AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology®

BIODEGRADATION

Isolation and Characterization of Levoglucosan-Metabolizing Bacteria

Ajay S. Arya,[a,b] Minh T. H. Hang,[a] Mark A. Eiteman[a,b]

[a]School of Chemical, Materials, and Biomedical Engineering, University of Georgia, Athens, Georgia, USA
[b]Department of Microbiology, University of Georgia, Athens, Georgia, USA

ABSTRACT Bacteria were isolated from wastewater and soil containing charred wood remnants based on their ability to use levoglucosan as a sole carbon source and on their levoglucosan dehydrogenase (LGDH) activity. On the basis of their 16S rRNA gene sequences, these bacteria represented the diverse genera *Microbacterium*, *Paenibacillus*, *Shinella*, and *Klebsiella*. Genomic sequencing of the isolates verified that two isolates represented novel species, *Paenibacillus athensensis* MEC069^T and *Shinella sumterensis* MEC087^T, while the remaining isolates were closely related to *Microbacterium lacusdiani* or *Klebsiella pneumoniae*. The genetic sequence of LGDH, *lgdA*, was found in the genomes of these four isolates as well as *Pseudarthrobacter phenanthrenivorans* Sphe3. The identity of the *P. phenanthrenivorans* LGDH was experimentally verified following recombinant expression in *Escherichia coli*. Comparison of the putative genes surrounding *lgdA* in the isolate genomes indicated that several other gene products facilitate the bacterial catabolism of levoglucosan, including a putative sugar isomerase and several transport proteins.

IMPORTANCE Levoglucosan is the most prevalent soluble carbohydrate remaining after high-temperature pyrolysis of lignocellulosic biomass, but it is not fermented by typical production microbes such as *Escherichia coli* and *Saccharomyces cerevisiae*. A few fungi metabolize levoglucosan via the enzyme levoglucosan kinase, while several bacteria metabolize levoglucosan via levoglucosan dehydrogenase. This study describes the isolation and characterization of four bacterial species that degrade levoglucosan. Each isolate is shown to contain several genes within an operon involved in levoglucosan degradation, furthering our understanding of bacteria that metabolize levoglucosan.

KEYWORDS levoglucosan dehydrogenase, *Pseudarthrobacter phenanthrenivorans*, pyrolysis, *Microbacterium*, *Klebsiella*, *Paenibacillus*, *Shinella*

During fast pyrolysis of lignocellulosic biomass, the complex carbohydrates and other compounds are dehydrated, and substantial water is generated as a by-product. At temperatures greater than 300°C, cellulose is readily depolymerized and dehydrated specifically to a mixture of anhydrosugars and their polysaccharides, the most abundant being 1,6-anhydro-β-D-glycopyranose, also known as levoglucosan (1–3), and the disaccharide of glucose and levoglucosan, cellobiosan (4). Although lignocellulose generates less levoglucosan than pure cellulose, impregnating woody lignocellulose with a 1% (wt/vol) phosphoric acid solution during treatment leads to a levoglucosan content as high as 33.6% (wt/wt) (5) partitioned primarily into an aqueous phase (6, 7).

Levoglucosan is a potentially fermentable sugar that is not naturally metabolized by typical production microorganisms (e.g., *Saccharomyces cerevisiae* and *Escherichia coli*). Chemically, levoglucosan can be hydrolyzed using sulfuric acid at 110°C to readily fermentable glucose (6, 8). Microorganisms metabolize levoglucosan by one of two

Editor Pablo Ivan Nikel, Novo Nordisk Foundation Center for Biosustainability

Address correspondence to Mark A. Eiteman, eiteman@engr.uga.edu.

The authors declare no conflict of interest.

Received 19 September 2021
Accepted 3 December 2021
Accepted manuscript posted online 15 December 2021
Published 22 February 2022

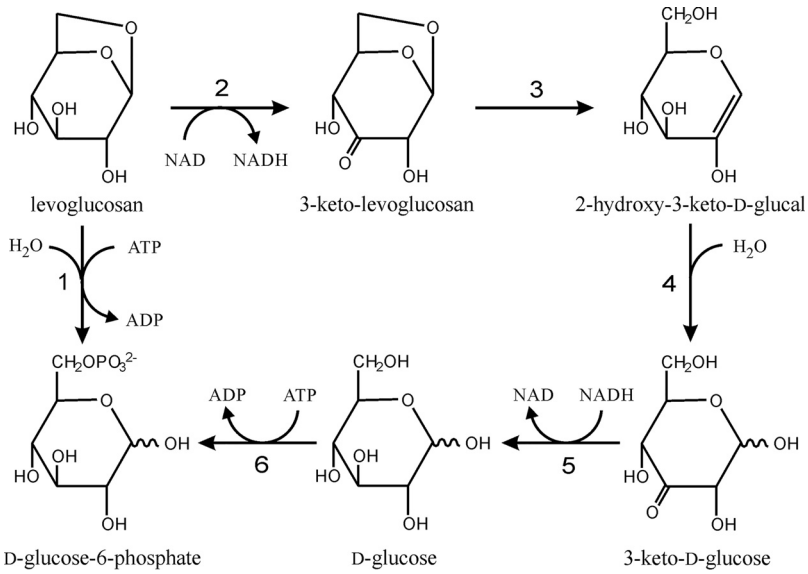

FIG 1 Metabolism of levoglucosan by levoglucosan kinase (enzyme 1) and proposed levoglucosan dehydrogenase (2) pathway leading to D-glucose (enzymes 2 to 5). Presumably glucokinase (6) converts intracellular D-glucose generated from levoglucosan to D-glucose-6-phosphate. The levoglucosan pathway is drawn as previously proposed for *B. smithii* (17).

strategies. In general, fungi use levoglucosan kinase (LGK) to hydrate and phosphorylate simultaneously the anhydro-sugar to glucose-6-phosphate (9–11), while bacteria are proposed first to oxidize levoglucosan to 3-keto-levoglucosan using levoglucosan dehydrogenase (LGDH) (12).

LGK is specific for levoglucosan (Fig. 1), and common hexoses, such as glucose, mannose, and galactose, are neither phosphorylated nor inhibit the enzyme (9). LGKs generally have poor affinity for levoglucosan, with a K_m of about 50 to 110 mM, while the K_m for ATP is 0.2 mM (9, 11). ADP, but not glucose-6-phosphate, is strongly inhibitory, having a K_i of 0.15 mM (9). Genes encoding LGK from *Aspergillus niger* (13) and *Lipomyces starkeyi* (14) have been expressed in *E. coli*. Multiple mutations in the LGK of *L. starkeyi* resulted in a 15-fold increase in growth rate on levoglucosan and an over 24-fold improvement in enzyme activity in an *E. coli* host (15).

Comparatively little is known about prokaryotic utilization of levoglucosan, and only *Arthrobacter* sp. strain I-552, *Pseudarthrobacter phenanthrenivorans* Sphe3, and *Bacillus smithii* have been studied in detail (12, 16–18). *B. smithii* has been proposed to follow a four-step conversion from levoglucosan to glucose in which levoglucosan is first oxidized via the NAD-dependent dehydrogenase (LGDH) to 3-keto-levoglucosan, a β-eliminase forms 2-hydroxy-3-keto-glucal, a hydratase forms 3-keto-glucose, and a NADH-dependent reductase finally produces glucose (Fig. 1) (17). Previously, *Arthrobacter* sp. strain I-552 and *P. phenanthrenivorans* Sphe3 were proposed to follow a three-step conversion in which LGDH oxidizes levoglucosan to form 3-keto-levoglucosan, which a hydratase converts to 3-keto-glucose and LGDH subsequently reduces using NADH to form glucose (12, 16). The partially purified LGDH from *Arthrobacter* sp. strain I-552 has a K_m of 14 mM for levoglucosan and a K_m of 0.47 mM for NAD (12). After cell lysis and debris removal by centrifugation, LGDH activity was detected in the supernatant fraction, suggesting that LGDH resides in the cytosol and that levoglucosan enters the cell membrane by some unknown mechanism. Furthermore, the expression of intracellular LGK in *E. coli* enables growth on levoglucosan as the sole carbon source (13, 14, 19), demonstrating that levoglucosan enters *E. coli* cells through existing nonspecific channels or transporters. Sugiura et al. identified the gene *lgdA* in *P. phenanthrenivorans* Sphe3, characterized the recombinant protein, and obtained crystal structures that resolved the active site of LGDH (18). Since the description of LGDH in *Arthrobacter* sp. strain I-552, additional bacteria

TABLE 1 Bacterial isolates able to metabolize levoglucosan as the sole carbon source at 37°C

Isolate designation	Source	Closest 16S rRNA BLAST match	% RNA sequence similarity
MEC069	Raw wastewater chemical plant, Athens, GA, USA	*Paenibacillus marchantiophytorum*	95.82
MEC084	Nantahala National Forest, NC, USA	*Microbacterium lacusdiani*	99.87
MEC087	Russell farmstead, SC, USA	*Shinella zoogloeoides*	98.70
MEC097	Raw wastewater, water reclamation center, Athens, GA, USA	*Klebsiella pneumoniae*	99.36

capable of growth on levoglucosan as their sole carbon source have been identified (20, 21). Based upon their 16S rRNA gene sequences, these isolates were identified as *Sphingobacterium multivorum*, *Acinetobacter oleivorans* JC3-1, *Enterobacter* sp. strain SJZ-6, *Microbacterium* sp. strain FXJ8.207, *Microbacterium* sp. strain FXJ8.203, *Enterobacter cloacae* DSM 16657, *Bacillus smithii* S-2701M, and *Parageobacillus thermoglucosidasius*. Recently, a gene cluster coding for four functional enzymes has been demonstrated in *B. smithii* S-2701M (17).

The goal of this study was to isolate and characterize additional bacteria that use levoglucosan as the sole carbon source. Furthermore, using the genomic sequence of each isolate, this study aimed to identify shared genes that likely play a role in levoglucosan catabolism. Levoglucosan has been previously reported to be a major component in the fine particulate emissions from the combustion of wood in residential fireplaces (22) and prescribed forest fires (23). We therefore reasoned that levoglucosan-degrading bacteria would be found at sites where lignocellulosic materials were burned and sought bacteria from soil samples near recent campfires and forest fires.

RESULTS

Isolation of levoglucosan-degrading bacteria. Upon enrichment in medium containing levoglucosan, samples from several soils associated with wood fires as well as from water treatment facilities yielded bacterial strains able to grow on levoglucosan as the sole carbon source (Table 1). Isolate MEC084 came from a mixture that was primarily charred wood with silty topsoil in a campfire pit (Nantahala National Forest, NC, USA). Isolate MEC087 was also from a sample of campfire sediment; however, this sample had browner silty topsoil and was heavy in organic matter (Russell Farmstead, Mountain Rest, SC, USA). The remaining two isolates came from wastewater: MEC069 was from the wastewater leaving a chemical plant (Athens, GA, USA), while MEC097 was isolated from the raw wastewater entering a wastewater treatment facility (Middle Oconee Wastewater Treatment Facility, Athens, GA, USA).

