Systems biology illuminates alternative metabolic niches in the human gut microbiome

Cecilia Noecker[1], Juan Sanchez[2], Jordan E. Bisanz[1+], Veronica Escalante[1], Margaret Alexander[1], Kai Trepka[1], Almut Heinken[3], Yuanyuan Liu[4], Dylan Dodd[4,5], Ines Thiele[3], Brian DeFelice[2], and Peter J. Turnbaugh[1,2*]

[1]Department of Microbiology & Immunology, University of California, San Francisco, San Francisco, CA 94143

[2]Chan-Zuckerberg Biohub, San Francisco, CA 94158

[3]School of Medicine, National University of Ireland, Galway, Ireland

[4]Department of Pathology, Stanford University, Stanford, CA 94305

[5]Department of Microbiology & Immunology, Stanford University, Stanford, CA 94305

[+]Present address: Department of Biochemistry and Molecular Biology, Pennsylvania State University, State College, PA

*Correspondence to: Peter J. Turnbaugh, Ph.D., peter.turnbaugh@ucsf.edu

SUMMARY

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.

KEYWORDS

Human gut microbiome; *Eggerthella lenta*; systems biology; metabolomics; stable isotope-resolved metabolomics; metabolic niche

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 difficult to culture (Lagkouvardos et al., 2017; Tramontano et al., 2018). 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 (Alexander et al., 2021; Medlock et al., 2018). 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* (Koropatkin et al., 2012) and *Clostridium sporogenes* (Liu et al., 2022), as well as large-scale efforts to characterize species-level metabolic activity using community multi-omic profiling (Franzosa et al.,

2018; Hertel et al., 2019). 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 (Almeida et al., 2019) 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 (Koppel et al., 2018). Although *E. lenta* is commonly found in healthy individuals, it can cause severe bacteremia (Gardiner et al., 2015) and is increased in abundance in the gut microbiota of patients with several autoimmune diseases (Cekanaviciute et al., 2017; Chen et al., 2016; Islam et al., 2021; Zhu et al., 2021).

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 (Bess et al., 2020; Devlin and Fischbach, 2015; Haiser et al., 2013; Koppel et al., 2018; Maini Rekdal et al., 2019). However, none of these compounds except dopamine have been reported to provide a growth or fitness advantage in any conditions tested to date. Genome analysis of *E. lenta* has also predicted that it may be able to perform autotrophic acetogenesis (Harris et al., 2018), 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 (Hylemon et al., 2018; Maini Rekdal et al., 2020; Tramontano et al., 2018), 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 two 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* (Tramontano et al., 2018). 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 (Noronha et al., 2018), *Methods*, **Table S1**]. 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; **Figure S1A-B**).

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 (**Figure 1A**). 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 > 3x blank sample mean, **Figure 1B**). 612 features (14.9% of features overall) were significantly different in abundance between sterile controls and supernatants at the final time point (FDR-adjusted p<0.1, **Figure 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 (Bisanz et al., 2020; Han et al., 2021) (**Figure S1C**). 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* (**Figure 1C-D**). 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 (Haiser et al., 2013). 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* (Beloborodov et al., 2009). 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, 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 three were depleted significantly in *E. lenta* cultures (**Figure S1D**, **Figure 1D**): arginine, riboflavin, and EDTA (which is likely reduced due to complexing with metal ions rather than from direct uptake or metabolism). This result suggested that

most compounds were included in excess, leading us to reduce the concentration of several non-depleted amino acids for subsequent experiments (**Table S1**). Interestingly, 5 of the identified metabolite features significantly depleted by *E. lenta* were not explicitly included in our defined media formulation, including guanine and five arginine dipeptides (**Figure 1D**). Since these compounds were found at low intensities, were annotated with high confidence, and are structurally related to intentionally included compounds, we inferred that they may be trace contaminants from commercial preparations of uracil and arginine (see *Methods*). Their rapid depletion 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 (**Figure 1D**, **Figure S1E**). This observation suggests that two 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 (**Figure 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 (*Methods*, **Table S2**). We collected growth curve data from EDM1 with and without each component and fit logistic growth models to the results, finding that 22 out of 41 compounds tested had a significant effect on at least one of the following growth parameters (Wilcoxon rank-sum test, FDR-adjusted $p<0.2$): carrying capacity (maximum density), growth rate, time to mid-exponential, and/or area under the growth curve (**Figure S2A**). 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 (**Figure S2B**).

