Title

Genetic basis for the cooperative bioactivation of plant lignans by Eggerthella lenta and other human gut bacteria

Journal

Nature Microbiology, 5(1)

ISSN

2058-5276

Authors

Bess, Elizabeth N

Bisanz, Jordan E

Yarza, Fauna

et al.

Publication Date

2020

DOI

10.1038/s41564-019-0596-1

Peer reviewed

HHS Public Access

Author manuscript

Nat Microbiol. Author manuscript; available in PMC 2020 May 04.

Published in final edited form as:

Nat Microbiol. 2020 January ; 5(1): 56–66. doi:10.1038/s41564-019-0596-1.

Genetic basis for the cooperative bioactivation of plant lignans by *Eggerthella lenta* and other human gut bacteria

Elizabeth N. Bess[1,2,3], **Jordan E. Bisanz**[1], **Fauna Yarza**[1], **Annamarie Bustion**[1], **Barry E. Rich**[2], **Xingnan Li**[4], **Seiya Kitamura**[5], **Emily Waligurski**[1], **Qi Yan Ang**[1], **Diana L. Alba**[6], **Peter Spanogiannopoulos**[1], **Stephen Nayfach**[7], **Suneil K. Koliwad**[6], **Dennis W. Wolan**[5], **Adrian A. Franke**[4], **Peter J. Turnbaugh**[1,8,*]

[1]Department of Microbiology & Immunology, University of California San Francisco, 513 Parnassus Avenue, San Francisco, CA 94143, USA.

[2]Department of Chemistry, University of California, Irvine, 1102 Natural Sciences 2, Irvine, CA 92617, USA.

[3]Department of Molecular Biology and Biochemistry, University of California, Irvine, 3205 McGaugh Hall, Irvine, CA 92697, USA.

[4]University of Hawaii Cancer Center, Honolulu, HI 96813, USA.

[5]Department of Molecular Medicine, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA.

[6]Diabetes Center, University of California San Francisco, 513 Parnassus Avenue, San Francisco, CA 94143, USA.

[7]United States Department of Energy Joint Genome Institute, Walnut Creek, CA 94598, USA

[8]Chan Zuckerberg Biohub, San Francisco, CA 94158, USA.

Abstract

Plant-derived lignans, consumed daily by most individuals, are thought to protect against cancer and other diseases[1]; however, their bioactivity requires gut bacterial conversion to enterolignans[2]. Here, we dissect a four-species bacterial consortium sufficient for all five reactions in this pathway. A single enzyme (benzyl ether reductase; *ber*), was sufficient for the first two biotransformations, variable between strains of *Eggerthella lenta,* critical for enterolignan

*Correspondence to: Peter J. Turnbaugh, Associate Professor, Department of Microbiology & Immunology, University of California, San Francisco, 513 Parnassus Avenue HSW 1529, San Francisco, CA 94143, peter.turnbaugh@ucsf.edu, (415) 502-3237.

Author contributions

E.N.B. performed or supervised all of the experimental work. J.E.B. performed the bioinformatic analyses and performed a subset of experimental work. S.N. developed the metagenome database. P.S. assisted with bacterial culturing and heterologous expression. F.Y. and A.B. designed and implemented the culture-independent assays for gene prevalence. F.Y. performed the *ex vivo* incubations. E.W. generated bacterial mutants. B.E.R. performed mass spectrometry on bacterial cultures. X.L. and A.A.F. performed mass spectrometry on mouse samples. Q.A. extracted DNA from human samples collected by D.L.A. and S.K.K. S.K. and D.W.W. synthesized dmSECO. P.J.T. supervised the study.

Competing interests

P.J.T. is on the scientific advisory boards for Kaleido, Pendulum, Seres, and SNIPRbiome; there is no direct overlap between the current study and these consulting duties. All other authors have no relevant declarations.

production in gnotobiotic mice, and unique to Coriobacteriia. Transcriptional profiling (RNAseq) independently identified *ber* and genomic loci upregulated by each of the remaining substrates. Despite their low abundance in gut microbiomes and restricted phylogenetic range, all of the identified genes were detectable in the distal gut microbiomes of most individuals living in Northern California. Together, these results emphasize the importance of considering strain-level variations and bacterial co-occurrence to gain a mechanistic understanding of the bioactivation of plant secondary metabolites by the human gut microbiome.

Plants have been consumed for centuries to prevent and treat disease, yet their medicinal properties may arise from the metabolites produced by the gut microbiome[3–5]. Here, we focus on lignans, due to their ubiquity in human diets[6] and reliance upon the gut microbiome for conversion to the active metabolites enterodiol (END) and enterolactone (ENL)[2,7,8], compounds with structural and functional similarities to the endogenous hormone estrogen[1,2,9–14]. Despite the discovery of enterolignans in humans nearly 40 years ago, the enzymes responsible for their production remained unknown[15,16]. The identification of a culturable bacterial consortium sufficient for the metabolism of dietary lignans to phytoestrogens[7,17] enabled us to link human gut bacterial genetic loci to all five biotransformations involved in the metabolism of the plant lignan pinoresinol (PINO) to its end-products END and ENL. These results emphasize the utility of functional studies for elucidating the genetic basis for the gut bacterial metabolism of dietary small molecules.

The human gut microbiome metabolizes the dietary lignan PINO via four distinct chemical reactions—two sequential benzyl ether reductions, guaiacol demethylation, catechol dehydroxylation, and diol lactonization (Fig. 1a)—yielding the enterolignans END and ENL[7,18]. We tested four bacterial species for their ability to perform each reaction during anaerobic growth. Consistent with prior studies[18–20], *Eggerthella lenta* DSM2243^T catalyzed the first two reactions, *Blautia producta* DSM3507 metabolized secoisolariciresinol (SECO), and *Lactonifactor longoviformis* DSM17459^T converted END to ENL (Extended Data Fig. 1a). We were unable to confirm prior reports suggesting *E. lenta* dehydroxylates didemethylsecoisolariciresinol (dmSECO) to END[7,18], both for the type strain (DSM2243^T) and the putative dmSECO-metabolizing isolate *E. lenta* SECO-MT75m2. To better mimic the previous co-incubations[7,18], we incubated a panel of 25 Coriobacteriia strains[21] (the class to which *E. lenta* belongs) with *B. producta* DSM3507 and SECO. None of the *E. lenta* strains produced END; however, strains from the neighboring genus *Gordonibacter* (strains 3C and 28C) produced END in co-culture (Supplementary Table 1) and in isolation (Extended Data Fig. 1a). No substantial changes in growth were observed upon incubation of each bacterium with its respective substrate (Extended Data Fig. 1b).

Next, we sought to identify the genes and enzymes responsible. Multiple studies have revealed variation between *E. lenta* strains in their ability to metabolize plant secondary metabolites[22,23] and endogenous compounds[24]. We screened our 25-strain Coriobacteriia collection[21] using HPLC (Fig. 1b). Metabolism varied across strains: 9 strains (36%) did not produce any detectable metabolites (non-metabolizers). Among the 16 metabolizers, the extent of metabolism varied considerably, from 4.8±0.5% to complete conversion of dosed PINO to the reduced products laricirisinol (LAR) and SECO (Supplementary Table 1). All

of the metabolizing strains were from the *Eggerthella* genus, including 15 *E. lenta* strains and 1 *E. sinensis* strain. Metabolizing strains could not be predicted based on phylogeny (Extended Data Fig. 2), suggesting the repeated gain and/or loss of the genes responsible.

Only a single genomic locus was conserved in all metabolizers and absent in all non-metabolizers, based on the ElenMatchR tool[21] (Fig. 1b and Supplementary Table 2). This locus consisted of two genes (Extended Data Fig. 3) encoding an enzyme homologous to fumarate reductase [benzyl ether reductase (*ber*)] and a gene homologous to transcriptional regulators from the LuxR family [*ber* regulator (*berR*)]. Consistent with its putative regulatory function[25], *berR* is transcribed in reverse orientation to *ber*. We used heterologous expression to test if Ber is sufficient to catalyze the benzyl ether reduction of lignans. Incubation of *E. coli* Rosetta 2(DE3) cells engineered to express Ber led to 97.2±1.9% conversion of PINO to the products LAR and SECO (Fig. 1c). Mutation of two tyrosine residues in Ber's putative active site to phenylalanine (Y383F and Y540F) altered conversion of PINO to SECO (Fig. 1c), consistent with a related enzyme in which these residues are critical for catalysis[26]. While there have been reports of genes from sphingomonads that enable conversion of PINO to LAR and SECO, these genes share no primary sequence homology with *ber*[27].

Comparative genomics was infeasible for the remaining three bacterial species due to the lack of sufficient strains in our collection or public repositories. Multiple recent studies have demonstrated that RNAseq can help identify genes involved in the metabolism of plant secondary metabolites and endogenous compounds[22–24]. We tested if it would have been possible to identify *ber* using only RNAseq data from a single strain. Incubation of *E. lenta* DSM2243^T with PINO during exponential growth resulted in a robust but specific transcriptional response; just 4 genes in *E. lenta* DSM2243^T were differentially expressed (Fig. 1d and Supplementary Table 3). *ber* was the most highly up-regulated gene by a factor of 500 [*ber,* 2670.9-fold; second most up-regulated gene (encoding a hypothetical protein), 5.4-fold]. The significant induction of *ber* transcription in response to PINO was confirmed by RT-qPCR on an independent set of cultures (Fig. 1e) and at concentrations of PINO ranging from 8 μM to 512 μM (Fig. 1f), demonstrating that the substrate-dependent expression of *ber* is robust, reproducible, specific, and sensitive to physiological relevant concentrations (see Methods)[28,29].

Following the benzyl ether reduction of PINO that produces LAR and then SECO, lignan metabolism proceeds with *Blautia producta* DSM3507 demethylating the 3-position methoxy groups of SECO, liberating two catechols to form dmSECO (Fig. 1a)[18,19]. Incubation of *B. producta* DSM3507 with SECO resulted in a robust but specific transcriptional response: 6 genes were up-regulated and 2 genes were down-regulated (Fig. 2a and Supplementary Table 3). The most up-regulated gene (12.6-fold) is the only differentially expressed gene with sequence homology to methyltransferases, prompting us to annotate it as guaiacol lignan methyltransferase (*glm*) (Fig. 2b and Extended Data Fig. 3a). Up-regulation of *glm* was validated by RT-qPCR using an independent set of cultures (Fig. 2b). Attempts at heterologous expression have been unsuccessful, likely due to the lack of closely-related genetically tractable host strains and the requirement of multiple proteins to mediate methyl transfer[30]. As an alternative approach, we performed *ex vivo* incubations

of three human stool samples with PINO and monitored the conversion to the downstream metabolites (Fig. 2c). Consistent with the hypothesis that *glm* catalyzes the conversion of SECO to dmSECO, we found that differences in *glm* presence/absence were predictive of the major metabolic end-product of these microbial communities.

Bacterial lignan metabolism pathway proceeds with *Gordonibacter pamelaeae* 3C dehydroxylating the 4-position hydroxyl group in the catechol moiety of dmSECO, forming END (Fig. 1a). Following incubation of *B. producta* DSM3507 with SECO, cell-free supernatants containing dmSECO were filter-sterilized and introduced to *G. pamelaeae* 3C. In contrast to the incubations with purified substrates, RNAseq analysis of *G. pamelaeae* 3C exposed to spent media from *B. producta* revealed a broader transcriptional response, including 37 differentially expressed genes, all of which were up-regulated relative to lignan-free spent media controls (Fig. 2a and Supplementary Table 3). We identified a large biosynthetic gene cluster, in which 15/19 genes were significantly up-regulated (Fig. 2a and 2b). Based on our recent identification and biochemical characterization of a similar catechol dehydroxylase active against dopamine (*dadh*)[24], we reasoned that the catechol lignan dehydroxylase (*cldh*) likely requires a molybdenum cofactor[31]. Consistent with this hypothesis, we identified a gene within the dmSECO-induced gene cluster predicted to encode a molybdopterin oxidoreductase (referred to herein as *cldh*) that was up-regulated 3893.8-fold. The only gene encoding a putative enzyme more up-regulated than *cldh* is an 115-amino acid 4Fe-4S ferredoxin-type binding domain (locus tag C1877_07245) (Fig. 2a and Supplementary Table 3). This protein is directly adjacent to *cldh* (Fig. 2b) and is common to molybdopterin enzymes[31]. Up-regulation of *cldh* was validated by RT-qPCR using an independent set of cultures and a purified synthetic standard of dmSECO (Fig. 2b). We further validated the proposed function of *cldh* using comparative genomics. A systematic screen of our 25-Coriobacteriia collection revealed that only *G. pamelaeae* 3C and *Gordonibacter* sp. 28C produce END (Supplementary Table 1). ElenMatchR[21] revealed that *cldh* is conserved in both *Gordonibacter* strains but missing in the other 23 strains and *E. lenta* SECO-MT75m2[7,18]. Although 378 genes were predictive of the metabolic activity of our strain collection due to shallow sampling of *Gordonibacter* spp. (Supplementary Table 2), only 13 genes were also significantly up-regulated in response to dmSECO (Fig. 2d), of which *cldh* is the only plausible candidate gene based on its annotation (Supplementary Table 3).

Finally, *Lactonifactor longoviformis* DSM17459^T catalyzes the diol lactonization of the enterolignan END to ENL (Fig. 1a). Incubation of *L. longoviformis* with END resulted in a robust but specific transcriptional response: 6 genes were up-regulated and 2 were down-regulated relative to vehicle controls (Fig. 2a and Supplementary Table 3). This included a single genomic locus comprising three genes that all exhibited transcript levels >300-fold higher in the presence of END (Fig. 2a). The next most up-regulated gene was only 5.3-fold, highlighting the specificity of the transcriptional response to END. One of the three up-regulated genes in this locus encodes a MFS transporter that may facilitate the uptake of END to the bacterial cytosol. The other two genes both encode cytosolic enzymes: 4Fe-4S NADP-dependent oxidoreductase (locus tag BUA56_RS20795; up-regulated 359.6-fold) and NAD(P)-dependent short-chain dehydrogenase/reductase (locus tag BUA56_RS20800, referred to herein as enterodiol lactonizing enzyme (*edl*); up-regulated 470.8-fold). Both

genes were expressed in *E. coli* Rosetta 2(DE3) cells. Only cells engineered to express Edl led to 22.4±0.9% conversion of END to ENL (Fig. 2e), demonstrating that *edl* is sufficient to catalyze the final step in the lignan metabolism pathway.