Initial characterization of isolates. The maximum specific growth rates on levoglucosan and levoglucosan-utilizing enzyme activities of the four isolates were determined (Table 2). MEC097 had a high growth rate on levoglucosan (~0.8 h^{-1}), while the remaining isolates achieved maximum specific growth rates of about 0.4 h^{-1}. All isolates possessed levoglucosan dehydrogenase (LGDH) activity, while no levoglucosan kinase activity was detected (<0.001 IU/mg). The greatest specific LGDH activity of 0.32 IU/mg dry cell weight was measured in the isolate showing the greatest specific growth rate (MEC097), while the remaining activities ranged between 0.024 and 0.090 IU/mg dry cell weight. We observed a correlation between the measured specific growth rate and the measured levoglucosan dehydrogenase specific activity (R^2 = 0.91).

TABLE 2 Maximum specific growth rates on levoglucosan and levoglucosan dehydrogenase activity for bacterial isolates

Isolate	Maximum specific growth rate[a] (h^{-1})	LGDH sp act[a] (IU/mg dry cell wt)
MEC069	0.47 (0.11)	0.057 (0.028)
MEC084	0.43 (0.02)	0.026 (0.006)
MEC087	0.40 (0.04)	0.090 (0.042)
MEC097	0.80 (0.10)	0.32 (0.16)

[a]Mean of 3 to 4 independent samples. Standard deviation of measurements is shown in parentheses.

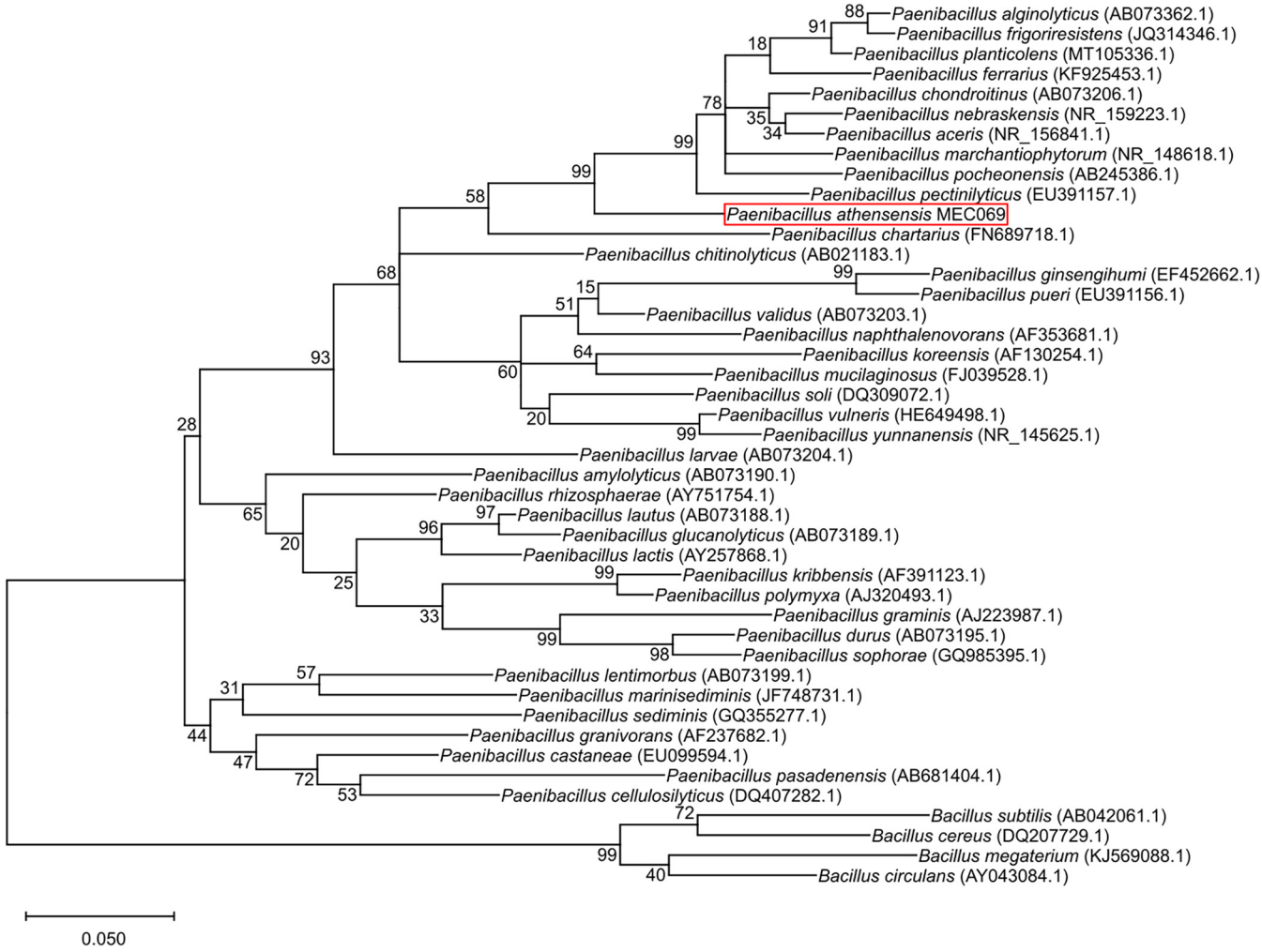

FIG 2 Phylogenetic tree of *Paenibacillus* 16S rRNA gene sequences constructed using the maximum-likelihood method. Members of the *Bacillus* were used as an outgroup. *Paenibacillus* presented were chosen from a tree of all *Paenibacillus* 16S rRNAs to represent major clades within the genus as well as to show the species with genome sequences closely related to *P. athensensis* MEC069. The tree with the highest log likelihood (−11,701.73) is drawn to scale, with the bar with units of the number of substitutions per site and bootstrap values presented next to tree nodes. There were a total of 1,623 positions in the final data set. Evolutionary analyses were conducted in MEGA X.

Taxonomic identification of isolates. To identify *lgdA* and other genes important in levoglucosan catabolism, the genomic DNA from each isolate was sequenced. Assembly statistics and GenBank accession numbers are presented in Table S1 in the supplemental material. Upon a BLAST search of the GenBank database, the 16S rRNA sequence obtained from the genome of MEC069 was closely related to *Paenibacillus marchantiophytorum* R55^T (NR_148618.1), *Paenibacillus nebraskensis* JJ-59^T (NR_159223.1), *Paenibacillus aceris* KUD4121^T (NR_156841.1), and *Paenibacillus planticolens* LMG 31457^T (MT105336.1), with 95.82%, 95.45%, 95.23%, and 95.06% sequence similarity, respectively (Fig. 2). These values were below the threshold of 98.5% for delineation of species (24), suggesting that MEC069 represented a novel species of *Paenibacillus.* This conclusion was confirmed by the average nucleotide identity (ANI) values well below the species threshold of 95% for comparison to genomes of the related species (25). For instance, the ANI values for comparison to *Paenibacillus ginsenghumi*, *Paenibacillus pasadenensis*, *Paenibacillus pectinolyticus*, and *Paenibacillus alginolyticus* were 71.7%, 71.5%, 71.4%, and 70.2%, respectively. Therefore, the name *Paenibacillus athensensis* is proposed below for this novel species.

Similarly, the 16S rRNA sequence from the genome of MEC087 was closely related to *Shinella zoogloeoides* DSM 287^T (AB238789.1), *Shinella yambaruensis* (AB285481.1), *Shinella fusca* (FM177879.1), and *Shinella kummerowiae* (EF070131.1), with 98.70%,

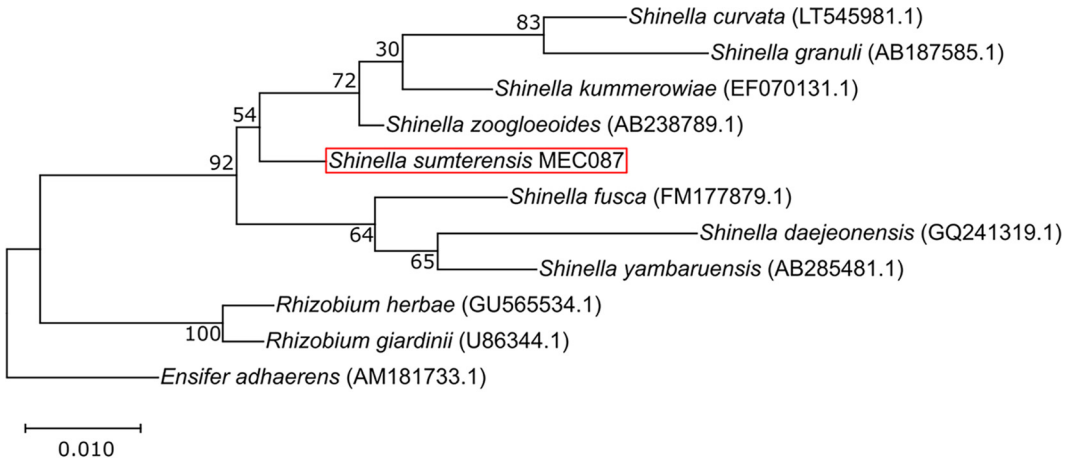

FIG 3 Phylogenetic tree of *Shinella* 16S rRNA gene sequences constructed using the maximum-likelihood method. *Ensifer adhaerens* LMG 20216^T, *Rhizobium giardinii*, and *Rhizobium herbae* were used as an outgroup. All 16S rRNA genes available from officially named *Shinella* organisms were used. The tree with the highest log likelihood (−3,378.77) is drawn to scale, with branch lengths measured in the number of substitutions per site and the bootstrap values shown next to the branches. There were a total of 1,535 positions in the final data set. Evolutionary analyses were conducted in MEGA X.

97.85%, 97.70%, and 97.67% sequence similarity (Fig. 3). Although near the threshold commonly used to delineate species, these values do not preclude MEC087 from being a novel species (26). In fact, the ANI values for comparison of the genome of MEC087 to the closely related genome of *S. zoogloeoides* was 83.5% and consistent with the assignment of MEC087 to a novel species of *Shinella* (27). Therefore, the name *Shinella sumterensis* is proposed below for this novel species.