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 (**Figure S2A**), 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 (van der Hee and Wells, 2021), dependence on acetate could shape the ecological interactions of *E. lenta* in the human gut microbiota. Although our untargeted LC-MS workflow was not able to quantify acetate, we had observed accumulation of several *N*-acetylated compounds in supernatant (**Figure 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 (**Figure 2A**). We therefore used a targeted derivatization and LC-MS/MS method to quantify acetate levels in supernatants from three 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 small quantity is required for the observed level of *E. lenta* growth (**Figure S3A**). 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 (**Figure 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 three *E. lenta* strains grown in EDM1 with different acetate concentrations (*E. lenta* DSM 2243 shown in **Figure 2C**, AB8n2 and Valencia in **Figure S3B-C**). 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-hydroxyglutaric acid, and 423 unidentified metabolite features (**Figure 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 three strain isolates of *E. lenta* with $^{13}C_2$ acetate provided as a stable isotope-labeled substrate (DSM 2243 in **Figures 2D-F**, 2 additional strains in **Figure S4**). We detected the incorporation of ^{13}C labeled atoms in 52 features in *E. lenta* supernatants at the final time point, of which 24 were previously identified as responsive to acetate concentrations (*Methods,* **Figures 2D**, **S4**). Acetate was incorporated into diverse products across metabolite classes, but was found at the highest enrichment levels in nucleotide and carbohydrate metabolites (**Data S1**).

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 (**Figure 2E**), as well as seven 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 two additional strains of *E. lenta*. While the overall rate of acetate incorporation differed between the three strains, the set of extracellular and intracellular labeled

compounds was fully consistent. Isotopic enrichment for two additional extracellular metabolites (malonic acid and 3-hydroxy-myristic acid) was identified in both of these strains as well as four additional intracellular metabolites in one or both strains (all of unknown identity), confirming that acetate is incorporated by *E. lenta* into varied biosynthetic pathways (**Figure S4**).

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 two routes: conversion to glutamate by a partial citric acid cycle, and synthesis of pyruvate by the enzyme pyruvate-ferredoxin oxidoreductase (PFOR, ELEN_RS10770) (**Figure 2F**). This hypothesis is consistent with the organization of the *E. lenta* DSM 2243 genome, as two of the three enzymes required for conversion of acetyl-CoA to glutamate are co-located (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 co-located in another region of the genome (ELEN_RS056[70-90]), suggesting 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 (**Data S1**). 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 (Sperry and Wilkins, 1976). We first confirmed that citrulline, but not ornithine, can replace arginine with nearly equivalent growth in EDM1, replicating a previous result in rich media [(Haiser et al., 2013), **Figure S5A**]. We then analyzed intracellular and extracellular metabolites from *E. lenta* DSM 2243 growing in EDM1, this time with $^{13}C_6$ 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 (**Figure S5B-E**), 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 (**Figure S5C-D**), suggesting biosynthesis from the carbamoyl phosphate intermediate. Labeled M+5 isotopologues of proline and prolinamide also appeared at low levels at later time points, likely indicating a slower flux producing these compounds from accumulated ornithine (**Figure S5D-E**). Yet in total, only 29/324 features were detected with ^{13}C enrichment for five or more carbon atoms in intracellular extracts, and most appeared closely related to arginine, citrulline, and ornithine (**Data S1**). These results confirm that arginine is primarily an energy source and not a major biosynthetic precursor for *E. lenta* (**Figure S5F**).

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

COnstraint-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 (Gu et al., 2019; Monk et al., 2017; Zhang et al., 2017). 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 (Chiu et al., 2014; Diener et al., 2020; Hertel et al., 2019). 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 (Heinken et al., 2020), 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 (**Table S3**). We curated the reconstruction based on genome annotations from multiple sources (Henry et al., 2010; Pascal Andreu et al., 2021; Price et al., 2022) 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 media concentrations into estimated maximum nutrient uptake rates for each compound. While these models are typically validated by comparison with gene essentiality data (Thiele and Palsson, 2010), 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.

This process resulted in a model with 1,244 reactions linked to 727 gene annotations and 1,218 metabolites (**Figure 3A**). The largest number of reactions were in the subsystems of fatty acid synthesis, extracellular transport, and glycerophospholipid biosynthesis (**Figure 3B**). Flux balance analysis of the final model estimated the maximum growth rate of *E. lenta* DSM 2243 in EDM1 to be 0.96 hr^{-1}, higher than experimental values (median 0.32 hr^{-1}, **Figure S2B**). The existence of a difference between these values is not surprising given that organisms do not necessarily grow at their theoretical maximum growth rate, and growth constraints may exist that are not encoded in the metabolic network model (Thiele and Palsson, 2010). However, the relatively large discrepancy indicates that additional modifications to the biomass equation may further improve the model.