Next, we sought to determine the *in vivo* relevance of the genes we had implicated in lignan metabolism, focusing on strain-level differences in the *E. lenta* population. Germ-free female BALB/c mice were colonized with either *ber*$^{+}$ *E. lenta* DSM2243^{T} or *ber*$^{-}$ *E. lenta* FAA1–3-56; a control group of mice was maintained germ-free (n=5 mice/group). Both colonized groups were also inoculated with *B. producta* DSM3507, *G. pamelaeae* 3C, and *L. longoviformis* DSM17459^{T}. Bacterial colonization was confirmed via qPCR (Fig. 3a) prior to oral gavage with PINO (10 mg/kg). Colonization level was comparable for all strains between the *ber*$^{+}$ and *ber*$^{-}$ communities, with the exception of *G. pamelaeae* 3C, which was significantly increased in the *ber*$^{+}$ gut microbiota (Fig. 3a). Following β-glucuronidase treatment, the administered lignan (PINO) and the two downstream enterolignans (END and ENL) were quantified in serum, urine, and intestinal contents via mass spectrometry (Supplementary Table 4). These data revealed a significantly higher level of the enterolignan END in the urine of *ber*$^{+}$ *E. lenta*-colonized mice relative to *ber*$^{-}$ and germ-free controls (Fig. 3b). Significant elevations in END were also detectable in the distal gut (Fig. 3c). Consistent with this observation, our *ex vivo* incubations of human stool samples (Fig. 2c) also accumulated END, perhaps due to the starting substrate used or a lack of sufficient time for the final reaction to run to completion in the context of a complex microbial community and, *in vivo*, prior to absorption or excretion. Surprisingly, ENL was detectable in GF mice (see Methods) and ENL levels were significantly increased above background levels in both the *ber*$^{+}$ and *ber*$^{-}$ colonization groups (Fig. 3d and Supplementary Table 4), suggesting a *ber*- and END-independent pathway for the production of ENL in this model system.

We also detected high serum levels of PINO (Fig. 3e), which could have further contributed to the lack of complete metabolism. We attempted to increase the residence time of PINO within the gastrointestinal tract by repeating this experiment with pinoresinol diglucoside (PINO-diGlc), which is the typical form of dietary substrate and more consistent with prior work in which lignans were dosed to rodent models using flaxseed[7,12]. Mice were colonized with *E. lenta* DSM2243^{T} (*ber*$^{+}$) or *E. lenta* FAA1–3-56 (*ber*$^{-}$), and both colonized groups were also inoculated with *B. producta* DSM3507, *G. pamelaeae* 3C, and *L. longoviformis* DSM17459^{T} in addition to *Clostridium saccharogumia* DSM17460^{T} to catalyze the hydrolysis of PINO-diGlc to free PINO. Bacterial colonization was confirmed via 16S rRNA gene sequencing of stool samples (Extended Data Fig. 4a). PINO was still detectable at high levels in serum and urine (Extended Data Fig. 4b) but did not exhibit any statistically significant differences with respect to colonization status. Both colonization groups tended to deplete PINO in the more proximal samples (ileum and cecum; Extended Data Fig. 4c), which we attribute to *C. saccharogumia* removing the glucoside moieties that otherwise impair intestinal absorption of PINO. No significant difference in the concentration of PINO in the more distal samples (colon and stool) was observed. Significantly higher levels of END were detected in the urine and serum of *ber*$^{+}$ *E. lenta*-colonized mice relative to germ-free controls (Fig. 3f) in addition to both the proximal and distal gut (Fig. 3g). ENL was significantly increased above background levels in the cecum of *ber*$^{+}$ but not *ber*$^{-}$ colonized mice (Fig. 3h).

Finally, we assessed the prevalence, abundance, and co-occurrence of the species and genes implicated in lignan metabolism across the gut microbiomes of 1870 individuals across 4 countries and 3 continents[32]. Gene abundance and host organism demonstrated a significant correlation between the first three genes in this pathway and their respective species: *ber* and *Eggerthella* spp. (ρ=0.58); *glm* and *Blautia spp.* (ρ=0.27); and *cldh* and *Gordonibacter* spp. (ρ=0.435; all three comparisons p-value $<1e^{-31}$ Spearman rank correlation) (Fig. 4a). Due to the absence of *Lactonifactor* in the taxonomic database used for this analysis, we were unable to analyze its abundance. We were also unable to detect *edl* even with a permissive homology cutoff (30% amino acid identity, 70% query coverage), consistent with prior culture-based analyses suggesting that bacteria capable of END production are found at low abundance relative to the bacteria that catalyze the earlier steps in the lignan metabolism pathway[17,20]. Despite the low abundance of these genes, we were still able to detect *ber* in the majority of the analyzed gut microbiome datasets (Fig. 4b).

There are multiple potential explanations for the partial correlation between the abundance of each of the bacterium and their respective genes, including strain-level variation within each taxa and the presence of these genes in other bacteria. To distinguish between these two alternative hypotheses, the amino acid sequences of Ber, Glm, Cldh, and Edl were queried (tBLASTn) against 23,790 genomes from the Integrated Gut Genomes database (IGGdb) comprising dereplicated species from both isolates and metagenome-assembled genomes[33]. Given the potential for even subtle differences in sequence identity to alter enzymatic activity, we used a stringent threshold to identify high-confidence homologs (>80% identity and >80% coverage; e-value$<10^{-50}$). This resulted in the identification of 20 gene-sequence homologs from 20 genomes largely representing the known host range for these genes (Supplementary Table 5). A relaxed cutoff (>50% identity and >80% coverage; e-value$<10^{-30}$) increased this number 8.8-fold to 176 (Fig. 4c and Supplementary Table 5), highlighting the diversity of the identified protein families. Multiple *glm* homologs were found across the Firmicutes phylum, consistent with reports that SECO demethylation is also catalyzed by *Eubacterium limosum*[20]. However, the restriction of high-stringency hits to the original taxonomic group of origin suggests that strain-level differences drive the observed inter-individual variations in gene abundance and may be more predictive of metabolic activity than species abundance.

A major caveat of sequencing-based approaches for determining the prevalence of gut bacterial genes is the high limit of detection, which is especially relevant here given the low abundance of the bacterial taxa involved in lignan metabolism (Fig. 4a). To better assess the prevalence of these genes in the human gut microbiome, we developed a culture-independent, PCR-based assay to determine the presence/absence of each gene from community DNA. We applied this assay to stool samples collected from 68 individuals residing in Northern California (Supplementary Table 6)[34]. In contrast to our sequencing-based analysis, all of the genes were found at high prevalence (range: 81–99%; Fig. 4d). While previous studies have suggested that lignan-activating bacteria are prevalent in the human gut[17,20], these results suggest that the same enzymes may be involved in lignan bioactivation in the majority of individuals.

These studies demonstrate the power of using natural bacterial genetic variations between strains coupled to RNAseq to identify functional links between genomic loci and metabolic activities and emphasize the importance of bacterial co-occurrence in shaping the multi-step biotransformation of dietary substrates, gut bacterial fitness, and likely downstream host phenotypes. Mechanistic insights into how these bacterial consortia coordinate their growth and metabolism will be essential to refine our ability to monitor the prevalence, abundance, and activity of these key bacterial strains in human cohorts. Our gnotobiotic mouse models provide a first step towards comparing the bioactivity of the enterolignans END and ENL and their precursors in the context of cancer and other diseases. More broadly, these results set the stage for an in-depth analysis of the biochemistry of each of the identified candidate enzymes (Extended Data Fig. 5), which together act on conserved chemical motifs that are found in numerous dietary, pharmaceutical, and endogenous small molecules.

Methods

Chemicals

Lignans (+)-pinoresinol (PINO), (+)-lariciresinol (LAR), and (−)-secoisolariciresinol (SECO) were obtained from Separation Research Ltd. Enterolignans (±)-enterodiol (END) and (±)-enterolactone (ENL), corticosterone (COR), and TRI Reagent were obtained from Sigma Aldrich. Didemethylsecoisolariciresinol (dmSECO) was synthesized by the method described previously[37] with slight modifications (see below). Chemicals were ≥ 95% pure.

General bacterial culturing conditions

The bacterial strains in our Coriobacteriia collection have previously been extensively described[26] and draft genomes are available in NCBI GenBank (see Supplementary Table 1 for accession numbers). *Blautia producta* DSM3507 and *Lactonifactor longoviformis* DSM17459^T were purchased from DSMZ and *de novo* sequenced as before[26] with draft genomes available (Supplementary Table 1). Culture media was composed of Bacto Brain Heart Infusion (BD Biosciences, 37 g/L) supplemented with *L*-cysteine-HCl (0.05%, w/v), *L*-arginine (1%, w/v), menadione (1 μg/mL), and hemin (5 μg/mL) (referred to herein as BHI++) and allowed to equilibrate in an anaerobic environment prior to use. Briefly, bacteria were cultured in BHI++ at 37 °C in an anaerobic chamber (Coy Laboratory Products) with an atmosphere composed of 2–3% H_2, 20% CO_2, and the balance N_2. Generally, to test lignan metabolism, a stock solution of the respective lignan (50 mM in methanol) was diluted (1:100) into BHI++. To harvest supernatant for HPLC analysis, cultures were centrifuged at 2000 rpm for 10 min at 4 °C, aspirating supernatant and storing at −20 °C until thawing prior to HPLC analysis. To harvest RNA, cultures were centrifuged at 2000 rpm for 10 min at 4 °C, decanting supernatant and flash-freezing the bacterial pellet in liquid nitrogen.

Screening bacterial strains for lignan metabolism

Wells of 96-well plates containing 99 μL of BHI++ media supplemented with PINO (510 μM) were inoculated (1:100), in triplicate, from dense cultures of each of the 25 strains of our Coriobacteriia collection. Plates were incubated at 37 °C for 48 hours. Next, the plates were centrifuged at 2000 rpm for 10 min at 4 °C, and the supernatant was harvested and

stored at −20 °C until analysis via HPLC, using Method A (see “HPLC methods” below). Results are summarized in Supplementary Table 1.

Wells of 96-well plates containing 98 μL of BHI++ media supplemented with SECO (505 μM) were each inoculated (1:100), in triplicate, from a dense culture of *B. producta* DSM3507 and from a dense culture of each of the 25 strains of our Coriobacteriia. Plates were incubated at 37 °C for 40 hours. Next, the plates were centrifuged at 2000 rpm for 10 min at 4 °C, and the supernatant was harvested and stored at −20 °C until analysis via HPLC, using Method A (see “HPLC methods” below). Results are summarized in Supplementary Table 1.

Culturing bacteria with lignans for time-course, RNAseq, and RT-qPCR experiments

Hungate tubes containing 5 mL of BHI++ were inoculated (1:100), in pairs, from dense cultures of *E. lenta* DSM2243^T, *B. producta* DSM3507, *G. pamelaeae* 3C, or *L. longoviformis* DSM17459^T. Culture tubes were incubated at 37 °C until mid-log-phase growth was achieved, at which time pairs of cultures were exposed to either 50 μL of the respective lignan (50 mM in methanol) or 50 μL of vehicle (methanol). Then, 1.5–2 hours following lignan (n=3) or vehicle (n=3) exposure, cultures were removed from the anaerobic chamber and centrifuged at 2000 rpm for 10 min at 4 °C. The supernatant was decanted, and the remaining cell pellet was either flash-frozen in liquid nitrogen and subsequently stored in a −80 °C freezer until RNA was extracted for RT-qPCR analysis, or 1 mL of TRI Reagent was added to the pellet for immediate extraction of RNA for RNAseq. To determine the dose-response of *ber* expression to PINO, experiments were performed as above with 2-fold dilution series of PINO in methanol (51.2, 25.6, 12.8, 6.4, 3.2, 1.6, 0.9, 0 mM).

Additional cultures differentially exposed to lignan or vehicle were incubated at 37 °C in the anaerobic chamber. Throughout the course of incubation, OD_{600} was periodically measured using a Hach spectrophotometer (model DR1900) that was housed in an anaerobic chamber. Cultures were periodically harvested. To assess SECO consumption, cultures were prepared in a 96-well plate (100 μL of BHI++) and exposed to SECO (500 μM). All cultures were periodically harvested. Supernatant was collected and stored at −20 °C until analysis via HPLC (see “HPLC methods” below).

During the course of the work reported herein, we obtained a pure sample of dmSECO (synthetic route reported in the Supplementary Methods). Prior to the availability of pure dmSECO, this compound was incubated with *G. pamelaeae* 3C by providing the spent, filter-sterilized supernatant afforded from the incubation of *B. producta* DSM3507 with SECO, which contained the metabolic byproduct of SECO, presumably didemethylsecoisolariciresinol (dmSECO).

RNA sequencing and analysis

To a bacterial pellet, 1 mL of TRI reagent was added. The sample was vortexed to resuspend the cells and incubated at room temperature for 10 minutes. The suspension was then transferred to a 2 mL screw-cap tube containing glass beads and lysing matrix (MP Biomedicals, catalog #: 116914050). Cells were lysed and the suspension homogenized for 5 minutes in a bead-beater (BioSpec Products, Mini-Beadbeater-96, catalog #: 1001).

Chloroform (200 μL) was added to the sample, which was then vortexed for 15 seconds before incubating at room temperature for 10 minutes. Next, the mixture was centrifuged at 16,000 × *g* for 15 minutes at 4 °C.

The upper aqueous phase (500 μL) was transferred to a new RNase-free microcentrifuge tube, and 500 μL of 100% ethanol was added, followed by vortexing to mix well. This mixture was transferred to a spin column (PureLink RNA Mini Kit, Life Technologies catalog #: 12183025) and centrifuged at ≥12,000 × *g* for 30 seconds, discarding flow-through, until all of the material had been added to the column. To the spin column, 350 μL of Wash Buffer I (PureLink RNA Mini Kit, Life Technologies, catalog #: 12183025) was added, and then the column was centrifuged at ≥12,000 × *g* for 30 seconds, discarding flow-through.