The remaining two isolates appeared to represent strains of previously described species. The 16S rRNA of MEC097 possessed 99.36% sequence similarity to that of *Klebsiella pneumoniae* DSM 30104^T (X87276.1). Consistent with its assignment to this species, the ANI value between the genomes of MEC097 and *K. pneumoniae* DSM 30104^T was 99.1% and well above the species threshold. In contrast, MEC084 was closely related to *Microbacterium lacusdiani* (NR_149217.2), *Microbacterium album* (NR_159264.1), *Microbacterium deserti* (NR_159265.1), and *Microbacterium marinilacus*^{T} (AB286020.1), with 99.87%, 99.18%, 99.12%, and 98.91% 16S rRNA sequence similarity, respectively (Fig. 4). However, high 16S rRNA sequence similarity does not confirm identity at the species level (24, 26), for which comparison of the whole-genome sequence is necessary. The genome of MEC084 possessed ANI values with *Microbacterium album*, *Microbacterium barkeri*, *Microbacterium indicum*, and *Microbacterium paludicola* of 81.49%, 77.1%, 77.1%, and 75.8%, respectively, confirming its assignment to this genus. However, it was not possible to determine if MEC084 was a novel species or belonged to one of the *Microbacterium* species whose genome sequences are not known. Because of this ambiguity, this isolate is not described as a novel species.

Phenotypic and biochemical comparison. To confirm these assignments and discover the general properties of the isolates, the phenotypic properties of these isolates and selected reference strains were determined (Table 3). *Paenibacillus* MEC069 and *P. polymyxa* DSM 36 exhibited many similar properties on a variety of carbohydrates. Both bacteria grew and produced gas on fructose, galactose, glucose, mannitol, and mannose, grew in the presence of cellobiosan, and did not utilize L-arabinose. While MEC069 grew in the presence of lactose and sucrose and produced acid in the presence of *myo*-inositol and levoglucosan, *P. polymyxa* DSM 36 did not grow in the presence of any of these carbohydrates. *P. polymyxa* DSM 36 also grew and produced acid in the presence of xylose, while MEC069 grew without producing acid.

Growth of *Microbacterium* MEC084 and *M. lacticum* DSM 20427 was similar on most of the carbohydrates examined. Both bacteria grew in the presence of cellobiosan, galactose, glucose, mannose, and xylose, and both bacteria produced acid in the presence

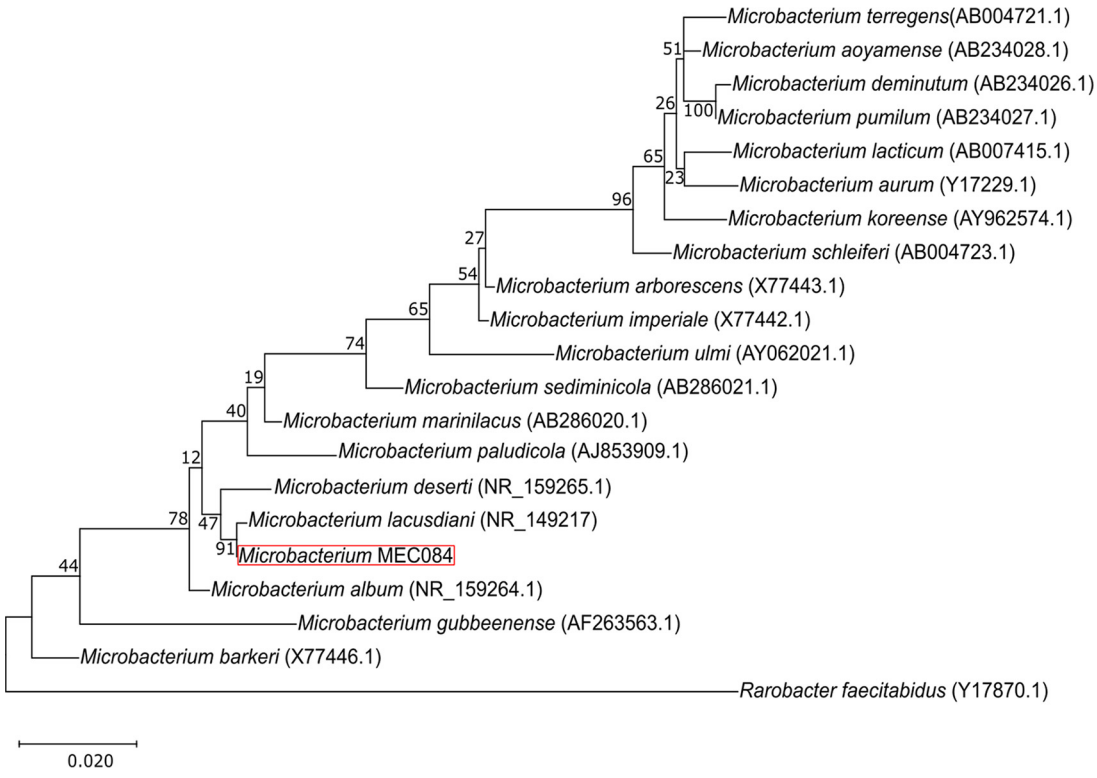

FIG 4 Phylogenetic tree of *Microbacterium* 16S rRNA gene sequences constructed using the maximum-likelihood method. *Rarobacter faecitabidus* DSM 4813^T was used as an outgroup. *Microbacterium* organisms presented were chosen from a tree of all *Microbacterium* 16S rRNAs to represent major clades within the genus as well as the species with genome sequences closely related to *Microbacterium* MEC084. The tree with the highest log likelihood (−4,487.68) is drawn to scale, with the bar with units of the number of substitutions per site and bootstrap values shown next to the branches. There were a total of 1,546 positions in the final data set. Evolutionary analyses were conducted in MEGA X.

of sucrose. MEC084 did not utilize L-arabinose or sorbitol, while *M. lacticum* DSM 20427 did. MEC084 utilized *myo*-inositol and levoglucosan, while *M. lacticum* DSM 20427 did not. Although both bacteria grew in the presence of lactose and mannitol, only *M. lacticum* DSM 20427 produced acid.

S. sumterensis MEC087 was compared to *S. zoogloeoides* DSM 287. *S. zoogloeoides* DSM 287 exhibited growth on all carbohydrates examined except lactose and levoglucosan. MEC087 grew on all the carbohydrates tested and alone produced acid from levoglucosan and sucrose. *S. zoogloeoides* DSM 287 produced acid in the presence of mannose.

K. pneumoniae MEC097 and *K. pneumoniae* DSM 30104 exhibited nearly identical morphologies and behaviors. For both strains, growth was observed on arabinose, cellobiosan, lactose, and levoglucosan, while growth with acid production occurred on fructose, galactose, *myo*-inositol, mannitol, sorbitol, sucrose, and xylose. With the exception of growth on levoglucosan as a sole carbon source, MEC097 exhibited identical phenotypes to *K. pneumoniae* DSM 30104 for all traits tested.

Comparison of the growth of *Klebsiella* strains on cellobiosan. The isolates and all corresponding reference strains grew on cellobiosan, but the isolates grew only when levoglucosan was the sole carbon source (Table 3), demonstrating that growth on cellobiosan was not a predictor for growth on levoglucosan. To examine this difference in greater detail, isolate MEC097 and *K. pneumoniae* subsp. *pneumoniae* DSM 30104 were grown on 10 mM cellobiosan (Fig. 5). Both strains reached their maximum optical density (OD) and exhausted most of the cellobiosan within 14 h. DSM 30104 reached a maximum OD of 2.3 ± 0.3 and accumulated 9.8 ± 1.1 mM levoglucosan, essentially equimolar to the amount of cellobiosan consumed. In contrast, MEC097 attained a maximum OD of 4.0 ± 0.4, and levoglucosan was not detected during the course of cellobiosan

TABLE 3 Phenotypic characterization of bacterial isolates that metabolize levoglucosan as the sole carbon source and reference strains used in this study[a]

Parameter	Value(s) for strain[b]:							
	1*	2	3*	4	5*	6	7*	8
DNA G+C (mol%) content	55.7	44.9	71.7	74.9	63.4	63.4	57.3	57.1
Gram stain	−	−	+	+	−	−	−	−
Cell morphology	b	b	cb	b	c	cb	b	b
Catalase reaction	−	+	−	+	+	+	+	+
Oxidase reaction	+	−	+	−	+	+	−	−
Utilization of:								
L-Arabinose	−	−	−	+	+	+	+	+
Cellobiosan	+	+	+	+	+	+	+	+
Fructose	+, a	+, a	+, a	+	+	+	+, a	−
D-Galactose	+, a	+, a	+	+	+	+	+, a	+, a
D-Glucose	+, a	+, a	+	+	+	+	+, a	+, a
myo-inositol	+	−	+	−	+	+	+, a	+, a
Lactose	+, a	−	+	+, a	+	−	+	+
Levoglucosan	+	−	+	−	+, a	−	+	−
Mannitol	+, a	+, a	+	+, a	+	+	+, a	+, a
D-Mannose	+, a	+, a	+	+	+	+, a	+	
Sorbitol	−	+, a	−	+	+	+	+, a	+, a
Sucrose	+, a	−	+, a	+, a	+, a	+	+, a	+, a
D-Xylose	+	+, a	+	+	+	+	+, a	−

[a]Symbols: +, grows on substrate or exhibits trait; −, does not grow on substrate or does not exhibit trait; a, produced acid; b, bacillus; c, coccus; cb, coccobacillus; *, isolates that metabolize levoglucosan as the sole carbon source.

[b]Strains were the following: 1, MEC069; 2, *Paenibacillus polymyxa* DSM 36^T; 3, MEC084; 4, *Microbacterium lacticum* DSM 20427^T; 5, MEC087; 6, *Shinella zoogloeoides* DSM 287^T; 7, MEC097; 8, *Klebsiella pneumoniae* subsp. *pneumoniae* DSM 30104^T.

degradation. Thus, MEC097 appeared to metabolize cellobiosan and any levoglucosan potentially formed during its catabolism, while DSM 30104 did not consume levoglucosan formed from cellobiosan degradation. No other products (e.g., acetate) were observed during the course of cellobiosan or levoglucosan consumption.

