

Expanded genome and proteome reallocation in a novel, robust *Bacillus coagulans* strain capable of utilizing pentose and hexose sugars

David Dooley,¹ Seunghyun Ryu,^{1,2} Richard J. Giannone,^{2,3} Jackson Edwards,⁴ Bruce S. Dien,⁴ Patricia J. Slininger,⁴ Cong T. Trinh^{1,2}

AUTHOR AFFILIATIONS See affiliation list on p. 21.

ABSTRACT *Bacillus coagulans*, a Gram-positive thermophilic bacterium, is recognized for its probiotic properties and recent development as a microbial cell factory. Despite its importance for biotechnological applications, the current understanding of *B. coagulans*' robustness is limited, especially for undomesticated strains. To fill this knowledge gap, we characterized the metabolic capability and performed functional genomics and systems analysis of a novel, robust strain, *B. coagulans* B-768. Genome sequencing revealed that B-768 has the largest *B. coagulans* genome known to date (3.94 Mbp), about 0.63 Mbp larger than the average genome of sequenced *B. coagulans* strains, with expanded carbohydrate metabolism and mobilome. Functional genomics identified a well-equipped genetic portfolio for utilizing a wide range of C5 (xylose, arabinose), C6 (glucose, mannose, galactose), and C12 (cellobiose) sugars present in biomass hydrolysates, which was validated experimentally. For growth on individual xylose and glucose, the dominant sugars in biomass hydrolysates, B-768 exhibited distinct phenotypes and proteome profiles. Faster growth and glucose uptake rates resulted in lactate overflow metabolism, which makes *B. coagulans* a lactate overproducer; however, slower growth and xylose uptake diminished overflow metabolism due to the high energy demand for sugar assimilation. Carbohydrate Transport and Metabolism (COG-G), Translation (COG-J), and Energy Conversion and Production (COG-C) made up 60%–65% of the measured proteomes but were allocated differently when growing on xylose and glucose. The trade-off in proteome reallocation, with high investment in COG-C over COG-G, explains the xylose growth phenotype with significant upregulation of xylose metabolism, pyruvate metabolism, and tricarboxylic acid (TCA) cycle. Strain B-768 tolerates and effectively utilizes inhibitory biomass hydrolysates containing mixed sugars and exhibits hierarchical sugar utilization with glucose as the preferential substrate.

IMPORTANCE The robustness of *B. coagulans* makes it a valuable microorganism for biotechnology applications; yet, this phenotype is not well understood at the cellular level. Through phenotypic characterization and systems analysis, this study elucidates the functional genomics and robustness of a novel, undomesticated strain, *B. coagulans* B-768, capable of utilizing inhibitory switchgrass biomass hydrolysates. The genome of B-768, enriched with carbohydrate metabolism genes, demonstrates high regulatory capacity. The coordination of proteome reallocation in Carbohydrate Transport and Metabolism (COG-G), Translation (COG-J), and Energy Conversion and Production (COG-C) is critical for effective cell growth, sugar utilization, and lactate production via overflow metabolism. Overall, B-768 is a novel, robust, and promising *B. coagulans* strain that can be harnessed as a microbial biomanufacturing platform to produce chemicals and fuels from biomass hydrolysates.

Editor Danielle Tullman-Ercek, Northwestern University, Evanston, Illinois, USA

Address correspondence to Cong T. Trinh, ctrinh@utk.edu.

The authors declare no conflict of interest.

See the funding table on p. 22.

Received 13 July 2024

Accepted 6 September 2024

Published 8 October 2024

Copyright © 2024 Dooley et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

KEYWORDS *Bacillus coagulans*, B-768, lactate overproducer, overflow metabolism, carbohydrate metabolism, robustness, biomass hydrolysates, switchgrass, thermophile, proteome reallocation

Lignocellulosic biomass derived from bioenergy crops (e.g., poplar, switchgrass) or agricultural residues (e.g., corn stover) are renewable and sustainable feedstocks (1, 2). These feedstocks can be produced at a scale of 1 billion tons per year in the United States, reducing reliance on petroleum to make fuels, chemicals, and materials (3). Lignocellulosic biomass is comprised of cellulose, hemicellulose, and lignin (4). These polymers can be isolated to make consumables (5, 6), advanced materials (7), and other valuable commodities. Alternatively, these polymers can be further broken down to monomers through chemical pretreatment followed by enzymatic saccharification to generate biomass hydrolysates that consist of pentose (C5) and hexose (C6) sugars and byproducts (e.g., organic acids, furans, and phenolics). Through fermentation, microbial cell factories can be harnessed as biocatalysts to convert these sugars into various classes of fuels, chemicals, and materials (2, 8, 9) including carboxylic acids (10), alcohols (11–13), aldehydes (14), ketones (15), amines (16), esters (17, 18), polyhydroxyalkanoates (19), and protein-based polymers (20). Biomanufacturing is thus important for a circular bioeconomy where sustainable feedstocks can be effectively utilized to make renewable goods (21).

One technical barrier to biomanufacturing is creating robust microbial catalysts capable of transforming biomass hydrolysates into fuels, chemicals, and materials. Robustness is defined as the efficacy by which microbes can convert biomass hydrolysates into a target chemical under harsh conditions that cause biocatalyst poisoning (22, 23). Robustness is a complex phenotype that requires controlled expression of multiple genes. For instance, the effective assimilation of C5 and C6 sugars in biomass hydrolysates is a complex phenotype, requiring a highly regulated gene network (24). Coping with chemical inhibitors such as pretreatment byproducts in biomass hydrolysates and target products generated during fermentation forces cells to optimize their cellular resources at the systems level (25). Therefore, deploying robust microbial catalysts, whether natural isolates or engineered microorganisms, for efficient conversion of lignocellulosic biomass to a target product, is critical for biomanufacturing.

The robustness and unique metabolic capability of *Bacillus coagulans* make it a promising biomanufacturing platform. *B. coagulans* (also known as *Weizmannia/Heyndrickxia coagulans*) is a GRAS (generally recognized as safe) Gram-positive bacterium and facultative thermophilic anaerobe. *B. coagulans* grows on a variety of carbohydrates (26–28) at a wide range of temperatures (29) and pHs (30) and is capable of overproducing lactate (31), thermostable enzymes (31), and antimicrobial peptides (31–33). Due to its unique metabolic capability and inherent robustness, *B. coagulans* is also utilized for probiotic formulations (34). Even though it is a promising biomanufacturing platform, the fundamental understanding of its robustness and underlying genetics is limited.

In this study, we characterized the robustness physiology of *B. coagulans* B-768, a novel, undomesticated strain. We first assessed its unique metabolic capability by growing it on various C5 (arabinose, xylose), C6 (glucose, galactose, mannose), and C12 (cellobiose) sugars at a wide range of temperatures and evaluated its tolerance to high lactate concentrations. Genome sequencing and annotation were performed to provide insights into its observed metabolic capability. Subsequent characterization of B-768 growing on xylose, glucose, a mixture of xylose and glucose, and biomass hydrolysate, coupled with quantitative proteomics, validated the functions of genes involved in sugar utilization and elucidated how cells dynamically reallocate their proteomes in response to different carbon sources. Finally, we examined the regulation of sugar utilization in B-768 and uncovered the cellular processes and molecular drivers responsible for B-768's robust growth on undetoxified hydrolysates.

RESULTS AND DISCUSSION

B. coagulans B-768 is a robust biomanufacturing platform microorganism

Utilization of a wide range of biomass-derived sugars is crucial for a robust biomanufacturing platform microorganism. B-768 was grown on glucose, xylose, or cellobiose—the dominant sugars present in biomass hydrolysates—at a range of temperatures (30°C–59°C, Fig. 1A). Optimum cell growth was observed between 45°C and 53°C with a maximum growth rate of 0.4 h⁻¹ on glucose and 0.2 h⁻¹ on xylose or cellobiose. Next, B-768 was grown in media containing galactose, mannose, or arabinose at 50°C in addition to the previous sugars (Fig. 1B). Glucose and mannose were the most efficiently used carbon sources, followed by galactose, xylose, arabinose, and cellobiose, which all yielded similar maximum specific growth rates and maximum cell biomasses (Fig. 1B; Fig. S1A). In addition to the assimilation efficiencies of various carbon sources, tolerance to high product concentrations is critical to determine cellular robustness. Since *B. coagulans* is well known for its high lactate production, we evaluated the growth of

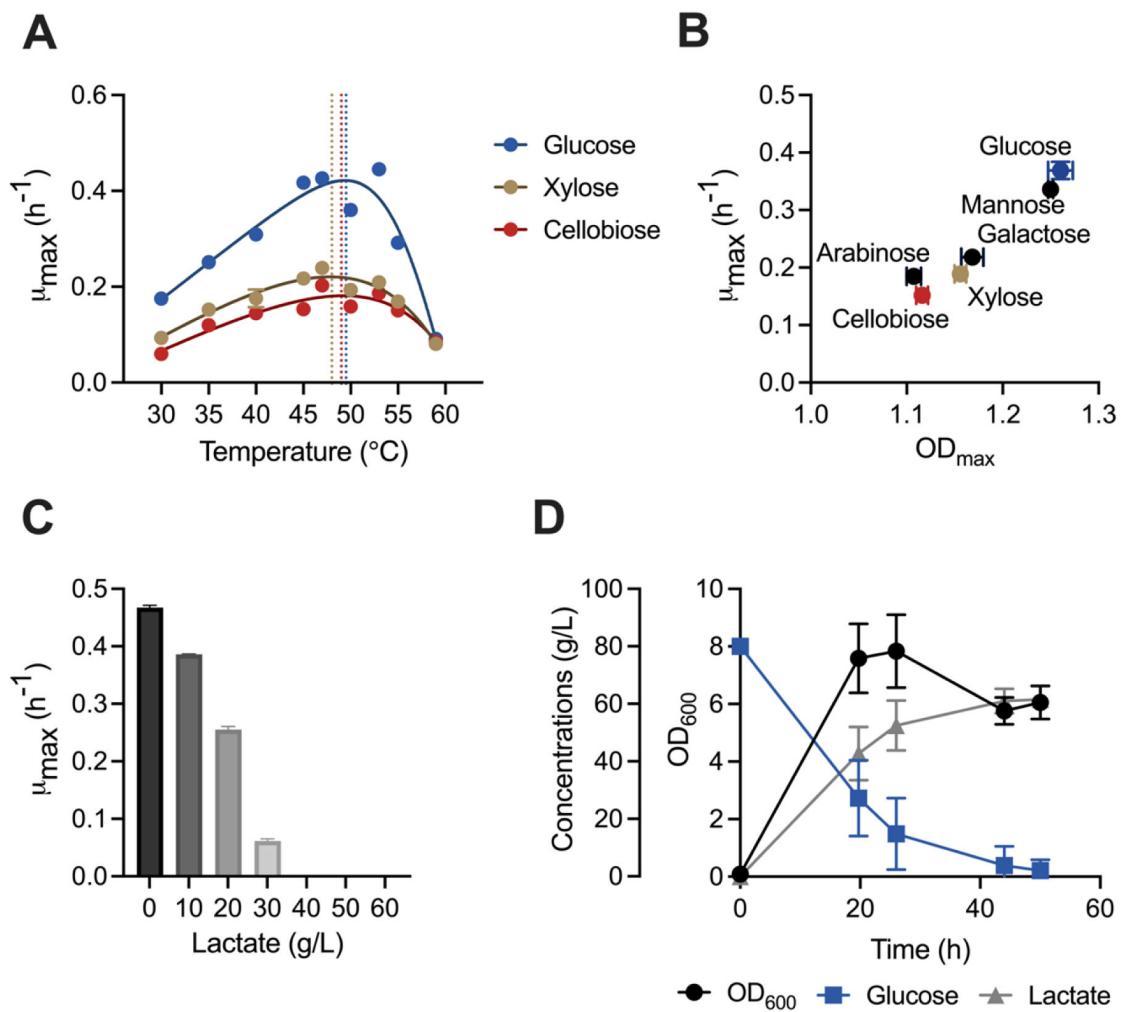


FIG 1 Physiological characterization of *B. coagulans* B-768 reveals cellular robustness and efficient utilization of various carbon sources at various temperatures. (A) Dependence of specific growth rates of B-768 on temperatures when grown on glucose, xylose, or cellobiose. The dotted lines indicate the temperature at which growth was optimal. (B) Carrying capacity and maximum specific growth rate of B-768 growing on various C5, C6, and C12 fermentable sugars at 50°C. (C) Effect of lactate concentrations on maximum specific growth rates of B-768 growing on glucose at 50°C. (D) Kinetic profiles of cell growth, glucose consumption, and lactate production in a pH-controlled bioreactor (pH = 6.5) at 50°C. Data points represent 1 mean \pm 1 standard deviation of three biological replicates.

TABLE 1 Summary of *B. coagulans* B-768 assembly and comparison with reference genomes

Feature	B-768	36D1	DSM1
Length (bp)	3,944,476	3,552,226	3,366,995
G+C content (%)	45.7	46.0	46.5
CDS	3,863	3,227	3,215
ncRNA	4	5	7
rRNA	30	30	30
tRNA	85	84	84
CRISPR arrays	5	4	3

B-768 in cultures containing various lactate concentrations with glucose as a sole carbon source (Fig. 1C; Fig. S1B). The maximum specific growth rate was reduced as lactate concentration was increased, and B-768 could tolerate up to at least 30 g/L lactate. Since this experiment was performed in a pH-uncontrolled environment, it was hypothesized that higher lactate tolerance could be achieved when the medium pH was maintained at an optimum level. To evaluate this hypothesis, a pH-controlled fermentation was performed with 80 g/L glucose as the sole carbon source. As expected, when the pH was maintained at 6.5, B-768 produced up to 60 g/L lactate at 92.7% of the theoretical maximum yield (Fig. 1D). These results highlight the robustness of B-768 for metabolizing biomass-derived sugars, growing over a broad range of temperatures, and selectively producing lactic acid at high yields.

Functional and comparative genomics of *B. coagulans* B-768

*Genome sequence of B-768 and its relatedness to other *B. coagulans* strains*

The genome of B-768 was sequenced to elicit the genetic basis of its physiological robustness. The genome assembly of B-768 is 3,944,476 base pairs long with 45.7% GC content (Fig. 2). Strain B-768 contains 4,059 genomic features, of which 3,863 (95.2%) are coding sequences (CDS). It should be noted that B-768 shares considerable average nucleotide identity (ANI) with *Heyndrickxia faecalis* and *Weizmannia agrestimuris* according to the Type Strain Genome Server (35). However, phylogenetic analysis of the core genomes of NCBI RefSeq assemblies for *B. coagulans*, *H. faecalis*, and *W. agrestimuris* showed that B-768's closest two neighbors were *B. coagulans* CSIL1 and 36D1, which was further confirmed by ANI analysis (36). Therefore, comparative and functional genomics of B-768 were performed against *B. coagulans*.

