

ARTICLE



Functional and structural diversification of incomplete phosphotransferase system in cellulose-degrading clostridia

Tao Xu^{1,2,3,11}, Xuanyu Tao^{3,11}, Hongxi He^{4,5}, Megan L. Kempher³, Siping Zhang⁶, Xiaochun Liu^{4,5}, Jun Wang^{4,5}, Dongyu Wang⁷, Daliang Ning³, Chongle Pan^{7,8}, Honghua Ge^{4,5}, Nannan Zhang^{4,5}, Yong-Xing He⁶ and Jizhong Zhou^{3,9,10}

© The Author(s), under exclusive licence to International Society for Microbial Ecology 2023

Carbohydrate utilization is critical to microbial survival. The phosphotransferase system (PTS) is a well-documented microbial system with a prominent role in carbohydrate metabolism, which can transport carbohydrates through forming a phosphorylation cascade and regulate metabolism by protein phosphorylation or interactions in model strains. However, those PTS-mediated regulated mechanisms have been underexplored in non-model prokaryotes. Here, we performed massive genome mining for PTS components in nearly 15,000 prokaryotic genomes from 4,293 species and revealed a high prevalence of incomplete PTSs in prokaryotes with no association to microbial phylogeny. Among these incomplete PTS carriers, a group of lignocellulose degrading clostridia was identified to have lost PTS sugar transporters and carry a substitution of the conserved histidine residue in the core PTS component, HPr (histidine-phosphorylatable phosphocarrier). *Ruminoclostridium cellulolyticum* was then selected as a representative to interrogate the function of incomplete PTS components in carbohydrate metabolism. Inactivation of the HPr homolog reduced rather than increased carbohydrate utilization as previously indicated. In addition to regulating distinct transcriptional profiles, PTS associated CcpA (Catabolite Control Protein A) homologs diverged from previously described CcpA with varied metabolic relevance and distinct DNA binding motifs. Furthermore, the DNA binding of CcpA homologs is independent of HPr homolog, which is determined by structural changes at the interface of CcpA homologs, rather than in HPr homolog. These data concordantly support functional and structural diversification of PTS components in metabolic regulation and bring novel understanding of regulatory mechanisms of incomplete PTSs in cellulose-degrading clostridia.

The ISME Journal (2023) 17:823–835; <https://doi.org/10.1038/s41396-023-01392-2>

INTRODUCTION

Microbial survival is largely determined by the ability to adapt to fluctuating nutrient supplies in natural environments. Carbohydrate utilization is one of the most important nutrient acquisition functions to microorganisms. The phosphoenolpyruvate (PEP):carbohydrate phosphotransferase system (PTS) is involved in the uptake of diverse carbohydrates and regulation of microbial metabolism, which has been extensively characterized in a few model microorganisms [1–4]. Its function has been expanded to nitrogen and phosphate metabolisms [5], antibiotic resistance [6, 7], biofilm formation [8, 9], expression of virulence genes [10, 11], endotoxin production [12], and regulation of quorum sensing [13]. Therefore, the PTS is important to biogeochemical carbon cycling, ecological homeostasis, and health status of higher life forms.

The complete PTS has three core components [2], including Enzyme I (EI), histidine-phosphorylatable phosphocarrier protein (HPr), and Enzyme II complex (EII) made up of EIIA, EIIB, and EIIC/D.

To transport sugars, the high-energy phosphoryl group donated by PEP is transferred from EI to HPr, then to EIIA/B, and finally to sugars during transport by EIIC [2, 4, 14]. To regulate carbon metabolism, the core PTS components can either transfer the high-energy phosphoryl group to downstream target proteins or interact with target proteins to form a complex, which results in activation/inactivation of those target proteins. For example, LevR in *Bacillus subtilis* can be activated by HPr (P~15His-HPr) through phosphorylation to regulate levan degradation [15] and the P-46Ser-HPr phosphorylated by HPr kinase/phosphorylase (HPrK) can biochemically interact with a LacI-family repressor, catabolite control protein A (CcpA), to cause carbon catabolite repression (CCR) in *B. subtilis* [16, 17]. However, these canonical mechanisms have only been characterized in a small set of model bacteria (e.g., *B. subtilis*) which all contain a complete set of the three core PTS components. Previous studies have found that many proteobacteria do not have EIIB or EIIC components, which might affect PTS-mediated carbohydrate metabolism [2, 3, 18]. With the

¹Section on Pathophysiology and Molecular Pharmacology, Joslin Diabetes Center, Boston, MA, USA. ²Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA, USA. ³Institute for Environmental Genomics, and Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK, USA. ⁴School of Life Sciences, Anhui University, Hefei 230601, PR China. ⁵Institutes of Material Science and Information Technology, Anhui University, Hefei 230601, PR China. ⁶Ministry of Education Key Laboratory of Cell Activities and Stress Adaptations, School of Life Sciences, Lanzhou University, Lanzhou 730000, PR China. ⁷Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK, USA. ⁸School of computer science, University of Oklahoma, Norman, OK, USA. ⁹School of Civil Engineering and Environmental Sciences, University of Oklahoma, Norman, OK, USA. ¹⁰Earth and Environmental Sciences, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. ¹¹These authors contributed equally: Tao Xu, Xuanyu Tao. ✉email: zhangnn@ahu.edu.cn; heyx@lzu.edu.cn; jzhou@ou.edu

Received: 22 September 2022 Revised: 21 February 2023 Accepted: 23 February 2023
Published online: 10 March 2023

advancements of microbial genome sequencing and extensive culturing studies of environmental and host-associated microorganisms [19, 20], it raises questions on the overall prevalence of incomplete PTSs in microbes and carbon regulatory mechanisms for those incomplete PTSs [2, 21].

Carbohydrate availability either in natural or artificial biosystems, such as gastrointestinal tracts [22–24] or microbial consortia-based bioprocessing systems [25], determines microbial diversity and functional outcomes. Understanding carbon utilization mechanisms would facilitate microbe-specific or microbiome-oriented modulation towards improved functions [24]. To do so, we made use of a large number of fully assembled prokaryotic genomes and performed computational and experimental investigation into PTS components. Distinct from canonical PTSs in bacteria [2–4], here we describe the architecture and distribution of incomplete PTSs with loss of conserved components through massive genome mining in prokaryotic genomes, and revealed novel functional diversification of incomplete PTS components by integrated multi-omics analyzes in a panel of PTS mutants, and solved the crystal structure of key components that are pivotal to carbohydrate metabolism in incomplete PTSs.

RESULTS

High prevalence and phylogeny-independent distribution of incomplete PTSs in prokaryotes

To systematically understand the distribution and configuration of PTSs, we performed genome mining for PTS core and associated components in 14,998 sequenced prokaryotic genomes from 4,293 species classified by GTDB-Tk [26], including archaea and bacteria (Fig. 1A). In total, 23 HMM profiles of PTS relevant components (Table S1), including HPr, EI, EIIA, EIIB, EIIC/D, HPrK, and LacI-family regulators, were used to search for homologs in each genome through hmm search against our inhouse database containing 56,869,107 proteins built from these genomes. HPr, EI, EIIA, EIIB, EIIC/D are considered as core components of PTS, and HPrK and LacI regulators are considered as associated components of PTS. Homologs are identified to be a single or fused architecture depending on the number of modules predicted. We then cataloged the copy number of each hmm module, the occurrence of each PTS component, and the completeness of a PTS in each genome along with corresponding phylogenetic information. PTSs consisting of at least one set of core components (EI, HPr, EIIA, EIIB, and EIIC/D) with singular architecture are considered as complete PTS, with the potential of direct sugar transport; in contrast, those lacking at least one of above core components are named incomplete PTS presumably incapable of sugar transport [2, 4].

At the genome level, we observed 3,164 prokaryotic species have an incomplete PTS due to the lack of one or more PTS core components (Table S1), accounting for 73.70% of species analyzed in this study. In particular, 981 species do not have an HPr homolog (Fig. 1B), 840 species without EI, 3,008 species without a complete EIABC complex (Figure S1A). It is notable that 5 of 243 archaeal species carry all core components of PTS (HPr, EI, EIIA, EIIB, and EIIC) that are required for sugar transport. Among them, *Haloferax mediterranei* [27] and *Haloferax volcanii* DS2 [28, 29], have been demonstrated to use the PTS to transport fructose. Moreover, extensively sequenced genomes of some bacterial species allowed us to dissect inter-strain variation of the configuration of PTS, which has been previously overlooked. For example, among 27 *B. subtilis* strains, 12 strains have all PTS components whereas the others are missing EI or EIIC. Inter-strain variations were also observed in *Escherichia coli* and *Clostridioides difficile* (Table S1). All these observations revealed novel insights into the unrecognized prevalence of incomplete PTSs in prokaryotes and the possibility of inter-strain variations in PTS configuration.

We reassessed the distribution of each PTS core or associated component to microbial phylogeny (Fig. 1C and Table S1). As EIIA, EIIB, and EIIC often form a protein complex that interacts with sugars that will be transported [2, 4, 30–32], we evaluated and grouped the overall completeness of EIABC into three categories (absent, partial or complete). In general, PTS components have little shared patterns for their occurrence across the phylogenetic tree. For example, most HPr-carrying species, especially in members of *Chlamydia*, *Actinobacteria* and *Gamma-proteobacteria*, lack HprK homologs; species with EI may not have the potential to assemble a complete EIABC complex. One exception is that HPr and EIABC completeness tend to have a similar distribution pattern (Fig. S1B). The Jaccard correlation analysis that only distinguishes the occurrence of the PTS components rather than the copy number in a given genome, revealed that HPr and EIIA have a high correlation coefficient (Figs. 1D; S1C), indicating they are likely to co-occur in the genome. Both HPr and EI, which are usually encoded by the *ptsHI* operon in several well-studied bacteria [2], have very low correlation in their occurrence in this large dataset. Although LacI-family regulators, such as CcpA, were reported to interact with HPr to regulate catabolic genes in Firmicutes [2, 4, 16, 17], we observed a large variation in the copy number of LacI-family members across the phylogeny but there is no clear association with the presence of HPr (Fig. 1D). These observations elicit new questions about the evolution and functional relevance of PTS components.

The presence of each PTS core or associated component has no clear distribution pattern across the phylogeny of its origin (Figs. 1C; S1D). For example, in the phylum of *Firmicutes*, HprK mainly exists in bacilli and clostridia rather than negativicutes. This variable inter-class distribution was also observed in *Proteobacteria*, where HprK is mostly absent in *Gamma-Proteobacteria* and *deltaproteobacteria* but present in *betaproteobacteria*. It is also notable that nearly all PTS components, except EI, are missing in *Campylobacterota* and a few clusters of species in other Classes of bacteria, including *Bacteroidia*, *Actinobacteria*, *Clostridia*, and *Gamma-Proteobacteria*; however, these absences are distributed with no clear patterns across the tree. The large compositional deficiency in *Campylobacterota* is consistent with its most likely autotrophic lifestyle [33]. This unpredictable distribution suggests the overall evolutionary relationship cannot reflect presence or absence of PTS components.

Compositional and motif-level features confine the potential role of PTSs in sugar transport

The genetic and molecular features of PTS components were further investigated to determine sugar transport. EI and EII components vary in the copy number of their homologs and their protein architecture (Fig. 1E). For instance, EI_PEP_utilisers_N, and EIIB_GUT_C have a low copy number in genomes and are often fused with other modules to form multi-modular proteins; PTS_EIIC have multiple copies and are also mainly present in a fused architecture. Compared to other components, EI most widely exists in both archaea and bacteria, except flavobacteria, cytophagia, and sphingobacteria in the phylum of *Bacteroidetes*. By examining the distribution of EIIC transporters (Fig. S1D), we found that *Actinobacteria*, *Clostridia*, *Bacilli*, *Negativicutes*, *Gamma-Proteobacteria*, *Alpha-Proteobacteria*, and *Fusobacteria* tend to have more diverse sugar transporters than other classes, which may indicate their genetic potential of transporting and utilizing different carbohydrates.

