Ribulokinase and Transcriptional Regulation of Arabinose Metabolism in *Clostridium acetobutylicum*

Lei Zhang,^a Semen A. Leyn,^{b,c} Yang Gu,^a Weihong Jiang,^a Dmitry A. Rodionov,^{b,c} and Chen Yang^a

Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China^a; Sanford-Burnham Medical Research Institute, La Jolla, California, USA^b; and Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia^c

The transcription factor AraR controls utilization of L-arabinose in *Bacillus subtilis*. In this study, we combined a comparative genomic reconstruction of AraR regulons in nine *Clostridium* species with detailed experimental characterization of AraR-mediated regulation in *Clostridium acetobutylicum*. Based on the reconstructed AraR regulons, a novel ribulokinase, AraK, present in all analyzed *Clostridium* species was identified, which was a nonorthologous replacement of previously characterized ribulokinases. The predicted function of the *araK* gene was confirmed by inactivation of the *araK* gene in *C. acetobutylicum* and biochemical assays using purified recombinant AraK. In addition to the genes involved in arabinose utilization and arabinoside degradation, extension of the AraR regulon to the pentose phosphate pathway genes in several *Clostridium* species was revealed. The predicted AraR-binding sites in the *C. acetobutylicum* genome and the negative effect of L-arabinose on DNA-regulator complex formation were verified by *in vitro* binding assays. The predicted AraR-controlled genes in *C. acetobutylicum* were experimentally validated by testing gene expression patterns in both wild-type and *araR*-inactivated mutant strains during growth in the absence or presence of L-arabinose.

The genus *Clostridium* is a diverse group of low-GC Gram-positive anaerobes that includes a large number of species important for cellulose degradation, development of renewable energy sources, and biotechnology. Many of these species are saprophytic organisms found in the soil (15). Among them, *Clostridium acetobutylicum* is one of the best-studied clostridia and was used to develop an industrial fermentation process for producing acetone and butanol (5, 22). This strain is known to utilize a broad range of monosaccharides, disaccharides, starches, and other polysaccharides (2, 20).

L-Arabinose is a major component of polysaccharides in plant cell walls, and its utilization pathway in bacteria has been investigated extensively (19). L-Arabinose can be transported into the cell through the ABC transport system AraFGH or arabinose-proton symporter AraE (Fig. 1A). L-Arabinose can also be obtained through hydrolysis of arabinosides. In some bacteria, such as *Bacillus subtilis* and *Geobacillus stearothermophilus*, arabinosides are transported into the cell and further degraded into L-arabinose by intracellular enzymes, including α -arabinofuranosidase AbfA and arabinosidase Arb43 (glycoside hydrolase family 43) (8, 36). L-Arabinose is consecutively converted to L-ribulose, L-ribulose-5-phosphate, and xylulose-5-phosphate by the action of L-arabinose isomerase AraA, L-ribulokinase AraB, and L-ribulose-5-phosphate 4-epimerase AraD, respectively, in many bacteria (e.g., *B. subtilis*). D-Xylulose-5-phosphate is further catabolized to form the central metabolic intermediates glyceraldehyde-3-phosphate and acetyl phosphate by the pentose phosphate pathway enzymes, including transketolase (Tkt), transaldolase (Tal), and phosphoketolase (Ptk).

Our current knowledge of transcriptional regulation of L-arabinose utilization pathway in low-GC, Gram-positive bacteria is mostly based on the studies in *B. subtilis*. The arabinose utilization in *B. subtilis* is controlled by the transcription factor AraR, which consists of a GntR-type DNA-binding domain in the N-terminal region and a C-terminal effector-binding domain ho-

mologous to the GalR/LacI family of repressors (21, 32). In the absence of the effector arabinose, AraR represses the expression of 13 genes, including the *araR* gene, the genes involved in arabinose catabolism (*araABD* and *araE*) and arabinoside degradation (*araNPQ*, *abfA*, *abnA*, and *abfA2*), and two genes with unknown functions (*araLM*) (27). The DNA-binding sites of AraR in the promoter regions of these genes have been characterized in *B. subtilis* (9).

Several *Clostridium* species have been shown to metabolize L-arabinose in early studies and our preliminary analysis (7, 26). The initial genomic survey of *Clostridium acetobutylicum* ATCC 824 identified all of the genes involved in arabinose utilization, except for the gene encoding ribulokinase AraB (22). Our previous study has tentatively predicted a candidate gene coding for an alternative ribulokinase (termed AraK), which is not orthologous to the known ribulokinases (30). However, this prediction has yet to be verified experimentally. Moreover, the regulatory mechanism of arabinose catabolism in *Clostridium* species remains unclear. Although the AraR-binding sites in the *C. acetobutylicum* genome have been previously predicted by bioinformatics analysis (30), the rapidly growing number of complete genomes in the *Clostridium* genus allows significant improvement of the accuracy of prediction of AraR-binding DNA motifs and expansion of AraR regulons. The bioinformatics predictions generated by such com-

Received 22 September 2011 Accepted 18 December 2011

Published ahead of print 22 December 2011

Address correspondence to C. Yang, chenyang@sibs.ac.cn, or D. A. Rodionov, rodionov@burnham.org.

L. Zhang and S. A. Leyn contributed equally to this work.

Supplemental material for this article may be found at <http://j.b.asm.org/>.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JB.06241-11

FIG 1 Reconstruction of AraR regulons in *Clostridium* species. (A) Metabolic context of the reconstructed clostridial Ara regulons. Matching colors are used to mark components of three distinct pathways related to arabinose metabolism. The AraR regulon members predicted in this work are in red boxes. (B) Maximum likelihood phylogenetic tree and inferred DNA recognition motifs of the AraR regulator. The Ara regulon was reconstructed in this work for the nine *Clostridium* genomes marked in red. The known AraR-binding motif in *B. subtilis* is also shown. (C) Genomic organization of the *araR*-containing loci and AraR-regulated genes in *Clostridium* genomes. Genes (shown by arrows) from the same metabolic pathway are marked in matching colors. The *araR* genes and candidate AraR-binding sites are shown by red arrows and circles, respectively. The predicted *araK* gene is marked by asterisks.

parative genomic analyses can be confirmed by further experimental studies. Previously, we have applied this integrated approach to predict and validate novel metabolic enzymes and transcriptional regulons involved in xylose utilization in *Clostridium* (12).

