

METHODS AND RESOURCES

Systems biology elucidates the distinctive metabolic niche filled by the human gut microbe *Eggerthella lenta*

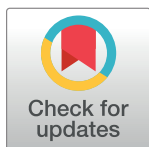
Cecilia Noecker¹, Juan Sanchez², Jordan E. Bisanz^{1,2a}, Veronica Escalante¹, Margaret Alexander¹, Kai Trepka¹, Almut Heinken^{3,4b}, Yuanyuan Liu⁴, Dylan Dodd^{4,5}, Ines Thiele^{3,6,7}, Brian C. DeFelice², Peter J. Turnbaugh^{1,2,*}

1 Department of Microbiology & Immunology, University of California, San Francisco, San Francisco, California, United States of America, **2** Chan Zuckerberg Biohub—San Francisco, San Francisco, California, United States of America, **3** School of Medicine, National University of Ireland, Galway, Ireland, **4** Department of Pathology, Stanford University, Stanford, California, United States of America, **5** Department of Microbiology & Immunology, Stanford University, Stanford, California, United States of America, **6** Ryan Institute, University of Galway, Galway, Ireland, **7** APC Microbiome Ireland, Cork, Ireland

^{2a} Current address: Department of Biochemistry and Molecular Biology, Pennsylvania State University, State College, Pennsylvania, United States of America

^{4b} Current address: INSERM UMRS 1256, Nutrition, Genetics, and Environmental Risk Exposure (NGERE), University of Lorraine, Nancy, France

* peter.turnbaugh@ucsf.edu



OPEN ACCESS

Citation: Noecker C, Sanchez J, Bisanz JE, Escalante V, Alexander M, Trepka K, et al. (2023) Systems biology elucidates the distinctive metabolic niche filled by the human gut microbe *Eggerthella lenta*. PLoS Biol 21(5): e3002125. <https://doi.org/10.1371/journal.pbio.3002125>

Academic Editor: Sebastian E. Winter, UT Southwestern: The University of Texas Southwestern Medical Center, UNITED STATES

Received: October 4, 2022

Accepted: April 14, 2023

Published: May 19, 2023

Copyright: © 2023 Noecker et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All resources and data associated with this publication are publicly available. RNA sequencing data are available in NCBI GEO accession number GSE212420. Metabolomics datasets are available in the Metabolomics Workbench repository, accession number PR001620. Genome-scale metabolic reconstructions of *E. lenta* strains and associated data are available from Zenodo at doi:[10.5281/zenodo.7063254](https://doi.org/10.5281/zenodo.7063254). Processed datasets underlying all figures are also available at Zenodo at doi:[10.5281/zenodo.7063254](https://doi.org/10.5281/zenodo.7063254).

Abstract

Human gut bacteria perform diverse metabolic functions with consequences for host health. The prevalent and disease-linked Actinobacterium *Eggerthella lenta* performs several unusual chemical transformations, but it does not metabolize sugars and its core growth strategy remains unclear. To obtain a comprehensive view of the metabolic network of *E. lenta*, we generated several complementary resources: defined culture media, metabolomics profiles of strain isolates, and a curated genome-scale metabolic reconstruction. Stable isotope-resolved metabolomics revealed that *E. lenta* uses acetate as a key carbon source while catabolizing arginine to generate ATP, traits which could be recapitulated in silico by our updated metabolic model. We compared these in vitro findings with metabolite shifts observed in *E. lenta*-colonized gnotobiotic mice, identifying shared signatures across environments and highlighting catabolism of the host signaling metabolite agmatine as an alternative energy pathway. Together, our results elucidate a distinctive metabolic niche filled by *E. lenta* in the gut ecosystem. Our culture media formulations, atlas of metabolomics data, and genome-scale metabolic reconstructions form a freely available collection of resources to support further study of the biology of this prevalent gut bacterium.

Introduction

Human gut bacteria perform diverse and specialized metabolic functions with consequences for host health. Yet the core metabolic strategies relied upon for growth by many commensal gut microbes remain unclear, which is reflected in the large number of gut taxa that remain

5281/zenodo.7779458. All original code used in this study is available on GitHub and at Zenodo (https://github.com/turnbaughlab/2022_Noecker_ElentaMetabolism, doi:10.5281/zenodo.7779454).

Funding: This work was supported by the National Institutes of Health (2R01HL122593; 1R01AT011117; 1R01DK114034 to P.J.T., F32GM140808; T32AI060537 to C.N., T32GM141323 to K.T., K08DK110335; R35GM142873; R01AT011396 to D.D., 1RF1AG058942; 1U19AG063744 I.T., R25GM056847 to V.E.) and the National Science Foundation (1650113 to V.E.). P.J.T. is a Chan Zuckerberg Biohub - San Francisco Investigator (7028823) and held an Investigators in the Pathogenesis of Infectious Disease Award (1017921) from the Burroughs Wellcome Fund. D. D. was also supported by the Stanford Microbiome Therapies Initiative and an OHF-ASN Foundation for Kidney Research Career Development Award. I. T. was supported by the European Research Council (757922) and a Science Foundation Ireland grant (12/RC/2273-P2). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: P.J.T. is on the scientific advisory boards for Pendulum, Seed, and SNIPRbiome; there is no direct overlap between the current study and these consulting duties. D.D. is a co-founder of Federation Bio, a company developing microbiome-based therapeutics. All other authors have no relevant declarations.

Abbreviations: COBRA, Constraint-based Reconstruction and Analysis; EDM1, *Eggerthella* Defined Media 1; FBA, flux balance analysis; FDR, false discovery rate; FVA, flux variability analysis; GF, germ-free; LC-MS, liquid chromatography–mass spectrometry; MSI, Metabolomics Standards Initiative; pFBA, parsimonious flux balance analysis; PFOR, pyruvate-ferredoxin oxidoreductase; PBS, phosphate-buffered saline; SIRM, stable isotope-resolved metabolomics.

difficult to culture [1,2]. The growth strategies of individual gut species and strains shape their ability to colonize a host and their potential chemical interactions with other community members and with the host [3,4]. Efforts to describe and model the metabolism and growth of various community members have included detailed biochemical studies of resource utilization by individual model species such as members of the genus *Bacteroides* [5] and *Clostridium sporogenes* [6], as well as large-scale efforts to characterize species-level metabolic activity using community multiomic profiling [7,8]. However, these efforts have been most fruitful for members of the microbiota that are found at high abundance and with prior knowledge of well-annotated metabolic pathways.

One key group of human gut microbes whose core metabolism remains particularly unclear are those that are fully asaccharolytic, i.e., derive no growth benefit from sugars and instead may rely on a range of more unconventional nutrients. Many of these taxa are members of the family *Eggerthellaceae*, which are widely found in mammalian gut microbiota [9] but rarely found in other environments. The species *Eggerthella lenta* is a notable example of this group. *E. lenta* is a gram-positive facultative anaerobe found at high prevalence in human gut microbiota [10]. Although *E. lenta* is commonly found in healthy individuals, it can cause severe bacteremia [11] and is increased in abundance in the gut microbiota of patients with several autoimmune diseases [12–15].

E. lenta has distinctive metabolic properties and a capacity for many unusual chemical transformations, but it remains unknown how these properties fit into its overall metabolic network and evolutionary strategy. *E. lenta* strains can metabolize varied mammalian and dietary substrates, including cardenolides, bile acids, plant lignans, and dopamine [10,16–19]. However, none of these compounds except dopamine have been reported to provide a growth or fitness advantage in any conditions reported to date. Genome analysis of *E. lenta* has also predicted that it may be able to perform autotrophic acetogenesis [20], but this prediction has not been biochemically validated. *E. lenta* culture conditions typically require rich media and high levels of the amino acid L-arginine. Past studies reported little to no growth of *E. lenta* in minimal or chemically defined media formulations [1,21,22], complicating mechanistic biochemical studies of its metabolism.

In this study, we first developed a chemically defined media that supports strong growth of *E. lenta* strains and described the metabolic footprint and growth determinants of *E. lenta* in this environment. We used stable isotope-resolved metabolomics (SIRM) to investigate the pathways by which *E. lenta* metabolizes 2 key nutrients, acetate and arginine. This platform allowed us to curate and interpret a genome-scale metabolic model of the *E. lenta*-type strain to make predictions about untested growth conditions and to identify gaps in the metabolic network representing novel enzymes or pathways. Extending this approach, we further documented extensive diversity in the metabolic footprint of a collection of *E. lenta* strain isolates. Finally, we evaluated the relevance of these findings to a host-associated context by profiling the metabolome of *E. lenta*-colonized gnotobiotic mice, defining shared and divergent metabolic activities between in vitro and in vivo environments. In total, we elucidate an unusual metabolic niche and lay a comprehensive foundation for future mechanistic studies of *E. lenta* metabolism.

Results

Extensive metabolite footprint of *Eggerthella lenta* in chemically defined media

To identify key nutrients and metabolic pathways required for growth of *E. lenta*, we first developed a custom chemically defined media formulation, referred to as *Eggerthella* Defined

Media 1 (EDM1). We designed the initial EDM1 formulation by making several modifications to a recipe previously reported to support growth of many human gut bacterial isolates but not *E. lenta* [1]. We increased the quantity of L-arginine, removed sugars, and ensured the availability of all amino acids and vitamins/cofactors with fragmented or missing biosynthetic pathways in the *E. lenta* DSM 2243 genome (Virtual Metabolic Human database annotations [23]; [Materials and methods, S1 Table](#)). The resulting media is composed of compounds typically present in the mammalian gut from microbial, host, and/or dietary sources. It supported robust *E. lenta* growth at a level comparable with standard culture conditions (Brain Heart Infusion media supplemented with 1% arginine; [S1A and S1B Fig](#)).

Using this platform, we sought to identify primary metabolites used and produced by *E. lenta* and the underlying core metabolic pathways active in the EDM1 condition. We used untargeted metabolomics to analyze culture supernatants of the type strain *E. lenta* DSM 2243 across 6 time points over its 50-hour growth curve in EDM1 batch culture ([Fig 1A](#)). We assigned putative identifications to metabolomics features by comparison with local library standards and public databases, with associated confidence levels for each identification based on a modified version of the Metabolomics Standards Initiative (MSI) framework ([Materials and methods](#)). After dereplication of features from positive and negative ionization modes, 4,095 features were detected, of which 636 (15.6%) were not detected in sterile control media (sample mean intensity $>3\times$ blank sample mean; [Fig 1B](#)). This large number of features likely represents a complex mixture of true known and unknown metabolites along with fragmentation products and other chemical artifacts, as is typically observed in untargeted metabolomics experiments [24]. A total of 612 features (14.9% of features overall) were significantly different in abundance between sterile controls and supernatants at the final time point (false discovery rate (FDR)-adjusted $p < 0.1$; [Fig 1C](#)), of which the majority (444, 72.5%) were increased in *E. lenta* cultures. Notably, the number of differentially abundant features at the final time point, both in total and among those assigned an identification, is substantially higher than previously reported metabolomic profiles of this species in ISP-2 and Mega media [25,26] ([S1C Fig](#)). This increased sensitivity was expected given our use of both chemically defined culture media and untargeted metabolomics.

Metabolites of diverse chemical classes are modified by *E. lenta* ([Fig 1C and 1D](#)). Compounds produced by *E. lenta* tended to be amino acid and nucleic acid metabolites. As expected, these included ornithine and citrulline, suggesting activity from the arginine deiminase pathway, which is highly expressed by *E. lenta* in the presence of arginine [16]. However, other arginine-related metabolites were also produced at lower levels, including N,N-dimethylarginine, N₅-(1-iminoethyl)-ornithine, and homocitrulline, suggesting that arginine may also be metabolized via other pathways. Several other metabolites produced at lower levels appeared to be products of metabolism of other amino acids in the media, including 4-methyl-2-hydroxy-pentanoic acid (from leucine), indole-3-acetate and indole-3-lactic acid (from tryptophan), and 3-phenyllactic acid (from phenylalanine), consistent with one previous report of production of indole-containing compounds and phenyl acids by *E. lenta* [27]. Other metabolites produced in supernatants included the amino acids alanine, glutamate, glutamine, histidine, and lysine, as well as several intermediates in biosynthesis of both purines and pyrimidines (inosine, orotic acid, hypoxanthine, uridine, and thymidine). Overall, the set of metabolites produced by *E. lenta* supports its previously reported dependence on arginine catabolism but is highly multifaceted.

Of the 54 compounds in our EDM1 recipe, 22 were detected by untargeted metabolomics but just 3 were depleted significantly in *E. lenta* cultures ([Figs 1D and S1D](#)): arginine, riboflavin, and EDTA (which is likely depleted due to complexing with metal ions rather than from direct uptake or metabolism). This result suggested that most compounds were included in

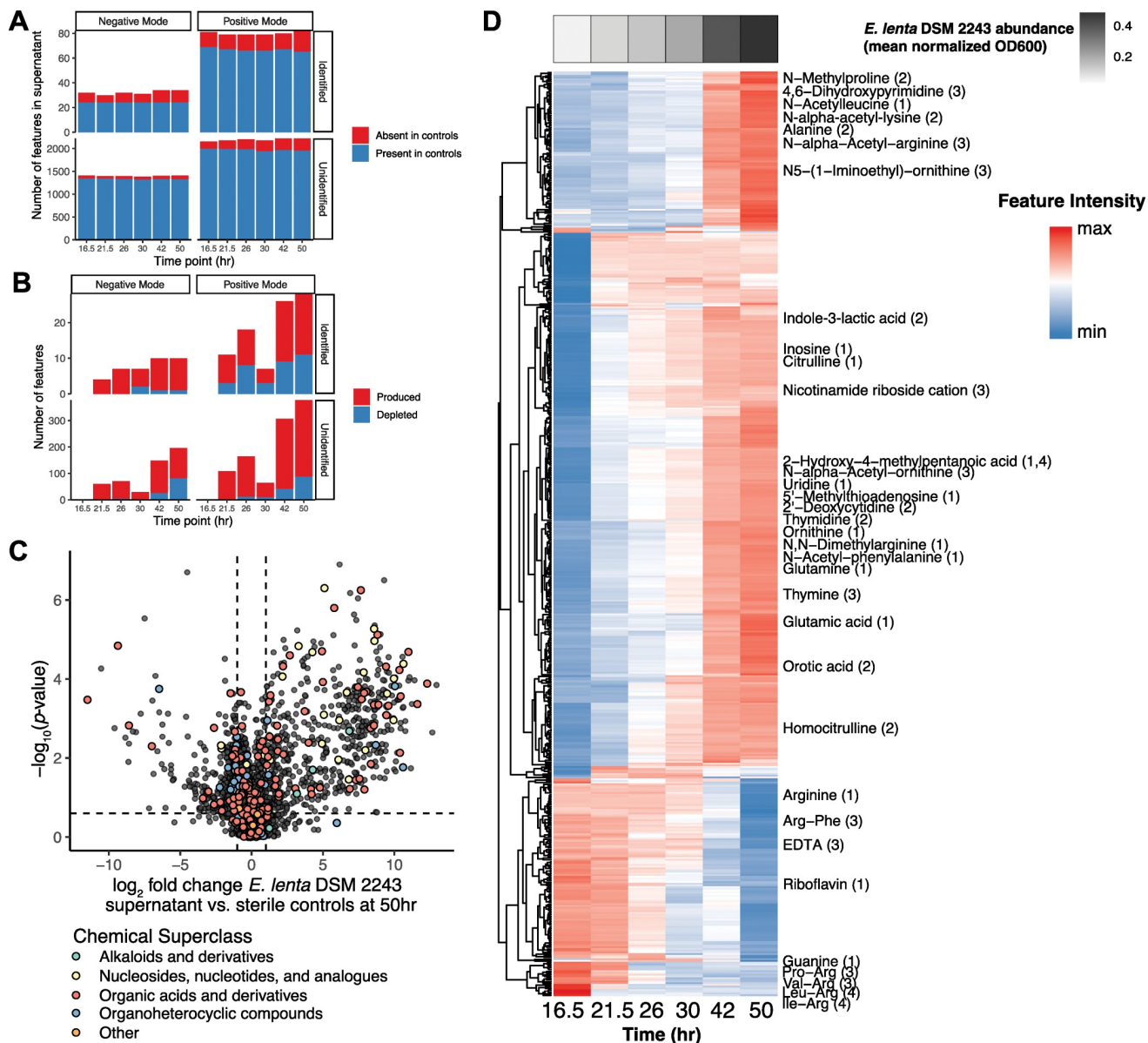


Fig 1. Production and depletion of diverse metabolites by *Eggerthella lenta* DSM 2243 in chemically defined media. (A) Number of metabolite features detected by tandem LC-MS in culture samples at each time point. Features are considered present if their average peak height in supernatant is greater than $3\times$ the average peak height in blank samples. Using both positive and negative ionization modes, an increasing number of features not found in controls appear in culture supernatants over time. (B) Number of differentially abundant metabolite features compared with sterile control media at each time point, based on FDR-adjusted t tests of log-transformed peak heights. (C) Volcano plot of differentially abundant metabolite features at the final time point (50 hours) compared with sterile controls. P values shown on the y axis are based on Welch's t tests comparing values at the final time point vs. sterile controls (Benjamini-Hochberg adjusted). (D) Heatmap of individual metabolite trajectories in cultures of *E. lenta* DSM 2243 grown in EDM1 batch culture. Features shown are those whose abundance was significantly different from controls (FDR-adjusted $p < 0.1$ and absolute \log_2 fold change > 0.75) at the final time point. Identified metabolites are labeled; the number in parentheses indicates the MSI confidence level for that identification (with 1 as highest confidence, see [Materials and methods](#)). Values shown are average log-transformed peak heights, scaled for each feature. The gray heatmap at the top indicates the average batch culture density at each time point of *E. lenta* DSM 2243 in EDM1 (normalized OD600). See also [S1 and S2 Figs](#) and [S1 and S2 Tables](#). EDM1, Eggerthella Defined Media 1; FDR, false discovery rate; LC-MS, liquid chromatography-mass spectrometry; MSI, Metabolomics Standards Initiative.

<https://doi.org/10.1371/journal.pbio.3002125.g001>

excess, leading us to reduce the concentration of several nondepleted amino acids for subsequent experiments (S1 Table). Interestingly, 5 of the identified metabolite features significantly depleted by *E. lenta* were not explicitly included in our defined media formulation, including guanine and 4 arginine dipeptides (Fig 1D). Since these compounds were found at low intensities and were annotated with high confidence, we inferred that they may be trace contaminants from commercial preparations of one or more EDM1 media components, which was partially supported by a preliminary follow-up analysis of fresh media (Materials and methods, S1E Fig). Regardless, the rapid depletion of these compounds indicates that their presence may influence growth and metabolic activity and reinforces the value of untargeted metabolomic profiling.

We examined the dynamics of metabolite production and depletion over the 50-hour growth of *E. lenta* in batch culture. Hierarchical clustering of metabolite trajectories indicated that among both produced and depleted features, some metabolites are produced/depleted rapidly early in growth, while others shift more dramatically later as the culture approaches stationary phase (Figs 1D and S1F). This observation suggests that 2 or more distinct growth phases may be occurring as resources are consumed from the media. Among identified metabolites, the trace guanine and arginine dipeptides are first depleted from the culture in early time points, while citrulline, inosine, and indole-3-lactic acid are produced at relatively higher rates (Fig 1D). In the later phase, arginine is depleted more rapidly, while alanine, 4,6-dihydroxypyrimidine, and various *N*-acetylated amino acid metabolites are produced.

To gain a better understanding of the contributions of individual nutrients to *E. lenta* growth, we systematically tested the effect of their removal from the media on growth of *E. lenta* DSM 2243 (Materials and methods, S2 Table). We collected growth curve data from EDM1 with and without each component and fit logistic growth models to the results, finding that 21 out of 41 compounds tested had a significant effect on the area under the growth curve (Wilcoxon rank-sum test, FDR-adjusted $p < 0.2$; S2A Fig). The only compounds whose individual removal fully prevented growth of *E. lenta* were arginine, tryptophan, riboflavin, biotin, and magnesium (although it is plausible that other compounds are required in trace amounts and were not fully removed by our preparation methods, particularly minerals such as iron). In general, removing amino acids most commonly tended to reduce carrying capacity, consistent with a role as carbon and/or energy sources, while removing vitamins had more varied effects on the growth curve (S2B Fig).