To the column, 80 μL of DNase mix (PureLink DNase Set, Life Technologies catalog #: 12185010; 8 μL 10× reaction buffer, 10 μL DNase, 62 μL RNase-free water) was added, and the column was incubated at room temperature for 15 minutes. Next, 350 μL of Wash Buffer I (PureLink RNA Mini Kit, Life Technologies, catalog #: 12183025) was added, and then the column was centrifuged at ≥12,000 × *g* for 30 seconds. The column was transferred to a new collection tube, and 500 μL of Wash Buffer II was added, followed by centrifugation at ≥12,000 × *g* for 30 seconds, discarding flow-through. The column was centrifuged at ≥12,000 × *g* for 60 seconds, drying the column, and then moved to a collection tube. Then, 50 μL of RNase-free water was added, and the column was incubated at room temperature for 1 minute. Finally, the column was centrifuged for 1 minute at ≥12,000 × *g*, retaining the flow-through, which contains total RNA.

A second, solution-phase DNase treatment was performed using TURBO-DNase (Ambion, ThermoFisher, catalog #: AM2238), adding 5 μL TURBO-DNase buffer and 1 μL TURBO-DNase to the 50 μL solution of total RNA. This solution was incubated at 37 °C for 30 minutes, after which 56 μL of Lysis Buffer (PureLink RNA Mini Kit, Life Technologies, catalog #: 12183025) and 56 μL of 100% ethanol were added. The solution was vortexed, and then the sample was transferred to a spin cartridge with collection tube and centrifuged at ≥12,000 × *g* for 30 seconds, discarding the flow-through. Wash Buffer I (350 μL; PureLink RNA Mini Kit, Life Technologies, catalog #: 12183025) was added, followed by centrifugation of the spin column for 30 seconds at ≥12,000 × *g*. The spin column was moved to a new collection tube, and 500 μL of Wash Buffer II (PureLink RNA Mini Kit, Life Technologies, catalog #: 12183025) was added. The sample was centrifuged (≥12,000 × *g* for 30 seconds), discarding flow-through. The column was centrifuged at ≥12,000 × *g* for 1 minute, drying the column, and then moved to a collection tube. Then, 30 μL of RNase-free water was added, and the column was incubated at room temperature for 1 minute. Finally, the column was centrifuged for 1 minute at ≥12,000 × *g*, retaining the flow-through, which contains purified total RNA.

Total RNA was subjected to rRNA depletion using Ribo-Zero Bacterial rRNA Depletion (Illumina, catalog #: MRZB12424), following the manufacturer's protocol. RNA fragmentation, cDNA synthesis, and library preparation proceeded using NEBNext Ultra RNA Library Prep Kit for Illumina (New England BioLabs, catalog #: E7530) and NEBNext

Multiplex Oligos for Illumina, Dual Index Primers (New England BioLabs, catalog #: E7600), following the manufacturer's protocol. All samples, except *Gordonibacter pamelaeae* 3C samples, were single-end sequenced (1×50 bp) using an Illumina HiSeq2500 platform (High Output, v4 chemistry). *Gordonibacter pamelaeae* 3C samples were single-end sequenced (1×100 bp) using an Illumina HiSeq2500 platform (Rapid Run, v2 chemistry). For each sample, fastq files are available in NCBI's Sequence Read Archive (SRA), BioProject accession number PRJNA450120. Briefly, reads were mapped to reference genomes using Bowtie2[35]. HTSeq was used to count the number of transcripts mapping to genes[36]. Finally, differential gene expression was assessed using DESeq[37].

Comparative genomics

Comparative genomics was performed using ElenMatchR, the development of which we have reported[21]. This tool can be found at https://jbisanz.shinyapps.io/elenmatchr/. For the discovery of *ber*, genes were clustered at a minimum of 60% amino acid identity and 80% query coverage. For the identification of *cldh* candidates, a 30% amino acid identity/50% query coverage cutoff was used to allow for better identification of orthologs across genus boundaries. For comparison of comparative genomic data and transcriptional data for *G. pamelaeae* 3C, alternate gene calls were to generate transcript counts to maintain compatibility with ElenMatchR (available as compressed serialized R object (RDS) for download at github.com/jbisanz/ElenMatchR/raw/master/resources/RDS/gtfs/Gordonibacter_pamelaeae_3C.gff.RDS) resulting in a loss of 2 differentially expressed hypothetical genes.

RT-qPCR analysis

From total RNA, cDNA was synthesized using iScript Reverse Transcriptase (BioRad, catalog #: 1708841). RT-qPCR assays were performed using SYBR Select Master Mix for CFX (Applied Biosystems) using Thermocycler C1000 CFX384 Real-Time System (BioRad). The primers used for amplification are listed in Supplementary Table 7. Fold changes in transcript levels were calculated as lignan exposure relative to vehicle exposure using the $\Delta\Delta C_T$ method.

Cloning ber

The *ber* sequence was PCR-amplified (forward primer: 5' ATGCATGGTACCGC AACCGCATCCGGCG 3'; reverse primer: 5' ATGCATGGATCCTTACGCTTTGTCGCCCAG CA 3') from *E. lenta* DSM2243^T, introducing KpnI and BamHI restriction sites with the forward and reverse primers, respectively. The PCR product was cloned into a pET19bTEV plasmid, affording a plasmid designated as pET19bTEV-*ber*. This plasmid was introduced into *E. coli* Rosetta 2(DE3) cells.

Mutagenizing ber

The *ber* gene was mutated using the QuikChange Lightning Site-Directed Mutagensis Kit (Agilent Technologies) following the manufacturer's protocol. The primers used to introduce point mutations are presented in Supplementary Table 7. This procedure afforded the

plasmids designated as pET19bTEV-*ber* A1148T (produces Ber Y383F) and pET19bTEV-*ber* A1619T (produces Ber Y540F). To generate the double mutant, the pET19bTEV-*ber* A1148T was mutagenized to introduce the A1619T mutation, affording pET19bTEV-*ber* A1148T A1619T.

Heterologous expression of Ber

E. coli Rosetta 2(DE3) cells bearing pET19bTEV with or without *ber* WT, A1148T (Y383F), A1619T (Y540F), A1148T and A1619T were aerobically cultured at 37 °C for 1 day in lysogeny broth (LB) that was supplemented with ampicillin (50 μg/mL) and chloramphenicol (15 μg/mL). The resulting cultures were used to inoculate (1:100) LB that was supplemented with 250 μM PINO (from a stock solution of 50 mM in MeOH). Cultures were aerobically incubated for 48 hours. Then, the bacterial cells were pelleted, and the supernatant was harvested and stored at −20 °C until analysis via HPLC, using Method B (see “HPLC methods” below).

HPLC methods

HPLC measurements of analyte concentrations were performed using an Agilent HPLC (1220 Infinity) equipped with a pump, degasser, autosampler, column oven, and diode array detector. Data was acquired using OpenLAB CDS (Agilent Technologies).

Method A: Solvent A, 0.1% formic $acid_{(aq)}$; Solvent B, 100% methanol; 0–7.5 min, 30% −48% B; 7.5–12 min, 48%−90% B; 12.5 min, 90% B; 12.5–12.6 min, 90%−30% B; 12.6–14 min, 30% B; Flow rate: 1 mL/min; Temperature: 37 °C; Column: C18 column (Kinetex 2.6 μM 100Å, 15 cm × 0.46 cm; Phenomenex: 00F-4462-E0); Guard column: SecurityGuard ULTRA cartridge (Phenomenex part #: AJ0–8768); Injection volume: 30 μL. At 275 nm, PINO (retention time (rt)=9.35 min), LAR (rt=7.49 min), SECO (rt=7.73 min), END (rt=9.97 min), and ENL (10.65 nm) were measured. At 240 nm, COR (rt=12.24 min) was measured.

Method B: Solvent A, 0.1% formic $acid_{(aq)}$; Solvent B, 100% acetonitrile; 0–1 min, 30% B; 1–5 min, 30%−80% B; 5–6 min, 80%−30% B; 6–10 min, 30% B; Flow rate: 1 mL/min; Temperature: 37 °C; Column: C18 column (Kinetex 2.6 μM 100Å, 15 cm × 0.46 cm; Phenomenex: 00F-4462-E0); Guard column: SecurityGuard ULTRA cartridge (Phenomenex part #: AJ0–8768); Injection volume: 30 μL. At 285 nm, PINO (rt=4.28 min), LAR (rt=3.17 min), SECO (rt=2.76 min), END (rt=3.53 min), and ENL (rt=4.81 min) were measured. At 240 nm, COR (rt=4.91 min) was measured.

Stock solutions of PINO, LAR, SECO, END, and ENL were prepared at 50 mM in DMSO. These solutions were used to generate a standard containing PINO, LAR, SECO, END, and ENL (HPLC Method A, each 750 μM in BHI++; HPLC Method B, each 1000 μM in BHI++), which was then serially diluted [2-fold dilutions to 5.9 μM (Method A) and 7.8 μM (Method B)]. These standards were each diluted 2-fold in a solution of the internal standard corticosterone (COR, 200 μM in water). The resulting samples were injected into an HPLC using conditions given above. Calibration curves were generated by measuring peak areas and performing linear regression against known concentrations.

Supernatant samples from the incubation of bacteria with lignan were prepared by diluting 2-fold in a solution of internal standard corticosterone (200 μM in water). The resulting samples were injected into an HPLC using conditions given below. Retention times and UV absorption traces of analytes were compared to lignan standards to identify lignans. Lignan concentrations were calculated from calibration curves.

Synthesis of edl construct

The genes *edl* (BUA56_RS20800) and BUA56_RS20795, codon-optimized for *E. coli*, were each commercially prepared and independently introduced into pET151/D-TOPO plasmids (Invitrogen GeneArt Synthesis, Thermo Fisher Scientific) behind a T7 promotor and with an N-terminal His-tag and V5 epitope followed by a TEV recognition site. A negative control vector (absent *edl*) was prepared by excising *edl* from the commercially synthesized pET151/D-TOPO-*edl* plasmid. The V5 epitope and TEV restriction site of this plasmid were PCR-amplified (forward primer: 5' ATGCGCCATGGTAAGCCTATCCCTAAC3'; reverse primer: 5'GAGCTCATCGTGAATTCGTGAAGGGATCAATTCCCTG3'), introducing NcoI and SacI restriction sites with the forward and reverse primers, respectively. The resulting amplicon and pET151/D-TOPO-*edl* were each digested with NcoI and SacI restriction enzymes (Thermo Scientific). Plasmid digestion excised the V5 epitope, TEV recognition site, and *edl.* Once the digested amplicon and plasmid fragments were ligated with T4 DNA ligase (New England BioLabs), a His-tag, V5 epitope, and TEV restriction site (but no *edl*) were maintained between the T7 promoter and T7 terminator. This plasmid (termed the empty vector) was introduced into *E. coli* Rosetta2(DE3) cells.

Heterologous expression of Edl

E. coli Rosetta 2(DE3) cells bearing pET151/D-TOPO with *edl* or BUA56_RS20795 or without (empty vector) and sterile controls were aerobically cultured at 37 °C overnight in lysogeny broth (LB). The resulting cultures were used to inoculate (1:100) LB that was supplemented with 10 μM END (from a stock solution of 5 mM in DMSO) and 25 μM IPTG. Cultures were aerobically incubated for 48 hours at 37 ^{o}C. Then, the bacterial cells were pelleted, and the supernatant was harvested and immediately prepared for analysis by LC-MS/MS, as described below. Prepared samples were then either analyzed by LC-MS/MS or frozen at −20 °C, thawed, and submitted for analysis within 6 hours of preparation.

LC-MS/MS analysis of samples from Edl heterologous expression

Concentrations of END and ENL from heterologous expression experiments were assessed by electrospray ionization (ESI) triple-quadrupole liquid chromatography mass spectrometry (Acquity UPLC and Quattro Premier XE, Waters Micromass, Milford, MA) in negative-ion mode with single reaction monitoring. Solvent A, 5 mM ammonium formate$_{(aq)}$; Solvent B, 5% 5 mM ammonium formate$_{(aq)}$:95% acetonitrile; 0–4 min, 40%−95% B. Temperature: 50 ^{o}C. Column: C18 Column (Acquity UPLC BEH 1.7 μM, 2.1 mm × 50 mm, Waters part #: 186002350). Injection volume: 10 μL. Retention times: END, 0.91 min; ENL: 0.67 min; HMR: 0.62 min. Capillary voltage: 3.30 kV. Cone voltage: 10 V. Source temperature: 125 ^{o}C. Desolvation gas flow: 800 L/hr. Desolvation temperature: 400 ^{o}C. $[M-H]^-$ m/z of parent/daughter ions: END, 301.08/252.90; ENL, 297.02/252.91; HMR, 372.99/172.66.

Stock solutions of END and ENL were prepared at 5 mM in DMSO. These were used to generate a standard containing END and ENL, each 20 μM in LB, which was then serially diluted three-fold to 3 nM. Each standard (125 μL) was diluted two-fold into hydroxymatairesinol (HMR, 2 μM in ddH_2O). Standards were extracted with 0.5 mL of ethyl acetate. The organic extracts were dried by centrifugal vacuum concentration (Refrigerated CentriVap Concentrator, Labconco part#: 7310020 and DryFast Ultra, Welch part#: 20422B-01) and resuspended in 250 μL 25% acetone in water. The calibration curves used for quantification were 1.5 nM to 10 μM for END and 4.5 nM to 10 μM for ENL with a limit of quantification set at a signal-to-noise ratio of 10:1. Calibration curves were generated by measuring peak areas and performing linear regression against known concentrations

Supernatant (50 μL) from incubations with END was diluted with 75 μL of LB and 125 μL of HMR (2 μM in ddH_2O). Samples were extracted with 0.5 mL of ethyl acetate. The organic extracts were dried by centrifugal vacuum concentration, resuspended in 250 μL 25% acetone in water, and injected into the LC-MS/MS instrument. END and ENL concentrations were quantitated using calibration curves.

Projecting in vivo lignan concentrations

Human data on lignan concentrations are not available; however, pigs are reasonable models of the human GI tract[38]. In pigs fed a lignan-rich diet, ~40% was excreted as enterolignans[39]. This suggests that at least 40% of an ingested dose of lignans comes in contact with gut bacteria. Using a reported model for predicting gut concentrations of orally ingested drugs[28] and based on measurement of lignan intake in a Western diet[29], we conservatively project that gut concentrations between 4 μM and 93 μM PINO are achieved in humans.