In canonical PTSs, the H15 in HPr is an indispensable amino acid for transferring the high-energy phosphate moiety from EI to the sugar transporting EIABC complex or other regulatory proteins [6, 16, 17]. Analysis of HPr homologs in 4293 prokaryotic species revealed a very conserved motif sequence around H15 (Fig. 1F). However, 10% of HPr homologs have H15 substituted by other

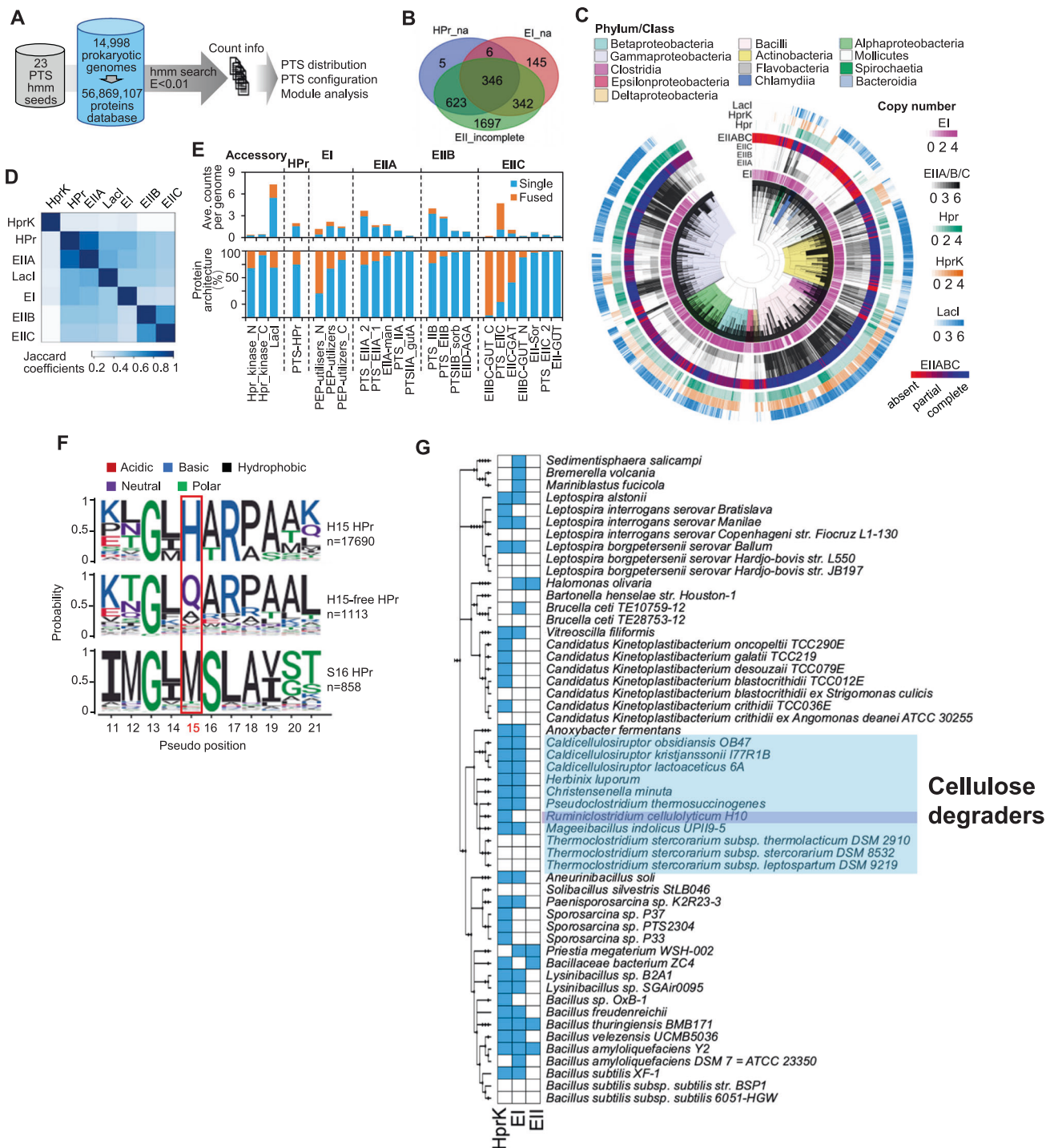


Fig. 1 Architectural and compositional diversification of phosphotransferase systems (PTS) in prokaryotes. **A** Schematic workflow for genome mining of PTS systems in 14,998 complete genomes of 4,293 species, including archaea and bacteria. **B** Venn diagram showing the number of microbial species that miss different components (na). Among 3164 species with incomplete PTS systems, 3008 species have an incomplete EII complex for sugar transport due to the absence of either EIIA, EIIB, or EIIC. **C** The distribution of PTS relevant components across the microbial phylogenetic tree. The copy number or completeness of PTS components are shown in different tracks. **D** Heatmap of Jaccard coefficients to correlate the co-occurrence of PTS components. **E** The averaged copy number (upper) and proportion (lower) of single or fused PTS components in prokaryotic genomes. **F** The probability of amino acids at positions 11–21 in HPr homologs. All sequences were aligned and split to three categories, H15-carrying HPr, H15-free HPr, and S16-carrying HPr. **G** The phylogenetic tree of microbial species that only carry H15-free HPr homologs and presumably lost the function of PTS-mediated sugar transport. The presence of HPrK and EI and the completeness of EIIABC (EII) were labeled blue for each species. A group of species in clostridia have been reported to be cellulose degraders. The genetically tractable *R. cellulolyticum* H10 was selected as a representative to dissect the function of non-canonical PTS.

amino acids, such as Q, and 85% of these come from 53 species with an incomplete PTS (Fig. 1G). Most HPr homologs lacking H15 have a phosphorylatable Serine at site 16 and presents a very distinct motif sequence (Fig. 1F). It indicates the substitution of H15 may disable PTS-controlled sugar transport and affect the regulatory role mediated by H15 phosphorylation. It is notable that a cluster of H15-free HPr carrying species in clostridia have been shown to degrade cellulose (Fig. 1G), including a model organism of mesophilic cellulolytic clostridia, *Ruminiclostridium cellulolyticum* [34]. It is known that lignocellulose biodegradation is essential to microbial homeostasis in animal guts and carbon recycling in the global ecosystem [35, 36]. These results raise questions regarding the role of the incomplete PTS in carbohydrate utilization, for instance, microbial degradation of plant fibers, and the mechanism of catabolite regulation in these heterotrophic microbes.

Novel metabolic profiles of *R. cellulolyticum* dependent of the mini “PTS” components

R. cellulolyticum is a natural producer of lignocellulosic biofuels and chemicals via consolidated bioprocessing [34]. It carries an incomplete PTS with the substitution of H15 in HPr and the absence of EI, EIIA, EIIB, and EIIC (Fig. 2A). We named this incomplete PTS a mini PTS and sought to understand its regulatory role in carbohydrate utilization. Most microbes exhibit CCR—preferring glucose over other sugars with a diauxic growth when multiple sugars are available [16]. There are three major mechanisms of CCR characterized in bacteria, including PTS-mediated cAMP-CRP mechanism in enteric bacteria, RNA-binding protein mediated mechanism in pseudomonads, PTS-mediated CcpA/HprK mechanism in Firmicutes [16]. When grown on a

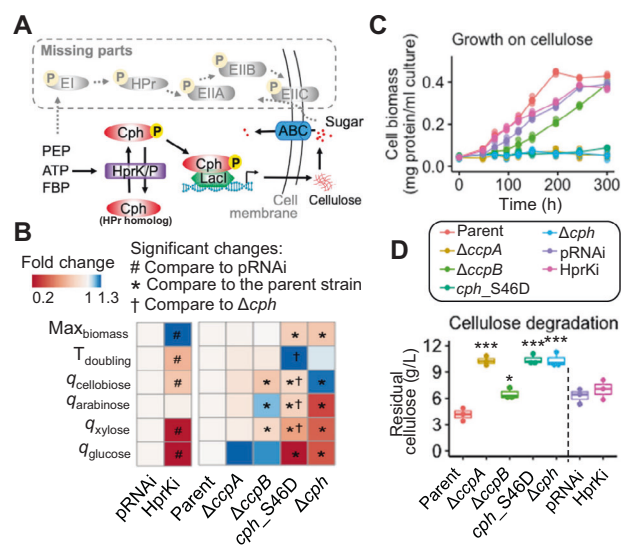


Fig. 2 The mini “PTS” of *R. cellulolyticum* is essential for microbial metabolism. **A** A hypothetical model of non-canonical PTS in absence of sugar transport components in *R. cellulolyticum* H10. **B** Comparison of physiological and catabolic features, including the maximal cell biomass (Max_{biomass}), the doubling time of growth (T_{doubling}), the utilization efficiency of cellobiose (q_{cellobiose}), arabinose (q_{arabinose}), xylose (q_{xylose}), and glucose (q_{glucose}), across mutants and control strains, when grown on a carbon mix consisting of cellobiose, arabinose, xylose, and glucose. Data represent the fold change of mean values between mutants and controls indicated in the plot, those of which present statistical significances are marked. **C** Growth of bacterial strains on 10 g/L Avicel cellulose (n = 3). **D** Determination of residual cellulose after 300 h fermentation with bacterial strains on 10 g/L Avicel cellulose (n = 3). Unpaired two-tailed t test, *p < 0.05, ***p < 0.001. Data (C and D) are represented as mean ± sd.

carbon mixture, *R. cellulolyticum* utilized cellobiose, glucose, and pentoses simultaneously (Figure S2A), at a >3-fold higher consumption rate for soluble sugars other than glucose (Fig. S2A, S2B; q_{cellobiose} = 27.7, q_{xylose} = 27.6, q_{arabinose} = 19.3, q_{glucose} = 5.8 mmol/g/h). Compared to the utilization of singular sugars, the presence of glucose increased, rather than decreased, the consumption rate of cellobiose and xylose (Fig. S2C, S2D). Similar reverse catabolite repression—preferring other organics over glucose, has been described in a few bacterial species, such as probiotic *Streptococcus thermophilus* [37] *Bifidobacterium longum* [38] and pathogenic *Pseudomonas aeruginosa/putida* [39, 40].

We performed genetic manipulation to interrogate the role of mini PTS components in *R. cellulolyticum* (Fig. S2E). We applied Cas9n-based genome editing [41] to precisely inactivate the *hpr* ortholog (named hereafter catabolite promotion HPr, or *cph*, the mutant of which is Δ*cph*) and two CcpA-like transcriptional regulators, *ccpA* (*ccpA* ortholog; Δ*ccpA*) and *ccpB* (*ccpA* paralog; Δ*ccpB*), and generate a site-specific *cph* mutant with the serine at 46 substituted by aspartate to mimic the phosphorylated form (*cph*_S46D). Since a *hprK* knockout was not successfully obtained, we knocked down the activity of *hprK* by antisense RNA molecules (*hprKi*). The functional relevance of each PTS component was evaluated by comparing carbohydrate utilization and cell growth of these mutants on different carbon sources. Detailed interpretation to phenotypic changes is shown in the supplementary text (Supplementary text A). On soluble carbon mix (CM, consisting of cellobiose, glucose, arabinose, and xylose), which mimics the co-presence of diverse sugars released during lignocellulose biodegradation, we observed that *cph* knockout and *cph*_S46D reduced the consumption rate of most sugars (Figs. 2B, S2B). The positive role of S46-carrying Cph in using simple sugars simultaneously is opposite to catabolite repression as described in model microbes [16, 21]. Since HPr proteins are cognate targets of HprK for phosphorylation [42], it is consistent that *hprK* knockdown also decreases sugar utilization (Figs. 2B, S2B). For the *ccpA*-like genes, loss of *ccpB* rather than *ccpA* affected the consumption of soluble sugars, suggesting functional divergence of these two LacI regulators in *R. cellulolyticum*. On cellulose, we observed that loss of *cph* and *ccpA*, as well as *cph*_S46D, abolished bacterial growth (Fig. 2C); *ccpB* disruption had both a reduced growth rate and cellulose degradation (Fig. 2C, D), whereas *hprK* knockdown had no significant impact on cellulose degradation at final time point. The indispensable role for Cph and CcpA in cellulose metabolism, instead of catabolite repression, has not been reported before. Compared to canonical PTS, these metabolic profiles together demonstrated unique characteristics of this non-canonical incomplete PTS, such as Cph-dependent catabolite promotion and functional specialization of CcpA on cellulose.

Functional diversification of the mini PTS components at the transcriptional, biochemical and structural level

To systematically understand the mechanism of how this mini PTS regulates catabolite metabolism, we investigated into transcriptomic changes in PTS mutants, examined the functional relevance and interactions of PTS components by correlating transcriptomic, DNA-binding, and protein-binding profiles, and finally performed protein crystallization for key players of this incomplete PTS to decipher the molecular-level mechanism.

Cph, *CcpA*, and *CcpB* mediates distinct transcriptomic profiles. On carbon mix, Δ*cph* had broad transcriptional changes in 286 genes that were enriched for diverse basic cellular functions and carbohydrate metabolism (Table S2); whereas *cph*_S46D presented very little impact on transcriptome (Fig. 3B). These results suggest that *R. cellulolyticum* H15-free Cph is associated with pleiotropic functions, not limited to carbohydrate utilization, which largely depend on the phosphorylation of Cph at S46.