Identification of *lgdA* in genome assemblies and comparison of neighboring genes. Locus tag Asphe3_10730 in *P. phenanthrenivorans* Sphe3 (CP002379.1) is a gene 1,173 bp in length that has recently been shown to encode an LGDH (18). Previously, the gene product had been annotated as a putative dehydrogenase containing an NAD-binding Rossmann fold. Expression of this gene in *E. coli* BL21(DE3)-RIPL cells produced an enzyme that converted NAD to NADH in the presence of levoglucosan. To determine if homologs of this gene were present in the levoglucosan-utilizing isolates, their genomes were searched by BLAST to the Asphe3_10730 sequence. Each genome contained a gene with 66 to 73% sequence identity (Fig. 6; see also Table S2 in the supplemental material). Expression of this putative gene, denoted *lgdA*,

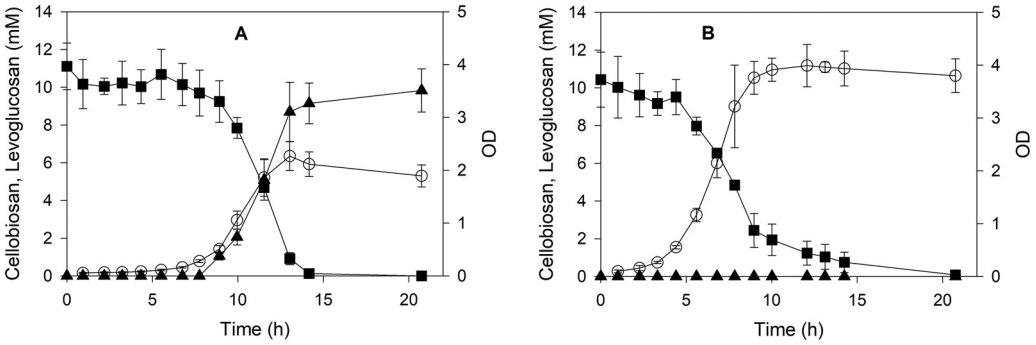

FIG 5 Growth of *K. pneumoniae* subsp. *pneumoniae* DSM 30104^T (A) and *K. pneumoniae* subsp. *pneumoniae* MEC097 (B) on 10 mM cellobiosan as the sole carbon source. Symbols: cellobiosan (■), levoglucosan (▲), and OD (○). Means and standard deviations from triplicates are indicated.

FIG 6 Alignment of LGDH from the bacterial isolates and *P. phenanthrenivorans* Sphe3. No highlighting indicates a residue conserved in at least 60% of the sequences, gray highlighting indicates similar residues, and black highlighting indicates residues that are chemically different from the majority at that position.

from *P. phenanthrenivorans* Sphe3 or any isolate in *E. coli* MG1655 did not confer the ability to grow on levoglucosan as a sole carbon source. Therefore, the product of this gene was not sufficient for levoglucosan catabolism. To identify other genes that might be required, the contigs containing each putative *lgdA* were aligned with Mauve (28), and several neighboring genes conserved among all the bacteria were identified (Fig. 7). These four genes were predicted to code for a monosaccharide ABC transporter substrate-binding protein of the CUT2 family (Table S3, sequence identity of 53 to 70%), a monosaccharide ABC transporter membrane protein of the CUT2 family (Table S4, sequence identity of 58 to 72%), a monosaccharide ABC transporter ATP-binding protein of the CUT2 family (Table S5, sequence identity of 45 to 64%), and a sugar phosphate isomerase/epimerase, denoted *lgdB* (Table S6, sequence identity of 44 to 56%), in addition to the dehydrogenase already known to be levoglucosan dehydrogenase coded by *lgdA* (Table S2).

Additionally, transposases were identified in several contigs near *lgdA*. The contig containing the *lgdA* of *Klebsiella* MEC097 had two IS*5*/IS*1182* transposases, gene locus B4U61_026080, 2,424 bases upstream of the gene cluster, and B4U61_026145, 4,112 bases downstream of the gene cluster. In addition, an IS*3* transposase (gene locus B4U61_26630) was also located 3,899 nucleotides downstream of the gene cluster. In *Shinella* MEC087, two IS*3* transposases, B5M41_025320 and B5M41_025315, were identified 8,656 and 9,859 bases, respectively, downstream from the gene cluster.

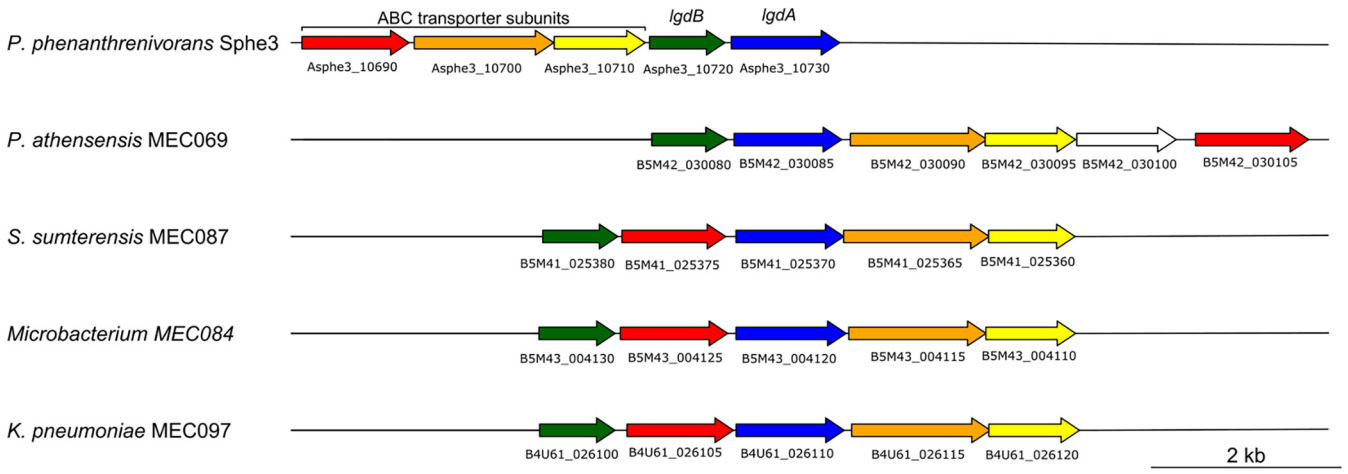

FIG 7 Conserved genes neighboring *lgdA* in the genomes of levoglucosan-utilizing isolates. These genes are annotated as a substrate-binding component (red), ATP-binding component (orange), and membrane component (yellow) of a monosaccharide ABC transporter of the CUT2 family, a sugar phosphate isomerase/epimerase (*lgdB*, green), and a putative dehydrogenase (*lgdA*, blue, shown in this work to be levoglucosan dehydrogenase).

TABLE 4 Kinetic parameters of purified recombinant LGDH from various bacteria[a]

Strain	k_{cat} (s^{-1})	K_m (levoglucosan) (mM)	K_m (NAD) (mM)	K_i (levoglucosan) (mM)	Reference or source
P. phenanthrenivorans	4.3 (0.1)	7.5 (0.6)	0.34 (0.08)	NA[b]	19[c]
P. phenanthrenivorans	52 (3)	27 (4)	1.0 (0.2)	NA	This study
P. athensensis MEC069	84 (4)	1.7 (0.4)	0.31 (0.05)	320 (50)	This study
Microbacterium MEC084	63 (7)	1.8 (0.8)	0.63 (0.18)	470 (150)	This study
S. sumterensis MEC087	27 (5)	2.1 (0.8)	9.1 (2.2)	93 (21)	This study
K. pneumoniae MEC097	58 (3)	22 (3)	0.79 (0.12)	NA	This study

[a]Values were calculated using R with the standard error in parentheses based on reaction rates measured under 27 different concentrations of levoglucosan and NAD^+ for each enzyme. When a substrate inhibition model provided the best fit to the data, the K_i is also shown.
[b]NA, not applicable.
[c]The parameters from reference 19 were obtained at pH 8.5 and 40°C. All reaction mixtures in this study were pH 9.0 and 30°C.

Determination of LGDH kinetic parameters. With a probable *lgdA* coding for LGDH identified in the genome of each isolate, the kinetic parameters of these enzymes were determined. The recombinant LGDHs were expressed with C-terminal His tags and purified on His-trap columns. The kinetic parameters of the five LGDHs varied (Table 4). For three of the enzymes, the best-fitting equations to the data were models that included substrate inhibition. These enzymes all possessed values of K_m for levoglucosan 1 to 2 orders of magnitude lower than the K_i, so it is unlikely that the substrate inhibition would be physiologically significant. In contrast, the values of K_m for levoglucosan for the enzymes from *P. phenanthrenivorans* and *Klebsiella* MEC097 were near 20 mM, suggesting that these enzymes were not well adapted to levoglucosan catabolism.

Comparison of *lgdA* and *lgdB* expression on levoglucosan consumption. As noted above, expression of *lgdA* alone in *E. coli* did not confer growth on levoglucosan as the sole carbon source. The original pathway hypothesized for the bacterial conversion of levoglucosan to glucose proposed a hydration of 3-keto-levoglucosan to 3-keto-glucose (12). Because each of the regions neighboring *lgdA* in the genomes contained a putative isomerase/epimerase, denoted *lgdB* (Fig. 7), we hypothesized that this second protein was also necessary for *E. coli* to metabolize levoglucosan. We therefore constructed BL21(DE3) cells expressing *P. phenanthrenivorans lgdA*, *lgdB*, or both genes on the pET-CDFDuet1 vector. During growth on PA-5052 autoinduction medium with 2 g/liter levoglucosan, the available glucose was consumed by 12 h of growth, glycerol by 18 h, and lactose by 58 h. During lactose consumption, the recombinant proteins were expressed. Cultures expressing both *lgdA* and *lgdB* consumed most of the available levoglucosan by 58 h (Fig. 8). Culture expressing only *lgdA* or *lgdB* consumed less than 20% of the levoglucosan. Thus, both genes were necessary for levoglucosan catabolism.

DISCUSSION

Proof of function of LGDH. Among the four isolates described in this study, *Klebsiella pneumoniae* subsp. *pneumoniae* MEC097 exhibited both the highest growth rate on levo-

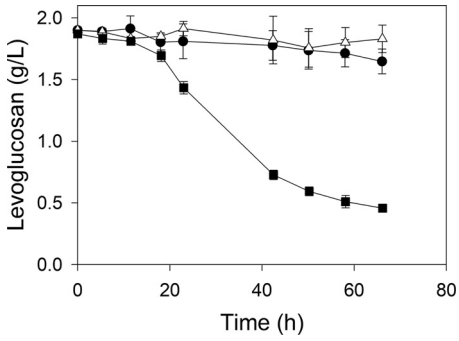

FIG 8 Consumption of levoglucosan by *E. coli* BL21(DE3) cells expressing the *P. phenanthrenivorans lgdA* (●), *lgdB* (△), or both (■) genes following growth on PA-5052 autoinduction medium with 2 g/liter levoglucosan at 37°C and 250 rpm. Means and standard deviations from triplicate cultures are shown.

glucosan (0.80 ± 0.10 h^{-1}) and specific LGDH activity (0.32 ± 0.16 IU/g dry cell weight). In contrast, *P. athensensis* MEC069, *Microbacterium* MEC084, and *S. sumterensis* MEC087 all had growth rates in the range of 0.30 to 0.50 h^{-1} with 15 to 30% the specific LGDH activity of *K. pneumoniae* subsp. *pneumoniae* MEC097.

Interestingly, there is no direct correlation (R^2 = 0.02) between the maximum specific growth rate of isolates on levoglucosan (Table 2) and the turnover (k_{cat}) of the purified LGDH (Table 4), suggesting that enzyme expression level, levoglucosan transport, or another factor is controlling the rate of cell growth on levoglucosan. For example, the measured LGDH activity during growth was greatest for *K. pneumoniae* subsp. *pneumoniae* MEC097, although the k_{cat} is an intermediate value compared to those of other LGDHs. In contrast, *P. athensensis* MEC069 and *Microbacterium* MEC084 contain the most active LGDH enzymes, but these strains exhibited the lowest specific LGDH activity. The K_m and k_{cat} observed in this study at pH 9.0 and 30°C were an order of magnitude greater than those observed at pH 8.5 and 40°C (18).