The initial model with nonzero growth in EDM1 did not recapitulate the experimentally observed dependencies on either arginine or acetate (**Figure 3C**). We

329 noticed that this lack of dependency was linked to the inclusion of Wood-Ljungdahl
330 acetogenesis reactions in the model, previously suggested to be present in *E. lenta*
331 (Harris et al., 2018; Hylemon et al., 2018). The presence of these reactions allowed the
332 model to draw on an effectively unlimited source of acetyl-CoA from CO_2 and H_2.
333 Regardless of whether the previous annotation of this pathway (which has not been
334 biochemically validated) is correct, reductive acetogenesis may not be
335 thermodynamically favorable during *in vitro* growth in our anaerobic chamber, where the
336 H_2 concentration is ≤ 5% (Smith et al., 2020). Blocking model flux through the carbon
337 monoxide dehydrogenase reaction of this pathway increased growth dependency on
338 uptake of both arginine and acetate, reflecting our experimental observations (**Figure**
339 **3C**). The model also found no growth benefit from pyruvate, citrate, and other fatty acids
340 based on a lack of annotated transporters for these compounds, consistent with
341 experimental results.
342 In another key curation step, required to enable biomass production by the model
343 in EDM1, we noticed that *E. lenta* lacks an annotated gene for the enzyme enoyl-acyl
344 protein carrier reductase, which performs the elongation in the typical type 2 fatty acid
345 synthesis pathway used in bacteria. Because fatty acid biosynthesis is essential and
346 previous studies have noted a high level of diversity in this essential step among
347 bacterial genomes (Massengo-Tiassé and Cronan, 2009), we preserved this step in the
348 model without any current gene annotation. This gap may indicate a novel enzyme
349 family performing this conversion (**Table S3**).
350 We applied the iEL2243_2 model to predict growth phenotypes across our leave-
351 one-out chemically defined media conditions, finding that these were generally
352 consistent with some remaining notable exceptions (**Figure 3D**, overall Matthews
353 correlation of 0.35, Fisher exact test odds ratio=9.1, p=0.06). Amino acid dependencies
354 matched well between the model and experimental data, with the exception of cysteine,
355 which likely provides a benefit as a reducing agent that is not accounted for by the
356 model (Strobel, 2009). Vitamin dependencies were also generally consistent, with the
357 notable exception of folate, which had no effect on growth despite the lack of several
358 genes for reactions in the canonical folate biosynthesis pathway and the absence of a
359 known dihydrofolate reductase enzyme (Rodionov et al., 2019). The phenomenon of

presumed-essential but absent folate genes in bacterial genomes has been recognized previously (de Crécy-Lagard et al., 2007; Levin et al., 2004; Rodionov et al., 2019), suggesting the possible existence of undiscovered alternative enzymes. Notably, growth was negatively affected by the removal of the folate precursor *p*-aminobenzoate (**Figure S2A**). 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 (**Figure 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 (**Figure 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 (*Methods*). 71.9% of genes linked to active reactions were in the top half of metabolic genes by expression level in the EDM1 condition (> 109 transcripts per million), and 91.9% were in the top 75%. Expression level and absolute flux magnitude were highly correlated across all genes linked to metabolic reactions (Spearman rho=0.34, $p<2.2\text{x}10^{-16}$, **Figure 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, **Figure 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 1st, 3rd, 4th, and 5th most highly expressed protein-coding genes in our RNA-Seq data, **Table S4**). 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 (Maini Rekdal et al., 2020; Ravcheev and Thiele, 2014). 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*. 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 (**Figure S6A**). 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, **Figure S6B**) 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 (Bisanz et al., 2020). 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 (**Figure S7A-B**) and used linear models to identify features with significant strain-associated differences in abundance. Over half of the features produced by the UCSF DSM 2243 type strain (52.8%) were variable across strains of *E. lenta* (**Figure 4A**), and 1,097 features produced by at least two 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, **Figure 4B**), consistent with previous findings from untargeted metabolomics profiling of these strains in rich media with a different metabolomics platform (Bisanz et al., 2020). Overall metabolite profiles were moderately associated with presence/absence patterns of variable gene families between strains (p=0.03, **Figure 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.

While strain-variable metabolites were quite diverse, they were enriched for certain chemical classes. 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 (**Figure 4C**). The share of strains producing any individual feature varied widely (**Figure**

449 **4C**), although the largest number of features (76.8%) were produced by either only a
450 few (<4) strains or nearly all (>27) strains (**Figure S7C**).