Phylogenomic analysis of a 55-strain *B. coagulans* pan-genome was performed, and the relationship of B-768 to model strains 36D1 and DSM1 was evaluated. B-768 belongs to the same clade as 36D1 (clade A) but is highly diverged from DSM1 (clade B, Fig. 3A). B-768's genome is 11% larger than 36D1 and 17% larger than DSM1 (Table 1). Concordantly, there are ~20% more CDS in B-768 compared to 36D1 or DSM1. On average, B-768's genome is 630 Kbp larger than other *B. coagulans* genomes and is the largest reported genome of this species to date. Strains B-768 and 36D1, but not DSM1, can metabolize C5 sugars (e.g., xylose and arabinose) (31, 37–39). Other strains reported to metabolize xylose, including NL01 (27, 40) and DSM 2314 (41), also belong to clade A. The ability to efficiently use C5 sugars could aid strains in occupying an ecological niche and incentivize the acquisition of relevant genomic material through vertical and horizontal gene transfers. This hypothesis is supported by the fact that genomes from clade A are on average ~171 Kbp larger than those for clade B. Though further characterization of xylose and arabinose usage in *B. coagulans* strains is needed to confirm this hypothesis, C5 metabolism could partially explain the genetic differences underlying the topology of the *B. coagulans* pan-genome.

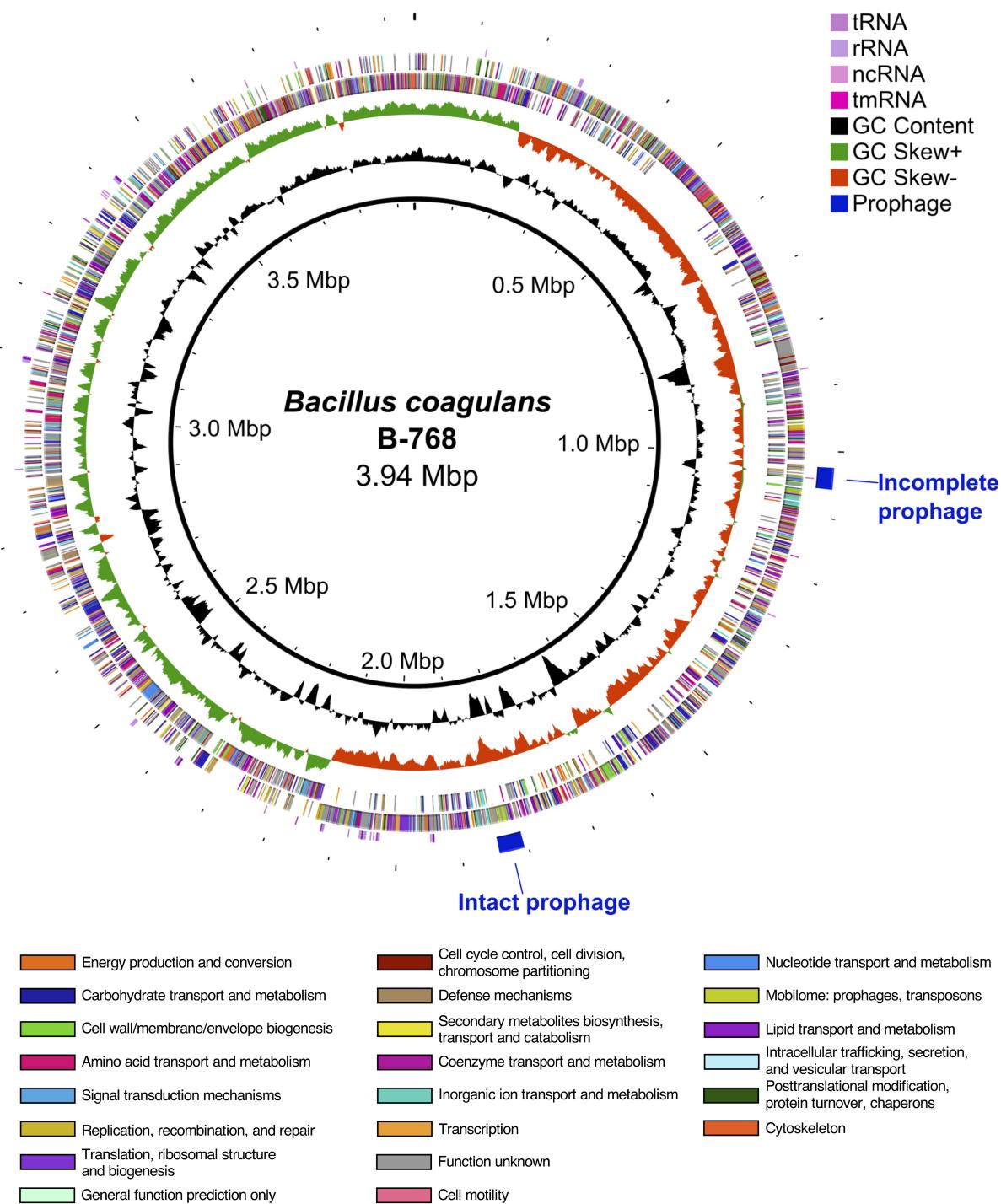


FIG 2 Circular genome map of *Bacillus coagulans* B-768. Tracks from outside to in are as follows: PHASTEST 3.0 prophage regions, non-coding RNAs, forward and reverse strand CDS colored by COG category, GC Skew = $(G - C)/(G + C)$, and GC content. Abbreviations: tRNA, transfer RNA; rRNA, ribosomal RNA; ncRNA, non-coding RNA; tmRNA, transfer-messenger RNA.

The larger genome of B-768 provides potential gains of function for enhanced cellular robustness

To better understand the biological functions that may be associated with the added genomic material of B-768, we conducted a comparative COG (Clusters of Orthologous Genes) analysis by calculating the percent change of CDS belonging to each COG

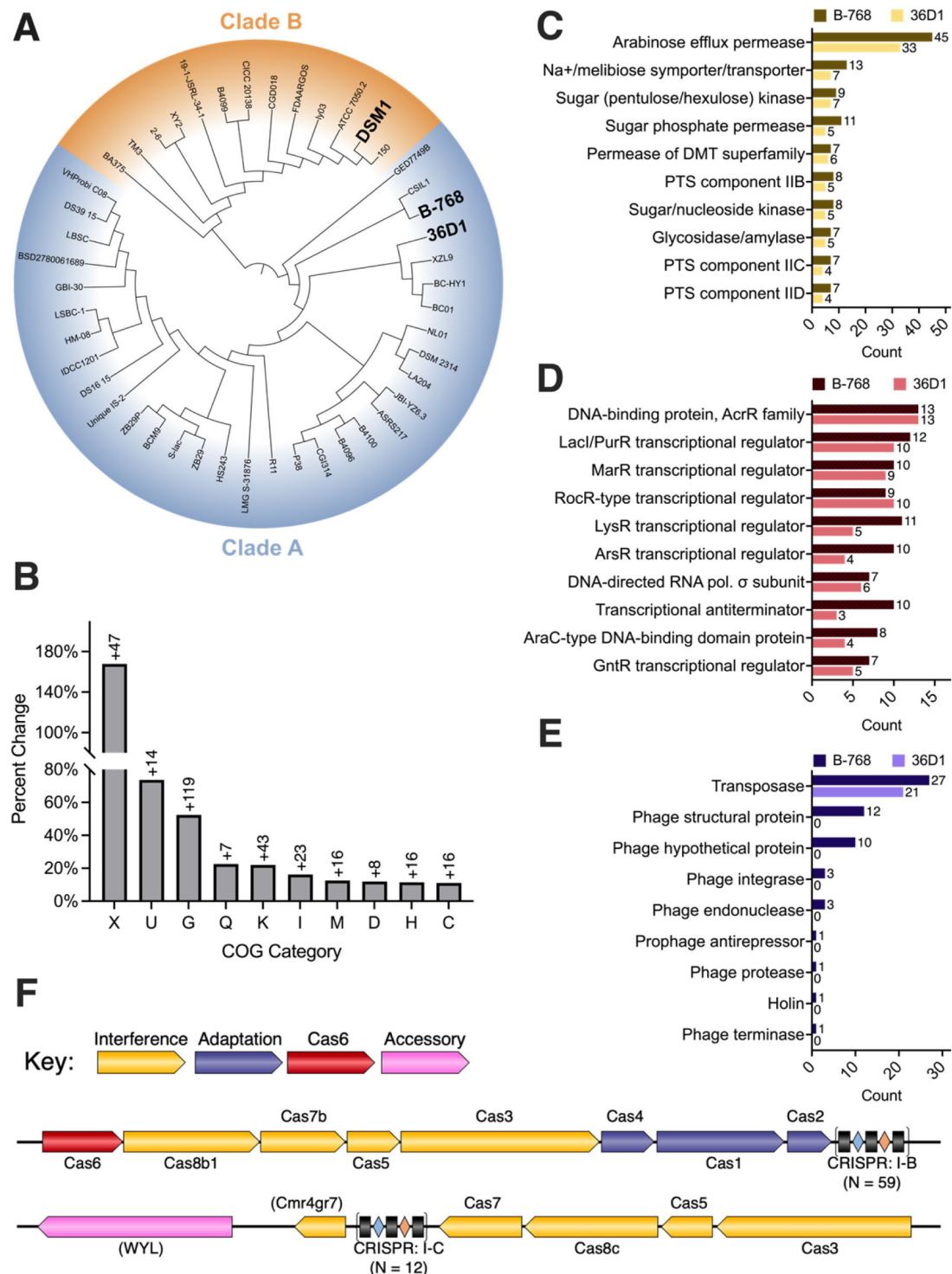


FIG 3 Comparative and functional genomics of *B. coagulans* B-768. (A) Core genome alignment of *B. coagulans* pan-genome ($n = 55$). (B) Percent change of COG categories between strains B-768 and 36D1. Numbers above the bars indicate the excess number of proteins in B-768 relative to 36D1. Full COG Category names are listed at the end of the figure caption. (C–E) Breakdown of most abundant proteins in (C) Carbohydrate Metabolism and Transport, (D) Transcription, and (E) Mobilome COG categories. (F) Predicted CRISPR-Cas operons of B-768. COG Categories: U = Intracellular Trafficking, Secretion, and Vesicular Transport; X = Mobilome; G = Carbohydrate Transport and Metabolism; Q = Secondary Metabolite Biosynthesis, Transport and Catabolism; K = Transcription; I = Lipid Transport and Metabolism; M = Cell Wall/Membrane/Envelope Biogenesis; D = Cell Cycle Control, Cell Division, Chromosome Partitioning; H = Coenzyme Transport and Metabolism; C = Energy Production and Conversion.

category between B-768 and its closest reference, 36D1 (Fig. 3B through E). The percent change was calculated as follows: $100 * \frac{N_{B-768} - N_{36D1}}{N_{36D1}}$, where N is the number of proteins belonging to the specified COG. Several putative phage proteins found using PHASTEST (42) not recognized by automated COG annotation (43) were manually added to the Mobilome COG. The percent change of the top 10 COGs is shown in Fig. 3B. B-768 contains more genes than 36D1 for nearly every COG category with several categories being heavily expanded, such as the Mobilome (COG-X), Intracellular Trafficking, Secretion, and Vesicular Transport (COG-U), Carbohydrate Transport and Metabolism (COG-G), and Transcription (COG-K, Fig. 3B). Detailed examination of the individual proteins involved in these COGs can shed light into the robustness of B-768, specifically COG-G and COG-K due to their relevance to carbon utilization (Fig. 3C and D) and COG-X due to its role in horizontal gene transfer and genome composition (Fig. 3E).

Unique carbohydrate metabolism, transport, and regulation of B-768

Carbohydrate Metabolism and Transport is the largest functional category in B-768, underlying the observed robust carbon metabolism. B-768 contains 52.4% more proteins in COG-G than 36D1, and more proteins were identified for all protein classes (Fig. 3C). Top protein classes include arabinose efflux permeases, Na⁺/melibiose symporters/transporters, sugar kinases, and sugar permeases. The overrepresentation of transporters among this group suggests B-768's vigorous carbon metabolism is partly due to improved uptake of carbon substrates. Transcriptional regulators are known to play key roles in regulating the use of carbon sources. A breakdown of the COG-K reveals B-768 has at least 16 more transcriptional regulators than 36D1, belonging to the LacI/PurR, LysR, ArsR, and GntR families (Fig. 3D). LacI and PurR regulators are mostly allosteric repressors of carbon and nucleotide synthesis, respectively, and may control one (e.g., LacI) or many (e.g., PurR) operons (44). Given the ability of B-768 to utilize various carbon sources (Fig. 1B), an enlarged complement of LacI/PurR regulators likely helps coordinate the use of available sugars. LysR and GntR regulators act globally as either activators or repressors of various cellular processes including, but not limited to, metabolism, virulence, motility, and intercellular communication (45, 46). ArsR regulators repress operons linked to metal stress under normal physiological conditions. Upon binding with multivalent metal or metalloid ions, the associated operons are derepressed and can provide resistance to otherwise toxic concentrations of metals/metalloids (47). An excess of six ArsR regulators in B-768 may enable it to survive in metal-rich environments where other *B. coagulans* strains cannot thrive.

Mobilome of B-768 reveals its capability to expand its genome

B-768 contains 168% more proteins in its Mobilome—which includes prophages, transposons, and plasmid mobilization and maintenance machinery—compared to 36D1 (Fig. 3E). Six more transposases, which are capable of rearranging, adding, or removing genomic material, were identified from B-768's chromosome. The most significant portion of B-768's expanded Mobilome, however, is attributed to two putative prophage regions that encode 32 CDS. By contrast, 36D1 has no prophages or phage-related proteins. Prophage-like entities are commonly found in probiotic strains and can provide resistance to similar phages (48, 49). The first prophage region is 32.7 kb long (genomic coordinates: 1,024,169–1,056,950) and is predicted to be incomplete due to the absence of various necessary phage genes such as portal, capsid, terminase, and tail genes. It does, however, encode for two integrases, one protease, one endonuclease, one antirepressor, and several other unannotated phage-like proteins which may contribute to bacterial fitness and/or horizontal gene transfer. The second prophage is 38.3 kb long (genomic coordinates: 1,805,785–1,844,080) and is putatively intact, possessing all major proteins required for phage infection. Many of the closest homologs for B-768's putative phage proteins are from *Bacillus* phages, but several come from *Corynebacterium*, *Agrobacterium*, and *Enterococcus* phages as well, meaning the exact origin and

classification of the prophage regions remain unclear. The existence of prophages in *B. coagulans* strains has been previously reported, but in all cases, the phages were deemed non-functional (50, 51). Given the well-established role of bacteriophages in horizontal gene transfer, the presence of two putative prophages could help explain the origin of B-768's expanded genetic repertoire. An operational phage in *B. coagulans* could potentially be useful for introducing DNA via transduction, enabling various genetic engineering approaches. Together, these findings provide a glimpse into the expanded genome of B-768 that facilitates its extraordinary cellular robustness.