In this study, we used a comparative genomic approach to identify AraR-binding DNA motifs and reconstruct AraR regulons in nine biotechnologically important *Clostridium* species. This approach combined metabolic reconstruction with analysis of conserved AraR-binding sites in multiple genomes (29). The novel ribulokinase AraK was identified in all analyzed *Clostridium* genomes, and this functional assignment was experimentally confirmed through a combination of genetic and biochemical techniques. The extension of the AraR regulon to the pentose phosphate pathway genes was identified in several *Clostridium* species and experimentally validated in *C. acetobutylicum*.

MATERIALS AND METHODS

Genome resources and bioinformatics tools. Genome sequences of *Clostridium* spp. analyzed in this study were obtained from GenBank (<http://www.ncbi.nlm.nih.gov/GenBank/>). Identification of orthologs and gene neighborhood analysis were performed in MicrobesOnline (<http://www.microbesonline.org/>) (4). Functional annotations of genes involved in arabinose metabolism and related pathways were derived from the SEED comparative genomic database (<http://theseed.uchicago.edu/FIG/index.cgi>) (25). Phylogenetic trees were built using the maximum likelihood method in the PHYLIP package and visualized with Dendroscope (13). Sequence alignments were made with MUSCLE (6). Comparative genomic reconstruction of regulons was performed using the RegPredict web server (<http://regpredict.lbl.gov>) (24). Sequence logos for regulatory motifs were constructed using the WebLogo package (3).

Reconstruction of regulons. For identification of regulatory DNA motifs of AraR, we started from training sets of known AraR-regulated genes in *B. subtilis* and their orthologs in multiple *Clostridium* genomes. An iterative motif detection algorithm implemented in RegPredict was used to identify common regulatory DNA motifs in upstream regions of these genes (reviewed in reference 29). For each clade of AraR proteins on the phylogenetic tree, a separate training gene set was used. A positional weight matrix was constructed for each identified motif and used to scan the genomes in this clade and identify candidate AraR-binding sites. Scores of candidate sites were calculated as the sum of positional nucleotide weights. The score threshold was defined as the lowest score observed in the training set. A gene having a site score over the threshold was included in the AraR regulon if its regulation was conserved in a set of related genomes. The details of the reconstructed AraR regulons are captured and displayed in the RegPrecise database (<http://regprecise.lbl.gov>) (23).

Strains and growth conditions. The *C. acetobutylicum* strains and plasmids used in this study are given in Table 1. *Escherichia coli* strains DH5 α (Invitrogen, Carlsbad, CA) and ER2275 (18) were used for gene cloning, and BL21(DE3)pLysS (Invitrogen, Carlsbad, CA) was used for protein overexpression. The *E. coli* K-12 strain that carries the plasmid harboring the *araB* (b0063) gene with the T5 promoter and the His₆ tag was a kind gift from H. Mori at the Nara Institute of Science and Technology, Japan (16). *E. coli* strains were grown on LB medium. Kanamycin (30 μ g ml⁻¹), chloramphenicol (30 μ g ml⁻¹), spectinomycin (100 μ g ml⁻¹), and isopropyl- β -D-thiogalactopyranoside (IPTG; 0.2 mM) were added as appropriate. *C. acetobutylicum* strains were routinely cultured at 37°C in CGM medium (38). Erythromycin (30 μ g ml⁻¹) was added when needed. Solid medium was made with CGM medium containing 2% (wt/vol) agar A (Sangong Corp., Shanghai, China). For reverse transcription (RT)-PCR analysis of gene expression, *C. acetobutylicum* strains were grown in a medium without regulatory carbon sources, which contains (per liter) 1.5 g of K₂HPO₄, 1.5 g of KH₂PO₄, 2.2 g of CH₃COONH₄, 30 g

TABLE 1 *C. acetobutylicum* strains and plasmids used in this study

Strain or plasmid	Description	Source or reference ^a
<i>C. acetobutylicum</i> strains		
ATCC 824	Wild type	ATCC
<i>araK</i> ::intron mutant	<i>araK</i> -inactivated mutant; Em ^r	This study
<i>araR</i> ::intron mutant	<i>araR</i> -inactivated mutant; Em ^r	This study
<i>araK</i> ::intron pSY9- <i>araK</i> strain	<i>araK</i> -complemented strain; Cm ^r	This study
Plasmids		
pWJ1	Group II intron	39
pWJ1- <i>araK</i>	pWJ1 derivative for inserting intron into <i>araK</i>	This study
pWJ1- <i>araR</i>	pWJ1 derivative for inserting intron into <i>araR</i>	This study
pSY9	pIMP1-derived vector; Amp ^r Cm ^r	28
pSY9- <i>araK</i>	pSY9 derivative carrying CAC1344 gene	This study
pET28a		Novagen
pET28a- <i>araK</i>	pET28a derivative carrying CAC1344 gene	This study
pET28a- <i>araR</i>	pET28a derivative carrying CAC1340 gene	This study

^a ATCC, American Type Culture Collection.

of glycerol, 30 g of sodium pyruvate, 0.1 g of L-cysteine, 0.38 g of MgSO₄ · H₂O, 20 mg of MnSO₄ · H₂O, 20 mg of FeSO₄ · 7H₂O, 1 mg of *p*-aminobenzoic acid, 1 mg of vitamin B₁, 1 mg of biotin, and 6 g of yeast extract. L-Arabinose was supplied at a concentration of 30 g liter⁻¹ as indicated. For growth phenotype assays, *C. acetobutylicum* strains were cultivated at 37°C in triplicates in 60 ml of P2 minimal medium (1) supplemented with 5 g liter⁻¹ L-arabinose or D-xylose as the sole carbon source. Cell growth was monitored spectrophotometrically at 600 nm. Arabinose and xylose were detected by high-pressure liquid chromatography (HPLC) using an Agilent model 1200 instrument equipped with a Waters Sugar Pak I column (6.5 by 300 mm) and a refractive index detector (Agilent). Double-distilled water was used as the mobile phase at a flow rate of 0.6 ml min⁻¹, and the column was operated at 60°C.