Acetate and arginine are key carbon and energy sources for *E. lenta*

Surprisingly, we found that sodium acetate contributed substantially to *E. lenta* growth in EDM1 (S2A Fig), even though it was included at a relatively low concentration (1 mM, compared to 57 mM arginine in EDM1). Since acetate is an abundant and variable metabolic byproduct of diverse human gut microbes [28], dependence on acetate could shape the ecological interactions of *E. lenta* in the human gut microbiota. Although our untargeted liquid chromatography–mass spectrometry (LC-MS) workflow was not able to quantify acetate, we had observed accumulation of several *N*-acetylated compounds in supernatant (Fig 1D), suggesting that the amount of acetate incorporated into core metabolic pathways may be relatively small. However, acetate provided a dose-dependent increase in carrying capacity for *E. lenta* up to a concentration of at least 10 mM in EDM1 (Fig 2A). We therefore used a targeted derivatization and LC-MS/MS method to quantify acetate levels in supernatants from 3 strains of *E. lenta* (DSM 2243, AB8n2, and Valencia) grown in EDM1 with different acetate concentrations (0, 1, or 10 mM). Acetate was depleted to approximately the limit of quantification in cultures from the 1 mM acetate group, but not the 10 mM acetate group, confirming that a relatively

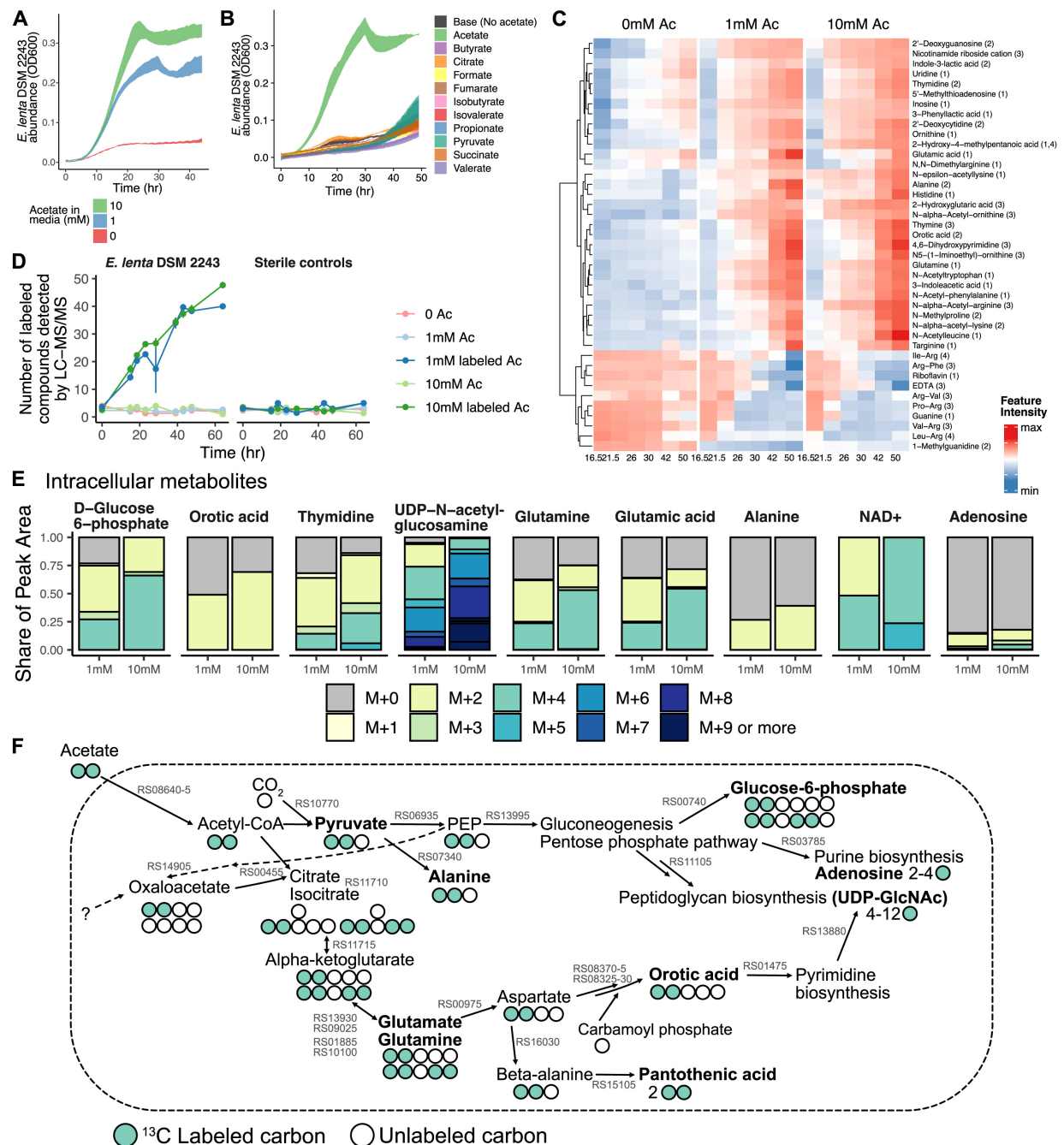


Fig 2. *E. lenta* uses acetate for nucleotide and peptidoglycan biosynthesis. (A) Growth of *E. lenta* DSM 2243 in EDM1 media with varying concentrations of sodium acetate. (B) Growth of *E. lenta* DSM 2243 in EDM1 media in which 1 mM sodium acetate is replaced by other small carbon compounds. (C) Trajectories of identified metabolite features responsive to acetate concentration in *E. lenta* EDM1 batch cultures. Values are scaled average log-transformed peak heights from untargeted metabolomics profiling of supernatants. Labels show metabolite identity and MSI confidence level in parentheses. Metabolites shown are those that were assigned an identity and that had significantly different trajectories in the 0 mM vs. 1 mM acetate group based on spline regression comparison with the R package *santaR* (FDR-adjusted $p < 0.25$). (D) SIRM profiling of *E. lenta* DSM 2243 in EDM1 media with $^{13}\text{C}_2$ labeled acetate. The number of compounds with labeled isotopologues detected at a peak area $> 10^5$ is shown for each sample group and time point, indicating incorporation of acetate into varied metabolites by *E. lenta*. (E) MIDs of intracellular metabolites. Each barplot shows the average isotopologue distribution in 1 mM and 10 mM acetate cultures. Compounds shown are those with an average labeled MID > 0.15 and a total peak area from labeled isotopologues of at least 10^4 in at least 1 *E. lenta* DSM 2243 labeled acetate condition. (F) Hypothesized pathways for incorporation of acetate into *E. lenta* central carbon metabolism and into biosynthetic pathways to produce labeled metabolites. Circles indicate the number of carbon atoms in selected compounds and are colored green to indicate incorporation of ^{13}C isotopes from external acetate. Compound names in bold were detected with the observed labeling patterns in either

intracellular metabolite extracts or culture supernatants. Corresponding enzymes are annotated in the *E. lenta* DSM 2243 genome for all reactions shown and labeled in gray with the NCBI locus tag number. For pathways shown at a summary level (gluconeogenesis, pentose phosphate pathway, purine and pyrimidine biosynthesis, peptidoglycan biosynthesis), only the first enzyme in the pathway is labeled on the plot. See also [S3–S5 Figs and S1 Data](#). EDM1, *Eggerthella* Defined Media 1; FDR, false discovery rate; MID, mass isotopologue distribution; MSI, Metabolomics Standards Initiative; SIRM, stable isotope-resolved metabolomics.

<https://doi.org/10.1371/journal.pbio.3002125.g002>

small quantity is required for the observed level of *E. lenta* growth ([S3A Fig](#)). We tested the effect of replacing acetate with equimolar amounts of 10 other small carbon compounds, finding that no tested alternative compound provided a comparable benefit ([Fig 2B](#)). Based on these results, we chose to further investigate *E. lenta*'s acetate utilization pathways.

First, we used our untargeted LC-MS metabolomics workflow to compare metabolites in supernatant over time from the same 3 *E. lenta* strains grown in EDM1 with different acetate concentrations (*E. lenta* DSM 2243 shown in [Fig 2C](#); AB8n2 and Valencia in [S3B and S3C Fig](#)). Using smoothing spline models, we found that many produced or depleted compounds had significantly different abundance trajectories across the growth phase (FDR-adjusted $p < 0.25$) depending on the presence of acetate. These included pyrimidine metabolites, *N*-acetylated amino acids, amino acid metabolites including indole-3-lactic acid and 2-hydroxy-glutaric acid, and 423 unidentified metabolite features ([Fig 2C](#)). Of the 612 features produced by *E. lenta*, 53.4% had significantly different trajectories in the no acetate condition. Most differentially abundant compounds were associated with cell density and produced by *E. lenta* at higher levels when grown with higher acetate concentrations, reinforcing the general loss of biomass production in the absence of acetate.

To identify the specific pathways by which acetate is metabolized by *E. lenta*, we next profiled metabolites in the supernatant across time during growth of the same 3 strain isolates of *E. lenta* with $^{13}\text{C}_2$ acetate provided as a stable isotope-labeled substrate (DSM 2243 in [Fig 2D–2F](#); 2 additional strains in [S4 Fig](#)). We detected the incorporation of ^{13}C labeled atoms in 52 features in *E. lenta* supernatants from any strain at the final time point, of which 24 were previously identified as responsive to acetate concentrations (Materials and methods, [Figs 2D and S4](#)). Acetate was incorporated into diverse products across metabolite classes but was found at the highest enrichment levels in nucleotide and carbohydrate metabolites ([S1 Data](#)).

Because many core metabolites are not produced in excess or secreted during growth, we also analyzed intracellular metabolites from extracts collected at a single time point in the late-exponential growth phase. Labeled intracellular compounds included glutamate, glutamine, sugars, nucleotide metabolites, and UDP-*N*-acetyl-glucosamine, a primary component of peptidoglycan ([Fig 2E](#)), as well as 7 labeled compounds of unknown identity. The signal from carbohydrate-related compounds including glucose-6-phosphate and UDP-*N*-acetyl-glucosamine was almost exclusively from labeled isotopologues (97.5% in 1 mM acetate and 100% in 10 mM acetate), indicating that synthesis of these compounds using acetate may be more efficient than any alternative non-acetate-dependent pathways available to *E. lenta* in the EDM1 condition.

Acetate-derived extracellular and intracellular metabolites were consistent across the 2 additional strains of *E. lenta*. While the overall rate of acetate incorporation differed between the 3 strains, the set of extracellular and intracellular labeled compounds was fully consistent. Isotopic enrichment for 2 additional extracellular metabolites (malonic acid and 3-hydroxymyristic acid) was identified in both of these strains as well as 4 additional intracellular metabolites in one or both strains (all of unknown identity), confirming that acetate is incorporated by *E. lenta* into varied biosynthetic pathways ([S4 Fig](#)).

Based on these results and metabolic gene annotations of the *E. lenta* DSM 2243 genome, we hypothesized that *E. lenta* converts acetate to acetyl-CoA via acetate kinase

(ELEN_RS08645) and phosphate acetyltransferase (ELEN_RS08640). Acetyl-CoA could then be used as a carbon source via 2 routes: conversion to glutamate by a partial citric acid cycle, and synthesis of pyruvate by the enzyme pyruvate-ferredoxin oxidoreductase (PFOR; ELEN_RS10770) (Fig 2F). This hypothesis is consistent with the organization of the *E. lenta* DSM 2243 genome, as 2 of the 3 enzymes required for conversion of acetyl-CoA to glutamate are colocated (aconitate hydratase and isocitrate dehydrogenase, ELEN_RS11710, ELEN_RS11715). Genes for another partial component of the citric acid cycle—fumarate hydratase and malate dehydrogenase—are colocated in another region of the genome ELEN_RS056[70-90], suggesting that they may act in a separate functional role. Taken together, these data suggest that *E. lenta* uses acetate as a key carbon source for synthesis of biomass components, in tandem with ATP generation from arginine catabolism, anaerobic respiration, and/or other unknown pathways.

However, we inferred that acetate is likely not the sole carbon source used by *E. lenta* in EDM1, given the relatively low concentration required for growth promotion and the abundance of unlabeled isotopologues detected for many produced compounds (S1 Data). We wondered whether arginine or ornithine may also be substrates for synthesis of biomass components, or if arginine is exclusively catabolized to ornithine for ATP production, as suggested by one previous study in rich media [29]. We first confirmed that citrulline, but not ornithine, can replace arginine with nearly equivalent growth in EDM1, replicating a previous result in rich media ([16]; S5A Fig). We then analyzed intracellular and extracellular metabolites from *E. lenta* DSM 2243 growing in EDM1, this time with $^{13}\text{C}_6$ L-arginine as a stable isotope-labeled substrate. We found by far the largest composition of ^{13}C enriched isotopologues in ornithine, citrulline, and other closely related compounds (S5B–S5E Fig), indicating that arginine is predominately processed by the arginine deiminase pathway. However, we observed M+1 enrichment (i.e., incorporation of a single ^{13}C carbon atom from arginine) in produced glutamine, orotic acid, and pyrimidines, among others (S5C and S5D Fig), suggesting biosynthesis from the carbamoyl phosphate intermediate. Labeled M+5 isotopologues of proline and prolina-mide also appeared at low levels at later time points, likely indicating a slower flux producing these compounds from accumulated ornithine (S5D and S5E Fig). Yet in total, only 29/324 features were detected with ^{13}C enrichment for 5 or more carbon atoms in intracellular extracts, and most appeared closely related to arginine, citrulline, and ornithine (S1 Data). These results confirm that arginine is primarily an energy source and not a major biosynthetic precursor for *E. lenta* (S5F Fig).

A genome-scale metabolic model of the *E. lenta* type strain recapitulates growth, metabolite, and gene expression phenotypes

COntstraint-based Reconstruction and Analysis (COBRA) is a set of computational tools that has been applied to interpret -omics data and optimize metabolic activities for various microbes of importance in basic science, metabolic engineering, and medicine [30–32]. It has been proposed as a promising strategy to predict phenotypes and design modifications to complex host-associated microbial communities by synthesizing information about the physiology of individual members and the available nutrients into a rational framework [8,33,34]. However, the value of such a framework is dependent on its ability to accurately describe the contributions of metabolically active community members. The reconstructions currently available for many anaerobic microbes have only been curated to a limited degree and remain minimally validated. Therefore, we used our in vitro platform to curate and analyze a genome-scale metabolic network model of *E. lenta* DSM 2243 growth in EDM1 and assessed the degree to which this model can explain *E. lenta* metabolic phenotypes across conditions.

We obtained a genome-scale metabolic reconstruction from the AGORA database version 2.0.0 [35], which we term iEL2243_2. Initial testing indicated that the model was incapable of biomass production in EDM1 media, so we performed additional curation of model reactions and transporters (S3 Table). We curated the reconstruction based on genome annotations from multiple sources [36–38] and added transporters for strongly depleted and produced compounds that were identified with high confidence in our metabolomics data. Throughout this process, we compared model results with experimentally observed growth in chemically defined media conditions, using these results to inform the curation process and add missing reactions where supported by experimental data. We simulated metabolic fluxes in different conditions by converting defined media concentrations into estimated maximum nutrient uptake rates for each compound, assuming a fixed and evenly mixed concentration of *E. lenta* cells (Materials and methods). Nutrient constraints for other metabolites detected in EDM1 media by untargeted metabolomics were applied qualitatively following previous work (Materials and methods; [39]). While these models are typically validated by comparison with gene essentiality data [40], the tools to generate such data are not yet available for *E. lenta*. We instead evaluated whether the model was consistent with observed metabolite utilization and production and with gene expression during exponential growth in EDM1, and whether predicted essential genes were conserved across strain genomes. We did not use metabolite production or transcriptomic data to refine the model so that we could instead use these datasets to independently evaluate the effectiveness of the growth-based curation.

This process resulted in a model with 1,244 reactions linked to 727 gene annotations and 1,218 metabolites (Fig 3A). The largest number of reactions were in the subsystems of fatty acid synthesis, extracellular transport, and glycerophospholipid biosynthesis (Fig 3B). Flux balance analysis of the final model estimated the maximum growth rate of *E. lenta* DSM 2243 in EDM1 to be 0.96 h^{-1} , higher than experimental values (median 0.32 h^{-1} ; S2B Fig). The difference between these values is not surprising given that bacteria do not necessarily grow at their theoretical maximum growth rate, and growth constraints may exist that are not encoded in the metabolic network model [40]. However, the relatively large discrepancy indicates that additional modifications to the biomass equation may further improve the model. In particular, we found that the maximum growth rate was sensitive to the biomass requirements for flavin compounds and for ATP (S6 Fig).

The initial model with nonzero growth in EDM1 did not recapitulate the experimentally observed dependencies on either arginine or acetate (Fig 3C). We noticed that this lack of dependency was linked to the inclusion of Wood–Ljungdahl acetogenesis reactions in the model, previously suggested to be present in *E. lenta* [20,22]. The presence of these reactions allowed the model to draw on an effectively unlimited source of acetyl-CoA from CO_2 and H_2 . Regardless of whether the previous annotation of this pathway (which has not been biochemically validated) is correct, reductive acetogenesis may not be thermodynamically favorable during in vitro growth in our anaerobic chamber, where the H_2 concentration is $\leq 5\%$ [41]. Blocking model flux through the carbon monoxide dehydrogenase reaction of this pathway increased growth dependency on uptake of both arginine and acetate, reflecting our experimental observations (Fig 3C). The model also found no growth benefit from pyruvate, citrate, and other fatty acids based on a lack of annotated transporters for these compounds, consistent with experimental results.

In another key curation step, required to enable biomass production by the model in EDM1, we noticed that *E. lenta* lacks an annotated gene for the enzyme enoyl-acyl protein carrier reductase, which performs the elongation in the typical type 2 fatty acid synthesis pathway used in bacteria. Because fatty acid biosynthesis is essential and previous studies have noted a high level of diversity in this essential step among bacterial genomes [42], we preserved this

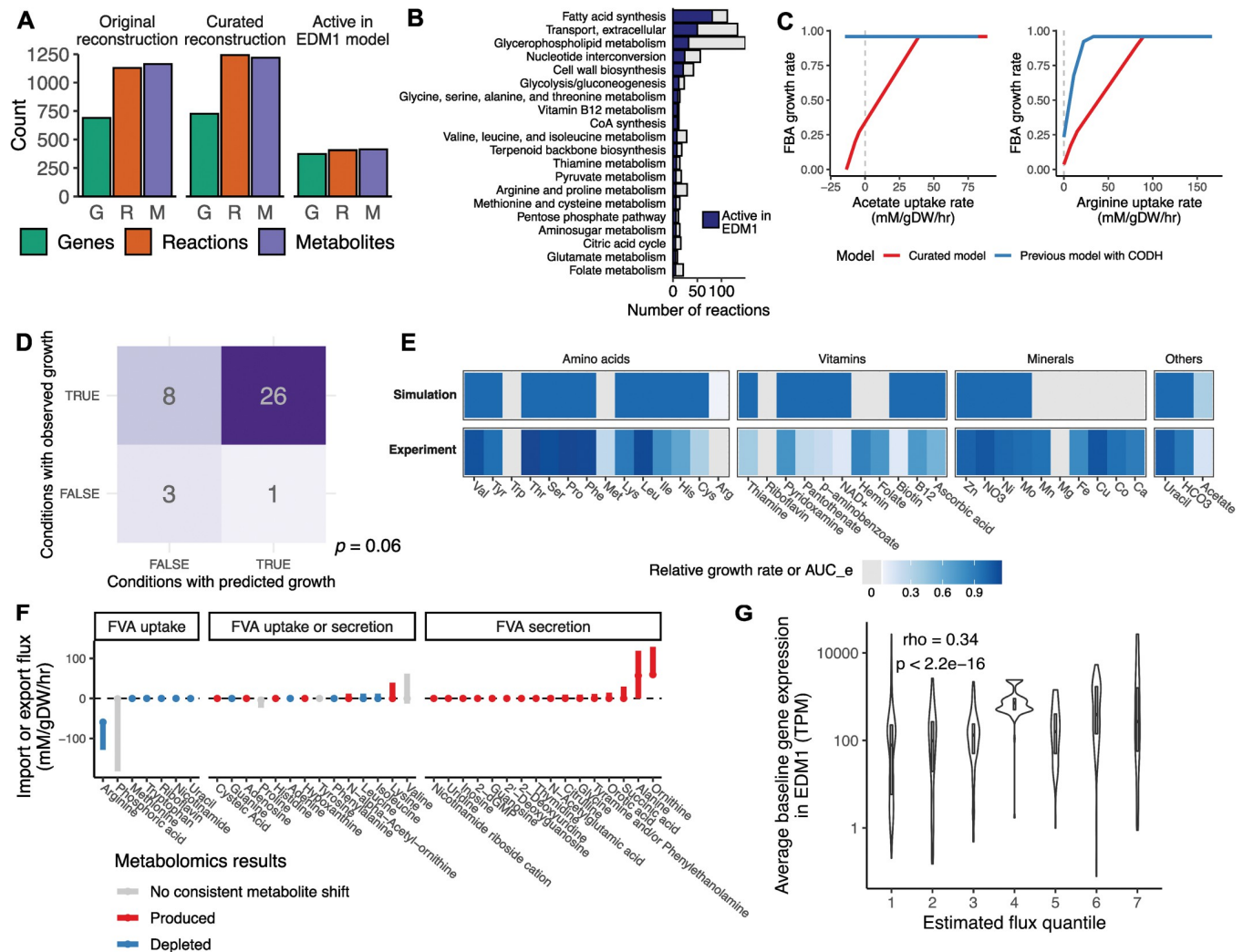


Fig 3. A curated genome-scale metabolic model of *E. lenta* DSM 2243 partly explains growth phenotypes across conditions. (A) Summary of the curated reconstruction of *E. lenta* DSM 2243 indicating the number of genes, reactions, and metabolites in the original and curated models, and the share of those required to be active for growth in EDM1 based on pFBA. (B) Summary of the total number of reactions by subsystem, and the share of each subsystem predicted to be active in EDM1 (only the top 20 subsystems are shown). (C) Acetate and L-arginine uptake dependencies inferred by the model. In the final curated model (red lines), the maximum growth rate decreases with decreasing availability of both L-arginine and acetate, qualitatively consistent with experimental data. A previous model incorporating a carbon monoxide dehydrogenase reaction based on [20] (blue lines) failed to recapitulate the expected dependencies. (D) Confusion matrix summarizing a comparison of growth/no growth between the iEL2243_2 model vs. experimental observations for leave-one-out media conditions. The labeled *p*-value is based on a Fisher exact test. (E) Full set of quantitative comparisons underlying panel (D). Each column shows the FBA-inferred maximum growth rate in the EDM1 condition with a media component removed, paired with the experimentally observed area under the empirical growth curve (AUC_e) for that condition. A gray tile indicates zero growth. (F) Comparison of shifts in metabolomics data with uptake and secretion rate ranges inferred for the same compounds by FVA. Metabolites that can only be imported according to FVA were decreased in metabolomics data, while those with potential for being produced were indeed produced. (G) Comparison of absolute fluxes inferred by pFBA with gene expression of linked enzymes of *E. lenta* DSM 2243 during exponential growth in EDM1. Within flux quantiles (on the x-axis), genes are expressed at a wide range of levels, but genes linked to reactions with the highest fluxes are generally highly expressed. See also [S6 and S7 Figs](#) and [S3 and S4 Tables](#). EDM1, *Escherichia* Defined Media 1; FBA, flux balance analysis; FVA, flux variability analysis; pFBA, parsimonious flux balance analysis.