Based on crystal structures showing the C-3 hydroxyl group of levoglucosan close to the nicotinamide ring of NAD and His-193 (18), the C-3 hydroxyl group is likely oxidized in a hydride transfer with NAD. We hypothesized that if LGDH is oxidizing the C-3 hydroxyl group of levoglucosan, then organisms capable of growth on levoglucosan might also be able to metabolize *myo*-inositol, which has a similar arrangement of hydroxyl groups. Each isolate was indeed able to use *myo*-inositol as the sole carbon source. Among the reference strains, however, all strains except *P. polymyxa* DSM 36 and *M. lacticum* DSM 20427 were also able to grow on *myo*-inositol. Since several reference strains grew on *myo*-inositol but none were able to metabolize the similarly structured levoglucosan and LGDH from all isolates showed no activity in the presence of various concentrations of *myo*-inositol and NAD, the reference strains use a distinct enzyme or pathway for the metabolism of *myo*-inositol.

Cellobiosan is a disaccharide consisting of one β-D-glucose subunit and a levoglucosan subunit bound by a 1,4-linkage. All bacteria examined could metabolize cellobiosan, even those unable to metabolize levoglucosan. The observation that reference strains showed a difference in cellobiosan and levoglucosan metabolism suggested two pathways to degrade cellobiosan: (i) bacteria convert cellobiosan to cellobiose, which is then directly metabolized, or (ii) bacteria cleave cellobiosan into levoglucosan and glucose monomers and then metabolize one or both of the resulting monosaccharides. Our study comparing the growth of *Klebsiella* strains MEC097 and DSM 30104 on cellobiosan (Fig. 5) supports the conclusion that these bacteria cleave cellobiosan into glucose and levoglucosan. In particular, the reference strain accumulated an equimolar amount of levoglucosan during cellobiosan degradation and only attained half the cell density as MEC097 on cellobiosan. Neither glucose nor levoglucosan accumulated during growth of MEC097 on cellobiosan, demonstrating that the two monosaccharides are metabolized simultaneously by this strain, and cellobiose was not detected during the growth of either strain. While the ability to catabolize levoglucosan may be relatively rare among bacteria, the β-glucosidase activity necessary to cleave cellobiosan into its component monomers appears to be much more common. This observation is worth noting as researchers continue the development of bacterial strains for the utilization of pyrolysis products.

Comparison of *lgdA* in the isolate genomes as well as comparison of the neighboring genes yields several insights into bacterial metabolism of levoglucosan. While the arrangement and total number of genes in each *lgdA* operon were variable between isolates, five genes were conserved among all isolates (Fig. 7) and were also recently observed in *B. smithii* (17). Previous studies demonstrated that *E. coli* grows on levoglucosan merely by overexpression of levoglucosan kinase (13, 14, 19). Thus, no specialized levoglucosan transporters are necessary for levoglucosan to cross the cell membrane of *E. coli*. The substrate-binding periplasmic protein, transmembrane protein, and ATP-binding protein components of ABC transporters consistently identified in the five *lgd* operons suggest that the genes clustered with *lgdA* and *lgdB* are components

of ABC transporters specific to levoglucosan. Indeed, these genes are orthologous to the components of the ribose ABC transporter of *Caldanaerobacter subterraneus* subsp. *tengcongensis* and the *myo*-inositol ABC transporter of *Caulobacter vibrioides* (29).

The presence of a conserved putative isomerase adjacent to *lgdA* suggests that during the catabolism of levoglucosan, this enzyme may be responsible for the hydration of 3-keto-levoglucosan. The product of the *lgdB* gene is homologous to tagatose-3-epimerase (30) and 5-keto-L-gluconate epimerase (31) *iolO* from *Thermotoga maritima*, D-tagatose epimerase from *Pseudomonas cichorii* (32), D-psicose 3-epimerase from *Clostridium cellulolyticum* (33), and inosose isomerase *iolI* from *B. subtilis* (34), among other isomerases. Many of these enzymes require a divalent cation such as Mn^{2+}, Mg^{2+}, or Zn^{2+} in their active site to isomerize a keto-sugar. Recently, the products of the *lgdB1* and *lgdB2* genes from *B. smithii* were demonstrated to be an Mn^{2+}-independent β-eliminase for 3-keto-levoglucosan and an Mn^{2+}-dependent bifunctional enzyme that is both a β-eliminase for 3-keto-levoglucosan and a hydratase for 2-hydroxy-3-keto-D-glucal (15). These proteins had a translated sequence identity of 46.52 to 64.34% and 17.02 to 18.75%, respectively, compared to *lgdB* gene products identified in this work. While *lgdB2* has a potential homolog in the genomes of *Microbacterium* MEC084, *S. sumterensis* MEC087, and *P. athensensis* MEC069, the activity of these proteins was not determined in this work. Several homologous enzymes do not require a phosphorylated substrate, an observation that is consistent with the proposed mechanism of catalyzing the conversion of 3-keto-levoglucosan to 3-keto-glucose (12). The observation that *E. coli* expressing both *lgdA* and *lgdB* consumes levoglucosan much faster than cells expressing *lgdA* alone supports the conclusion that the *lgdB* product is important for levoglucosan consumption (Fig. 8) and is likely a dehydratase converting 3-keto-levoglucosan to 3-keto-glucose (or 2-hydroxy-3-keto-D-glucal). Previous work suggests four enzymes are necessary for the conversion of levoglucosan to glucose (17), whereas our *in vivo* work suggests only two are needed. The previous study used purified enzymes from *B. smithii*, while in this study, *P. phenanthrenivorans* genes were expressed in *E. coli*, suggesting native *E. coli* enzymes mediate the final steps in the conversion of levoglucosan to glucose and allow for levoglucosan consumption. Additional experiments will be necessary to confirm the function of all gene products.

In the genomes of *K. pneumoniae* subsp. *pneumoniae* MEC097 and *S. sumterensis* MEC087, the proximity between transposases and the five genes described above suggests that the genes entered the genome by horizontal gene transfer. A putative oxidoreductase from *Klebsiella pneumoniae* (PDB code 4GQA) was proposed to have LGDH activity (18). However, from a comparison of this gene with all the *lgdA* genes identified in this work as well as those of *P. phenanthrenivorans* Sphe3 and *B. smithii*, this enzyme shares 39.13 to 41.43% translated sequence identity. The enzyme is likely an oxidoreductase but might not be an LGDH, as these proteins share 66.07 to 72.82% sequence identity. Assuming no available *Klebsiella* genome contains a functional *lgdA*, it can be reasoned that *lgdA* and other genes that contribute to levoglucosan consumption were integrated into the MEC097 genome by horizontal gene transfer and not a group of genes that likely evolved with *K. pneumoniae*. In the genome assemblies of *P. athensensis* MEC069 and *Microbacterium* MEC084, no transposases were found in the contigs containing the five conserved genes.

Taxonomy of bacteria with LGDH activity. The four bacterial isolates enriched in this study on the basis of their ability to grow on levoglucosan as the sole carbon source represent four distinct species, two of which are novel. The isolates are proposed to be *Paenibacillus athensensis* MEC069^T, *Microbacterium* MEC084, *Shinella sumterensis* MEC087^T, and *Klebsiella pneumoniae* subsp. *pneumoniae* MEC097.

P. athensensis was determined to be a novel species based on its 16S rRNA sequence and the ANI of its genome compared to closely related *Paenibacillus*. Among officially named *Paenibacillus*, *P. pectinilyticus* and *P. alginolyticus* were the closest relatives, with 95.77% and 94.95% 16S rRNA gene similarity, respectively, suggesting that *P. athensensis*

MEC069^T is a unique species (Fig. 1). The ANI between these bacteria and MEC069 were 71.35% and 70.18%, confirming that they are distinct species of *Paenibacillus*.

While *S. sumterensis* MEC087^T exhibited many similarities with *S. zoogloeoides*, including 99.86% 16S rRNA sequence similarity, the ANI between these two bacteria was 83.53%, indicating that the two organisms are likely of the same genus but of distinct species.

MEC084 may be a novel species among the *Microbacterium*, but this strain could be an isolate of *M. lacusdiani*. From the 16S rRNA sequences, MEC084 has 99.86% similarity to *M. lacusdiani* and 98.90% sequence similarity with *M. marinilacus*. Notable differences between *M. lacusdiani* and MEC084 are that *M. lacusdiani* does not utilize D-galactose, D-glucose, and sorbitol, utilizes L-arabinose and sorbitol, was isolated from a freshwater environment, and has a G+C content of 70.4 mol% (35). With no genetic sequence from *M. lacusdiani* available for comparison other than the 16S rRNA gene, one cannot definitively conclude these are unique bacteria.

Isolate MEC097 is the only known strain of *K. pneumoniae* subsp. *pneumoniae* capable of utilizing levoglucosan as its sole carbon source. Comparing 16S rRNA sequences, there is 99.25% sequence similarity between MEC097 and *K. pneumoniae* subsp. *pneumoniae* DSM 30104^T (X87276.1); the ANI between the two strains (using assembly GCA_000281755.1) is 0.9905, making it clear they are the same species. On plate media, both organisms form the same mucoid, white colonies, and they consume many of the same carbohydrates. MEC097 is a unique strain of *K. pneumoniae*.

Although isolates exhibited biochemical activities similar to those of their closest match by phylogenetic analysis, surprisingly none of the reference strains were capable of growth on levoglucosan. This difference in substrate utilization between close phylogenetic neighbors implies that the isolates have a novel metabolic pathway that may have been shared through horizontal gene transfer, a conclusion supported by the high sequence similarity between any two translated *lgdA* (66.07 to 75.19%).

Description of *Paenibacillus athensensis* sp. nov. *Paenibacillus athensensis* (a.thens.en'sis. N.L. masc. adj. athensensis, in recognition of Athens, Georgia, United States, the location from which the type strain was isolated).

Cells are Gram-positive staining coccobacilli, facultatively anaerobic, motile, catalase and oxidase positive. Colonies are small, cream-beige colored, opaque, dull, with entire margins after 24 h of growth on levoglucosan isolation medium (LIM) at 37°C. Growth was observed on cellobiosan, *myo*-inositol, levoglucosan, and xylose. Growth and acid production occur on fructose, galactose, glucose, lactose, mannitol, mannose, and sucrose. Growth was not observed with arabinose or sorbitol. The DNA G+C content is 55.7 mol%. The type strain (MEC069^T = DSM 28474^T = ATCC TSD-105) was isolated from the wastewater of a chemical manufacturing company (Athens, GA; GPS location 33.98023, −83.32673). The accession number for the genome assembly is MYFO00000000, and the 16S rRNA sequence is locus tag B5M42_24910.

