Green Chemistry

Metabolic modification of *Sphingobium lignivorans* SYK-6 for lignin valorization through the discovery of an unusual transcriptional repressor of lignin-derived dimer catabolism

Journal:	*Green Chemistry*
Manuscript ID	GC-ART-10-2024-005328.R1
Article Type:	Paper
Date Submitted by the Author:	17-Dec-2024

Complete List of Authors:	Kato, Ryo; Nagaoka University of Technology, Department of Materials Science and Bioengineering Kuatsjah, Eugene; National Renewable Energy Laboratory, Renewable Resources and Enabling Sciences Center Fujita, Masaya; Nagaoka University of Technology, Department of Materials Science and Bioengineering Bleem, Alissa; National Renewable Energy Laboratory, Bioenergy Hishiyama, Shojiro; Forestry and Forest Products Research Institute, Department of Forest Resource Chemistry Katahira, Rui; National Renewable Energy Laboratory, Renewable Resources and Enabling Sciences Center Senda, Toshiya; High Energy Accelerator Research Organization; The Graduate University for Advanced Studies, Beckham, Gregg; National Renewable Energy Laboratory, Bioenergy Kamimura, Naofumi; Nagaoka University of Technology, Department of Materials Science and Bioengineering Masai, Eiji; Nagaoka University of Technology, Department of Materials Science and Bioengineering

Green Foundation

1. In this study, we enhanced the metabolic capacity of *Sphingobium lignivorans* SYK-6, a promising strain for biological lignin valorization, to facilitate the rapid production of a polymer building block, 2-pyrone-4,6-dicarboxylic acid (PDC), from lignin-derived dimeric compounds.
2. We identified and characterized *ligS*, an unusual transcriptional repressor of lignin-derived aromatic dimer catabolism genes in SYK-6. Disrupting *ligS* in a PDC-producing strain derived from SYK-6 resulted in a 1.5-6.0-fold increase in the rate of PDC production from β-O-4- and β-1-type dimers and a metabolite of the β-O-4 dimer. Additionally, the time to achieve 100% yield of PDC production was reduced by 9-11 h when growing cells were used.
3. We aim further to elucidate the SYK-6 metabolic system and develop a metabolically engineered SYK-6 that will contribute to biological funneling at the industrial level.

Metabolic modification of *Sphingobium lignivorans* SYK-6 for lignin valorization through the discovery of an unusual transcriptional repressor of lignin-derived dimer catabolism

Ryo Kato [a], Eugene Kuatsjah [b†], Masaya Fujita [a], Alissa C. Bleem [b], Shojiro Hishiyama [c], Rui Katahira [b], Toshiya Senda [d], Gregg T. Beckham [b], Naofumi Kamimura [a], Eiji Masai [a,*]

[a] *Department of Materials Science and Bioengineering, Nagaoka University of Technology, Nagaoka, Niigata, Japan*
[b] *Renewable Resources and Enabling Sciences Center, National Renewable Energy Laboratory, Golden, CO, USA*
[c] *Department of Forest Resource Chemistry, Forestry and Forest Products Research Institute, Tsukuba, Ibaraki, Japan*
[d] *Structural Biology Research Center, Institute of Materials Structure Science, High Energy Accelerator Research Organization (KEK), Tsukuba, Ibaraki, Japan*

[†]Present address: Eugene Kuatsjah, Institute of Sustainability for Chemicals, Energy, and Environment (ISCE^2), Agency for Science, Technology, and Research (A*STAR), 138665 Singapore, Republic of Singapore

[*]Corresponding author: emasai@vos.nagaokaut.ac.jp.

Abstract

Sphingobium lignivorans SYK-6 catabolizes guaiacylglycerol-β-guaiacyl ether (GGE, a β-O-4-type dimer) and 1,2-diguaiacylpropane-1,3-diol (DGPD, a β-1-type dimer) derived from lignin. Recently, SLG_35860 containing TetR- and MarR-type transcriptional regulator motifs was suggested to be involved in the regulation of GGE and DGPD catabolism. Here we investigated the role of SLG_35860 in the transcriptional regulation of GGE and DGPD catabolism genes. SLG_35860 designated *ligS* repressed 11 genes involved in GGE and DGPD catabolism. LigS binds directly to specific sequences in the promoter region of each gene. The MarR domain was shown to be involved in these bindings; however, GGE, DGPD, and their metabolites did not function as effectors of LigS. We discovered unidentified compound(s) in the black liquor of oxygen-soda anthraquinone pulping of Japanese cedar that SYK-6 cannot metabolize and that acted as effector(s). Therefore, LigS constantly represses the transcription of the GGE and DGPD catabolism genes to low levels. Based on these findings, we examined the productivity of a polymer building block, 2-pyrone-4,6-dicarboxylic acid (PDC), from GGE, DGPD, and a GGE metabolite using an engineered *ligS* mutant. The rates of PDC production from each compound by this strain were 1.5–7.1 times higher than those of a PDC-producing strain carrying *ligS*.

Introduction

Lignin, a heterogeneous polymer composed of phenylpropanoid units, is a major component of plant cell walls and comprises up to ~30% of plant biomass [1]. Lignin is second only to cellulose in abundance and is the most abundant natural resource of aromatic compounds. Moreover, it is now well recognized that valorization of lignin is essential to establish a lignocellulosic bioeconomy [2]. However, due to its heterogeneous macromolecular structure and complexity, lignin cannot be utilized in its current form. Even if lignin is depolymerized, it yields heterogeneous low-molecular-weight aromatic compounds, which presents a challenge to its valorization [1, 3]. Recently, there has been considerable interest in approaches for lignin utilization that produce value-added products from lignin through a process that combines chemical depolymerization of lignin and biological funneling, which converts lignin-derived heterogeneous aromatic compounds into single target compounds [4-10]. The production of new polymer building blocks such as muconic acid, 2-pyrone-4,6-dicarboxylic acid (PDC), pyridine dicarboxylic acid, and vanillic acid from lignin through such processes has been reported [9, 11-14]. To utilize the metabolic capabilities of bacteria to produce target compounds from various lignin-derived aromatic compounds in high yields, these metabolic systems must be optimized for chemical production through genetic and metabolic engineering approaches. Therefore, a thorough characterization of the bacterial catabolic system of lignin-derived aromatic compounds, including regulatory systems, is crucial [4, 5, 8-10, 15].

In lignin, the phenylpropane units are connected through various C–O–C and C–C bonds, such as β-O-4, β-5, β-β, 5-5, and β-1 linkages [16, 17]. Among these intermolecular linkages, the β-O-4 linkage is the most abundant, accounting for ~45–50% and ~60% of the inter-monomer linkages in softwood and hardwood lignin, respectively [3]. Although the recalcitrant β-1 linkage is present in only up to ~9% of lignin, the degradation of compounds with a β-1 linkage is vital for the complete utilization of lignin [18-21]. *Sphingobium lignivorans* SYK-6 (hereafter, SYK-6) can catabolize β-O-4, β-1, β-5, and 5-5 type dimers, and its catabolic system has been almost wholly elucidated [22, 23]. Among these compounds, a β-O-4 dimer, guaiacylglycerol-β-guaiacyl ether (GGE), and a β-1 dimer, 1,2-diguaiacylpropane-1,3-diol (DGPD), have *threo* and *erythro* diastereomers, each containing enantiomers (**Fig. 1A–B**). In GGE catabolism, the four stereoisomers of GGE are converted to the achiral β-hydroxypropiovanillone (HPV) through oxidation and ether cleavage by multiple stereoselective Cα-dehydrogenases and glutathione *S*-transferases [24-27]. HPV is then oxidized at the Cγ position and converted to vanilloyl acetaldehyde, which is further metabolized [28]. In DGPD

catabolism, *threo*-DGPD undergoes continuous oxidation and reduction by multiple stereoselective Cα-dehydrogenases to *erythro*-DGPD with inverted stereochemistry [29]. The resulting *erythro*-DGPD is converted to achiral DGPD-stilbene by Cγ-formaldehyde lyase and further metabolized [30]. In this way, SYK-6 catabolizes GGE and DGPD stereoisomers into achiral compounds by a diverse group of stereoselective enzymes.

In SYK-6, the catabolism of lignin-derived aromatic monomers is controlled by specific transcriptional regulators: FerC for ferulic acid, DesR for vanillic acid and syringic acid, DesX for syringic acid, and LigR for protocatechuic acid [31-34]. These regulators recognize feruloyl-CoA, vanillic acid/syringic acid, vanillic acid/syringic acid, and protocatechuic acid/gallic acid as effectors to release transcriptional repression or activate the transcription of each transcriptional unit. In addition to SYK-6, transcriptional regulatory systems for ferulic acid, vanillic acid, protocatechuic acid, and 4-hydroxybenzoic acid catabolism genes have been reported in *Rhodococcus jostii*, *Corynebacterium glutamicum*, *Acinetobacter baylyi*, *Caulobacter crescentus*, *Agrobacterium tumefaciens*, and *Pseudomonas putida* [35-42]. There are very few studies on the transcriptional regulation of lignin-derived dimer catabolism, and the only reported example to our knowledge is the transcriptional regulatory system of 5,5′-dehydrodivanillic acid catabolism in SYK-6 [43]. Hence, transcriptional regulatory systems for β-O-4-, β-1-, and β-5-type dimer catabolism have not yet been demonstrated in any bacteria.

Recently, in randomly barcoded transposon insertion sequencing (RB-TnSeq) of SYK-6, we identified the SLG_35860 gene, whose disruption by transposon insertions led to markedly increased strain fitness during growth in the presence of GGE and DGPD [44]. SLG_35860 is predicted to be a transcriptional regulator with an unusual chimeric structure consisting of a TetR domain with a TetR-type helix-turn-helix (HTH) motif and a TetR C-terminal motif, and a MarR domain containing two MarR-type HTH motifs (**Fig. 2**; UniProt, C0SUK5). In this study, we investigated the involvement of SLG_35860 in the transcriptional regulation of GGE and DGPD catabolism genes. Transcriptional analysis of an SLG_35860 disruption mutant and biochemical analyses of the SLG_35860 gene product revealed that it comprehensively represses transcription of the GGE and DGPD catabolism genes to low levels. Surprisingly, the SLG_35860 gene product continued to bind to their promoter regions during GGE and DGPD catabolism. Based on these findings, a PDC-producing strain was created using the SLG_35860 disruption mutant as a platform, which significantly increased the rate of PDC production from GGE and DGPD compared with a PDC-producing strain with intact SLG_35860.

Materials and methods

Bacterial strains, plasmids, and culture conditions

The strains and plasmids used in this study are listed in **Table S1**. SYK-6 is our laboratory stock strain and has been deposited in the National Institute of Technology and Evaluation (NITE) Biological Resource Center and the Japan Collection of Microorganisms (JCM) under the accession numbers NBRC 103272 and JCM 17495. All strains were cultured in lysogeny broth (LB). SYK-6 and its mutants were cultured in LB containing 12.5 mg/L of nalidixic acid or Wx minimal medium containing 10 mM sucrose, 10 mM glutamate, 0.13 mM methionine, and 10 mM proline (SEMP) [34]. Complementary strains of SYK-6 and Δ*ligS* were cultured in LB containing 50 mg/L of kanamycin (Km) and 1 mM *m*-toluate (an inducer of the expression from P_m promoter in pJB861). The media for *E. coli* transformants were supplemented with 25 mg/L Km or 100 mg/L ampicillin (Amp). SYK-6, its mutants, and complementary strains were cultured at 30°C and *E. coli* at 37°C with shaking at 160 rpm.