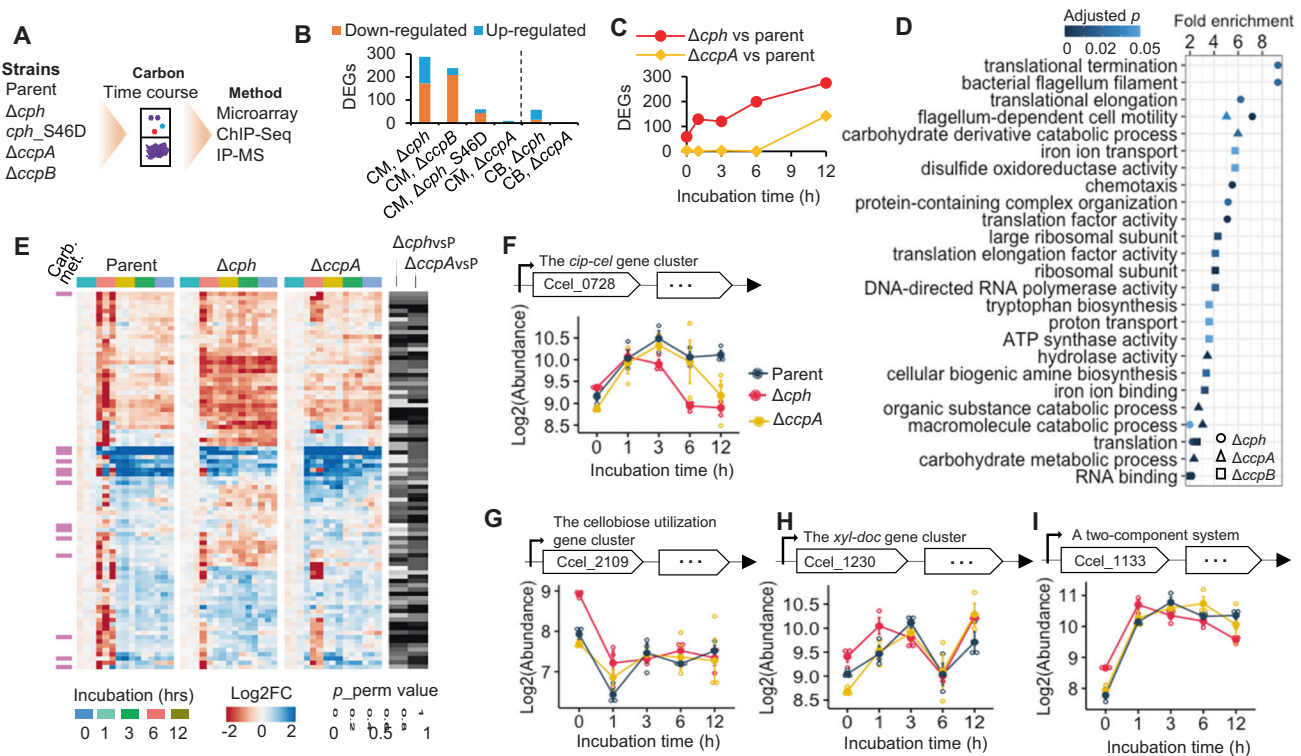


Fig. 3 Distinct roles of the mini “PTS” components in carbohydrate dependent transcriptional regulation. **A** Experimental design for functional characterization of the mini “PTS” genes in *R. cellulolyticum* H10 by genetic disruption, physiological and transcriptomic profiling under different carbon sources, as well as ChIP-Seq and IP-MS. **B** The number of DEGs across mutants grown with either cellobiose (CB) or carbon mix (CM). All mutants were compared to the parental strain to count DEGs ($p < 0.01$ and $\log_2 |FC| > 2$, $n = 3$). **C** Changes in the number of DEGs in mutants versus the parental strain at 0, 1, 3, 6, 12 h of cellulose incubation experiments ($n = 3$). **D** GO enrichment analysis of DEGs in Δcph , $\Delta ccpA$, and $\Delta ccpB$. Δcph , $\Delta ccpA$, and the parent strain after 12 h cellulose incubation were compared. $\Delta ccpB$ and the parental strain at the mid log phase of growth on cellulose were compared. **E** Heatmap of transcriptional abundance of select genes in mutants during cellulose incubation. The transcript abundance in each genetic background was normalized to the mean of non-incubated (0 h) conditions. Select genes were shared DEGs in Δcph vs parent and $\Delta ccpA$ vs parent after 12 h cellulose incubation. The p_{perm} values indicate whether gene-specific response patterns between mutants and parental strains on cellulose are significantly similar. Genes associated carbohydrate transport and metabolism (Carb. met.) are annotated. Transcriptomic responses of select genes in key gene clusters related to cellulose degradation (**F**), cellobiose catabolism (**G**), xylan degradation (**H**), and a cellulose-responsive TCS (**I**). Data represents mean $\pm$ se ($n = 3$).

$\Delta ccpB$ presented 239 differentially expressed genes (DEGs) on carbon mix versus 10 DEGs in $\Delta ccpA$ (Fig. 3B; Table S2). DEGs in $\Delta ccpB$ were involved in a variety of basic biological processes (Table S2), consistent with its wide impact on the consumption of soluble sugars (Figs. 2, S2D). As $\Delta ccpA$ had a limited impact on transcriptome (Fig. 3B), irrespective of soluble sugars, *R. cellulolyticum* CcpB is more functionally similar to reported CcpA in *B. subtilis* and *Staphylococcus mutans* [43, 44]. It is known that HPr interacts with CcpA and thereby modulates the DNA binding affinity of CcpA [2, 16, 17, 45, 46]. Among more than 200 DEGs in Δcph and $\Delta ccpB$, respectively (Table S2), only 71 genes were shared and enriched for some biological processes (e.g., mono-carboxylic acid biosynthetic process, ATPase-coupled transmembrane transporter activity and quorum sensing) (Fig. S3B). Plus differential catabolic and phenotypic features as shown above (Fig. 2B), it seems that Cph and CcpB regulate transcription mainly through independent pathways.

Considering the essentiality of Cph and CcpA on cellulose, we performed incubation experiments in which cellobiose grown mutants were incubated in cellulose medium for longitudinal profiling of transcriptomes (Fig. S3C). Δcph and $\Delta ccpA$ quickly responded to cellulose (Fig. S3A) but exhibited a unique set of DEGs at distinct emerging rates across time points (Fig. 3C). In terms of GO enrichment, $\Delta ccpA$ are highly associated with carbohydrate/cellulose metabolism (Fig. 3D); whereas $\Delta ccpB$ and Δcph have broad impacts on basic biological processes and

molecular functions. These results further support functional divergence of CcpA and CcpB, where CcpA exclusively regulates cellulose metabolism. Analysis of the dynamic response of 88 shared DEGs between Δcph and $\Delta ccpA$ (Table S2 and Figure S3B) revealed that most DEGs were quickly suppressed but then restored or even increased in $\Delta ccpA$ and the parent strain, whereas strong suppression persisted in Δcph (Fig. 3E). We observed that the sole promoter-driven *cip-cel* gene cluster, which is indispensable for cellulose metabolism [47–49], was able to quickly respond to cellulose incubation in Δcph , $\Delta ccpA$, and the parent strain (Fig. 3F). However, the enhancement decayed soon in both mutants. Loss of sustainable transcriptional enhancement in the *cip-cel* gene cluster could lead to a lower expression of pivotal cellulose degrading enzymes for the formation of cellulosomal complexes to hydrolyze cellulose efficiently. Other gene clusters involved in cellobiose utilization (Fig. 3G), xylan degradation (Fig. 3H), and a cellulose-responsive TCS (Fig. 3I), did not show substantial differences across these strains in response to cellulose. Therefore, Cph and CcpA are important to sustain a long-term high expression of pivotal cellulose degrading genes, rather than responding to cellulose.

CcpA and CcpB from R. cellulolyticum diverged from reported CcpAs with distinct DNA binding motifs. We then performed ChIP-seq and biochemical assays to understand the regulatory mechanisms of CcpA and CcpB. ChIP-seq analysis revealed that both regulators

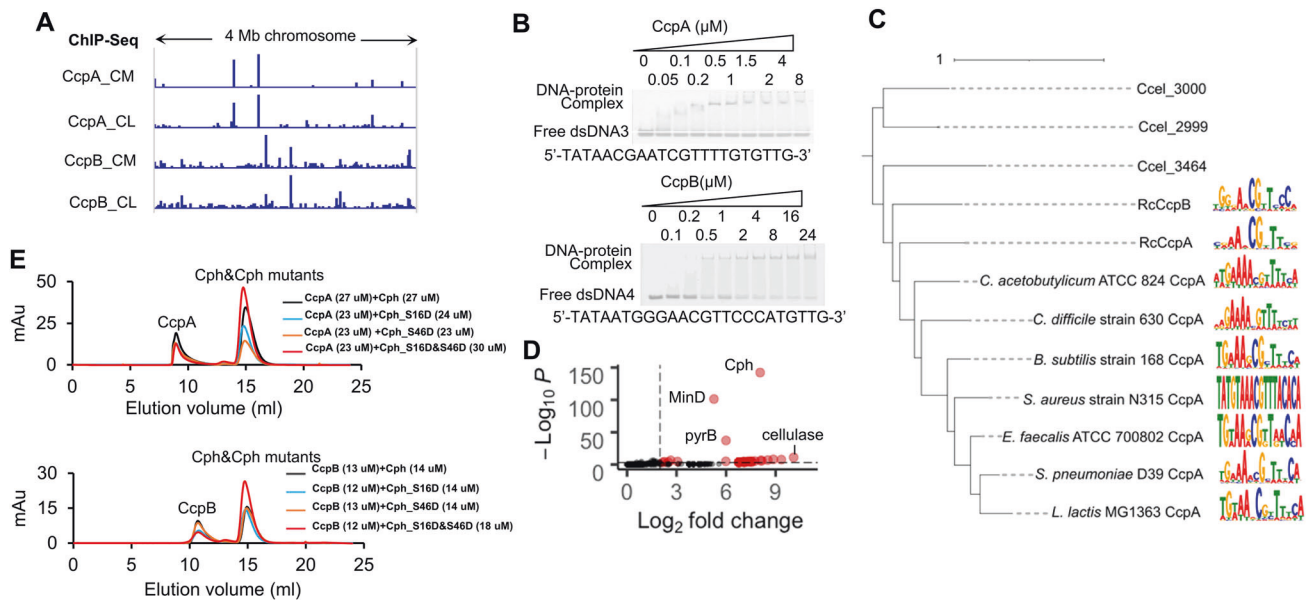


Fig. 4 Distinct DNA binding profiles of *R. cellulolyticum* CcpA homologs that are independent of Cph and its phosphorylation status. **A** Genome-wide binding sites of CcpA and CcpB in corresponding complementation strains ($\Delta ccpA$ with pEr-CcpA-FLAG and $\Delta ccpB$ with pEr-CcpB-FLAG) grown on carbon mix (CM) or cellulose (CL). Each condition has two biological replicates for chromatin immunoprecipitation sequencing (ChIP-Seq), which were combined for binding site calling (fold change > 2, p adj < 0.05). **B** Electrophoretic mobility shift assays with purified proteins and select DNA sequences identified by ChIP-Seq. **C** Phylogeny of previously characterized CcpA proteins and CcpA-like homologs in *R. cellulolyticum* (RcCcpA and RcCcpB). Their DNA recognition profiles are shown on the right. **D** Interactors of Cph identified by immunoprecipitation-mass spectrometry (IP-MS). The significant bindings were indicated by the fold change of identified abundance (red, FC > 2, p adj < 0.0001) between the complementation (Δcph with pEr-0806) and control (Δcph with pEr-empty) strains. **E** Examination of protein-protein interactions by gel filtration. Purified CcpA (upper) and CcpB (lower) were incubated with Cph (black), Cph_S16D (blue), Cph_S46D (orange), and Cph_S16&46D (red).

could not bind to the upstream of the *cip-cel* operon, indicating PTS-mediated regulation of cellulose degradation in *R. cellulolyticum* is not through direct CcpA binding of the canonical CCR mechanism [49, 50] (Supplementary text B). They also had distinct binding spectra across the genome (Fig. 4A; Table S3), irrespective of carbon sources. By integrating transcriptomic and ChIP-seq data, we found that CcpA may directly regulate itself and a neighbor gene (Ccel_1006, a FGGY-family carbohydrate kinase), by which cellulose degradation could be regulated indirectly (Supplementary text B). In comparison, there are more binding sites identified in CcpB and more genes potentially directly regulated by CcpB. They were significantly enriched for transporter activity when grown on carbon mixture, and enriched for carbohydrate metabolic and amino acid biosynthetic processes when grown on cellulose (Fig. S4A and Table S3), which are consistent with observed metabolic phenotypes in $\Delta ccpB$ - an increase in arabinose consumption rate when grown on carbon mixture, and a reduced growth rate and cellulose degradation when grown on cellulose (Fig. 2). Changing the carbon source from cellobiose to cellulose slightly broadened the binding of CcpA and CcpB (Figs. 4A, S4A). To validate the binding, we purified His-tagged CcpA and CcpB and performed electrophoretic mobility shift assays (Fig. 4B, Figs. S4B, C). Results showed that CcpA and CcpB efficiently bound corresponding DNA fragments at 0.2 and 0.4 μ M, respectively. MEME-ChIP analysis indicated that binding sites of CcpA and CcpB have different palindromic consensus motifs mostly differing in terminal nucleic acids (Fig. 4C), where CcpA tends to have 5'-GGN but CcpB prefers 5'-CGA. Compared to previously characterized CcpA proteins, CcpA and CcpB from *R. cellulolyticum* have diverged and have distinct binding motifs from those previously described.