Description of *Shinella sumterensis* sp. nov. *Shinella sumterensis* (sum.ter.en'sis N.L. fem. adj. sumterensis, in recognition of Sumter National Forest, South Carolina, United States, the location from which the type strain was isolated).

Cells are Gram-negative staining cocci arranged in long chains, facultatively anaerobic, motile, catalase and oxidase positive. Colonies are small, pinkish-white, glistening, smooth, with entire margins after growth for 24 h on levoglucosan isolation medium at 37°C. Growth was observed on arabinose, cellobiosan, fructose, galactose, glucose, *myo*-inositol, lactose, levoglucosan, mannitol, mannose, sorbitol, sucrose, and xylose. The DNA G+C content is 63.4 mol%. The type strain (MEC087^T = DSM 27294^T = ATCC TSD-106) was isolated from silty topsoil heavy in organic matter from a campfire pit (Russell Homestead, SC; GPS location 34.91107, −83.17371). The accession number for the genome assembly is MYFN00000000, and the 16S rRNA sequence is locus tag B5M41_026195.

MATERIALS AND METHODS

Liquid medium. Defined levoglucosan isolation medium (LIM) was based on ATCC 1190 medium (36) and consisted of (per liter) 2.00 g levoglucosan, 1.50 g $KH_2PO_4 \cdot 7H_2O$, 3.10 g $Na_2HPO_4 \cdot 7H_2O$, 0.50 g NH_4Cl, 0.50 g $(NH_4)_2SO_4$, 47 mg $CaCl_2 \cdot 2H_2O$, 20 mg $FeSO_4 \cdot 7H_2O$, 90 mg $MgCl_2 \cdot 6H_2O$, 7.5 mg nitriloacetic

acid, 15 mg $MgSO_4 \cdot 7H_2O$, 2.5 mg $MnSO_4 \cdot H_2O$, 5.0 mg NaCl, 0.41 mg $CoCl_2 \cdot 2H_2O$, 0.50 mg $ZnSO_4 \cdot H_2O$, 0.46 mg $NiCl_2 \cdot 6H_2O$, 50 μg $CuSO_4 \cdot 5H_2O$, 50 μg H_3BO_3, 50 μg $Na_2MoO_4 \cdot 2H_2O$, 50 μg $Na_2WO_4 \cdot 2H_2O$, 40 μg $Al_2(SO_4)_3$, 5 μg Na_2SeO_3, 120 μg pyridoxine·HCl, 60 μg thioctic acid, 60 μg *p*-aminobenzoic acid, 24 μg folic acid, 60 μg calcium pantothenate, 60 μg nicotinic acid, 60 μg thiamine·HCl, 24 μg biotin, 6 μg riboflavin, and 1.2 μg cyanocobalamin. The medium was brought to pH 7.0 with NaOH. Agar plates contained LIM with 18 g/liter agar.

The expression of recombinant enzymes in *Escherichia coli* under the control of a *lac* promoter was accomplished in PA-5052 autoinduction medium (37). For experiments testing the consumption of levoglucosan with one or more enzymes expressed, PA-5052 medium was modified to contain 2.5 g/liter glycerol, 0.25 g/liter glucose, 0.5 g/liter α-lactose, and 2 g/liter levoglucosan.

Isolation of levoglucosan-degrading bacteria. Samples were collected from several sources: silty topsoil and charred wood from a campfire pit (Nantahala National Forest, NC; GPS location 35.01050, −83.25258), silty topsoil heavy in organic matter from a campfire (Russell Farmstead, Mountain Rest, SC; GPS location 34.91107, −83.17371), raw wastewater from a chemical company (Athens, GA; GPS location 33.98023, −83.32673), and raw wastewater (Middle Oconee Water Reclamation Center, Athens, GA; GPS location 33.90955, −83.39059). Duplicates of each sample, 0.3 g or 0.3 ml, were inoculated into 3 ml of LIM in sterile, screw-top tubes, which were continuously shaken at 37°C. Twice after consecutive 24-h periods, 0.10 ml from each enrichment was transferred into 3 ml fresh LIM and similarly incubated. After the third transfer, samples were streaked onto LIM-agar plates, and individual colonies were isolated.

Analyses. The optical density (OD) of cultures was measured at 600 nm (DU-650 spectrophotometer; Beckman Instruments, San Jose, CA). Growth rates were measured of cultures in 50 ml of LIM in 250 ml baffled flasks shaken at 250 rpm (19 mm pitch) at 37°C. About 6 to 8 samples were withdrawn during exponential growth to measure the OD and calculate the growth rate.

High-performance liquid chromatography using refractive index detection on a Shimadzu Prominence system (Shimadzu Scientific Instruments, Columbia, MD) and a Coregel 64-H ion-exclusion column (Transgenomic Ltd., Glasgow, United Kingdom) at 60°C with a mobile phase of 4 mN H_2SO_4 and a flow rate of 0.6 ml/min were used for analysis of organic compounds (38).

Enzyme assays. Levoglucosan kinase activity was measured by a coupled assay with pyruvate kinase and lactate dehydrogenase (39). The assay mixture contained 0.1 M levoglucosan, 50 mM triethanolamine-HCl (pH 8.0), 0.1 M KCl, 1.12 U/ml pyruvate kinase (Sigma-Aldrich Co., St. Louis, MO), 1.6 U/ml lactate dehydrogenase (Sigma-Aldrich Co., St. Louis, MO), 2 mM ATP, 2 mM phosphoenolpyruvate, 10 mM $MgCl_2$, and 160 μM NADH. A codon-optimized levoglucosan kinase gene from *Lipomyces starkeyi* YZ-215 was expressed in *E. coli* XL1-Blue cells and used as a positive control (18). Levoglucosan dehydrogenase activity was measured by the direct formation of NADH (12), and the assay mixture contained 0.1 M Tris (pH 9.0), 2 mM NAD, and 100 mM levoglucosan. For either enzyme, one unit (IU) of activity was defined as the amount of enzyme required to produce 1 μmol glucose-6-phosphate or NADH per min. Activity and specific enzyme activity (IU per dry cell weight) were measured by growing a 50-ml culture in LIM in a 250-ml baffled shake flask to an OD of 1.0. A portion of the culture was centrifuged, and the pellet was washed with water twice before drying in an oven at 60°C for 24 h for determination of the dry cell weight.

Purified recombinant LGDH was assayed for activity using variations in the spectrophotometric assay (12) with 0.1 M Tris (pH 9.0 at room temperature), 0.1 to 10.0 mM NAD^+, and 1 to 300 mM levoglucosan. Each reaction mixture was incubated at 30°C in a Peltier heater for 2 min, and then the reaction was initiated by the addition of 100 μl of 0.05 to 0.10 mg LGDH/ml, which had separately been incubated at 30°C. Absorbance at 340 nm was recorded every 4 s for 2 min for the calculation of the reaction rate. Kinetic parameters were determined with an R script utilizing the minpack.lm plug-in for the Levenberg-Marquardt nonlinear least-squares algorithm (40).

Illumina sequencing and genome assembly. Overnight cultures of each isolate grown in LIM were used to isolate genomic DNA as detailed in the instructions of the Promega Wizard genomic DNA purification kit (Promega Corporation, Madison, WI). Genomic DNA in EB buffer was then used to create NGS DNA libraries, which were then sequenced using paired-end 2 × 250-bp reads on an Illumina MiSeq sequencer (Georgia Genomics Facility, Athens, GA).

Low-quality reads were removed from the sequencing data using Bowtie 2 (41). Genomes were then assembled using the A5-miseq pipeline (42). The mol% G+C of each organism was determined from the final genome assemblies. The 16S rRNA gene sequences were identified in the genome assemblies and used for a phylogenetic analysis of each isolate. Average nucleotide identity between isolate genomes and other species was determined with the ANI Calculator (www.ezbiocloud.net [43]). All draft genomes were submitted to GenBank and annotated with the NCBI Prokaryotic Genome Annotation Pipeline (44).

Phenotypic and biochemical comparison. Based on the results of the genomic analysis, reference strains were purchased from the German Collection of Microorganisms and Cell Cultures (DSMZ) for phenotypic comparison to the isolates. These include *Paenibacillus polymyxa* DSM 36^T, *Microbacterium lacticum* DSM 20427^T, *Klebsiella pneumoniae* subsp. *pneumoniae* DSM 30104^T, and *Shinella zoogloeoides* DSM 287^T.

Growth and acid production of the isolates and reference strains were tested on several carbon and energy sources. LIM was used (3 ml), except levoglucosan was replaced by 20 mM target substrate, and 4 mg/liter bromocresol purple was added to detect acid production. Cultures were incubated at 37°C for 48 h with the following compounds: D-arabinose, cellobiosan, D-fructose, D-galactose, D-glucose, *myo*-inositol, lactose, levoglucosan, D-mannitol, mannose, D-sorbitol, sucrose, and D-xylose. The following conventional growth media were also examined: SIM stab for the determination of H_2S production, indole production, and motility and lysogeny broth (LB) agar in a GasPAK EZ pouch system (Becton, Dickinson and Company, NJ, USA) to test for anaerobic growth. The catalase test (bubble formation) was performed by growing each isolate overnight on LIM agar at 37°C and then streaking one colony onto a

TABLE 5 Primers used in this study[a]

Organism and primer name	Sequence	Melting point (°C)
P. phenanthrenivorans Sphe3		
MEC189 C-term F	GGG*CTCTTC*AATGATGCAGAACCTCAACGTCG	67
MEC189 C-term R	GGG*CTCTTC*AGTGGGCGGAGATCTGCGGAAC	
pCDF-lgdAF	GAGATATACCATGCAGAACCTCAACGTCGG	62
pCDF-lgdAR	TTAAGCATTATCACGCGCTGATCTGCG	
pCDF-MCS1F	GATCAGCGCGTGATAATGCTTAAGTCGAACAGAAAG	59
pCDF-MCS1R	GGTTCTGCATGGTATATCTCCTTATTAAAGTTAAACAAAA	
pCDF-lgdBF	GAAGGAGATATACATATGAAACTCGGTTACTGTTCCATC	62
pCDFD-lgdBR	AGCCTAGGTTAATTATCAGAGGAGTTTCTCGAGGTACTT	
pCDF-MCS2F	CGAGTTTCATATGTATATCTCCTTCTTATACTTAACTAAT	60
pCDF-MCS2R	GAGAAACTCCTCTGATAATTAACCTAGGCTGCTGCC	
P. athensensis MEC069^T		
MEC069 C-term F	GGG*CTCTTC*GATGGGAAAGGCGCATTC	62
MEC069 C-term R	GGG*CTCTTC*AGTGGATCGTTTTCACCCATTGC	
MEC069-SOE-F-GALG	GGAAGGCGCCCTCGGCGCTATTCTTAGCTTCC	60
MEC069-SOE-R-GALG	AATAGCGCCGAGGGCGCCTTCCTCAATATACTTTTTCGCC	
Microbacterium MEC084		
MEC084 C-term-F	GGG*CTCTTC*CATGCAGGACCTCAACATCG	63
MEC084 C-term-R	GC*GCTCTTC*AGTGGCCATCGGAGGAGCTCTC	
S. sumterensis MEC087^T		
MEC087 C-term-F	GGG*CTCTTC*GATGACCAAAGTGATGAACGTC	60
MEC087 C-term-R	GGG*CTCTTC*CGTGTCCCACCTTCTCCCACAG	
K. pneumoniae MEC097		
MEC097 C-term-F	GGG*CTCTTC*GATGAAAACACTGAATGTAGGTATG	58
MEC097 C-term-R	CC*GCTCTTC*CGTGGATTTCGTCAACCGAAACC	