Preparation of substrates

threo-DGPD, *erythro*-DGPD, DGPD-keto, α-(2-methoxyphenoxy)-β-hydroxypropiovanillone (MPHPV), HPV, and vanilloyl acetic acid were prepared as described in a previous study [28, 29, 45]. Vanillic acid was purchased from Sigma-Aldrich Co., LLC. GGE, guaiacol, acetovanillone, protocatechuic acid, syringic acid, and vanillin were purchased from Tokyo Chemical Industry Co., Ltd. PDC was prepared as described previously [46].

Construction of plasmids

Reporter plasmids were constructed using DNA fragments containing the putative promoter regions of *ligD*, *ligL*, *ligP*, *hpvZ*, *ligO*, and *ldpA*, which were PCR amplified from SYK-6 total DNA using Q5 high-fidelity DNA polymerase (New England Biolabs) and the primer pairs listed in **Table S2**. Each promoter region was predicted based on RT-PCR and RNA-Seq results (**Fig. S1**). The PCR products were inserted into the SphI and HindIII sites upstream of *lacZ* of pSEVA225 using NEBuilder HiFi DNA Assembly cloning kit (New England Biolabs). Expression plasmids for *ligS*, dTetR, and dMarR were constructed using DNA fragments carrying each region amplified through PCR using SYK-6 total DNA and the primer pairs listed in **Table S2**. The PCR products were inserted into the NdeI and BamHI sites downstream of the T7 promoter of pET-16b (pET*ligS*, pETdTetR, and pETdMarR) or NdeI and HindIII sites downstream of the *malE* of pMAL-c5X (pMAL*ligS*, pMALdTetR, and pMALdMarR). Nucleotide and amino acid sequences of dTetR and dMarR are shown in **Table S3**. For creating complementary plasmids, DNA fragments were prepared through PCR using SYK-6 total DNA, pMAL*ligS*, and pMALdMarR as templates and the primers

(**Table S2**). The PCR products were inserted into the BamHI site of pJB861 to create pJB*ligS*, pJBMBP*ligS*, and pJBMBPdMarR. The fidelity of nucleotide sequences of the inserts were confirmed by sequencing.

Construction of SYK-6 mutants

To construct the mutants, the upstream and downstream regions of the genes were amplified through PCR from SYK-6 total DNA using the primer pairs listed in **Table S2**. The resulting fragments were cloned into pAK405GFP using an NEBuilder HiFi DNA assembly cloning kit [30]. Each resulting plasmid was introduced into SYK-6 cells by triparental mating, and the resulting mutants were selected as described previously [47]. Gene deletion was confirmed through colony PCR using the primer pairs listed in **Table S2**.

Sequence analysis and LigS structure prediction

Nucleotide sequences were determined by DNA sequencing services (Eurofins Genomics). Sequence analysis and sequence similarity searches were performed using the MacVector program (MacVector, Inc.) and the BLASTP program, respectively (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Consensus sequences were aligned using ClustalW (https://www.genome.jp/tools-bin/clustalw) and displayed using weblogo3 (https://weblogo.threeplusone.com/create.cgi). The motif search and structure prediction were performed using InterProScan (https://www.ebi.ac.uk/interpro/search/sequence/) and Alphafold2, respectively [48].

Inducing and noninducing culture conditions and total RNA isolation

Cells of SYK-6, Δ*ligS*, SYK-6[pJB861], Δ*ligS*[pJB861], Δ*ligS*[pJB*ligS*], Δ*ligS*[pJBMBP*ligS*], and Δ*ligS*[pJBMBPdMarR] were grown in 5 mL LB for 24 h, harvested by centrifugation at 5,000 *g* for 5 min at 4°C, washed twice with Wx medium, and resuspended in the same medium. The resulting cells were inoculated into 5 mL of Wx-SEMP to an optical density at 600 nm (OD_{600}) of 0.2 and incubated until the OD_{600} of the culture reached 0.5 to 0.6. The cells were cultured for an additional 2 h as noninducing conditions and 6 h after adding 5 mM GGE or DGPD as inducing conditions. For the strains harboring the above plasmids, 1 mM *m*-toluate was added. Total RNA was isolated using Illustra RNAspin (Cytiva) and purified using Monarch RNA Cleanup Kit (New England Biolabs).

RNA-Seq, reverse transcription-PCR (RT-PCR), and quantitative RT-PCR (qRT-PCR)

Total RNA was obtained from SYK-6 and Δ*ligS* cells cultured in Wx-SEMP or Wx-SEMP containing 5 mM GGE or DGPD. RNA-Seq analysis of SYK-6 was performed on the Illumina platform (Novogene). Read quality was filtered using Trimmomatic version 0.39 [49] and mapped to the reference genome of SYK-6 (GCA_000283515.1) by bowtie2 version 2.5.0 [50]. Relative transcript

abundance was calculated as transcripts per million (TPM). For RT-PCR and qRT-PCR, cDNAs were synthesized from each total RNA by reverse transcription using a SuperScript IV Reverse Transcriptase (Invitrogen). The synthesized cDNA was purified using a NucleoSpin gel and PCR clean-up kit (Takara Bio). RT-PCR reactions were performed using purified cDNA synthesized from total RNA extracted from SYK-6 cells cultured in Wx-SEMP or Wx-SEMP containing 5 mM GGE, the primers (**Table S2**), and a Q5 high-fidelity DNA polymerase (New England Biolabs). qRT-PCRs reactions were performed with a StepOne Real-time PCR system (Applied Biosystems) using purified cDNA synthesized from each total RNA, the primers (**Table S2**), and a PowerUp SYBR Green Master Mix for qPCR (Thermo Fisher Scientific).

Promoter assay

Each reporter plasmid shown in **Table S1** was introduced into cells of SYK-6, Δ*ligS*, or Δ*ligDLN* through triparental mating. The resulting cells were inoculated into 5 mL Wx-SEMP containing Km and 5 mL Wx-SEMP containing 5 mM GGE and Km to an OD_{600} of 0.2 and incubated until the OD_{600} of the culture reached 0.5 to 0.6. Cells were harvested by centrifugation, washed twice with Wx medium (without solution II) [34], and resuspended in the same medium to an OD_{600} of 1.0. The β-galactosidase activity of the cells was measured using 2-nitrophenyl-β-D-galactopyranoside according to a modified Miller assay [https:// openwetware.org/wiki/Beta-Galactosidase_Assay_(A_better_Miller)] [51, 52].

Preparation of LigS, dTetR, and dMarR

pET-16b, pET*ligS*, pETdTetR, pMAL-c5X, pMAL*ligS*, and pMALdMarR (**Table S1**) were each introduced into *E. coli* BL21(DE3) cells. The transformed cells were grown in 5 mL LB, and each gene expression was induced for 4 h at 30°C by adding 1 mM isopropyl-β-D-thiogalactopyranoside when the OD_{600} of the cultures reached 0.5. Cells were collected by centrifugation (14,000 *g* for 1 min) and washed twice with 50 mM Tris-HCl buffer (pH 7.5; buffer A). The cells were resuspended in the same buffer and used as resting cells. Cells were broken by an ultrasonic disintegrator at an output ample of 80%, cycle for 10 s with 1.0-s intervals every second. This was done three times, and the supernatants of cell lysates were obtained as cell extracts after centrifugation (19,000 *g* for 15 min). LigS and dTetR were purified on a Ni Sepharose High Performance column (Cytiva). The His-tag was removed by an overnight incubation with Factor Xa (0.2 U/mg protein; Promega or New England Biolabs) at 20°C. The purified fractions were desalted and concentrated using an Amicon Ultra centrifugal filter unit (LigS: 30-kDa cutoff, TetR: 10-kDa cutoff; Merck Millipore). The samples obtained were stored at −80°C. For the preparation of MBP, MBPLigS, and MBPMarR, streptomycin (1 mg/mL) was added to cell extracts of *E. coli*[pMAL], *E. coli*[pMAL*ligS*], *E. coli*[pMALdMarR], and then the cell extracts were centrifuged to remove nucleic acids (19,000 *g*,

10 min). The resulting supernatant was used as MBP, MBPLigS, and MBPdMarR. Gene expressions and the purity of the enzymes were examined using SDS-12% polyacrylamide gel electrophoresis. The protein bands in gels were stained with Coomassie Brilliant Blue.

Molecular mass analysis of LigS

The molecular mass of LigS was determined using mass photometry [53]. The samples were diluted immediately before measurements in buffer A to 10–20 nM. The measurements were conducted using a Refeyn TwoMP mass photometer (Refeyn) with a data acquisition time of 60 s. DiscoverMP software (Refeyn) was used for data collection and analysis. Contrast-to-mass conversion was performed using a calibration with ovalbumin, bovine serum albumin, and aldolase. The recorded events were fitted to Gaussian distributions to calculate the molecular mass of the Gaussian fit.

Electrophoretic mobility shift assay (EMSA)

EMSA was performed according to a previous study [33, 54]. DNA fragments were prepared by amplification from the SYK-6 total DNA, oligomer annealing, or using synthetic DNA from Twist Bioscience (**Table S2, Table S4)**. For the reactions, purified LigS (0.01–1 μM), purified dTetR (17 μM), and cell extracts of *E. coli* harboring pMAL-c5X, pMAL*ligS*, and pMALMBPdMarR (10 ng/μL) were employed. To examine the association of LigS with effector molecules, a solution of GGE, MPHPV, HPV, DGPD, DGPD-keto, acetovanillone, guaiacol, vanillin, vanillic acid, protocatechuic acid, PDC (MPHPV, 25 mM; others, 50 mM), a black liquor (0.05–5% [vol/vol]), and a black liquor incubated with SYK-6 (5% [vol/vol]) was added to the reaction mixtures. To fractionate the black liquor, 100 μL of black liquor was first acidified to pH 1 by adding 45 μL of 12 *N* HCl. After saturating the solution with NaCl, an equivalent volume of ethyl acetate was added. After vigorous mixing and holding, the ethyl acetate layer was separated. After repeating this operation five times, the ethyl acetate was evaporated, and the extract was dissolved in 50% dimethyl sulfoxide to use the black liquor extract. After incubation, 1 μL loading buffer (50% [vol/vol] glycerol and 0.2% [vol/vol] bromophenol blue) was added, and samples were separated on a 2% agarose gel electrophoresis in 0.5× Tris/Borate/EDTA **(**TBE) buffer. After electrophoresis, the DNAs were detected using a SYBR Gold Nucleic Acid Gel Stain (Invitrogen) and photographed under a 470 nm blue LED. The black liquor obtained from oxygen-soda anthraquinone pulping of Japanese cedar was used after being neutralized with phosphoric acid [55].

Conversion of GGE, DGPD, and their metabolites by Δ*ligS*

Cells of SYK-6 and Δ*ligS* were grown in 5 mL LB for 24 h and collected by centrifugation (14,000 *g* for 1 min) and then washed twice with buffer A. The cells were resuspended in the same buffer

with an OD_{600} of 0.2–10 and used as resting cells. Resting cells were incubated in a 400 μL reaction mixture containing 80 μM GGE, 200 μM β*R*-MPHPV, 180 μM β*S*-MPHPV, 120 μM HPV, 40 μM DGPD I, 50 μM DGPD II, 70 μM DGPD-keto I, 70 μM DGPD-keto II, 100 μM *erythro*-DGPD,100 μM vanilloyl acetic acid, 100 μM vanillic acid, and 100 μM syringic acid (not a metabolite of GGE and DGPD) for 8 h at 30 °C. For each substrate conversion, cells were used with an OD_{600} of 10 (DGPD I), 1.0 (β*S*-MPHPV, β*R*-MPHPV, DGPD II, DGPD-keto I, DGPD-keto II, and *erythro*-DGPD, vanillic acid, and syringic acid), 0.5 (GGE and vanilloyl acetic acid), and 0.2 (HPV), respectively. After the reactions, the supernatants of the reaction mixtures were analyzed by high-performance liquid chromatography (HPLC).