<https://doi.org/10.1371/journal.pbio.3002125.g003>

step in the model without any current gene annotation. This gap may indicate a novel enzyme family performing this conversion (**S3 Table**).

We applied the iEL2243_2 model to predict growth phenotypes across our leave-one-out chemically defined media conditions, finding that these were generally consistent with some remaining notable exceptions (Fig 3D; overall Matthews correlation of 0.35, Fisher exact test

odds ratio = 9.1, $p = 0.06$). Amino acid dependencies matched well between the model and experimental data, with the exception of cysteine, which likely provides a benefit as a reducing agent that is not accounted for by the model [43]. Vitamin dependencies were also generally consistent, with the notable exception of folate, which had no effect on growth despite the lack of several genes for reactions in the canonical folate biosynthesis pathway and the absence of a known dihydrofolate reductase enzyme [44]. The phenomenon of presumed-essential but absent folate genes in bacterial genomes has been recognized previously [44–46], suggesting the possible existence of undiscovered alternative enzymes. Notably, growth was negatively affected by the removal of the folate precursor *p*-aminobenzoate (S2A Fig). Most of the remaining discrepancies between the model and the growth data are in conditions in which metal ions were removed, which were expected to be required by the model (Cu^{2+} , Ca^{2+}) but were not essential based on our experiments (Fig 3E). However, these likely reflect difficulties in fully removing trace minerals in our experiment rather than errors in the model reconstruction.

While we curated the model based on growth data, we did not incorporate our metabolomics data except to add transporters for highly differentially abundant metabolites. Even so, we found that there was a high correspondence between observed metabolite shifts and the possible uptake and secretion fluxes inferred by flux variability analysis (FVA) of the model. FVA identifies the range of fluxes for each reaction that are compatible with near-maximum growth. All 37 identified metabolites present in both the model and our metabolomics data displayed experimental shifts in abundance qualitatively compatible with inferred flux ranges (Fig 3F), providing additional support for model quality.

We further compared the iEL2243_2 inferred flux profile with RNA-Seq data from *E. lenta* growing in this condition, which was not used for model curation (Materials and methods). Genes linked to active reactions were 3.4 times more likely to be in the top half of metabolic genes by expression level than those linked to inactive reactions (71.9% of active versus 42.9% inactive, $p < 2.2 \times 10^{-16}$). Expression level and absolute flux magnitude were correlated across all genes linked to metabolic reactions (Spearman $\rho = 0.34$, $p < 2.2 \times 10^{-16}$; Fig 3G). While we would not expect a perfect correlation between expression and metabolic flux, correspondence between the two provides support that our model has correctly identified pathways with high activity.

Having established consistency with experimental data, we next examined overall reaction fluxes and key pathways in the final model. We found that fewer than half of reactions were predicted to be active in EDM1 by parsimonious flux balance analysis (pFBA; Fig 3A). In the pFBA solution, acetate is incorporated into a partial reductive citric acid cycle via pyruvate formate oxidoreductase (PFOR), which then feeds lipid and carbohydrate biosynthesis pathways, consistent with our SIRM results and with our RNA-Seq data, where PFOR was one of the most highly expressed genes. The vast majority (99.6%) of arginine uptake flux was directed to ATP generation, and 58.4% of ATP generation was sourced from the arginine deiminase pathway (which contains the first, third, fourth, and fifth most highly expressed protein-coding genes in our RNA-Seq data; S4 Table). The remainder of ATP generation in the pFBA solution was attributed to anaerobic respiration via an ATP synthase reaction, although the specific electron transport chain substrates were not clear. However, consistent with this hypothesis, genes linked to respiration were expressed at moderate levels, including ATP synthase subunits and an Rnf electron transport complex, and *E. lenta* is known to have a large number of poorly characterized enzymes potentially involved in electron transfer [21,47]. The model also identified the regeneration of NADP⁺ via transaminase reactions (using mainly pyruvate and/or branched chain amino acids) and glutamate dehydrogenase as a key high-flux pathway.

Finally, we applied the model to predict the effects of knocking out individual reactions on growth of *E. lenta*. We found that 15.3% of all reactions in iEL2243_2 were predicted to be essential in any condition and 19.4% to be essential in EDM1. These reactions tended to be involved in lipid metabolism, cell wall biosynthesis, and transport of essential metabolites (S7A Fig). Genes linked to reactions whose removal reduced growth to <70% of wild-type levels were found in a greater number of *E. lenta* strain genomes than other genes (Wilcoxon rank-sum test, $p = 0.001$; S7B Fig) and were more likely to be part of the core genome (found in all strains; Fisher exact test odds ratio = 1.74, $p = 0.0002$). Overall, while significant manual curation was required for the model to recapitulate realistic growth in EDM1, our updated model is able to predict and interpret many aspects of *E. lenta* growth and metabolic activity across conditions.

The strain-variable *E. lenta* metabolome is enriched for nucleotides and cell wall metabolites and can be linked to genome variation

Our initial efforts to characterize *E. lenta* core metabolism focused mainly on the type strain. However, *E. lenta* has an open pan-genome and established variability in secondary and xenobiotic metabolism [26]. We therefore evaluated the extent to which the metabolic profile of this species is conserved across a larger number of strain isolates. We used untargeted metabolomics to profile stationary phase supernatants of 30 strains grown in EDM1 (S5 Table, S8A and S8B Fig) and used linear models to identify features with significant strain-associated differences in abundance. Over half of the features produced by the UCSF stock of the DSM 2243 type strain (52.8%) were variable across strains of *E. lenta* (Fig 4A), and 1,097 features produced by at least 2 other strains were not produced by the type strain. Divergence in metabolite profiles between strains was not associated with phylogenetic divergence based on an alignment of core genes (Procrustes analysis, $p = 0.31$; Fig 4B), consistent with previous findings from untargeted metabolomics profiling of these strains in rich media with a different metabolomics platform [26]. Overall metabolite profiles were moderately associated with presence/absence patterns of variable gene families between strains ($p = 0.04$; Fig 4B), indicating that the presence or absence of biosynthetic genes and pathways only partly explains variation in the metabolome and that other factors like gene regulation and enzymatic activity may also play a substantial role. Interestingly, both metabolite and gene presence/absence profiles could be partly explained by the geographic region of origin of each strain (PERMANOVA; metabolites $R^2 = 0.22$, $p = 0.04$; genes $R^2 = 0.36$, $p = 0.02$; S8B Fig), suggesting a possible influence of geography on the distribution of accessory metabolism. Other metadata variables, including host health status and isolation body site, were not significantly associated with metabolite profiles.

While strain-variable metabolites were quite diverse, they were enriched for certain chemical classes. The vast majority (92.0%) of strain-variable metabolites had no identity information, a similar ratio to the total number of metabolite features (91.8% of features in the whole dataset). Among other features, organic acids (which included many amino acid metabolites) were the least likely to be strain-variably produced. In contrast, organic oxygen compounds (which included several features identified as sugars) and nucleotide metabolites were more likely to be strain-variably produced, and organic heterocyclic compounds and benzenoids were enriched for strain-variable depletion (Fig 4C). The share of strains producing any individual feature varied widely (Fig 4C), although the largest number of features (76.8%) were produced by either only a few (<4) strains or nearly all (>27) strains (S8C Fig).

Given the large share of unidentified metabolites in our dataset, we evaluated whether linking strain-variable metabolites with strain-variable genes could inform metabolite annotations.

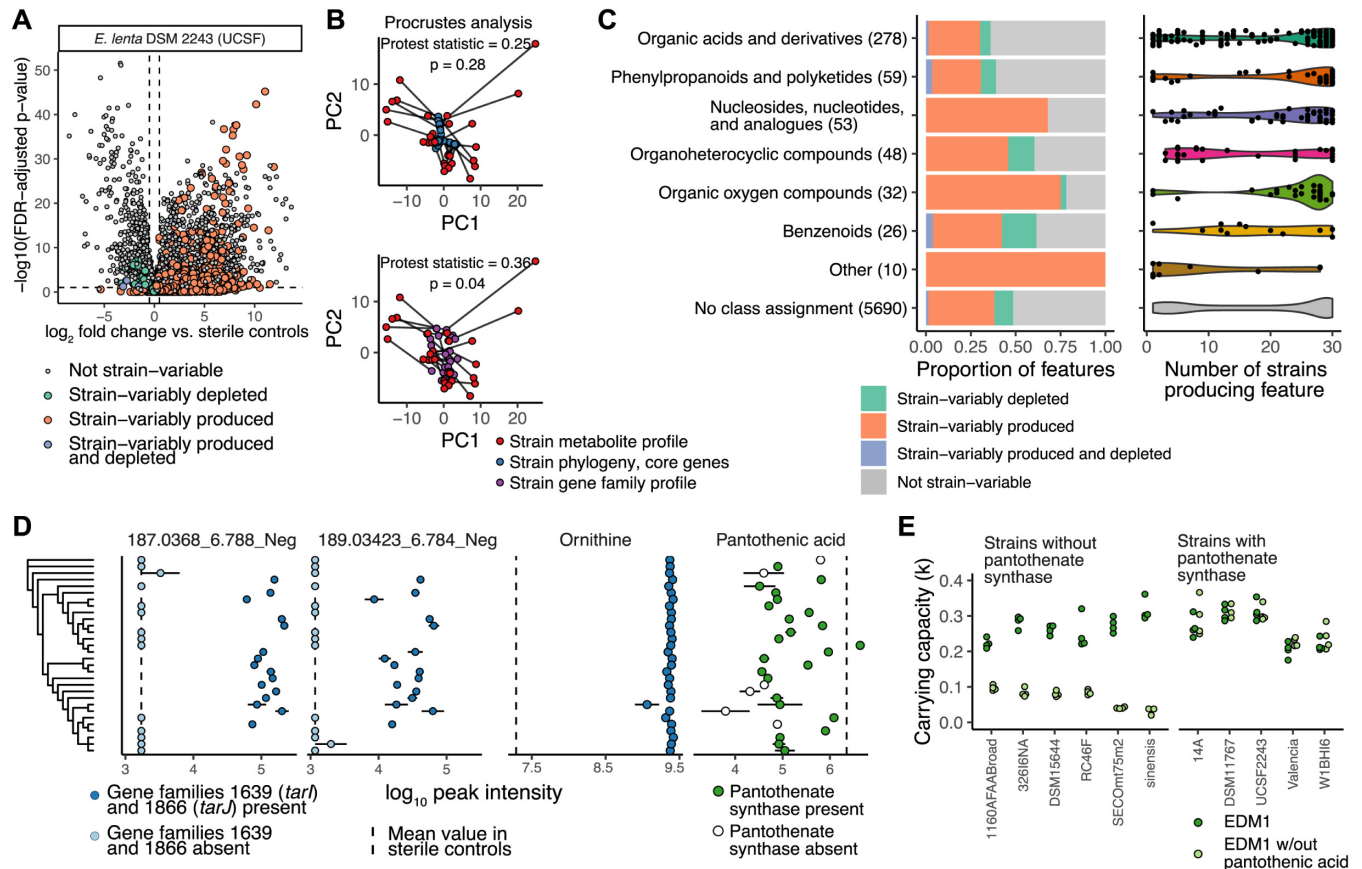


Fig 4. Extensive within-species variation in *E. lenta* metabolites can be linked to variable gene families. (A) Volcano plot of metabolite features detected in stationary phase supernatants of *E. lenta* DSM 2243 (UCSF lab stock) vs. sterile controls. P -values are based on Benjamini–Hochberg corrected Welch’s t tests. Features are colored based on whether their classification as significantly produced or depleted (increased or decreased based on linear regression models of peak area by strain, with FDR-adjusted p -value < 0.1 and $\log_2\text{FC} > 0.5$) is consistent across 28 other *E. lenta* isolates and 1 isolate of *Eggerthella sinensis* profiled in the same experiment. (B) Procrustes analysis of overall metabolite profiles compared with genome features. The upper plot shows a rotated Procrustes superimposition of average metabolite profiles for each isolate (red points) and the phylogenetic distance between them based on an alignment of core genes (blue points). The lower plot shows a superimposition of metabolite profiles and profiles based on the presence/absence of variable gene clusters (purple points). (C) The left-hand plot shows the distribution of strain-variable features in various ClassyFire chemical superclasses, based on Feature-based Molecular Networking with GNPS. The number in parentheses for each class indicates the total number of features with that assignment. The right-hand plot shows the number of strains producing a given feature within each superclass, among features produced by any *Eggerthella* isolate. Each point represents a single feature, and its position on the x-axis indicates the number of strains for which that feature was significantly increased (FDR-adjusted p -value < 0.1 and \log_2 fold change > 0.5) in supernatants compared with controls. (D) Feature abundances of example metabolites across strains. The first 2 panels show 2 strain-variable unidentified features associated with the presence of specific strain-variable gene families—putatively identified as the 2 dominant naturally occurring isotopes of an $[M+Cl]^-$ adduct of the teichoic acid component ribitol. The points indicate the log-transformed abundances of these features for each strain. The dotted line in each panel indicates the average level of that feature in sterile controls. Points in dark blue represent strains whose genomes contain genes for a ribitol-5-phosphate cytidyltransferase (*tarJ*) and ribulose-5-phosphate reductase (*tarI*) not found in other genomes. The third and fourth panels show a highly conserved identified metabolite (ornithine) compared with a strain-variable identified metabolite (pantothenic acid). Points in white in the pantothenic acid panel indicate strains whose genome lacks the final step in the biosynthetic pathway for this metabolite. Points are shown as mean and standard error across 3 replicates. The order of strains on the y-axis matches their phylogeny, shown in S8A Fig. (E) Strains lacking a gene annotated as pantothenate synthetase deplete pantothenic acid completely from media (previous panel) and have a substantial growth defect when grown in the absence of pantothenic acid (left plot). Closely related strains that do possess this gene are unaffected by the removal of pantothenic acid (right plot). Carrying capacity is estimated based on a logistic growth model fit by the R package *growthcurver*. See also S8 and S9 Figs and S5–S7 Tables.

<https://doi.org/10.1371/journal.pbio.3002125.g004>

We performed an association analysis between metabolite feature abundances and the presence of specific accessory gene families, applying a method developed for previous analysis of this *E. lenta* strain collection [26]. A full 39.0% of metabolite features were significantly associated with the presence of one or more variable gene families (FDR-adjusted $p < 10^{-4}$). Using stricter filtering criteria for significance, effect size, and separability, 84 metabolite features

(1.3%), of which 80 had no annotation, were linked with the presence of variable genes (S6 Table, Materials and methods). Gene families linked to these features were enriched for KEGG annotations in sulfur metabolism ($q = 0.00017$), ABC transporters ($q = 0.02$), porphyrin metabolism ($q = 0.03$), and biosynthesis of nucleotide sugars ($q = 0.049$), consistent with the profile of identified variable metabolites.

As a case study, we further examined 2 of the top hits from this analysis, 2 closely related but unidentified metabolite features highly associated with the presence of 2 adjacent gene families (Fig 4D). These gene families were annotated by Prokka [48] as ribulose-5-phosphate reductase 1 (*tarI*) and a ribitol-5-phosphate cytidylyltransferase (*tarJ*), which are essential enzymes in the biosynthesis of CDP-ribitol teichoic acid. Teichoic acids are an abundant component of the cell wall of gram-positive bacteria that can take multiple forms and can be synthesized with either CDP-glycerol or CDP-ribitol subunits [49–51]. Interestingly, the m/z values and retention times of the linked features were consistent with an annotation as the 2 dominant $[M+Cl]^-$ naturally occurring isotope adducts of a 5-carbon sugar alcohol—i.e., potentially ribitol, xylitol, or a related compound. The annotation of these features as a sugar alcohol was further supported by comparison with a ribitol standard spiked into sterile EDM1 (S8D Fig).

Further examination of the biosynthetic gene cluster in which these genes are located revealed extensive strain diversity, with 10 different gene arrangements across the 30 isolates (S8E Fig), suggesting recent positive selection possibly as a form of phage defense [52,53] or host immune interaction [54]. Most genomes have one or more genes with homology to *E. coli* *arnC* genes in this region, indicating that the products may be lipoteichoic acids anchored to the cell membrane rather than wall teichoic acids [49]. Among *E. lenta* genomes without *tarI* and *tarJ*, all except the type strain have *tarD* and *tagF* genes in the same region instead, which are involved in the synthesis of CDP-glycerol teichoic acid instead of CDP-ribitol (S8E Fig) and would be consistent with the absence of extracellular ribitol in those strains. Two other metabolite features were associated with the presence of other members of this gene cluster, possibly indicative of other strain-variable cell wall components (S6 Table). This example illustrates that comparative multiomics can be a powerful strategy to identify and begin to decipher the functional consequences of strain variation, even when metabolite identities are not confirmed.

In addition to the unbiased association analysis above, we also assessed whether strain variation in metabolites of known identity could be predicted based on relevant gene annotations. We created genome-scale metabolic reconstructions of a subset of strains included in this experiment ($n = 24$, using the DEMETER pipeline), curated them using a limited version of the process applied to the type strain (Materials and methods), and again predicted growth and reaction fluxes in EDM1 and in leave-one-out media conditions using flux balance analysis. Across the metabolic networks of *E. lenta* strains, most reactions were conserved, including arginine metabolism and central carbon metabolism (S7 Table). Tryptophan and riboflavin auxotrophies were also predicted to be conserved across strains. Variable reactions tended to be in the subsystems of transport, fatty acid biosynthesis, cell wall biosynthesis, and nucleotide interconversion (S9A Fig). Consistent with the central role of arginine metabolism, ornithine and citrulline levels in our metabolomics dataset were very consistent across strains. Ornithine was among the least variable metabolite features (Fig 4D), and one of the most correlated with biomass (as estimated by optical density, Spearman $\rho = 0.36$, FDR-adjusted $p = 0.1$).

While the predicted effects of most compounds on growth were similar or identical across strains (S9B Fig), we noticed a clear difference in pantothenic acid dependence, as a subset of strains were predicted to be unable to grow in its absence. These strains lack the final enzyme in the biosynthesis pathway for pantothenic acid, which is itself a precursor of coenzyme A.

Pantothenic acid was depleted to varying degrees in our metabolomics data, reaching the lowest levels in strains that lack pantothenic acid synthase (Fig 4D). Notably, M+2 isotopologues of pantothenic acid were also detected in supernatants from the acetate SIRM experiment, corroborating that at least 3 *E. lenta* strains synthesize this vitamin de novo (S1 Data and Figs 2F and S4B). We tested growth of pantothenate synthase-lacking strains in comparison with a subset of genetically similar strains in EDM1 with or without pantothenic acid, confirming that strains without this gene family had a greatly reduced carrying capacity in the absence of pantothenic acid (Fig 4E) and highlighting the ability of curated genome-scale models to predict phenotypic differences. Overall, our analysis of strain variation in metabolite profiles is consistent with a model in which *E. lenta*'s distinctive central carbon and energy metabolism is a core species trait, while more peripheral biosynthetic pathways including synthesis of cofactors and cell surface components can vary freely to adapt to specific microenvironments [55].