Defense systems in B-768: restriction modification and CRISPR-Cas

Bacteriophage contamination is a common problem in bacterial fermentation and can lower product yields and either slow or halt fermentations (52). Bacteria have developed numerous defense systems for rebuffing bacteriophage infection, including innate restriction modification (R-M) and adaptive CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR associated). To elucidate the robustness of B-768, we evaluated its innate and adaptive defense systems to assess its potential resistance to phage infection and suitability for industrial application.

Restriction-modification systems recognize and cut foreign DNAs having different methylation patterns than the host. R-M systems are categorized into four types, Types I–IV, based on genetic structure and biochemical function. Type I systems are multi-subunit complexes composed of the R (restriction endonuclease), M (methyltransferase), and S (DNA-specificity) subunits expressed from a single operon. Type II and III systems are comprised of restriction endonuclease and methyltransferase subunits, but while Type II R and M proteins function independently, Type III restriction enzyme activity requires an R-M complex. Type IV systems are effectors that may be encoded by either one or two genes (53). Type I, II, and III restriction endonucleases recognize and cleave unmethylated DNA at a specific DNA motif, while Type IV restriction endonucleases cleave DNA that is methylated differently from the host. Our genomic analysis shows that B-768 has eleven predicted R-M proteins including seven Type I, two Type II, and two Type IV proteins (Table 2). No Type III R-M proteins were found. In B-768, a complete Type I R-M system is encoded within a single operon spanning from 2,754,339 to 2,761,240. We found that B-768 also has four additional Type I R submodules co-localized from 3,201,231 to 3,202,333, but it is unclear whether these isolated proteins participate in Type I R-M activity in B-768. Both Type II R-M proteins in B-768 are contained in an operon from 2,170,437 to 2,176,691. B-768's Type IV genes are co-localized, meaning they likely encode for a two-subunit effector. In total, B-768 has three distinct and intact R-M systems that enable restriction of foreign DNAs.

CRISPR-Cas systems are functionally distinct immune systems found in archaea, bacteria, and even huge bacteriophages (54–56). These systems work in a two-step process: adaptation and interference. In adaptation, stretches of foreign DNA (protospacers) are integrated into CRISPR arrays within the host genome and stored for future use. These short fragments are called spacers. During interference, spacers are transcribed, processed, and form a complex with Cas effector proteins to guide them to a cognate protospacer, where the target DNA is cleaved. Our genomic analysis found that B-768 possesses five putative CRISPR arrays containing 145 spacers. Three CRISPR arrays have no associated Cas proteins near them and two have accompanying Type I-B and I-C CRISPR machinery (Fig. 3F). The Type I-B CRISPR system is predicted to have complete

TABLE 2 Number of genes in restriction-modification (RM) systems

R-M	B-768	36D1 (31)	DSM1 (31)
Type I	7	19	6
Type II	2	0	1
Type III	0	2	2
Type IV	2	3	3
Total	11	24	12

TABLE 3 Identification of spacers in B-768 CRISPR arrays

Spacer	Hit ID	Genus	Type	Method
GGTTGGGTAGCCGGTTGATATTCTGATTAATCAA	JQ362498.1	<i>Sphingomonas</i>	Phage	SpacePHARER
GGTTGGGTAGCCGGTTGATATTCTGATTAATCAA	KC481682.1	<i>Bacillus</i>	Phage	SpacePHARER
GCTTTAAACTCCATCAGGGTTTATTTGGGT	KX965989.1	<i>Aeribacillus</i>	Phage	SpacePHARER
TATAGAAACACGTTTCGGCGACAAGGCCTGGCATG	KU665491.1	<i>Bacillus</i>	Phage	SpacePHARER
ATCCCGGCGGTGCTGGCTATGTCGAAATATGC	NZ_CP022365.1	<i>Azospirillum</i>	Plasmid	Blastn_short
CTTTAAAAAAATCACAAATGAAAAATTACGTGCGG AATATGCGCAAATCATATTCAACTCCAATGTCAT	NZ_CP075898.1	<i>Exiguobacterium</i>	Plasmid	Blastn_short
AAAACTAAGTAAAAAGAATTAGCCGTATGCC	NZ_CP015595.1	<i>Acinetobacter</i>	Plasmid	Blastn_short
TGAAACCGCGGGAAAGCCAATACGGCGCGAACCA	NZ_CP032924.1	<i>Agrobacterium</i>	Plasmid	Blastn_short
GCGATATACTGCCGTAGGGCATGTCGCCGTCAATAAA	NZ_CP024924.1	<i>Sphingomonas</i>	Plasmid	Blastn_short

interference and adaptation arsenals, while the Type I-C system appears to only possess interference-related genes. Analysis of B-768's entire set of spacers reveals CRISPR RNAs targeting both phages and plasmids, which indicates B-768 has a history of deploying CRISPR-Cas against foreign invaders of various origins (Table 3). In addition to robust and programmable defense against deleterious phages, B-768's endogenous CRISPR systems could hold promise for editing *B. coagulans* and other industrially relevant thermophiles.

Lignocellulosic biomass-derived sugar metabolism in B-768

The C6 sugars (e.g., glucose) are mainly transported inside the cells using the phosphoenolpyruvate (PEP):carbohydrate phosphotransferase system (PTS) (Data Set S1). The phosphorylated sugars (e.g., glucose-6-phosphate) are assimilated via the Embden-Meyerhof-Parnas (EMP) and/or Entner-Doudoroff (ED) pathways. We identified the full set of necessary genes for the EMP pathway in B-768 (Fig. 4; Data Set S2). On the other hand, B-768 contains an incomplete ED pathway due to the absence of 6-phosphogluconate dehydratase (EDD), which converts 6-P-gluconate into 2-keto-3-deoxy-6-phosphogluconate (Fig. 4; Data Set S2). B-768 does, however, possess 6-phosphogluconate dehydrogenase (6PGD) that converts 6-phosphogluconate into ribulose-5-phosphate, one of the end products in the pentose phosphate pathway. In summary, glycolysis is the primary pathway for C6 sugar degradation in B-768, with possible metabolic flux through a hybrid ED and pentose phosphate pathway route.

Since B-768 grows on xylose (Fig. 1A and B), it is expected that B-768 contains xylose assimilation genes. Transport is the first step of carbon utilization and B-768 contains two putative xylose transporters: xylose H+-symporter (*XYL*, encoded by *xylT*) and an ABC-type xylose transporter (*xylFGH*). These were identified using 36D1 xylose transporters as reference genes (Data Set S1). Xylose transported inside of the cell via xylose transporters is metabolized by two key enzymes: xylose isomerase (*XI*, encoded by *xylA*), which converts D-xylose into D-xylulose, and xylulokinase (*XK*, encoded by *xylB*), which converts D-xylulose into D-xylulose-5-phosphate. Typically, *xylA*, *xylB*, and *xylT* are encoded alongside *xylR*—a transcriptional repressor of the *xyl* operon (57, 58)—in a single operon, but in B-768 *xylT* is located separately over 1 Mbp away (Fig. 4; Data Set S2). Once xylose is converted into D-xylulose-5-phosphate, it can be further assimilated by either the pentose phosphate pathway (PPP) or the phosphoketolase pathway (PKP). The PKP results in equimolar production of lactate and acetate in lactic acid bacteria (59), while the PPP yields homolactic acid, making the PPP desirable for lactate production from pentoses (38). B-768 contains a full complement of PPP genes but lacks the critical PKP enzyme, phosphoketolase (Data Set S2). Given the lack of phosphoketolase, B-768 mainly utilizes the homolactic acid PPP to degrade xylose, setting it apart from most other lactic acid bacteria that prefer the PKP (60). This is an essential trait for ensuring high lactic acid yields on cellulosic sugar mixtures.

B-768 also contains genes that putatively allow it to use xylosides. Xylosides are released when xylan or xyloglucan is depolymerized and may be converted into D-xylose

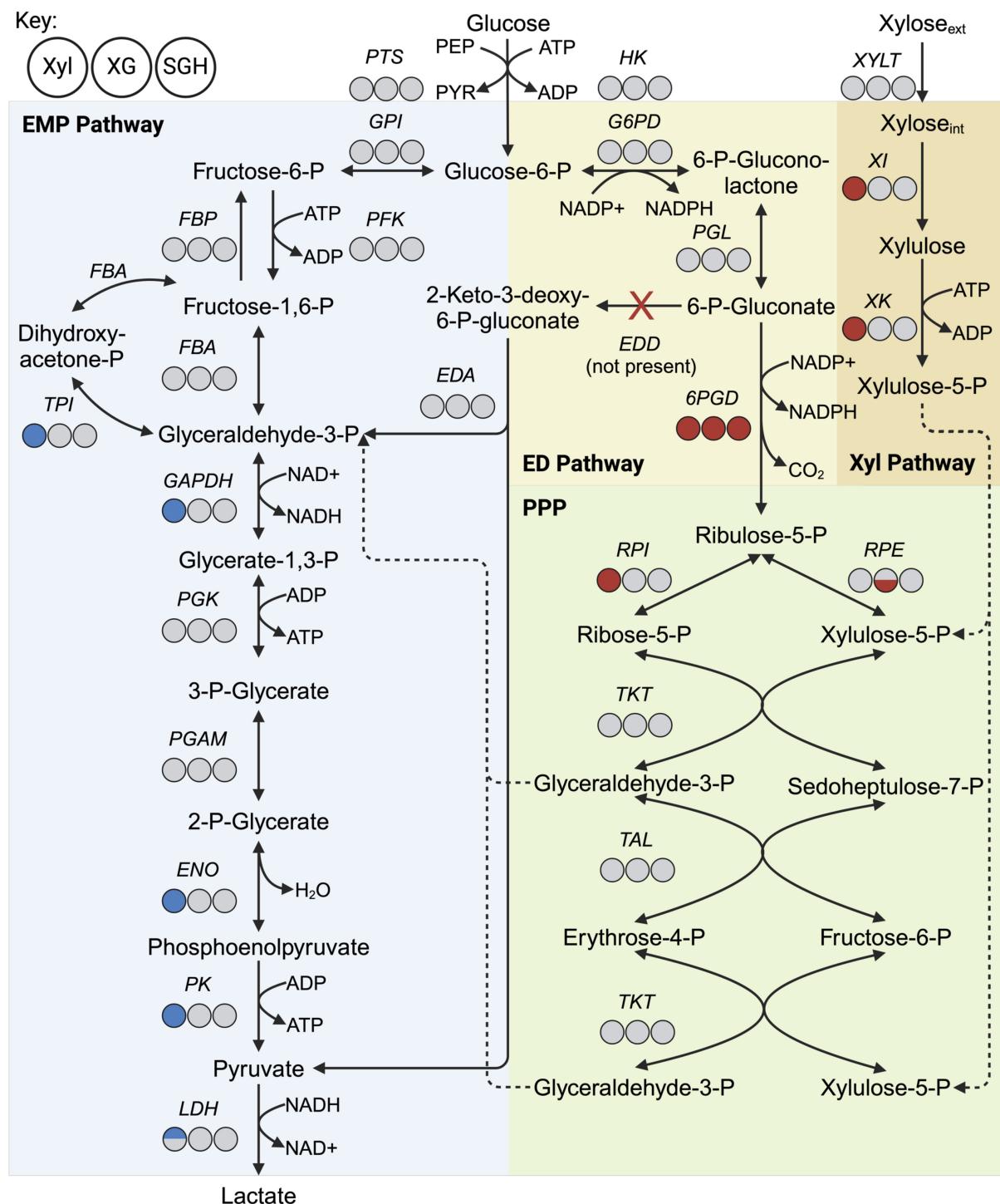


FIG 4 Central carbon metabolism of B-768. Circles from left to right represent the differential expression of a given protein in (i) xylose vs glucose, (ii) xylose and glucose mixture vs glucose, and (iii) switchgrass hydrolysate vs glucose conditions. Red = upregulated with \log_2 -fold change ≥ 1 and $q < 0.05$; blue = downregulated with \log_2 -fold change ≤ -1 and $q < 0.05$; Gray = not significant. Paralogs of the same enzyme were plotted as half circles with the most abundant paralog on top.

by α - or β -xylosidases (61). Putative xyloside symporter (*xynT*) and α -xylosidase (*xyIS*) genes were found in the *xyI* operon in B-768, and three β -xylosidases were scattered throughout the rest of the genome. While closely related strain 36D1 has a copy of *xynT*, it does not possess any known xylosidases (31). To determine the origin of the

TABLE 4 BLASTp top hits for B-768 xylosidase proteins

Query	Subject	Description	Organism	Query cover	E-value	Percent identity
Alpha-xylosidase	WP_063613105.1	Alpha-xylosidase	<i>Bacillus sp.</i>	1	0	98.9%
Beta-xylosidase/AraC-like DNA-binding protein	WP_161788365.1	Helix-turn-helix domain-containing protein	<i>Bacillus coagulans</i>	1	0	98.1%
GH43 family beta-xylosidase	WP_071451997.1	Family 43 glycosyl hydrolase	<i>Bacillus coagulans</i>	1	0	95.7%
4-beta-xylosidase	WP_014868562.1	Glycoside hydrolase family 43 protein	<i>Melissococcus plutonius</i>	0.98	0	64.9%

four xylosidase proteins, BLASTp analysis was performed against the non-redundant protein database (Table 4). This search revealed homologs in closely related species for all but one of the xylosidases. Enzyme 4-beta-xylosidase located from 665440 to 666999 had top hits primarily belonging to *Enterococcus* spp., with the top hit occurring in *Melissococcus plutonius*. This provides another example of *B. coagulans* B-768 utilizing horizontal gene transfer to expand its carbon-utilizing potential. This is a useful trait because xylan oligomers are competitive inhibitors of cellulases.