PDC production using resting cells and growing cells

For PDC production using resting cells, the cells of Δ*ligI* and Δ*ligI ligS* were grown in 5 mL LB for 24 h, collected by centrifugation (14,000 *g* for 1 min), and then washed twice with buffer A. The cells were resuspended in the same buffer and used as resting cells. Resting cells were incubated in a 400 μL reaction mixture containing 100 μM GGE (OD_{600} of 0.5), 80 μM HPV (OD_{600} of 0.2), 80 μM DGPD I (OD_{600} of 10), or 60 μM DGPD II (OD_{600} of 1) and buffer A for 24 h at 30 °C. For PDC production using growing cells, the cells of Δ*ligI* and Δ*ligI ligS* were grown in 5 mL LB for 24 h, collected by centrifugation (14,000 *g* for 1 min), and then washed twice with buffer A. Washed cells were inoculated into 10 mL Wx-SEMP containing 1 mM GGE or 1 mM HPV to an OD_{600} of 0.2 and incubated with shaking for 24 h at 30°C. Cell growth was measured by OD_{600}. Portions of the cultures were periodically collected, and the reactions were stopped by centrifugation. The resultant supernatants were diluted, filtered, and analyzed using HPLC.

HPLC analysis

HPLC analysis was performed with the ACQUITY UPLC system (Waters). The sample solution was filtered through a PTFE filter (Captiva Econofilter, Agilent) with a pore size of 0.20 μm. The conversion of GGE, MPHPV, HPV, DGPD, and DGPD-keto and the PDC production from GGE and DGPD were analyzed using a TSKgel ODS-140HPT column (particle size, 2.3 μm; 2.1 × 100 mm, Tosoh). The mobile phase was a mixture of solution A (acetonitrile containing 0.1% formic acid or 0.5% phosphoric acid) and B (water containing 0.1% formic acid or 0.5% phosphoric acid) under the following conditions: 0–1.5 min, 12% A; 1.5–3.0 min, linear gradient from 12 to 60% A; 3.0–3.1 min, decreasing gradient from 60 to 12% A; 3.1–4.0 min, 12% A. The flow rate was 0.5 mL/min, and the column temperature was 30°C. Phosphate-containing solvents were used for the detection of PDC, while formic acid-containing solvents were used for the detection of GGE, MPHPV, HPV, DGPD, and DGPD-keto. GGE, MPHPV, DGPD, DGPD-keto, and vanilloyl acetic

acid were detected at 280 nm. HPV and syringic acid were detected at 275 nm. Vanillic acid and PDC were detected at 260 nm and 315 nm, respectively.

Results

Transcriptional inducibility and transcriptional units of β-O-4- and β-1-type dimer catabolism genes

We first investigated the transcriptional inducibility of nine genes involved in GGE catabolism (*ligD*, *ligL*, *ligN*, *ligF*, *ligE*, *ligP*, *ligG*, *ligQ*, and *hpvZ*) by RNA-Seq. SYK-6 cells were cultured in Wx-SEMP medium in the presence or absence of 5 mM GGE. Total RNA was isolated from cultured cells, and RNA-Seq analysis was performed. The transcripts per million (TPM) values of eight genes other than *ligQ* increased 3–41-fold under GGE-inducing conditions compared with noninducing conditions (**Table S5**). These results indicated that the transcription of GGE catabolism genes, with the exception of *ligQ*, is induced in the presence of GGE. In addition to *ligD*, *ligL*, and *ligN*, the other four DGPD catabolism genes, *ligO*, *ldpC*, *ldpB*, and *ldpA*, were also induced under GGE-inducing conditions with 4–17-fold increases in the TPM value.

In SYK-6, *ligDFEG* has been shown to form an operon in GGE catabolism genes (**Fig. 1C**) [25], and here we determined the transcriptional units of other genes. RT-PCR analysis was performed using total RNA from the GGE-inducing cells and showed amplification between the genes SLG_08630–*ligDFEG*–SLG_08680, *ldpB*–SLG_12630, *ldpA*–*ldpC*, SLG_32570–*ligP*, SLG_33680–*ligL*–SLG_33650, and SLG_35890–*ligN*–SLG_35920 (**Fig. S1A and B**). In contrast, *ligQ*, *hpvZ*, and *ligO* were shown to be transcribed as monocistronic units. To validate these RT-PCR results, we examined the RNA-Seq read coverage around the GGE and DGPD catabolism genes of SYK-6 samples cultured in the presence of GGE. For the *ldpB*–SLG_12630 and *ldpA*–*ldpC* regions, we confirmed increased transcription just upstream of *ldpB* and *ldpA* (**Fig. S1C**). However, in the SLG_32570–*ligP*, SLG_33680–*ligL*–SLG_33650, and SLG_35890–*ligN*–SLG_35920 regions, increased transcription was observed just upstream of *ligP*, *ligL*, and *ligN*. DNA amplification of SLG_32570–*ligP*, SLG_33680–*ligL*–SLG_33650, and SLG_35890–*ligN*–SLG_35920 appeared to be due to read-through from upstream, indicating that the promoters are located just upstream of *ligP*, *ligL,* and *ligN* (**Fig. 1C and Fig. S1C**).

Involvement of SLG_35860 in the transcription of GGE and DGPD catabolism genes

qRT-PCR was performed to investigate the precise transcriptional inducibility of GGE catabolism genes (*ligD*, *ligL*, *ligN*, *ligF*, *ligE*, *ligP*, *ligG*, *ligQ*, and *hpvZ*) and DGPD catabolism genes (*ligD*, *ligL*, *ligN*, *ligO*, *ldpB*, and *ldpA*; underlined genes indicate commonly used in both

catabolic pathways) in SYK-6. Total RNA was isolated from SYK-6 cultured in Wx-SEMP (noninducing conditions) and Wx-SEMP containing 5 mM GGE or 5 mM DGPD (GGE- or DGPD-inducing conditions), and the transcript amount was determined via qRT-PCR. The transcript amount of GGE catabolism genes other than *ligN*, *ligG*, and *ligQ* increased 2.3–17-fold under GGE-inducing conditions compared with noninducing conditions (**Fig. 3A**). The transcription of *ldpB* and *ldpA*, which are involved in DGPD catabolism was induced 27- and 14-fold, respectively, during DGPD-inducing conditions compared with noninducing conditions (**Fig. 3A**). However, *ligD*, *ligL*, *ligN*, and *ligO* did not show transcriptional inducibility under DGPD-inducing conditions.

In a previous RB-TnSeq experiment, we found the transposon insertional mutation of SLG_35860 caused a significant increase in fitness when it was grown on GGE and DGPD as a carbon source [44]. InterProScan search predicted that the SLG_35860 product consists of an N-terminal TetR domain (dTetR) composed of a TetR-type DNA-binding HTH motif and a TetR C-terminal motif, and a C-terminal MarR domain (dMarR) composed of two MarR-type HTH motifs (**Fig. 2**). Here, we created an SLG_35860 disruption mutant (**Fig. S2A**), and qRT-PCR analysis was performed using total RNA isolated from SLG_35860 mutant cells cultured in Wx-SEMP (noninducing conditions). The transcript levels of all GGE and DGPD catabolism genes except *ligQ* in this mutant increased 7.7–245-fold compared with those in SYK-6 grown under the same conditions (**Fig. 3B**). These results indicate that the SLG_35860 gene product comprehensively represses the transcription of both GGE and DGPD catabolism genes, and we designated this gene *ligS*.

Identification of promoter regions and inducer molecules

To identify the promoter regions of the GGE and DGPD catabolism genes, we performed promoter analysis of *ligD*, *ligL*, *ligP*, *hpvZ*, and *ldpA*, which are the first genes of each transcriptional unit and are transcriptionally inducible in the presence of GGE- or DGPD- inducing conditions. Each putative promoter region was cloned upstream of *lacZ* in the promoter probe vector pSEVA225 to create reporter plasmids (pD1, pL1, pP1, pZ1, and pA1), and the resulting plasmids were introduced into SYK-6. Each strain was grown under Wx-SEMP or Wx-SEMP containing 5 mM GGE, and promoter activity was measured. The promoter activities of *ligD* and *ligP* increased 3.2-fold and 5.7-fold, respectively, under GGE-inducing conditions compared with noninducing conditions. The *hpvZ* promoter activity slightly increased by 1.4-fold (**Fig. 4A**). To identify an inducer molecule, we generated a GGE-accumulating mutant of SYK6 (Δ*ligDLN*; **Fig. S2B**) and evaluated the promoter

activities of Δ*ligDLN* harboring pD1 or pP1. These strains were cultured in Wx-SEMP containing 5 mM GGE and showed equal or greater activity than SYK-6 incubated in the same medium (**Fig. 4A**). Thus, the GGE was shown to function as an inducer of *ligD* and *ligP*.

Next, each reporter plasmid was introduced into the *ligS* disruption mutant (Δ*ligS*), and promoter activity was measured in cells grown under noninducing conditions. All promoter activities in Δ*ligS* increased 2.5–13-fold compared with the wild-type strain grown under noninducing conditions (**Fig. 4A**). These results indicate that LigS negatively regulates the promoter activity of GGE and DGPD catabolism genes. It was also assumed that the GGE could function as an effector of LigS. Promoter regions were identified from promoter assays using various regions upstream of each gene (**Fig. 4B**). Putative −35 and −10 elements were found in the identified *ligD*, *ligL*, *ligP*, and *hpvZ* promoter regions (**Fig. S3**); however, they were not found in the *ldpA* promoter region. To examine whether the putative −35 elements upstream of *ligD*, *ligL*, *ligP*, and *hpvZ* have function as promoters, the promoter activities of each upstream region carrying mutations in the putative −35 elements were evaluated (**Fig. S3**). Consequently, the promoter activity disappeared by mutagenesis, indicating that these sequences are the −35 element (**Fig. 4B**).

Binding of LigS to each promoter region

EMSA was performed using purified LigS and DNA fragments from each promoter region (*ligD*, *ligL*, *ligP*, *hpvZ*, and *ldpA*) to evaluate direct binding. His-tag fused *ligS* was expressed in *E. coli* under the control of the T7 promoter, the gene product was purified by Ni-affinity chromatography, and the His-tag was removed to obtain purified LigS (**Fig. S4A and B**). The molecular mass of LigS was determined by mass photometry analysis (**Fig. S4C**). The main peak was 53 kDa, close to the theoretical molecular weight of His-tagged LigS (Mw, 59,095). This result suggests that LigS exists mainly as a monomer in solution. EMSAs revealed that LigS-DNA complexes were formed when probes of the upstream regions of *ligD*, *ligL*, *ligP*, and *hpvZ* (Dp1, Lp1, Pp1, and Zp1) were incubated with LigS at concentrations of 10–1,000 nM (**Fig. S5A**). In contrast, the upstream region probe of *ldpA* (Ap1) required 1,000 nM LigS to form a clearly shifted band, suggesting that this region has a lower affinity for LigS.