Comparison of *E. lenta*'s metabolic profile in vitro and in vivo identifies shared signatures and usage of a novel nutrient

Having characterized the metabolic profile of the *E. lenta* species in a simplified in vitro environment, we next asked how these findings compare with its metabolic activity in a host, and whether our in vitro platform could help identify metabolic processes performed by *E. lenta* within the gastrointestinal tract. We monocolonized germ-free (GF) mice with one of 3 strains of *E. lenta* by oral gavage, collected serum and intestinal contents after 2 weeks of colonization, and profiled metabolites using the same LC-MS/MS untargeted metabolomics workflow as above. We identified features that were significantly differentially abundant in *E. lenta*-colonized mice versus their GF counterparts using linear mixed models and compared these features with our in vitro metabolomics datasets, identifying metabolites consistently shifted by the presence of *E. lenta* across environments. After data processing, quality filtering, and dereplication, we obtained a dataset of 19,714 metabolite features from intestinal samples. Of these, 9.5% were significantly differentially abundant (FDR-adjusted $p < 0.1$) in response to colonization with at least 1 strain in at least 1 segment of the intestinal tract, indicating a substantial metabolic impact of *E. lenta* on the intestinal environment (S10A Fig). Interestingly, only 0.95% of features were significantly shifted in the ileum, compared with 6.3% in colon and 4.8% in the cecum. In a previous monocolonization experiment, *E. lenta* DSM 2243 colonized the ileum at roughly an order of magnitude lower abundance than the cecum and colon of GF mice [26]. Additionally, only 13 features (0.26%) were differentially abundant in serum in response to any of the 3 strains. Overall separability of metabolite profiles between GF and colonized was also highest in the cecum and colon (Fig 5A). These results indicate that *E. lenta*'s strongest metabolic effects are restricted to the lower intestinal tract.

We assessed the extent to which metabolite features produced by *E. lenta* in cell culture are detectably shifted by the presence of *E. lenta* in mice. To do so, we integrated our processed metabolomics datasets by linking metabolite features across datasets with highly similar m/z , retention time, and MS2 spectra (see Materials and methods). Based on this analysis, 37.2% of identified metabolite features in intestinal contents and 12.2% of features overall were also detected in vitro (S10B Fig). We compared the estimated \log_2 fold change of each linked feature in vitro with the corresponding shifts in vivo (full set in S2 Data; Fig 5B shows the comparison with the strain collection dataset in Fig 4; S10C Fig shows a comparison with the dataset in Fig 1). A total of 202 features significantly increased by the presence of *E. lenta* DSM 2243 in cecal contents were also increased in one of our EDM1 in vitro datasets, providing support that they are directly produced by *E. lenta* in vivo. These features represented 78.9% of the set that could be linked across datasets and 20.9% of the full set of *E. lenta* DSM 2243-increased

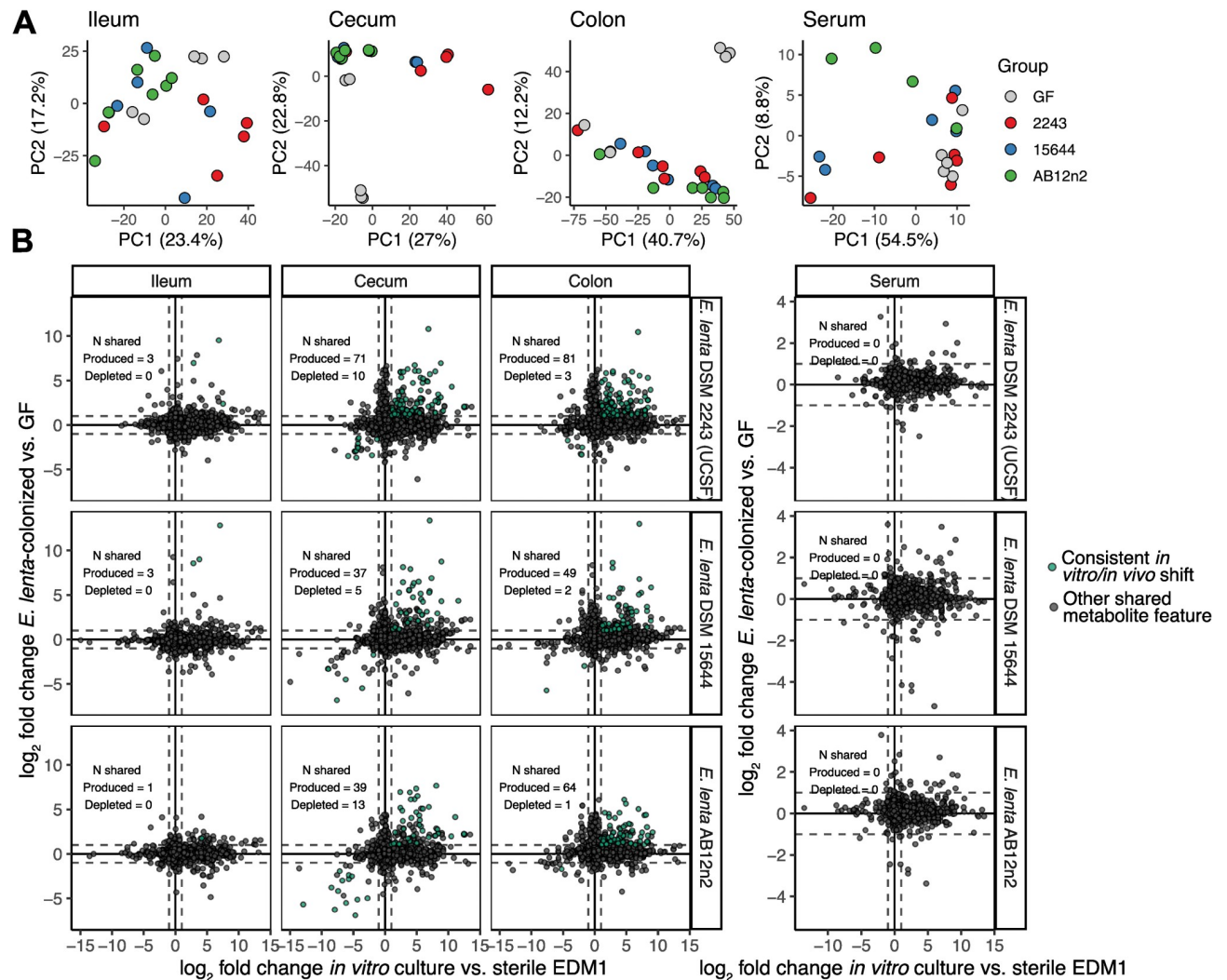


Fig 5. Comparison between *E. lenta*'s metabolic footprint in vivo and in vitro reveals shared metabolite signatures. (A) Principal component analysis of untargeted metabolomics profiles of intestinal contents and of serum. (B) Comparison of the effect of *E. lenta* on metabolite features detected in both EDM1 cultures and monocolonized mice. Each point represents a metabolite feature detected in both datasets. The x axis indicates the \log_2 fold change of each feature in supernatants compared with sterile controls, compared with the estimated \log_2 fold change of that feature in monocolonized mice compared with GF mice, based on linear mixed effects models after adjusting for covariates. Points are colored green if the feature is significantly differentially abundant in gnotobiotic mice and is shifted in the same direction by the corresponding strain in the stationary phase in vitro experiment. See also [S10 Fig](#) and [S2 Data](#). EDM1, *Eggerthella* Defined Media 1; GF, germ-free.

<https://doi.org/10.1371/journal.pbio.3002125.g005>

features in cecal contents. Only 18 metabolites depleted in cecal contents were similarly depleted in vitro, but only 3 of the other 405 depleted features were detected in vitro at all, indicating that *E. lenta* likely uses a much richer set of nutrients in vivo than those available in EDM1. Overlapping produced and depleted metabolites were found in greater abundance in the cecum and colon than the ileum and serum ([Fig 5B](#)), again suggesting a greater metabolic footprint of *E. lenta* in the lower gastrointestinal tract relative to other sites.

Ornithine was among the most increased features across sampling sites and strains, consistent with our in vitro data ([Figs 6A and S11A](#)). Other features that were consistently increased across strains included 5-methyluridine, citrulline, glutamine, and lysine ([S2 Data](#)). Interestingly, arginine was only significantly reduced by colonization with one of the 3 *E. lenta* strains in this experiment ([S11B Fig](#)). However, most other proteinogenic amino acids were increased

in abundance in intestinal contents in colonized mice compared with GF (S11C Fig), likely due to differences in host activity, so the absence of an increase in arginine may be consistent with arginine usage by *E. lenta*. Arginine levels were also more variable within strain colonization groups compared with GF (S11B Fig), even though levels of the primary products ornithine and citrulline were relatively consistent. This observation may reflect heterogeneity in host amino acid absorption and metabolism in response to *E. lenta* colonization.

Given these results, we evaluated what other substrates may be used as carbon or energy sources by *E. lenta* in vivo. The metabolites most strongly depleted by the presence of *E. lenta* DSM 2243 in the intestinal tract included several fatty acids conjugated with carnitine as well as multiple other nitrogen-containing metabolites: saccharopine and agmatine (Fig 6A). We chose to investigate agmatine usage further for several reasons: its chemical similarity to arginine, the presence of known agmatine utilization genes in the *E. lenta* type strain genome, evidence of a consistent decrease across all 3 strain colonization groups (S11A Fig), and its multiple roles as a microbial metabolite and a host metabolite involved in regulation of cell division and neural signaling [56]. The *E. lenta* DSM 2243 genome contains 2 complete and 2 partial operons encoding genes for the agmatine deiminase pathway. This pathway operates analogously to the arginine deiminase pathway, with ATP production via carbamate kinase as the final step (Fig 6B). Despite this similarity, the agmatine deiminase enzyme family is highly structurally distinct from arginine deiminase [57]. Presence of this pathway is conserved across strains, as other *E. lenta* genomes contain anywhere between 1 and 4 copies of the key genes for agmatine deiminase and putrescine carbamoyltransferase (KEGG; S8 Table). Additionally, a transcriptional regulator found adjacent to this operon in some strains was previously associated with *E. lenta* competitive fitness in vivo [26].

Based on these observations, we predicted that *E. lenta* may be able to grow in the absence of arginine if it is supplied with agmatine as an alternative energy source. A flux balance analysis simulation of *E. lenta* in agmatine-based EDM1 predicted a somewhat reduced maximum growth rate (0.54 versus 0.96 h⁻¹) in this condition, with arginine synthesized for protein via its annotated biosynthetic pathway from glutamate. Indeed, we found that replacing arginine with agmatine introduced a growth lag but resulted in a slightly higher final carrying capacity than the equivalent amount of arginine (Fig 6C). We additionally investigated agmatine-responsive genes using RNA-Seq. We grew *E. lenta* DSM 2243 in a formulation of EDM1 with 70% of the standard levels of arginine and acetate, treated cultures with either concentrated agmatine solution or water, and extracted RNA for sequencing. The genes most strongly induced by treatment with agmatine were 2 copies of putrescine carbamoyltransferase, 1 copy of agmatine deiminase, and a transporter in the same operon (Fig 6D and S9 Table). Genes in the second complete agmatine deiminase operon (ELEN_RS110[5-15]) were not differentially expressed, suggesting that the annotation of this second operon may be incorrect and/or may be involved in metabolism of a related compound. Interestingly, the most strongly down-regulated genes were 2 transport-related genes adjacent to the energy-conserving hydrogenase (Ech) complex (ELEN_RS078[45-50]), one of which has structural homology to the arginine-ornithine antiporter found in the arginine deiminase operon (ELEN_RS09745, 27.6% identity). These results indicate that *E. lenta* can generate ATP from agmatine as a distinct alternative to arginine both in vitro and in vivo and has extensive genetic machinery to efficiently and specifically use each of these compounds.

Discussion

In this study, we used custom growth media and untargeted metabolomics to profile the metabolism of a poorly understood gut microbe at a systems level. Although *E. lenta* is found

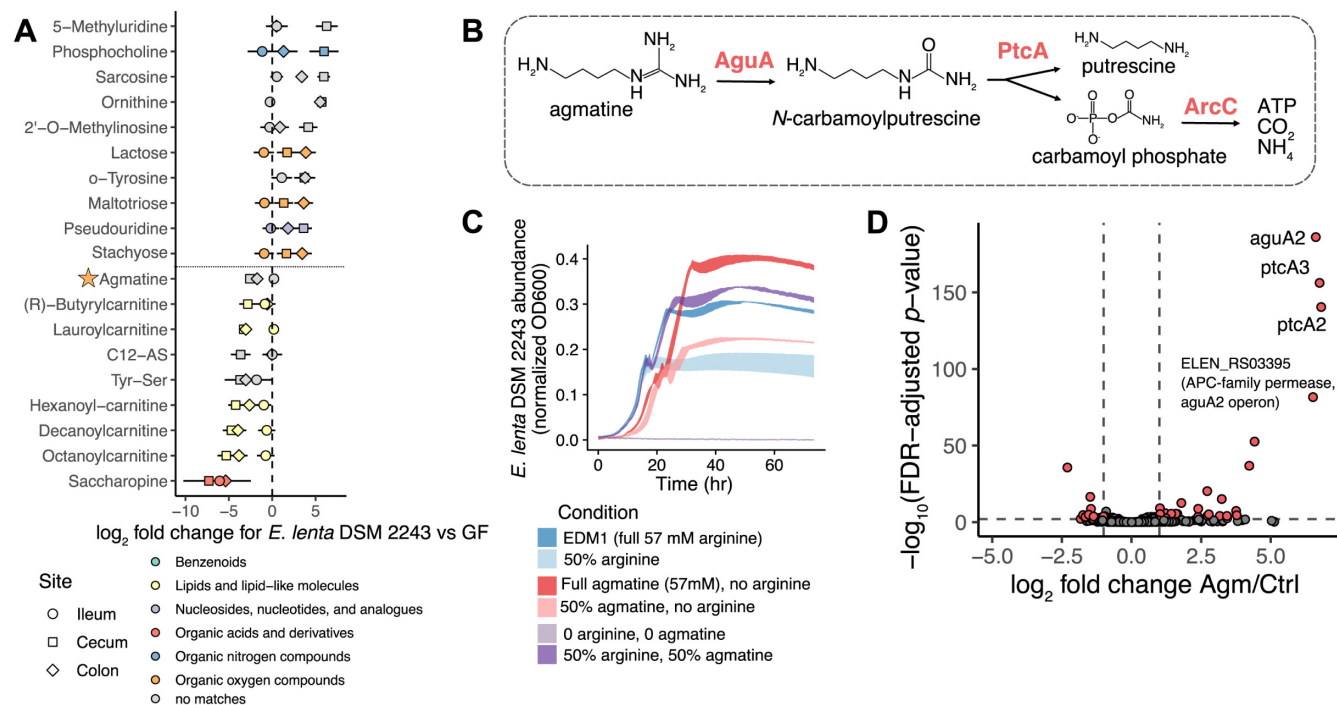


Fig 6. Agmatine can replace arginine as an energy source for *E. lenta*. (A) Identified metabolite features with the highest estimated effects in *E. lenta* DSM 2243-colonized mice compared with GF. Each point indicates the effect size of that feature in a particular sample site (denoted by shape). (B) Model of the agmatine deiminase ATP-generating pathway [57]. Multiple copies of an operon containing genes for all 3 of the labeled enzymes are annotated in the *E. lenta* DSM 2243 genome. (C) Growth of *E. lenta* DSM 2243 in EDM1 where arginine has been fully or partially replaced with agmatine sulfate. Curves show mean \pm standard error for 4 replicates. (D) Induction of the agmatine deiminase pathway in *E. lenta* DSM 2243 cultures in response to the addition of agmatine. The volcano plot shows the \log_2 fold change and FDR-adjusted *p*-values of agmatine-treated cultures compared to vehicle (as estimated by negative binomial differential abundance models with DESeq2). See also [S11 Fig](#) and [S8 and S9 Tables](#). EDM1, *Eggerthella* Defined Media 1; FDR, false discovery rate; GF, germ-free.

<https://doi.org/10.1371/journal.pbio.3002125.g006>

at >50% prevalence in gut microbiota of North American adults [10] and linked to acute and chronic disease [4], very little is known about its core metabolic properties. We documented an unusual set of carbon sources, nutrient dependencies, and secreted metabolites, and incorporated these into a genome-scale metabolic model that better recapitulated pathway activity and response to new environments. We further identified core and strain-variable properties across a large collection of strain isolates. Finally, we evaluated the extent to which these in vitro and in silico findings can inform our understanding of the host-associated in vivo metabolic activity of this organism. This broad strategy uncovered several specific new findings on *E. lenta*'s role in the gut microbial ecosystem and its potential effects on human hosts.

We first analyzed *E. lenta*'s metabolic footprint in a sensitive chemically defined environment using untargeted metabolomics. The extent and variety of compounds produced by *E. lenta* across multiple growth phases is consistent with previous experimental and theoretical work on “costless” metabolite secretions by diverse microbes [58–60]. In particular, many nucleotides and nucleic acid intermediates are synthesized by *E. lenta* and secreted without any apparent cost to growth. Secretion of these broadly useful metabolites may contribute to a previously observed outsized impact of *E. lenta* on the composition of synthetic communities [61]. Interestingly, several small molecules produced by *E. lenta* in EDM1 and in mice are known to impact host immune signaling, including indole-3-acetate [62] and inosine [63,64]. Notably, the relative level of production of these metabolites and others varied widely across *E.*

lenta strain isolates. Teichoic acids, identified here as another strain-variable feature, are also key targets of host innate immunity, with differential responses depending on their composition [54]. While much focus has been deservedly paid to individual specialized immunomodulatory transformations performed by *E. lenta* [4,65], our results suggest that *E. lenta*'s effects on host immunity may be multifaceted.

We elucidated the roles of 3 common gut metabolites in the metabolic network of *E. lenta*: arginine, acetate, and agmatine. First, we confirmed that conversion of arginine to ornithine is a core property of the *E. lenta* species. Production of ornithine was the most consistent metabolic feature across strains and environments. Our stable isotope analysis indicated that ornithine is primarily an end product of growth and is relatively inaccessible as a carbon source for *E. lenta*. However, ornithine is a favorable carbon and/or energy source for numerous other gut microbes [23], including as a substrate for Stickland metabolism by gut bacteria including *Clostridioides difficile* [6,66,67]. Therefore, production of ornithine by *E. lenta* may promote the growth of other proteolytic bacteria in the surrounding gut ecosystem.

We also found that the presence of acetate has a dramatic effect on *E. lenta* growth and metabolism in vitro. Acetate is a ubiquitous microbial metabolite in the mammalian gut that varies in concentration [28]. Previous studies have speculated that *E. lenta* may produce acetate via autotrophic acetogenesis [20,22]. While our study does not resolve the question of whether *E. lenta* has a functional acetogenic Wood–Ljungdahl pathway, we found that environmental acetate is an important biosynthetic precursor for *E. lenta*, incorporated partially via a distinctive bifurcated citric acid cycle [68,69]. If *E. lenta* is, in fact, an acetate consumer in vivo, as we have observed in vitro, this role may have ecological consequences. For example, *E. lenta* may compete for cross-fed acetate with other gut microbes, including the abundant, health-linked members of the Firmicutes that metabolize acetate to butyrate at high rates [70,71]. However, while we did not identify any compound that can replace the role of acetate in *E. lenta*'s metabolic network, the observation that *E. lenta* can grow to high carrying capacities in rich media and in GF mice presumably lacking acetate indicates that other undetermined compounds may be able to serve as equivalent carbon sources.

Finally, we identified agmatine as an alternate energy source for *E. lenta* in vivo. Agmatine is a host metabolite with multiple roles as a neurotransmitter, regulator of nitric oxide synthesis, and regulator and precursor of polyamine metabolism [56]. Although agmatine can be synthesized at low levels by the host, particularly in the brain, the gastrointestinal tract is thought to be a major source of systemic agmatine [72]—sourced either directly from the diet and/or from microbial metabolism. Dietary sources of agmatine include a variety of plant and animal products, with the highest known levels in fermented foods and alcoholic beverages [73]. Altered agmatine levels have been associated with a range of diseases, including depression and diabetes [56]. Notably, reduced agmatine levels in the gut have been linked to cell proliferation and cancer [74]. Therefore, depletion of gastrointestinal agmatine by gut microbes including *E. lenta* has the potential to impact host health and disease. Further work is needed to clarify the roles of both production and degradation by gut microbes in regulation of host agmatine metabolism.

Overall, our analysis of *E. lenta* nutrient dependencies revealed that this species occupies a metabolic niche that is distinct from canonically described roles in the gut ecosystem, such as primary and secondary carbohydrate degraders or conventional methanogens and acetogens. *E. lenta* relies heavily on ATP generation from arginine and/or agmatine catabolism, uses acetate as a key carbon source, and likely performs anaerobic respiration with unknown and potentially diverse substrates. The carbon and energy sources and auxotrophies that we identified were highly conserved across the *E. lenta* species, with the exception of pantothenate. Knowledge of these conserved metabolic dependencies may be an important tool in future

therapeutic attempts to engineer or modify *E. lenta* abundance, metabolic activity, and community interactions. In addition, the resources described here, together with the development of tools for genetic manipulation of *E. lenta*, may provide a basis for further investigation of the biochemical and physiological mechanisms underlying its distinctive metabolic strategy.