451 Given the large share of unidentified metabolites in our dataset, we evaluated
452 whether linking strain-variable metabolites with strain-variable genes could inform
453 metabolite annotations. We performed an association analysis between metabolite
454 feature abundances and the presence of specific accessory gene families, applying a
455 method developed for previous analysis of this *E. lenta* strain collection (Bisanz et al.,
456 2020). A full 39.0% of metabolite features were significantly associated with the
457 presence of one or more variable gene families (FDR-adjusted $p<10^{-4}$). Using stricter
458 filtering criteria for significance, effect size, and separability, 84 metabolite features
459 (1.3%), of which 80 had no annotation, were linked with the presence of variable genes
460 (**Table S5**, *Methods*). Gene families linked to these features were enriched for KEGG
461 annotations in sulfur metabolism (q = 0.00017), ABC transporters (q = 0.02), porphyrin
462 metabolism (q = 0.03), and biosynthesis of nucleotide sugars (q = 0.049), consistent
463 with the profile of identified variable metabolites.

464 As a case study, we further examined two of the top hits from this analysis, two
465 closely related but unidentified metabolite features highly associated with the presence
466 of two adjacent gene families (**Figure 4D**). These gene families were annotated by
467 Prokka (Seemann, 2014) as ribulose-5-phosphate reductase 1 (*tarI*) and a ribitol-5-
468 phosphate cytidylyltransferase (*tarJ*), which are essential enzymes in the biosynthesis
469 of CDP-ribitol teichoic acid. Teichoic acids are an abundant component of the cell wall
470 of gram-positive bacteria that can take multiple forms and can be synthesized with
471 either CDP-glycerol or CDP-ribitol subunits (Brown et al., 2013; Percy and Gründling,
472 2014; Weidenmaier and Peschel, 2008). Interestingly, the *m/z* value and MS2 spectrum
473 of the linked features were consistent with an annotation as the two dominant [M+Cl]-
474 naturally occurring isotope adducts of a 5-carbon sugar alcohol - *i.e.* potentially ribitol,
475 xylitol, or a related compound.

476 Further examination of the *tar/tag* biosynthetic gene cluster in which these genes
477 are located revealed extensive strain diversity, with 10 different gene arrangements
478 across the 30 isolates (**Figure S7D**), suggesting recent positive selection possibly as a
479 form of phage defense (Buttimer et al., 2022; Soto-Perez et al., 2019) or host immune

interaction (van Dalen et al., 2020). 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 (Percy and Gründling, 2014). Among *E. lenta* genomes without *tarI* and *tarJ*, all except the type strain have a *tagD* gene in the same region instead, which catalyzes the synthesis of CDP-glycerol subunits instead of CDP-ribitol (**Figure S7D**) 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 (**Table S5**). This example illustrates that comparative multi-omics 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 (*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 (**Table S6**). 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 (**Figure S7E**). 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 (**Figure 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 (**Figure S7F**), 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 (**Figure 4D**). Notably, M+2 isotopologues of pantothenic acid were also detected in supernatants from the acetate SIRM experiment, corroborating that at least three *E. lenta* strains synthesize this vitamin *de novo* (**Data S1**, **Figure S4B, Figure 2F**). 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 (**Figure 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 (Monk et al., 2013).

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 three strains of *E. lenta* by oral gavage, collected serum and intestinal contents after two 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 vs. 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, 16.7% were significantly differentially abundant (FDR-adjusted

542 p<0.2) in response to colonization with at least one strain in at least one segment of the
543 intestinal tract, indicating a substantial metabolic impact of *E. lenta* on the intestinal
544 environment (**Figure S8A**). Interestingly, despite previous data showing colonization of
545 *E. lenta* DSM 2243 at similar levels from the ileum to the colon in GF mice (Bisanz et
546 al., 2020), only 1.6% of features were significantly shifted in the ileum, compared with
547 11.5% in colon and 8.1% in the cecum. Additionally, only 21 features (0.41%) were
548 differentially abundant in serum in response to any of the three strains. Overall
549 separability of metabolite profiles between germ-free and colonized was also highest in
550 the cecum and colon (**Figure 5A**). These results indicate that *E. lenta*'s strongest
551 metabolic effects are restricted to the lower intestinal tract.