Cellobiose assimilation is key to bypassing carbon catabolite repression (CCR), wherein xylose assimilation is inhibited by glucose. Cellobiose utilization in *B. coagulans* is mainly mediated by the five-member CELO1 operon encoding three phosphotransferase system (PTS) subunits EI_A (*celA*), EI_B (*celB*), EI_C (*celC*), a 6-phospho-β-glucosidase (*pbgI*), and a transcriptional regulator (*celR*) (27). Some strains like *B. coagulans* NL01, 36D1, 2–6, and HM08 have an additional cellobiose operon, CELO2, which encodes paralogs of *celABC* along with an unknown protein (*celX*) (27). Using 36D1 as a reference strain, we identified cellobiose assimilation pathway genes in B-768. In B-768, both CELO1 and CELO2 are fully intact and located from 1361260 to 1367923 and 1464816 to 1469898, respectively. Notably, B-768 contains two additional copies of 6-phospho-β-glucosidase—the enzyme catalyzing the rate-limiting step in the breakdown of cellobiose to glucose (62, 63)—which may improve cellobiose catabolism.

Due to B-768's metabolic capability to assimilate galactose, mannose, and arabinose (Fig. 1B), we also investigated the catabolic pathways of these monomeric sugars. Galactose metabolism in bacteria is mediated by the ubiquitous four-enzyme Leloir pathway, containing the enzymes galactose mutarotase, galactokinase, galactose-1-phosphate uridylyltransferase, and UDP-galactose 4-epimerase (64). B-768 possesses all four genes in an operon from 1161267 to 1165598 with an additional copy of galactose mutarotase (Data Set S1). Mannose metabolism in *Bacillus subtilis* is mediated by a mannose transporter (*manP*), a mannose-6-phosphate isomerase (*manA*), and a mannose activator (*manR*) (65). In *B. coagulans*, mannose metabolism is much the same, except the mannose transporter is mediated by the mannose PTS system encoded by four subunit genes (*manEIIABCD*) instead of a single gene (Data Set S1). B-768 possesses homologs for these genes, but *manA* is absent from the operon containing *manR* and *manEIIABCD* (Data Set S1). All canonical arabinose utilization proteins are present in B-768, many of which possess several putative paralogs (Data Set S1). Most of these paralogs are either for arabinose/arabinan permeases or transcriptional regulators of the *ara* regulon. This is consistent with the finding that most of B-768's genes involved in carbon metabolism putatively encode for C5/C6 sugar kinases, permeases, and transcriptional regulators (Fig. 3C). Overall, B-768 has accrued an extensive array of genes dedicated to metabolizing a wide variety of sugars.

Distinct phenotypes and proteomes of B-768 for utilization of biomass-derived sugars

B-768 grew faster on glucose than xylose, resulting in lactate overflow metabolism

Since xylose and glucose are the most dominant C5 and C6 sugars in biomass hydrolysates, we first aimed to elucidate the phenotypes and metabolic responses of B-768

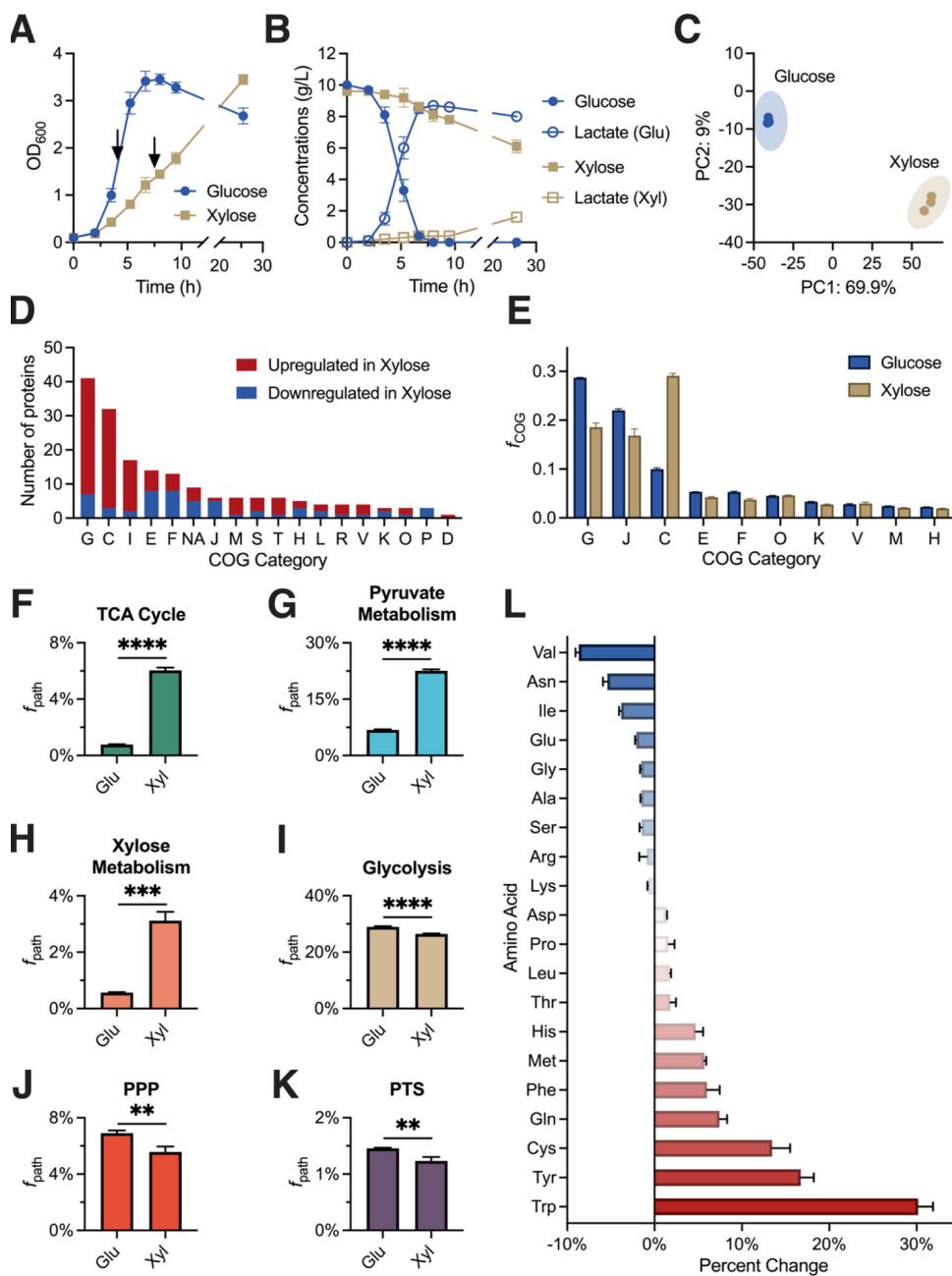


FIG 5 *B. coagulans* B-768 exhibited distinct proteomes for growing on glucose and xylose. (A–B) Kinetic profiles of (A) cell growth and (B) glucose/xylose consumption and lactate production. Proteomics samples were collected from parallel cultures at the time points indicated by black arrows. (C) Principal component analysis of *B. coagulans* B-768 proteomes for growth on glucose and xylose. (D) Number of differentially expressed proteins between glucose and xylose conditions for most perturbed COG categories. (E) Mass fraction of proteomes in top COG categories (f_{COG}) for glucose and xylose conditions. (F–K) Proteome allocation by pathway for glucose and xylose conditions. (L) Percent change in amino acid allocation of proteomes between glucose and xylose conditions. COG Categories: G = Carbohydrate Transport and Metabolism; C = Energy Production and Conversion; I = Lipid Transport and Metabolism; E = Amino Acid Transport and Metabolism; F = Nucleotide Transport and Metabolism; NA = No COG annotation; J = Translation; M = Cell Wall/Membrane/Envelope Biogenesis; S = Function Unknown; T = Signal Transduction Mechanisms; H = Coenzyme Transport and Metabolism; L = Replication, Recombination, and Repair; R = General Function Prediction Only; V = Defense Mechanisms; K = Transcription; O = Post-translation Modification, Protein Turnover, Chaperones; P = Inorganic Ion Transport and Metabolism; D = Cell Cycle Control, Cell Division, Chromosome Partitioning. Statistics for F–K were performed using an unpaired t-test with default settings in GraphPad Prism v10.2.3. Error bars represent the standard deviation of three biological replicates. ****P-value <0.0001; ***P-value <0.001; **P-value <0.01.

growing on these sugars. When supplied with 10 g/L glucose, B-768 grew at a rate of $0.59 \pm 0.01 \text{ h}^{-1}$. Within 8 h, it achieved a maximum growth of $3.45 \pm 0.11 \text{ OD}_{600}$ and exhausted all the glucose with a glucose uptake rate of $1.67 \pm 0.15 \text{ g glucose/OD}_{600}/\text{h}$ (Fig. 5A and B). B-768 produced lactate as the major fermentative product at a titer of $8.72 \pm 0.05 \text{ g lactate/L}$, a productivity of $1.46 \pm 0.03 \text{ g lactate/OD}_{600}/\text{h}$, and a yield of $0.89 \pm 0.01 \text{ g lactate/g glucose}$ ($\sim 90\%$ of the maximum theoretical yield). By contrast, when growing on xylose, B-768 exhibited a much slower growth rate of $0.32 \pm 0.01 \text{ h}^{-1}$. After $\sim 24 \text{ h}$, cells reached the growth of $3.45 \pm 0.10 \text{ OD}_{600}$ and only consumed $3.52 \pm 0.45 \text{ g/L}$ out of $\sim 10 \text{ g/L}$ xylose available with a xylose uptake rate of $0.35 \pm 0.04 \text{ g xylose/OD}_{600}/\text{h}$ (Fig. 5A and B). B-768 only produced lactate at a titer of $1.57 \pm 0.12 \text{ g lactate/L}$, a productivity of $0.07 \pm 0.01 \text{ g lactate/OD}_{600}/\text{h}$, and a yield of $0.18 \pm 0.02 \text{ g lactate/g xylose}$. Overall, faster growth and glucose uptake rate made B-768 reroute high carbon flux toward lactate biosynthesis. Even though B-768 reached the same cell biomass for growth on xylose after 24 h, slower growth and xylose uptake rate shifted its carbon flux mainly toward biomass synthesis at the expense of lactate production.

B-768 exhibited distinct proteomes for growth on glucose and xylose

To elucidate the phenotypic differences for B-768 grown on glucose and xylose at the cellular level, cultures were sampled at the mid-exponential phase for quantitative proteomics. Principal component analysis of glucose and xylose samples revealed highly distinct proteomes (Fig. 5C). Differential expression analysis was then performed on the glucose- and xylose-only samples and the number of significantly altered proteins was tallied for each COG category (Fig. 5D). The results showed that proteomes had shifted across the COG categories to varying degrees. Most significantly perturbed proteins in cells growing on xylose compared to on glucose belong to the Carbohydrate Transport and Metabolism (COG-G) and Energy Production and Conversion (COG-C) categories. As expected, both the major xylose utilization proteins—xylose isomerase and xylulokinase—were among the most xylose-upregulated proteins by $5.99 \log_2$ -fold ($q\text{-value} = 4.17E-14$) and $7.28 \log_2$ -fold ($q\text{-value} = 4.17E-14$), respectively, in the COG-G category (Fig. 4; Data Set S3). Among the PPP proteins, ribose-5-phosphate isomerase (RPI) was significantly upregulated by $1.29 \log_2$ -fold ($q\text{-value} = 0.0478$), while the rest were not significantly perturbed (Fig. 4; Data Set S3). Most of the COG-G proteins downregulated in the xylose condition were involved in the EMP pathway, such as triose phosphate isomerase (TPI), glyceraldehyde-3-phosphate dehydrogenase (GAPDH), enolase (ENO), and pyruvate kinase (PK) by $-1.9 \log_2$ -fold ($q\text{-value} = 8.45E-3$), $-1.29 \log_2$ -fold ($q\text{-value} = 3.53E-3$), $-1.36 \log_2$ -fold ($q\text{-value} = 0.012$), and $-1.48 \log_2$ -fold ($q\text{-value} = 2.93E-4$), respectively (Fig. 4; Data Set S3). Lactate dehydrogenase (LDH) is also significantly downregulated in xylose relative to glucose by $-2.63 \log_2$ -fold ($q\text{-value} = 1.83E-4$) at the measured timepoint (Fig. 4; Data Set S3), which is consistent with the observed low lactate production (Fig. 5B). In COG-C, several enzymes involved in pyruvate metabolism, such as formate C-acetyltransferase (or pyruvate formate lyase, PFL), dihydrolipoamide dehydrogenase (DLD), and pyruvate dehydrogenase subunits E1 α and E1 β (PDH) for converting pyruvate to acetyl-CoA, were significantly upregulated in xylose by $5.71 \log_2$ -fold ($q\text{-value} = 4.17E-14$), $3.71 \log_2$ -fold ($q\text{-value} = 4.17E-14$), $3.44 \log_2$ -fold ($q\text{-value} = 2.52E-6$), and $3.1 \log_2$ -fold ($q\text{-value} = 1.37E-13$), respectively (Data Set S3). Enzymes that convert acetyl-CoA into acetate and ethanol were upregulated in xylose, including acetyl-CoA synthetase (ACS), aldehyde dehydrogenase (ALDH), and alcohol dehydrogenase (ADH), by $3.7 \log_2$ -fold ($q\text{-value} = 1.24E-4$), $4.86 \log_2$ -fold ($q\text{-value} = 4.17E-14$), and $3.6 \log_2$ -fold ($q\text{-value} = 1.03E-7$), respectively (Data Set S3; Fig. S4). Acetyl-CoA is also passed into the TCA cycle, which is highly upregulated on xylose (Data Set S3; Fig. S5).

Tradeoffs of proteome reallocation between carbohydrate metabolism and energy generation contributed to the differential growth on glucose and xylose

We then investigated how B-768 re-allocated its resources in response to glucose or xylose by summing the mass fractions of all expressed proteins for each COG (Fig. 5E).