Another resource generated by this study is a curated constraint-based genome scale metabolic model of *E. lenta*. Constraint-based modeling is a promising approach for predicting community interactions and ecosystem engineering [75], but to date, community metabolic modeling tools have been difficult to validate and have generated relatively limited insights beyond what could be obtained with simpler annotation methods. Our analysis highlights the importance of phenotype-based curation of individual reconstructions. Specifically, the initial semi-curated AGORA model of *E. lenta* did not support any growth in EDM1 and lacked a complete version of the agmatine deiminase pathway. Yet, analysis of the more fully curated reconstruction enabled us to confirm key reactions during growth with arginine and agmatine in vitro, identify gaps representing potential novel enzymes, and uncover strain differences in vitamin dependence. These results suggest that the quality and predictive power of community metabolic models of the gut microbiota could be greatly improved by systematic data generation and refinement of reconstructions for a metabolically diverse sample of common taxa. Comparisons with growth in defined media conditions, -omics data, and strain conservation can assist with model validation even when genetic tools are not available.

Even so, the growth rate in EDM1 predicted by iEL2243_2 was still substantially higher than observed experimentally, suggesting that our reconstructions of the *E. lenta*-type strain and other strains may be further improved by additional curation. This curation could include quantitatively integrating transcriptomic and metabolomic datasets and/or biochemical characterization and curation of pathways and dependencies that were not fully explored in our study, including synthesis pathways, energy requirements, and demand for multiple vitamins and cofactors. Notably, our finding that strain variation in metabolite production was only moderately associated with accessory gene presence/absence also suggests that strain-specific curation may be required for quantitative prediction of some metabolic traits, such as production of nucleotide metabolites. These additional improvements may further advance the application of constraint-based modeling to accurately predict *E. lenta*'s metabolic activity in the complex and dynamic mammalian gut ecosystem.

Our approach combining untargeted metabolomics, genome-driven media development, computational modeling, and gnotobiotic experiments may be a useful strategy for accelerating scientific understanding of the biology of other understudied microbes. Each of these model systems and data types produced a broadly useful resource that partially supported findings from the others while also revealing novel facets of *E. lenta* metabolism. Together, our study sheds light on the unusual metabolic profile of an important member of the human gut microbiota, establishes a foundation for future mechanistic studies of this organism, and demonstrates a generalizable multidisciplinary approach to decipher the metabolic strategies of understudied microbes.

Materials and methods

Experimental model and subject details

Mouse husbandry and experiments. Mouse samples analyzed in this study were collected and described previously in [4]. The mouse experiment was approved by the University of California San Francisco Institutional Animal Care and Use Committee (protocol AN184143). The mice were housed at temperatures ranging from 67 to 74°F and humidity ranging from 30% to 70% light/dark cycle 12 hours/12 hours. LabDiet 5021 chow was used. No mice were

involved in previous procedures before experiments were performed. Mice were assigned to groups to achieve similar age distribution between groups. Researchers were not blinded to group assignment.

C57BL/6J mice (males, ages 4 to 8 weeks) were obtained from the University of California, San Francisco Gnotobiotics core facility (gnotobiotics.ucsf.edu; Jackson Labs Cat# 000664) and housed in Iso-positive cages (Tecniplast). Mice were colonized via oral gavage with *E. lenta* monocultures (10^9 CFU/mL, 200 μ L gavage, $n = 6$ *E. lenta* DSM 2243, $n = 5$ *E. lenta* DSM 15644, $n = 4$ *E. lenta* AB12n2) or control media ($n = 5$ GF). Colonization was confirmed via anaerobic culturing and/or qPCR for an *E. lenta*-specific marker (*elnmrk1*) [10,26]. Mice were colonized for 2 weeks prior to euthanasia and sample collection.

Bacterial strains. Strain isolates analyzed in this work are described in [26]. All experiments were performed in an anaerobic chamber with 2% to 5% hydrogen gas, 20% carbon dioxide, and the balance nitrogen, with growth in a 37°C incubator. Standard BHI media (VWR 90003–040) supplemented with 1% L-arginine (referred to below as BHI+) was used for culturing outside of defined media experiments.

Method details

Defined media formulations and preparation. Standard composition of the EDM1 media and related formulations are provided in [S1 Table](#). As specified in [S1 Table](#), some experiments were performed using the initial formulation of the media, and others using a simplified form based on the results of leave-one-out growth experiments. For most components, 30 to 1,000 \times stock solutions were prepared following [76]. Stock solutions were sterilized with a 0.22- μ m syringe filter and stored at –20°C. Amino acids and uracil were typically added together directly from powder into a combined 2 \times stock solution, which was then filter sterilized with a 0.22- μ m vacuum filter, except when preparing individual leave-one-out amino acid growth experiments. Most versions used ATCC Trace Mineral and Vitamin Mix Supplements (MD-TMS and MD-VS), except for experiments to test leaving out individual components of these mixes. Media formulations were allowed to equilibrate in an anaerobic chamber (Coy) for at least 24 hours prior to use. To investigate the sources of trace background compounds observed in untargeted metabolomics profiling of defined media, we also prepared fresh solutions of arginine, uracil, and the combined amino acid stock described above, collected them directly with the sample collection, preservation, and extraction methods described below, and analyzed them with the same metabolomics workflow.

Bacterial culture and growth assays. For growth and metabolomics experiments, glycerol stocks were first streaked on BHI+ agar plates and incubated at 37°C for 2 to 3 days. Individual colonies were inoculated into 3 to 4 mL liquid BHI+ and incubated at 37°C for 40 to 48 hours, or until approximately early stationary phase. Culture optical density (600 nm wavelength absorbance, OD600) was measured using a Hach DR1900 spectrophotometer. A sample of BHI starter culture (1 mL) was then centrifuged at 1,568 rcf for 4 minutes in a microcentrifuge (Thermo Scientific mySpin 12) in the anaerobic chamber and resuspended in 1 mL sterile phosphate-buffered saline (PBS). For leave-one-out experiments, the resuspended cells were washed by centrifuging and resuspending in PBS again. The resulting suspension was vortexed and diluted to an approximate OD600 of 0.1 and used as inoculum into defined experimental conditions.

Growth assays were performed in standard 96-well microplates (Corning) at 37°C with a microplate reader (Biotek Eon or PowerWave). Defined media was pipetted into each well (180 μ L), followed by 20 μ L of inoculum. All experiments included at least 3 sterile control wells for each condition, into which 20 μ L of sterile PBS was pipetted to establish consistent

background OD600 measurements. Replicate wells were distributed pseudorandomly across the plate to control for plate layout effects, and inoculated wells were always paired with an adjacent control well of the same condition. For each condition, 3 to 6 replicates were included. Plates were sealed with a transparent Breathe-Easy sealing gas exchange membrane (RPI). Every 30 minutes, plates were shaken at medium speed for 40 seconds, after which OD600 readings were performed.

After large metabolomics and RNA-Seq experiments (see below), culture purity was checked by plating and 16S rRNA gene Sanger sequencing, using standard primers (8F AGAGTTTGATCCTGGCTCAG and 1542R AAGGAGGTGATCCAGCCGCA). In the time course metabolomics experiment shown in Fig 1, cross-contamination of sterile control tubes occurred after the fourth time point. Growth in both culture and control tubes was confirmed as pure *E. lenta* by plating and 16S rRNA sequencing.

Sample collection for metabolomics. Time course experiments were conducted in tubes in the anaerobic chamber in a 37°C incubator. For all metabolomics experiments, 3 independent culture replicates were included for each condition, with an equal number of uninoculated control tubes. Starter cultures and inocula were prepared as described above for growth assays. Approximately 5 mL of defined media was added to VWR glass culture tubes (53283–800) with screw caps. The PBS-washed inoculum was added to culture tubes to obtain an approximate starting OD600 of 0.001. A preliminary growth assay was conducted to define time points spanning the exponential growth phase in the tested conditions. At each time point, OD600 measurements of all inoculated tubes were first measured using a Hach DR1900 spectrophotometer, with a paired control tube to normalize for the background. A 100 µL sample from each tube was then transferred into a 96-well microplate, which was sealed and removed from the anaerobic chamber. Plates were centrifuged at 1,928 rcf at 4°C for 8 minutes, after which supernatants were collected into fresh polypropylene tubes or plates, sealed, and flash-frozen in liquid nitrogen.

Two time course experiments were carried out with stable isotope-labeled substrates. Experimental groups included conditions in which sodium acetate in the defined media was replaced with $^{13}\text{C}_2$ labeled sodium acetate (Sigma-Aldrich 282014), along with a matched experimental group with the same concentration of unlabeled substrate. The same procedure was followed for the arginine labeling experiment, using $^{13}\text{C}_6$ labeled L-arginine HCl (Sigma-Aldrich 643440).

For the comparative strain metabolomics experiment, 96-well polypropylene deep well plates were prepared with 800 µL of fresh media in each well. Starter cultures and inocula for 29 isolates of *Eggerthella lenta* and 1 isolate of *Eggerthella sinensis* [26] (S5 Table) were prepared as described above for growth assays, except without final dilution, and 80 µL was used to inoculate wells, leaving a blank well in between every culture well to prevent cross-contamination. After 72 hours, OD600 measurements were taken, plates were centrifuged, and supernatants were collected as described above.

Targeted quantification of acetate. A subset of unlabeled supernatant samples from the acetate labeling time course were shipped to Stanford University on dry ice for targeted quantification of acetate.

Samples (20 µL) were first mixed with an internal standard solution (30 µL, 1 mM phenylpropionate-d9) in a V-bottomed, poly(propylene), 96-well plate, and extracted by mixing with 3 sample volumes of extraction solution (75% acetonitrile, 25% methanol). The plate was covered with a lid and centrifuged at 5,000 rcf for 15 minutes at 4°C. Supernatant was collected for derivatization before subjecting to LC-MS analysis.

Samples were processed using a derivatization method targeting compounds containing a free carboxylic acid. Extracted samples were mixed with 3-nitrophenylhydrazine (NPH; 200

mM in 50% acetonitrile) and N-(3-dimethylaminopropyl)-N'-ethylcarbodiimide (120 mM in 6% pyridine) at a 2:1:1 ratio. The plate was sealed with a plastic sealing mat (Thermo Fisher Scientific, #AB-0566) and incubated at 40°C, 600 rpm in a thermomixer for 60 minutes to derivatize the carboxylate-containing compounds. The reaction mixture was quenched with 0.02% formic acid in 10% acetonitrile:water before LC-MS.

Samples were injected via refrigerated autosampler into mobile phase and chromatographically separated by an Agilent 1290 Infinity II UPLC and detected using an Agilent 6545XT Q-TOF (quadrupole time of flight) mass spectrometer equipped with a dual jet stream electrospray ionization source, operating under extended dynamic range (1,700 m/z). Chromatographic separation was performed using an ACQUITY Bridged Ethylene Hybrid (BEH) C18 column 2.1 × 100 mm, 1.7-μm particle size, (Waters, Milford, MA), using chromatographic conditions published elsewhere [6]. MS1 spectra were collected in centroid mode, and peak assignments in samples were made based on comparisons of retention times and accurate masses from authentic standards using MassHunter Quantitative Analysis v10.0 software from Agilent Technologies. Acetate was quantified from calibration curves constructed with acetate-d4 as a standard using isotope-dilution MS with phenylpropionate-d9 as the internal standard. Calibration curves were performed in a modified base form of EDM1 lacking amino acids and other carboxylic acids. A background level of 1.05 mM of acetate was subtracted to obtain the final quantities.

A plate layout error for supernatant samples from time points 4 to 7 in this experiment was noted based on the resulting acetate concentrations and corrected across datasets.

Untargeted metabolomics. Bacterial culture supernatant and sterile media, used in culture, were thawed on wet ice. Once thawed, samples were homogenized by inversion 5 times. Extracellular culture supernatant samples were prepared as follows: 20 μL of culture supernatant were extracted using 80 μL of a chilled extraction solvent at -20°C (1:1 acetonitrile:methanol, 5% water containing stable isotope-labeled internal standards). Samples were homogenized via pipette action, incubated for 1 hour at -20°C, centrifuged at 4°C at 6,000 rcf for 5 minutes. The supernatant was transferred to a new plate and immediately sealed and kept at 4°C prior to prompt analysis via LC-MS/MS.

Intestinal samples (colon, cecum, and ileum) were prepared individually using a single protocol as follows. Samples were kept frozen on dry ice and massed to at least 10 mg. Approximately 4 mL of -20°C extraction solvent (2:2:1 methanol:acetonitrile:water + stable isotope labeled internal standards) were added per milligram of intestinal sample. Six to eight 1 mm zirconia silica beads were added to each sample followed by prompt bead beating (15 Hz, for 10 minutes). Following a 1-hour incubation in the -20°C freezer, samples were centrifuged at 4°C at 18,407 rcf for 5 minutes. Supernatant was collected and stored at -20°C prior to centrifugal plate filtration (0.2 μm polyvinylidene difluoride (PVDF), Agilent Technologies, Santa Clara CA) at 4°C at 4,122 rcf for 3 minutes. Collection plate was sealed and maintained at 4°C prior to prompt analysis.

Serum samples were first thawed on wet ice. Each 20 μL sample of serum was extracted with 4 volumes of methanol, containing stable isotope-labeled internal standards. Samples were homogenized by vortexing for 20 seconds and placed in a -20°C for 1 hour to maximize protein precipitation. After freezer incubation, samples were centrifuged at 4°C at 18,407 rcf for 5 minutes. Supernatant was removed and dried under vacuum via centrivap (Labconco). Dried samples were then resuspended in 30 μL of 80% acetonitrile in water containing exogenous standard CUDA at 60 ng/mL. Samples were maintained at 4°C prior to prompt analysis.

Within each analysis batch, a small amount of each sample was removed and combined to create multiple technical replicate “pools,” which were analyzed intermittently throughout the analysis. These pools were used as external standards to ensure instrument stability across the

batch. Additionally, method blanks were created using LC–MS-grade water in place of supernatant, sterile media, serum, or intestinal contents. These blanks were used to ensure that reported metabolites were not inadvertently added during sample preparation.

Samples, sterile media, pools, and blanks were promptly added to a Thermo Vanquish Autosampler at 4°C in a Vanquish UHPLC (Thermo Fisher Scientific, Waltham, MA). Chromatographic separation was performed using an ACQUITY Bridged Ethylene Hybrid (BEH) Amide column 2.1 × 150 mm, 1.7-μm particle size, (Waters, Milford, MA), using chromatographic conditions published elsewhere (HILIC method described in the Supplementary Methods of [25]). Samples were analyzed on a Thermo Q-Exactive HF orbitrap mass spectrometer operated utilizing data-dependent acquisition of MS2. Data were acquired independently in positive and negative modes via subsequent injections.

Stable isotope-resolved metabolomics (SIRM). Intracellular extract samples were prepared with the following procedure, which was optimized for lysis of thick gram-positive cell walls: 600 μL of culture was transferred to an Eppendorf tube in anaerobic conditions and subsequently centrifuged at 10,000 rcf for 3 minutes at 4°C, after which the supernatant was removed and the samples were immediately flash frozen to quench metabolites. To each pellet, 300 μL of cold methanol was then added, followed by sonication on ice for 5 minutes and then shaking at 4°C for 4 to 12 hours. Samples were then centrifuged at 4°C at 15,000 rcf for 8 minutes, after which 120 μL of supernatant was transferred to fresh tubes and stored at –80°C until analysis. Prior to analysis, intracellular samples were dried at room temperature via Centriva Benchtop Concentrator (Labconco). Samples were resuspended in 60 μL of a chilled solution of 1:1 methanol and acetonitrile, with 24% water at –20°C containing the internal standards CUDA and VAL-TYR-VAL each at 60 ng/mL. Samples were centrifuged at 4°C, 4,122 rcf for 5 minutes, and the supernatant transferred to a vial and immediately capped for LC-MS analysis.

Extracellular supernatant extraction for SIRM was performed as described above (Untargeted metabolomics section) with one modification. In SIRM samples, deuterated internal standards were replaced with CUDA and Val-Tyr-Val to enable untargeted enrichment analysis. LC-MS/MS analysis conditions for SIRM were identical to those used for standard untargeted metabolomics.

Untargeted metabolomics data processing. Untargeted metabolomics datasets were processed using MS-DIAL version 4.60 [77]. Metabolite features with intensity not greater than 3-fold elevated in samples compared to mean blank intensity were removed. Annotations were assigned using both local [25] and global (Mass Bank of North America) tandem mass spectral libraries. Annotation confidence scores were assigned based on MSI best practices [78,79]. Briefly, MSI level 1 denotes library matches of accurate mass (m/z), retention time (RT), and tandem mass spectra (MS2). MSI level 2 follows the same rules as MSI 1 but allows for partial matching of MS2 spectra—as is prone to occur when experimental spectra are convoluted. MSI level 3 denotes a high scoring and visually confirmed match of MS2 spectra. In a modification from the original definition of the MSI, we assign MSI level 4 when exact stereospecificity cannot be determined by MS2 and chromatographic separation. MSI level 4 is often assigned to sugars, lipids, and polyphenols. MSI levels 1 and 2 could only be assigned to metabolites in our local library, for which authentic standards have been analyzed in the same chromatographic conditions as the samples being annotated. Library matches were within 0.0025 m/z and 0.25 minutes retention time for all annotations assigned MSI levels 1 and 2 (comparisons are shown for the time course dataset in S12 Fig). Postprocessing was performed using MS-FLO [80] for removal of erroneous features.

Processed datasets were further analyzed using Feature-based Molecular Networking and MolNetEnhancer in the GNPS web platform version 22 [81–84], which assigned ClassyFire

chemical classes to features based on molecular networking, independently of whether they were assigned a library identity.

To merge positive and negative ionization mode datasets from the same samples, duplicate features across datasets were identified as those with an expected mass difference of less than 0.02, a retention time difference of less than 0.1, and a Pearson correlation across samples of at least 0.7. If one or both members of a pair of duplicate features were assigned an identification, the feature with lower (more confident) MSI score was retained in the merged dataset. Otherwise, the positive mode feature was retained. The other feature was removed for downstream analysis.

Prior to statistical analysis, initial untargeted metabolomics feature tables were filtered to remove features with a high coefficient of variation across replicate samples (>50%) and to remove potential technical outlier samples where the total signal from all features differed from the assay median by >50%. Log₁₀-transformed intensities were used for most statistical analysis, with the exceptions of SIRM datasets and the comparative strains dataset (for which values were approximately normally distributed without transformation). A pseudocount equal to 0.25 times the minimum nonzero value was added to the peak intensities for each feature before log transformation. Heatmaps of metabolite abundances were generated using the *ComplexHeatmap* package v2.13.1 [85].

SIRM data processing. Intra- and extracellular untargeted data generated from SIRM experiments was analyzed separately using *Compound Discoverer* version 3.3 (Thermo Scientific, Bremen, Germany). Samples treated with labeled compounds were always paired with matched samples treated with unlabeled compounds in order to correct for naturally occurring isotope abundances. Unlabeled samples were used for compound detection and formula assignment via isotope pattern-based prediction, spectral library matches, or mass lists matches. The isotope patterns and formulas from the sample files then served as a reference for the detection of potential isotopologues per compound in the labeled sample type.

Specifically, the workflow consisted of the following nodes in *Compound Discoverer*: *Input Files*→*Select Spectra*→*Align Retention Times (ChromAlign)*→*Detect Compounds (Legacy)*→*Group Compounds*→*Predict Compositions*→*Search Mass Lists*→*Search mzCloud*→*Mark Background Compounds*→*Assign Compound Annotations*→*Analyze Labeled Compounds*→*Descriptive Statistics*→*Differential analysis*.

The default settings from the “Stable Isotope Labeling w Metabolika Pathways and ID using Online Databases” workflow were used, with the following modifications:

1. Detect Compounds (Legacy): General–Min.Peak Intensity: 10000; Ions: [M+H]⁺+1 or [M-H][−]-1 for positive and negative mode experiments respectively.
2. Group Compounds: Peak Rating Filter–Peak Rating Threshold: 4; Number of Files: 3.
3. Search Mass Lists: Search Settings–Mass Lists: Combined Hilic Mass mzRT library; Use Retention Time: True; RT Tolerance: 0.3 min; Mass Tolerance: 5 ppm.
4. Search mzCloud: DDA Search–Match Factor Threshold: 85
5. Mark Background Compounds: General–Max. Sample/Blank: 3
6. Assign Compound Annotations: Data Sources–Data Source #1: mzCloud Search; Data Source #2: MassList Search; Data Source #3: Predicted Compositions.
7. Analyze Labeled Compounds: Pattern Analysis–Intensity Threshold [%]: 2

Positive and negative polarity files were analyzed initially as separate studies with the following study definitions: Study factors including strain, replicate, substrate concentration,

sample type, and time point were assigned. Sample types were assigned as either sample (unlabeled), labeled, or blank. These study factors interfaced with several nodes to reduce undesirable features and maximize reporting of quality high-intensity peaks with potential for accurate measurement of ^{13}C incorporation.