Gnotobiotic mouse experiments

Protocols for all experiments involving mice were approved by the University of California, San Francisco Institutional Animal Care and Use Committee (IACUC) and performed accordingly. Germ-free female adult BALB/C mice, 6–8 weeks of age, were housed in sterile isolators (CBC flexible, softwall isolator) under a strict 12-hour light cycle. Mice were fed *ad libitum* Lab Diet 5021. Sample size was chosen based on the results of *in vitro* experiments such that the study would be powered (95%) to detect significant (α=0.05) differences between experimental groups. Randomization was performed, to the best of our ability given the constraints posed by gnotobiotic isolators, based on age and cage of origin. Blinding was implemented for sample analysis.

Mice (n=15) were divided into three groups (n=5 per group), and each group was housed in one of three germ-free isolators. One group was maintained germ-free; the other two groups were colonized via oral gavage with a suspension of 5.8×10^6 CFUs of *B. producta* DSM3507, 1.0×10^7 CFUs of *G. pamelaeae* 3C, 6.3×10^6 CFUs of *L. longoviformis* DSM17459^T and either 1.1×10^7 CFUs of *E. lenta* DSM2243^T (*ber*$^+$ group) or 5.7×10^6 CFUs of *E. lenta* 1–3-56 (*ber*$^-$ group) in 200 μL of BHI++. Seven days following bacterial colonization, one mouse from each colonization group was orally gavaged (~100 μL) with

10 mg/kg of PINO (2 mg PINO per mL of 1:2 DMSO:H_2O). Maintaining mice in the gnotobiotic isolators, mice (one per isolator) were placed in metabolic cages (MMC100 Metabolic Cage, Hatteras Instruments) for 18 hours to collect total urine and feces, after which time mice were returned to their standard cages. For each of the next four days this procedure was repeated for the remaining mice, dosing mice at the same time each day, until all mice had been dosed. Once collected, fecal pellets were stored in a −80 °C freezer, and urine was stored in a −20 °C freezer; both sample types were kept frozen until analysis by mass spectrometry.

For *E. lenta* DSM2243^T, *E. lenta* 1–3-56, *B. producta* DSM3507, *G. pamelaeae* 3C, and *L. longoviformis* DSM17459^T, gDNA was extracted from pelleted mono-cultures of each bacterium using UltraClean Microbial DNA Isolation Kit (MO BIO Laboratories, catalog # 12224). To quantify the number of each of these bacteria colonizing each mouse, a real-time PCR detection system (Bio Rad CFX384 C1000), SYBR Select for CFX (Applied Biosystems, catalog #: 4472942) master mix, and the primers listed in Supplemental Table 7 (each of which amplify portions of single-copy genes in respective bacteria) were used. For each bacterium, a standard curve was generated (3×10^6–3×10^2 genome copies, prepared by 10-fold serial dilutions) by measuring Cq values and performing linear regression against standard solutions with known numbers of genome copies.

The morning before each mouse received its initial dose of PINO, fecal pellets were collected to assess bacterial colonization status. DNA was extracted from 49 mg to 105 mg of fecal pellet using DNeasy PowerLyzer Powersoil Kit (Qiagen, catalog #: 12855) following the manufacturer's protocol, except using MP Biomedicals Lysing Matrix E tubes (MP Biomedicals, catalog #: 116914050) for homogenization and incubating samples at 65 °C for 10 min after the addition of C1 and prior to homogenization. DNA was eluted in 100 μL of Solution C6.

DNA extracted from fecal pellets was diluted a further 50-fold in H_2O prior to amplification of each extract with each of the primers listed in Supplementary Table 7 for each of the five bacteria with which mice were colonized. Genome equivalents in mouse fecal pellets were quantified by using the standard curve to equate measured Cq values with known genome equivalents. Fig. 3a presents the number of genome equivalents for each bacterium per gram of feces, as measured by qPCR. All no-template controls and extracts from pellets in the germ-free group were below the limit of quantitation (3×10^2 genome copies).

The gnotobiotic experiment with PINO-diglucoside was performed as above with the following modifications: 20 mg/kg PINO-diglucoside was administered, mice were placed in metabolic cages for 20 hours, and *Clostridium saccharogumia* DSM17460^T was administered (1.3×10^6 CFUs) to both synthetic communities (1.5×10^6 CFUs *G. pamelaeae* 3C, 4.3×10^6 CFUs *B. producta* DSM3507, 1.8×10^6 CFUs *L. longoviformis* DSM17459^T, and 4.8×10^6 CFUs *E. lenta* 1–3-56 or 2.0×10^6 CFUs *E. lenta* DSM2243^T), as determined by drop-plate method. The relative proportions of taxa in these communities were assessed via 16S rRNA gene sequencing of the V4 variable region according to the methods of Gohl *et al.*[40] and processed using the DADA2 package[41].

LC-MS/MS analysis of mouse samples

Amounts of END, ENL, and PINO in mouse urine and stool samples were assessed by electrospray ionization (ESI) orbitrap liquid chromatography mass spectrometry (model Q-Exactive, Thermo Scientific Inc., Waltham, MA) in negative-ion mode with targeted single-ion monitoring (tSIM) as modified from our previous tandem mass spectrometry based method[42,43]. Briefly, 0.15 mL of urine or diluted urine (in 0.2 M trimethylamine buffer, pH 6) or lyophilized fecal pellets suspended (0.1–0.4 mg stool/μL) in trimethylamine buffer (0.2 M, pH 6) were mixed with enterolactone-$^{13}C_3$ (internal standard, 0.6 μg/mL in methanol) and incubated at 37 °C for 2 hours with 0.015 mL β-glucuronidase (#03 707 598 001; Roche Diagnostics, Indianapolis, IN) and 0.075 mL arylsulfatase (type H-1 #9626; Sigma, St. Louis, MO; 3 mg/mL in pH 6 triethylamine buffer) and then extracted with 1 mL methyl *tert*-butyl ether. The organic phase was dried under nitrogen and redissolved in 0.1 mL of a 1:1 mixture of MeOH:0.1% aq. formic acid. Of this solution, 0.025 mL was separated on a Kinetex C18 column (150 × 3.0 mm; 2.6 μm; Phenomenex) with a Kinetex C18 pre-column (Thermo) using a linear gradient at 0.3 mL/min of MeOH:MeCN:water: 15:15:70 to 25:25:50 in 6 min, then to 45:45:10 in 6 min, holding for 2 min, followed by equilibration at the initial ratio before subsequent injections. Orbitrap mass spectrometry was performed by ESI in negative-ion mode with the following settings: spray voltage, 3.8 kV; capillary transfer temperature, 280 ºC; sheath gas flow rate, 35; auxiliary gas, 5; in-source CID, 5 eV; AGC target, 5e5; maximum injection time, 100 ms; resolution, 35,000; microscan, 1. The divert valve was opened to the MS from 6.18 to 11.0 min to avoid excess interference. All analytes were measured at the monoisotopic masses of their deprotonated forms±5ppm and quantitated by external calibration ($[M-H]^-$ m/z: END, 301.14398; ENL 297.11268; PINO 357.13381; ENL-$^{13}C_3$ 300.12275).

Enterolignans (END and ENL) were unexpectedly observed in germ-free and *ber*$^-$ mice dosed with PINO (Supplementary Table 4). Analysis of mass spectrometry quality control samples ruled out instrument and reagent contamination. Another potential source of contamination was mouse chow or environment. To test this, we analyzed cecum, urine, stool, and serum from germ-free mice never dosed with PINO. Enterolignan levels in germ-free mice with and without dosed PINO are similar. This suggests dietary or environmental contamination of enterolignans is responsible for the enterolignan levels observed in germ-free mice.

Enterolignan levels in *ber*$^-$ mice are higher than levels observed in germ-free animals not dosed with PINO. We attribute this to dietary lignin. Crudely, lignin is a biopolymer from which lignan can be extracted. Among the most common sources of dietary lignin is grains, which are a component of the mouse chow used in this study. A reported study indicated that ENL can be produced from lignin. In rats dosed with isotope-labelled lignin, isotope-labelled ENL was produced[44]. This indicates that, in addition to lignan, lignin is also a dietary source of ENL. We suspect that while our *ber*$^-$ bacterial community could not convert PINO to downstream products, this bacterial consortium may be capable of converting dietary lignin to ENL.

Ex vivo incubations of human stool samples with PINO

Stool samples were obtained from three healthy individuals. All subjects consented to participate in the study, which was approved by the University of California, San Francisco (UCSF) Institutional Review Board. Stool samples were introduced into an anaerobic chamber and each was suspended in BHI++ (1 g stool/9 mL BHI++). Suspensions were vortexed and allowed to settle for 5 minutes. Next, supernatant was diluted 10-fold in fresh BHI++ followed by another 15-fold dilution with 1.5 mL of BHI++ supplemented with PINO (500 μM). Cultures were anaerobically incubated at 37 °C. Approximately every 24 hours, 50 μL of culture was harvested and centrifuged. Supernatant was aspirated and stored at −20 ºC until samples were prepared for analysis by HPLC, as described above.

PCR assays for detection of *ber*, *glm*, and *cldh* in DNA extracts from stool samples were performed using the forward and reverse primers listed in Supplementary Table 7. Reactions were performed using either a C1000 Touch™ Thermal Cycler or a S1000™ Thermal Cycler (BioRad) and AmpliTaq Gold 360 Master Mix (Applied Biosystems).

Culture-independent analysis of gene prevalence and abundance in the human gut microbiome

Human stool samples were collected as part of a multi-ethnic clinical cohort study, termed Inflammation, Diabetes, Ethnicity, and Obesity (IDEO; ClinicalTrials.gov identifier), consisting of 25–65-year-old men and women residing in Northern California and recruited from medical and surgical clinics at UCSF and the Zuckerberg San Francisco General Hospital, or through local public advertisements. The host phenotypic data from this cohort has been described in detail[34]. Informed consent was provided for all subjects participating in the study, which was approved by the UCSF Institutional Review Board. To extract DNA, stool samples were homogenized with bead beating for 5 min (Mini-Beadbeater-96, BioSpec, catalog #: 1001) using beads of mixed size and material (Lysing Matrix E 2 mL Tube, MP Biomedicals, catalog #: 116914500) using the digestion solution and lysis buffer of a Wizard SV 96 Genome DNA kit (Promega, catalog #: 2371). The samples were then centrifuged for 10 min at 16,000 × *g*, and then the supernatant was transferred to the binding plate. The DNA was then purified according to the manufacturer's instructions. DNA was further diluted 10-fold in water.

Genomic DNA was extracted from mono-cultured lignan-metabolizing bacteria: *Eggerthella lenta* DSM2243^T, *Blautia producta* DSM3507, *Gordonibacter pamelaeae* 3C, and *Lactonifactor longoviformis* DSM17459^T, each cultured in BHI++ in an anaerobic chamber, as described above. Dense cultures were centrifuged to pellet bacterial cells. Subsequently, cells were lysed and genomic DNA was extracted using UltraClean Microbial DNA Isolation Kit (MO BIO Laboratories, catalog # 12224), according to the manufacturer's instructions.

PCR assay for detection of genomic loci associated with lignan metabolism: DNA extractions from IDEO stool samples were diluted 20-fold in a mixture of nuclease-free water, AmpliTaq Gold 360 Master Mix (Applied Biosystems), and forward and reverse primers (0.5 μM). Forward and reverse primers used for amplification are listed in

Supplementary Table 7. Reactions were performed using either a C1000 Touch™ Thermal Cycler or a S1000™ Thermal Cycler (BioRad). Reactions were amplified over 30 cycles (40 cycles for *edl*) with the following parameters: denaturation at 95 °C for 30 sec., annealing at 59.7 °C for 30 sec., and extension at 72 °C for one min. with a final extension at 72 °C for 7 min. Genomic DNA extracted from mono-cultures of each bacterium served as positive controls and nuclease-free water served as a negative control. PCR amplification products were subjected to gel electrophoresis with agarose gels [1.5% agarose with SYBR Safe DNA gel stain (Invitrogen)]. Gels were visualized using a ChemiDoc™ Imaging System (BioRad). In the PCRs in which DNA amplified, only one amplicon was visualized per reaction. Amplicons matching positive controls in length were classified as having the target gene present; a subset of amplicons were Sanger sequenced to confirm their identity (described below). Genes detected for each individual are tabulated in Supplementary Table 6. Across 68 individuals, all four genes were detected in 70.6% of the cohort, three genes in 22.1%, two genes in 4.4%, one gene in 1.5%, and no genes in 1.5%. We did not detect any significant associations between the presence or absence of each gene and the following host factors: sex, ethnicity, diabetes status, or body mass index (obese versus lean) (p-value>0.05, chi-square test of independence).

Of the 48 individuals in which all four genes associated with lignan metabolism were detected, all of the resultant amplicons from a subset of five individuals (designated in Supplementary Table 6) and positive controls (genomic DNA extracted from mono-cultured organisms served as PCR template) were purified using Agencourt AMPure PCR Purification kit (Beckman Coulter) according to the manufacturer's instructions for a 96-well format. After purification, the amplicon was subjected to Sanger sequencing (GENEWIZ) using the same forward primer that was used for PCR amplification. The resulting sequences were manually evaluated for correct base-calling and trimmed of unassignable bases ("N"s) at both 5' and 3' ends of the reads. The resulting sequence reads were then aligned to gene sequences obtained from draft genomes (NCBI RefSeq locus tags: *ber,* ELEN_RS01850; *glm*, C3R19_12985; *cldh*, C1877_07250; *edl*, BUA56_RS20800) via NCBI's nucleotide BLAST web tool. Percent identity and coverage of query sequences relative to subject sequences for is reported in Supplementary Table 8.

To query gut bacterial genomes, the integrated gut genomes (IGG) database was obtained for all species (N=23,790). The database is available at github.com/snayfach/IGGdb[33]. BLAST databases were generated (BLAST+ 2.6.0) on a per-assembly basis. A translated BLAST (tBLASTn) was then carried out using the protein sequences of our genes of interest (*ber, glm, cldh,* and *edl*), against each database with the following parameters: -evalue 1e-8, -max_target_seqs 1. Per-species BLAST results were then aggregated and filtered such that hits had an e-value <1e-30, pident>50, and query coverage >80. Hits were visualized on the IGG all bacterial species tree using ggtree[45].