552 We assessed the extent to which metabolite features produced by *E. lenta* in cell
553 culture are detectably shifted by the presence of *E. lenta* in mice. To do so, we
554 integrated our processed metabolomics datasets by linking metabolite features across
555 datasets with highly similar *m/z*, RT, and MS2 spectra (see *Methods*). Based on this
556 analysis, 37.2% of identified metabolite features in intestinal contents and 12.2% of
557 features overall were also detected *in vitro* (**Figure S8B**). We compared the estimated
558 $\log_2$ fold change of each linked feature *in vitro* with the corresponding shifts *in vivo* (full
559 set in **Data S2**; **Figure 5B** shows the comparison with the strain collection dataset in
560 **Figure 4**, **Figure S8C** shows a comparison with the dataset in **Figure 1**). 202 features
561 significantly increased by the presence of *E. lenta* DSM 2243 in cecal contents were
562 also increased in one of our EDM1 *in vitro* datasets, providing support that they are
563 directly produced by *E. lenta in vivo*. These features represented 78.9% of the set that
564 could be linked across datasets and 20.9% of the full set of *E. lenta* DSM 2243-
565 increased features in cecal contents. Only 18 metabolites depleted in cecal contents
566 were similarly depleted *in vitro*, but only three of the other 405 depleted features were
567 detected *in vitro* at all, indicating that *E. lenta* likely uses a much richer set of nutrients *in*
568 *vivo* than those available in EDM1. Overlapping produced and depleted metabolites
569 were found in greater abundance in the cecum and colon than the ileum and serum
570 (**Figure 5B**), again suggesting a greater metabolic footprint of *E. lenta* in the lower
571 gastrointestinal tract relative to other sites.

572 Ornithine was among the most increased features across sampling sites and
573 strains, consistent with our *in vitro* data (**Figure 6A**, **Figure S9A**). Other consistently
574 increased features included 5-methyluridine, citrulline, glutamine, and lysine (**Data S2**).
575 Interestingly, arginine was only significantly reduced by colonization with one of the
576 three *E. lenta* strains in this experiment (**Figure S9B**). However, most other
577 proteinogenic amino acids were increased in abundance in intestinal contents in
578 colonized mice compared with GF (**Figure S9C**), likely due to differences in host
579 activity, so the absence of an increase in arginine may be consistent with arginine
580 usage by *E. lenta*.
581 Given these results, we evaluated what other substrates may be used as carbon
582 or energy sources by *E. lenta in vivo*. The metabolites most strongly depleted by the
583 presence of *E. lenta* DSM 2243 in the intestinal tract included several fatty acids
584 conjugated with carnitine as well as multiple other nitrogen-containing metabolites:
585 saccharopine and agmatine (**Figure 6A**). We chose to investigate agmatine usage
586 further for several reasons: its chemical similarity to arginine, the presence of known
587 agmatine utilization genes in the *E. lenta* type strain genome, evidence of a consistent
588 decrease across all three strain colonization groups (**Figure S9A**), and its multiple roles
589 as a microbial metabolite and a host metabolite involved in regulation of cell division
590 and neural signaling (Piletz et al., 2013). The *E. lenta* DSM 2243 genome contains two
591 complete and two partial operons encoding genes for the agmatine deiminase pathway.
592 This pathway operates analogously to the arginine deiminase pathway, with ATP
593 production via carbamate kinase as the final step (**Figure 6B**). Despite this similarity,
594 the agmatine deiminase enzyme family is highly structurally distinct from arginine
595 deiminase (Llácer et al., 2007). Presence of this pathway is conserved across strains,
596 as other *E. lenta* genomes contain anywhere between one and four copies of the key
597 genes for agmatine deiminase and putrescine carbamoyltransferase (KEGG, **Table**
598 **S7A**). Additionally, a transcriptional regulator found adjacent to this operon in some
599 strains was previously associated with *E. lenta* competitive fitness *in vivo* (Bisanz et al.,
600 2020).
601 Based on these observations, we predicted that *E. lenta* may be able to grow in
602 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 vs. 0.96 hr^{-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 (**Figure 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 two copies of putrescine carbamoyltransferase, one copy of agmatine deiminase, and a transporter in the same operon (**Figure 6D** and **Table S7B**). Genes in the second complete agmatine deiminase operon (ELEN_RS110[05-15]) 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 downregulated genes were two 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 at > 50% prevalence in gut microbiota of North American adults (Koppel et al., 2018) and linked to acute and chronic disease (Alexander et al., 2021), 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 accurately 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 (Chodkowski and Shade, 2020; Dunphy et al., 2021; Pacheco et al., 2019). 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 (Venturelli et al., 2018). Interestingly, several small molecules produced by *E. lenta* in EDM1 and in mice are known to impact host immune signaling, including indole-3-acetate (Roager and Licht, 2018) and inosine (Li et al., 2021; Mager et al., 2020). 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 (van Dalen et al., 2020). While much focus has been deservedly paid to individual specialized immunomodulatory transformations performed by *E. lenta* (Alexander et al., 2021; Paik et al., 2022), our results suggest that *E. lenta*'s effects on host immunity may be multifaceted.