The top seven COGs, including COG-G, J (Translation), C, E (Amino Acid Transport and Metabolism), F (Nucleotide Transport and Metabolism), O (Post-translation Modification, Protein Turnover, Chaperons), K, V (Defense Mechanisms), M (Cell Wall/Membrane/Envelope Biogenesis), and H (Coenzyme Transport and Metabolism), contributed up to 80% of the total measured proteomes as shown in Fig. 5E. The top three COGs, including G, J, and C, shared up to 60%–65% of each measured proteome with a high degree of variability for growth on glucose and xylose. Despite most significantly differential proteins in COG-G being upregulated in xylose relative to glucose (Fig. 5D), the proteome mass fraction allocated to COG-G was much higher when B-768 was grown on glucose compared to xylose (29% Glu vs 19% Xyl, Fig. 5E). This result indicates that protein expression was higher even with fewer proteins in cells grown in glucose compared to cells grown in xylose. Consistent with this hypothesis, B-768 also devoted significantly more resources to COG-J when grown on glucose compared to xylose (22% Glu vs 17% Xyl). Xylose-grown cells disproportionately invested their resources in COG-C, presumably to offset the higher energy cost of transporting and assimilating xylose (29% Xyl vs 10% Glu, Fig. 5E).

To better elucidate how proteome reallocation is linked to the distinct phenotypes for B-768 growing on glucose and xylose, the allocation of proteomes in COG-G and COG-C was analyzed at a more granular resolution based on the relevant pathways (Fig. 5F through K). Enriched pathways in the cells grown on xylose include the TCA cycle (6.78-fold, *P*-value < 0.0001, Fig. 5F), pyruvate metabolism (2.32-fold, *P*-value < 0.0001, Fig. 5G), and xylose metabolism (4.52-fold, *P*-value = 0.0001, Fig. 5H), reiterating the trends observed from differential expression analyses. Glucose-grown cells had slightly more investments into glycolysis (0.095-fold, *P*-value < 0.0001, Fig. 5I), the pentose phosphate pathway (0.24-fold, *P*-value = 0.0063, Fig. 5J), and the phosphotransferase system (0.18-fold, *P*-value = 0.0054, Fig. 5K). Overall, these findings reveal that B-768 redistributed its proteome to accommodate the increased energy demand of metabolizing xylose. To do this, various carbohydrate metabolism and energy production genes were overexpressed, which direct carbon flux from xylose through the PPP to yield pyruvate. Pyruvate is then converted into energy by channeling carbon and electron fluxes through pyruvate metabolism, then the TCA cycle, followed by the electron transport system for ATP synthesis. Although ATP synthase was not significantly different between xylose and glucose conditions, other key proteins in the electron transport system were upregulated in xylose, such as NADH dehydrogenase by 8.93 log₂-fold (*q*-value = 3.63E-9) and succinate dehydrogenase by 3.72 log₂-fold (*q*-value = 2.42E-5). A significant reduction in LDH expression and lactate production, and an increase in pyruvate metabolism correlates with the decreased lactate metabolism for growth on xylose.

Lastly, amino acid reallocation in the measured proteomes was performed as previously described to investigate the metabolic cost associated with switching to a xylose-utilizing proteome (66) (Fig. 5L). Higher amounts of costly amino acids like tryptophan, tyrosine, and phenylalanine are used in xylose-only conditions compared to glucose-only conditions (67). Consistent with this finding, the 10 most highly expressed proteins in xylose possess 107% more tryptophan, 37% more tyrosine, and 21% more phenylalanine than the top 10 proteins for glucose. The greater demand for these costly amino acids could help explain the marked investment into energy generation processes when B-768 is grown on xylose. This finding suggests that supplementation of *B. coagulans* B-768 with metabolically expensive amino acid(s) may alleviate part of the metabolic burden due to amino acid biosynthesis and improve growth on xylose.

Carbon catabolite repression is dominant in B-768 growing on glucose-xylose mixture and switchgrass biomass hydrolysate

As cellulosic hydrolysates contain a mixture of C5 and C6 sugars, it is important to understand how B-768 regulates its metabolism for growth on mixed sugars. Carbon catabolite repression (CCR) is a common regulatory mechanism used by bacteria to

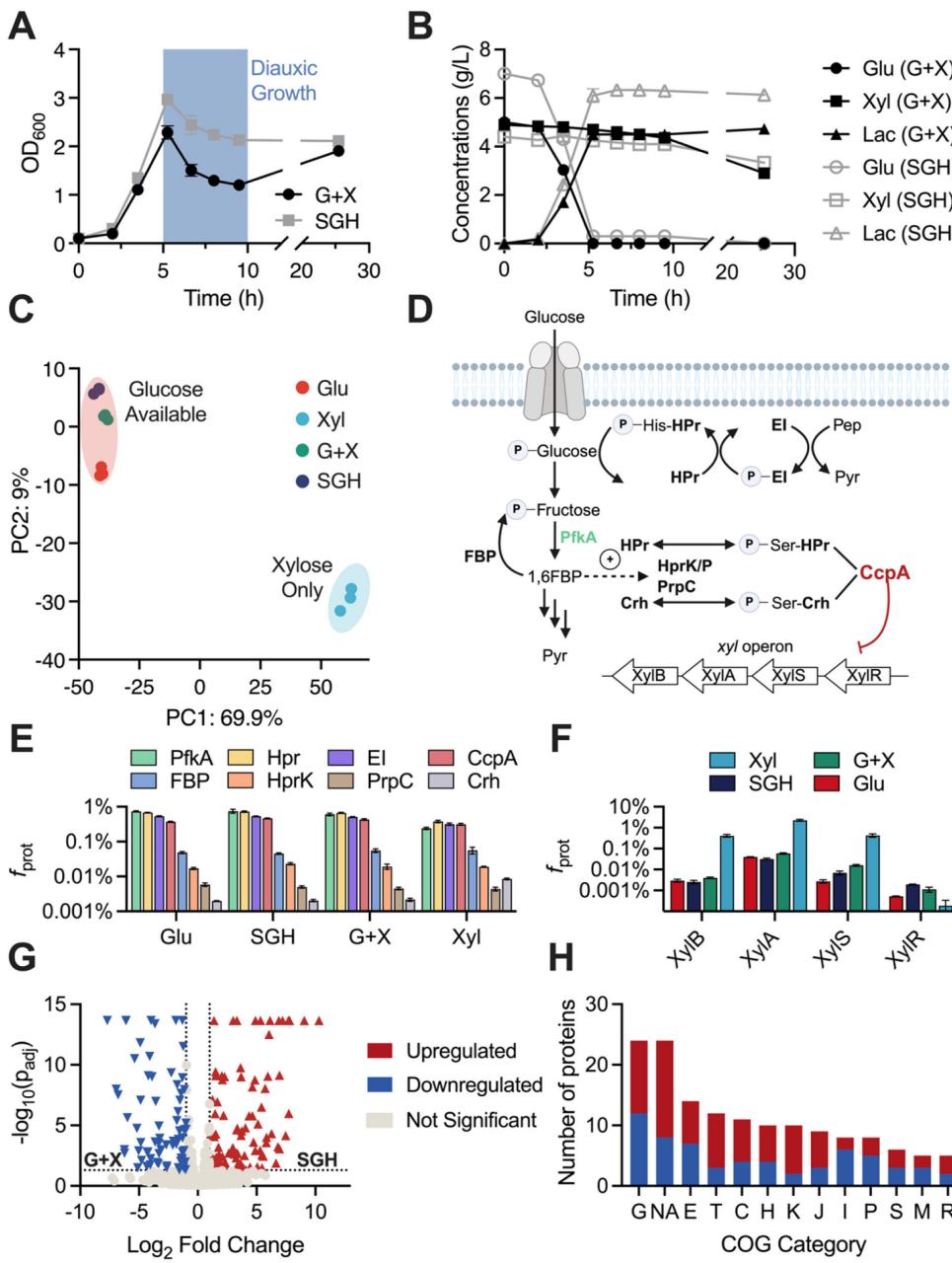


FIG 6 Growth of B-768 in refined sugars and raw hydrolysate reveals mechanisms of cellular robustness and carbon catabolite repression. (A) Growth profiles of B-768 grown on glucose + xylose (G + X) mixture and switchgrass hydrolysate (SGH). (B) Glucose and xylose consumption and lactate production profiles of B-768 grown on G + X and SGH. (C) Principal component analysis plot of B-768 proteomes grown on different sugars. (D) Schematic of CCR in *Bacillus* species. (E) Levels of CCR proteins in B-768 proteomes grown on different sugars. (F) Levels of *xyl* proteins in B-768 proteomes grown on different sugars. (G) Volcano plot of differentially expressed proteins between SGH and G + X conditions. (H) Number of significantly perturbed proteins by COG category. COG Categories: G = Carbohydrate Transport and Metabolism; E = Amino Acid Transport and Metabolism; T = Signal Transduction Mechanisms; C = Energy Production and Conversion; H = Coenzyme Transport and Metabolism; NA = No COG annotation; K = Transcription; J = Translation; I = Lipid Transport and Metabolism; P = Inorganic Ion Transport and Metabolism; S = Function Unknown; M = Cell Wall/Membrane/Envelope Biogenesis; R = General Function Prediction Only.

inhibit the metabolism of sub-optimal carbon sources (e.g., xylose) in the presence of a preferred substrate (e.g., glucose). The diauxic growth observed when B-768 is grown on either a mixture of refined glucose and xylose (G + X) or switchgrass hydrolysate (SGH) is evidence for a strong CCR on xylose assimilation by glucose (Fig. 6A and B). Consistent with this observation (Fig. 6B), the expression of the *xyI* operon was heavily repressed when glucose was available (Fig. 6D and F). The effects of the observed mechanism can be captured at the cellular level. Principal component analysis shows that all proteomes of cultures growing in the presence of glucose were clustered tightly together, while those grown on xylose were highly diverged (Fig. 6C).

To characterize CCR in *B. coagulans* B-768 and to select genetic targets for eliminating it, canonical CCR proteins were quantified across a variety of growth conditions based on the CCR model of the well-studied *B. subtilis* (Fig. 6D and E). To enable CCR, the master regulator CcpA and its primary corepressor, histidine-containing phosphocarrier protein (HPr), or the secondary catabolite corepressor (Crh), bind to catabolite-responsive elements (*cre*) located in the open reading frame or promoter region of the target genes and block transcription (68) (Fig. 6D). Hpr and Crh bind CcpA when phosphorylated at the Ser46 residue by the bifunctional HPr kinase/phosphorylase (HPrK/P), whose activity is increased by the glycolytic intermediate, fructose-1,6-bisphosphate (FBP) (69–71). CCR is further modulated by glucose-6-phosphate (G6P) and FBP, which improve the DNA-binding affinity of CcpA-(Hpr-Ser46-P), but not CcpA-(Crh-Ser46-P) (72, 73). The results show that protein levels of CcpA, HPr, and HprK/P did not significantly change between glucose-available and xylose-only conditions, agreeing with a previous report (Fig. 6E) (74). While not statistically significant, phosphofructokinase-1 (PfkA), which converts fructose-6-phosphate into FBP, is downregulated by 1.47 log₂-fold changes in xylose cultures relative to glucose. For growth on xylose, the FBP node of the EMP pathway can be effectively bypassed because carbon passed through the PPP connects to glycolysis downstream of FBP at glyceraldehyde-3-P. Lower concentrations of FBP would, in turn, downregulate the kinase activity of HprK and DNA-binding affinity of the CcpA effector complex (Fig. 6D). Therefore, in B-768, relief of CCR on xylose appears to be likely due to reduced intracellular levels of FBP.

Growth hardiness of *B. coagulans* B-768 on complex biomass hydrolysate is explained by proteomics

A major challenge to utilizing lignocellulosic sugars is overcoming microbially inhibitory chemicals formed during pretreatment and hydrolysis of the biomass. Depending on their concentration, they can lead to extended lag phases, reduced product yields, or stalled fermentations. While there are methods to mitigate these inhibitors, their drawbacks include added cost, sugar loss, and reduced overall conversion efficiency (75, 76). Therefore, using microbial biocatalysts that are robust to the presence of inhibitory compounds is paramount (77). To elucidate the cellular robustness of B-768, its growth and lactate production were compared for undetoxified dilute-acid switchgrass hydrolysate and refined sugar control (Fig. 6A and B). Surprisingly, no inhibitory effect on growth and lactate production was detected for B-768 grown on SGH compared to refined sugars. This result indicates that B-768 can utilize glucose and—to a much lesser extent—xylose present in undetoxified biomass hydrolysate without a reduction in growth or lactate production, which makes this strain particularly attractive as a microbial biomanufacturing platform.

To investigate the molecular drivers for the observed cellular robustness, comparative proteomics was performed between cells grown in SGH and X + G (Fig. 6G). COG analysis was performed on the 161 significantly differential proteins and all but 24 were assigned to a COG category (Fig. 6H). Most proteins that could not be assigned a COG were annotated as hypothetical proteins, but a few notable proteins included a putative transposase, a putative type VII secretion system protein, a flagellar protein, and two sporulation proteins. All these proteins are stress-implicated and probably play a role in helping B-768 adapt to hydrolysate toxicity. The top three COG categories were COG-G (*n*

= 24), COG-E ($n = 14$), and COG-T ($n = 12$). In terms of carbon metabolism, all three major arabinose utilization proteins, L-arabinose isomerase, L-ribulokinase, and L-ribulose-5-P 4-epimerase were significantly upregulated in SGH relative to X + G by 5.33 log₂-fold (q-value = 2.15E-14), 6.82 log₂-fold (q-value = 5.98E-4), and 7.47 log₂-fold (q-value = 6.88E-5), respectively. In addition, 6-phospho-β-glucosidase (cellobiose utilization) and galactokinase (galactose utilization) were upregulated in SGH by 9.06 log₂-fold (q-value = 2.15E-14) and 4.07 log₂-fold (q-value = 1.17E-4), respectively (Data Set S4). This result indicates that B-768 actively metabolizes arabinose, cellobiose, and galactose when grown on unrefined biomass hydrolysates.

In addition to these proteins, several transporters of both the ABC and PTS varieties ($n = 9$) and chemotaxis-implicated proteins ($n = 4$) were significantly altered (Table 5). Similar proteins were previously found to be significantly perturbed in *Clostridium acetobutylicum* in response to phenolic inhibitors—common toxins found in biomass hydrolysates (78, 79). Another inhibitor found in biomass hydrolysate is methylglyoxal (80, 81); therefore, it is noteworthy that the most upregulated protein in SGH relative to X + G was glyoxylase II, a crucial enzyme for detoxifying methylglyoxal (82). Finally, 10 putative transcriptional regulators were differentially expressed (eight up, two down), which may actuate transcriptional changes to accommodate growth in hydrolysates (Table 5). Together, these results provide promising engineering targets for further improving B-768's exceptional hydrolysate tolerance and demonstrate that B-768 displays a robust and diverse proteomic response to hydrolysate toxicity, enabling favorable growth kinetics and lactate titers in comparison to other industrial biocatalysts.