Gene disruption in *C. acetobutylicum*. Gene disruption in *C. acetobutylicum* ATCC 824 was performed by using group II intron-based targetron technology as described previously (35). Briefly, the 350-bp fragments for retargeting introns to insert within the *araK* (CAC1344) or *araR* (CAC1340) genes, respectively, were generated by one-step assembly PCR using the primers shown in Table S1 in the supplemental material according to the protocol of the TargeTron gene knockout system (Sigma). The PCR products were then digested and ligated to a targetron plasmid, pWJ1 (39), yielding the plasmids pWJ1-*araK* and pWJ1-*araR*. Each plasmid was methylated *in vivo* in *E. coli* ER2275(pAN1) (18) and electroporated into *C. acetobutylicum* ATCC 824. The transformants were selected on CGM plates supplemented with erythromycin. The resulting mutants with intron insertion in *araK* or *araR* genes were confirmed by PCR (see Fig. S1 in the supplemental material). Southern blot analysis with probes against the intron was used to confirm that the intron was incorporated only once into the genome (see Fig. S1 in the supplemental material).

For genetic complementation experiments, the *araK* gene from *C. acetobutylicum* was cloned into the pSY9 vector (28) under the control of the constitutive *P*_{ptb} promoter (37). PCR was carried out using the *C. acetobutylicum* ATCC 824 genomic DNA and the primers shown in Table S1 in the supplemental material. The obtained plasmid, pSY9-*araK*, was electroporated into the *araK*-inactivated mutant, generating the *araK*-complemented strain. As a negative control, the pSY9 vector was also expressed in the *araK*-inactivated mutant.

RNA isolation and real-time RT-PCR. Total RNA was isolated from *C. acetobutylicum* ATCC 824 grown in medium without regulatory carbon sources or with addition of arabinose, as described above. Cells were harvested at mid-exponential growth phase (~24 h), frozen immediately in liquid nitrogen, and ground into powder. RNA was isolated using TRIzol (Invitrogen, Carlsbad, CA) by following the manufacturer's instructions. Contaminant DNA was removed by DNase I (Takara) digestion, which was verified by performing the PCR under identical conditions without adding reverse transcriptase. cDNA was generated by reverse transcription reactions using random hexamers as primers, 1 μ g purified RNA, and Moloney murine leukemia virus (MMLV) (RNase H⁻) reverse transcriptase (Takara). The cDNA was amplified using the Applied Biosystems 7300 real-time PCR system. The reaction mixture (20 μ l) contained 50 to 100 ng cDNA, 0.2 μ M gene-specific primers (as shown in Table S1 in the supplemental material), and Power SYBR green PCR master mix (Applied Biosystems). The PCR parameters were 1 cycle of 95°C for 30 s, followed by 40 cycles of 95°C for 5 s and 60°C for 31 s. Melt curves were analyzed to ensure specificity of primer annealing and lack of primer secondary structures. Data analysis was performed with the 7300 system software (Applied Biosystems). The CAC2679 gene encoding a pullulanase was used as an internal control for normalization, because its transcript level did not change when *C. acetobutylicum* was grown on various carbohydrates, including arabinose (34). The expression level of each gene was presented as the average of six measurements from two biological replicates, with the corresponding standard deviation.

Northern blot analysis. Total RNA was isolated from *C. acetobutylicum* ATCC 824 grown in P2 minimal medium supplemented with 30 g liter⁻¹ L-arabinose. The RNA (20 μ g) was subjected to electrophoresis on 1.5% agarose gel containing 0.55 M formaldehyde and transferred to positively charged nylon membranes (Roche Applied Science) according to standard methods (31). Digoxigenin-labeled DNA probes for *araK* or *tal* were prepared by PCR using the primers shown in Table S1 in the supplemental material and the Roche digoxigenin (DIG) gel shift kit (Roche Applied Science). After hybridization and washes, an anti-DIG antibody was incubated with the blots and then detected by chemiluminescence with the CDP-STAR reagent kit (Roche Applied Science).

Protein overexpression and purification. The *araR* (CAC1340) and *araK* (CAC1344) genes from *C. acetobutylicum* were PCR amplified (primers shown in Table S1 in the supplemental material) and cloned into the expression vector pET28a. The recombinant proteins were overexpressed as N-terminal fusions with a His₆ tag in *E. coli* BL21(DE3)pLysS. The cells were grown on LB medium to an optical density at 600 nm (OD₆₀₀) of 0.8 at 37°C, induced by 0.2 mM IPTG, and harvested after 12 h of shaking at 16°C. Protein purification was performed using a rapid Ni-nitrilotriacetic acid (NTA) agarose minicolumn protocol as described previously (40). Briefly, harvested cells were resuspended in 20 mM HEPES buffer (pH 7) containing 100 mM NaCl, 0.03% Brij-35, 2 mM β -mercaptoethanol, and 2 mM phenylmethylsulfonyl fluoride. Lysozyme was added to a concentration of 1 mg ml⁻¹, and the cells were lysed by freeze-thawing followed by sonication. After centrifugation, the supernatant was loaded onto a Ni-NTA agarose column (0.2 ml). After bound proteins were washed with 50 mM Tris-HCl buffer (pH 8) containing 1 M NaCl, 0.3% Brij-35, and 2 mM β -mercaptoethanol, they were eluted with 0.3 ml of the same buffer supplemented with 250 mM imidazole. The buffer was then changed to 10 mM Tris-HCl (pH 7.4) containing 0.3 mM dithiothreitol (DTT), 1 mM EDTA, and 10% glycerol by using Bio-Spin columns (Bio-Rad). The purified proteins were run on a 12% sodium dodecyl sulfate-polyacrylamide gel to monitor their size and purity (see Fig. S2 in the supplemental material).

