanillin, J benzaldehyde, K syringaldehyde, L catechol. Colors show inhibitor concentrations (g L⁻¹): 0 (gray), 1
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 L\(^{-1}\) 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. mRNA expression from cultures at **B** t=0 h, **C** t=0.5 h, **D** t=4 h. Differentially-expressed genes in **B-D** are red triangles; unchanged genes are gray circles. Five most abundant COG functional categories (64) of differentially-expressed genes at **E** t=0.5 h and **F** t=4 h. Positive y-axis is up-regulated genes and negative y-axis is repressed genes.

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 otherwise with relaxing medium. **B** Growth rate was improved using turbidostat mode in which the culture was diluted each time it reached the threshold. In **A** and **B**, dashed red lines show cell density threshold (OD\(_{880}\) 30) and red arrows are 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 6h generation time. Black line shows average daily generation time (h) during turbidostat growth selection. Red points are sample times for physiology and genome sequences (CFY1, CFY2, CFY3). Batch culture growth (OD\(_{600}\)) of **D** wild-type and clones **E** CFY1A, **F** CFY2C, and **G** CFY3E in GS2 glucose medium containing either 0 (gray), 1 (purple), 2 (blue), 4 (yellow), or 6 (red) g L\(^{-1}\) ferulate. Data shows mean cell density (OD\(_{600}\)) of 4 cultures ± sd.

Fig 4 Growth physiology and membrane fatty acid composition of *C. phytofermentans* WT and GM3 strains. Rate of cellulose degradation by CFY3E (blue triangles) and WT (red circles) in **A** medium lacking ferulate and **B** medium supplemented with 2 g L\(^{-1}\) ferulate. **C** Plasmalogen content expressed as percent of total fatty acids of WT and GM3 strains grown in presence (+F) or absence (-F) of ferulate. **D** Cellular fatty acid profiles of log phase WT, CFY1B, CFY2C, and CFY3F cultures in -ferulate medium. Fatty acids classified by acyl chain length (C12-20) and whether acyl chains were saturated (dark blue), unsaturated (light blue), hydroxylated (green), cyclopropane (red), branched (yellow). Data in **A-C** show mean of duplicate cultures ± sd.

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 CFY1-3 time points. mRNA expression of the **B** *cphy2241-3* operon and the **C** *cphy1459-cphy1465* genes 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\(^{-\Delta\Delta C_{T}}\) normalized to 16S rRNA expression; bars show means of triplicate measurements ± sd.
Fig 6 ISL3-associated genome changes in CFY strains. A Read coverage showing deletion of the ISL3-1 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 2,689,393-3,023,191). C The cphy2178-cphy2461 region is a tandem duplication joined by an ISL3-2 element. Positions of single molecule restriction fragments >400 kb shown by BioNano optical mapping to span the duplicated region are shown. D Relative abundance of the cphy2178-ISL3-cphy2461 junction in the GM3 culture from day 6 to 83. Genomic locations of the PCR primers are shown in C. Abundance was measured by qPCR, calculated as $2^{-Ct}$ of the junction fragment in the GM3 culture relative to the purified CFY2C, which has the junction present in 100% cells.