Conclusions

Through detailed phenotype characterization, comparative functional genomics, and quantitative proteomic analyses, this study elucidates the mechanisms underlining the robustness of undomesticated *B. coagulans* strain B-768. Highlighted is the metabolic capability of B-768 to utilize a wide range of C5 and C6 sugars at various temperatures, with glucose as the preferential substrate. We created a high-quality genome assembly and annotation of *B. coagulans* B-768. Essential to the robustness of B-768 is its larger genome including the expanded Carbohydrate Transport and Metabolism and Mobilome, which enables improved sugar utilization and horizontal gene transfer. A large set of genes devoted to its innate and adaptive defense mechanisms should help *B. coagulans* resist phage attacks, which are a risk for bacterial industrial fermentation processes. B-768 exhibits distinct proteomes for growth on glucose and xylose, consistent with its observed phenotypes. Faster growth and glucose uptake rate resulted in lactate overflow metabolism while slower growth and xylose uptake rate diminished lactate production. This difference could be explained by the tradeoff in proteome reallocation between the Carbohydrate Transport and Metabolism process and the Energy Generation and Production process. B-768 can utilize biomass hydrolysates with various C5 and C6 sugars controlled by a tightly regulated gene network. Overall, B-768 is a robust microbial biomanufacturing platform that has potential beyond lactate fermentation. A fundamental understanding of its robustness revealed in this study will be critical for rechanneling and tuning its metabolism for optimized conversion of lignocellulosic biomass to chemicals, fuels, and materials beyond lactate such as designer bioesters (5, 66, 83). Even though genetic modification of *B. coagulans* is feasible (84), available genetic tools are much less developed and efficient for *B. coagulans* than *B. subtilis*. Developing robust genetic tools to manipulate *B. coagulans* will expand our understanding of its genetics, metabolism, and regulation and harness it for biomanufacturing applications.

TABLE 5 Differential protein expression for candidate proteins contributing to cellular robustness of B-768 growing in SGH

Protein ID	Function	Class	Log ₂ -fold change	q-value
Ga0586967_01_1426112_1427749	Methyl-accepting chemotaxis protein	Chemotaxis	7.21	2.2E-14
Ga0586967_01_2953390_2954649	Methyl-accepting chemotaxis protein	Chemotaxis	2.1	0.00377
Ga0586967_01_1922219_1923952	Methyl-accepting chemotaxis protein	Chemotaxis	2.02	0.0171
Ga0586967_01_3602104_3603432	Methyl-accepting chemotaxis protein	Chemotaxis	-6.24	0.00197
Ga0586967_01_1885126_1886544	PTS system trehalose-specific IIC component	Transport	-1.06	0.0103
Ga0586967_01_1886575_1887057	PTS system glucose-specific IIA component	Transport	-1.18	0.00121
Ga0586967_01_1467232_1467561	PTS system cellobiose-specific IIA component	Transport	-1.38	0.00901
Ga0586967_01_1541669_1542151	Fructoselysine and glucoselysine-specific PTS system IIB component	Transport	-2.64	0.0316
Ga0586967_01_2529421_2529912	Mannose/fructose/N-acetylgalactosamine-specific phosphotransferase system component IIB	Transport	-3.63	0.00023
Ga0586967_01_1189660_1191024	PTS system ascorbate-specific IIC component	Transport	-3.89	9.7E-10
Ga0586967_01_1452905_1453891	LacI family transcriptional regulator	Regulator	7.71	2.2E-14
Ga0586967_01_1692988_1693203	Putative transcriptional regulator	Regulator	5.56	0.00345
Ga0586967_01_633467_634027	TetR/AcrR family transcriptional repressor of ImrAB and yxaGH operons	Regulator	3.97	0.00309
Ga0586967_01_3908323_3909216	Phenylacetic acid degradation operon negative regulatory protein	Regulator	3.62	0.0132
Ga0586967_01_2622621_2623172	DNA-binding transcriptional ArsR family regulator	Regulator	3.09	0.00232
Ga0586967_01_1173785_1174315	DNA-binding MarR family transcriptional regulator	Regulator	2.63	0.0019
Ga0586967_01_3081124_3082011	DNA-binding transcriptional LysR family regulator	Regulator	2.27	9.3E-10
Ga0586967_01_3363633_3364772	GntR family transcriptional regulator of arabinose operon	Regulator	2.03	7.9E-10
Ga0586967_01_1915388_1915990	AcrR family transcriptional regulator	Regulator	-3.24	0.00012
Ga0586967_01_3584855_3586105	DNA-binding transcriptional ArsR family regulator	Regulator	-4.86	8.4E-08

MATERIALS AND METHODS

Cell culture

Strains

B. coagulans strain B-768 (ATCC 10545 = NRS-784) was obtained from the NRRL Agriculture Research Service strain collection of the United States Department of Agriculture (USDA) located in Peoria, IL. The strain was stored as glycerol stocks at -80°C.

Media

The basal medium contained 1.5 g/L yeast extract, 100 mM sodium MES buffer (pH 6.5), and mineral solution. The mineral solution contained the following: 2 g/L (NH₄)₂SO₄, 2 g/L KH₂PO₄, 2 g/L NaCl, 0.2 g/L MgSO₄·7H₂O, 5 mg/L MnSO₄·H₂O, and 10 mg/L FeSO₄·7H₂O. The phosphate-buffered saline (PBS) contained per liter: 8 g NaCl, 0.2 g KCl, and 1.44 g Na₂HPO₄, adjusted to pH 7.4 with NaOH.

The switchgrass hydrolysate (SGH) medium contained per liter: 10% (vol/vol) of concentrated dilute acid pretreated SGH and 100 mM MES buffer (pH 6.5). The concentrated SGH solution was prepared as previously described (85), which contained ~70 g/L glucose and ~44 g/L xylose as the main C5/C6 sugars of biomass hydrolysates. SGH was used at 10% wt/w for bottle cultures.

High-throughput microplate growth kinetic experiments

Growth of B-768 on various carbon sources (glucose, xylose, cellobiose, galactose, mannose, or arabinose) at various temperatures (37°C, 40°C, 45°C, 47°C, 53°C, or 55°C) was measured high-throughput using Bioscreen Pro (Oy Growth Curves Ab Ltd, Finland).

B-768 was inoculated in basal medium containing no sugar and grown at 50°C overnight. The next day, overnight grown cells were diluted in fresh medium with an initial OD₆₀₀ of 0.1 in total 300 µL and growth was monitored at 580 nm for 24 or 48 h. For the lactate tolerance experiment, cells were prepared in a basal medium with 8 g/L glucose, and growth was monitored at 50°C using the Bioscreen Pro.

Bottle cultures

Seed cultures and media were prepared as described for the high-throughput studies. Fermentations were conducted in 25 mL Pyrex bottles (Corning Life Sciences, NY) filled with 20 mL of medium and incubated and mixed at 50°C and 100 rpm, respectively. Cultures were sampled for sugars (glucose and xylose), lactate, and OD₆₀₀ using a 1 mL sample. Proteomic samples (1 mL) were withdrawn, centrifuged, and washed with chilled dH₂O, and the pellet was stored at -80°C.

pH-controlled fermentations

Fermentations were maintained at pH 6.5 and 50°C (unless stated otherwise) and stirred at 300 rpm. Cells were inoculated to an initial OD₆₀₀ of 0.1. The pH was maintained by adding 4 N KOH. Cultures were sampled for OD₆₀₀, sugars, and fermentation products.

Analytical methods

High-performance liquid chromatography

Metabolites such as sugars and acids (e.g., lactate and acetate) were measured by Bio-Rad HPX-87H column at 65°C with 5 mM sulfuric acid as a mobile phase using a Thermo high-performance liquid chromatography (Ultimate 3000) system equipped with a refractive index (RI).

Liquid chromatography with tandem mass spectrometry

B. coagulans strain B-768 cultured in the presence of various sugars ($n = 3$ per glucose, xylose, a mixture of glucose and xylose, and switchgrass hydrolysate) were collected by centrifugation and cell pellets prepared for liquid chromatography with tandem mass spectrometry (LC-MS/MS)-based proteomics as previously described (86).

Genome sequencing

The genomic DNA of *B. coagulans* B-768 was subjected to Pacific Biosciences (PacBio) long-read sequencing by DOE Joint Genome Institute (JGI) (87). A >10 kb Pacbio SMRTbell library was constructed and sequenced on the PacBio Sequel platform, which generated 78,854 high-fidelity CCS reads totaling 696,902,661 bp. Reads >5 kb were assembled with Flye v2.8.3 using default settings (88, 89). The final draft assembly contained one contig in one scaffold, totaling 3,944,476 bp in size, and was 99.95% complete according to CheckM (90). The input read coverage was 181.4x. The final draft assembly was annotated using the Integrated Microbial Genomes Annotation Pipeline (IMGAP) v5.1.1.

Bioinformatics and data analysis

CRISPR system and spacer identification

CRISPR system prediction was performed using CRISPRCasTyper v1.8.0 with default settings (91). For identification of spacer hits on plasmids, nucleotide sequences for the entire NCBI collection of bacterial plasmid sequences were downloaded from PLSDB v. 2021_06_23_v2 (92) and used to construct a BLASTn database against which the JGI-predicted CRISPR spacers were blasted using the blastn-short algorithm with default settings. Hits with e-value <0.05 were retained and the top hit for each spacer was

chosen as the target source. For identification of spacer hits on phages, SpacePHARER v5.c2e680a (93) was used with the GenBank_phage_2018_09 database and default settings.

Restriction-modification annotation

All B-768 protein sequences were searched against the complete REBASE database using BLASTp v2.15.0 + with default settings. Hits were filtered down to those with a percent identity >60% and alignment coverage >80% (alignment length/query length). Hits were then sorted by ascending e-value and the top hit for each query was retained. 11 hits remained and all had e-values < 7.5E-42. These were considered bona fide R-M homologs and the REBASE classification was used to assign the type.

Circular genome map

Proksee v5.2.1 (94) was used to generate a circular genome map for *B. coagulans* B-768 from functional annotation provided by JGI. GC content and skew were calculated using the built-in function in Proksee. COG mapping was performed manually using the feature upload tool and the output from COGclassifier v1.0.5 (95). Prophage regions were predicted using PHASTEST v3.0 and mapped using the feature upload tool (42).

Pan-genome phylogeny

Genome sequences of all RefSeq *B. coagulans* strains, except RUG1433, were downloaded from NCBI. A core-genome alignment was performed on the pan-genome using Parsnp v1.7.4 with default settings (96). A recombination-aware maximum likelihood tree was then constructed from the core-genome alignment using Gubbins v3.2.1 with default settings (97).

Proteomics

The resulting MS/MS spectra acquired from the Q Exactive MS were searched against the *B. coagulans* B-768 proteome database appended with common protein contaminants using the SEQUEST HT algorithm in Proteome Discoverer v.2.5 (Thermo Scientific) and peptide spectrum matches (PSMs) scored and filtered with Percolator [settings and dynamic/static mods described in Walker et al's study (86)]. Peptides (FDR < 1%) were then quantified by chromatographic peak area and derived abundances mapped back to each peptide's respective protein and peptide abundance summed to estimate protein-level abundances. Protein abundances were then log₂ transformed, normalized by LOESS, and median centered using InfernoRDN (98). Missing values were imputed, and statistical tests (T-test and ANOVA) were performed using Perseus v2.0.7.0 (99). Log₂ transformed and imputed protein abundance data were analyzed using the DEP R package v1.24.0 (100). Significant proteins were filtered out based on a 5% FDR and minimum log₂ fold-change of 1.

Proteome reallocation

Based on the mass fractions of each protein per sample, proteome reallocation was calculated by dividing each raw protein abundance by the sum of all protein abundances for a given sample and summed up across the indicated categories of interest (66, 86). COG annotations were mapped onto predicted protein sequences from JGI using COGclassifier v1.0.5 (95). All calculations were performed in R v4.3.0.

Amino acid allocation

Analyses of amino acid allocation in the measured proteomes were determined based on the formulation previously described (66, 86). All the calculations were performed in Python v3.8.16 and the outputs were analyzed in R v4.3.0.

Kinetics

The specific growth rate was determined by assuming that cell growth follows the first-order kinetics (Equation 1).

$$\frac{dX}{dt} = \mu \cdot X \quad (1)$$

where X (g DCW/L) is cell mass, t (h) is the culturing time, and μ (1/h) is the specific growth rate. Note that DCW represents dry cell weight and is linearly correlated with the measured cell optical density (OD_{600}). The substrate uptake rate was determined as follows (Equation 2):

$$-r_S = \mu \cdot Y_{S/X} \quad (2)$$

where r_S (g/g DCW/h) is the specific substrate uptake rate and $Y_{S/X}$ (g substrate/g CDW) is the yield of a substrate (e.g., glucose) to cell mass. The product formation rate was determined as follows (Equation 3):

$$r_P = \mu \cdot Y_{P/X} \quad (3)$$

where r_P (g/g DCW/h) is the specific substrate uptake rate and $Y_{P/X}$ (g substrate/g CDW) is the yield of a product (e.g., lactate) to cell mass.

Calculation of optimal growth temperatures

Optimal growth temperatures on various sugars were determined by fitting maximum growth rate versus temperature with a non-linear model (continuous hinge function, segmental regression lines with gentle connection, least squares fit) using GraphPad Prism v10.2.3.

ACKNOWLEDGMENTS

This research was funded by the DOE BER award (DE-SC0022226 to C.T.T., R.J.G., and B.S.D.). The sequencing project was supported by the Center for Bioenergy Innovation (CBI), which is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the US DOE under Contract Number DE-AC05-00OR22725. The work (Project IDs # 510751, 509417) conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported under Contract No. DE-AC02-05CH11231. B.S.D., J.E., and P.J.S. received financial support from the U.S. Department of Agriculture, Agricultural Research Service (CRIS Numbers 5010-41000-189).

The mention of trade names or commercial products in this article is solely to provide scientific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer.