EMSA. For the electrophoretic mobility shift assay (EMSA), the 180-bp DNA fragments from the upstream region of individual genes were PCR-amplified from *C. acetobutylicum* ATCC 824 genomic DNA using the primers shown in Table S1 in the supplemental material. Both forward and reverse primers were Cy5 fluorescence labeled at the 5' end (Sangong Corp., Shanghai, China). The PCR products were purified with

the PCR purification kit, and their concentration was determined spectrophotometrically. The fluorescence-labeled DNA (1 nM) was incubated with the indicated amount of purified AraR protein in 20 μ l of binding buffer containing 20 mM Tris (pH 7.5), 0.25 mM DTT, 10 mM MgCl₂, 5% glycerol, 0.8 μ g bovine serum albumin (BSA), and 1 μ g salmon sperm DNA (nonspecific competitor). As potential effectors of AraR-DNA binding, L-arabinose or D-xylose was added as indicated. After incubation at 30°C for 20 min, the reaction mixture was electrophoresed at 4°C on a 6% native polyacrylamide gel in 0.5 $\times$ Tris-borate-EDTA for 1.5 h at 100 V. Fluorescence-labeled DNA on the gel was then detected with the Starion FLA-9000 (FujiFilm, Japan). Specificity of the AraR-DNA interactions was tested by including a 400-fold molar excess of nonlabeled target DNA (specific competitor) in binding reaction mixtures.

Enzyme assays. Ribulokinase activity was assayed by coupling the formation of ADP to the oxidation of NADH to NAD⁺ via pyruvate kinase and lactate dehydrogenase. Briefly, 0.5 μ g of purified enzyme was added to 200 μ l of 50 mM Tris buffer (pH 7.5) containing 20 mM MgCl₂, 1 mM ATP, 2 mM phosphoenolpyruvate, 0.3 mM NADH, 1.2 U of pyruvate kinase, 1.2 U of lactate dehydrogenase, and 1 mM L-ribulose (Sigma-Aldrich). The change in NADH absorbance was monitored at 340 nm at 30°C by using a Beckman DU-800 spectrophotometer. Another protein (CAC1342) overexpressed and purified in the same way was used as a negative control. To test substrate specificity, L-ribulose was replaced by 5 mM D-glucose, D-fructose, D-ribose, and D-xylulose in the assay mixture. For determination of the apparent k_{cat} and the K_m values for L-ribulose, its concentration was varied in the range of 0.5 to 10 mM in the presence of 2 mM (saturating) ATP. Kinetic data were analyzed using Graphpad Prism 5.0 software. Protein concentrations were measured by using the Bradford reagent (Sangong Corp., Shanghai, China) with BSA as a standard.

The reaction product of ribulokinase was identified by HPLC using an Agilent model 1200 system. The reaction mixture (200 μ l) containing 200 mM HEPES (pH 7.5), 20 mM MgCl₂, 1 mM ATP, and 5 mM L-ribulose, and the purified ribulokinase (5 μ g) was incubated for 2 h at 37°C. Samples were applied at 70°C into a Waters Sugar Pak I column linked to a refractive index detector and eluted with double-distilled water at a flow rate of 0.6 ml min⁻¹.

RESULTS

Comparative genomic reconstruction of AraR regulons in *Clostridium* spp. AraR has been characterized in *B. subtilis* as a repressor of the L-arabinose metabolic operons (32). Orthologs to the *B. subtilis* *araR* were identified by BLAST searches against a nonredundant set of sequenced bacterial genomes (Fig. 1). A strong tendency of *araR* genes to cluster on the chromosome with arabinose utilization genes was observed. The phylogenetic tree was constructed for representative AraR proteins from *Bacillales*, *Lactobacillales*, and *Clostridiales* (Fig. 1B). Two separate groups were observed for clostridial AraR. AraR proteins in *C. acetobutylicum* and *Clostridium ljungdahlii* were located on a distinct branch from AraR in other clostridia such as *Clostridium beijerinckii* and *Clostridium cellulolyticum*. Two diverged AraR paralogs in *Clostridium carboxidivorans* belonged to these two clostridial branches, respectively.

To reconstruct the AraR regulons in *Clostridium* species, we applied the integrative comparative genomics approach (as implemented in RegPredict web server) that combines identification of candidate transcription factor-binding sites with cross-genomic comparison of regulons and with the functional context analysis of candidate target genes. The upstream regions of arabinose utilization genes in nine *Clostridium* genomes were analyzed using a motif recognition program to identify conserved AraR-binding DNA motifs. After construction of a positional weight matrix for each identified motif, we searched for additional AraR-binding sites in the analyzed

FIG 2 Maximum likelihood phylogenetic tree of selected FGGY family sugar kinases. The colors reflect chromosomal clustering of kinase genes with the genes from the respective sugar utilization pathways. Experimentally characterized sugar kinases are indicated.

Clostridium genomes and finally performed a consistency check or cross-species comparison of the predicted AraR regulons.

As shown in Fig. 1B, the identified palindromic AraR-binding DNA motif in *C. acetobutylicum* and *C. ljungdahlii* was partially similar to the AraR-binding sequences experimentally determined in *B. subtilis* (5'-ATTGTACGTACAAAT-3') (9). A relatively divergent binding motif was identified for AraR from other *Clostridium* species, including *C. beijerinckii*, *C. cellulolyticum*, and *Clostridium saccharolyticum* (5'-ATGTTGTACGTACAAGTT-3'), although the central core region of the motif was conserved. The obtained AraR-binding motifs were used to detect candidate members of the AraR regulons in the nine *Clostridium* genomes (Fig. 1; see Table S2 in the supplemental material). The most conserved part of the AraR regulons included the arabinose utilization genes *araA*, *araB*, *araD*, and transporter genes *araE* or *araFGH*. The presence of AraR-binding sites in the upstream region of *araR* gene suggested autoregulation of its expression in *Clostridium*.