We elucidated the roles of three 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

665 carbon and/or energy source for numerous other gut microbes (Noronha et al., 2018),
666 including as a substrate for Stickland metabolism by gut bacteria including
667 *Clostridioides difficile* (Girinathan et al., 2021; Liu et al., 2022; Pruss et al., 2022).
668 Therefore, production of ornithine by *E. lenta* may promote the growth of other
669 proteolytic bacteria in the surrounding gut ecosystem.

670 We also found that the presence of acetate has a dramatic effect on *E. lenta*
671 growth and metabolism *in vitro.* Acetate is a ubiquitous microbial metabolite in the
672 mammalian gut that varies in concentration (van der Hee and Wells, 2021). Previous
673 studies have speculated that *E. lenta* may produce acetate via autotrophic acetogenesis
674 (Harris et al., 2018; Hylemon et al., 2018). While our study does not resolve the
675 question of whether *E. lenta* has a functional acetogenic Wood-Ljungdahl pathway, we
676 found that environmental acetate is an important biosynthetic precursor for *E. lenta*,
677 incorporated partially via a distinctive bifurcated citric acid cycle (Amador-Noguez et al.,
678 2010; Huynen et al., 1999). If *E. lenta* is in fact an acetate consumer *in vivo*, as we have
679 observed *in vitro*, this role may have ecological consequences. For example, *E. lenta*
680 may compete for cross-fed acetate with other gut microbes, including the abundant,
681 health-linked members of the Firmicutes that metabolize acetate to butyrate at high
682 rates (Duncan et al., 2002; Muñoz-Tamayo et al., 2011). However, while we did not
683 identify any compound that can replace the role of acetate in *E. lenta*'s metabolic
684 network, the observation that *E. lenta* can grow to high carrying capacities in rich media
685 and in germ-free mice presumably lacking acetate indicates that other undetermined
686 compounds may be able to serve as equivalent carbon sources.

687 Finally, we identified agmatine as an alternate energy source for *E. lenta in vivo.*
688 Agmatine is a host metabolite with multiple roles as a neurotransmitter, regulator of
689 nitric oxide synthesis, and regulator and precursor of polyamine metabolism (Piletz et
690 al., 2013). Although agmatine can be synthesized at low levels by the host, particularly
691 in the brain, the gastrointestinal tract is thought to be a major source of systemic
692 agmatine (Haenisch et al., 2008)—sourced either directly from the diet and/or from
693 microbial metabolism. Dietary sources of agmatine include a variety of plant and animal
694 products, with the highest known levels in fermented foods and alcoholic beverages
695 (Galgano et al., 2012). Altered agmatine levels have been associated with a range of

diseases, including depression and diabetes (Piletz et al., 2013). Notably, reduced agmatine levels in the gut have been linked to cell proliferation and cancer (Molderings et al., 2004). 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 (Heinken et al., 2021a), 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.

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.

ACKNOWLEDGEMENTS

We thank the UCSF Gnotobiotics Core staff (Jessie Turnbaugh, Kimberly Ly) for animal care support and assistance. This work was supported by the National Institutes of Health (2R01HL122593; 1R01AT011117; 1R01DK114034 to P.J.T., F32GM140808 to C.N.). P.J.T. is a Chan Zuckerberg Biohub Investigator and held an Investigators in the Pathogenesis of Infectious Disease Award from the Burroughs Wellcome Fund. We acknowledge SeqCenter for assistance with RNA-Seq library preparation and sequencing.

AUTHOR CONTRIBUTIONS

Conceptualization, C.N. and P.J.T.; Methodology, C.N., J.S., J.E.B., B.D., and P.J.T.; Software, C.N., A.H., and I.T.; Formal Analysis, C.N., J.S., K.T., A.H., Y.L., and B.D.; Investigation, C.N., J.S., J.E.B., V.E., M.A., Y.L., and B.D.; Data Curation, C.N., J.S., Y.L., D.D., and B.D; Writing - Original Draft, C.N.; Writing - Review & Editing, C.N., J.S., J.E.B., V.E., M.A., K.T., A.H., Y.L., D.D., I.T., B.D., and P.J.T.; Visualization, C.N.; Supervision, I.T., D.D., B.D., and P.J.T.