AUTHOR AFFILIATIONS

¹Department of Chemical and Biomolecular Engineering, University of Tennessee Knoxville, Knoxville, Tennessee, USA

²Center for Bioenergy Innovation, Oak Ridge, Tennessee, USA

³Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

⁴United States Department of Agriculture (USDA), Agricultural Research Service (ARS), National Center for Agricultural Utilization Research (NCAUR), Bioenergy Research Unit, Peoria, Illinois, USA

AUTHOR ORCIDs

David Dooley  <http://orcid.org/0000-0001-6793-4463>
 Seunghyun Ryu  <http://orcid.org/0000-0003-0089-7922>
 Cong T. Trinh  <http://orcid.org/0000-0002-8362-725X>

FUNDING

Funder	Grant(s)	Author(s)
U.S. Department of Energy (DOE)	DE-SC0022226	Richard J. Giannone
		Bruce S. Dien
		Cong T. Trinh
Joint Genome Institute (JGI)	DE-AC02-05CH11231	Seunghyun Ryu
		Cong T. Trinh
DOE SC Center for Bioenergy Innovation (CBI) (DE-AC05-00OR22725)		Seunghyun Ryu
		Richard J. Giannone
		Cong T. Trinh

AUTHOR CONTRIBUTIONS

David Dooley, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review and editing | Seunghyun Ryu, Formal analysis, Investigation, Validation, Visualization, Writing – original draft, Writing – review and editing | Richard J. Giannone, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Software, Writing – review and editing | Jackson Edwards, Data curation, Investigation | Bruce S. Dien, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Writing – review and editing | Patricia J. Slininger, Methodology, Writing – review and editing | Cong T. Trinh, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Visualization, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

The whole-genome assemblies and annotation of *B. coagulans* B-768 were deposited at DDBJ/EMBL/GenBank under Bioproject ID [PRJNA1123721](#). All raw mass spectra for quantification of proteins used in this study have been deposited in the MassIVE and ProteomeXchange data repositories under accession numbers [MSV000095240](#) (MassIVE) and [PXD053622](#) (ProteomeXchange).

ADDITIONAL FILES

The following material is available [online](#).

Supplemental Material

Data Set S1 (mSystems00952-24-s0001.xlsx). Annotated carbohydrate transport in *B. coagulans* B-768.

Data Set S2 (mSystems00952-24-s0002.xlsx). Annotation of representative core metabolism of *B. coagulans* B-768.

Data Set S3 (mSystems00952-24-s0003.xlsx). Differentially expressed proteins between xylose and glucose.

Data Set S4 (mSystems00952-24-s0004.xlsx). Differentially expressed proteins between switchgrass hydrolysate and xylose + glucose.

Supplemental Material (mSystems00952-24-s0005.pdf). Figures S1-S5.

REFERENCES

1. Lynd LR, Beckham GT, Guss AM, Jayakody LN, Karp EM, Maranas C, McCormick RL, Amador-Noguez D, Bomble YJ, Davison BH, Foster C, Himmel ME, Holwerda EK, Laser MS, Ng CY, Olson DG, Román-Leshkov Y, Trinh CT, Tuskan GA, Upadhyay V, Vardon DR, Wang L, Wyman CE. 2022. Toward low-cost biological and hybrid biological/catalytic conversion of cellulosic biomass to fuels. *Energy Environ Sci* 15:938–990. <https://doi.org/10.1039/D1EE02540F>
2. Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, Frederick WJ, Hallett JP, Leak DJ, Liotta CL, Mielenz JR, Murphy R, Templer R, Tschaplinski T. 2006. The path forward for biofuels and biomaterials. *Science* 311:484–489. <https://doi.org/10.1126/science.1114736>
3. Langholtz MH. 2024. 2023 billion-ton report. <https://doi.org/10.23720/BT2023/2316165>.
4. Zoghlami A, Paës G. 2019. Lignocellulosic biomass: understanding recalcitrance and predicting hydrolysis. *Front Chem* 7:874. <https://doi.org/10.3389/fchem.2019.00874>
5. Lee J-W, Trinh CT. 2020. Towards renewable flavors, fragrances, and beyond. *Curr Opin Biotechnol* 61:168–180. <https://doi.org/10.1016/j.copbio.2019.12.017>
6. Bond JQ, Upadhye AA, Olcay H, Tompsett GA, Jae J, Xing R, Alonso DM, Wang D, Zhang T, Kumar R, Foster A, Sen SM, Maravelias CT, Malina R, Barrett SRH, Lobo R, Wyman CE, Dumesic JA, Huber GW. 2014. Production of renewable jet fuel range alkanes and commodity chemicals from integrated catalytic processing of biomass. *Energy Environ Sci* 7:1500–1523. <https://doi.org/10.1039/C3EE43846E>
7. Holbery J, Houston D. 2006. Natural-fiber-reinforced polymer composites in automotive applications. *JOM* 58:80–86. <https://doi.org/10.1007/s11837-006-0234-2>
8. Lee JW, Na D, Park JM, Lee J, Choi S, Lee SY. 2012. Systems metabolic engineering of microorganisms for natural and non-natural chemicals. *Nat Chem Biol* 8:536–546. <https://doi.org/10.1038/nchembio.970>
9. Keasling JD, Chou H. 2008. Metabolic engineering delivers next-generation biofuels. *Nat Biotechnol* 26:298–299. <https://doi.org/10.1038/nbt0308-298>
10. Napora-Wijata K, Strohmeier GA, Winkler M. 2014. Biocatalytic reduction of carboxylic acids. *Biotechnol J* 9:822–843. <https://doi.org/10.1002/biot.201400012>
11. Trinh CT, Unrean P, Srienc F. 2008. Minimal *Escherichia coli* cell for the most efficient production of ethanol from hexoses and pentoses. *Appl Environ Microbiol* 74:3634–3643. <https://doi.org/10.1128/AEM.02708-07>
12. Weber C, Farwick A, Benisch F, Brat D, Dietz H, Subtil T, Boles E. 2010. Trends and challenges in the microbial production of lignocellulosic bioalcohol fuels. *Appl Microbiol Biotechnol* 87:1303–1315. <https://doi.org/10.1007/s00253-010-2707-z>
13. Lynd LR, Liang X, Biddy MJ, Allee A, Cai H, Foust T, Himmel ME, Laser MS, Wang M, Wyman CE. 2017. Cellulosic ethanol: status and innovation. *Curr Opin Biotechnol* 45:202–211. <https://doi.org/10.1016/j.copbio.2017.03.008>
14. Kunjapur AM, Prather KLJ. 2015. Microbial engineering for aldehyde synthesis. *Appl Environ Microbiol* 81:1892–1901. <https://doi.org/10.1128/AEM.03319-14>
15. Yuzawa S, Mirsiaghi M, Jocic R, Fujii T, Masson F, Benites VT, Baidoo EEK, Sundstrom E, Tanjore D, Pray TR, George A, Davis RW, Gladden JM, Simmons BA, Katz L, Keasling JD. 2018. Short-chain ketone production by engineered polyketide synthases in *Streptomyces albus*. *Nat Commun* 9:4569. <https://doi.org/10.1038/s41467-018-07040-0>
16. Kim DI, Chae TU, Kim HU, Jang WD, Lee SY. 2021. Microbial production of multiple short-chain primary amines via retrobiosynthesis. *Nat Commun* 12:173. <https://doi.org/10.1038/s41467-020-20423-6>
17. Layton DS, Trinh CT. 2014. Engineering modular ester fermentative pathways in *Escherichia coli*. *Metab Eng* 26:77–88. <https://doi.org/10.1016/j.jymben.2014.09.006>
18. Thompson RA, Trinh CT. 2014. Enhancing fatty acid ethyl ester production in *Saccharomyces cerevisiae* through metabolic engineering and medium optimization. *Biotechnol Bioeng* 111:2200–2208. <https://doi.org/10.1002/bit.25292>
19. Anjum A, Zuber M, Zia KM, Noreen A, Anjum MN, Tabassum S. 2016. Microbial production of polyhydroxyalkanoates (PHAs) and its copolymers: a review of recent advancements. *Int J Biol Macromol* 89:161–174. <https://doi.org/10.1016/j.ijbiomac.2016.04.069>
20. Cappello J. 1990. The biological production of protein polymers and their use. *Trends Biotechnol* 8:309–311. [https://doi.org/10.1016/0167-7799\(90\)90207-e](https://doi.org/10.1016/0167-7799(90)90207-e)
21. Ubando AT, Felix CB, Chen W-H. 2020. Biorefineries in circular bioeconomy: a comprehensive review. *Bioresour Technol* 299:122585. <https://doi.org/10.1016/j.biortech.2019.122585>
22. Almeida JR, Modig T, Petersson A, Hähn - Hägerdal B, Lidén G, Gorwa - Grauslund MF. 2007. Increased tolerance and conversion of inhibitors in lignocellulosic hydrolysates by *Saccharomyces cerevisiae*. *J Chem Tech Biotech* 82:340–349. <https://doi.org/10.1002/jctb.1676>
23. Wang S, Sun X, Yuan Q. 2018. Strategies for enhancing microbial tolerance to inhibitors for biofuel production: a review. *Bioresour Technol* 258:302–309. <https://doi.org/10.1016/j.biortech.2018.03.064>
24. Chubukov V, Gerosa L, Kochanowski K, Sauer U. 2014. Coordination of microbial metabolism. *Nat Rev Microbiol* 12:327–340. <https://doi.org/10.1038/nrmicro3238>
25. Stephanopoulos G. 2007. Challenges in engineering microbes for biofuels production. *Science* 315:801–804. <https://doi.org/10.1126/science.1139612>
26. Zheng Z, Lin X, Jiang T, Ye W, Ouyang J. 2016. Genomic analysis of a xylose operon and characterization of novel xylose isomerase and xylulokinase from *Bacillus coagulans* NL01. *Biotechnol Lett* 38:1331–1339. <https://doi.org/10.1007/s10529-016-2109-0>
27. Zheng Z, Jiang T, Zou L, Ouyang S, Zhou J, Lin X, He Q, Wang L, Yu B, Xu H, Ouyang J. 2018. Simultaneous consumption of cellobiose and xylose by *Bacillus coagulans* to circumvent glucose repression and identification of its cellobiose-assimilating operons. *Biotechnol Biofuels* 11:320. <https://doi.org/10.1186/s13068-018-1323-5>
28. Yao K, Zhou Q, Liu D, Chen S, Yuan K. 2019. Comparative proteomics of the metabolic pathways involved in l-lactic acid production in *Bacillus coagulans* BCS13002 using different carbon sources. *LWT* 116:108445. <https://doi.org/10.1016/j.lwt.2019.108445>
29. Aulitto M, Fusco S, Bartolucci S, Franzén CJ, Contursi P. 2017. *Bacillus coagulans* MA-13: a promising thermophilic and cellulolytic strain for the production of lactic acid from lignocellulosic hydrolysate. *Biotechnol Biofuels* 10:210. <https://doi.org/10.1186/s13068-017-0896-8>
30. Aulitto M, Fusco S, Nickel DB, Bartolucci S, Contursi P, Franzén CJ. 2019. Seed culture pre-adaptation of *Bacillus coagulans* MA-13 improves lactic acid production in simultaneous saccharification and fermentation. *Biotechnol Biofuels* 12:45. <https://doi.org/10.1186/s13068-019-1382-2>
31. Su F, Xu P. 2014. Genomic analysis of thermophilic *Bacillus coagulans* strains: efficient producers for platform bio-chemicals. *Sci Rep* 4:3926. <https://doi.org/10.1038/srep03926>
32. Riazi S, Wirawan RE, Badmaev V, Chikindas ML. 2009. Characterization of lactosporin, a novel antimicrobial protein produced by *Bacillus coagulans* ATCC 7050. *J Appl Microbiol* 106:1370–1377. <https://doi.org/10.1111/j.1365-2672.2008.04105.x>
33. Hyronimus B, Le Marrec C, Urdaci MC. 1998. Coagulin, a bacteriocin-like inhibitory substance produced by *Bacillus coagulans* I4. *J Appl Microbiol* 85:42–50. <https://doi.org/10.1046/j.1365-2672.1998.00466.x>
34. Khatri I, Sharma S, Ramya TNC, Subramanian S. 2016. Complete genomes of *Bacillus coagulans* S-lac and *Bacillus subtilis* TO-A JPC, two phylogenetically distinct probiotics. *PLoS One* 11:e0156745. <https://doi.org/10.1371/journal.pone.0156745>
35. Meier-Kolthoff JP, Göker M. 2019. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat Commun* 10:2182. <https://doi.org/10.1038/s41467-019-10210-3>
36. Jain C, Rodriguez-R LM, Phillip AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 9:5114. <https://doi.org/10.1038/s41467-018-07641-9>
37. De Clerck E, Rodriguez-Diaz M, Forsyth G, Lebbe L, Logan NA, DeVos P. 2004. Polyphasic characterization of *Bacillus coagulans* strains, illustrating heterogeneity within this species, and emended description