To identify LigS binding sequences, EMSA was performed using DNA probes with upstream regions of various lengths for each gene (**Fig. 5A, Fig. S5B**). DNA regions from 41 bp to 66 bp thought to bind LigS were found. Sequence alignment revealed that the 5'-**TTGA**NNN**TGT**NN**G**N**AA**NN**G**-3' consensus sequence was present in all four regions except the

ldpA upstream region (**Fig. 5B, Fig. S5C**). EMSA was performed using mutated DNA probes (m1, m2, and m3) of the upstream regions of *ligD*, *ligL*, *ligP*, and *hpvZ* with mutations introduced into consensus 1 (c1; TTGA) on the 5' side, c2 (TGT) in the middle, and c3 (GNAANNG) on the 3' side (**Fig. 5C**). The binding capacity of LigS was lost in all the mutant probes except the Pp3m2 probe (*ligP* promoter region), which presented a slightly shifted band (**Fig. 5C**). Reporter plasmids with mutations in the consensus sequence were also created and introduced into SYK-6 to evaluate promoter activity under noninducing conditions. The activities of the promoters with m2 and m3 mutations showed significantly greater compared with those of the wild type (**Fig. 5D**). In contrast, the activities of all the promoters with m1 mutation were lost except the *ligP* promoter (**Fig. 5D**). The reason for this would be that the C1 sequences of *ligD*, *ligL*, and *hpvZ* overlap with the −35 element, so the mutation rendered −35 inactive. The above in vitro and in vivo experiments demonstrated that LigS recognizes and binds to the consensus sequence.

Investigation of the binding of the TetR and MarR domains of LigS to the promoter regions

Each domain polypeptide of LigS was prepared and EMSA was performed to investigate which domain recognizes and binds to the consensus sequence. The His-tag fused TetR domain (His-dTetR) was produced in *E. coli* under the control of the T7 promoter, and it was purified by Ni-affinity chromatography (**Fig. S6**). Since the His-tag fused MarR domain polypeptide (His-dMarR) produced in *E. coli* was insoluble, the tag was changed to maltose-binding protein (MBP). However, MBP-dMarR was not adsorbed on the amylose resin; therefore, an *E. coli* cell extract containing MBP-dMarR was used for EMSA (**Fig. S6**). EMSA was performed using the *ligD*, *ligL*, *ligP*, and *hpvZ* upstream regions as probes and dTetR or MBP-dMarR. MBP-dMarR bound to the *ligD*, *ligL*, *ligP*, and *hpvZ* upstream regions, whereas dTetR did not bind to any probe (**Fig. 6A and B**). It was thus concluded that the MarR domain of LigS recognizes and binds the consensus sequence.

Next, qRT-PCR analysis of the GGE and DGPD catabolism genes in Δ*ligS* containing MBP-tag fused LigS (MBP-LigS) or MBP-dMarR was performed. The transcript amounts of *ligD*, *ligL*, *ligP*, and *hpvZ* were decreased in both strains than in the vector control strain (**Fig. 6C**). No difference in the *ligL* and *ligP* transcript levels was observed. However, the *ligD* and *hpvZ* transcript levels were significantly lower in the strain containing MBP-LigS. These results may suggest that the TetR domain is required for tighter repression of *ligD* and *hpvZ* transcription.

Exploring effectors of LigS

The results of the promoter assay suggested that LigS may recognize GGE as an effector (**Fig. 4A**). EMSA was performed to investigate whether the binding of LigS to the promoter regions of *ligD*, *ligL*, *ligP*, and *hpvZ* is released in the presence of GGE. The reaction of DNA fragments with LigS was performed in the presence of 50 mM GGE; however, LigS-DNA binding was not released (**Fig. S7A**). Other concentrations of GGE (1, 2, 5, 10, and 25 mM) were assayed, but no unbound species were observed (data not shown). Therefore, we investigated the effects of DGPD and metabolites of GGE and DGPD, including MPHPV, HPV, *threo*-DGPD, *erythro*-DGPD, DGPD-keto, vanillin, vanillic acid, acetovanillone, guaiacol, protocatechuic acid, and PDC. However, no release of LigS-DNA binding was observed under any of the conditions (**Fig. S7B**). Finally, considering the possibility that LigS has multiple substrate binding sites, we also examined the effects of combinations of GGE and other compounds, such as GGE + DGPD, but did not observe any effect on LigS-DNA binding under any conditions (**Fig. S7C**).

GGE, DGPD, and their metabolites were determined not to act as effectors of LigS. Therefore, we considered the possibility that the actual effector was present in the pulp effluent (kraft pulping) from which SYK-6 was isolated. As a first step, we focused on black liquor obtained from oxygen-soda anthraquinone pulping of Japanese cedar, which is stored in our laboratory. EMSA was performed using the upstream region of *ligD* in the presence of 0.5–5% neutralized black liquor, and the shifted bands disappeared (**Fig. S8A**). To narrow down the effector molecules, EMSA was performed in the presence of an ethyl acetate extract of black liquor acidified with hydrochloric acid, which resulted in the disappearance of the band shift (**Fig. S8B**). To determine whether this effector molecule is a compound that SYK-6 can metabolize, EMSA was performed in the presence of culture medium obtained by incubating the neutralized black liquor with SYK-6 for 24 h (equivalent to 1% neutralized black liquor solution). As a result, the shifted band caused by LigS binding disappeared (**Fig. S8C**). HPLC analysis of black liquor incubated with SYK-6 revealed that the major components of the black liquor, vanillic acid, vanillin, and acetovanillone, were consumed entirely (**Fig. S8D**). Therefore, lignin-derived dimers or oligomers, which cannot be catabolized or incorporated by SYK-6 cells, are likely to act as effectors of LigS.

PDC production from lignin-derived aromatic compounds using Δ*ligS* as a platform strain

Since LigS strongly represses the transcription of eight genes involved in β-O-4 dimer catabolism and seven genes involved in β-1 dimer catabolism, *ligS* disruption was expected to

markedly enhance the conversion capacity towards GGE and DGPD (**Fig. 7A**). The conversion rates of the resting cells of Δ*ligS* toward metabolites of GGE (vanilloyl acetic acid and vanillic acid) and syringic acid were comparable to those of SYK-6 (**Fig. S9**). In contrast, compared with the resting SYK-6 cells, the resting Δ*ligS* cells exhibited 3.2-, 2.1-, 2.4-, 2.2-, 1.5-, 3.3-, 2.8-, 3.7-, and 2.4-fold increased conversion of GGE, DGPD I, DGPD II, DGPD-keto II, *erythro*-DGPD (for 2 h of reaction), β*S*-MPHPV (for 1 h of reaction), β*R*-MPHPV, HPV, and DGPD-keto I (for 4 h of reaction), respectively (**Fig. 7B**). Therefore, Δ*ligS* could potentially be used as a platform strain to improve the productivity of polymer building blocks from these lignin-derived dimeric compounds.

PDC accumulation was engineered in SYK-6 by deleting the PDC hydrolase gene (SLG_12570; *ligI*) in wild-type cells and in strain Δ*ligS*, resulting in strains Δ*ligI* and Δ*ligI ligS* (**Fig. 8A and Fig. S2**). Resting cells prepared from Δ*ligI ligS* and Δ*ligI* grown in LB were incubated with 100 μM GGE, 80 μM HPV, 80 μM DGPD I, and 60 μM DGPD II. For PDC production from GGE, Δ*ligI* reached only 37% molar yield after 24 h, whereas Δ*ligI ligS* reached 100% molar yield after 8 h (**Fig. 8B**). Compared with the PDC production rate, when PDC was produced linearly (for 4 h of incubation), the rate increased ca. 6-fold in Δ*ligI ligS*. For PDC production from HPV, Δ*ligI* reached 100% molar yield after 8 h of reaction, while Δ*ligI ligS* reached 100% molar yield after 4 h (**Fig. 8B**). The PDC production rate (for 4 h of incubation) increased ca. 1.5-fold in Δ*ligI ligS*. For PDC production from DGPD I and DGPD II, Δ*ligI* showed 78% and 68% molar yields after 24 h, whereas Δ*ligI ligS* reached 100% and 90% molar yields after 4 h. The PDC production rate in Δ*ligI ligS* (for 2 h of incubation) increased ca. 1.9-fold and 1.6-fold for DGPD I and DGPD II conversion, respectively.

Next, PDC was produced from 1 mM GGE and 1 mM HPV using Δ*ligI ligS* and Δ*ligI* cells growing in Wx-SEMP. For PDC production from GGE, Δ*ligI* reached 100% molar yield at 24 h, whereas Δ*ligI ligS* reached 100% molar yield at 15 h of cultivation (**Fig. 8C**). For PDC production from HPV, Δ*ligI* reached 100% molar yield after 21 h of cultivation, whereas Δ*ligI ligS* reached 100% molar yield after 10 h (**Fig. 8C)**. In a previous study, *Pseudomonas putida* PpY1100 carrying the SYK-6 HPV catabolism genes produced PDC in 100% molar yield from 1 mM HPV after 24 h of cultivation [56]. By using Δ*ligI ligS*, this study could produce PDC from HPV more efficiently than the previous study.

Discussion and Conclusions

This study revealed that an unusual transcriptional regulator, LigS, which has both TetR and MarR domains and cannot be classified into any known family of transcriptional regulators, comprehensively represses the transcription of GGE and DGPD catabolism genes to low levels. LigS represses the transcription of eight out of nine genes (*ligD*, *ligL*, *ligN*, *ligF*, *ligE*, *ligP*, *ligG*, and *hpvZ*) and seven genes (*ligD*, *ligL*, *ligN*, *ligO*, *ldpB*, *ldpC*, and *ldpA*) involved in the conversion of GGE to vanilloyl acetaldehyde and in the conversion of *threo*-DGPD to DGPD-stilbene, respectively. Since the disruption of *ligS* did not affect the ability to convert vanilloyl acetic acid, vanillic acid, and syringic acid, it seems that the transcriptional repression by LigS is specific to the GGE and DGPD catabolic genes (Fig. S9). In a previous RB-TnSeq analysis, the *ligS* mutation significantly increased fitness when GGE and DGPD were used as carbon sources [44]. Conversely, there was a slight decline in most cases when other lignin-derived compounds were utilized as carbon sources. These results support that LigS specifically regulates the genes involved in GGE and DGPD catabolism. BLASTP searches revealed orthologs of *ligS* in 53 strains belonging to alphaproteobacteria (**Table S6**). Among these, the full length of LigS was conserved in 20 strains, all demonstrating high genomic similarity to SYK-6 (78%–96% identity). The total conservation of the GGE and DGPD catabolic systems in these 20 strains implies that the regulation of these genes by LigS was acquired through evolution within a specific group of Sphingomonadaceae.

TetR- and MarR-type transcriptional regulators are known to negatively regulate transcription by forming homodimers and binding to promoter regions, inhibiting RNA polymerase binding or transcription elongation [57-59]. Molecular mass analysis and EMSA showed that LigS exists as a monomer and that only the MarR domain recognizes the LigS binding sequence (**Fig. 6B, Fig. S4C**). Since the MarR domain of LigS consists of two MarR motifs, it is likely that this domain can bind to DNA as efficiently as the dimer of MarR-type transcriptional regulators. However, qRT-PCR analysis revealed that, unlike those of *ligL* and *ligP*, the levels of transcriptional repression of *ligD* and *hpvZ* in the Δ*ligS* that produced dMarR were lower than those in the Δ*ligS* that produced full-length LigS (**Fig. 6C**). No known transcriptional regulator resembles LigS; however, PobR, a dimeric protein involved in the catabolism of 4-hydroxybenzoic acid in *Streptomyces coelicolor*, has been reported to contain two IclR-type transcriptional regulator motifs in each monomer. It has been suggested that PobR cannot repress using only one motif and functions solely in the presence of both motifs [60]. The TetR domain of LigS may also be partially involved in DNA binding, which enhances

LigS binding. The binding sequences of MarR-type transcriptional regulators are generally inverted repeat sequences [58]. However, the binding sequence of LigS was 5'-**TTGA**NNN**TGT**NN**G**N**AA**NN**G**-3', and no inverted repeat sequence was observed. This is likely due to the low identity of the amino acid sequence between the MarR motifs in LigS (19.2%), resulting in a nonrepeating binding sequence. The LigS binding sequences upstream of *ligD*, *ligL*, and *hpvZ* overlapped with the −35 elements (**Fig. S3**). Thus, LigS is thought to inhibit the recruitment of RNA polymerase to the promoter region, consistent with the mode of action of MarR-type transcriptional regulators. Unlike with the other promoters studied, in the case of the *ligP* upstream region, LigS was found to bind 28 bp upstream of the −35 element. The binding of LigS to this site is also thought to affect RNA polymerase binding. Although the transcription of *ligN*, *ligO,* and *ldpA* was repressed by LigS, no consensus sequence was found upstream of each gene (**Fig. 3B**). However, at least LigS showed weak binding to the upstream region of *ldpA* (**Fig. S5A**). These results suggest another weak binding sequence for LigS upstream of these genes.