Results were filtered for nonblank formula assignment and absence in background samples. The MSI levels and labeling status for persisting entries were manually inspected for each compound and annotated onboard via custom tags. MSI levels were assigned based on the criteria previously described to match MS-Dial output. The mass isotopologue distributions were plotted to ensure reproducibility between replicates of various time points and detect anomalous labeling trends. The absence of reported enrichment in control samples processed as labeled samples was verified. A minimum threshold of 3% combined enrichment across all isotopologues other than M+0 was applied. This threshold was necessary for less abundant peaks where the ^{13}C natural isotopic abundance correction introduces uncertainty in the M+1 and M+2 isotopologues.

A specification of the full Compound Discoverer workflow is available at https://github.com/turnbaughlab/2022_Noecker_ElentaMetabolism.

RNA-Seq. *E. lenta* DSM 2243 glycerol stocks were plated on BHI+ and incubated anaerobically at 37°C for 3 days. A single colony was then inoculated into 5 mL of BHI+ liquid culture and incubated at 37°C for 48 hours. A volume of 1 mL of the resulting culture was then centrifuged, washed once, and resuspended in equal volume PBS; all in anaerobic conditions. To obtain a starting OD₆₀₀ of 0.01, 220 μL of this inoculum were transferred into culture flasks containing 20 mL of EDM1 (70% carbon source reduced version; see [S1 Table](#)). After 20 hours of growth (early or mid-exponential phase), these cultures were treated with an additional 1.9 mL of sterile water or filter-sterilized solution containing either L-arginine (to reach a final concentration of 86 mM), sodium acetate (final concentration of 14.5 mM), or agmatine sulfate (final concentration of 30 mM). After 18 more hours (late exponential phase), 7.5 mL of each culture was collected into 15 mL conical tubes containing 5 mL of Qiagen Bacterial RNA-Protect (#76506). Cultures were centrifuged at 2,800 rcf at 4°C for 10 minutes, after which the supernatant was carefully removed. Pellets were extracted directly using the Qiagen RNeasy Mini kit (#74104) with modifications for difficult-to-lyse gram-positive bacteria. Samples were maintained on ice in between all disruption and centrifugation steps throughout the protocol. Briefly, 200 μL of TE buffer containing lysozyme (15 mg/mL, #L4919) and 20 μL of Qiagen Proteinase K (#19131) were added to each pellet, vortexed gently, and incubated at room temperature for 10 minutes with shaking on an Eppendorf ThermoMixer at 900 rpm. Buffer RLT (700 μL) was then added to each tube and vortexed, after which the full contents were transferred to MP Biomedical Lysing Matrix E tubes (#116914500) and disrupted mechanically in a BioSpec Mini-Beadbeater-96 for 50 seconds. After disruption, tubes were centrifuged for 3 minutes at 15,000 rcf, and 850 μL of supernatant was transferred to fresh tubes. To each sample, 590 μL of 80% ethanol was added and mixed by pipetting, after which lysates were transferred to Qiagen RNeasy spin columns and washed, following the RNeasy Mini kit QuickStart protocol including a single on-column DNase digestion (DNase #79254). After purification, RNA was eluted twice into 30 μL of nuclease-free water. RNA integrity was checked using an Agilent TapeStation 4150 and stored at -80°C.

RNA library preparation and sequencing was performed by the Microbial Genome Sequencing Center/SeqCenter. Samples were DNase treated with Invitrogen DNase (RNase free, #AM2222). Library preparation was performed using Illumina's Stranded Total RNA Prep Ligation with Ribo-Zero Plus kit (#20040529) and 10 bp IDT for Illumina indices. Supplementary oligonucleotide probes specific to *E. lenta* rRNA and other highly expressed non-coding RNAs were incorporated during Ribo-Zero depletion ([S10 Table](#)). Sequencing was

done on a NextSeq 2000 with 2×50 bp reads. Demultiplexing, quality control, and adapter trimming was performed with bcl-convert v3.9.3.

Reads were trimmed and quality filtered using *fastp* v0.20.0 [86] with the following parameters: `—trim_poly_g—cut_front—cut_tail—cut_window_size 4—cut_mean_quality 20—length_required 15`. The Hisat2 aligner v2.2.1 [87] was used to map reads to the *E. lenta* DSM 2243 reference genome, downloaded from NCBI RefSeq (GCF_000024265.1). Gene-level read counts were obtained using the corresponding NCBI annotations and the *featureCounts* function in the R package *Rsubread* v2.6.4 [88], with the minimum quality score set to 1.

Construction, curation, and analysis of metabolic reconstructions. Genome-scale metabolic reconstructions were created from genome sequences of 25 *E. lenta* strains [26] using the DEMETER pipeline [89]. Briefly, DEMETER performs systematic refinement of a draft genome-scale reconstruction, in this case generated through KBase [90]. Based on manually gathered experimental data, gap-filling solutions that had been manually determined in a subset of reconstructions are propagated by DEMETER to newly reconstructed strains. Moreover, DEMETER ensures correct reconstruction structure through use of a curated reaction and metabolite database and removes futile cycles resulting in unrealistically high ATP production. A test suite ensures agreement with the input experimental data and verifies model features such as mass and charge balance and feasible ATP production. The *E. lenta* DSM 2243 reconstruction [35] underwent additional refinement of reactions and gene annotations against manually performed comparative genomics analyses.

Reconstructions were analyzed using the Cobra Toolbox version 3.0 [91] in Matlab version 2018b, with the IBM Cplex solver version 128. Defined media concentrations were mapped from compound names to BiGG metabolite IDs [92] and converted to cell uptake rates over the duration of *E. lenta*'s exponential growth phase in batch culture (S1 Table) using the *concToCellRate* function in the Cobra Toolbox and an approximate cell dry weight of 3.3×10^{-13} g, calculated based on colony forming units and dry biomass quantification from 2 aliquots of a late-exponential phase EDM1 culture. Additional compounds detected in sterile culture media with high confidence based on untargeted metabolomics were included in the simulation media with a fixed maximum uptake rate of 1 mM/gDW/h.

The collection of *E. lenta* strain reconstructions included 2 reconstructions of the type strain: the DSM 2243 reconstruction, which had undergone additional comparative genomics curation with PubSeed [93], and a slightly smaller and less refined reconstruction included in the AGORA2 collection [35] based on genome resequencing of the ATCC 25559 version of the type strain. Neither reconstruction initially displayed nonzero growth in EDM1 using flux balance analysis. In order to facilitate interpretation of flux balance analysis (FBA) results and avoid excess gap-filled reactions, we used the simpler *E. lenta* ATCC 25559 type strain reconstruction as the basis for subsequent curation and analysis. We transferred reactions present in the DSM 2243 reconstruction into this version if they were supported by genome annotations from other sources (Prokka [48], GapMind [37]) and/or by experimental growth or metabolomics data. We also performed several additional custom curations. Transporters were added in a strain-specific manner for metabolites identified with high confidence (MSI level 1) and detected as secreted or depleted with a \log_2 fold change greater than 2 in the stationary phase strain collection metabolomics dataset (Fig 4). Several pathways were also modified based on growth assay results and/or pathway annotation software [36,37]. Curations were checked for viable growth in EDM1 using flux balance analysis. Reconstructions for the other 23 strains were only curated to ensure growth in EDM1 and to allow import/export based on metabolomics data, but not based on genome analysis with GapMind [37] or the results of leave-one-out growth experiments, since those were only performed using the type strain. A complete summary of all curation steps is found in S3 Table.

FBA, pFBA, and FVA were performed using the Cobra Toolbox functions *optimizeCbModel*, *minimizeModelFlux*, and *fastFVA*, respectively. Flux variability ranges are reported for 99% of the maximum growth rate.

Metabolite uptake and secretion ranges estimated by FVA were compared with the stationary phase strain collection metabolomics dataset (shown in Fig 4). To compare metabolite data with FVA estimates, identified metabolites were first mapped from InChIKey metabolite IDs to KEGG IDs using the CTS Convert utility [94] implemented in the R package *webchem* v1.1.2 [95]. KEGG IDs were then mapped to BiGG IDs using the BiGG database [92] and manually checked for consistency with compound IDs in the AGORA models. For purposes of this analysis, metabolites were considered produced if they had a \log_2 fold change greater than 0.5 in supernatants from at least one of the 3 type strain isolates included in the experiment (*E. lenta* DSM 2243—UCSF, *E. lenta* ATCC 25559, and *E. lenta* DSM 2243—DSMZ), and depleted if the \log_2 fold change was less than -0.5 .

To compare gene expression values with model flux estimates, we first ran pFBA and FVA for the modified EDM1 condition used for RNA-Seq (with 70% of the standard levels of arginine and acetate). We obtained the set of genes linked to reactions in the iEL2243_2 reconstruction, using NCBI BLASTn to map genes between different sets of annotations. Genes linked to multiple reactions were counted multiple times for each reaction, and vice versa. Only genes linked to reactions in the original ATCC 25559 reconstruction were included.

Similarly, to compare reaction knockout predictions with strain variation, genes linked to reactions in the original ATCC 25559 reconstruction were mapped to annotations used in a previous pan-genome analysis of 31 nonclonal *E. lenta* genomes [26]. In this previous analysis, amino acid sequence families were clustered across genomes using ProteinOrtho [96] with cutoffs of 60% identity and 80% coverage. This analysis was then used to determine the number of strains in which each gene family in the ATCC 25559 reconstruction was present and to compare that distribution with the effects predicted by knockout analysis of the unconstrained model.

Strain comparative metabolomics analysis. Metabolites were classified as strain-variably produced/depleted if they were differentially increased/decreased (FDR-adjusted $p < 0.1$ and absolute \log_2 fold change > 0.5) in supernatants from at least 1 isolate strain but fewer than 29 (of the 30 isolates included in this experiment).

The phylogenetic and comparative genomics analyses used in this study were previously reported, including a core gene phylogenetic tree (PhyloPhlan), gene family clustering across strains (ProteinOrtho) and Prokka and GhostKoala annotation of all genomes [26].

Procrustes analysis was performed using the R package *vegan* v2.6–2, with evaluation of significance using the *protest* function [97]. The *E. sinensis* isolate was excluded from Procrustes analysis to avoid skewing the distribution of phylogenetic distances.

The gene-metabolite association analysis was performed as described previously [26], with different cutoffs for prioritization. Briefly, all observed patterns of gene family presence-absence (based on clusters of 60% identity and 80% coverage) were enumerated across the collection of genomes. Log-transformed metabolite intensities were then tested for association with each presence-absence pattern using Welch's *t* tests. Using an initial cutoff of an FDR-adjusted *p*-value of 10^{-4} , 39.0% of metabolite features were significantly associated with a gene cluster by this method. To further restrict results to those features most likely to depend on the presence of a gene, we filtered gene-metabolite links using 2 additional separability criteria. First, the difference in median \log_{10} metabolite values between strain samples with and without the gene was required to be at least 0.4. Secondly, the 10th percentile \log_{10} metabolite value for strains with the gene was required to be at least 0.4 above the maximum value in controls, and the 90th percentile value for strains without the gene was required to be lower than that

value. Finally, only the highest association for each metabolite feature was retained. This additional filtering resulted in the final table of 84 gene family-metabolite links. KEGG pathway enrichment analysis of the final gene set was performed using *clusterProfiler* v4.0.5 [98] with a *p*-value cutoff of 0.1.

Cross-dataset untargeted metabolomics analysis. As described above, untargeted metabolomics datasets from supernatant, mouse intestinal contents, and serum were collected using the same chromatography and mass spectrometry methods. Pairs of features were compared across these datasets and linked if they were within 0.007 *m/z*, 0.5 minutes retention time and had a cosine similarity of at least 0.205 between their MS2 spectra for positive ionization mode and 0.251 for negative ionization mode. Features for which MS2 spectra were not collected were linked to other features within 0.001 *m/z* and 0.2 retention time. Linked feature pairs were also required to be annotated as the same adduct. These *m/z* and retention time thresholds were chosen based on examination of the distributions of pairwise differences between features. The cosine similarity cutoffs were chosen as the 99.5th percentile of cosine similarity between a large sample of unrelated feature pairs: specifically, all pairwise comparisons of 2 sets of 200 randomly sampled features with retention times differing by at least 1 and *m/z* differing by at least 0.01. This procedure was repeated separately for positive and negative ionization mode datasets. Under these criteria, only approximately 0.5% of linked features assigned an identity were linked to features with a conflicting identity. Linked pairs of features were merged into shared metabolite IDs that were carried forward for cross-dataset analysis and comparison.

Quantification and statistical analysis

All statistical analyses were performed in R v4.1.1, with data visualizations generated using the *ggplot2* package v3.3.6 [99]. Statistical tests, sample size, and standard error are reported in the figures and figure legends. Benjamini–Hochberg FDR correction was used to adjust for multiple comparisons in all cases.

Untargeted metabolomics statistical analysis. Differential trajectories across time series datasets were assessed using spline models implemented in the *santaR* package v1.2.3 [100]. Differential abundance analysis between supernatant samples and sterile controls at the final time point was performed using Welch's *t* tests after checking assumptions of normality. In one case where cross-contamination of sterile control tubes occurred at later time points, features at those time points were compared with control samples from the last uncontaminated time point.

Differential abundance analysis for the comparative strains dataset was performed using linear models with each strain identity as a covariate. In the gnotobiotic mouse experiment, each mouse was considered an independent experimental unit; however, hierarchical random effects statistical models were used to account for shared covariance between mice from the same cage. Specifically, differential abundance analysis for the gnotobiotic mouse intestinal dataset was performed using linear mixed models with the R package *lmerTest* v3.1–3 [101], incorporating fixed effects for intestinal site, colonization group, and the interaction between them, and nested random effects for cage and animal. The *diffsmeans* function with Benjamini–Hochberg multiple hypothesis adjustment was used to evaluate the statistical significance of differences of each colonization group versus GF under this model.

Statistical analysis of growth curves. Growth curves were normalized based on average time-matched readings from blank control wells. Normalized values were used to fit logistic growth models for each well using the R package *growthcurver* v0.3.1 [102]. Low-quality model fits ($\sigma > 0.1$) were removed prior to calculation of summarized parameter values.

Targeted metabolomics statistical analysis. Differential abundance analysis was performed using a linear model with terms for time point, strain, and their interaction. Differences from controls under the resulting model were estimated using Dunnett's method as implemented in the package *emmeans* v1.7.5 [103].

RNA-Seq statistical analysis. Differential expression analysis was performed using negative binomial generalized linear models implemented in the *DESeq2* package v1.32.0 [104].

Supporting information

S1 Fig. EDM1 chemically defined media supports robust growth of *Eggerthella lenta* and enables sensitive metabolomics profiling. Related to Fig 1. (A) Summary of the composition of EDM1 media. The number in parentheses indicates the number of specific compounds in each category. (B) Growth of *E. lenta* DSM 2243 in 2 commonly used media conditions (Brain Heart Infusion supplemented with L-arginine, and ISP-2 media supplemented with L-arginine), compared with 3 initial defined media formulations. (C) Comparison of total number of differentially abundant features and identified differentially abundant features in this experiment compared to previous metabolomics profiling of *E. lenta*. The combination of chemically defined culture media and untargeted metabolomics methods used in this experiment allowed for greater detection of metabolites produced by *E. lenta*. Produced and depleted features were defined as having an absolute \log_2 fold change vs. sterile control media greater than 0.5. (D) Metabolomics profiling of compounds known to be present in the chemically defined media formulation EDM1. A total of 22 media compounds were detected, most of which were not significantly depleted in *E. lenta* cultures over time. (E) Identified compounds detected by untargeted metabolomics in freshly prepared EDM1 media stock solutions. The heatmap shows observed peak heights for features detected from either freshly prepared sterile stock of 14 amino acids plus uracil ("Combined AA stock"), sterile L-arginine solution, or sterile uracil solution. Values are only shown for samples where the intensity was more than 3 times the average value in blank control samples. Guanine is indicated with a star based on its rapid depletion in *E. lenta* cultures. Its detection in the amino acid solution suggests that it may be a background contaminant from an undetermined amino acid component. The yellow-orange dots indicate compounds that are defined components of the media recipe. The dark red dots denote metabolite features that were previously observed in sterile controls in the time course experiment shown in Fig 1, based on strict matching of *m/z* and retention time. (F) Hierarchical clustering of metabolite trajectories reveals distinct growth phases. Scaled average metabolite intensities across time points during growth in EDM1 were hierarchically clustered with complete linkage and cut into discrete clusters with a height of 1.6, distinguishing early-, mid-, and late-produced and depleted metabolites. Cluster order is arbitrary. Annotated metabolites are listed below each cluster along with their MSI confidence level. Colors indicate ClassyFire metabolite classes as assigned by GNPS. Only clusters with at least 1 identified metabolite and at least 5 total features are shown. EDM1, *Eggerthella* Defined Media 1; MSI, Metabolomics Standards Initiative. (PDF)

S2 Fig. Effects of individual media components on growth of *E. lenta* DSM 2243. Related to Fig 1. (A) Growth curves for *E. lenta* DSM2243 growth in EDM1 media with individual media components removed. Gray curves indicate growth in full EDM1 media in the same experiment. Curves are shown as mean \pm standard error. A blue asterisk indicates a significance difference in the area under the empirical growth curve with and without the compound (Wilcoxon rank-sum test, FDR-adjusted $p < 0.2$). (B) Distribution of median effects of removal of all tested compounds on growth parameters estimated by a logistic model. The

dotted line indicates the median parameter estimate for the full EDM1 media across all experiments. Parameters were fitted with a logistic model implemented by the R package *growthcurver*. EDM1, *Eggerthella* Defined Media 1; FDR, false discovery rate. (PDF)

S3 Fig. Environmental acetate concentrations affect growth and metabolite production of 3 *E. lenta* strains. Related to Fig 2. (A) Targeted quantification of acetate depletion in *E. lenta* EDM1 cultures. Acetate was measured at 2–3 time points in supernatant samples from 3 *E. lenta* strains during growth in EDM1 as well as sterile controls. Quantification was performed using a method for derivatization of carboxylic acids with 3-nitrophenylhydrazine and N-(3-dimethylaminopropyl)-N'-ethylcarbodiimide followed by targeted LC-MS/MS. Error bars show mean \pm standard error. Linear models of acetate concentration versus strain and time point were inferred for each media group, and differences from controls under the resulting model were estimated using Dunnett's method. * indicates $p < 0.05$, *** indicates $p < 0.001$. (B) Growth of 3 *E. lenta* strain isolates in EDM1 with 0, 1, or 10 mM sodium acetate. Mean \pm standard error across 3 replicates is shown. (C) Acetate-responsive metabolites in supernatants from *E. lenta* AB8n2 and *E. lenta* Valencia. Metabolites shown are those that were assigned an identification, were differentially abundant compared with sterile controls (FDR-adjusted $p < 0.2$), and had significantly different trajectories over time in the presence vs. absence of acetate in either strain (based on smoothing spline regression with the R package *santaR*, FDR-adjusted $p < 0.25$). Values shown are scaled log-transformed peak heights. The number in parentheses indicates the MSI confidence level for each metabolite annotation (see [Materials and methods](#)). EDM1, *Eggerthella* Defined Media 1; FDR, false discovery rate; MSI, Metabolomics Standards Initiative. (PDF)

S4 Fig. Consistent incorporation of acetate across 3 *E. lenta* strains based on SIRM. Related to Fig 2. (A) Growth of *E. lenta* strains in EDM1 with varying levels of sodium acetate (either stable isotope-labeled $^{13}\text{C}_2$ or unlabeled). Optical density measurements were taken, and supernatant samples were collected at each indicated time point. Mean \pm standard error across 3 replicates is shown. (B) Average trajectories of labeled extracellular metabolites in 3 different strains of *E. lenta*. Metabolites shown are those with $>50\%$ and $>5 \times 10^4$ average peak area from labeled isotopologues in at least 1 time point in the 10 mM labeled acetate group. For metabolites detected in both positive and negative ionization mode, only positive mode is shown. The value in parentheses indicates the MSI annotation confidence level for each metabolite. (C) Labeled metabolites of known identity in intracellular extracts across 3 strains of *E. lenta* (data for DSM 2243 matches [Fig 2E](#)). Each panel shows the average mass isotopologue distribution across 3 replicates for a single metabolite in intracellular extracts from time point 5 (39 hours, late exponential phase). Metabolites are labeled with the compound name and MSI annotation confidence level in parentheses. Metabolites included are those with $>15\%$ and $>10^4$ average peak area from labeled isotopologues in either the 1 mM or 10 mM labeled acetate group. *N*-acetylated amino acids are excluded for space and reported in [S1 Data](#). The isotopologue color legend is the same as in panel (B). (D) Labeled metabolites of unknown identity across 3 strains of *E. lenta*. Each panel shows the average mass isotopologue distribution (across 3 replicates) for a single metabolite in intracellular extracts from time point 5 (39 hours, late exponential phase). Metabolites are labeled with their estimated exact mass, retention time, and ionization mode. Metabolites included are those with $>15\%$ and $>10^4$ average peak area from labeled isotopologues in either the 1 mM or 10 mM labeled acetate group. The isotopologue color legend is the same as in panels (B) and (C). EDM1, *Eggerthella* Defined Media 1; MSI, Metabolomics Standards Initiative; SIRM, stable isotope-

resolved metabolomics.
(PDF)