- of the species. *Syst Appl Microbiol* 27:50–60. <https://doi.org/10.1078/0723-2020-00250>
38. Patel MA, Ou MS, Harbrucker R, Aldrich HC, Buszko ML, Ingram LO, Shanmugam KT. 2006. Isolation and characterization of acid-tolerant, thermophilic bacteria for effective fermentation of biomass-derived sugars to lactic acid. *Appl Environ Microbiol* 72:3228–3235. <https://doi.org/10.1128/AEM.72.5.3228-3235.2006>
39. Patel MA, Ou MS, Ingram LO, Shanmugam KT. 2005. Simultaneous saccharification and co-fermentation of crystalline cellulose and sugar cane bagasse hemicellulose hydrolysate to lactate by a thermotolerant acidophilic *Bacillus* sp. *Biotechnol Prog* 21:1453–1460. <https://doi.org/10.1021/bp0400339>
40. Ouyang J, Cai C, Chen H, Jiang T, Zheng Z. 2012. Efficient non-sterilized fermentation of biomass-derived xylose to lactic acid by a thermotolerant *Bacillus coagulans* NL01. *Appl Biochem Biotechnol* 168:2387–2397. <https://doi.org/10.1007/s12010-012-9944-9>
41. Ahorsu R, Cintorino G, Medina F, Constantí M. 2019. Microwave processes: a viable technology for obtaining xylose from walnut shell to produce lactic acid by *Bacillus coagulans*. *J Clean Prod* 231:1171–1181. <https://doi.org/10.1016/j.jclepro.2019.05.289>
42. Wishart DS, Han S, Saha S, Oler E, Peters H, Grant JR, Stothard P, Gautam V. 2023. PHASTEST: faster than PHASTER, better than PHAST. *Nucleic Acids Res* 51:W443–W450. <https://doi.org/10.1093/nar/gkad382>
43. Shimoyama Y. 2022. COGclassifier: a tool for classifying prokaryote protein sequences into COG functional category, vv1.0.5. GitHub. <https://github.com/moshi4/COGclassifier>.
44. Swint-Kruse L, Matthews KS. 2009. Allostery in the LacI/GalR family: variations on a theme. *Curr Opin Microbiol* 12:129–137. <https://doi.org/10.1016/j.mib.2009.01.009>
45. Maddocks SE, Oyston PCF. 2008. Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiol (Reading)* 154:3609–3623. <https://doi.org/10.1099/mic.0.2008/022772-0>
46. Hoskisson PA, Rigali S. 2009. Chapter 1: variation in form and function the helix-turn-helix regulators of the GntR superfamily. *Adv Appl Microbiol* 69:1–22. [https://doi.org/10.1016/S0065-2164\(09\)69001-8](https://doi.org/10.1016/S0065-2164(09)69001-8)
47. Busenlehner LS, Pennella MA, Giedroc DP. 2003. The SmtB/ArsR family of metalloregulatory transcriptional repressors: structural insights into prokaryotic metal resistance. *FEMS Microbiol Rev* 27:131–143. [https://doi.org/10.1016/S0168-6445\(03\)00054-8](https://doi.org/10.1016/S0168-6445(03)00054-8)
48. Pei Z, Sadiq FA, Han X, Zhao J, Zhang H, Ross RP, Lu W, Chen W. 2021. Comprehensive scanning of prophages in *Lactobacillus*: distribution, diversity, antibiotic resistance genes, and linkages with CRISPR-Cas systems. *mSystems* 6:e0121120. <https://doi.org/10.1128/mSystems.01211-20>
49. Bondy-Denomy J, Qian J, Westra ER, Buckling A, Guttman DS, Davidson AR, Maxwell KL. 2016. Prophages mediate defense against phage infection through diverse mechanisms. *ISME J* 10:2854–2866. <https://doi.org/10.1038/ismej.2016.79>
50. Salvetti E, Orrù L, Capozzi V, Martina A, Lamontanara A, Keller D, Cash H, Felis GE, Cattivelli L, Torriani S, Spano G. 2016. Integrate genome-based assessment of safety for probiotic strains: *Bacillus coagulans* GB1-30, 6086 as a case study. *Appl Microbiol Biotechnol* 100:4595–4605. <https://doi.org/10.1007/s00253-016-7416-9>
51. Saroj DB, Gupta AK. 2020. Genome based safety assessment for *Bacillus coagulans* strain LBSC (DSM 17654) for probiotic application. *Int J Food Microbiol* 318:108523. <https://doi.org/10.1016/j.ijfoodmicro.2020.108523>
52. Samson JE, Moineau S. 2013. Bacteriophages in food fermentations: new frontiers in a continuous arms race. *Annu Rev Food Sci Technol* 4:347–368. <https://doi.org/10.1146/annurev-food-030212-182541>
53. Roberts RJ, Belfort M, Bestor T, Bhagwat AS, Bickle TA, Bitinaite J, Blumenthal RM, Degtyarev SK, Dryden DT, Dybvig K. 2003. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res* 31:1805–1812. <https://doi.org/10.1093/nar/gkg274>
54. Mojica FJM, Díez-Villaseñor C, García-Martínez J, Soria E. 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* 60:174–182. <https://doi.org/10.1007/s00239-004-0046-3>
55. Karginov FV, Hannon GJ. 2010. The CRISPR system: small RNA-guided defense in bacteria and archaea. *Mol Cell* 37:7–19. <https://doi.org/10.1016/j.molcel.2009.12.033>
56. Al-Shayeb B, Sachdeva R, Chen L-X, Ward F, Munk P, Devoto A, Castelle CJ, Olm MR, Bouma-Gregson K, Amano Y, et al. 2020. Clades of huge phages from across Earth's ecosystems. *Nature New Biol* 578:425–431. <https://doi.org/10.1038/s41586-020-2007-4>
57. Kraus A, Hueck C, Gärtner D, Hillen W. 1994. Catabolite repression of the *Bacillus subtilis* xyl operon involves a cis element functional in the context of an unrelated sequence, and glucose exerts additional xylR-dependent repression. *J Bacteriol* 176:1738–1745. <https://doi.org/10.1128/jb.176.6.1738-1745.1994>
58. Dahl MK, Schmiedel D, Hillen W. 1995. Glucose and glucose-6-phosphate interaction with xyl repressor proteins from *Bacillus* spp. may contribute to regulation of xylose utilization. *J Bacteriol* 177:5467–5472. <https://doi.org/10.1128/jb.177.19.5467-5472.1995>
59. Walton SL, Bischoff KM, van Heiningen ARP, van Walsum GP. 2010. Production of lactic acid from hemicellulose extracts by *Bacillus coagulans* MXL-9. *J Ind Microbiol Biotechnol* 37:823–830. <https://doi.org/10.1007/s10295-010-0727-4>
60. Kandler O. 1983. Carbohydrate metabolism in lactic acid bacteria. *Antonie Van Leeuwenhoek* 49:209–224. <https://doi.org/10.1007/BF00399499>
61. Gu Y, Ding Y, Ren C, Sun Z, Rodionov DA, Zhang W, Yang S, Yang C, Jiang W. 2010. Reconstruction of xylose utilization pathway and regulons in Firmicutes. *BMC Genomics* 11:1–14. <https://doi.org/10.1186/1471-2164-11-255>
62. Ong SA, Ng ZJ, Wu JC. 2016. Production of high concentration of L-lactic acid from cellobiose by thermophilic *Bacillus coagulans* WCP10-4. *Appl Microbiol Biotechnol* 100:6501–6508. <https://doi.org/10.1007/s00253-016-7600-y>
63. Sternberg D, Vijayakumar P, Reese ET. 1977. β-Glucosidase: microbial production and effect on enzymatic hydrolysis of cellulose. *Can J Microbiol* 23:139–147. <https://doi.org/10.1139/m77-020>
64. Frey PA. 1996. The Leloir pathway: a mechanistic imperative for three enzymes to change the stereochemical configuration of a single carbon in galactose. *FASEB J* 10:461–470. <https://doi.org/10.1096/fasebj.10.4.8647345>
65. Sun T, Altenbuchner J. 2010. Characterization of a mannose utilization system in *Bacillus subtilis*. *J Bacteriol* 192:2128–2139. <https://doi.org/10.1128/JB.01673-09>
66. Seo H, Giannone RJ, Yang Y-H, Trinh CT. 2022. Proteome reallocation enables the selective *de novo* biosynthesis of non-linear, branched-chain acetate esters. *Metab Eng* 73:38–49. <https://doi.org/10.1016/j.ymben.2022.05.003>
67. Akashi H, Gojobori T. 2002. Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proc Natl Acad Sci U S A* 99:3695–3700. <https://doi.org/10.1073/pnas.062526999>
68. Görke B, Stölke J. 2008. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat Rev Microbiol* 6:613–624. <https://doi.org/10.1038/nrmicro1932>
69. Galinier A, Haiech J, Kilhoffer M-C, Jaquinod M, Stölke J, Deutscher J, Martin-Verstraete I. 1997. The *Bacillus subtilis* crh gene encodes a HPr-like protein involved in carbon catabolite repression. *Proc Natl Acad Sci U S A* 94:8439–8444. <https://doi.org/10.1073/pnas.94.16.8439>
70. Galinier A, Kravanza M, Engelmann R, Hengstenberg W, Kilhoffer M-C, Deutscher J, Haiech J. 1998. New protein kinase and protein phosphatase families mediate signal transduction in bacterial catabolite repression. *Proc Natl Acad Sci U S A* 95:1823–1828. <https://doi.org/10.1073/pnas.95.4.1823>
71. Deutscher J, Küster E, Bergstedt U, Charrier V, Hillen W. 1995. Protein kinase-dependent HPr/CcpA interaction links glycolytic activity to carbon catabolite repression in Gram-positive bacteria. *Mol Microbiol* 15:1049–1053. <https://doi.org/10.1111/j.1365-2958.1995.tb02280.x>
72. Schumacher MA, Seidel G, Hillen W, Brennan RG. 2007. Structural mechanism for the fine-tuning of CcpA function by the small molecule effectors glucose 6-phosphate and fructose 1,6-bisphosphate. *J Mol Biol* 368:1042–1050. <https://doi.org/10.1016/j.jmb.2007.02.054>
73. Seidel G, Diel M, Fuchsauer N, Hillen W. 2005. Quantitative interdependence of effectors, CcpA and cre in carbon catabolite regulation

- of *Bacillus subtilis*. *FEBS J* 272:2566–2577. <https://doi.org/10.1111/j.1742-4658.2005.04682.x>
74. Singh KD, Schmalisch MH, Stölke J, Görke B. 2008. Carbon catabolite repression in *Bacillus subtilis*: quantitative analysis of repression exerted by different carbon sources. *J Bacteriol* 190:7275–7284. <https://doi.org/10.1128/JB.00848-08>
75. 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>
76. Pienkos PT, Zhang M. 2009. Role of pretreatment and conditioning processes on toxicity of lignocellulosic biomass hydrolysates. *Cellulose* 16:743–762. <https://doi.org/10.1007/s10570-009-9309-x>
77. Ujor VC, Okonkwo CC. 2022. Microbial detoxification of lignocellulosic biomass hydrolysates: biochemical and molecular aspects, challenges, exploits and future perspectives. *Front Bioeng Biotechnol* 10:1061667. <https://doi.org/10.3389/fbioe.2022.1061667>
78. Jönsson LJ, Martín C. 2016. Pretreatment of lignocellulose: formation of inhibitory by-products and strategies for minimizing their effects. *Bioresour Technol* 199:103–112. <https://doi.org/10.1016/j.biortech.2015.10.009>
79. Luo H, Zheng P, Bilal M, Xie F, Zeng Q, Zhu C, Yang R, Wang Z. 2020. Efficient bio-butanol production from lignocellulosic waste by elucidating the mechanisms of *Clostridium acetobutylicum* response to phenolic inhibitors. *Sci Total Environ* 710:136399. <https://doi.org/10.1016/j.scitotenv.2019.136399>
80. Jayakody LN, Turner TL, Yun EJ, Kong II, Liu J-J, Jin Y-S. 2018. Expression of Gre2p improves tolerance of engineered xylose-fermenting *Saccharomyces cerevisiae* to glycolaldehyde under xylose metabolism. *Appl Microbiol Biotechnol* 102:8121–8133. <https://doi.org/10.1007/s00253-018-9216-x>
81. Bonn G, Bobleter O. 1984. HPLC-analyses of plant biomass hydrolysis and fermentation solutions. *Chromatographia* 18:445–448. <https://doi.org/10.1007/BF02262975>
82. Chandrangsu P, Dusi R, Hamilton CJ, Helmann JD. 2014. Methylglyoxal resistance in *Bacillus subtilis*: contributions of bacillithiol-dependent and independent pathways. *Mol Microbiol* 91:706–715. <https://doi.org/10.1111/mmi.12489>
83. Lee J-W, Trinh CT. 2019. Microbial biosynthesis of lactate esters. *Biotechnol Biofuels* 12:226. <https://doi.org/10.1186/s13068-019-1563-z>
84. Kovács AT, van Hartskamp M, Kuipers OP, van Kranenburg R. 2010. Genetic tool development for a new host for biotechnology, the thermotolerant bacterium *Bacillus coagulans*. *Appl Environ Microbiol* 76:4085–4088. <https://doi.org/10.1128/AEM.03060-09>
85. Walker C, Dien B, Giannone RJ, Slininger P, Thompson SR, Trinh CT. 2021. Exploring proteomes of robust *Yarrowia lipolytica* isolates cultivated in biomass hydrolysate reveals key processes impacting mixed sugar utilization, lipid accumulation, and degradation. *mSystems* 6:e0044321. <https://doi.org/10.1128/mSystems.00443-21>
86. Walker C, Mortensen M, Poudel B, Cotter C, Myers R, Okekeogbu IO, Ryu S, Khomami B, Giannone RJ, Laursen S, Trinh CT. 2023. Proteomes reveal metabolic capabilities of *Yarrowia lipolytica* for biological upcycling of polyethylene into high-value chemicals. *mSystems* 8:e0074123. <https://doi.org/10.1128/msystems.00741-23>
87. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, et al. 2009. Real-time DNA sequencing from single polymerase molecules. *Science* 323:133–138. <https://doi.org/10.1126/science.1162986>
88. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 37:540–546. <https://doi.org/10.1038/s41587-019-0072-8>
89. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10:563–569. <https://doi.org/10.1038/nmeth.2474>
90. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>
91. Russel J, Pinilla-Redondo R, Mayo-Muñoz D, Shah SA, Sørensen SJ. 2020. CRISPRCasTyper: automated identification, annotation, and classification of CRISPR-Cas loci. *CRISPR J* 3:462–469. <https://doi.org/10.1089/crispr.2020.0059>
92. Galata V, Fehlmann T, Backes C, Keller A. 2019. PLSDB: a resource of complete bacterial plasmids. *Nucleic Acids Res* 47:D195–D202. <https://doi.org/10.1093/nar/gky1050>
93. Zhang R, Mirdita M, Levy Karin E, Norroy C, Galiez C, Söding J. 2021. SpacePHARER: sensitive identification of phages from CRISPR spacers in prokaryotic hosts. *Bioinformatics* 37:3364–3366. <https://doi.org/10.1093/bioinformatics/btab222>
94. Grant JR, Enns E, Marinier E, Mandal A, Herman EK, Chen C-Y, Graham M, Van Domselaar G, Stothard P. 2023. Proksee: in-depth characterization and visualization of bacterial genomes. *Nucleic Acids Res* 51:W484–W492. <https://doi.org/10.1093/nar/gkad326>
95. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Ciuffo S, Li W. 2013. Prokaryotic genome annotation pipeline, the NCBI handbook. 2nd ed. National Center for Biotechnology Information (US).
96. Treangen TJ, Ondov BD, Koren S, Phillippy AM. 2014. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol* 15:524. <https://doi.org/10.1186/s13059-014-0524-x>
97. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR. 2015. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* 43:e15. <https://doi.org/10.1093/nar/gku1196>
98. Polpitiya AD, Qian W-J, Jaitly N, Petyuk VA, Adkins JN, Camp DG, Anderson GA, Smith RD. 2008. DAnTE: a statistical tool for quantitative analysis of -omics data. *Bioinformatics* 24:1556–1558. <https://doi.org/10.1093/bioinformatics/btn217>
99. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J. 2016. The Perseus computational platform for comprehensive analysis of (proteo)omics data. *Nat Methods* 13:731–740. <https://doi.org/10.1038/nmeth.3901>
100. Zhang X, Smits AH, van Tilburg GB, Ovaai H, Huber W, Vermeulen M. 2018. Proteome-wide identification of ubiquitin interactions using UbIA-MS. *Nat Protoc* 13:530–550. <https://doi.org/10.1038/nprot.2017.147>