GGE was shown to function as an inducer of expression from the *ligD* and *ligP* promoters (**Fig. 4A**). However, the effector molecules of LigS were not GGE, DGPD, or their metabolites. These facts indicate that there is a GGE-inducible transcriptional regulatory system for transcription from the above promoters that is independent of LigS regulation, which remains to be elucidated. The search for molecules that function as effectors of LigS discovered that the black liquor of oxygen-soda anthraquinone pulping contained the target molecules (**Fig. S8A**). LigS-DNA binding was released in the presence of the ethyl acetate extract of black liquor and black liquor incubated with SYK-6. These results suggest that the effector molecules are lignin-derived dimers or oligomers that cannot be catabolized or incorporated by SYK-6 cells (**Fig. S8B–D**). These facts may be related to the environment in which SYK-6 initially inhabited: the pond used to treat effluent from the kraft pulp mill from which SYK-6 was isolated was presumably rich in lignin-derived dimers and oligomers. The environment may contain effector molecules that SYK-6 can take up and metabolize. Further identification of effector molecules is expected in the future.

The conversion rates of GGE, DGPD, and their metabolites by the resting cells of Δ*ligS* were 1.5- to 3.7-fold higher than those of SYK-6 (**Fig. 7B**). The rate of PDC production from GGE, HPV, and DGPD using resting cells was 1.5- to 6.0-fold faster with Δ*ligI ligS* than with Δ*ligI* (**Fig. 8B**). Furthermore, the time required to produce PDC from GGE and HPV in 100% molar yield using growing cells of Δ*ligI ligS* was reduced by 9 and 11 h, respectively, compared with growing cells of Δ*ligI* (**Fig. 8C**). Since LigS independently and comprehensively represses the GGE and DGPD

catabolism genes to low levels, the disruption of *ligS* increased the conversion rates for each metabolite in addition to GGE and DGPD. The lack of derepression by LigS when SYK-6 catabolizes GGE and DGPD indicates that LigS consistently represses both GGE and DGPD catabolism genes. Therefore, the use of Δ*ligI ligS* resulted in a significant increase in the rate of PDC production from GGE, HPV, and DGPD. However, when cultured in SEMPs containing GGE or HPV, the growth of Δ*ligI ligS* was reduced compared to Δ*ligI* (**Fig. 8C**). Additionally, in the RB-TnSeq analysis, a slight decrease in fitness was observed when lignin-derived aromatic compounds other than GGE and DGPD were utilized as carbon sources (the reduction was more significant for syringic acid) [44]. Thus, the disruption of *ligS* appears to have implications beyond the transcriptional regulation of the catabolism of lignin-derived aromatic compounds. Further investigation into this matter may provide insights into the physiological importance of LigS.

This study characterized LigS, an unusual regulator that constantly represses the transcription of a series of catabolism genes to low levels. To our knowledge, no such transcriptional repressors have been reported. The presence of LigS in SYK-6 was first discovered via RB-TnSeq analysis, and it is almost impossible to determine its presence via general transcriptional analysis [44]. This study implies that unknown regulatory systems may be hidden even in well-characterized microbial catabolic systems. A comprehensive analysis of genome-wide genetic mutants is instrumental for fully understanding catabolic systems and developing bacterial strains maximally optimized for use in biological lignin valorization.

Authorship contribution statement

Ryo Kato: Conceptualization, Investigation, Writing - original draft, Writing - Review & Editing. **Eugene Kuatsjah:** Investigation, Writing - Review & Editing. **Masaya Fujita:** Investigation, Writing - Review & Editing. **Alissa Bleem:** Investigation, Writing - Review & Editing. **Shojiro Hishiyama:** Resources. **Rui Katahira:** Resources, Writing - Review & Editing. **Toshiya Senda:** Supervision, Writing - Review & Editing. **Gregg T. Beckham:** Conceptualization, Writing - Review & Editing, Supervision, Funding acquisition. **Naofumi Kamimura:** Conceptualization, Writing - Review & Editing, Supervision. **Eiji Masai**: Conceptualization, Writing - original draft, Writing – Review & Editing, Supervision, Project administration, Funding acquisition.

Conflicts of interest

There are no conflicts to declare.

Data availability statements

The data supporting this article have been included as part of the Supplementary Information.

Acknowledgements

We thank Drs. Kengo Magara and Satoshi Kubo at the Forestry and Forest Products Research Institute for providing the black liquor, and Dr. Takuya Akiyama at the University of Tokyo for valuable discussions. This work was supported in part by JST Grant Number JPMJPF2104 and JST CREST Grant Number JPMJCR23L4. This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. For EK, ACB, RK, and GTB, this material is based upon work supported by the Center for Bioenergy Innovation (CBI), U.S. Department of Energy, Office of Science, Biological and Environmental Research Program under Award Number ERKP886. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

References

1. A. J. Ragauskas, G. T. Beckham, M. J. Biddy, R. Chandra, F. Chen, M. F. Davis, B. H. Davison, R. A. Dixon, P. Gilna, M. Keller, P. Langan, A. K. Naskar, J. N. Saddler, T. J. Tschaplinski, G. A. Tuskan and C. E. Wyman, *Science*, 2014, **344**, 1246843.
2. L. R. Lynd, G. T. Beckham, A. M. Guss, L. N. Jayakody, E. M. Karp, C. Maranas, R. L. McCormick, D. Amador-Noguez, Y. J. Bomble, B. H. Davison, C. Foster, M. E. Himmel, E. K. Holwerda, M. S. Laser, C. Y. Ng, D. G. Olson, Y. Román-Leshkov, C. T. Trinh, G. A. Tuskan, V. Upadhayay, D. R. Vardon, L. Wang and C. E. Wyman, *Energ Environ Sci*, 2022, **15**, 938-990.
3. J. Zakzeski, P. C. Bruijnincx, A. L. Jongerius and B. M. Weckhuysen, *Chem Rev*, 2010, **110**, 3552-3599.
4. J. G. Linger, D. R. Vardon, M. T. Guarnieri, E. M. Karp, G. B. Hunsinger, M. A. Franden, C. W. Johnson, G. Chupka, T. J. Strathmann, P. T. Pienkos and G. T. Beckham, *Proc Natl Acad Sci U S A*, 2014, **111**, 12013-12018.
5. T. D. Bugg and R. Rahmanpour, *Curr Opin Chem Biol*, 2015, **29**, 10-17.
6. O. Y. Abdelaziz, D. P. Brink, J. Prothmann, K. Ravi, M. Z. Sun, J. Garcia-Hidalgo, M. Sandahl, C. P. Hulteberg, C. Turner, G. Liden and M. F. Gorwa-Grauslund, *Biotechnol Adv*, 2016, **34**, 1318-1346.
7. G. T. Beckham, C. W. Johnson, E. M. Karp, D. Salvachúa and D. R. Vardon, *Curr Opin Biotechnol*, 2016, **42**, 40-53.
8. J. Becker and C. Wittmann, *Biotechnol Adv*, 2019, **37**, 107360.
9. C. W. Johnson, D. Salvachua, N. A. Rorrer, B. A. Black, D. R. Vardon, P. C. St John, N. S. Cleveland, G. Dominick, J. R. Elmore, N. Grundl, P. Khanna, C. R. Martinez, W. E. Michener, D. J. Peterson, K. J. Ramirez, P. Singh, T. A. VanderWall, A. N. Wilson, X. N. Yi, M. J. Biddy, Y. J. Bomble, A. M. Guss and G. T. Beckham, *Joule*, 2019, **3**, 1523-1537.
10. F. Weiland, M. Kohlstedt and C. Wittmann, *Metab Eng*, 2022, **71**, 13-41.
11. Z. Mycroft, M. Gomis, P. Mines, P. Law and T. D. H. Bugg, *Green Chem*, 2015, **17**, 4974-4979.
12. D. Salvachúa, C. W. Johnson, C. A. Singer, H. Rohrer, D. J. Peterson, B. A. Black, A. Knapp and G. T. Beckham, *Green Chem*, 2018, **20**, 5007-5019.
13. Y. Higuchi, H. Ishimaru, T. Yoshikawa, T. Masuda, C. Sakamoto, N. Kamimura, E. Masai, D. Takeuchi and T. Sonoki, *Bioresour Technol*, 2023, **385**, 129450.

14. Y. Otsuka, T. Araki, Y. Suzuki, M. Nakamura, N. Kamimura and E. Masai, *Bioresour Technol*, 2023, **377**, 128956.

15. E. Erickson, A. Bleem, E. Kuatsjah, A. Z. Werner, J. L. DuBois, J. E. McGeehan, L. D. Eltis and G. T. Beckham, *Nat Catal*, 2022, **5**, 86-98.

16. R. Vanholme, B. Demedts, K. Morreel, J. Ralph and W. Boerjan, *Plant Physiol*, 2010, **153**, 895-905.

17. J. Ralph, C. Lapierre and W. Boerjan, *Curr Opin Biotechnol*, 2019, **56**, 240-249.

18. M. P. Pandey and C. S. Kim, *Chem Eng Technol*, 2011, **34**, 29-41.

19. K. M. Torr, D. J. van de Pas, E. Cazeils and I. D. Suckling, *Bioresour Technol*, 2011, **102**, 7608-7611.

20. S. Van den Bosch, W. Schutyser, R. Vanholme, T. Driessen, S. F. Koelewijn, T. Renders, B. De Meester, W. J. J. Huijgen, W. Dehaen, C. M. Courtin, B. Lagrain, W. Boerjan and B. F. Sels, *Energy Environ Sci*, 2015, **8**, 1748-1763.

21. K. Saito, A. Kaiho, R. Sakai, H. Nishimura, H. Okada and T. Watanabe, *J Agric Food Chem*, 2016, **64**, 9152-9160.

22. E. Masai, Y. Katayama and M. Fukuda, *Biosci Biotechnol Biochem*, 2007, **71**, 1-15.

23. N. Kamimura, K. Takahashi, K. Mori, T. Araki, M. Fujita, Y. Higuchi and E. Masai, *Environ Microbiol Rep*, 2017, **9**, 679-705.

24. E. Masai, A. Ichimura, Y. Sato, K. Miyauchi, Y. Katayama and M. Fukuda, *J Bacteriol*, 2003, **185**, 1768-1775.

25. Y. Sato, H. Moriuchi, S. Hishiyama, Y. Otsuka, K. Oshima, D. Kasai, M. Nakamura, S. Ohara, Y. Katayama, M. Fukuda and E. Masai, *Appl Environ Microbiol*, 2009, **75**, 5195-5201.