Phosphorylatable Cph does not interact with either CcpA or CcpB. As HprK modulates the phosphorylation status of HPr

and our genetic studies have demonstrated the involvement of HprK, Cph, and *cph_S46D* in carbohydrate metabolism (Fig. 2), we investigated the biochemical properties of Cph during transcriptional and post-translational regulation. By performing immunoprecipitation-mass spectrometry (IP-MS), we found both Ser16 and Ser46 of Cph were phosphorylated, with phosphorylation ratios of 28% and 38%, respectively (Table S4). To confirm if both sites could be phosphorylated by HprK, in vitro incubation assays were carried out using purified HprK, Cph, Cph_S46D, and Cph_S16A/S46A mutant proteins, with supplemented ATP. Most Cph proteins were phosphorylated by HprK whereas the two mutant proteins were not (Fig. S4D), indicating that HprK only targets the Ser46 site of Cph. The in vitro phosphorylated Cph was also subjected to MS/MS analysis and ~93% Ser46 and only ~1% Ser16 was phosphorylated, further confirming that HprK specifically targets Ser46 rather than Ser16 (Table S4). Apart from using ATP as the phosphate donor, Cph phosphorylation at S46 also occurred when pyrophosphate (PPi) and fructose-1,6-bisphosphate (FBP) were co-provided for HprK (Fig. S4D and Table S4), which is different from using either ATP when FBP is co-present or just PPi to phosphorylate HPr in *B. subtilis* [42]. Therefore, Cph in *R. cellulolyticum* can be phosphorylated by HprK upon sensing a high energy status (Supplementary text C).

To explore potential interactors of Cph, we performed IP-MS analysis and identified 60 proteins that were significantly enriched (FDR-adjusted p < 0.05 and fold change > 2, Fig. 4D). These proteins have diverse functions, including pyrimidine ribonucleotide biosynthetic process, cellulose catabolic process, signal transduction, and ribosome biogenesis (Table S4 and Supplementary text B). Although HPr (P-Ser46) interacts with CcpA in the canonical PTS in Firmicutes [2, 4, 16, 17], both CcpA and CcpB were identified to have no interaction with Cph (Table S4). Undetectable interactions between unphosphorylated/phosphorylated Cph and CcpA or CcpB were further supported by

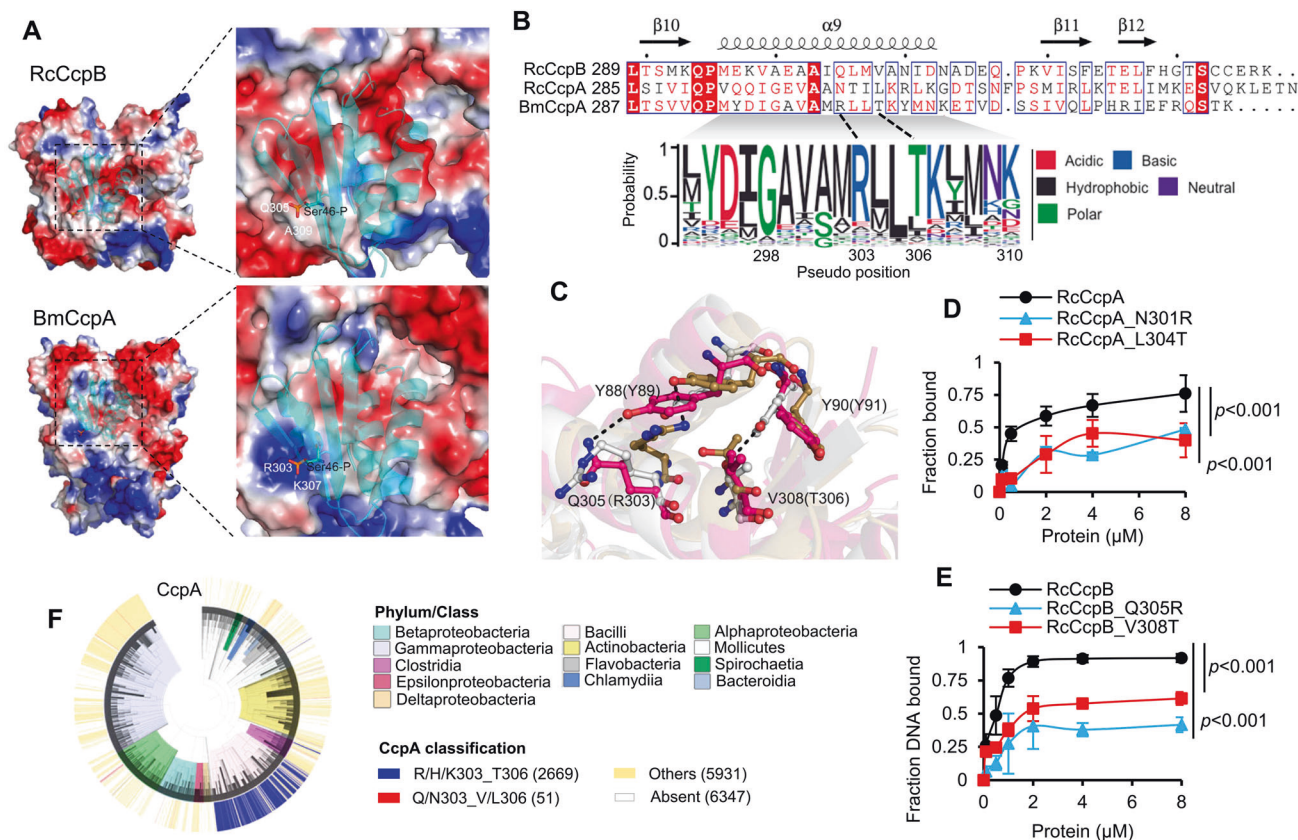


Fig. 5 Structural basis of Cph-independent regulation by CcpA and CcpB in *R. cellulosilyticum*. **A** Surface potential of the CcpB dimer from *R. cellulosilyticum* (RcCcpB without the DNA binding domain) and CcpA from *B. megaterium* (BmCcpA). The RcCcpB dimer was superimposed to the BmCcpA dimer of BmCcpA-HprSer46P complex and the structure of Hpr and Cph was shown in a cyan cartoon representation, with the Ser46P shown in sticks and labeled. **B** Multiple sequence alignment showing residues involved in conformational switching are different between BmCcpA and *R. cellulosilyticum* CcpA homologs (RcCcpA and RcCcpB). The sequence logo of the conserved segment (294–310) in CcpA homologs ($n = 8651$) depicts the usage frequency of amino acids, in reference to the relative position in BmCcpA. **C** Hydrogen bonding network essential for CcpA conformational switching in *B. megaterium* apo-CcpA (white), CcpA-HprSer46P complex (yellow) and *R. cellulosilyticum* apo-CcpB (magenta). The residue labels for BmCcpA were shown in parentheses. Effect of site-specific mutations in RcCcpA (**D**) and RcCcpB (**E**) on DNA binding ($n = 2-3$ EMSA gels). The significance was determined by ANOVA. **F** Distribution of 8651 CcpA orthologs predicted in 14,998 bacterial genomes. The classic HPr-dependent CcpA orthologs (with R/H/K at 303 and T at 306) and RcCcpA/RcCcpB-like orthologs (with Q/N at 303 and V/L at 306) were colored in blue and red, respectively.

isothermal titration calorimetry (Fig. S4E), gel filtration (Fig. 4E), and IP-MS analyses of CcpA and CcpB (Table S4). These biochemical observations together support the unique and independent functionalization of these miniPTS components in transcriptional regulation and microbial metabolism.

Structural basis of Cph-independent DNA binding for CcpA and CcpB. To unravel the Cph-independent DNA binding mechanism of CcpA homologs in *R. cellulosilyticum*, we solved the crystal structures of CcpB (Fig. S5A) and Cph (Fig. S5B). CcpA crystallization was not successful. The 2.2 Å and 2.4 Å resolution structures of CcpB and Cph allowed structural comparisons with previously characterized apo-CcpA and HPr-46P-CcpA-DNA complex from *Bacillus megaterium* (Fig. 5). *R. cellulosilyticum* CcpB adopted a dimeric structure, with the N-terminal DNA binding domain (1–58) missing (Fig. S5C). The absence of DNA binding domains also occurred in apo form structures of other LacI regulators [17, 51, 52], probably due to the highly dynamic nature of DNA binding domains of the LacI regulators in the absence of bound DNA. The CcpB dimer was optimally superimposed onto the *B. megaterium* CcpA (BmCcpA) dimer with a root-mean-square deviation (RMSD) of 2.7 Å for 449 corresponding Ca atoms. However, the conserved residues in BmCcpA that are involved in the binding of classic HPr (eg. R303, K307 and D296) were mostly

changed in CcpB (Fig. 5A). Most prominently, the R303 and K307 residues in BmCcpA, which are essential to interact with phosphorylated S46 in HPr [17], were replaced by Q305 and A309 in RcCcpB (Fig. 5A). A structural model of RcCcpA was also predicted using AlphaFold2 and there was a similar change at the interface, where the basic R303 in BmCcpA was substituted by the neutral Gln residue (Fig. S5C, S5D). As a result, the surface potential of RcCcpB/RcCcpA is mostly neutral at the HPr-46P binding interface, rather than a basic interface in BmCcpA (Figs. 5A, 5B, S5D). Therefore, RcCcpB/RcCcpA presented a distinct structural interface to interact with potential ligands (Figs. 5A, 5B, S5D). Comparison of the DNA binding domains of RcCcpA with BmCcpA revealed residues I16–N24 constitute the DNA-recognizing helix, which is well conserved in CcpA orthologous proteins (Fig. S5C). On the other hand, overlay of RcCph onto BmHPr-46P revealed very small structural variations (RMSD of 1.2 Å for 66 Ca atoms) and similar residues at the CcpA binding interface (Fig. S5B). These structural comparisons suggested that the neutrally charged interface on RcCcpB and RcCcpB would disengage the binding of negatively charged RcCph-Ser46-P and supports our in vitro biochemical results that RcCcpA/RcCcpB do not interact with phosphorylated Cph in *R. cellulosilyticum*.

We then compared the structure of RcCcpB/RcCcpA with apo-BmCcpA and CcpA-HPr-Ser46-P-DNA ternary complex, which

represent DNA-free and DNA-bound conformations, respectively (Fig. 5C, S5E). Two pairs of hydrogen bonds, namely Y89-R303 and Y91-T306, which were uniquely formed in each of the two distinct conformations, could serve as the potential allosteric switch of BmCcpA. In the DNA-bound conformation, HPr-Ser46-P drew the sidechain of R303 close to Y89, resulting in the formation of a hydrogen bond between Y89 and R303, and this conformational change propagates to Y91, leading to the disruption of the hydrogen bond formed between Y91 and T306 [17]. In the apo-RcCcpB (Fig. 5C), the R303 equivalent residue Q305 hydrogen-bonds to Y88 (equivalent to the Y89 residue in BmCcpA) in the absence of HPr-Ser46-P and the T306 equivalent site was replaced by Val308, making it impossible to form the hydrogen bond with Y90 (equivalent to the Y91 residue in BsCcpA). Similar situation also applied to the predicted RcCcpA (Fig. S5E). These results corroborate that the solved structure of apo-RcCcpB or predicted structure of RcCcpA is in a DNA-binding competent conformation and suggest that Q305 (N301 in RcCcpA) and V308 (L304 in RcCcpA) sites are critical for maintaining HPr/Cph-independent DNA binding conformation of RcCcpB/RcCcpA. Single site mutations of Q305 and V308 in RcCcpB, N301 and L304 in RcCcpA, resulted in diminished DNA binding capacities, confirming the role of these residues in supporting the DNA-binding conformations of RcCcpB and RcCcpA (Figs. 5D, 5E, S5F). Examination of CcpA orthologs from 14,998 microbial genomes revealed that residues R303 and T306 critical for the HPr/Cph-dependent allosteric switch, are not conserved in the majority of species except for most bacilli genomes (Fig. 5F). This suggests that HPr/Cph-dependent allosteric switching of CcpA orthologs in the canonical CCR is not a universal mechanism across the bacterial kingdom but mainly evolved in certain bacilli species.