S5 Fig. Stable isotope profiling of *E. lenta* arginine metabolism confirms that arginine is primarily converted to ornithine as an energy source. Related to Fig 2. (A) Citrulline, but not ornithine, has a similar effect as L-arginine on *E. lenta* growth. Growth curves of *E. lenta* grown in EDM1 media where the 1% L-arginine (red) has been replaced with an equimolar quantity of either L-citrulline (blue) or L-ornithine (green). Curves show mean \pm standard error across 4 replicates. (B) In *E. lenta* DSM 2243 cultures grown with 1% $^{13}\text{C}_6$ labeled arginine, correspondingly labeled citrulline and ornithine accumulate in supernatants over the course of growth. Curves show mean \pm standard error across 3 replicates. (C) Mass isotopologue distributions of extracellular metabolites. Each barplot shows the isotopologue mean peak areas for each feature over time. Compounds shown are those of known identity that increase by a factor of at least 2^4 , have at least 1 isotopologue with a peak area of greater 10^6 in at least 1 time point, and have a labeled isotopologue with $>3\%$ abundance in at least 1 time point. (D) Mass isotopologue distributions of intracellular metabolites. Each barplot shows the mean peak areas of isotopologues for each feature at 2 time points. Compounds shown are those of known identity with an average labeled MID > 0.1 and a total peak area from labeled isotopologues of at least 10^5 in at least 1 time point. The isotopologue color legend is the same as in panel (C). (E) Distribution of total signal of extracellular metabolites across labeling patterns. While signal from numerous unlabeled compounds is detected over time (left), compounds with M+5 labeling patterns are mainly restricted to ornithine, citrulline, and a compound of unknown identity (middle), and compounds found with high signal as M+6 isotopologues are mainly arginine and citrulline (right). Compounds shown are those with the highest peak areas at the final time point in positive ionization mode. (F) Hypothesized pathways for metabolism of L-arginine by *E. lenta*. Circles indicate the number of carbon atoms in selected compounds and are colored blue to indicate incorporation of ^{13}C isotopes from external arginine. Compound names in bold were detected with the observed labeling patterns in either intracellular metabolite extracts or culture supernatants right. MID, mass isotopologue distribution.
(PDF)

S6 Fig. Sensitivity of iEL2243_2 maximum growth rate to biomass equation coefficients. Related to Fig 3. Each panel shows the effect of varying the coefficient of the biomass equation for selected components on FBA-estimated maximum growth rate in EDM1 media. Compounds shown are those whose biomass coefficients have the largest effects on estimated growth rate. EDM1, *Eggerthella* Defined Media 1; FBA, flux balance analysis.
(PDF)

S7 Fig. Single-reaction knockout analysis of iEL2243_2 identifies conserved genes across metabolic subsystems. Related to Fig 3. (A) Predicted effects of knocking out reactions in the top 20 largest subsystems on growth of *E. lenta*, according to pFBA analysis of the iEL2243_2 model. Reactions designated “Has effect” are those for which the knockout has a predicted maximum growth rate less than wild-type but greater than 0. Essential reactions are those that reduced biomass flux to 0 when removed from the model. (B) Reactions linked to more conserved gene families are more likely to have substantial effects on growth when removed. Each point represents a reaction, separated on the x axis by whether the model without that reaction grew at $>70\%$ of the wild-type model. The y axis indicates the fraction of *E. lenta* strain genomes in which gene families (defined using ProteinOrtho clustering) linked to that reaction were present.
(PDF)

S8 Fig. Within-species variation in *E. lenta* metabolic profiles across genomes and metabolomes. Related to Fig 4. (A) Phylogeny of 30 *Eggerthella* strains analyzed in this study. This phylogeny was previously constructed based on core gene alignments using PhyloPhlan [26]. (B) PCA of log-transformed metabolite intensity profiles of stationary phase supernatants from 30 *Eggerthella* isolates in EDM1. In the left-hand panel, color indicates the geographic origin of each isolate. The right panel shows the largest feature loadings for the PCA and their corresponding chemical classes as assigned by GNPS, where available. Dereplicated metabolite features with an average value $> 10^5$ in at least 1 strain were included. (C) Distribution of the number of strains producing or depleting each metabolite feature. Features included are those that were significantly modified by at least 1 *Eggerthella* isolate in this experiment (FDR-adjusted p -value < 0.1 and \log_2 fold change > 0.5). (D) Comparison of MS1 spectra and extracted ion chromatograms for the unidentified features shown in Fig 4D in a representative sample, compared with a pure ribitol standard spiked into sterile EDM1 at a concentration of 100 $\mu\text{g/mL}$. The data shown in the top row are from the biological sample, while the data in the bottom row are from the ribitol standard. (E) Map of the teichoic acid biosynthesis region of the genome of representative *Eggerthella* strains. Genes outlined in bold are the gene families associated with the unidentified metabolite features shown in Fig 4D. Gene regions were defined in each genome based on the location of the genes annotated as *tagG* and *tagH* by Prokka. EDM1, *Eggerthella* Defined Media 1; FDR, false discovery rate; PCA, principal component analysis.

(PDF)

S9 Fig. Genome-scale metabolic modeling of *E. lenta* strain variation. (A) Distribution of core and accessory reactions across subsystems, based on comparative analysis of metabolic reconstructions of 24 *E. lenta* strain genomes. (B) Predicted maximum growth rate inferred by flux balance analysis of each of the 24 *E. lenta* strain reconstructions in 52 leave-one-out media conditions based on EDM1. Gray tiles indicate predicted cases of zero growth.

(PDF)

S10 Fig. Differential abundance analysis of intestinal and serum metabolites of *E. lenta*-monocolonized mice compared to GF. Related to Fig 5. (A) Volcano plots of differential abundance analysis of metabolite features in intestinal contents and serum of gnotobiotic mice monocolonized with one of 3 *E. lenta* strains. Effect sizes and significance are estimated from group comparisons based on linear mixed models of log-transformed metabolite abundances, accounting for animal and cage random effects. (B) Total number of untargeted metabolomics features in intestinal contents and serum of gnotobiotic mice that could be linked to features in either of 2 in vitro EDM1 metabolomics datasets, based on high similarity of m/z , retention time, and MS2 spectra. (C) Comparison of the effect of *E. lenta* DSM 2243 on metabolites detected in both EDM1 cultures in the untargeted time course experiment and monocolonized mice. Each point represents a metabolite feature detected in both datasets. The x axis indicates the \log_2 fold change of each feature in supernatants from the *E. lenta* DSM 2243 time course experiment compared with sterile controls, compared with the covariate-adjusted \log_2 fold change of that feature in monocolonized mice compared with GF mice. Points are colored green if the feature is significantly differentially abundant in gnotobiotic mice and is shifted in the same direction by the corresponding strain in the time course in vitro experiment. EDM1, *Eggerthella* Defined Media 1; GF, germ-free.

(PDF)

S11 Fig. Shifts in intestinal amino acid metabolites of *E. lenta*-monocolonized mice compared to GF. Related to Fig 6. (A) Annotated metabolites with the largest shifts in intestinal

contents of *E. lenta*-colonized mice compared with GF. Metabolites are shown if they were identified based on library comparison and were among the 600 most strongly shifted features in any individual site or colonization group, based on linear mixed models. Each point shows the effect size in a single site, and color indicates chemical class where available (assigned using feature-based molecular networking with GNPS). **(B)** Abundance of arginine and agmatine-related metabolites in gnotobiotic mice. Arginine is only slightly depleted by *E. lenta*, although its expected products, ornithine and citrulline, are greatly increased. Agmatine is significantly depleted, while its expected product, putrescine, is not significantly increased. “.” indicates Benjamini–Hochberg adjusted $p < 0.1$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$. **(C)** Volcano plots illustrating shifts in the abundance of proteinogenic amino acids in *E. lenta*-colonized mice. Arginine is colored in green. Effect sizes and significance are estimated from group comparisons based on linear mixed models of log-transformed metabolite abundances, accounting for animal and cage random effects.

(PDF)

S12 Fig. Annotation of untargeted metabolomics features by comparison with library standards. Related to Materials and methods. **(A)** Difference between library standard m/z and average observed m/z for annotated metabolite features in the EDM1 time course dataset, separated by MSI annotation confidence level. **(B)** Difference between library standard retention time and average observed retention time for annotated metabolite features in the EDM1 time course dataset, separated by MSI annotation confidence level. Features assigned MSI of 3 or 4 did not necessarily have retention times matched with a library standard. EDM1, *Eggerthella* Defined Media 1; MSI, Metabolomics Standards Initiative.

(PDF)

S1 Table. Chemically defined media formulations used in this study. Related to Fig 1 and Materials and methods. Recipes used for preparation of chemically defined media used for experiments in this study. The first 2 columns indicate the manufacturer information for each compound and the concentration of working solution prepared for that compound. Unless otherwise specified, reference to EDM1 indicates that the “Standard EDM1” preparation was used.

(XLSX)

S2 Table. Summarized results of media leave-one-out growth experiments. Related to Fig 1. Parameters were fit by logistic growth models using the R package *growthcurver*. A separate model was fit for each replicate in each experiment, and the average and standard deviation for each parameter across replicates are reported. Average growth rate r was calculated as a harmonic mean.

(XLSX)

S3 Table. Curation steps applied to *E. lenta* DSM 2243/ATCC 25559 AGORA reconstruction. Related to Fig 3. Summary of curation steps, supporting data, and gene annotations for each reaction added or modified in the iEL2243_2 reconstruction.

(XLSX)

S4 Table. Most highly expressed genes by *E. lenta* DSM 2243 during growth in EDM1. Related to Fig 3. Locus tags, gene annotation, and average and standard deviation of the 100 most highly expressed transcripts during *E. lenta* growth in the baseline EDM1 condition.

(XLSX)

S5 Table. *Eggerthella* strains profiled with untargeted metabolomics. List of strains for which EDM1 culture supernatants were analyzed with untargeted metabolomics, along with

corresponding culture collection identifiers.
(XLSX)

S6 Table. Metabolite features associated with variable *E. lenta* gene families across strains. Related to Fig 4. Results of association analysis linking patterns of strain-variable genes with strain-variable metabolite features. Associations listed are those that met the strictest significance and separability criteria (see [Materials and methods](#)). Gene annotations are listed for association patterns with 20 or fewer candidate gene families.
(XLSX)

S7 Table. Summary of conserved and strain-variable reactions by subsystem in *E. lenta* strain metabolic reconstructions. Related to Fig 4. Statistics on the distribution of core and accessory reactions across *E. lenta* strain metabolic reconstructions.
(XLSX)

S8 Table. Genes linked to agmatine utilization by *E. lenta*. Related to Fig 6. KEGG annotations of gene families in the agmatine deiminase pathway in *E. lenta* genomes, as previously obtained using GhostKoala.
(XLSX)

S9 Table. *E. lenta* genes differentially expressed in response to the presence of agmatine. Related to Fig 6. *E. lenta* DSM 2243 genes with differential expression in response to agmatine sulfate treatment (FDR-adjusted $p < 0.1$ and absolute \log_2 fold change > 1), as estimated by DESeq2.
(XLSX)

S10 Table. Supplementary oligonucleotide probes used for depletion of highly abundant *E. lenta* noncoding RNAs. Related to Materials and methods. Probes designed for depletion of *E. lenta* ribosomal RNA and highly abundant *ssrA* and *rnpB* noncoding RNAs, used in Illumina Ribo-Zero library preparation.
(XLSX)

S1 Data. Labeled features detected in stable isotope experiments. Related to Fig 2. Summary of labeled isotopologues detected by untargeted metabolomics. Each tab includes data for a single experiment and sample type: extracellular metabolites with labeled acetate, intracellular metabolites with labeled acetate, extracellular metabolites with labeled arginine, and intracellular metabolites with labeled arginine. In addition to basic properties of each compound/feature, the average peak area, standard error in peak area, and average fractional distribution are reported for each detected isotopologue. Compounds were filtered based on the same criteria as in [Figs 2, S4 and S5](#).
(XLSX)

S2 Data. Differentially abundant features across in vivo and in vitro untargeted metabolomics datasets. Related to Fig 5. Each tab lists the set of untargeted metabolomics features that were differentially abundant (linear mixed effects models, absolute \log_2 fold change estimate > 1 and FDR-adjusted p -value < 0.2) in at least 1 intestinal site between *E. lenta*-colonized and GF mice and that were also detected in in vitro untargeted metabolomics experiments, separated by strain and by feature annotation status (identified/unknown). For each feature, the corresponding \log_2 fold change and significance in the in vitro dataset(s) are listed for comparison. Features are ordered by their effect size in cecal contents.
(XLSX)

Acknowledgments

We thank the UCSF Gnotobiotics Core staff (Jessie Turnbaugh, Kimberly Ly) for animal care support and assistance and SeqCenter for assistance with RNA-Seq library preparation and sequencing.

Author Contributions

Conceptualization: Cecilia Noecker, Peter J. Turnbaugh.

Data curation: Cecilia Noecker, Juan Sanchez, Yuanyuan Liu, Dylan Dodd.

Formal analysis: Cecilia Noecker, Juan Sanchez, Kai Trepka, Almut Heinken, Yuanyuan Liu, Brian C. DeFelice.

Funding acquisition: Cecilia Noecker, Peter J. Turnbaugh.

Investigation: Cecilia Noecker, Juan Sanchez, Jordan E. Bisanz, Veronica Escalante, Margaret Alexander, Yuanyuan Liu.

Methodology: Cecilia Noecker, Juan Sanchez, Jordan E. Bisanz, Brian C. DeFelice, Peter J. Turnbaugh.

Software: Cecilia Noecker, Almut Heinken, Ines Thiele.

Supervision: Dylan Dodd, Ines Thiele, Brian C. DeFelice, Peter J. Turnbaugh.

Visualization: Cecilia Noecker.

Writing – original draft: Cecilia Noecker.

Writing – review & editing: Cecilia Noecker, Juan Sanchez, Jordan E. Bisanz, Veronica Escalante, Margaret Alexander, Kai Trepka, Almut Heinken, Yuanyuan Liu, Dylan Dodd, Ines Thiele, Brian C. DeFelice, Peter J. Turnbaugh.

References

1. Tramontano M, Andrejev S, Pruteanu M, Klünemann M, Kuhn M, Galardini M, et al. Nutritional preferences of human gut bacteria reveal their metabolic idiosyncrasies. *Nat Microbiol*. 2018. <https://doi.org/10.1038/s41564-018-0123-9> PMID: 29556107
2. Lagkouvardos I, Overmann J, Clavel T. Cultured microbes represent a substantial fraction of the human and mouse gut microbiota. *Gut Microbes*. 2017; 8:493–503. <https://doi.org/10.1080/19490976.2017.1320468> PMID: 28418756
3. Medlock GL, Carey MA, McDuffie DG, Mundy MB, Giallourou N, Swann JR, et al. Inferring Metabolic Mechanisms of Interaction within a Defined Gut Microbiota. *Cell Systems*. 2018; 7:245–257.e7. <https://doi.org/10.1016/j.cels.2018.08.003> PMID: 30195437
4. Alexander M, Ang QY, Nayak RR, Bustion AE, Sandy M, Zhang B, et al. Human gut bacterial metabolism drives Th17 activation and colitis. *Cell Host Microbe*. 2021. <https://doi.org/10.1016/j.chom.2021.11.001> PMID: 34822777
5. Koropatkin NM, Cameron EA, Martens EC. How glycan metabolism shapes the human gut microbiota. *Nat Rev Microbiol*. 2012; 10:323–335. <https://doi.org/10.1038/nrmicro2746> PMID: 22491358
6. Liu Y, Chen H, Van Treuren W, Hou B-H, Higginbottom SK, Dodd D. *Clostridium sporogenes* uses reductive Stickland metabolism in the gut to generate ATP and produce circulating metabolites. *Nat Microbiol*. 2022; 7:695–706. <https://doi.org/10.1038/s41564-022-01109-9> PMID: 35505245
7. Franzosa EA, Sirota-Madi A, Avila-Pacheco J, Fornelos N, Haiser HJ, Reinker S, et al. Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nat Microbiol*. 2018. <https://doi.org/10.1038/s41564-018-0306-4> PMID: 30531976
8. Hertel J, Harms AC, Heinken A, Baldini F, Thinnies CC, Glaab E, et al. Integrated Analyses of Microbiome and Longitudinal Metabolome Data Reveal Microbial-Host Interactions on Sulfur Metabolism in Parkinson's Disease. *Cell Rep*. 2019; 29:1767–1777.e8. <https://doi.org/10.1016/j.celrep.2019.10.035> PMID: 31722195

9. Almeida A, Mitchell AL, Boland M, Forster SC, Gloor GB, Tarkowska A, et al. A new genomic blueprint of the human gut microbiota. *Nature*. 2019. <https://doi.org/10.1038/s41586-019-0965-1> PMID: 30745586
10. 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*. 2018; 7:e33953. <https://doi.org/10.7554/eLife.33953> PMID: 29761785
11. Gardiner BJ, Tai AY, Kotsanas D, Francis MJ, Roberts SA, Ballard SA, et al. Clinical and Microbiological Characteristics of *Eggerthella lenta* Bacteremia. *J Clin Microbiol*. 2015; 53:626–635. <https://doi.org/10.1128/JCM.02926-14> PMID: 25520446
12. Zhu Q, Hou Q, Huang S, Ou Q, Huo D, Vázquez-Baeza Y, et al. Compositional and genetic alterations in Graves' disease gut microbiome reveal specific diagnostic biomarkers. *ISME J*. 2021. <https://doi.org/10.1038/s41396-021-01016-7> PMID: 34079079
13. Chen J, Wright K, Davis JM, Jeraldo P, Marietta EV, Murray J, et al. An expansion of rare lineage intestinal microbes characterizes rheumatoid arthritis. *Genome Med*. 2016; 8:43. <https://doi.org/10.1186/s13073-016-0299-7> PMID: 27102666
14. Islam MZ, Tran M, Xu T, Tierney BT, Patel C, Kostic AD. Reproducible and opposing gut microbiome signatures distinguish autoimmune diseases and cancers: a systematic review and meta-analysis. *Microbiome*. 2022; 10:218. <https://doi.org/10.1186/s40168-022-01373-1> PMID: 36482486
15. Cekanaviciute E, Yoo BB, Runia TF, Debelius JW, Singh S, Nelson CA, et al. Gut bacteria from multiple sclerosis patients modulate human T cells and exacerbate symptoms in mouse models. *Proc Natl Acad Sci U S A*. 2017; 114:10713–10718. <https://doi.org/10.1073/pnas.1711235114> PMID: 28893978
16. Haiser HJ, Gootenberg DB, Chatman K, Sirasani G, Balskus EP, Turnbaugh PJ. Predicting and Manipulating Cardiac Drug Inactivation by the Human Gut Bacterium *Eggerthella lenta*. *Science*. 2013; 341:295–298. <https://doi.org/10.1126/science.1235872> PMID: 23869020
17. Devlin AS, Fischbach MA. A biosynthetic pathway for a prominent class of microbiota-derived bile acids. *Nat Chem Biol*. 2015; 11:685–690. <https://doi.org/10.1038/nchembio.1864> PMID: 26192599
18. Bess EN, Bisanz JE, Yarza F, Bustion A, Rich BE, Li X, et al. Genetic basis for the cooperative bioactivation of plant lignans by *Eggerthella lenta* and other human gut bacteria. *Nat Microbiol*. 2020; 5:56–66. <https://doi.org/10.1038/s41564-019-0596-1> PMID: 31686027
19. Maini Rekdal V, Bess EN, Bisanz JE, Turnbaugh PJ, Balskus EP. Discovery and inhibition of an interspecies gut bacterial pathway for Levodopa metabolism. *Science*. 2019; 364:eaau6323. <https://doi.org/10.1126/science.aau6323> PMID: 31196984
20. Harris SC, Devendran S, Méndez- García C, Mythen SM, Wright CL, Fields CJ, et al. Bile acid oxidation by *Eggerthella lenta* strains C592 and DSM 2243^T. *Gut Microbes*. 2018;1–17. <https://doi.org/10.1080/19490976.2018.1458180> PMID: 29617190
21. Maini Rekdal V, Nol Bernadino P, Luescher MU, Kiamehr S, Le C, Bisanz JE, et al. A widely distributed metalloenzyme class enables gut microbial metabolism of host- and diet-derived catechols. *eLife*. 2020;9. <https://doi.org/10.7554/eLife.50845> PMID: 32067637
22. Hylemon PB, Harris SC, Ridlon JM. Metabolism of hydrogen gases and bile acids in the gut microbiome. *FEBS Lett*. 2018; 592:2070–2082. <https://doi.org/10.1002/1873-3468.13064> PMID: 29683480
23. Noronha A, Modamio J, Jarosz Y, Guerard E, Sompairac N, Preciat G, et al. The Virtual Metabolic Human database: integrating human and gut microbiome metabolism with nutrition and disease. *Nucleic Acids Res*. 2019; 47(D1):D614–D624. <https://doi.org/10.1093/nar/gky992> PMID: 30371894
24. Benton HP, Ivanisevic J, Mahieu NG, Kurczy ME, Johnson CH, Franco L, et al. Autonomous metabolomics for rapid metabolite identification in global profiling. *Anal Chem*. 2015; 87:884–891. <https://doi.org/10.1021/ac5025649> PMID: 25496351
25. Han S, Van Treuren W, Fischer CR, Merrill BD, DeFelice BC, Sanchez JM, et al. A metabolomics pipeline for the mechanistic interrogation of the gut microbiome. *Nature*. 2021; 595:415–420. <https://doi.org/10.1038/s41586-021-03707-9> PMID: 34262212
26. Bisanz JE, Soto-Perez P, Noecker C, Aksenov AA, Lam KN, Kenney GE, et al. A Genomic Toolkit for the Mechanistic Dissection of Intractable Human Gut Bacteria. *Cell Host Microbe*. 2020;1–14. <https://doi.org/10.1016/j.chom.2020.04.006> PMID: 32348781
27. Beloborodov NV, Khodakova AS, Bairamov IT, Olenin AY. Microbial origin of phenylcarboxylic acids in the human body. *Biochemistry*. 2009; 74:1350–1355. <https://doi.org/10.1134/s0006297909120086> PMID: 19961416
28. van der Hee B, Wells JM. Microbial Regulation of Host Physiology by Short-chain Fatty Acids. *Trends Microbiol*. 2021; 29:700–712. <https://doi.org/10.1016/j.tim.2021.02.001> PMID: 33674141
29. Sperry JF, Wilkins TD. Arginine, a growth-limiting factor for *Eubacterium lentum*. *J Bacteriol*. 1976; 127:780–784. <https://doi.org/10.1128/jb.127.2.780-784.1976> PMID: 182668