26. K. Tanamura, T. Abe, N. Kamimura, D. Kasai, S. Hishiyama, Y. Otsuka, M. Nakamura, S. Kajita, Y. Katayama, M. Fukuda and E. Masai, *Biosci Biotechnol Biochem*, 2011, **75**, 2404-2407.

27. Y. Higuchi, D. Sato, N. Kamimura and E. Masai, *Sci Rep*, 2020, **10**, 20614.

28. Y. Higuchi, S. Aoki, H. Takenami, N. Kamimura, K. Takahashi, S. Hishiyama, C. S. Lancefield, O. S. Ojo, Y. Katayama, N. J. Westwood and E. Masai, *Appl Environ Microbiol*, 2018, **84**.

29. R. Kato, K. Maekawa, S. Kobayashi, S. Hishiyama, R. Katahira, M. Nambo, Y. Higuchi, E. Kuatsjah, G. T. Beckham, N. Kamimura and E. Masai, *Appl Environ Microbiol*, 2023, **89**, e0017123.

30. E. Kuatsjah, M. Zahn, X. Chen, R. Kato, D. J. Hinchen, M. O. Konev, R. Katahira, C. Orr, A. Wagner, Y. Zou, S. J. Haugen, K. J. Ramirez, J. K. Michener, A. R. Pickford, N. Kamimura,

E. Masai, K. N. Houk, J. E. McGeehan and G. T. Beckham, *Proc Natl Acad Sci U S A*, 2023, **120**, e2212246120.

31. N. Kamimura, K. Takamura, H. Hara, D. Kasai, R. Natsume, T. Senda, Y. Katayama, M. Fukuda and E. Masai, *J Bacteriol*, 2010, **192**, 3394-3405.
32. D. Kasai, N. Kamimura, K. Tani, S. Umeda, T. Abe, M. Fukuda and E. Masai, *FEMS Microbiol Lett*, 2012, **332**, 68-75.
33. T. Araki, S. Umeda, N. Kamimura, D. Kasai, S. Kumano, T. Abe, C. Kawazu, Y. Otsuka, M. Nakamura, Y. Katayama, M. Fukuda and E. Masai, *Sci Rep*, 2019, **9**, 18036.
34. T. Araki, K. Tanatani, N. Kamimura, Y. Otsuka, M. Yamaguchi, M. Nakamura and E. Masai, *Appl Environ Microbiol*, 2020, **86**, e01712-01720.
35. A. A. DiMarco, B. Averhoff and L. N. Ornston, *J Bacteriol*, 1993, **175**, 4499-4506.
36. S. Romero-Steiner, R. E. Parales, C. S. Harwood and J. E. Houghton, *J Bacteriol*, 1994, **176**, 5771-5779.
37. D. Parke, *J Bacteriol*, 1996, **178**, 266-272.
38. U. Gerischer, A. Segura and L. N. Ornston, *J Bacteriol*, 1998, **180**, 1512-1524.
39. B. Morawski, A. Segura and L. N. Ornston, *FEMS Microbiol Lett*, 2000, **187**, 65-68.
40. K. X. Zhao, Y. Huang, X. Chen, N. X. Wang and S. J. Liu, *J Bacteriol*, 2010, **192**, 1565-1572.
41. K. Morabbi Heravi, J. Lange, H. Watzlawick, J. Kalinowski and J. Altenbuchner, *J Bacteriol*, 2015, **197**, 959-972.
42. H. Otani, P. J. Stogios, X. Xu, B. Nocek, S. N. Li, A. Savchenko and L. D. Eltis, *Nucleic Acids Res*, 2016, **44**, 595-607.
43. K. Mori, K. Niinuma, M. Fujita, N. Kamimura and E. Masai, *Appl Environ Microbiol*, 2018, **84**, e01314-01318.
44. A. Bleem, R. Kato, Z. A. Kellermyer, R. Katahira, M. Miyamoto, K. Niinuma, N. Kamimura, E. Masai and G. T. Beckham, *Cell Rep*, 2023, **42**, 112847.
45. S. Hishiyama, Y. Otsuka, M. Nakamura, S. Ohara, S. Kajita, E. Masai and Y. Katayama, *Tetrahedron Lett*, 2012, **53**, 842-845.
46. Y. Otsuka, M. Nakamura, K. Shigehara, K. Sugimura, E. Masai, S. Ohara and Y. Katayama, *Appl Microbiol Biotechnol*, 2006, **71**, 608-614.
47. A. Kaczmarczyk, J. A. Vorholt and A. Francez-Charlot, *Appl Environ Microbiol*, 2012, **78**, 3774-3777.
48. K. Tunyasuvunakool, J. Adler, Z. Wu, T. Green, M. Zielinski, A. Zidek, A. Bridgland, A. Cowie, C. Meyer, A. Laydon, S. Velankar, G. J. Kleywegt, A. Bateman, R. Evans, A. Pritzel,

M. Figurnov, O. Ronneberger, R. Bates, S. A. A. Kohl, A. Potapenko, A. J. Ballard, B. Romera-Paredes, S. Nikolov, R. Jain, E. Clancy, D. Reiman, S. Petersen, A. W. Senior, K. Kavukcuoglu, E. Birney, P. Kohli, J. Jumper and D. Hassabis, *Nature*, 2021, **596**, 590-596.

49. A. M. Bolger, M. Lohse and B. Usadel, *Bioinformatics*, 2014, **30**, 2114-2120.
50. B. Langmead and S. L. Salzberg, *Nat Methods*, 2012, **9**, 357-359.
51. J. H. Miller, *Experiments in molecular genetics*, 1972.
52. X. Zhang and H. Bremer, *J Biol Chem*, 1995, **270**, 11181-11189.
53. A. Sonn-Segev, K. Belacic, T. Bodrug, G. Young, R. T. VanderLinden, B. A. Schulman, J. Schimpf, T. Friedrich, P. V. Dip and T. U. Schwartz, *Nat Commun*, 2020, **11**, 1772.
54. J. A. Ream, L. K. Lewis and K. A. Lewis, *Anal. biochem.*, 2016, **511**, 36-41.
55. K. Magara and S. Kubo, *Japan TAPPI J*, 2019, **73**, 249-258.
56. Y. Higuchi, R. Kato, K. Tsubota, N. Kamimura, N. J. Westwood and E. Masai, *Metab Eng*, 2019, **55**, 258-267.
57. J. L. Ramos, M. Martinez-Bueno, A. J. Molina-Henares, W. Teran, K. Watanabe, X. Zhang, M. T. Gallegos, R. Brennan and R. Tobes, *Microbiol Mol Biol Rev*, 2005, **69**, 326-356.
58. I. C. Perera and A. Grove, *J Mol Cell Biol*, 2010, **2**, 243-254.
59. J. Guo, X. Zhang, S. Luo, F. He, Z. Chen, Y. Wen and J. Li, *PloS One*, 2013, **8**, e71330.
60. R. Zhang, D. M. Lord, R. Bajaj, W. Peti, R. Page and J. K. Sello, *Nucleic Acids Res*, 2018, **46**, 1501-1512.

Figure legends

Fig. 1. Catabolic pathways of GGE and DGPD in SYK-6 and the organization of their catabolism genes. (A) The GGE catabolic pathway. Enzymes: LigD, LigL, and LigN, Cα-dehydrogenases; LigF, LigE, and LigP, β-etherases; LigG and LigQ, glutathione removing enzymes; HpvZ, HPV oxidase. Abbreviations: GGE, guaiacylglycerol-β-guaiacyl ether; MPHPV, α-(2-methoxyphenoxy)-β-hydroxypropiovanillone; GS-HPV, α-glutathionyl-β-hydroxypropiovanillone; HPV, β-hydroxypropiovanillone; VAL, vanilloyl acetaldehyde; GS^-, reduced glutathione; GSSG, oxidized glutathione. (B) The DGPD catabolic pathway. Enzymes: LigD, LigL, LigN, LigO, LdpB, and LdpC, Cα-dehydrogenases; LdpA, *erythro*-DGPD Cγ-formaldehyde lyase. Abbreviations: DGPD, 1,2-diguaiacylpropane-1,3-diol. (C) Organization of the GGE and DGPD catabolism genes. The arrows indicate transcription units. The ORF shown in blue and red indicate that the respective genes are involved in GGE and DGPD catabolism, respectively.

Fig. 2. Structure and motif prediction of the SLG_35860 (*ligS*) gene product. The SLG_35860 (*ligS*) gene product comprises four motifs: a TetR-type DNA binding helix-turn-helix (HTH) motif, a TetR-type C-terminal motif, and two MarR-type HTH motifs. The structure of the SLG_35860 (*ligS*) gene product was predicted by AlphaFold2.

Fig. 3. Transcription levels of the GGE and DGPD catabolism genes in SYK-6 and the SLG_35860 mutant. (A) qRT-PCR analysis of the GGE and DGPD catabolism genes in SYK-6. Total RNAs were isolated from SYK-6 cells grown in Wx-SEMP, Wx-SEMP + 5 mM GGE, and Wx-SEMP + 5 mM DGPD. (B) qRT-PCR analysis of the GGE and DGPD catabolism genes in SLG_35860 mutant. Total RNAs were isolated from SLG_35860 mutant cells grown in Wx-SEMP. The value for each amount of mRNA was normalized to the level of 16S rRNA. Each value is the average ± standard deviation (error bars) of three independent experiments. Statistical differences were determined by one-way ANOVA with Dunnett's multiple-comparison test (A) or Student's t-test (B). The asterisks of one-way ANOVA with Dunnett's multiple-comparison test indicate statistically significant differences between the values linked by brackets (ns, $P > 0.01$; *, $P < 0.01$; **, $P < 0.001$; ***, $P < 0.0001$), and the asterisks of Student's t-test indicate statistically significant differences between the values linked by brackets (ns, $P > 0.05$; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$).

Fig. 4. Identification of the promoter regions of the GGE and DGPD catabolism genes. (A) Promoter activities of the cells of SYK-6, *ligD ligL ligN* triple mutant (Δ*ligDLN*), and SLG_35860 mutant (Δ*ligS*) harboring pD1 (carrying upstream regions of *ligD*), pL1 (carrying upstream regions of *ligL*), pP1 (carrying upstream regions of *ligP*), pZ1(carrying upstream regions of *hpvZ*), pA1 (carrying upstream regions of *ldpA*) or pSEVA225 (vector). The cells used for the assay were grown in Wx-SEMP or Wx-SEMP + 5 mM GGE. Each bar value indicates Miller units, and each value is the average ± the standard deviation (error bars) of three independent experiments. (B) Promoter activities of SYK-6 cells harboring each reporter plasmid carrying deleted promoter regions or mutated promoters. The cells used for the assay were grown in Wx-SEMP. Putative −35 elements

are shown in green boxes. The mutated sequences in the putative −35 sequences are shown in Fig. S3. Each value is the average ± standard deviation (error bars) of three independent experiments.

Fig. 5. Identification of LigS binding sequences. (A) EMSAs of LigS (100 nM) binding to DNA probes of deleted upstream regions of *ligD*, *ligL*, *ligP*, and *hpvZ* (40 fmol). (B) The consensus sequences found in the promoter regions of *ligD*, *ligL*, *ligP*, and *hpvZ* are shown. Probes with mutations introduced into the consensus sequence were used for EMSA. (C) EMSAs of LigS (100 nM) binding to Dp6 (a *ligD* promoter region probe), Lp6 (a *ligL* promoter region probe), Pp3 (a *ligP* promoter region probe), Zp6 (an *hpvZ* promoter region probe), and these mutant probes (40 fmol). Each probe was incubated with (+) or without (-) LigS. (D) Promoter activities of SYK-6 cells harboring each reporter plasmid with mutations in the consensus sequence. The cells used for the assay were grown in Wx-SEMP. Each value is the average ± standard deviation (error bars) of three independent experiments.