DISCUSSION

The PTS is prominently known for its role in microbial carbon metabolism through direct sugar transport and regulatory functions. Since PTS-mediated sugar uptake has only been characterized in a few microbial representatives, our study provides, for the first time to our knowledge, a systematic view of its prevalence and distribution in nearly 15,000 genomes from 4293 prokaryotic species. We revealed that the canonical PTS is present in only 26.3% of sequenced microbial species; the incomplete PTS is prevalent and distributed in a phylogeny-independent manner. These data suggest that the PTS is neither a dominant nor conserved mechanism for prokaryotes to transport carbohydrates, supporting the hypothesis that PTS-mediated sugar transport evolved recently [2, 16, 53]. Key players of PTS-mediated CCR in *Enterobacteria* and *Firmicutes* are EI_A and HPr, respectively [2, 16, 21]. However, 27% and 23% of microbial species in our large data set have no predictable homologs of EI_A and HPr, suggesting the well-known CCR mechanism mediated by PTS is not applicable to most prokaryotes.

This study raises a question about the role of incomplete PTSs in CCR, particularly the relationship between CCR and the completeness of PTS in carbohydrate utilization. It is known that canonical PTS-mediated CCR allows microbes to efficiently use sugars, grow fast and remain competitive in natural environments [2, 16]. However, our results indicated that bacterial strains carrying incomplete PTS have a different strategy for carbon utilization, for example, co-consumption of multiple carbohydrates in *R. cellulolyticum*. If all bacteria in a microbial community only contain canonical PTS components and strictly execute PTS-mediated CCR, competition for preferred carbon could happen between bacteria, leading to inefficient carbon utilization and inhibition of carbon degradation. However, with diversified carbon utilization strategies depending on PTS regulatory machineries, the cooperation and efficiency of carbon degradation can be greatly improved within a microbial community that constantly faces complex and

dynamic supply of carbon sources. On the other hand, given that PTS-mediated CCR presumably improves bacterial competitiveness, bacteria with canonical CCR may be more competitive when facing stresses such as warming or drying. For example, most species in *Bacilli* and *Clostridia* are endospore-forming and can remain in the dormant state for years [54, 55], which could offer competitive advantages for both bacilli and clostridia under selective pressure from increased temperature or drying. However, our previous study found that warming can only increase the relative abundance of *Bacilli* rather than *Clostridia* and the warming-enhanced homogeneous selection was mainly attributed to the positive responses of *Bacilli* [56]. This differential impact on *Bacilli* and *Clostridia* in the warming field can be explained by the bacterial PTS completeness because the completeness of core PTS components in bacilli is much higher than clostridia (Fig. 1C), and the HPr/Cph-dependent allosteric switching mainly exists in bacilli (Fig. 5F). These indicate that more species in *Bacilli* may execute canonical PTS-mediated CCR, which can finally enhance the competitiveness of *Bacilli* in response to warming stress compared to *Clostridia*.

Our study in *R. cellulolyticum*, which was selected as a representative of incomplete PTS carriers in cellulose-degrading clostridia, demonstrated a mild reverse CCR and catabolite promotion by incomplete PTS components, such as Cph and CcpA. The incomplete PTS regulates metabolism through molecular mechanisms distinct from the long-standing model established in *B. subtilis*, where the CcpA-HPr-S46-P complex represses the catabolism of unfavorable sugars. First, *R. cellulolyticum* CcpA plays a counterintuitive role in metabolic regulation since it is not involved in the utilization of co-existing soluble sugars both metabolically and transcriptionally, but it is essential for cellulose metabolism by maintaining a sustainable high expression of the cellulose degrading gene cluster through yet to be characterized mechanisms (Supplementary text B). Second, CcpA and CcpB in *R. cellulolyticum* have mechanistically diverged from previously characterized CcpA orthologs because of distinct DNA binding motifs identified experimentally and changes in ligand-dependent allosteric regulation. In this incomplete PTS model, CcpA and CcpB also have been functionally specified, differing in metabolic regulation from the aspect of carbohydrate consumption, transcriptional association, and DNA binding profiles. Third, H15-free Cph in *R. cellulolyticum* has a similar structure to other HPr but does not interact with either CcpA or CcpB to alter DNA binding, regardless of its phosphorylation status mediated by HprK using either ATP or newly discovered Ppi as the energy donor (Supplementary text C). Fourth, Cph and its phosphorylated form (e.g., cph_S46D) regulate carbohydrate metabolism and affect broad transcriptional profiles, which are distinct from CcpA-/CcpB-associated metabolic and transcriptional regulation. Instead of interacting with CcpA regulators, we found Cph binds to a variety of proteins, such as ribosomal components, anti-sigma factor antagonists, and transcriptional termination factors, which could link to transcriptional regulation and physiological adaptation to changing carbon sources (Supplementary text B).

In terms of sugar uptake in anaerobic cellulose degraders with an incomplete PTS, we would consider the incomplete PTS as a beneficial evolutionary outcome for economical and efficient carbohydrate transport for cellulose-degrading bacteria. First, inducer exclusion led by PTS sugar transport would not exist such that these microbes can consume multiple sugars simultaneously as shown in *R. cellulolyticum*. Second, ABC transporters that preferably pump disaccharides or oligosaccharides, such as the preference for cellobiose over glucose in *R. cellulolyticum*, stoichiometrically consume less energy to internalize sugar units for central metabolism, when compared to monosaccharide ABC transporters. This is of more significance to cellulose degraders that require energy-demanding synthesis and secretion of cellulose degrading enzymes.

Taken together, these genetic, metabolic, transcriptional, and biochemical data concordantly support that incomplete PTS components independently regulate carbohydrate metabolism. Given that lignocellulose utilization is critical to carbon cycling in the global ecosystem, our findings can also provide new insights into selections of functional gene information for incorporating into ecosystem models for improving accuracy in carbon decomposition and reducing model uncertainty (e.g., selecting both incomplete PTS genes and carbon degradation genes instead of only choosing carbon degradation genes).

CONCLUSION

Spanning from massive genome mining to mechanistic dissection, this study brings novel understanding of the limited applicability of the long-standing canonical PTS in prokaryotes and the diversified regulatory mechanisms of PTS components for carbohydrate metabolism. Distinct from catabolite repression of canonical PTS, the representative incomplete PTS in *R. cellulolyticum* promotes carbohydrate utilization, especially the indispensable role for Cph and CcpA in cellulose metabolism. Systematic investigation into *R. cellulolyticum* PTS components revealed independent functioning of Cph and CcpA, which is determined by structural changes in CcpA. These fundamental discoveries allow us to rethink the role of PTS in microbial competitiveness in microbiomes and will guide the design of microbiome-oriented nutrient intervention for improved functions in different settings.

METHODS

Bacterial strains and plasmid construction

Strains, plasmids, and primers used in this study are listed in the Table S5. Bacterial mutant strains of *R. cellulolyticum* were constructed in the $\Delta mspI$ background using an all-in-one CRISPR-Cas9 nickase editing system [41]. By inserting a 7-bp DNA fragment to cause frameshift and introduction of early stop codons, the knockout mutants of Ccel_0806, Ccel_1005, and Ccel_1438 were generated and named Δcph , $\Delta ccpA$, and $\Delta ccpB$, respectively. We also performed point mutagenesis in Ccel_0806, yielding *cph*_S46D in which Ser at 46 was substituted by Asp to simulate phosphorylation [57]. However, we failed to create a S46A mutant in which Ser at 46 would have been substituted by Ala to mimic loss of phosphorylation. The genome editing plasmid in *R. cellulolyticum* mutants was cured as previously reported [58]. As we failed to knock out the *hprK* gene using either Clostron or Cas9 nickase genome editing, RNA interference (RNAi) was applied to downregulate the expression of the *hprK* gene by plasmid-based overexpression of antisense RNAs that are complementary to *hprK* transcripts (pRNAi-*hprK*). The empty plasmid transformant (pRNAi) was used as an experimental control.

Measurement of bacterial growth and carbohydrate utilization

All *R. cellulolyticum* mutants and control strains were revived in complex VM medium [59] with 5 g/L cellobiose. Unless otherwise indicated, all strains were cultured anaerobically at 34 °C. To test microbial physiology and metabolism, a carbon mixture or Avicel PH101 crystalline cellulose was used as the carbon source. The carbon mixture consisted of 3 g/L cellobiose, 2.5 g/L glucose, 2 g/L xylose, and 1 g/L arabinose to mimic the complexity of released sugars during lignocellulose degradation. When using carbon mixture as a carbon source, we cultured bacterial strains to OD₆₀₀ 0.5–0.6 in the defined VM medium, and then transferred into the same fresh medium for sample collection and growth monitoring. Cellobiose-grown cultures at the exponential growth phase (OD₆₀₀ 0.5–0.6) were inoculated into 50 ml of defined VM medium with 10 g/L Avicel PH101 crystalline cellulose. Unless otherwise indicated, 15 µg/ml thiampenicol was added to maintain the plasmid in pRNAi-*hprK* and pRNAi strains. Three biological replicates of each strain were used for analyses.

Bacterial growth on carbon mixture and cellulose was estimated by optical density and total protein measurement, respectively [58]. High-performance liquid chromatography was used to measure cellobiose, glucose, xylose, and arabinose in the supernatant of spent medium [58].

The specific rate of sugar consumption ($q_{\text{substrate}}$) in mmol per gram of cells per hour was the derivative of the time course plots [60]. The remaining cellulose during fermentation was measured by the phenol-sulfuric acid method [61].

Microarray analysis

R. cellulolyticum strains (i.e., the parent strain, $\Delta ccpA$, $\Delta ccpB$, Δcph , and *cph*_S46D) were cultivated in the defined VM medium with either carbon mixture or cellulose as mentioned above. On carbon mixture, each strain had three biological replicates and 10 ml cell samples were collected at three time points (16, 20, and 24 h) during the exponential phase. On 10 g/L cellulose, $\Delta ccpB$ and the parent strain were cultivated with six biological replicates and collected at mid-exponential growth phase. Since $\Delta ccpA$ and Δcph could not grow on cellulose, we performed an incubation experiment (Fig. S3C) to understand how $\Delta ccpA$ and Δcph respond to cellulose. Specifically, the parent strain, $\Delta ccpA$, and Δcph were grown to an OD₆₀₀ of 0.5–0.6 in 50 ml defined VM media with cellobiose as the carbon source. Each strain had three biological replicates. Bacterial cells from each biological replicate were then collected by centrifugation at 4000 g and washed twice with the defined VM medium (no carbon added). Finally, washed cells from each biological replicate were inoculated into the defined VM medium with 10 g/L cellulose. During shaking incubation at 34 °C, samples were collected at five time points (0, 1, 3, 6, 12 h). After centrifugation at 4 °C, 5000 × g for 10 min, cell pellets were immediately flash frozen with liquid nitrogen and then stored at –80 °C for further use.

Total RNA extraction, microarray hybridization, and data analysis were performed as described previously [62]. Venn diagram graphs were generated with the online tool (<http://bioinformatics.psb.ugent.be/webtools/Venn/>). ClueGO v 2.5.5 and CluePedia v 1.5.5 were used for functional enrichment analysis with a two-sided minimal-likelihood test, and *p* values were corrected using the Benjamini-Hochberg method [63]. PCA analysis was conducted with the “prcomp” function in R using the normalized data.