[a]Primers used for the recombinant expression of levoglucosan dehydrogenase from isolates in pET28b-SapKO vector with a C-terminal His tag and those used for the expression of *P. phenanthrenivorans* Sphe3 *lgdA* and *lgdB* in pCDF-Duet1 are shown. BspQI recognition sequences are italicized. Note that in the middle of *P. athensensis lgdA*, the DNA sequence coding for the residues GALG make a strong hairpin structure, which appears to inhibit protein production in BL21 cells. Primers MEC069-SOE-F-GALG and MEC069-SOE-R-GALG are used to alter the DNA sequence of MEC069 *lgdA* once it has already been ligated to the pET28bCH-SapKO vector, creating one double-stranded piece of DNA that is recircularized using NEBuilder HiFi master mix.

clean microscope slide and adding several drops of 3% H_2O_2. The oxidase test was performed in a similar manner except that cells were collected on a sterile cotton swab before addition of a few drops of freshly prepared Kovac's oxidase solution (1% tetra-methyl-*p*-phenylenediamine dihydrochloride in deionized water). A change in color to purple-black was indicative of a positive test.

To examine growth on cellobiosan as the sole carbon source, *Klebsiella pneumoniae* subsp. *pneumoniae* DSM 30104^T and MEC097 were each grown from a single colony overnight in 5.0 ml LIM containing 10 mM glucose instead of levoglucosan. These overnight cultures were used to inoculate 50 ml LIM containing 10 mM cellobiosan in 250-ml shake flasks. Samples were examined for OD and cellobiosan, levoglucosan, and other products by HPLC until no increase in OD was observed.

Identification of *lgdA* in genome assemblies and comparison of neighboring genes. Previously published partial peptide sequences of LGDH (12) were used in a tBLASTn search of GenBank. The closest result, gene locus Asphe10730 of assembly GCA_000189535.1 of *Pseudarthrobacter phenanthrenivorans* Sphe3, was cloned into pET-28b and recombinantly expressed in *E. coli* BL21(DE3)-RIPL cells (Agilent Technologies, Santa Clara, CA). Cell-free extracts were assayed for LGDH activity as described above. Genomes from levoglucosan isolates were then searched by BLAST using the experimentally identified *lgdA* gene.

Cloning. A modified version of the pET28b vector [pET28b(+)-CHSapKO-CH.BspQI] was used for the expression of *lgdA* under the control of a T7 promoter (45). Briefly, the modifications within pET28b-CHSap include deletion of a SapI site at position 3108, addition of two overlapping SapI sites positioned between the start codon and the double stop codon, and a polyhistidine tag at the 3′ location relative to the reading frame for the production of C-terminally His-tagged protein. Identified *lgdA* genes were PCR amplified from genomic DNA using Phusion polymerase (New England Biolabs Inc., Ipswich, MA) and the primers listed in Table 5. PCR fragments were purified on DNA clean and concentrate columns (Zymo Research Co., Irvine, CA) and eluted using nuclease-free water. Volumes of 50 ng of pET28b (+)-CHSapKO-CH.BspQI vector (45) and 27 ng of PCR product (1:2 molar ratio) were combined and digested with 1 U of BspQI (New England Biolabs Inc., Ipswich, MA) and subsequently ligated with Fast-Link T_4 DNA ligase (Epicentre, Madison, WI) as previously described (45). After ligation, 1 μl of the ligation reaction was transformed into 40 μl *E. coli* strain XL1-Blue (Stratagene, San Diego, CA) by electroporation, and cells were grown on LB agar containing kanamycin. Colonies were screened for the presence of pET28b-CHSap containing *lgdA* by restriction analysis and PCR verification. DNA sequencing verified the sequence of *lgdA* in each construct (pET28b-CH*lgdA*).

The pET-CDFDuet1 vector (Novagen Inc., Madison, WI) was used to express *P. phenanthrenivorans* Sphe3 *lgdA*, *lgdB*, or both genes in BL21(DE3) cells (New England Biolabs, Ipswich, MA). *lgdA* and *lgdB* were inserted into MCS1 and MCS2, respectively, to form pET-CDFDuet1-*lgdA* and pET-CDFDuet1-*lgdB*. Each gene and the vector backbone were PCR amplified with primers (Table 5) producing ~25-bp overhangs complementary between the gene insert and the vector so they could be assembled with HiFi master mix (New England Biolabs, Ipswich, MA). Vector pET-CDFDuet1-*lgdA-lgdB* was created by PCR amplification of pET-CDFDuet1-*lgdA* around MCS2 and inserting *lgdB*. After incubation for 1 h, 2 μl of master mix was used to transform chemically competent DH5α cells, which were plated on LB plates containing 50 μg/ml streptomycin and 50 μg/ml spectinomycin and incubated at 37°C overnight. One colony of each transformant was grown in 3 ml LB supplemented with streptomycin and spectinomycin at 37°C for 6 h. Plasmid was isolated from the cells using the Zyppy plasmid miniprep kit (Zymo Research Corp., Irvine, CA) and subsequently used to transform chemically competent BL21(DE3) cells. Five colonies of each transformant were used to inoculate one test tube containing 3 ml LB with streptomycin and spectinomycin. After incubation at 37°C and 250 rpm for 6 h, 0.25 ml of the LB culture was used to inoculate triplicate 250-ml shake flasks containing 50 ml of PA-5052 autoinduction medium (37) modified to contain 2.5 g/liter glycerol, 0.25 g/liter glucose, 0.5 g/liter α-lactose, and 2 g/liter levoglucosan incubated at 37°C and 250 rpm.

Recombinant expression of enzymes. To generate recombinant LGDH, 40-μl aliquots of *E. coli* strain BL21(DE3)-RIPL cells (Agilent Technologies, Santa Clara, CA) were transformed with each pET28b-CH*lgdA* by electroporation and subsequently plated on LB plus Kan agar. After growing overnight, 5 to 10 colonies were inoculated into 2.5 ml Superbroth (35 g/liter tryptone, 20 g/liter yeast extract, 5 g/liter NaCl, pH adjusted to 7.3 with NaOH) with 100 μg/ml kanamycin in a test tube, and then the culture was incubated at 37°C and 300 rpm for 4 to 6 h. A volume of 5 ml was used to inoculate 2.5-liter Ultra Yield flasks (Thomson Scientific Co., Oceanside, CA) containing 500 ml PA-5052 autoinduction medium (37) with 100 μg/ml kanamycin, which were incubated at 22 to 24°C and 300 rpm for 18 to 20 h. Cells were harvested by centrifugation (11,000 $\times$ *g* for 20 min at 4°C) and resuspended in ~30 ml of 50 mM Tris, 25 mM imidazole, 500 mM NaCl, pH 8.0.

Purification of His-tagged LGDH. Resuspended cells were lysed with a Thermo Electron FA-078A French press using an FA-032 cell cooled to 4°C at 8,000 lb/in^2. Cell-free extract was then centrifuged at 60,000 $\times$ *g* for 20 min at 4°C. The supernatant was purified with a Pharmacia ÄKTA purifier HPLC system at room temperature. The cell extract was first loaded (2 ml/min) on a 5-ml HisTrap HP column (GE Life Sciences, Marlborough, MA) equilibrated with 50 mM Tris, 25 mM imidazole, 500 mM NaCl, pH 8.0. After washing unbound protein from the column, the His-tagged LGDH was eluted with a gradient of imidazole from 25 mM to 500 mM over 15 column volumes. The protein eluted as one large peak, which was collected and dialyzed, first into 1 liter of 50 mM Tris buffer (pH 8.0), 250 mM NaCl, 5 mM β-mercaptoethanol (BME) for 4 h at 4°C and then into 2 liters of 50 mM Tris buffer (pH 8.0), 5 mM BME for 4 h at 4°C. LGDH was further purified by anion-exchange chromatography using a 5 ml HiTrap Q column (GE Life Sciences, Marlborough, MA) equilibrated with 50 mM Tris, pH 8.0. Protein was eluted by using a linear gradient of 0 to 1.0 M NaCl over 15 column volumes. LGDH eluted as one large, sharp peak, which was dialyzed first in 1 liter of 50 mM Tris (pH 8.0), 250 mM NaCl, 5 mM BME for 4 h at 4°C and then in 2 liters of 50 mM Tris (pH 8.0), 5 mM BME for 4 h at 4°C. Fractions from both purifications were stored at −20°C until SDS-PAGE analysis of protein purity.

Data availability. The newly determined genome assemblies for *Paenibacillus athensensis* MEC069, *Microbacterium* MEC084, *Shinella sumterensis* MEC087, and *Klebsiella pneumoniae* MEC097 were deposited in GenBank under the accession numbers MYFO00000000, MYFP00000000, MYFN00000000, and MWQX00000000, respectively.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 0.2 MB.

ACKNOWLEDGMENTS

Sarah Lee, for technical assistance, and Mary-Ann Moran and Brandon Satinsky, for assistance in obtaining sequencing of isolate genomes, are gratefully acknowledged. We also thank William B. Whitman and Cory Momany for helpful discussion.