Fig. 6. Characterization of the TetR and MarR domains of LigS. (A) EMSAs of dTetR (17 μM) binding to *ligD*, *ligL*, *ligP*, and *hpvZ* promoter region probes (40 fmol). Each probe was incubated with (+) and without (-) purified dTetR. (B) EMSAs of dMarR binding to *ligD*, *ligL*, *ligP*, and *hpvZ* promoter region probes. Crude MBP, MBP-LigS, and dMarR (10 μg/mL) were incubated with the above DNA probes. (C) Transcription of *ligD*, *ligL*, *ligP*, and *hpvZ* in Δ*ligS* cells complemented with MBP-LigS and dMarR. SYK-6[pJB861 (vector)] cells, Δ*ligS*[pJB861] cells, Δ*ligS*[pJB*ligS*] cells, Δ*ligS*[pJBMBP*ligS*] cells, and Δ*ligS*[pJBMBPdMarR] cells were grown in Wx-SEMP + 1 mM *m*-toluic acid, and total RNAs were isolated from the cells. The value for each amount of mRNA was normalized to the level of 16S rRNA. Each value is the average ± standard deviation (error bars) of three independent experiments.

Fig. 7. Disruption of *ligS* improved the capacity to convert GGE, DGPD, and their metabolites. (A) LigS-repressed GGE and DGPD catabolic pathways. (B) Conversion of GGE, DGPD, and these intermediate metabolites by resting cells of SYK-6 (black) and Δ*ligS* (orange), respectively. Cells with an OD_{600} of 10 (DGPD I), 1.0 (β*S*-MPHPV, β*R*-MPHPV, DGPD II, DGPD-keto I, DGPD-keto II, and *erythro*-DGPD), 0.5 (GGE), and 0.2 (HPV), respectively, were used.

Fig. 8. Drastic increase in PDC production rate in Δ*ligI ligS*. (A) Schematic of the catabolic system of Δ*ligI* (left) and Δ*ligI ligS* (right) used for PDC production. (B) Conversion of 100 μM GGE, 80 μM HPV, 80 μM DGPD I, and 60 μM DGPD II by resting cells of Δ*ligI* and Δ*ligI ligS*. For the conversion of GGE, HPV, DGPD I, and DGPD II, cells with an OD_{600} of 0.5, 0.2, 10, and 1, respectively, were used. (C) Conversion of 1 mM GGE and HPV by Δ*ligI* and Δ*ligI ligS* cells during growth in Wx-SEMP. Portions of the reaction mixtures were collected and analyzed by HPLC. All experiments were performed in triplicate, and each value represents the means ± standard deviation (error bars).

Fig. 1

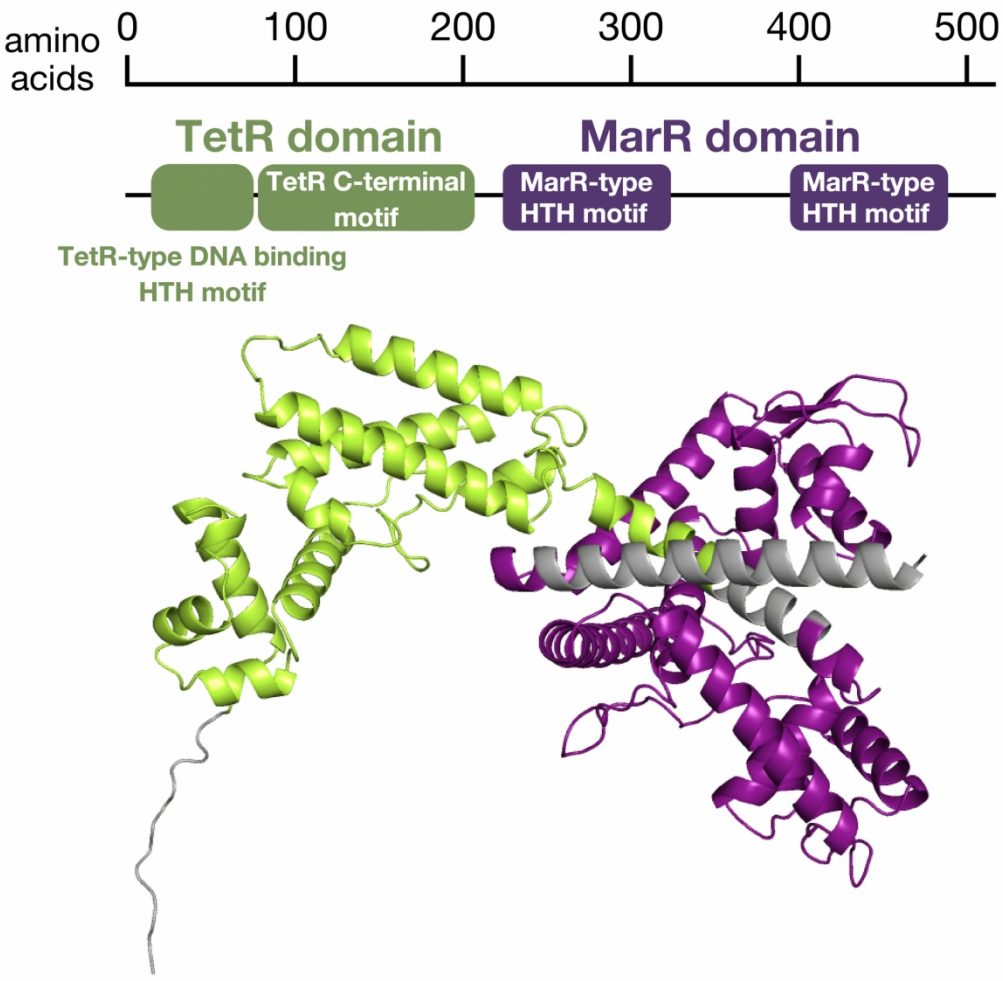

Fig. 2

79x78mm (600 x 600 DPI)

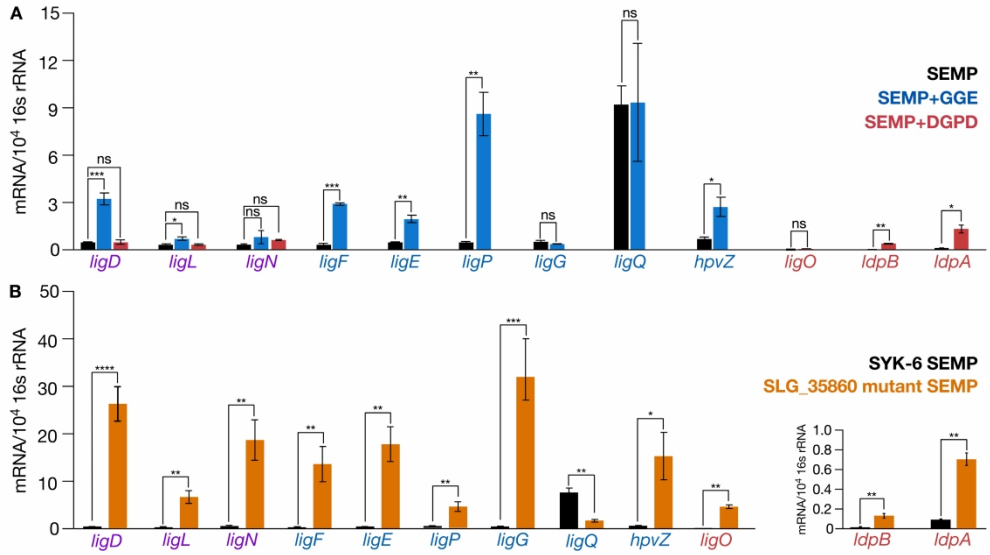

Fig. 3

177x97mm (600 x 600 DPI)

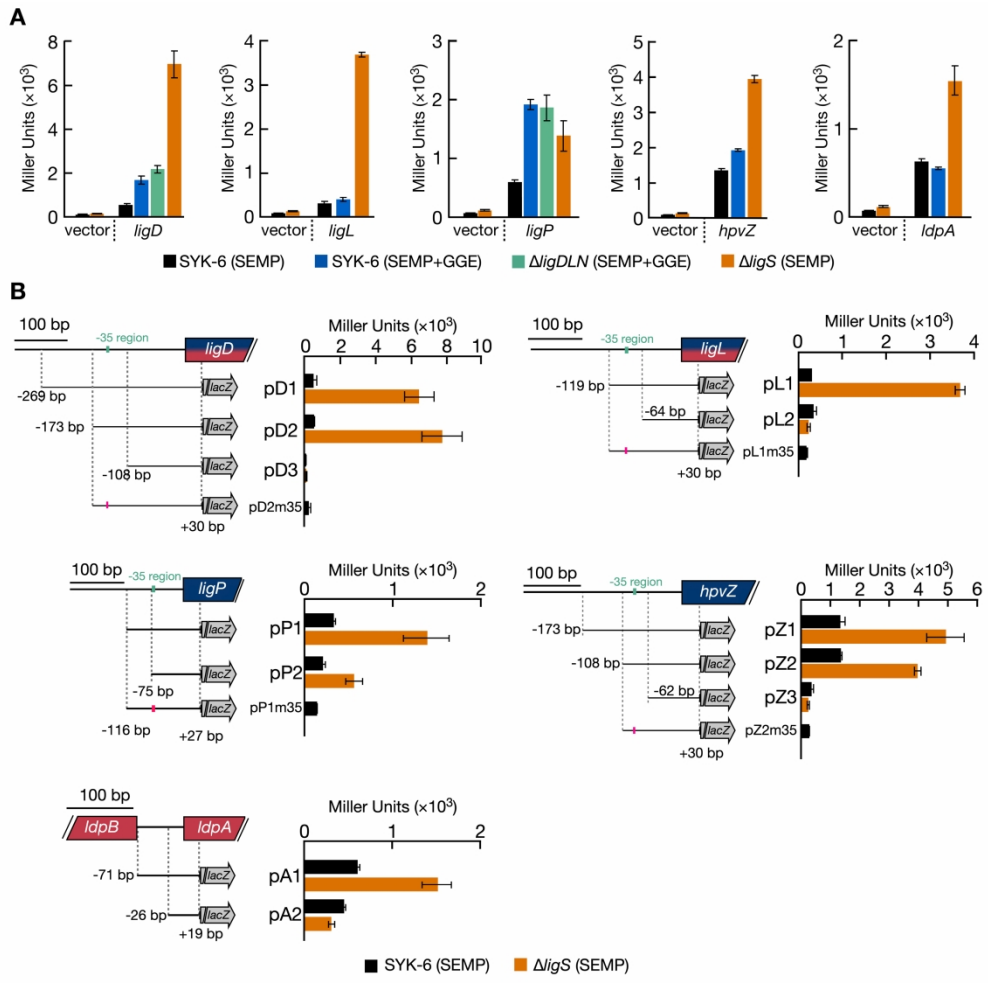

Fig. 4

150x147mm (600 x 600 DPI)