ChIP-seq

$\Delta ccpA$ and $\Delta ccpB$ mutants were transformed with pEr-CcpA-FLAG and pEr-CcpB-FLAG respectively. The parent strain carrying pEr-Empty-FLAG was used as the negative control. The ChIP-seq experiment was performed as reported previously with some minor modifications [64, 65]. In brief, bacterial cells grown on carbon mixture in 50 ml of the defined VM medium were harvested at OD₆₀₀ 0.5; to keep consistent with transcriptomic analysis on cellulose, bacterial cells grown on cellobiose were incubated with 10 g/L cellulose for 8 h (Fig. S3C). Each strain has two biological replicates under each condition. Cell pellets from 50 ml of these cultures were treated with 1% formaldehyde for 10 min at 34 °C, then quenched by 250 mM glycine. Cells were then collected by centrifugation, washed twice with the PBS buffer (pH 7.4) and resuspended in 750 µL of the IP-buffer (100 mM Tris-HCl pH 8.0, 1% Triton X-100, 500 mM NaCl, and 1 mM PMSF). Cells were disrupted by bead beating and then sonicated to shear DNA/protein to an average size of 300–500 bp. After centrifugation, the supernatant was incubated with anti-FLAG M2 Magnetic Beads at 4 °C overnight. In the next day, the anti-FLAG M2 Magnetic Beads were washed twice with wash buffer 1 (100 mM Tris-HCl pH 8.0 and 600 mM NaCl), wash buffer 2 (100 mM Tris-HCl pH 8.0 and 300 mM NaCl) and wash buffer 3 (100 mM Tris-HCl pH 8.0 and 100 mM NaCl), and finally eluted by the ChIP elution buffer (100 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS) with 50 µg proteinase K at 56 °C for 2 h and then at 65 °C overnight. Immunoprecipitated DNA samples were purified using the ChIP DNA Clean & Concentrator (ZYMO). DNA was used for the construction of sequencing libraries using a KAPA DNA Library Preparation kit (Kapa Biosystems) and then sequenced on MiSeq (Illumina) with the MiSeq Reagent Kit V2.

ChIP-seq reads were mapped to the genome of *R. cellulolyticum* H10, using BWA-MEM (Galaxy Version 0.7.17.1) with Illumina mode [66], and then filtered (MAPQ quality score > 20) by Filter SAM or BAM (Galaxy Version 1.8+) [67]. The enriched peaks were identified using MACS2 callpeak (Galaxy Version 2.1.1) [68]. The consensus motifs for Ccel_1438 and Ccel_1005 were predicted by MEME-ChIP (version 5.3.3) [69].

Immunoprecipitation-mass spectrometry (IP-MS)

The parent strain w/pEr-Empty-FLAG (negative control), $\Delta ccpA$ w/pEr-CcpA-FLAG, $\Delta ccpB$ w/pEr-CcpB-FLAG, and Δcph w/pEr-Cph-FLAG were used for the IP-MS experiment to detect potential proteins that interact with CcpA, CcpB, and Cph. The IP-MS was performed using two

independent cultures for each strain. All IP-MS experiments were conducted in medium with 10 g/L cellulose using the same incubation experiment design (Fig. S3C). All strains were incubated with cellulose for 8 h and then were collected by centrifugation, washed twice with PBS buffer and resuspended in 750 μ L IP-buffer (100 mM Tris-HCl pH 8.0, 1% Triton X-100, 500 mM NaCl, and 1 mM PMSF). The cells were disrupted by bead-beating and then centrifugation at 13,000 *g* for 10 min. The supernatant was incubated with Anti-FLAG M2 Magnetic Beads (Sigma Aldrich) at 4 °C overnight. The next day, the Anti-FLAG M2 Magnetic Beads were washed twice with wash buffer 1 (100 mM Tris-HCl pH 8.0 and 600 mM NaCl), wash buffer 2 (100 mM Tris-HCl pH 8.0 and 300 mM NaCl) and wash buffer 3 (100 mM Tris-HCl pH 8.0 and 100 mM NaCl), and finally target proteins were eluted from the antibody using 100 μ L elution buffer (100 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS) at 65 °C for 2 h. The eluted proteins were prepared by filter aided sample preparation for mass spectrometry as previously reported [70]. Tryptic peptides were separated by reverse phase XSelect CSH C18 2.5 μ m resin (Waters) on an in-line 150 $\times$ 0.075 mm column using an UltiMate 3000 RSLCnano system (Thermo). Peptides were eluted using a 90 min gradient from 98:2 to 65:35 buffer A:B ratio. Eluted peptides were ionized by electrospray (2.2 kV) followed by mass spectrometric analysis on an Orbitrap Fusion™ Tribrid™ Mass Spectrometer (Thermo). MS data were acquired using the FTMS analyzer in profile mode at a resolution of 240,000 over a range of 375–1500 *m/z*. Following HCD activation, MS/MS data were acquired using the ion trap analyzer in centroid mode and normal mass range with normalized collision energy of 28–31% depending on charge state and precursor selection range. The acquired MS/MS data were searched using Sipsos algorithm against the *R. cellulolyticum* H10 protein database with the FDR < 0.01 at protein level [71].

Protein expression and purification

The Cph, CcpA, CcpB, and HprK genes were amplified by PCR from *R. cellulolyticum* H10 genomic DNA, and cloned into the modified pET28a vector (Novagene) which contains a C-terminal 6xHis tag. The cloning junctions were confirmed by DNA sequencing. Each recombinant plasmid was transformed into the *E. coli* Rosetta (DE3) strain (Novagen). Cells were grown at 37 °C in Luria Bertani (LB) medium containing 100 μ g/ml kanamycin and 40 μ g/ml chloramphenicol. Expression of Cph, CcpA, CcpB, and HprK were induced at an OD_{600 nm} of 0.8 by adding 0.2 mM isopropyl- β -D-1-thiogalactopyranoside followed by incubation at 16 °C for 20 h. The cells encoding Cph or HprK were collected and lysed by sonication in 20 ml buffer A (20 mM Tris-HCl buffer pH 8.0, 100 mM NaCl) in an ice-water bath. After centrifugation, the His-tagged fusion protein Cph or HprK was isolated with Ni-NTA affinity column (GE Healthcare) and further purified by gel filtration (Superdex 75, GE Healthcare) in buffer A. The cells encoding CcpA were sonicated in 20 ml buffer B (20 mM Tris-HCl buffer pH 8.0 containing 300 mM NaCl) in an ice-water bath. After lysis by sonication, CcpA were isolated with Ni-NTA affinity column (GE Healthcare) and further purified by gel filtration (Superdex 200, GE Healthcare) in buffer B. The cells expressing CcpB were lysed by sonication in 20 ml buffer C (20 mM Tris-HCl buffer pH 7.0 containing 300 mM NaCl) in an ice-water bath. After centrifugation, the CcpB were isolated with Ni-NTA affinity column (GE Healthcare) and further purified by gel filtration (Superdex 200, GE Healthcare) in buffer C. The Cph, CcpA, and CcpB mutants were obtained by site-directed mutagenesis using the plasmid encoding the wild-type proteins as a template, overexpressed, and purified the same way as the wild-type protein. All eluted proteins were collected, concentrated using centrifugal ultrafiltration respectively for further study. The purity of protein samples was assessed by denaturing polyacrylamide gel electrophoresis (SDS-PAGE). Purified CcpA ($\epsilon_{280} = 15,000 \text{ M}^{-1} \text{ cm}^{-1}$) and CcpB ($\epsilon_{280} = 24,000 \text{ M}^{-1} \text{ cm}^{-1}$) were incubated with Cph, Cph_S16D, Cph_S46D, and Cph_S16&46D ($\epsilon_{280} = 8500 \text{ M}^{-1} \text{ cm}^{-1}$). CcpA/CcpB with Cph and Cph mutant mixtures (1:10 molar ratio) were analyzed on analytical gel filtration chromatography (Superdex 75 HR 10/300 GL) in buffer B or buffer C, respectively. The elution volumes for CcpA, CcpB, and Cph (Cph mutants) were 8.5 ml, 11 ml, and 15 ml, respectively.

Isothermal titration calorimetry (ITC) assay

The ITC experiments were performed with a MicroCal PEAQ-ITC instrument (Malvern, USA) and all statistical analyses were performed with the origin ITC analysis software package (MicroCal). All samples were centrifuged and degassed extensively prior to the experiments. Cph (400 μ M; Syringe) was firstly titrated into CcpA (20 μ M; sample cell) and CcpB (20 μ M; sample cell), respectively. Then, the Tris solution (20 mM Tris-HCl, pH 7.5, and 100 mM

NaCl) was also titrated into the CcpA sample or CcpB sample, and the resulting heat was subtracted from the experimental curve. The titration was performed with a total of 19 injections with constant stirring. The data was finally fitted to the one-site binding model to generate thermodynamic parameters.

In vitro phosphorylation of Cph

The phosphorylation assays were performed in the presence of His₆-Cph (1 mM) and His₆-HprK/P (500 μ M) in phosphorylation buffer (10 mM Tris-HCl, pH 7.4, 100 mM NaCl, 5 mM MgCl₂, 1 mM dithiothreitol (DTT), and 10% glycerol, 20 mM ATP or pyrophosphate) with or without FBP (10 mM). The reactions were carried out at 25 °C for 60 min. The samples were loaded onto either nondenaturing 17% polyacrylamide gels for separation of unphosphorylated Cph and phosphorylated Cph or 12% Native-PAGE for further mass spectrometry analysis of phosphorylated Cph. The phosphorylation ratios were estimated using the formula:

$$\text{Phosphorylation ratio} = \frac{\text{spectral counting (SC) of the phosphorylated peptide}}{\text{SC of the phosphorylated peptide} + \text{SC of the non-phosphorylated counterparts}} \times 100\%$$

Phosphorylation site determination by liquid chromatography-tandem mass spectrometry (LC-MS/MS)

The In-gel digestion of the phosphorylated Cph were performed as described previously [72]. Briefly, the band slices were reduced with 10 mM Tris (2-carboxyethyl) phosphine in 100 mM NH₄HCO₃, alkylated with 40 mM 2-chloroacetamide in 100 mM NH₄HCO₃, and digested with trypsin (Sigma) at a 1:50 enzyme-to-substrate ratio. Digested peptides were purified via homemade C18 Spin Tips, lyophilized, dissolved in 0.1% formic acid (10 μ L), and measured by LC-MS/MS. 2 μ g of each sample peptides were loaded onto an Easy-nL 1200 ultra-high-performance liquid chromatography (UHPLC) instrument (Thermo Fisher Scientific) and transferred into an online coupled Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific) by nanoelectrospray ionization (nanoESI). Peptides were eluted from a 20-cm-long analytical column packed with 1.9- μ m reverse-phase C18 particles using a 60-min segmented gradient of 5% to 45% solvent B (90% [vol/vol] acetonitrile, 0.1% [vol/vol] formic acid) at a constant flow rate of 300 nL/min. For the MS/MS measurement, a data dependent Top Speed (3 s) method was used to collect spectra: high resolution MS1 spectra (Orbitrap resolution: 120,000; mass range: 350–1400 Th; and automatic gain control (AGC) target: 4×10^5 ; maximum injection time 50 ms) and high resolution MS2 spectra (Quadrupole isolation window: 1.6 Th; Orbitrap resolution: 30,000; HCD energy: 30%; AGC target: 5×10^2 ; maximum injection time: 22 ms). Dynamic exclusion was enabled with a duration time of 120 s. Acquired raw data were processed using MaxQuant software [73] with Methionine oxidation and serine, threonine, and tyrosine residues phosphorylation set as variable modifications, and cysteine carbamidomethylation set as fixed modification. All the other parameters were set as default. The raw data file was searched against a reference *E. coli* K-12 proteome and the protein sequences of Cph and HprK from *R. cellulolyticum* H10 using a built-in Andromeda search engine [74]. The maximum number of missed cleavages allowed was set to 3 for trypsin. Phosphopeptides with an Andromeda score of >70 and a localization probability of >0.75 were considered, and their MS/MS spectra were then analyzed manually.

Electrophoretic mobility shift assay

6-carboxyfluorescein (FAM)-labeled DNA fragments were obtained by annealing oligos (Table S5). The DNA fragments (potential binding motif for CcpA) were added at a concentration of 0.05 μ M and incubated at room temperature for 10 min with an excess of CcpA, CcpAN301R, or CcpAL304T in a total volume of 20 μ L mixture. The mixture contained 20 mM Tris pH 7.5, 500 mM KCl, 10 mM MgCl₂, 10% (vol/vol) glycerol, 2 mM DTT, and 2 mM EDTA. Bovine serum albumin (BSA) was added at 3 μ M to prevent nonspecific binding. The DNA fragments (potential binding motif for CcpB) were added at a concentration of 0.1 μ M and incubated at room temperature for 10 min with an excess of CcpB, CcpBQ305R, and CcpBV308T in a total volume of 20 μ L mixture. The mixture contained 20 mM Tris pH 7.5, 250 mM KCl, 10 mM MgCl₂, 10% (vol/vol) glycerol, 2 mM DTT, and 2 mM EDTA. BSA was added at 3 μ M to prevent nonspecific binding. After incubation for 30 min at 25 °C, the protein-DNA mixtures were directly subjected to 6% native PAGE with 1xTris acetate-EDTA

buffer. For FAM-labeled probes, the images were collected and analyzed on a Typhoon FLA-9500 imaging system (GE Healthcare).