DECLARATION OF 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. All other authors have no relevant declarations.

FIGURE TITLES AND LEGENDS

Figure 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 3x 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 Metabolomics Standards Initiative confidence level for that identification (with 1 as highest confidence, see 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 *Eggerthella lenta* DSM 2243 in EDM1 (normalized OD600). See also **Figure S1-2**, **Table S1-2**.

Figure 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 with 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)** Stable isotope-resolved metabolomics profiling of *E. lenta* DSM 2243 in EDM1 media with $^{13}C_2$ 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)** Mass isotopologue distributions (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 one *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 **Figure S3-5**, **Data S1**.

Figure 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 parsimonious flux balance analysis (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 (Harris et al, 2018) (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. **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 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 flux variability analysis (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 **Figure S6**, **Table S3-4**.

Figure 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 strain) 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 with FDR-adjusted p-value<0.1 and log2FC>0.5) is consistent across 28 other *E. lenta* isolates and one 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 panel 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 panel shows the number of strains producing a given feature, 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. Superclasses (y-axis labels) are the same as in panel C. **E)** Feature abundances of example metabolites across strains. The first two panels show two strain-variable unidentified features associated with the presence of specific strain-variable gene families - putatively identified as the two 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 cytidylyltransferase (*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 three replicates. The order of strains on the y-axis matches their phylogeny, shown in **Figure S7A**. **F)** 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 panel). Closely related strains that do possess this gene are unaffected by removal of pantothenic acid (right panel). Carrying capacity is estimated based on a logistic growth model fit by the R package *growthcurver.* See also **Figure S7**, **Table S5-6.**

Figure 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 germ-free 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 stationary phase *in vitro* experiment. See also **Figure S8**, **Data S2.**

Figure 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 germ-free. 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 (Llácer et al., 2007). Three copies of an operon containing genes for all three of the labeled proteins 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 ± standard error for four 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 **Figure S9**, **Table S7.**

STAR METHODS

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to the Lead Contact Peter Turnbaugh (Peter.Turnbaugh@ucsf.edu).

Materials availability

This study does not contain newly generated materials.

Data and code availability

RNA sequencing data has been deposited in NCBI GEO and are publicly available as of the date of publication. Metabolomics datasets have been deposited in Metabolomics Workbench and are publicly available as of the date of publication. Processed metabolomics datasets, growth data, and metabolic reconstructions are available from Zenodo and are publicly available as of the date of publication. Accession numbers and DOIs are listed in the key resources table.

All original code has been deposited at Zenodo and GitHub (https://github.com/turnbaughlab/2022_Noecker_ElentaMetabolism) and is publicly available as of the date of publication. DOIs are listed in the key resources table.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mouse husbandry and experiments

Mouse samples analyzed in this study were collected and described previously in (Alexander et al., 2021). The mouse experiment was approved by the University of California San Francisco Institutional Animal Care and Use Committee. The mice were housed at temperatures ranging from 67-74°F and humidity ranging from 30-70% light/dark cycle 12hr/12hr. 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.

C57BL/6J mice (males, ages 4-8 weeks) were obtained from the University of California, San Francisco Gnotobiotics core facility (gnotobiotics.ucsf.edu) and housed in Iso positive cages (Tecniplast). Mice were colonized via oral gavage with *E. lenta* monocultures (10^9 CFU/mL, 200 µl gavage) and colonization was confirmed via anaerobic culturing and/or qPCR for an *E. lenta* specific marker (*elnmrk1*) (Bisanz et al., 2020; Koppel et al., 2018). Mice were colonized for 2 weeks prior to sacrifice and sample collection.

Bacterial strains

Strain isolates analyzed in this work are described in (Bisanz et al., 2020). All experiments were performed in an anaerobic chamber with 2-5% 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 **Table S1**. As specified in **Table S1**, 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, 30-1000x stock solutions were prepared following (Zhang et al., 2009)**.** Stock solutions were sterilized with a 0.22 µm syringe filter and stored at -20°C. Amino acids were typically added together directly from powder into a combined 2x stock solution which was then filter sterilized with a 0.22 µm 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. Individual replacement components are specified in **Table S1.** Media formulations were allowed to equilibrate in an anaerobic chamber (Coy) for at least 24 hours prior to use.