Interestingly, a conserved member of AraR regulons in the nine *Clostridium* species was a candidate gene encoding an alternative ribulokinase, termed AraK. This novel ribulokinase is a nonorthologous replacement of the ribulokinase AraB from *B. subtilis*. Both AraK and AraB proteins belong to the FGGY sugar kinase family (Pfam accession number PF00294), which displays remarkable diversity in substrate specificity (Fig. 2). Phylogenetic analysis of the FGGY kinase family showed that the branches containing orthologs of AraB and AraK, respectively, were highly diverged from each other (Fig. 2). Because of the homology with some xylulokinases, the AraK protein in *C. acetobutylicum* (CAC1344) was incorrectly annotated as xylulokinase in the pub-

lic databases (e.g., GenBank). Here the functional assignment of ribulokinase in the nine *Clostridium* genomes was supported by chromosomal clustering with arabinose utilization genes and by identification as AraR regulon members (Fig. 1). Additionally, an alternative arabinose isomerase (AraA-II) from the fucose isomerase FucI family was predicted for *C. cellulolyticum*, *C. carboxidivorans*, and *Clostridium papyrosolvens* (Fig. 1).

The AraR regulons in *C. acetobutylicum*, *C. carboxidivorans*, *C. beijerinckii*, and *C. cellulovorans* extended to the pentose phosphate pathway in the central carbon metabolism, in addition to the known AraR targets in *B. subtilis* (Fig. 1). Candidate AraR-binding sites were identified in the regulatory region of *tkt*, *tal*, and *ptk* genes that were involved in conversion of xylulose-5-phosphate to central metabolic intermediates glyceraldehyde-3-phosphate and/or acetyl phosphate (Fig. 1A). The AraR regulon members also include the genes involved in uptake and degradation of arabinosides (*araT*, *abfA*, and *arb43*). A member of the aldose 1-epimerase family in the AraR regulon was tentatively assigned the functional role of arabinose mutarotase (Epi), which interconverts alpha and beta anomers of L-arabinose.

In summary, the comparative genomics analysis allowed us to reconstruct the AraR regulons in *C. acetobutylicum* and eight other *Clostridium* genomes and to tentatively assign the candidate genes for an alternative ribulokinase. We then performed experimental validation of the predicted function of *araK* and characterization of the AraR regulon in *C. acetobutylicum* as described below.

Experimental validation of the predicted ribulokinase in *C. acetobutylicum*. To validate the inferred function of the *araK* gene in *C. acetobutylicum*, we disrupted this gene by inserting an intron (confirmed by PCR and Southern blot analysis, as shown in Fig. S1 in the supplemental material) and tested the ability of the mutant to grow on arabinose as the sole carbon source. Inactivation of the *araK* gene severely impaired the growth on arabinose (Fig. 3A). The growth rate of the *araK*-inactivated mutant (0.0014 OD h^{-1}) was about 90-fold lower than that of the wild type (0.13 OD h^{-1}). Complementation of the *araK*-inactivated mutant by using a plasmid construct constitutively expressing *araK* restored the growth on arabinose. On the other hand, the growth rate of the *araK*-inactivated mutant on xylose was not diminished (Fig. 3B). Therefore, the phenotype of the *C. acetobutylicum* *araK*-inactivated mutant confirmed the predicted physiological role of the *araK* gene in arabinose utilization. There may be a second low-activity kinase that could also be used for arabinose utilization.

To extend the genetic findings and provide biochemical evidence to the proposed gene assignment, we used the recombinant AraK from *C. acetobutylicum*, which was overexpressed in *E. coli* with the N-terminal His₆ tag and purified using Ni-NTA affinity chromatography, to test for ribulokinase activity (Fig. 4A). As expected, the *C. acetobutylicum* AraK displayed a ribulokinase activity with the k_{cat} of 41 s^{-1} . The apparent K_m value for L-ribulose (0.96 mM) was higher than the value reported for *E. coli* ribulokinase AraB (0.14 mM) (17). Furthermore, the substrate specificity of *C. acetobutylicum* AraK was tested. It was highly specific to L-ribulose and incapable of phosphorylating D-xylulose (Fig. 4A), D-glucose, D-fructose, and D-ribose (data not shown). Finally, the conversion of L-ribulose to L-ribulose 5-phosphate by AraK was directly detected using HPLC (Fig. 4B). The reaction mixture contained L-ribulose, ATP, and purified recombinant AraK from *C. acetobutylicum*, and the formation of L-ribulose 5-phosphate was identified by HPLC. The ribulokinase from *E. coli* was assayed in

FIG 3 Effect of *araK* gene disruption on cell growth of *C. acetobutylicum* on L-arabinose (A) and D-xylose (B). Cells were grown in P2 minimal medium containing 5 g liter⁻¹ of L-arabinose or D-xylose as the sole carbon source. The optical density at 600 nm (open symbols) and the sugar concentration in the medium (filled symbols) were monitored.

parallel as a control. These biochemical assays provided an independent verification of the predicted enzymatic activity of AraK from *C. acetobutylicum*.

Experimental characterization of the predicted AraR regulon in *C. acetobutylicum*. To validate the regulatory role of the inferred *araR* gene in L-arabinose metabolism and the predicted AraR regulon, the *C. acetobutylicum* *araR*-inactivated mutant was constructed. Total RNA was isolated from both the wild-type and *araR*-inactivated mutant strains grown in the medium without regulatory carbon sources (as described in Materials and Meth-

ods) or with addition of arabinose. By using quantitative RT-PCR, we examined the relative transcript levels of predicted AraR-controlled genes in both strains during growth in the absence or presence of arabinose (Fig. 5A). In the absence of arabinose, the expression levels of the *araK* (CAC1344), *araA* (CAC1342 and CAC1346), *araD* (CAC1341), and *araE* (CAC1339 and CAC1345) genes in the *araR*-inactivated mutant were elevated more than 17-fold compared to that of the wild-type strain. Several of these genes were expressed at high levels in the *araR*-inactivated mutant regardless of the presence or absence of arabinose, whereas addi-

FIG 4 Biochemical characterization of a recombinant AraK protein. (A) Ribulokinase activity was assayed by coupling the formation of ADP to the oxidation of NADH to NAD⁺ and monitoring the absorbance at 340 nm. For testing substrate specificity, L-ribulose was replaced by D-xylulose in the assay mixture. No activity was detected in a control experiment using AraA (CAC1342). (B) HPLC analysis of a reaction mixture containing L-ribulose and ATP before (0 h) and 2 h after adding recombinant AraK from *C. acetobutylicum* or ribulokinase (AraB) from *E. coli*. RIU, refractive index units.