Crystallization and data collection

The purified Cph and CcpB were concentrated by ultrafiltration (Millipore Amicon) to 10 mg/ml and 3 mg/ml, respectively. Initial crystal screening was set up with Crystal Screen, Crystal Screen 2, and PEG/Ion Screen reagent kits (Hampton Research) at 16 °C by using the hanging-drop vapor-diffusion method. Crystals appeared within 2 weeks. Further crystal optimization experiments were performed by systematic variation of the precipitant concentration and protein concentration. The Cph crystals were produced by mixing 1 µl protein solution (~10 mg/ml) and an equal volume of reservoir solution containing 0.2 M sodium sulfate, 20% w/v PEG 3350, 10% v/v ethylene glycol, and incubating at 15 °C. Crystallization of CcpB was performed at 16 °C using the hanging-drop vapor diffusion method. The appropriate crystals of CcpB were obtained from the reservoir solution of 0.1 M Bis Tris propane pH 6.5, 0.2 M trisodium citrate dihydrate, 20% w/v PEG 3350, 12% v/v ethylene glycerol after 3 weeks.

The crystals were harvested using cryo-loops and immersed briefly in a cryoprotectant solution consisting of 80% (v/v) reservoir solution and 20% (v/v) glycerol. The crystals were subsequently flash-cooled and stored in liquid nitrogen and transferred to beamline BL18U1 of the SSRF (Shanghai Synchrotron Radiation Facility) for X-ray diffraction analysis and data collection. The data set of Cph and CcpB were processed using HKL3000. Crystals of Cph and CcpB belong to *P4₁2₁2* and *H3*, respectively. Table S5 gives a summary of the data-collection statistics.

Structure determination, refinement, and prediction

The structures of Cph and CcpB were determined by the Molecular Replacement method using the structures of Crh (PDB, 1mo1) from *B. subtilis* and CcpA (PDB, 1rzt) from *Priestia megaterium* as the search models, respectively. The final structures of Cph and CcpB were solved at a resolution of 2.4 and 2.2 Å, respectively. The resultant model was refined using the maximum likelihood method implemented in REFMAC5 as part of the CCP4i program suite and the PHENIX program refinement and rebuilt interactively using the program COOT. Both the structures were assessed with MolProbity, and the Ramachandran outliers were less than 1% in the final refined structures. Crystallographic parameters are listed in Table S5. The predicted structure of CcpA was obtained by submitting its sequence to the AlphaFold2 server at <https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb> using the default settings [75]. All structure figures were prepared with PyMOL.

Genome mining

We downloaded both genomic sequences and protein annotation files (*.genomic.fna.gz and *.protein.faa.gz) of 14,998 prokaryotic genomes from the NCBI database, which were labeled as complete genomes in December 2019. 99.37% of their genomes have completeness reported to be >90% in the GTDB database (<https://data.gtdb.ecogenomic.org/releases/release207/207.0/>) [76]. Each genome was then taxonomically classified by GTDB-Tk v2.1.0 and R207v2 reference database using the *classify_wf* function with default setting [76]. All protein sequences were indexed by adding corresponding GenBank accession and assembly name to their headers, then pooled together to generate a single in-house protein database with 56,869,107 sequences, which is publicly available (<http://www.ou.edu/ieg/publications/datasets>). All hmm files of reported PTS components were downloaded from the Pfam database (Table S1). The number of homologs in each genome was examined by hmm search from the hmmer-3.3 package against the protein database. The E-value threshold we applied for hmm search was at 0.01, which was the gathering threshold for building Pfam families [77]. Protein homologs were grouped into two categories based on protein architecture, including singular or modular/fused forms. Counts of each architecture-specific homolog in each genome, along with corresponding genome identification and taxonomic information, were summarized in a single master file for downstream analysis (Table S1).

For each hmm profile of PTS components, we calculated its averaged copy number per genome with protein architecture features considered, and further examined the constitution of singular or fused architecture for each hmm profile across genomes. To understand the presence of PTS components, including HPr, EI, EIIA, EIIB, EIIC, HPrK, and LacI family regulators, we summed the count number of hmm profiles of each PTS component in each genome. Since the EI complex consists of EIIA, EIIB,

and EIIC, the completeness of EI is determined by the co-presence of all three components in the same genome. We then determined microbial species with unique PTS configurations according to their taxonomic information. This also allowed us to interrogate the possibility of inter-strain variations in PTS configuration in some relevant microbes. A Venn diagram was used to illustrate the number of microbial species with some unique or shared PTS configurations.

To understand the association between microbial phylogeny and PTS configuration, we generated a phylogenetic tree for all species at the class level with at least 50 genomes included. The taxid information of downloaded genomes was applied to infer the phylogenetic relationship between microbial strains using the online phyloT tree generator (<https://phylot.biobyte.de/>). For each class on the phylogenetic tree, we calculated the percentage of genomes that carry a certain copy number of PTS components or different completeness of EI (including absent, partial, and complete). The phylogenetic tree from phyloT was visualized by iTOL (<https://itol.embl.de/>). The integrated visualization of a heatmap showing averaged counts of each hmm profiles in genomes of all classes, and their phylogeny displayed by cladogram, was plotted by complex heatmap [78]. The Shannon index of PTS configuration was calculated for each genome by the “diversity()” function of the vegan package [79] in R and then statistically summarized to class levels to indicate the diversity of PTS configuration within each class. The higher Shannon index indicates likely coexistence and similar copy number of homologs in a genome, thereby reflecting the completeness of all hmm profiles. The variation of Shannon index indicates the consistency of PTS configuration within each class.

To evaluate the co-occurrence of PTS components, we computed Jaccard coefficients in R. At the module level, the copy number of each hmm profile was converted to a binary data before computing; at the protein level, copy numbers of PTS components, or the total copy number of hmm profiles belonging to specific PTS components was converted binarily for computing.

All HPr homologs predicted in 8651 genomes by hmm search were pooled together for motif analyses. Sequences were grouped into three HPr categories (H15, H15-free, and S16), by the amino acid at the pseudo position 15, which was determined by searching for similar conserved motifs. Since Crh and Npr are too similar to be distinguished from canonical HPr, here we named all of them as HPr homologs. The probability of amino acids used in the conserved motif was plotted by ggseqlog [80]. For those H15-free HPr homologs, we further examined the presence of other PTS components, including HPrK, EI, and completeness of EI components, then projected the PTS configuration data to the phylogenetic tree of these H15-free carrying microbes. Cellulose degraders were manually curated.

DATA AVAILABILITY

Raw sequences of the ChIP-seq experiments are available in the NCBI SRA database (<http://www.ncbi.nlm.nih.gov/sra>) under accession number PRJNA748961. The MS proteomic data of IP-MS or phosphorylation determination are deposited in the ProteomeXchange consortium (<http://www.proteomexchange.org/>) via the iProX partner repository with the accession number PXD027554. Microarray data are deposited in the ArrayExpress under accession number E-MTAB-11755 and E-MTAB-11774. The structures of CcpB and Cph have been deposited to the Protein Data Bank (PDB) with accession numbers 7FF4 and 7FF5, respectively. Our in-house protein database can be accessed through <http://www.ou.edu/ieg/publications/datasets>. All other relevant data are available in Supplementary Information. Source data are provided with this paper.

CODE AVAILABILITY

The genome mining and other analysis codes have been deposited in GitHub (<https://github.com/adarobot/PTS>).

REFERENCES

- Josef D, Galinier A, Martin-Verstraete I. Carbohydrate uptake and metabolism. *Bacillus subtilis and its closest relatives*. 2014. ASM Press, Washington, DC, USA, 129–50.
- Deutscher J, Aké FMD, Derkaoui M, Zébré AC, Cao TN, Bouraoui H, et al. The bacterial phosphoenolpyruvate:carbohydrate phosphotransferase system: regulation by protein phosphorylation and phosphorylation-dependent protein-protein interactions. *Microbiol Mol Biol Rev*. 2014;78:231–56.