30. Zhang Y, Cai J, Shang X, Wang B, Liu S, Chai X, et al. A new genome-scale metabolic model of *Corynebacterium glutamicum* and its application. *Biotechnol Biofuels*. 2017; 10:169. <https://doi.org/10.1186/s13068-017-0856-3> PMID: 28680478
31. Monk JM, Lloyd CJ, Brunk E, Mih N, Sastry A, King Z, et al. iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat Biotechnol*. 2017; 35:904–908. <https://doi.org/10.1038/nbt.3956> PMID: 29020004
32. Gu C, Kim GB, Kim WJ, Kim HU, Lee SY. Current status and applications of genome-scale metabolic models. *Genome Biol*. 2019; 20:121. <https://doi.org/10.1186/s13059-019-1730-3> PMID: 31196170
33. Chiu H-C, Levy R, Borenstein E. Emergent Biosynthetic Capacity in Simple Microbial Communities. *PLoS Comput Biol*. 2014; 10:e1003695. <https://doi.org/10.1371/journal.pcbi.1003695> PMID: 24992662
34. Diener C, Gibbons SM, Resendis-Antonio O. MICOM: Metagenome-Scale Modeling To Infer Metabolic Interactions in the Gut Microbiota. *mSystems*. 2020; 5: e00606–19, /msystems/5/1/msys.00606-19.atom. <https://doi.org/10.1128/mSystems.00606-19> PMID: 31964767
35. Heinken A, Hertel J, Acharya G, Ravcheev DA, Nyga M, Okpala OE, et al. Genome-scale metabolic reconstruction of 7,302 human microorganisms for personalized medicine. *Nat Biotechnol*. 2023. <https://doi.org/10.1038/s41587-022-01628-0> PMID: 36658342
36. Pascal Andreu V, Roel-Touris J, Dodd D, Fischbach MA, Medema MH. The gutSMASH web server: automated identification of primary metabolic gene clusters from the gut microbiota. *Nucleic Acids Res*. 2021:gkab353. <https://doi.org/10.1093/nar/gkab353> PMID: 34019648
37. Price MN, Deutschbauer AM, Arkin AP. Filling gaps in bacterial catabolic pathways with computation and high-throughput genetics. *PLoS Genet*. 2022; 18:e1010156. <https://doi.org/10.1371/journal.pgen.1010156> PMID: 35417463
38. Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat Biotechnol*. 2010; 28:977–982. <https://doi.org/10.1038/nbt.1672> PMID: 20802497
39. Aurich MK, Fleming RMT, Thiele I. MetaboTools: A Comprehensive Toolbox for Analysis of Genome-Scale Metabolic Models. *Front Physiol*. 2016;7. <https://doi.org/10.3389/fphys.2016.00327> PMID: 27536246
40. Thiele I, Palsson BØ. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc*. 2010; 5:93–121. <https://doi.org/10.1038/nprot.2009.203> PMID: 20057383
41. Smith NW, Shorten PR, Altermann E, Roy NC, McNabb WC. Mathematical modelling supports the existence of a threshold hydrogen concentration and media-dependent yields in the growth of a reductive acetogen. *Bioprocess Biosyst Eng*. 2020; 43:885–894. <https://doi.org/10.1007/s00449-020-02285-w> PMID: 31982985
42. Massengo-Tiassé RP, Cronan JE. Diversity in enoyl-acyl carrier protein reductases. *Cell Mol Life Sci*. 2009; 66:1507–1517. <https://doi.org/10.1007/s00018-009-8704-7> PMID: 19151923
43. Strobel HJ. Basic laboratory culture methods for anaerobic bacteria. *Methods Mol Biol*. 2009; 581:247–261. https://doi.org/10.1007/978-1-60761-214-8_16 PMID: 19768627
44. Rodionov DA, Arzamasov AA, Khoroshkin MS, Iablokov SN, Leyn SA, Peterson SN, et al. Micronutrient Requirements and Sharing Capabilities of the Human Gut Microbiome. *Front Microbiol*. 2019; 10:1316. <https://doi.org/10.3389/fmicb.2019.01316> PMID: 31275260
45. de Crécy-Lagard V, El Yacoubi B, de la Garza RD, Noirié A, Hanson AD. Comparative genomics of bacterial and plant folate synthesis and salvage: predictions and validations. *BMC Genomics*. 2007; 8:245. <https://doi.org/10.1186/1471-2164-8-245> PMID: 17645794
46. Levin I, Giladi M, Altman-Price N, Ortenberg R, Mevarech M. An alternative pathway for reduced folate biosynthesis in bacteria and halophilic archaea. *Mol Microbiol*. 2004; 54:1307–1318. <https://doi.org/10.1111/j.1365-2958.2004.04339.x> PMID: 15554970
47. Ravcheev DA, Thiele I. Systematic genomic analysis reveals the complementary aerobic and anaerobic respiration capacities of the human gut microbiota. *Front Microbiol*. 2014;5. <https://doi.org/10.3389/fmicb.2014.00674> PMID: 25538694
48. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014; 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153> PMID: 24642063
49. Percy MG, Gründling A. Lipoteichoic acid synthesis and function in gram-positive bacteria. *Annu Rev Microbiol*. 2014; 68:81–100. <https://doi.org/10.1146/annurev-micro-091213-112949> PMID: 24819367
50. Brown S, Santa Maria JP, Jr WS. Wall teichoic acids of gram-positive bacteria. *Annu Rev Microbiol*. 2013; 67:313–336. <https://doi.org/10.1146/annurev-micro-092412-155620> PMID: 24024634

51. Weidenmaier C, Peschel A. Teichoic acids and related cell-wall glycopolymers in Gram-positive physiology and host interactions. *Nat Rev Microbiol*. 2008; 6:276–287. <https://doi.org/10.1038/nrmicro1861> PMID: 18327271
52. Soto-Perez P, Bisanz JE, Berry JD, Lam KN, Bondy-Denomy J, Turnbaugh PJ. CRISPR-Cas System of a Prevalent Human Gut Bacterium Reveals Hyper-targeting against Phages in a Human Virome Catalog. *Cell Host Microbe*. 2019; 26:325–335.e5. <https://doi.org/10.1016/j.chom.2019.08.008> PMID: 31492655
53. Buttner C, Bottacini F, Shkorporov AN, Draper LA, Ross P, Hill C. Selective Isolation of *Eggerthella lenta* from Human Faeces and Characterisation of the Species Prophage Diversity. *Microorganisms*. 2022; 10:195. <https://doi.org/10.3390/microorganisms10010195> PMID: 35056644
54. van Dalen R, Peschel A, van Sorge NM. Wall Teichoic Acid in *Staphylococcus aureus* Host Interaction. *Trends Microbiol*. 2020; 28:985–998. <https://doi.org/10.1016/j.tim.2020.05.017> PMID: 32540314
55. Monk JM, Charusanti P, Aziz RK, Lerman JA, Premyodhin N, Orth JD, et al. Genome-scale metabolic reconstructions of multiple *Escherichia coli* strains highlight strain-specific adaptations to nutritional environments. *Proc Natl Acad Sci U S A*. 2013; 110:20338–20343. <https://doi.org/10.1073/pnas.1307797110> PMID: 24277855
56. Piletz JE, Aricioglu F, Cheng J-T, Fairbanks CA, Gilad VH, Haenisch B, et al. Agmatine: clinical applications after 100 years in translation. *Drug Discov Today*. 2013; 18:880–893. <https://doi.org/10.1016/j.drudis.2013.05.017> PMID: 23769988
57. Ll acer JL, Polo LM, Tav  rez S, Alarc  n B, Hilario R, Rubio V. The Gene Cluster for Agmatine Catabolism of *Enterococcus faecalis*: Study of Recombinant Putrescine Transcarbamylase and Agmatine Deiminase and a Snapshot of Agmatine Deiminase Catalyzing Its Reaction. *J Bacteriol*. 2007; 125:1254–1265. <https://doi.org/10.1128/jb.01216-06> PMID: 17028272
58. Pacheco AR, Moel M, Segr   D. Costless metabolic secretions as drivers of interspecies interactions in microbial ecosystems. *Nat Commun*. 2019; 10:103. <https://doi.org/10.1038/s41467-018-07946-9> PMID: 30626871
59. Chodkowski JL, Shade A. Exometabolite Dynamics over Stationary Phase Reveal Strain-Specific Responses. *mSystems*. 2020; 5:e00493–20. <https://doi.org/10.1128/mSystems.00493-20> PMID: 33361318
60. Dunphy LJ, Grimes KL, Wase N, Kolling GL, Papin JA. Untargeted Metabolomics Reveals Species-Specific Metabolite Production and Shared Nutrient Consumption by *Pseudomonas aeruginosa* and *Staphylococcus aureus*. *mSystems*. 2021; 6. <https://doi.org/10.1128/mSystems.00480-21> PMID: 34156287
61. Venturelli OS, Carr AC, Fisher G, Hsu RH, Lau R, Bowen BP, et al. Deciphering microbial interactions in synthetic human gut microbiome communities. *Mol Syst Biol*. 2018; 14:e8157. <https://doi.org/10.15252/msb.20178157> PMID: 29930200
62. Roager HM, Licht TR. Microbial tryptophan catabolites in health and disease. *Nat Commun*. 2018; 9:3294. <https://doi.org/10.1038/s41467-018-05470-4> PMID: 30120222
63. Mager LF, Burkhard R, Pett N, Cooke NCA, Brown K, Ramay H, et al. Microbiome-derived inosine modulates response to checkpoint inhibitor immunotherapy. *Science*. 2020:eabc3421. <https://doi.org/10.1126/science.abc3421> PMID: 32792462
64. Li D, Feng Y, Tian M, Ji J, Hu X, Chen F. Gut microbiota-derived inosine from dietary barley leaf supplementation attenuates colitis through PPAR  signaling activation. *Microbiome*. 2021; 9:83. <https://doi.org/10.1186/s40168-021-01028-7> PMID: 33820558
65. Paik D, Yao L, Zhang Y, Bae S, D'Agostino GD, Zhang M, et al. Human gut bacteria produce TH17-modulating bile acid metabolites. *Nature*. 2022; 603:907–912. <https://doi.org/10.1038/s41586-022-04480-z> PMID: 35296854
66. Girinathan BP, DiBenedetto N, Worley JN, Peltier J, Arrieta-Ortiz ML, Immanuel SRC, et al. In vivo commensal control of *Clostridioides difficile* virulence. *Cell Host Microbe*. 2021; 29:1693–1708.e7. <https://doi.org/10.1016/j.chom.2021.09.007> PMID: 34637781
67. Pruss KM, Enam F, Battaglioli E, DeFeo M, Diaz OR, Higginbottom SK, et al. Oxidative ornithine metabolism supports non-inflammatory *C. difficile* colonization. *Nat Metab*. 2022; 4:19–28. <https://doi.org/10.1038/s42255-021-00506-4> PMID: 34992297
68. Amador-Noguez D, Feng X-J, Fan J, Roquet N, R  bitz H, Rabinowitz JD. Systems-level metabolic flux profiling elucidates a complete, bifurcated tricarboxylic acid cycle in *Clostridium acetobutylicum*. *J Bacteriol*. 2010; 192:4452–4461. <https://doi.org/10.1128/JB.00490-10> PMID: 20622067
69. Huynen MA, Dandekar T, Bork P. Variation and evolution of the citric-acid cycle: a genomic perspective. *Trends Microbiol*. 1999; 7:281–291. [https://doi.org/10.1016/s0966-842x\(99\)01539-5](https://doi.org/10.1016/s0966-842x(99)01539-5) PMID: 10390638

70. Duncan SH, Barcenilla A, Stewart CS, Pryde SE, Flint HJ. Acetate Utilization and Butyryl Coenzyme A (CoA):Acetate-CoA Transferase in Butyrate-Producing Bacteria from the Human Large Intestine. *Appl Environ Microbiol.* 2002; 68:5186–5190. <https://doi.org/10.1128/AEM.68.10.5186-5190.2002> PMID: 12324374
71. Muñoz-Tamayo R, Laroche B, Walter E, Doré J, Duncan SH, Flint HJ, et al. Kinetic modelling of lactate utilization and butyrate production by key human colonic bacterial species. *FEMS Microbiol Ecol.* 2011; 76:615–624. <https://doi.org/10.1111/j.1574-6941.2011.01085.x> PMID: 21388423
72. Haenisch B, von Kügelgen I, Bönisch H, Göthert M, Sauerbruch T, Schepke M, et al. Regulatory mechanisms underlying agmatine homeostasis in humans. *Am J Physiol Gastrointest Liver Physiol.* 2008; 295:G1104–G1110. <https://doi.org/10.1152/ajpgi.90374.2008> PMID: 18832451
73. Galgano F, Caruso M, Condelli N, Favati F. Focused review: agmatine in fermented foods. *Front Microbiol.* 2012; 3:199. <https://doi.org/10.3389/fmicb.2012.00199> PMID: 22701114
74. Molderings GJ, Kribben B, Heinen A, Schröder D, Brüß M, Göthert M. Intestinal tumor and agmatine (decarboxylated arginine): low content in colon carcinoma tissue specimens and inhibitory effect on tumor cell proliferation in vitro. *Cancer.* 2004; 101:858–868. <https://doi.org/10.1002/cncr.20407> PMID: 15305420
75. Heinken A, Basile A, Thiele I. Advances in constraint-based modelling of microbial communities. *Curr Opin Syst Biol.* 2021; 27:100346. <https://doi.org/10.1016/j.coisb.2021.05.007>
76. Zhang G, Mills DA, Block DE. Development of Chemically Defined Media Supporting High-Cell-Density Growth of Lactococci, Enterococci, and Streptococci. *Appl Environ Microbiol.* 2009; 75:1080–1087. <https://doi.org/10.1128/AEM.01416-08> PMID: 19074601
77. Tsugawa H, Cajka T, Kind T, Ma Y, Higgins B, Ikeda K, et al. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat Methods.* 2015; 12(6):523–526. <https://doi.org/10.1038/nmeth.3393> PMID: 25938372
78. Schymanski EL, Jeon J, Gulde R, Fenner K, Ruff M, Singer HP, et al. Identifying small molecules via high resolution mass spectrometry: communicating confidence. *Environ Sci Technol.* 2014; 48:2097–2098. <https://doi.org/10.1021/es5002105> PMID: 24476540
79. Fiehn O, Robertson D, Griffin J, van der Werf M. The metabolomics standards initiative (MSI). *Metabolomics.* 2007. Available from: <https://link.springer.com/article/10.1007/s11306-007-0070-6>.
80. DeFelice BC, Mehta SS, Samra S, Čajka T, Wancewicz B, Fahrman JF, et al. Mass Spectral Feature List Optimizer (MS-FLO): A tool to minimize false positive peak reports in untargeted liquid chromatography-mass spectrometry (LC-MS) data processing. *Anal Chem.* 2017; 89:3250–3255. <https://doi.org/10.1021/acs.analchem.6b04372> PMID: 28225594
81. Nothias L-F, Petras D, Schmid R, Dührkop K, Rainer J, Sarvepalli A, et al. Feature-based molecular networking in the GNPS analysis environment. *Nat Methods.* 2020; 17:905–908. <https://doi.org/10.1038/s41592-020-0933-6> PMID: 32839597
82. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol.* 2016; 34:828–837. <https://doi.org/10.1038/nbt.3597> PMID: 27504778
83. Ernst M, Kang KB, Caraballo-Rodríguez AM, Nothias L-F, Wandy J, Chen C, et al. MolNetEnhancer: Enhanced Molecular Networks by Integrating Metabolome Mining and Annotation Tools. *Metabolites.* 2019;9. <https://doi.org/10.3390/metabo9070144> PMID: 31315242
84. Djoumbou Feunang Y, Eisner R, Knox C, Chepelev L, Hastings J, Owen G, et al. ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *J Cheminform.* 2016; 8:61. <https://doi.org/10.1186/s13321-016-0174-y> PMID: 27867422
85. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics.* 2016; 32:2847–2849. <https://doi.org/10.1093/bioinformatics/btw313> PMID: 27207943
86. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 2018; 34:i884–i890. <https://doi.org/10.1093/bioinformatics/bty560> PMID: 30423086
87. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* 2019; 37:907–915. <https://doi.org/10.1038/s41587-019-0201-4> PMID: 31375807
88. Liao Y, Smyth GK, Shi W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res.* 2019; 47:e47. <https://doi.org/10.1093/nar/gkz114> PMID: 30783653
89. Heinken A, Magnúsdóttir S, Fleming RMT, Thiele I. DEMETER: efficient simultaneous curation of genome-scale reconstructions guided by experimental data and refined gene annotations. *Bioinformatics.* 2021;btab622. <https://doi.org/10.1093/bioinformatics/btab622> PMID: 34473240

90. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, et al. KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nat Biotechnol.* 2018; 36:566–569. <https://doi.org/10.1038/nbt.4163> PMID: 29979655
91. Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A, Heinken A, et al. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat Protoc.* 2019; 14:639–702. <https://doi.org/10.1038/s41596-018-0098-2> PMID: 30787451
92. King ZA, Lu J, Dräger A, Miller P, Federowicz S, Lerman JA, et al. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res.* 2016; 44:D515–D522. <https://doi.org/10.1093/nar/gkv1049> PMID: 26476456
93. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* 2014; 42:D206–D214. <https://doi.org/10.1093/nar/gkt1226> PMID: 24293654
94. Wohlgemuth G, Haldiya PK, Willighagen E, Kind T, Fiehn O. The Chemical Translation Service—a web-based tool to improve standardization of metabolomic reports. *Bioinformatics.* 2010; 26:2647–2648. <https://doi.org/10.1093/bioinformatics/btq476> PMID: 20829444
95. Szöcs E, Stirling T, Scott ER, Scharmüller A, Schäfer RB. webchem: An R Package to Retrieve Chemical Information from the Web. *J Stat Softw.* 2020;93. <https://doi.org/10.18637/jss.v093.i13>
96. Lechner M, Findeiß S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: Detection of (Co-) orthologs in large-scale analysis. *BMC Bioinformatics.* 2011; 12:124. <https://doi.org/10.1186/1471-2105-12-124> PMID: 21526987
97. Oksanen J, Simpson GL, Blanchet FG, Kindt R, Legendre P, Minchin PR, et al. vegan: Community Ecology Package. 2022. Available from: <https://CRAN.R-project.org/package=vegan>.
98. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb).* 2021; 2:100141. <https://doi.org/10.1016/j.xinn.2021.100141> PMID: 34557778
99. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York; 2016. Available from: <https://ggplot2.tidyverse.org>.
100. Wolfer A. Short Asynchronous Time-Series Analysis [R package santaR version 1.2.3]. 2022 [cited 2022 Jul 5]. Available from: <https://CRAN.R-project.org/package=santaR>.
101. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest Package: Tests in Linear Mixed Effects Models. *J Stat Softw.* 2017; 82:1–26. <https://doi.org/10.18637/jss.v082.i13>
102. Sprouffske K, Wagner A. Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves. *BMC Bioinformatics.* 2016; 17:172. <https://doi.org/10.1186/s12859-016-1016-7> PMID: 27094401
103. Lenth RV. emmeans: Estimated Marginal Means, aka Least-Squares Means. 2022. Available from: <https://CRAN.R-project.org/package=emmeans>.
104. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15. <https://doi.org/10.1186/s13059-014-0550-8> PMID: 25516281