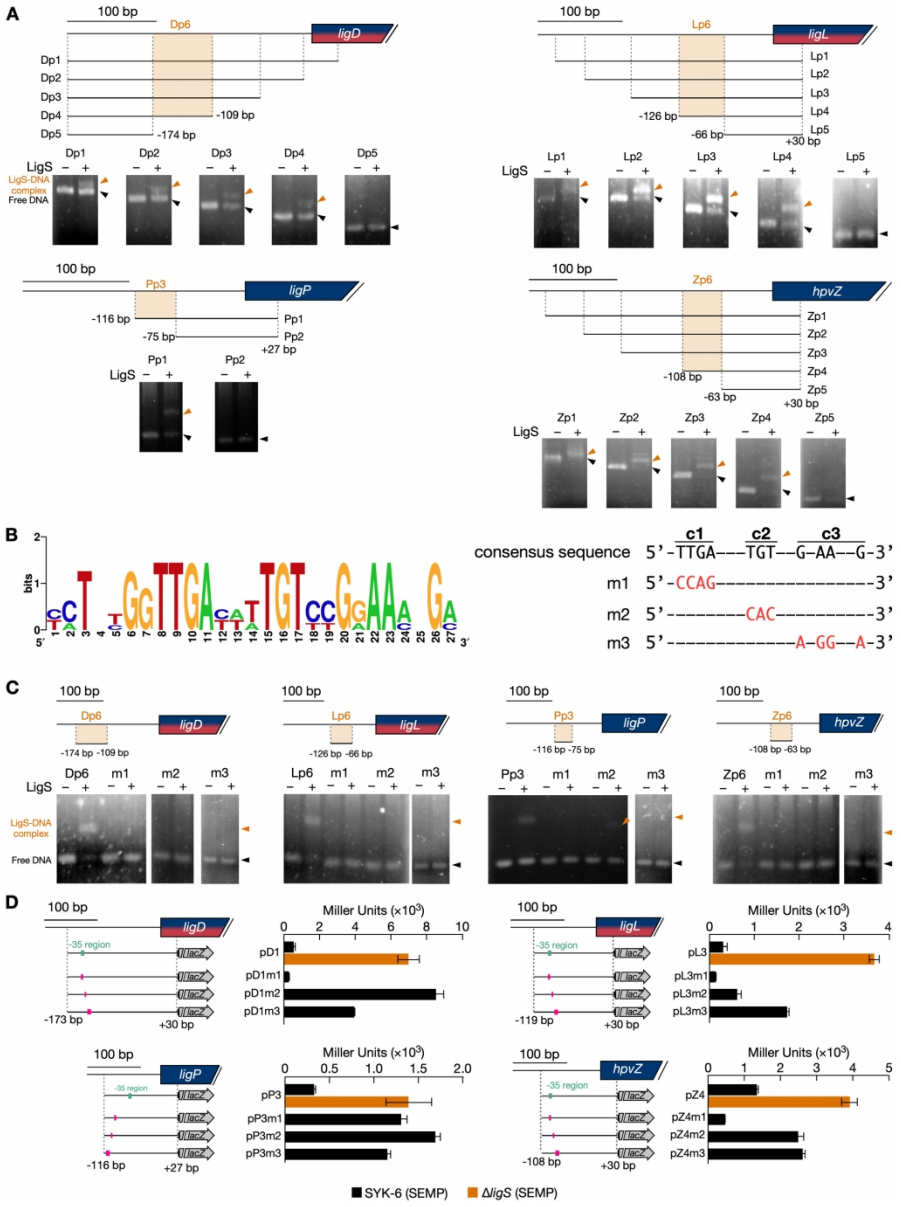

Fig. 5

170x228mm (300 x 300 DPI)

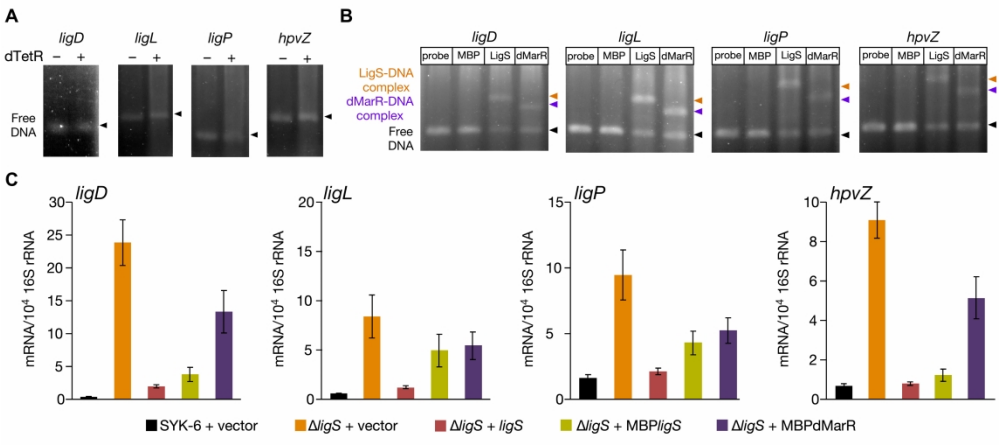

Fig. 6

170x76mm (600 x 600 DPI)

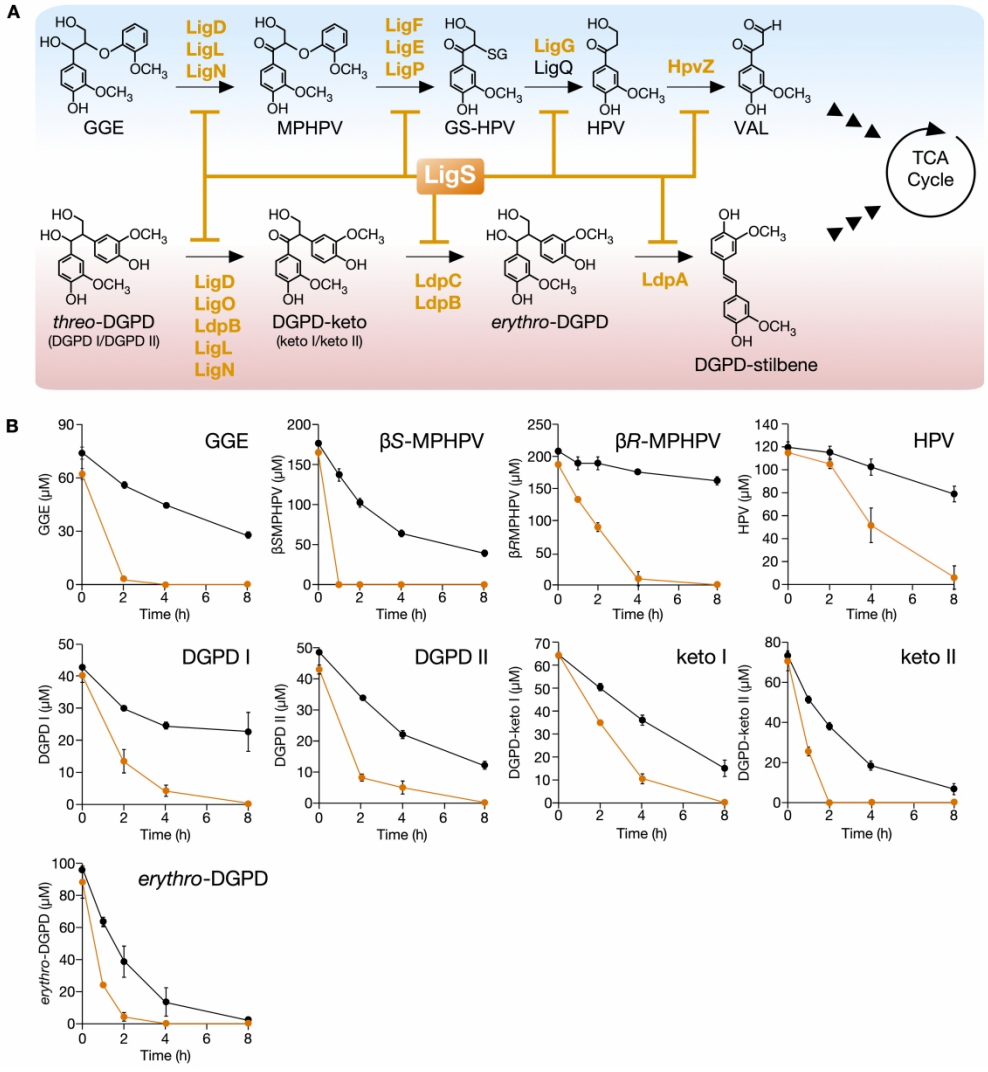

Fig. 7

170x182mm (600 x 600 DPI)

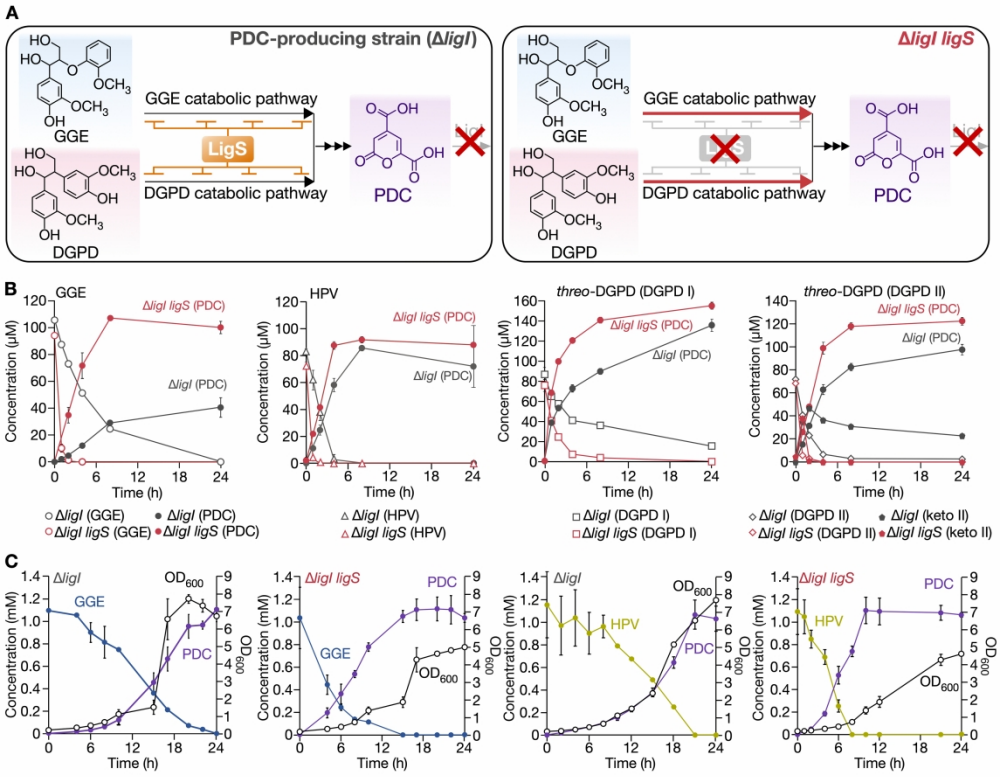

Fig. 8

170x135mm (600 x 600 DPI)

Metabolic modification of *Sphingobium lignivorans* SYK-6 for lignin valorization through the discovery of an unusual transcriptional repressor of lignin-derived dimer catabolism

Ryo Kato [a], Eugene Kuatsjah [b†], Masaya Fujita [a], Alissa C. Bleem [b], Shojiro Hishiyama [c], Rui Katahira [b], Toshiya Senda [d], Gregg T. Beckham [b], Naofumi Kamimura [a], Eiji Masai [a,*]

[a] *Department of Materials Science and Bioengineering, Nagaoka University of Technology, Nagaoka, Niigata, Japan*
[b] *Renewable Resources and Enabling Sciences Center, National Renewable Energy Laboratory, Golden, CO, USA*
[c] *Department of Forest Resource Chemistry, Forestry and Forest Products Research Institute, Tsukuba, Ibaraki, Japan*
[d] *Structural Biology Research Center, Institute of Materials Structure Science, High Energy Accelerator Research Organization (KEK), Tsukuba, Ibaraki, Japan*

Data availability statements

The data supporting this article have been included as part of the Supplementary Information.