FIG 5 Transcriptional analysis of AraR regulon genes. (A) Quantitative RT-PCR analysis of gene expression. Total RNA was isolated from *C. acetobutylicum* ATCC 824 wild-type and *araR*-inactivated mutant strains grown in the medium without regulatory carbon sources or with addition of L-arabinose (30 g liter⁻¹). The expression levels of each gene were normalized to the gene expression in the wild-type strain grown in the absence of arabinose. (B) Northern blot analysis of the *araK-araE-araA-tal-tkt-epi* transcript. Total RNA was extracted from wild-type *C. acetobutylicum* grown in P2 minimal medium supplemented with L-arabinose (30 g liter⁻¹). The RNA was subjected to electrophoresis and analyzed with digoxigenin-labeled DNA probes for *araK* or *tal*. The open arrowheads indicate the positions of 23S and 16S rRNAs.

tion of arabinose to the culture caused a significant induction of the genes' expression in the wild-type strain.

In addition to the genes of arabinose utilization pathway, the transcription of the pentose phosphate pathway genes *tkt* (CAC1348), *tal* (CAC1347), and *ptk* (CAC1343) and the arabinoside degradation genes *epi* (CAC1349) and *arb43* (CAC1529) was also controlled by AraR (Fig. 5A). In fact, the most prominent effect of *araR* mutation was observed for the *ptk* gene, which was upregulated over 1,000-fold in the *araR*-inactivated mutant compared to the wild type when grown without arabinose. The expression levels of the *tkt* and *tal* genes in the *araR*-inactivated mutant in the absence of arabinose were over 23-fold higher than those in the wild-type strain grown in the same medium and about 3- to 4-fold higher than the levels measured in the arabinose-grown wild-type cells. All of the genes studied, especially the *tkt* and *tal* genes, in the *araR*-inactivated mutant exhibited reduced expression levels when arabinose was present in the medium. Although the underlying mechanisms are yet to be studied, these genes may also be controlled by another transcriptional factor. The impaired growth of the *araR*-inactivated mutant on arabinose was observed (data not shown), which may be caused by accumulation of toxic metabolic intermediates due to constitutive expression of arabinose utilization genes. Therefore, the quantitative RT-PCR results confirm that AraR is a negative regulator of 11 genes involved in the arabinose utilization, arabinoside degradation, and pentose phosphate pathway in *C. acetobutylicum*.

To determine whether the *araK-araE-araA-tal-tkt-epi* cluster constituted a transcriptional unit, Northern blot analysis was per-

formed. Total RNA was isolated from wild-type *C. acetobutylicum* grown with arabinose as the sole carbon source and was hybridized to DNA probes for *araK* or *tal* (Fig. 5B). A transcript of about 8 kb was obtained, which matched the expected size of the full-length *araK-araE-araA-tal-tkt-epi* transcript. In addition, the probes also hybridized to smaller transcripts, suggesting that there might be internal promoters within the gene cluster. The cotranscription of *araK* with *araE*, *araA*, *tal*, *tkt*, and *epi* genes is in line with the quantitative RT-PCR results.

To confirm the ability of AraR protein to specifically bind to the predicted DNA sites, EMSA was performed using the purified recombinant AraR protein from *C. acetobutylicum*. Five predicted target DNA fragments (180 bp) from the upstream region of the *C. acetobutylicum* *araE*, *araR*, *araD*, *ptk*, and *araK* genes, respectively, were tested by EMSA (Fig. 6A to E). A shifted band was observed upon incubation of AraR protein with the target DNA fragments, and its intensity was enhanced in the presence of increasing amounts of AraR protein. The band shift for all five target sites was essentially complete at an AraR concentration above 0.5 μ M. Different binding affinities of AraR protein to individual DNA fragments were observed. The AraR protein exhibited higher affinity for the tested upstream fragments of *ptk* and *araK* genes than for that of the *araR* gene. However, AraR may bind to multiple sites in the regulatory region between divergently transcribed *araR* and *araD* genes (Fig. 1C), and the cooperative binding could result in a high level of repression (21). To test the effector of AraR-DNA binding, different concentrations of arabinose were used in EMSA. As shown in Fig. 6F, the AraR protein was released from

FIG 6 Binding of AraR to the promoter region of the *ptk* (A), *araK* (B), *araE* (C), *araR* (D), and *araD* (E) genes. EMSA was performed by incubating the indicated amounts of purified *C. acetobutylicum* AraR protein with a Cy5 fluorescence-labeled 180-bp DNA fragment (1 nM). (F) EMSA of the *ptk* promoter region with AraR in the presence of increasing amounts of L-arabinose.

the AraR-DNA complex when the arabinose concentration was higher than 1 mM, whereas xylose (20 mM) had no effect on the binding of AraR to target DNA (see Fig. S3 in the supplemental material). The formation of the AraR-DNA complex was suppressed in the presence of 400-fold excess unlabeled DNA fragments but not in the presence of nonspecific competitor, salmon sperm DNA (see Fig. S3 in the supplemental material). These observations confirm that AraR is an arabinose-responsive regulator that binds specifically to the predicted DNA-binding sites in the *C. acetobutylicum* genome.