To query gut metagenomes, Metaquery2 was used to determine abundances of the relevant genera by name with our genes of interest requiring a minimum of 70% identity and 70% query and target coverage by BLAST[32]. Where multiple samples were collected per subject, as identified from the subject_id, the median abundance was taken and only subjects with

matching taxonomic and gene abundance profiles were considered. Presence of the genera or gene was determined based on a non-zero count for the feature.

Didemethylsecoisolariciresinol (dmSECO) synthesis

General—All reagents and solvents were purchased from commercial suppliers and were used without further purification. Arctigenin was purchased from Ark Pharm (purity ≥ 98%). All reactions were performed in an inert atmosphere of argon. ^{1}H and ^{13}C NMR spectra were collected using a Bruker 600 MHz spectrometer with chemical shifts reported relative to residual deuterated solvent peaks or a tetramethylsilane internal standard. Accurate masses were measured using an ESI-TOF (HRMS, Agilent MSD) or MSQ Plus mass spectrometer (LRMS, Thermo Scientific). Reactions were monitored on TLC plates (silica gel 60, F254 coating, EMD Millipore, 1057150001), and spots were either monitored under UV light (254 mm) or stained with phosphomolybdic acid. The same TLC system was used to test purity, and the final product showed a single spot on TLC with both phosphomolybdic acid and UV absorbance. The purity of the final product was >95% based on ^{1}H NMR and reverse phase HPLC-UV on monitoring absorption at 254 nm.

Analytical LC method to determine the purity of synthetic compounds—Purity determination of the final product was performed on a Thermo Scientific Accela HPLC system using Accela 1250 pump. The Thermo Accucore C18 RP HPLC column (150 mm × 2.1 mm, particle size 2.6 μm) was used. The UV absorption between 190 nm and 400 nm was monitored, and the purity was determined by the peak area at 254 nm. A gradient method was used. Solvent A, 0.1% formic acid$_{(aq)}$; Solvent B, 99.9% acetonitrile with 0.1% formic acid; 0–1 min, 1% B; 1–16 min., 1–100% B; 16–19 min, 100% B; 19.01–22 min, 1% B; Flow rate: 0.5 mL/min.

Synthesis—Didemethylsecoisolariciresinol, **4**, was synthesized by the method described previously[36], with slight modifications.

(3*R*,4*R*)-3,4-bis(3,4-dihydroxybenzyl)dihydrofuran-2(3H)-one, 8—To a dichloromethane (DCM) solution of arctigenin, 7, (100 mg, 269 μmol) was added a DCM solution of BBr_3 (2M, 1.0 mL, 2 mmol) and stirred overnight. The reaction mixture was slowly added to ice-cold water and stirred vigorously for 60 min at room temperature. This solution was extracted with diethyl ether three times. The organic layer was combined and dried over $MgSO_4$, filtered and concentrated *in vacuo* to give fairly pure **8** as an off-white solid (90 mg, 1.7 mmol, quant.). ^{1}H NMR (600 MHz, Acetone-d_6) δ 6.79 – 6.75 (m, 2H), 6.73 (d, J = 8.0 Hz, 1H), 6.63 (d, J = 2.1 Hz, 1H), 6.58 (dd, J = 8.0, 2.1 Hz, 1H), 6.46 (dd, J = 8.0, 2.1 Hz, 1H), 4.01 (dd, J = 8.9, 7.6 Hz, 1H), 3.84 (t, J = 8.8 Hz, 1H), 2.90 – 2.79 (m, 2H), 2.64 – 2.57 (m, 2H), 2.53 – 2.44 (m, 1H), 2.41 (dd, J = 13.3, 9.3 Hz, 1H). ^{13}C NMR (151 MHz, Acetone) δ 179.1, 145.8, 145.8, 144.6, 144.4, 131.3, 130.7, 121.7, 120.8, 117.2, 116.5, 116.1, 116.0, 71.6, 57.8, 47.0, 42.1, 38.1, 34.7, 18.7. LRMS (−) calcd for $(M-H)^-$ 329.1. Found 329.2.

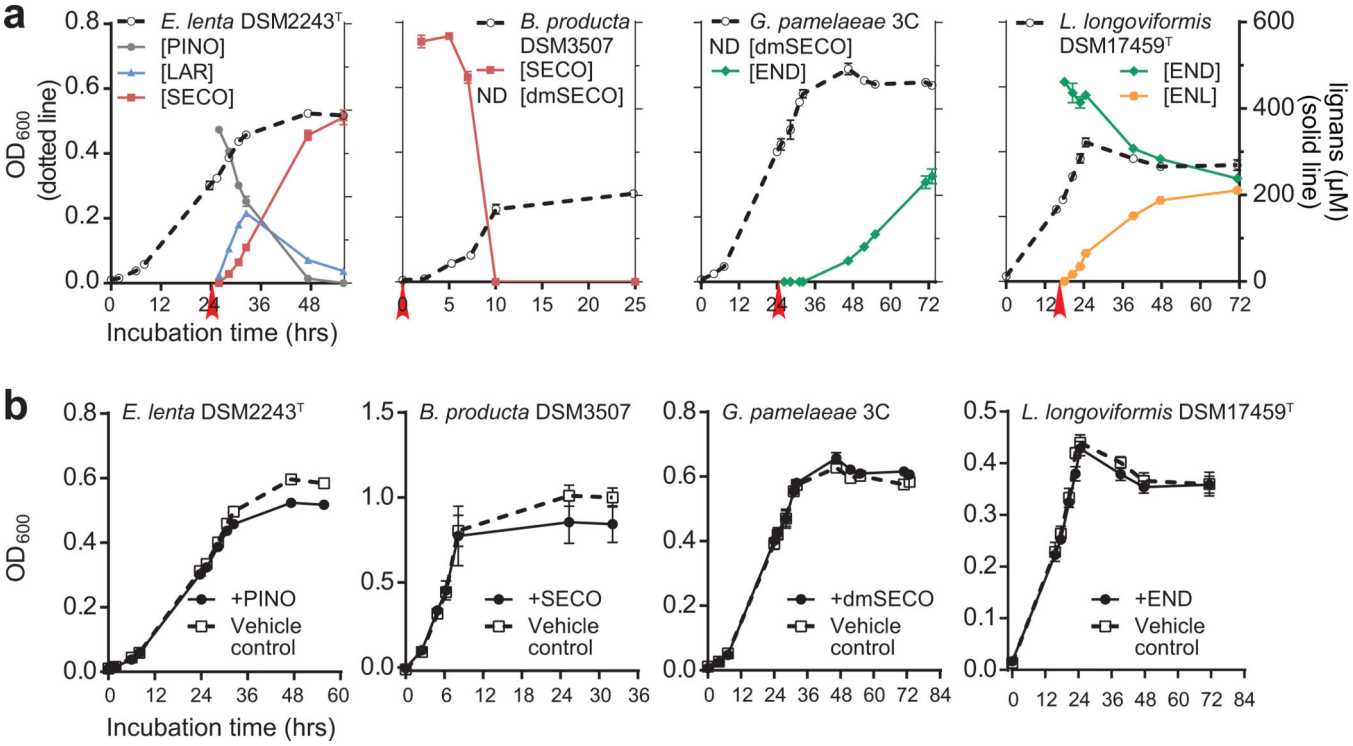
4,4'-((2*R*,3*R*)-2,3-bis(hydroxymethyl)butane-1,4-diyl)bis(benzene-1,2-diol), 4—To a tetrahydrofuran (THF) solution of **8** (50 mg, 152 μmol) was slowly added a THF

solution of lithium aluminum hydride (2M, 0.2 mL, 400 μmol), and stirred 2 hours. To this mixture was slowly added additional THF solution of lithium aluminum hydride (2M, 0.2 mL, 400 μmol), and stirred 2 hours. To this solution was slowly added aqueous H_2SO_4 solution (2N, 10 mL), which was then extracted with diethyl ether three times. The organic layer was combined, dried over $MgSO_4$, filtered and concentrated *in vacuo*. Purification by reverse-phase HPLC using an XBridge™ Prep C18 column (Waters, 5 μm, 19 × 150 mm) was repeated until the purity of **4** exceeded 95%. HPLC fraction containing the target compound was lyophilized to give **4** as an off-white solid (7.1 mg, 21 μmol, 14%).^{1}H NMR (600 MHz, Acetone-d_6) δ 7.70 (s, 2H), 7.63 (s, 2H), 6.71 – 6.68 (m, 4H), 6.51 (dd, J = 8.0, 2.1 Hz, 2H), 4.09 (br s, 2H), 3.66 (d, J = 11.2 Hz, 2H), 3.52 – 3.44 (m, 2H), 2.62 – 2.59 (m, 4H), 1.91 – 1.85 (m, 2H). ^{13}C NMR (151 MHz, Acetone) δ 145.6, 143.9, 134.0, 121.2, 117.0, 115.8, 61.0, 45.1, 35.7. HRMS (+) calcd for $(M+H)^+$ 335.1489. Found 335.1491.

Data Availability

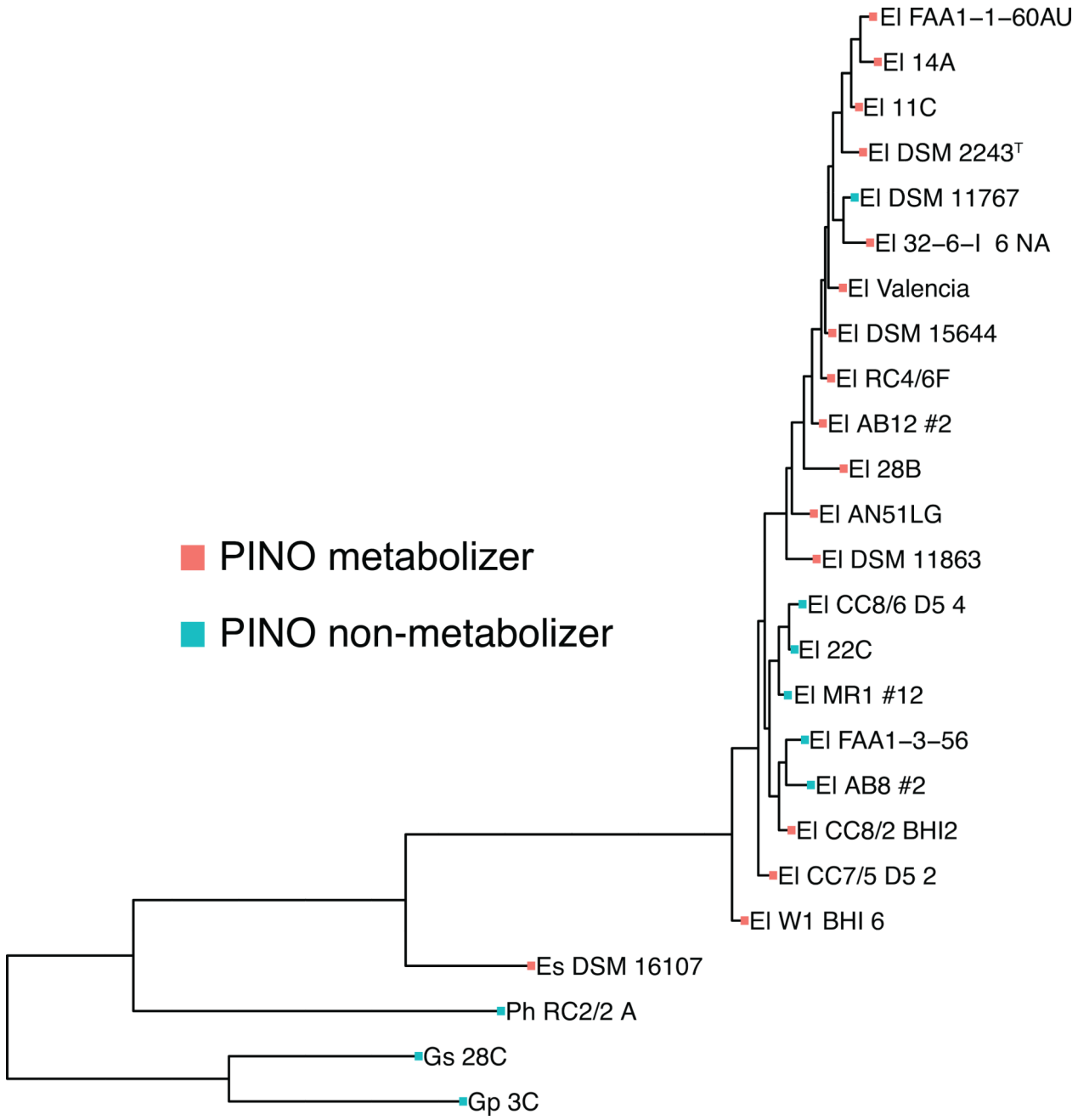
16S rDNA and RNA sequencing data have been deposited in the NCBI Sequence Read Archive under BioProject accession numbers PRJNA450120 and PRJNA412637. Figure source data and additional study data are available by request (peter.turnbaugh@ucsf.edu).