REFERENCES

1. Shafizadeh F, Fu YL. 1973. Pyrolysis of cellulose. Carbohydr Res 29:113–122. https://doi.org/10.1016/s0008-6215(00)82074-1.
2. Maduskar S, Maliekkal V, Neurock M, Dauenhauer PJ. 2018. On the yield of levoglucosan from cellulose pyrolysis. ACS Sustainable Chem Eng 6:7017–7025. https://doi.org/10.1021/acssuschemeng.8b00853.
3. Junior II, do Nascimento MA, Alves de Souza ROM, Dufour A, Wojcieszak R. 2020. Levoglucosan: a promising platform molecule? Green Chem 22:5859–5880. https://doi.org/10.1039/D0GC01490G.
4. Lomax JA, Commandeur JM, Arisz PW, Boon JA. 1991. Characterization of oligomers and sugar ring-cleavage products in the pyrolysate of cellulose. J Anal App Pyrol 19:65–79. https://doi.org/10.1016/0165-2370(91)80035-7.
5. Dobele G, Rossinskaja G, Dizhbite T, Telysheva G, Meier D, Faix O. 2005. Application of catalysis for obtaining 1,6-anhydrosaccharides from cellulose and wood by fast pyrolysis. J Anal Appl Pyrolysis 74:401–405. https://doi.org/10.1016/j.jaap.2004.11.031.
6. Bennet NM, Duff SJB. 2005. Fortification of spent sulphite pulping liquor with levoglucosan derived from pyrolysis oil. Am Soc Agric Biol Eng 2005:PNW05-1015.
7. Vitasari CR, Meindersma GW, de Haan AB. 2011. Water extraction of pyrolysis oil: the first step for the recovery of renewable chemicals. Bioresour Technol 102:7204–7210. https://doi.org/10.1016/j.biortech.2011.04.079.
8. Bennet NM, Helle SS, Duff SJB. 2009. Extraction and hydrolysis of levoglucosan from pyrolysis oil. Bioresour Technol 100:6059–6063. https://doi.org/10.1016/j.biortech.2009.06.067.

9. Kitamura Y, Yasui T. 1991. Purification and some properties of levoglucosan (1,6-anhydro-β-D-glucopyranose) kinase from the yeast *Sporobolomyces salmonicolor*. Agric Biol Chem 55:523–529. https://doi.org/10.1271/bbb1961.55.523.
10. Prosen EM, Radlein D, Piskorz J, Scott DS, Legge RL. 1993. Microbial utilization of levoglucosan in wood pyrolysate as a carbon and energy source. Biotechnol Bioeng 42:538–541. https://doi.org/10.1002/bit.260420419.
11. Xie H, Zhuang X, Bai Z, Qi H, Zhang H. 2006. Isolation of levoglucosan-assimilating microorganisms from soil and an investigation of their levoglucosan kinases. World J Microbiol Biotechnol 22:887–892. https://doi.org/10.1007/s11274-006-9133-5.
12. Nakahara K, Kitamura Y, Yamagishi Y, Shoun H, Yasui T. 1994. Levoglucosan dehydrogenase involved in the assimilation of levoglucosan in *Arthrobacter* sp. I-552. Biosci Biotechnol Biochem 58:2193–2196. https://doi.org/10.1271/bbb.58.2193.
13. Zhuang X, Zhang H. 2002. Identification, characterization of levoglucosan kinase, and cloning and expression of levoglucosan kinase cDNA from *Aspergillus niger CBX-209* in *Escherichia coli*. Protein Expr Purif 26:71–81. https://doi.org/10.1016/s1046-5928(02)00501-6.
14. Dai J, Yu Z, He Y, Zhang L, Bai Z, Dong Z, Du Y, Zhang H. 2009. Cloning of a novel levoglucosan kinase gene from *Lipomyces starkeyi* and its expression in *Escherichia coli*. World J Microbiol Biotechnol 25:1589–1595. https://doi.org/10.1007/s11274-009-0048-9.
15. Klesmith JR, Bacik JP, Michalczyk R, Whitehead TA. 2015. Comprehensive sequence-flux mapping of a levoglucosan utilization pathway in *E. coli*. ACS Synth Biol 4:1235–1243. https://doi.org/10.1021/acssynbio.5b00131.
16. Yasui T, Kitamura Y, Nakahara K, Abe Y. 1991. Metabolism of levoglucosan (1,6-anhydro-β-D-glucopyranose) in bacteria. Agric Biol Chem 55:1927–1929. https://doi.org/10.1080/00021369.1991.10870881.
17. Kuritani Y, Sato K, Dohra H, Umemura S, Kitaoka M, Fushinobu S, Yoshida N. 2020. Conversion of levoglucosan into glucose by the coordination of four enzymes through oxidation, elimination, hydration, and reduction. Sci Rep 10:20066. https://doi.org/10.1038/s41598-020-77133-8.
18. Sugiura M, Nakahara M, Yamada C, Arakawa T, Kitaoka M, Fushinobu S. 2018. Identification, functional characterization, and crystal structure determination of bacterial levoglucosan dehydrogenase. J Biol Chem 293:17375–17386. https://doi.org/10.1074/jbc.RA118.004963.
19. Layton DS, Ajjarapu A, Choi DW, Jarboe LR. 2011. Engineering ethanologenic *Escherichia coli* for levoglucosan utilization. Bioresour Technol 102:8318–8322. https://doi.org/10.1016/j.biortech.2011.06.011.
20. Lian J, Choi J, Tan YS, Howe A, Wen Z, Jarboe LR. 2016. Identification of soil microbes capable of utilizing cellobiosan. PLoS One 11:e0149336. https://doi.org/10.1371/journal.pone.0149336.
21. Iwazaki S, Hirai H, Hamaguchi N, Yoshida N. 2018. Isolation of levoglucosan-utilizing thermophilic bacteria. Sci Rep 8:4066. https://doi.org/10.1038/s41598-018-22496-2.
22. Schauer JJ, Kleeman MJ, Cass GR, Simoneit BRT. 2001. Measurement of emissions from air pollution sources. 3. C1–C29 organic compounds from fireplace combustion of wood. Environ Sci Technol 35:1716–1728. https://doi.org/10.1021/es001331e.
23. Lee S, Baumann K, Schauer JJ, Sheesley RJ, Naeher LP, Meinardi S, Blake DR, Edgerton ES, Russell AG, Clements M. 2005. Gaseous and particulate emissions from prescribed burning in Georgia. Environ Sci Technol 39:9049–9056. https://doi.org/10.1021/es051583l.
24. Stackebrandt E, Ebers J. 2006. Taxonomic parameters revisited: tarnished gold standards. Microbiol Today 33:152–155.
25. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol 57:81–91. https://doi.org/10.1099/ijs.0.64483-0.
26. Keswani J, Whitman WB. 2001. Relationship of 16S rRNA sequence similarity to DNA hybridization in prokaryotes. Int J Syst Evol Microbiol 51:667–678. https://doi.org/10.1099/00207713-51-2-667.
27. Chun J, Oren A, Ventosa A, Christensen H, Arahal DR, da Costa MS, Rooney AP, Yi H, Xu XW, Meyer SD, Trujillo ME. 2018. Propose minimal standards for the use of genome data for the taxonomy of prokaryotes. Int J Syst Evol Microbiol 68:461–466. https://doi.org/10.1099/ijsem.0.002516.
28. Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394–1403. https://doi.org/10.1101/gr.2289704.
29. Herrou J, Crosson S. 2013. *myo*-inositol and D-ribose ligand discrimination in an ABC periplasmic binding protein. J Bacteriol 195:2379–2388. https://doi.org/10.1128/JB.00116-13.
30. Shin S-M, Cao T-P, Choi JM, Kim S-B, Lee S-J, Lee SH, Lee D-W. 2017. TM0416, a hyperthermophilic promiscuous nonphosphorylated sugar isomerase, catalyzes various C5 and C6 epimerization reactions. Appl Environ Microbiol 83:e03291-16. https://doi.org/10.1128/AEM.03291-16.
31. Rodionova IA, Leyn SA, Burkart MD, Boucher N, Noll KM, Osterman AL, Rodionov DA. 2013. Novel inositol catabolic pathway in *Thermotoga maritima*. Environ Microbiol 15:2254–2266. https://doi.org/10.1111/1462-2920.12096.
32. Ishida Y, Kamiya T, Itoh H, Kimura Y, Izumori K. 1997. Cloning and characterization of the D-tagatose 3-epimerase gene from *Pseudomonas cichorii* ST-24. J Ferm Bioeng 83:529–534. https://doi.org/10.1016/S0922-338X(97)81132-4.
33. Chan HC, Zhu Y, Hu Y, Ko TP, Huang CH, Ren F, Chen CC, Ma Y, Guo RT, Sun Y. 2012. Crystal structures of D-psicose 3-epimerase from *Clostridium celluloyticum* H10 and its complex with ketohexose sugars. Protein Cell 3:123–131. https://doi.org/10.1007/s13238-012-2026-5.
34. Zhang RG, Dementieva I, Duke N, Collart F, Quaite-Randall E, Alkire R, Dieckman L, Maltsev N, Korolev O, Joachimiak A. 2002. Crystal structure of *Bacillus subtilis* loll shows endonuclease IV fold with altered Zn binding. Proteins 48:423–426. https://doi.org/10.1002/prot.10159.
35. Zhang BH, Salam N, Cheng J, Li HQ, Yang JY, Zha DM, Guo QG, Li WJ. 2017. *Microbacterium lacusdiani* sp. nov., a phosphate-solubilizing novel actinobacterium isolated from mucilaginous sheath of *Microcystis*. J Antibiot (Tokyo) 70:147–151. https://doi.org/10.1038/ja.2016.125.
36. Wiegel J, Ljungdahl LG, Rawson JR. 1979. Isolation from soil and properties of the extreme thermophile *Clostridium thermohydrosulfuricum*. J Bacteriol 139:800–810. https://doi.org/10.1128/jb.139.3.800-810.1979.
37. Studier FW. 2005. Protein production by auto-induction in high-density shaking cultures. Protein Expr Purif 41:207–234. https://doi.org/10.1016/j.pep.2005.01.016.
38. Eiteman MA, Chastain MJ. 1997. Optimization of the ion-exchange analysis of organic acids from fermentation. Anal Chim Acta 338:69–75. https://doi.org/10.1016/S0003-2670(96)00426-6.
39. Kitamura Y, Abe Y, Yasui T. 1991. Metabolism of levoglucosan (2,6-anhydro-β-D-glucopyranose) in microorganisms. Agric Biol Chem 55:515–521. https://doi.org/10.1271/bbb1961.55.515.
40. Elzhov TV, Mullen KM, Spiess AN, Bolker B. 2016. Package minpack.lm. R interface to the Levenburg-Marquardt nonlinear least-squares algorithm found in MINPACK, plus support for bounds. https://cran.r-project.org/web/packages/minpack.lm/minpack.lm.pdf.
41. Langmead B, Salzberg S. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.
42. Coil D, Guillaume J, Darling AE. 2015. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. Bioinformatics 31:587–589. https://doi.org/10.1093/bioinformatics/btu661.
43. Yoon SH, Ha SM, Kwon S, Lim J, Kim Y, Seo H, Chun J. 2017. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int J Syst Evol Microbiol 67:1613–1617. https://doi.org/10.1099/ijsem.0.001755.
44. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki E, Zaslavsky L, Lomsadze A, Pruit KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.
45. Galloway NR, Toutkoushian H, Nune M, Bose N, Momany C. 2013. Rapid cloning for protein crystallography using type IIs restriction enzymes. Cryst Growth Des 13:2833–2839. https://doi.org/10.1021/cg400171z.