DISCUSSION

In this study, we performed comparative genomic reconstruction of AraR regulons in nine biotechnologically important *Clostridium* species by combining the identification of candidate AraR-binding sites with functional analysis of target genes. Although

two different types of AraR-binding DNA motifs were identified for *Clostridium* species, the consensus sequences shared the conserved bases (G₉, A₁₁, and T₁₆) that have been shown to be critical for AraR binding in *B. subtilis* (9). The combined bioinformatics with *in vitro* and *in vivo* characterization of AraR regulon in *C. acetobutylicum* confirmed that AraR repressed the expression of genes involved in arabinose utilization, arabinoside degradation, and the pentose phosphate pathway in the absence of arabinose. When arabinose was added to the culture, binding of arabinose to AraR inhibited the binding of AraR to target DNAs, leading to derepression of transcription of the above genes.

Based on the reconstructed AraR regulons, we have identified a novel ribulokinase AraK present in all analyzed *Clostridium* species, which is a nonorthologous replacement of previously characterized ribulokinase AraB. Disruption of the *araK* gene in *C.*

acetobutylicum and detailed biochemical characterization of purified recombinant AraK allowed us to confirm the predicted function. Interestingly, unlike in other *Clostridium* species, where AraK is the only ribulokinase, both AraK and AraB are present in *C. carboxidivorans*. This organism also possesses two AraR paralogs, and their DNA-binding sites were detected in the upstream region of the *araK* and *araB* genes, respectively. AraK and AraB, together with *E. coli* glycerol kinase, fuculokinase, gluconokinase, xylulokinase, and rhamnulokinase, belong to the FGGY sugar kinase family, which represents a striking example of variations in substrate specificity. Comparative sequence analysis and structural modeling could help us identify specificity-determining regions and to assess evolution of substrate specificity in this important yet poorly explored family (Y. Zhang, personal communication).

The CAC1339 to -1349 gene cluster in the *C. acetobutylicum* genome includes genes for two AraA paralogs, CAC1342 and CAC1346. Previous DNA microarray analysis has showed that the CAC1346 gene was induced by both arabinose and xylose, while the CAC1342 gene was strongly induced by arabinose and weakly induced by xylose (34). Here our results indicated that both arabinose isomerase genes are controlled by AraR. Moreover, we overexpressed both genes in *E. coli*, purified the recombinant proteins, and performed the biochemical assays. Both AraA proteins were highly specific to L-arabinose and cannot convert D-xylose to D-xylulose (data not shown). In addition, the CAC1339 to -1349 region also contains genes for two putative arabinose proton-symporters (CAC1339 and CAC1345), which are considered to be responsible for arabinose uptake. Our previous genetic mutagenesis and complementation experiments have showed that the CAC1345 gene product may also contribute to xylose transport (12). The exact interpretation of the observed functional redundancy in *C. acetobutylicum* requires further investigation.

Due to the absence of glucose-6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase, the pentose phosphate pathway in *C. acetobutylicum* mainly serves to synthesize biomass precursors, ribose-5-phosphate and erythrose-4-phosphate (33). Regulation of pentose phosphate pathway genes by AraR allows the cells to activate this pathway and catabolize arabinose efficiently when arabinose is available in the culture. Moreover, expression of arabinose transporters and metabolic enzymes in *C. acetobutylicum* is also under the control of AraR. Thus, the coordinated control of different interrelated pathways by AraR could result in efficient arabinose utilization. On the other hand, the transcriptional factor XylR controls the expression of only metabolic enzymes XylAB for xylose utilization, while the xylose transporter gene and pentose phosphate pathway genes are not members of the XylR regulon in *C. acetobutylicum* (12). This difference between AraR and XylR regulons may partly explain the much shorter lag phase of *C. acetobutylicum* grown on arabinose compared to growth on xylose. In addition to AraR regulation, expression of arabinose utilization genes in *B. subtilis* is also subject to catabolite repression mediated by the transcriptional factor CcpA, which binds to the catabolite responsive element (*cre*) (10, 14). The inhibition of arabinose metabolism in the presence of glucose (26) and induction of arabinose utilization genes upon glucose depletion (11) were also observed for *C. acetobutylicum*. Studies on the involvement of CcpA and *cre* in regulation of clostridial arabinose metabolism are now under way.

ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China (31070033 and 31121001), the National Basic Research Program of China (973: 2012CB721101), and the Knowledge Innovation Program of the Chinese Academy of Sciences (KSCX2-EW-G-5 and KSCX1-YW-11C3). The work of D.A.R. and S.A.L. was supported by the U.S. Department of Energy (DE-SC0004999) and Russian Academy of Sciences (program "Molecular and Cellular Biology").