Extended Data



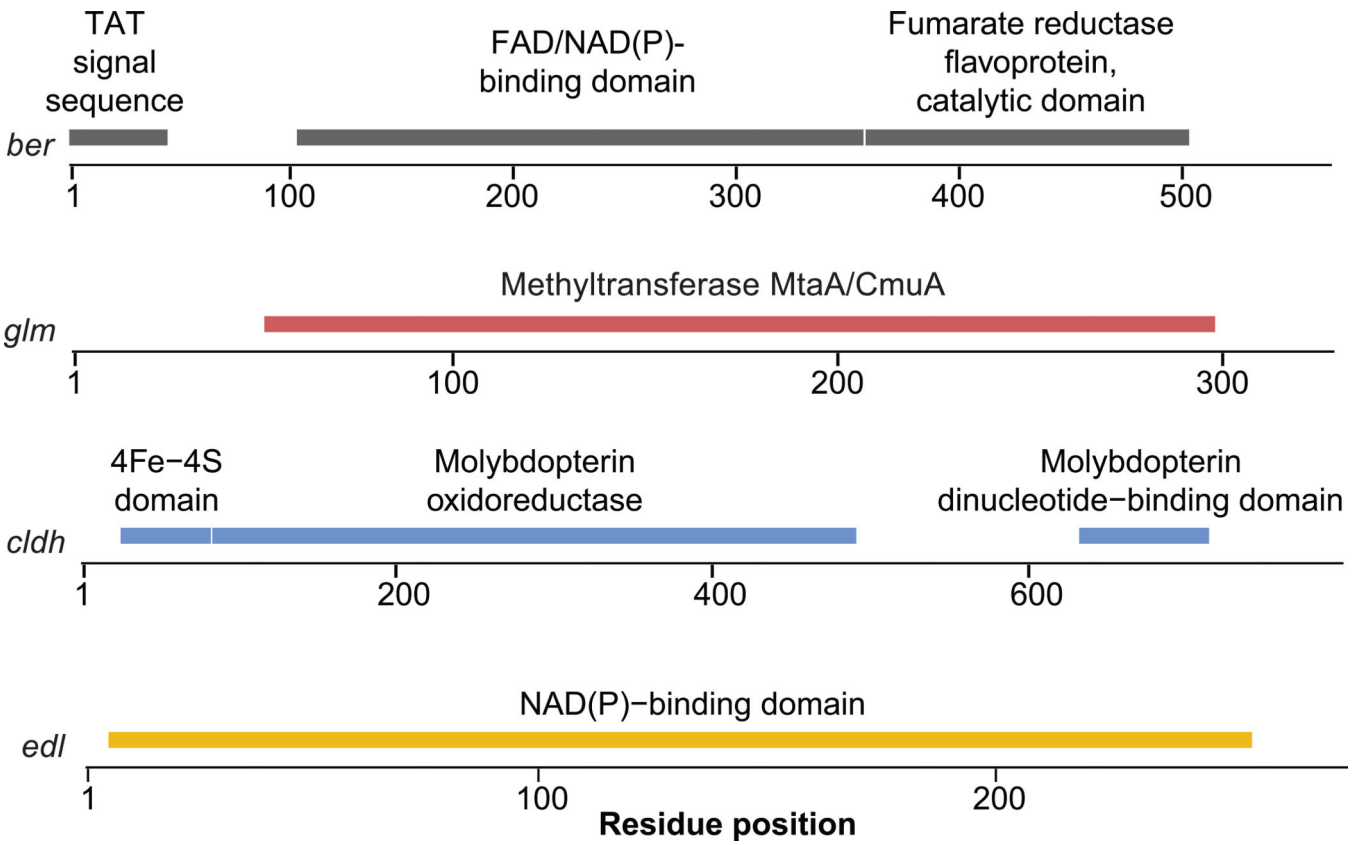
Extended Data Fig. 1. A four-member gut bacterial consortium is capable of converting dietary lignans to phytoestrogenic enterolignans.

a, Time-course experiments exhibiting the conversion of PINO to ENL and the growth profile of each bacterium with the lignan it metabolizes. Due to the chemical instability of dmSECO, this compound could not be accurately measured. Red arrows indicate time at which culture was exposed to lignan. Lignan concentrations were measured by HPLC. Culture turbidity, measured as optical density at 600 nm (OD_{600}), is plotted. Values are mean±SEM (n=3 biological replicates). **b**, Growth profiles of each bacterium cultured with and without lignan. Culture turbidity, measured as optical density at 600 nm (OD_{600}), is plotted. Values are mean±SEM (n=3 biological replicates).



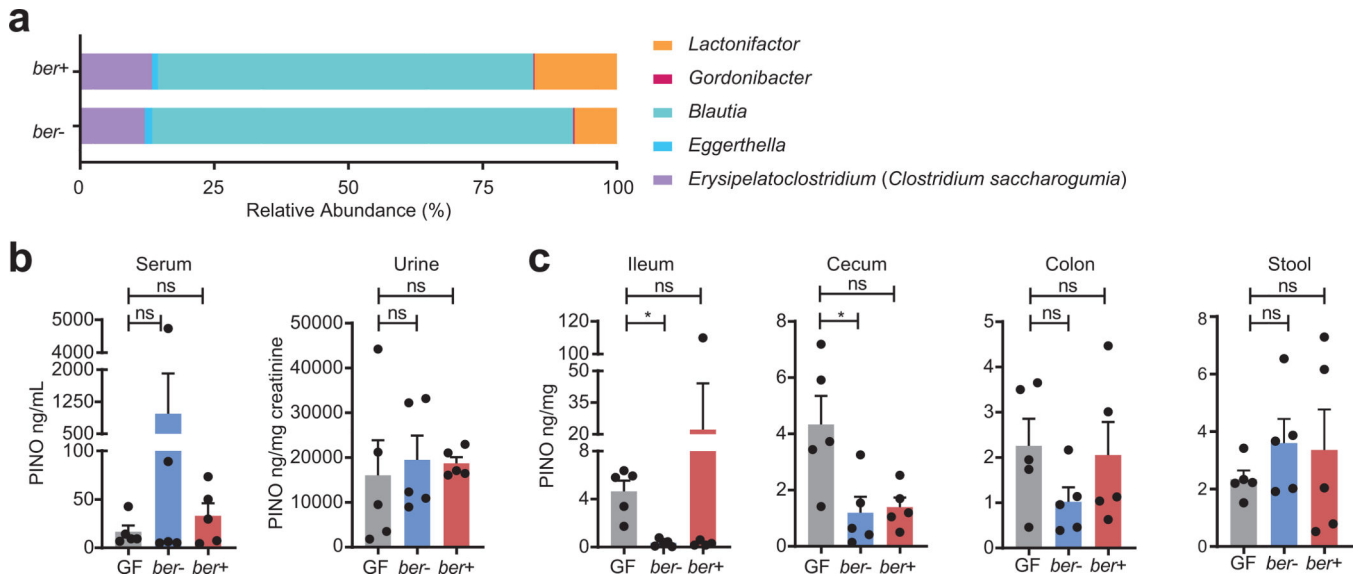
Extended Data Fig. 2. PINO-metabolizing Coriobacteriia strains cannot be predicted based on phylogeny.

Phylophlan-based phylogenetic tree produced using ElenMatchR: Comparative Genomics Tool v0.321. This tree demonstrates the non-monophyletic nature of PINO metabolism across the strain collection and suggests that this phenotypic trait is decoupled from bacterial evolutionary history, suggesting the repeated gain or loss of the genes responsible.



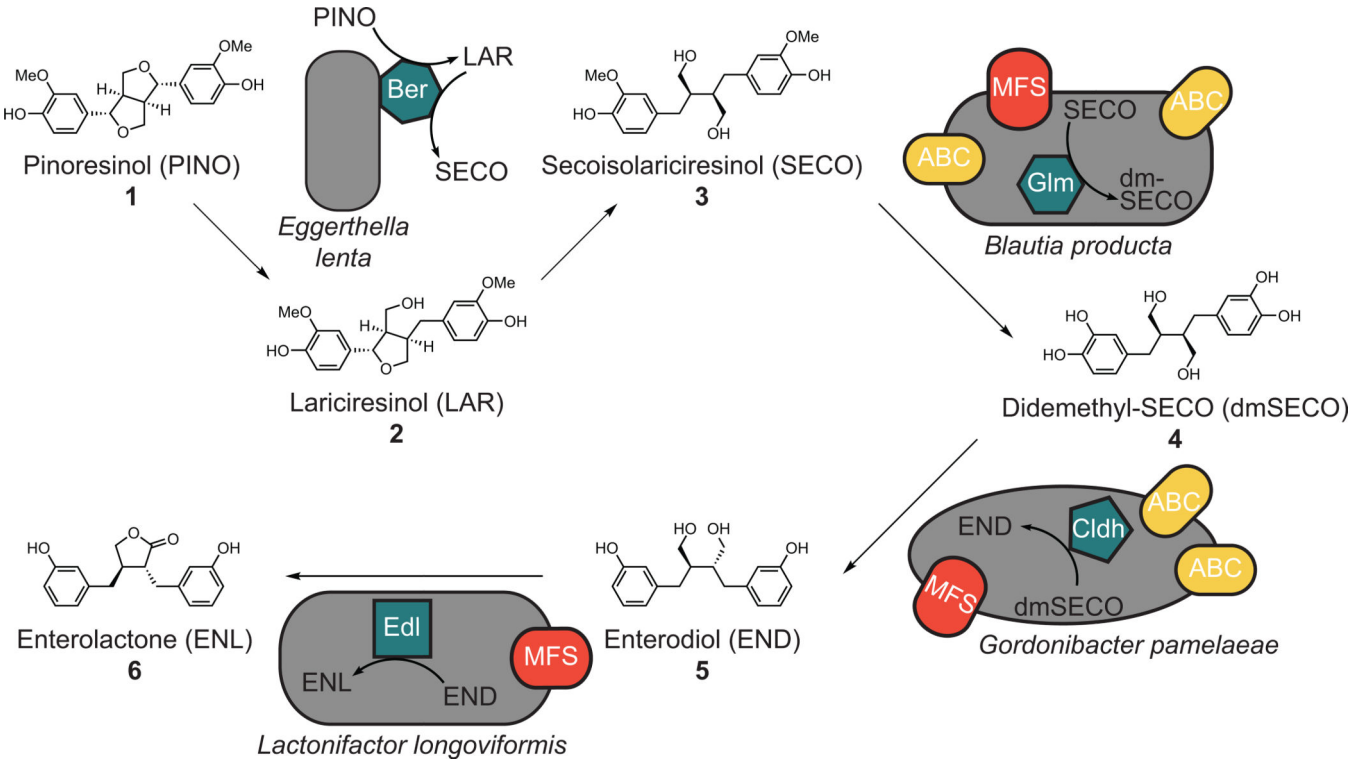
Extended Data Fig. 3. Domain maps for gut bacterial genes implicated in the lignan metabolism pathway.

Annotations, assigned by homology, of the domains that constitute the putative lignan-metabolizing enzymes are presented and provide support for the inferred biochemical functions. All proteins are predicted to be cytoplasmic with the exception of Ber, which has an N-terminal TAT signal sequence, targeting Ber for secretion.



Extended Data Fig. 4. Colonization and PINO levels for gnotobiotic mice dosed with PINO-diGlc.

a, Relative abundance of bacterial genera for each of the strains used to colonize mice, as measured by 16S rRNA gene sequencing. **b-c**, Lignan levels in mice dosed with PINO-diglucoside (20 mg/kg) measured by Orbitrap mass spectrometry. Bars are mean±SEM (n=5 biologically independent samples/colonization group, except in the ileum samples where *ber*$^{+}$ n=4 biologically independent samples). Kruskal-Wallis with Dunn's multiple comparisons test: *p<0.05. ns: not significant. *ber*$^{+}$ and *ber*$^{-}$: germ-free mice colonized with *E. lenta* DSM2243^{T} (*ber*$^{+}$ group) or *E. lenta* 1–3-56 (*ber*$^{-}$ group) and *B. producta* DSM3507, *G. pamelaeae* 3C, and *L. longoviformis* DSM17459^{T}; mice dosed with PINO-diglucoside were also colonized with *C. saccharogumia* DSM17460^{T}. GF: germ-free mice.



Extended Data Fig. 5. The putative enzymes mediating bacterial metabolism of dietary lignans.

A working model of the bacterial lignan metabolism pathway is presented. Several transporters, which traffic small molecules (ABC transporters) or ions (MFS transporters) across bacterial membranes, were significantly up-regulated in response to lignan doses and may be responsible for funneling substrates or products across cell membranes. Ber: benzyl ether reductase; Glm: guaiacol lignan methyltransferase; Cldh: catechol lignan dehydroxylase; Edl: enterodiol lactonizing enzyme; ABC: ATP-binding cassette; MFS: major facilitator superfamily.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Special thanks to Emily Balskus, Andrew Patterson, and Katie Pollard for comments on the manuscript. We are indebted to Michael Blaut for providing *E. lenta* SECO-Mt75m2; Lizett Ortiz de Ora for assistance generating the control construct for Edl expression; Felix Grun, Karly Torii, and Laurie Custer for technical assistance with mass spectrometry assays; and to Separation Research Ltd. (Turku, Finland) for the donation of chemicals. This work was supported by the National Institutes of Health (R01HL122593; R21CA227232), the Searle Scholars Program (SSP-2016–1352), and the University of California, Irvine, Department of Chemistry. P.J.T. is a Chan Zuckerberg Biohub investigator and a Nadia's Gift Foundation Innovator supported, in part, by the Damon Runyon Cancer Research Foundation (DRR-42–16). Fellowship support was provided by the Natural Sciences and Engineering Research Council of Canada (J.E.B.), the Canadian Institutes of Health and Research (P.S.), the Agency for Technology, Science and Research (Q.A.), and the Life Sciences Research Foundation and Howard Hughes Medical Institute (E.N.B.).