3. Galinier A, Deutscher J. Sophisticated regulation of transcriptional factors by the bacterial phosphoenolpyruvate: sugar phosphotransferase system. *J Mol Biol.* 2017;429:773–89.
4. Deutscher J, Francke C, Postma PW. How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiol Mol Biol Rev.* 2006;70:939–1031.
5. Pflüger-Grau K, Görke B. Regulatory roles of the bacterial nitrogen-related phosphotransferase system. *Trends Microbiol.* 2010;18:205–14.
6. Snyder H, Kellogg SL, Skarda LM, Little JL, Kristich CJ. Nutritional control of antibiotic resistance via an interface between the phosphotransferase system and a two-component signaling system. *Antimicrob Agents Chemother.* 2014;58:957–65.
7. Sánchez S, Chávez A, Forero A, García-Huante Y, Romero A, Sánchez M, et al. Carbon source regulation of antibiotic production. *J Antibiot.* 2010;63:442–59.
8. Houot L, Chang S, Pickering BS, Absalon C, Watnick PI. The phosphoenolpyruvate phosphotransferase system regulates *Vibrio cholerae* biofilm formation through multiple independent pathways. *J Bacteriol.* 2010;192:3055–67.
9. Houot L, Chang S, Absalon C, Watnick PI. *Vibrio cholerae* phosphoenolpyruvate phosphotransferase system control of carbohydrate transport, biofilm formation, and colonization of the germfree mouse intestine. *Infection Immunity.* 2010;78:1482–94.
10. Wu M-C, Chen Y-C, Lin T-L, Hsieh P-F, Wang J-T. Cellobiose-specific phosphotransferase system of *Klebsiella pneumoniae* and its importance in biofilm formation and virulence. *Infect Immun.* 2012;80:2464–72.
11. Kok M, Bron G, Erni B, Mukhija S. Effect of enzyme I of the bacterial phosphoenolpyruvate: sugar phosphotransferase system (PTS) on virulence in a murine model. *Microbiology.* 2003;149:2645–52.
12. Kim H-J, Lee C-R, Kim M, Peterkofsky A, Seok Y-J. Dephosphorylated NPr of the nitrogen PTS regulates lipid A biosynthesis by direct interaction with LpxD. *Biochem Biophys Res Commun.* 2011;409:556–61.
13. Ha J-H, Hauk P, Cho K, Eo Y, Ma X, Stephens K, et al. Evidence of link between quorum sensing and sugar metabolism in *Escherichia coli* revealed via cocrystal structures of LsrK and HPr. *Sci Adv.* 2018;4:eaar7063.
14. Kundig W, Ghosh S, Roseman S. Phosphate bound to histidine in a protein as an intermediate in a novel phospho-transferase system. *Proc Natl Acad Sci USA.* 1964;52:1067–74.
15. Martin-Verstraete I, Charrier V, Stülke J, Galinier A, Erni B, Rapoport G, et al. Antagonistic effects of dual PTS-catalysed phosphorylation on the *Bacillus subtilis* transcriptional activator LevR. *Mol Microbiol.* 1998;28:293–303.
16. Görke B, Stülke J. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat Rev Microbiol.* 2008;6:613–24.
17. Schumacher MA, Allen GS, Diel M, Seidel G, Hillen W, Brennan RG. Structural basis for allosteric control of the transcription regulator CcpA by the phosphoprotein HPr-Ser46-P. *Cell.* 2004;118:731–41.
18. Barabote RD, Saier MH Jr. Comparative genomic analyses of the bacterial phosphotransferase system. *Microbiol Mol Biol Rev.* 2005;69:608–34.
19. Lagier J-C, Khelaifia S, Alou MT, Ndongo S, Dione N, Hugon P, et al. Culture of previously uncultured members of the human gut microbiota by culturomics. *Nat Microbiol.* 2016;1:16203.
20. Nayfach S, Roux S, Seshadri R, Udway D, Varghese N, Schulz F, et al. A genomic catalog of Earth's microbiomes. *Nat Biotechnol.* 2021;39:499–509.
21. Deutscher J. The mechanisms of carbon catabolite repression in bacteria. *Curr Opin Microbiol.* 2008;11:87–93.
22. Delannoy-Bruno O, Desai C, Raman AS, Chen RY, Hibberd MC, Cheng J, et al. Evaluating microbiome-directed fibre snacks in gnotobiotic mice and humans. *Nature.* 2021;7865:91–95.
23. Patnode ML, Beller ZW, Han ND, Cheng J, Peters SL, Terrapon N, et al. Interspecies competition impacts targeted manipulation of human gut bacteria by fiber-derived glycans. *Cell.* 2019;179:59–73.e13.
24. Wastyk HC, Fragiadakis GK, Perelman D, Dahan D, Merrill BD, Yu FB, et al. Gut-microbiota-targeted diets modulate human immune status. *Cell.* 2021;184:4137–53.
25. Pereira FC, Wasmund K, Cobankovic I, Jehmlich N, Herbold CW, Lee KS, et al. Rational design of a microbial consortium of mucosal sugar utilizers reduces *Clostridiaceae* colonization. *Nat Commun.* 2020;11:5104.
26. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, Hugenholtz P. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.* 2022;50:D785–D794.
27. Cai L, Cai S, Zhao D, Wu J, Wang L, Liu X, et al. Analysis of the transcriptional regulator GlpR, promoter elements, and posttranscriptional processing involved in fructose-induced activation of the phosphoenolpyruvate-dependent sugar phosphotransferase system in *Haloferax mediterranei*. *Appl Environ Microbiol.* 2014;80:1430–40.
28. Pickl A, Johnsen U, Schonheit P. Fructose degradation in the haloarchaeon *Haloferax volcanii* involves a bacterial type phosphoenolpyruvate-dependent phosphotransferase system, fructose-1-phosphate kinase, and class ii fructose-1,6-bisphosphate Aldolase. *J Bacteriol.* 2012;194:3088–97.
29. Hartman AL, Norais C, Badger JH, Delmas S, Haldenby S, Madupu R, et al. The complete genome sequence of *Haloferax volcanii* DS2, a model archaeon. *PLoS One.* 2010;5:e9605.
30. Erni B. The bacterial phosphoenolpyruvate: sugar phosphotransferase system (PTS): an interface between energy and signal transduction. *J Iran Chem Soc.* 2013;10:593–630.
31. Plumbridge J. Regulation of gene expression in the PTS in *Escherichia coli*: the role and interactions of Mlc. *Curr Opin Microbiol.* 2002;5:187–93.
32. Saier MH Jr, Reizer J. Domain shuffling during evolution of the proteins of the bacterial phosphotransferase system. *Res Microbiol.* 1990;141:1033–8.
33. Waite DW, Vanwongterghem I, Rinke C, Parks DH, Zhang Y, Takai K, et al. Comparative genomic analysis of the class *Epsilonproteobacteria* and proposed reclassification to *Epsilonbacteraeota* (phyl. nov.). *Front Microbiol.* 2017;8:682.
34. Desvaux M. *Clostridium cellulolyticum*: model organism of mesophilic cellulolytic clostridia. *FEMS Microbiol Rev.* 2005;29:741–64.
35. Wilhelm RC, Singh R, Eltis LD, Mohn WW. Bacterial contributions to delignification and lignocellulose degradation in forest soils with metagenomic and quantitative stable isotope probing. *ISME J.* 2019;13:413–29.
36. Seshadri R, Leahy SC, Attwood GT, Teh KH, Lambie SC, Cookson AL, et al. Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. *Nat Biotechnol.* 2018;36:359–67.
37. van den Bogaard PTC, Kleerebezem M, Kuipers OP, de Vos WM. Control of lactose transport, β -galactosidase activity, and glycolysis by CcpA in *Streptococcus thermophilus*: evidence for carbon catabolite repression by a non-phosphoenolpyruvate-dependent phosphotransferase system sugar. *J Bacteriol.* 2000;182:5982–9.
38. Parche S, Belet M, Rezzonico E, Jacobs D, Arigoni F, Titgemeyer F, et al. Lactose-over-glucose preference in *Bifidobacterium longum* NCC2705: glcP, encoding a glucose transporter, is subject to lactose repression. *J Bacteriol.* 2006;188:1260–5.
39. Collier DN, Hager PW, Phibbs PV Jr. Catabolite repression control in the *Pseudomonas*. *Res Microbiol.* 1996;147:551–61.
40. Park H, McGill SL, Arnold AD, Carlson RP. *Pseudomonas* reverse carbon catabolite repression, interspecies metabolite exchange, and consortial division of labor. *Cell Mol Life Sci.* 2020;77:395–413.
41. Xu T, Li Y, Shi Z, Hemme CL, Li Y, Zhu Y, et al. Efficient genome editing in *Clostridium cellulolyticum* via CRISPR-Cas9 Nickase. *Appl Environ Microbiol.* 2015;81:4423–31.
42. Mijakovic I, Poncet S, Galinier A, Monedero V, Fieulaine S, Janin J, et al. Pyrophosphate-producing protein dephosphorylation by HPr kinase/phosphorylase: a relic of early life? *Proc Natl Acad Sci USA.* 2002;99:13442–7.
43. Tobisch S, Zühlke D, Bernhardt J, Stülke J, Hecker M. Role of CcpA in regulation of the central pathways of carbon catabolism in *Bacillus subtilis*. *J Bacteriol.* 1999;181:6996–7004.
44. Abranches J, Nascimento MM, Zeng L, Browngardt CM, Wen ZT, Rivera MF, et al. CcpA regulates central metabolism and virulence gene expression in *Streptococcus mutans*. *J Bacteriol.* 2008;190:2340–9.
45. Jones BE, Dossanet V, Küster E, Hillen W, Josef Deutscher, Klevit RE. Binding of the catabolite repressor protein CcpA to its DNA target is regulated by phosphorylation of its corepressor HPr. *J Biol Chem.* 1997;272:26530–5.
46. Deutscher J, Josef D, Küster E, Bergstedt U, Charrier V, Hillen W. Protein kinase-dependent HPr/CcpA interaction links glycolytic activity to carbon catabolite repression in Gram-positive bacteria. *Mol Microbiol.* 1995;15:1049–53.
47. Xu C, Huang R, Teng L, Jing X, Hu J, Cui G, et al. Cellulosome stoichiometry in *Clostridium cellulolyticum* is regulated by selective RNA processing and stabilization. *Nat Commun.* 2015;6:1–13.
48. Maamar H, Abdou L, Boileau C, Valette O, Tardif C. Transcriptional analysis of the cip-cel gene cluster from *Clostridium cellulolyticum*. *J Bacteriol.* 2006;188:2614–24.
49. Abdou L, Boileau C, de Philip P, Pagès S, Fiérobé H-P, Tardif C. Transcriptional regulation of the *Clostridium cellulolyticum* cip-cel operon: a complex mechanism involving a catabolite-responsive element. *J Bacteriol.* 2008;190:1499–506.
50. Xu C, Huang R, Teng L, Wang D, Hemme CL, Borovok I, et al. Structure and regulation of the cellulose degradome in *Clostridium cellulolyticum*. *Biotechnol Biofuels.* 2013;6:1–15.
51. Ye F, Wang C, Fu Q, Yan X-F, Bharath SR, Casanas A, et al. Structural basis of a novel repressor, SghR, controlling *Agrobacterium* infection by cross-talking to plants. *J Biol Chem.* 2020;295:12290–304.
52. Huffman JL, Lu F, Zalkin H, Brennan RG. Role of residue 147 in the gene regulatory function of the *Escherichia coli* purine repressor. *Biochemistry.* 2002;41:511–20.

53. Cases I, Velázquez F, de Lorenzo V. The ancestral role of the phosphoenolpyruvate–carbohydrate phosphotransferase system (PTS) as exposed by comparative genomics. *Res Microbiol.* 2007;158:666–70.
54. Dürre P. Physiology and Sporulation in *Clostridium*. *Microbiol Spectr.* 2014;2:2–4.
55. Errington J. Regulation of endospore formation in *Bacillus subtilis*. *Nat Rev Microbiol.* 2003;1:117–26.
56. Ning D, Yuan M, Wu L, Zhang Y, Guo X, Zhou X, et al. A quantitative framework reveals ecological drivers of grassland microbial community assembly in response to warming. *Nat Commun.* 2020;11:1–12.
57. Wittekind M, Reizer J, Deutscher J, Saier MH, Klevit RE. Common structural changes accompany the functional inactivation of HPr by seryl phosphorylation or by serine to aspartate substitution. *Biochemistry.* 1989;28:9908–12.
58. Li Y, Tschaplinski TJ, Engle NL, Hamilton CY, Rodriguez M Jr, Liao JC, et al. Combined inactivation of the *Clostridium cellulolyticum* lactate and malate dehydrogenase genes substantially increases ethanol yield from cellulose and switchgrass fermentations. *Biotechnol Biofuels.* 2012;5:2.
59. Higashide W, Li Y, Yang Y, Liao JC. Metabolic engineering of *Clostridium cellulolyticum* for production of isobutanol from cellulose. *Appl Environ Microbiol.* 2011;77:2727–33.
60. Desvaux M, Guedon E, Petitdemange H. Cellulose catabolism by *Clostridium cellulolyticum* growing in batch culture on defined medium. *Appl Environ Microbiol.* 2000;66:2461–70.
61. Hemme CL, Fields MW, He Q, Deng Y, Lin L, Tu Q, et al. Correlation of genomic and physiological traits of thermoanaerobacter species with biofuel yields. *Appl Environ Microbiol.* 2011;77:7998–8008.
62. Tao X, Xu T, Kempfer ML, Liu J, Zhou J. Precise promoter integration improves cellulose bioconversion and thermotolerance in *Clostridium cellulolyticum*. *Metab Eng.* 2020;60:110–8.
63. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics.* 2009;25:1091–3.
64. Davis SE, Mooney RA, Kanin EI, Grass J, Landick R, Ansari AZ. Mapping *E. coli* RNA polymerase and associated transcription factors and identifying promoters genome-wide. *Methods Enzymol.* 2011;498:449–71.
65. Nelson JD, Denisenko O, Bomsztyk K. Protocol for the fast chromatin immunoprecipitation (ChIP) method. *Nat Protoc.* 2006;1:179–85.
66. Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics.* 2010;26:589–95.
67. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAM tools. *Bioinformatics.* 2009;25:2078–9.
68. Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. *Nat Protoc.* 2012;7:1728–40.
69. Machanick P, Bailey TL. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics.* 2011;27:1696–7.
70. Wiśniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. *Nat Methods.* 2009;6:359–62.
71. Guo X, Li Z, Yao Q, Mueller RS, Eng JK, Tabb DL, et al. Sipsos Ensemble improves database searching and filtering for complex metaproteomics. *Bioinformatics.* 2018;34:795–802.
72. Wang Y, Zhang S-P, Zhang M-Y, Kempfer ML, Guo D-D, Han J-T, et al. The antitoxin MqsA homologue in *Pseudomonas fluorescens* 2P24 has a rewired regulatory circuit through evolution. *Environ Microbiol.* 2019;21:1740–56.
73. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol.* 2008;26:1367–72.
74. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res.* 2011;10:1794–805.
75. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. *Nat Methods.* 2022;19:679–82.
76. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics.* 2022;38:5315–6.
77. Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, et al. The Pfam protein families database. *Nucleic Acids Res.* 2012;40:D290–301.
78. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics.* 2016;32:2847–9.
79. Dixon P. VEGAN, a package of R functions for community ecology. *J Veg Sci.* 2003;14:927–30.
80. Wagih O. ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics.* 2017;33:3645–7.

ACKNOWLEDGEMENTS

Genome mining, construction of mutant strains, phenotypic and omics experiments were supported by the Office of the Vice President for Research at the University of Oklahoma to JZ. Biochemical assays, crystallization, and structure determination were supported by the National Natural Science Foundation of China (Grant No. 31971422) to YH and (Grant No.31970103) to HG, and Key Research and Development Plan Project of Anhui Province (grant No. 202004a06020035) to NZ. We thank SSRF beamline BL18U1 for synchrotron data collection. We also thank the staff for providing technical support by using the facility of the Institute of Health Sciences and Technology of Anhui University.

AUTHOR CONTRIBUTIONS

TX, XT, and JZ developed the original concepts. TX, XT, YH, and NZ designed the experiments. TX conducted the genome mining analyses. XT and TX carried out the experiments for mutant construction, phenotype determination, microarray and ChIP-seq. XT, DW, and CP carried out the IP-MS experiment. HH, SZ, XL, JW, NZ, and YH carried out all biochemical experiments including protein purification, EMSA, ITC, and phosphorylation site determination. HH, HG, and NZ carried out the crystallization and collected the data. NZ and YH solved the structures. DN assisted with the statistical analysis. TX, XT, YH, and NZ wrote the paper. MLK and JZ edited the manuscript. All authors were given the opportunity to review the results and comment on the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41396-023-01392-2>.

Correspondence and requests for materials should be addressed to Nannan Zhang, Yong-Xing He or Jizhong Zhou.

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

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.