REFERENCES

- Baer SH, Blaschek HP, Smith TL. 1987. Effect of butanol challenge and temperature on lipid composition and membrane fluidity of butanol-tolerant *Clostridium acetobutylicum*. Appl. Environ. Microbiol. 53:2854–2861.
- Bronnenmeier K, Staudenbauer WL. 1993. Molecular biology and genetics of substrate utilization in clostridia. Butterworth-Heinemann, Boston, MA.
- Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. Genome Res. 14:1188–1190.
- Dehal PS, et al. 2010. MicrobesOnline: an integrated portal for comparative and functional genomics. Nucleic Acids Res. 38:D396–D400.
- Dürre P. 2007. Biobutanol: an attractive biofuel. J. Biotechnol. 2:1525–1534.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.
- Ezeji T, Qureshi N, Blaschek HP. 2007. Butanol production from agricultural residues: impact of degradation products on *Clostridium beijerinckii* growth and butanol fermentation. Biotechnol. Bioeng. 97:1460–1469.
- Ferreira MJ, de Sá-Nogueira I. 2010. A multitask ATPase serving different ABC-type sugar importers in *Bacillus subtilis*. J. Bacteriol. 192:5312–5318.
- Franco IS, Mota LJ, Soares CM, de Sá-Nogueira I. 2007. Probing key DNA contacts in AraR-mediated transcriptional repression of the *Bacillus subtilis* arabinose regulon. Nucleic Acids Res. 35:4755–4766.
- Fujita Y. 2009. Carbon catabolite control of the metabolic network in *Bacillus subtilis*. Biosci. Biotechnol. Biochem. 73:245–259.
- Grimmler C, Held C, Liebl W, Ehrenreich A. 2010. Transcriptional analysis of catabolite repression in *Clostridium acetobutylicum* growing on mixtures of D-glucose and D-xylose. J. Biotechnol. 150:315–323.
- Gu Y, et al. 2010. Reconstruction of xylose utilization pathway and regulons in Firmicutes. BMC Genomics 11:255.
- Huson DH, et al. 2007. Dendroscope: an interactive viewer for large phylogenetic trees. BMC Bioinformatics 8:460.
- Inacio JM, de Sá-Nogueira I. 2007. Trans-acting factors and cis elements involved in glucose repression of arabinan degradation in *Bacillus subtilis*. J. Bacteriol. 189:8371–8376.
- Keis S, Bennett CF, Ward VK, Jones DT. 1995. Taxonomy and phylogeny of industrial solvent-producing clostridia. Int. J. Syst. Bacteriol. 45:693–705.
- Kitagawa M, et al. 2005. Complete set of ORF clones of *Escherichia coli* ASKA library (a complete set of *E. coli* K-12 ORF archive): unique resources for biological research. DNA Res. 12:291–299.
- Lee LV, Gerratana B, Cleland WW. 2001. Substrate specificity and kinetic mechanism of *Escherichia coli* ribulokinase. Arch. Biochem. Biophys. 396:219–224.
- Mermelstein LD, Papoutsakis ET. 1993. In vivo methylation in *Escherichia coli* by the *Bacillus subtilis* phage phi 3T I methyltransferase to protect plasmids from restriction upon transformation of *Clostridium acetobutylicum* ATCC 824. Appl. Environ. Microbiol. 59:1077–1081.
- Michal G. 1999. Biochemical pathways: an atlas of biochemistry and molecular biology. John Wiley & Sons, Inc., New York, NY.
- Mitchell WJ. 1998. Physiology of carbohydrate to solvent conversion by clostridia. Adv. Microb. Physiol. 39:31–130.
- Mota LJ, Tavares P, de Sá-Nogueira I. 1999. Mode of action of AraR, the key regulator of L-arabinose metabolism in *Bacillus subtilis*. Mol. Microbiol. 33:476–489.
- Nölling J, et al. 2001. Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. J. Bacteriol. 183:4823–4838.
- Novichkov PS, et al. 2010. RegPrecise: a database of curated genomic

- inferences of transcriptional regulatory interactions in prokaryotes. *Nucleic Acids Res.* **38**:D111–D118.
24. Novichkov PS, et al. 2010. RegPredict: an integrated system for regulon inference in prokaryotes by comparative genomics approach. *Nucleic Acids Res.* **38**:W299–W307.
 25. Overbeek R, et al. 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* **33**:5691–5702.
 26. Qureshi N, Li XL, Hughes S, Saha BC, Cotta MA. 2006. Butanol production from corn fiber xylan using *Clostridium acetobutylicum*. *Biotechnol. Prog.* **22**:673–680.
 27. Raposo MP, Inácio JM, Mota LJ, de Sá-Nogueira I. 2004. Transcriptional regulation of genes encoding arabinan-degrading enzymes in *Bacillus subtilis*. *J. Bacteriol.* **186**:1287–1296.
 28. Ren C, et al. 2010. Identification and inactivation of pleiotropic regulator CcpA to eliminate glucose repression of xylose utilization in *Clostridium acetobutylicum*. *Metab. Eng.* **12**:446–454.
 29. Rodionov DA. 2007. Comparative genomic reconstruction of transcriptional regulatory networks in bacteria. *Chem. Rev.* **107**:3467–3497.
 30. Rodionov DA, Mironov AA, Gelfand MS. 2001. Transcriptional regulation of pentose utilisation systems in the *Bacillus/Clostridium* group of bacteria. *FEMS Microbiol. Lett.* **205**:305–314.
 31. Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: a laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
 32. Sá-Nogueira I, Mota LJ. 1997. Negative regulation of L-arabinose metabolism in *Bacillus subtilis*: characterization of the *araR* (*araC*) gene. *J. Bacteriol.* **179**:1598–1608.
 33. Senger RS, Papoutsakis ET. 2008. Genome-scale model for *Clostridium acetobutylicum*. Part II. Development of specific proton flux states and numerically determined sub-systems. *Biotechnol. Bioeng.* **101**:1053–1071.
 34. Servinsky MD, Kiel JT, Dupuy NE, Sund CJ. 2010. Transcriptional analysis of differential carbohydrate utilization by *Clostridium acetobutylicum*. *Microbiology* **156**:3478–3491.
 35. Shao L, et al. 2007. Targeted gene disruption by use of a group II intron (targetron) vector in *Clostridium acetobutylicum*. *Cell Res.* **17**:963–965.
 36. Shulami S, et al. 2011. The L-arabinan utilization system of *Geobacillus stearothermophilus*. *J. Bacteriol.* **193**:2838–2850.
 37. Tummala SB, Welker NE, Papoutsakis ET. 1999. Development and characterization of a gene expression reporter system for *Clostridium acetobutylicum* ATCC 824. *Appl. Environ. Microbiol.* **65**:3793–3799.
 38. Wiesenborn DP, Rudolph FB, Papoutsakis ET. 1988. Thiolase from *Clostridium acetobutylicum* ATCC 824 and its role in the synthesis of acids and solvents. *Appl. Environ. Microbiol.* **54**:2717–2722.
 39. Xiao H, et al. 2011. Confirmation and elimination of xylose metabolism bottlenecks in glucose phosphoenolpyruvate-dependent phosphotransferase system-deficient *Clostridium acetobutylicum* for simultaneous utilization of glucose, xylose, and arabinose. *Appl. Environ. Microbiol.* **77**:7886–7895.
 40. Yang C, Rodionov DA, Rodionova IA, Li X, Osterman AL. 2008. Glycerate 2-kinase of *Thermotoga maritima* and genomic reconstruction of related metabolic pathways. *J. Bacteriol.* **190**:1773–1782.