References

1. Adlercreutz H Lignans and human health. Crit. Rev. Clin. Lab. Sci. 44, 483–525 (2007). [PubMed: 17943494]
2. Clavel T, Doré J & Blaut M Bioavailability of lignans in human subjects. Nutr. Res. Rev. 19, 187–196 (2006). [PubMed: 19079885]
3. Spanogiannopoulos P, Bess EN, Carmody RN & Turnbaugh PJ The microbial pharmacists within us: a metagenomic view of xenobiotic metabolism. Nat. Rev. Microbiol. 14, 273–287 (2016). [PubMed: 26972811]
4. Carmody RN & Turnbaugh PJ Host-microbial interactions in the metabolism of therapeutic and diet-derived xenobiotics. J. Clin. Invest. 124, 4173–4181 (2014). [PubMed: 25105361]
5. Koppel N, Maini Rekdal V & Balskus EP Chemical transformation of xenobiotics by the human gut microbiota. Science 356, eaag2770 (2017). [PubMed: 28642381]
6. Valsta LM et al. Phyto-oestrogen database of foods and average intake in Finland. Br. J. Nutr. 89, S31–S38 (2011).
7. Woting A, Clavel T, Loh G & Blaut M Bacterial transformation of dietary lignans in gnotobiotic rats. FEMS Microbiol. Ecol. 72, 507–514 (2010). [PubMed: 20370826]
8. Clavel T & Mapesa JO Phenolics in Human Nutrition: Importance of the Intestinal Microbiome for Isoflavone and Lignan Bioavailability in Natural Products: Phytochemistry, Botany and Metabolism of Alkaloids, Phenolics and Terpenes (eds. Ramawat KG & Mérillon J-M) 2433–2463 (Springer Berlin Heidelberg, 2013).
9. Olsen A et al. Plasma enterolactone and breast cancer incidence by estrogen receptor status. Cancer Epidemiol. Biomarkers Prev. 13, 2084–2089 (2004). [PubMed: 15598765]
10. Sonestedt E et al. Enterolactone is differently associated with estrogen receptor β–negative and –positive breast cancer in a Swedish nested case-control study. Cancer Epidemiol. Biomarkers Prev. 17, 3241–3251 (2008). [PubMed: 18990767]
11. Seibold P et al. Enterolactone concentrations and prognosis after postmenopausal breast cancer: Assessment of effect modification and meta-analysis. International Journal of Cancer 135, 923–933 (2014). [PubMed: 24436155]
12. Mabrok HB et al. Lignan transformation by gut bacteria lowers tumor burden in a gnotobiotic rat model of breast cancer. Carcinogenesis 33, 203–208 (2012). [PubMed: 22080573]
13. Saarinen NM et al. Enterolactone inhibits the growth of 7,12-dimethylbenz(a) anthracene-induced mammary carcinomas in the rat. Mol. Cancer Ther. 1, 869–876 (2002). [PubMed: 12492120]
14. Zaineddin AK et al. Serum enterolactone and postmenopausal breast cancer risk by estrogen, progesterone and herceptin 2 receptor status. International Journal of Cancer 130, 1401–1410 (2012). [PubMed: 21544804]
15. Stitch SR et al. Excretion, isolation and structure of a new phenolic constituent of female urine. Nature 287, 738–740 (1980). [PubMed: 7432490]
16. Setchell KDR et al. Lignans in man and in animal species. Nature 287, 740–742 (1980). [PubMed: 6253812]
17. Clavel T et al. Intestinal bacterial communities that produce active estrogen-like compounds enterodiol and enterolactone in humans. Appl. Environ. Microbiol. 71, 6077–6085 (2005). [PubMed: 16204524]
18. Clavel T, Borrmann D, Braune A, Doré J & Blaut M Occurrence and activity of human intestinal bacteria involved in the conversion of dietary lignans. Anaerobe 12, 140–147 (2006). [PubMed: 16765860]
19. Wang L-Q, Meselhy MR, Li Y, Qin G-W & Hattori M Human intestinal bacteria capable of transforming secoisolariciresinol diglucoside to mammalian lignans, enterodiol and enterolactone. Chem. Pharm. Bull. 48, 1606–1610 (2000). [PubMed: 11086885]
20. Clavel T, Henderson G, Engst W, Doré J & Blaut M Phylogeny of human intestinal bacteria that activate the dietary lignan secoisolariciresinol diglucoside. FEMS Microbiol. Ecol. 55, 471–478 (2006). [PubMed: 16466386]

21. Bisanz JE et al. Illuminating the microbiome's dark matter: a functional genomic toolkit for the study of human gut Actinobacteria. bioRxiv (2018). doi:10.1101/304840
22. Haiser HJ et al. Predicting and manipulating cardiac drug inactivation by the human gut bacterium Eggerthella lenta. Science 341, 295–298 (2013). [PubMed: 23869020]
23. Koppel N, Bisanz JE, Pandelia M-E, Turnbaugh PJ & Balskus EP Discovery and characterization of a prevalent human gut bacterial enzyme sufficient for the inactivation of a family of plant toxins. eLife 7, e33953 (2018). [PubMed: 29761785]
24. Maini Rekdal V, Bess EN, Bisanz JE, Turnbaugh PJ & Balskus EP Discovery and inhibition of an interspecies gut bacterial pathway for Levodopa metabolism. Science 364, eaau6323 (2019). [PubMed: 31196984]
25. Davis RM, Muller RY & Haynes KA Can the natural diversity of quorum-sensing advance synthetic biology? Frontiers in Bioengineering and Biotechnology 3, (2015).
26. Rohman A, van Oosterwijk N, Thunnissen A-MWH & Dijkstra BW Crystal Structure and Site-directed Mutagenesis of 3-Ketosteroid Δ1-Dehydrogenase from Rhodococcus erythropolis SQ1 Explain Its Catalytic Mechanism. J. Biol. Chem. 288, 35559–35568 (2013). [PubMed: 24165124]
27. Fukuhara Y et al. Discovery of pinoresinol reductase genes in sphingomonads. Enzyme Microb. Technol. 52, 38–43 (2013). [PubMed: 23199737]
28. Maier L et al. Extensive impact of non-antibiotic drugs on human gut bacteria. Nature 555, 623 (2018). [PubMed: 29555994]
29. Milder IEJ et al. Intake of the Plant Lignans Secoisolariciresinol, Matairesinol, Lariciresinol, and Pinoresinol in Dutch Men and Women. J. Nutr. 135, 1202–1207 (2005). [PubMed: 15867304]
30. Matthews RG Cobalamin- and corrinoid-dependent enzymes. Met. Ions Life Sci. 6, 53–114 (2009). [PubMed: 20877792]
31. Hille R The molybdenum oxotransferases and related enzymes. Dalton Trans. J. Inorg. Chem. 42, 3029–3042 (2013).
32. Nayfach S, Fischbach MA & Pollard KS MetaQuery: a web server for rapid annotation and quantitative analysis of specific genes in the human gut microbiome. Bioinformatics 31, 3368–3370 (2015). [PubMed: 26104745]
33. Nayfach S, Shi ZJ, Seshadri R, Pollard KS & Kyrpides NC New insights from uncultivated genomes of the global human gut microbiome. Nature 568, 505–510 (2019). [PubMed: 30867587]
34. Alba DL et al. Subcutaneous fat fibrosis links obesity to insulin resistance in Chinese-Americans. J. Clin. Endocrinol. Metab. jc.2017–02301 (2018).
35. Langmead B & Salzberg SL Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357 (2012). [PubMed: 22388286]
36. Anders S, Pyl PT & Huber W HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169 (2015). [PubMed: 25260700]
37. Anders S & Huber W Differential expression analysis for sequence count data. Genome Biol. 11, R106 (2010). [PubMed: 20979621]
38. Tang H & Mayersohn M Porcine Preduction of Pharmacokinetic Parameters in People: A Pig in a Poke? Drug Metab. Dispos. dmd.118.083311 (2018).
39. Bolvig AK, Adlercreutz H, Theil PK, Jørgensen H & Bach Knudsen KE Absorption of plant lignans from cereals in an experimental pig model. Br. J. Nutr. 115, 1711–1720 (2016). [PubMed: 27001342]
40. Gohl DM et al. Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. Nat. Biotechnol. 34, 942–949 (2016). [PubMed: 27454739]
41. Callahan BJ et al. DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods 13, 581–583 (2016). [PubMed: 27214047]
42. Franke AA et al. Liquid chromatographic–photodiode array mass spectrometric analysis of dietary phytoestrogens from human urine and blood. Journal of Chromatography B 777, 45–59 (2002).
43. Franke AA, Halm BM, Kakazu K, Li X & Custer LJ Phytoestrogenic isoflavonoids in epidemiologic and clinical research. Drug Test. Anal. 1, 14–21 (2009). [PubMed: 20355154]
44. Begum AN et al. Dietary Lignins Are Precursors of Mammalian Lignans in Rats. J. Nutr. 134, 120–127 (2004). [PubMed: 14704303]

45. Yu G, Smith DK, Zhu H, Guan Y & Lam TT-Y ggtree : an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol. Evol. 8, 28–36 (2017).

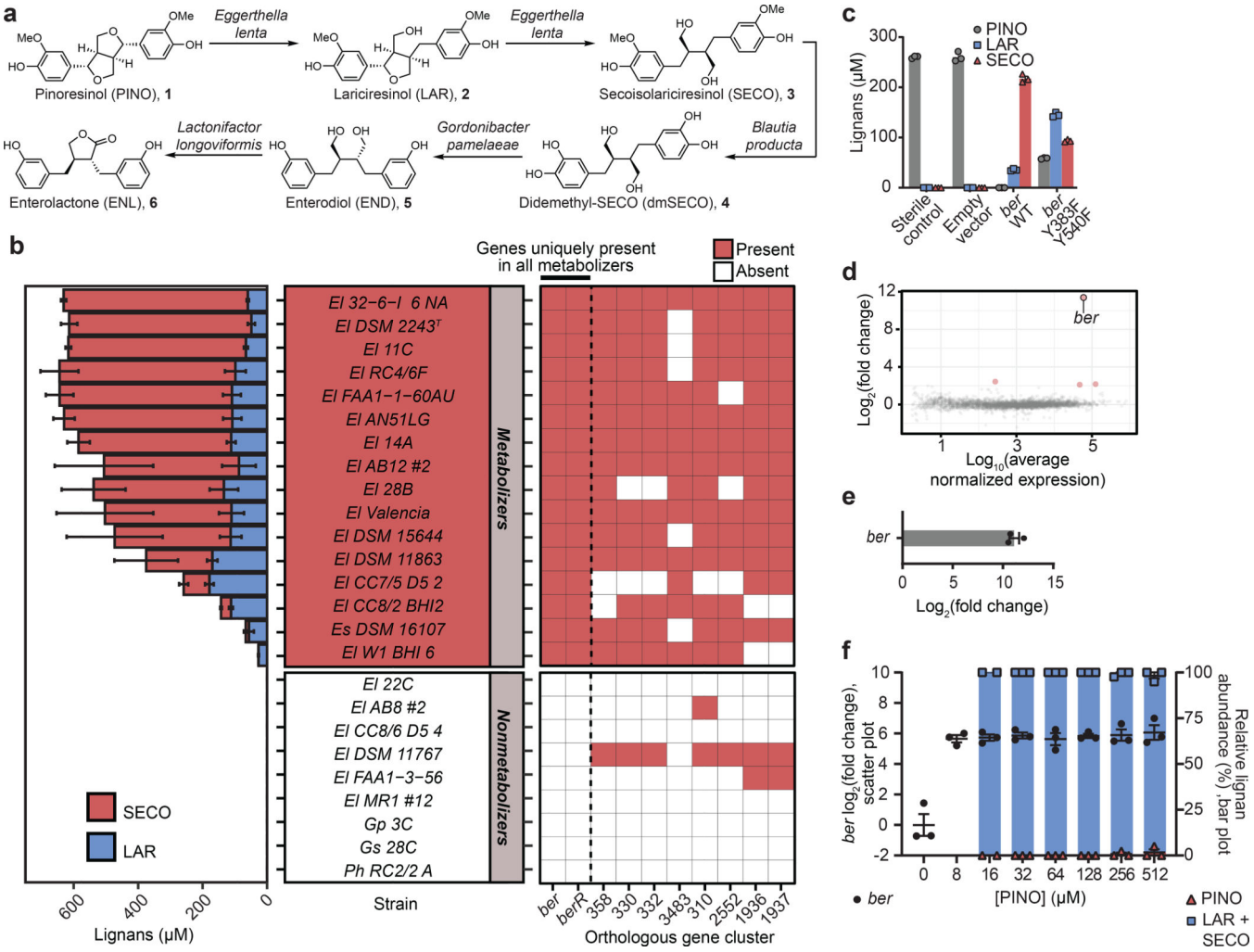


Figure 1. Lignan metabolism varies between Coriobacteriia strains and enables identification of a single enzyme sufficient to catalyze the first two reactions in the lignan metabolism pathway.

a, Known biotransformations by which gut-residing bacteria can convert pinoresinol (PINO) into the enterolignan products enterodiol (END) and enterolactone (ENL). **b**, Evaluation of PINO-metabolizing ability across our collection of strains from the Coriobacteriia class, of which *E. lenta* is a member. Lignans were quantified by HPLC (values are mean±SEM, n=3 biologically independent samples). *ber*: benzyl ether reductase; *berR*: benzyl ether reductase regulator; El: *Eggerthella lenta*; Es: *Eggerthella sinesis*; Gp: *Gordonibacter pamelaeae*; Gs: *Gordonibacter* species; Ph: *Paraeggerthella hongkongensis*. **c**, Incubation of PINO (250 μM) with *E. coli* Rosetta 2(DE3) expressing WT or mutagenized Ber on a pET19btev plasmid. PINO and the products of its benzyl ether reduction, lariciresinol (LAR) and secoisolaricerisinol (SECO), were quantified by HPLC (bars are mean±SEM, n=3 biologically independent samples). **d**, RNAseq results obtained upon exposure of *E. lenta* DSM2243^T to PINO (n=3 biologically independent samples) relative to vehicle (n=3 biologically independent samples). Points in pink are genes for which was observed a fold-change>|4| and FDR<0.01 (Wald test, Benjamini–Hochberg multiple testing correction). **e**, RT-qPCR confirmation of *ber* up-regulation upon exposure of *E. lenta* DSM2243^T to PINO

(500 μM) relative to vehicle (bar is mean±SEM, n=3) and **f**, at varying concentrations of PINO (mean±SEM, n=3 biologically independent samples), along with conversions of PINO to LAR and SECO at respective doses of PINO (bar is mean±SEM, n=3 biologically independent samples; except 16 μM, n=2 biologically independent samples). Lignan concentrations measured by HPLC. Lignan conversion data not available for 8 μM of PINO due to instrument's limit of quantitation.

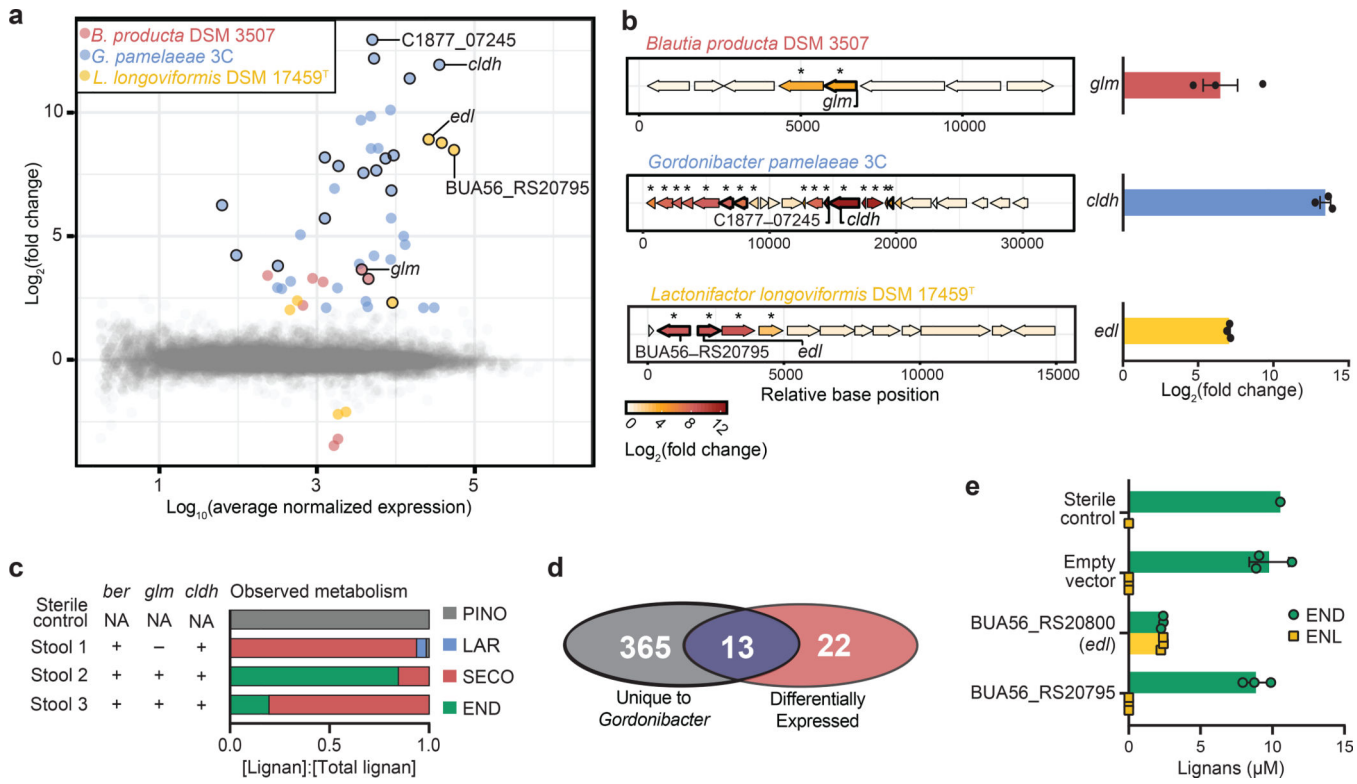


Figure 2. Genomic loci are up-regulated in response to each substrate in the lignan metabolism pathway.

a, Plot of RNAseq results observed from the exposure of *B. producta* DSM3507 to SECO, *G. pamelaeae* 3C to dmSECO, and *L. longoviformis* DSM17459^T to END (n=3 biologically independent samples per organism) relative to respective vehicle exposures (n=3 biologically independent samples per organism). Colored points have a fold-change>|4| and FDR<0.01 (Wald test, Benjamini–Hochberg multiple testing correction). Points encircled with black designate clusters of genes that are represented in the locus diagrams in Fig. 2b. **b**, The most up-regulated genomic loci in *B. producta* DSM3507, *G. pamelaeae* 3C, and *L. longoviformis* DSM17459^T. Genes that are outlined in bold demonstrate homology to enzymes (Supplementary Table 3). The genes implicated in lignan-metabolizing biotransformations are annotated: guaiacol lignan methyltransferase (*glm*), catechol lignan dehydroxylase (*cldh*), enterodiol lactonizing enzyme (*edl*). Asterisks denote that the gene is up-regulated in the presence of lignan (n=3 biologically independent samples) relative to vehicle (n=3 biologically independent samples) with a fold-change>|4| and FDR<0.01. Up-regulation of *glm*, *cldh*, and *edl* upon exposure to respective lignan substrates relative to vehicle was confirmed through RT-qPCR (bar is mean±SEM, n=3 biologically independent samples). **c**, Gene presence of *ber*, *glm*, and *cldh* (assessed by PCR) in human stool samples. Plot of metabolism resulting from *ex vivo* incubation of human stool samples with PINO for 7 days. **d**, Venn diagram showing the overlap between genes implicated in dmSECO metabolism by comparative genomics and RNA sequencing (reanalyzed using alternate gene-calling methodology, see Methods). **e**, Incubation of END (10 μM) with *E. coli* Rosetta 2(DE3) expressing Edl on a pET151/D-TOPO plasmid. END and ENL were quantified by mass spectrometry (bars are mean±SEM, n=3 biologically independent samples).

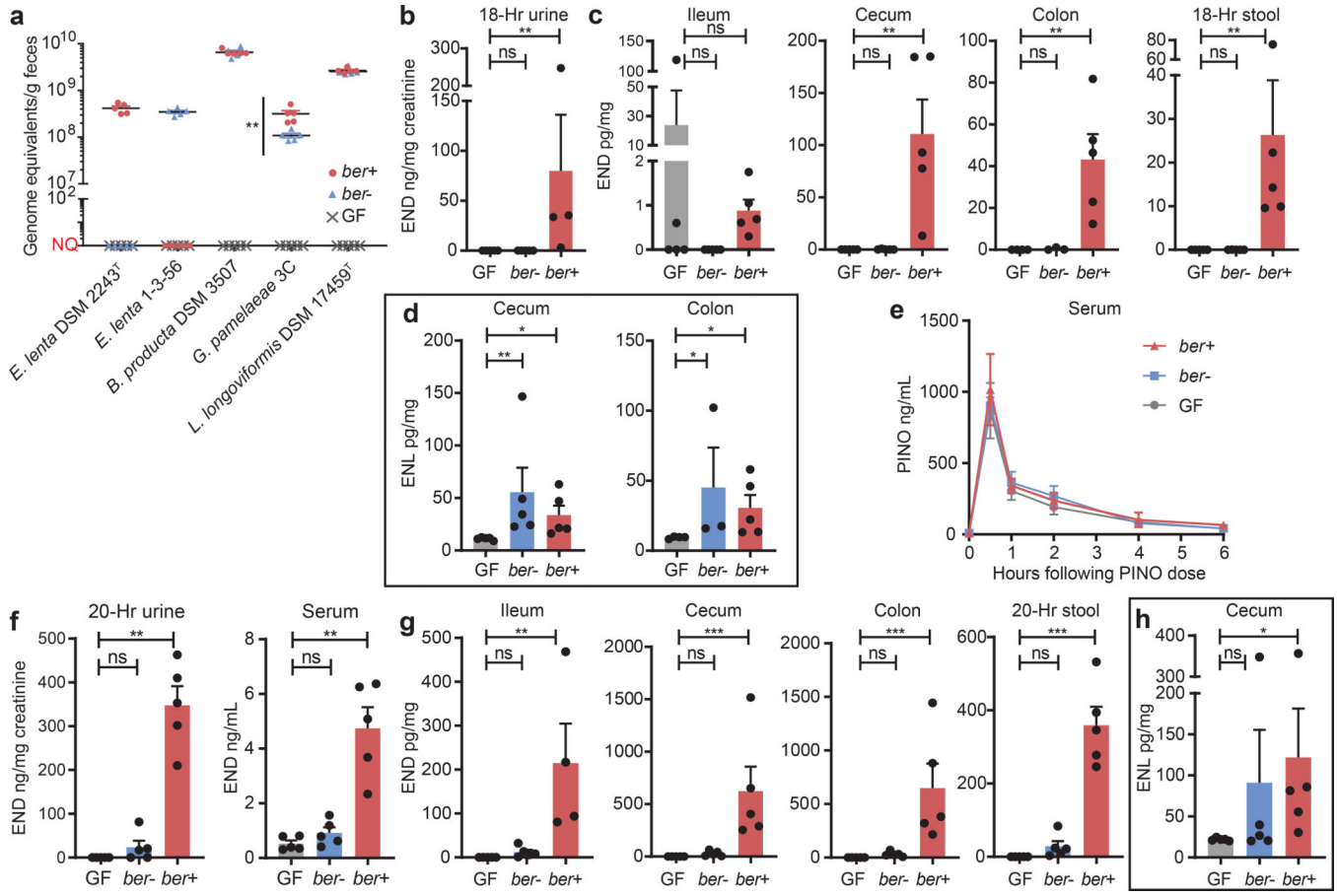


Figure 3. *Ber* significantly alters enterolignan production in gnotobiotic mice.

a, Mouse bacterial colonization levels were assessed by qPCR. Genome equivalents not detected or below the limit of quantitation (3×10^2 genome equivalents) are represented as NQ (not quantifiable). Center lines are mean±SEM (n=5 biologically independent samples). Mann-Whitney test for significance: $^{**}p=0.0079$, two-tailed. **b-h**, Lignan levels in mice dosed with **b-e**, PINO (10 mg/kg) or **f-h**, PINO-diglucoside (20 mg/kg) measured by Orbitrap mass spectrometry. Panels **b,c,f,g** represent END levels. Panels **d,h** represent ENL levels. Panel **e** represents PINO levels. Values are mean±SEM (n=5 biologically independent samples/colonization group with the following exceptions: **b-e**, 18-hr urine (*ber*$^+$, n= 4 biologically independent samples), colon (GF, n=4 biologically independent samples; *ber*$^-$, n=3 biologically independent samples), and serum at 6-hr timepoint (*ber*$^+$, n= 4 biologically independent samples); **g**, ileum (*ber*$^+$, n= 4 biologically independent samples). Kruskal-Wallis with Dunn's multiple comparisons test: $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$. ns: not statistically significant. *ber*$^+$ and *ber*$^-$: germ-free mice colonized with *E. lenta* DSM2243^T (*ber*$^+$ group) or *E. lenta* 1–3-56 (*ber*$^-$ group) and *B. producta* DSM3507, *G. pamelaeae* 3C, and *L. longoviformis* DSM17459^T; mice dosed with PINO-diglucoside were also colonized with *C. saccharogumia* DSM17460^T. GF: germ-free mice.

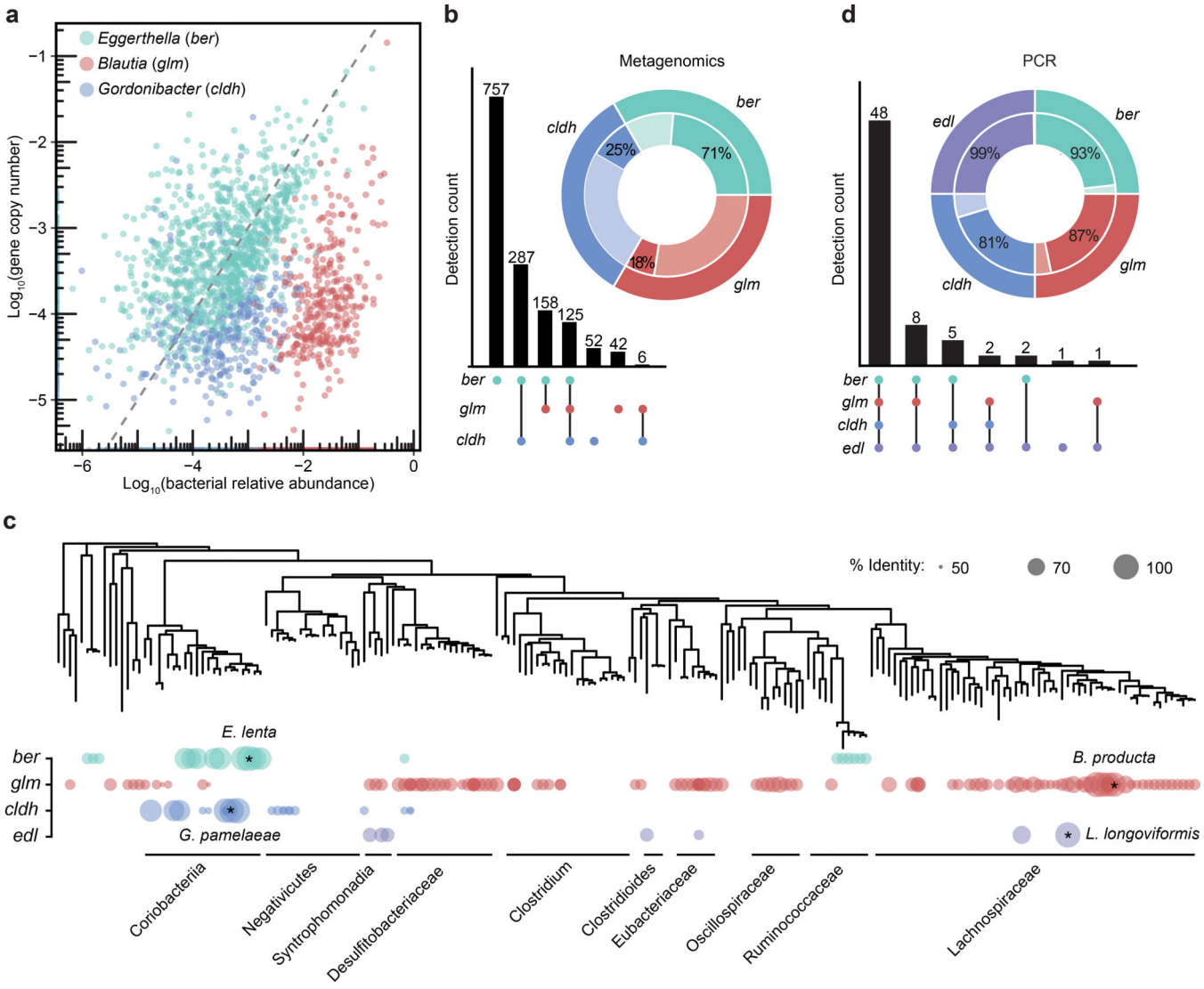


Figure 4. Genes linked to lignan metabolism are significantly correlated with bacterial host genera and prevalent in human gut microbiomes.

a, Correlation of putative lignan-metabolizing genes with host genera in diverse metagenomes (total subjects=1870). The dotted line (x=y) represents a perfect linear correlation. **b,** Co-occurrence of the lignan metabolizing pathway in metagenomes. The UpSet plot represents within-sample patterns in gene co-occurrence and the number of samples in which that pattern is detected (detection count). Inset: calculated prevalence. **c,** Phylogenetic distribution of proteins linked to lignan metabolism across gut microbial genomes. A reported hit is denoted by a colored circle with size proportional to the percentage identity to each query sequence. **d,** PCR detection of genes implicated in lignan metabolism in the stool samples of 68 adults living in Northern California (for one individual, no genes were identified; Supplementary Table 6).