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 2-3 days. Individual colonies were inoculated into 3-4 mL liquid BHI+ and incubated at 37°C for 40-48 hours, or until approximately early stationary phase. Culture optical density (600 nm wavelength absorbance, OD600) was measured using a Hach DR1900 spectrophotometer. 1 mL samples of BHI starter cultures were then centrifuged at 1,568 rcf for 4 minutes in a microcentrifuge (ThermoScientific 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). 180 µL of defined media were pipetted into each well, followed by 20 µL of inoculum. All experiments included at least three sterile control wells for each condition, into which 20 µL 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. 3-6 replicates were included for each condition. 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).

Sample collection for metabolomics

Time course experiments were conducted in tubes in the anaerobic chamber in a 37°C incubator. For all metabolomics experiments, three 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. 5mLs 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. 100 µL from each tube were 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 $^{13}C_2$ 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 $^{13}C_6$ labeled L-arginine HCl (Sigma-Aldrich 643440).

For the comparative strain metabolomics experiment, 96-well polypropylene deep well plates were prepared with 800µL of fresh media in each well. Starter cultures and inocula for 29 isolates of *Eggerthella lenta* and 1 isolate of *Eggerthella sinensis* (Bisanz et al., 2020) 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 min 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 min 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 x 100 mm, 1.7-micron particle size, (Waters Corp. Milford, MA), using chromatographic conditions published elsewhere (Liu et al., 2022). 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 v.10.0 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 1.05mM of acetate was subtracted to obtain the final quantities.

A plate layout error for supernatant samples from time points 4-7 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 five 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^{o}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^{o}C, centrifuged at 4^{o}C at 6000 rcf for 5 min. The supernatant was transferred to a new plate and immediately sealed and kept at 4^{o}C prior to prompt analysis via LC-MS/MS.

Intestinal samples (colon, cecum, ileum) were prepared individually using a single protocol as follows. Samples were kept frozen on dry ice and massed to at least 10 mg. Four microliters of -20^{o}C extraction solvent (2:2:1 methanol:acetonitrile:water + stable isotope labeled internal standards) were added per milligram of intestinal sample. Six to eight 1mm 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^{o}C freezer, samples were centrifuged at 4^{o}C at 18,407 rcf for 5 minutes. Supernatant was collected and stored at -20^{o}C prior to centrifugal plate filtration (0.2 micron polyvinylidene difluoride (PVDF) Agilent Technologies, Santa Clara CA) at 4^{o}C at 4,122 rcf for 3 min. Collection plate was sealed and maintained at 4^{o}C prior to prompt analysis.

Serum samples were first thawed on wet ice. 20 µL 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^{o}C for 1 hour to maximize protein precipitation. After freezer incubation, samples were centrifuged at 4^{o}C at 18,407 rcf for 5 minutes. Supernatant was removed and dried under vacuum via centrivap (Labconco Corp.). 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^{o}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 x 150 mm, 1.7-micron particle size, (Waters Corp. Milford, MA), using chromatographic conditions published elsewhere (Lai et al., 2018). Samples were analyzed on a Thermo Q-Exactive HF orbitrap mass spectrometer operated utilizing data dependent acquisition of MS2. Data was acquired independently in positive and negative modes via subsequent injections.

SIRM metabolomics

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 10,000rcf for three minutes at 4°C, after which the supernatant was removed and the samples were immediately flash frozen to quench metabolites. 300 µL of cold methanol was then added to each pellet, followed by sonication on ice for 5 minutes and then shaking at 4°C for 4-12 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 Centrivap Benchtop Concentrator (Labconco Corp.). Samples were re-suspended in 60 µL of a chilled solution of 1:1 methanol and acetonitrile, with 24% water at -20^{o}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 metabolomics 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 metabolomics were identical to those used for standard untargeted metabolomics.

Untargeted metabolomics data processing

Untargeted metabolomics datasets were processed using MS-DIAL version 4.60 (Tsugawa et al., 2015). 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 (Han et al., 2021) and global (Mass Bank of North America) tandem mass spectral libraries. Annotation confidence scores were assigned based on Metabolomics Standards Initiative (MSI) best practices (Fiehn et al., 2007; Schymanski et al., 2014). 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. MSI level 4 is assigned 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. Post processing was performed using MS-FLO (DeFelice et al., 2017) for removal of erroneous features.

Processed datasets were further analyzed using Feature-based Molecular Networking and MolNetEnhancer in the GNPS web platform (Djoumbou Feunang et al., 2016; Ernst et al., 2019; Nothias et al., 2020; Wang et al., 2016), 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%. Log_{10}-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 non-zero value was added to the peak intensities for each feature before log transformation. Heatmaps of metabolite abundances were generated using the *ComplexHeatmap* package (Gu et al., 2016).

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 non-blank 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 three days. A single colony was then inoculated into 5mLs of BHI+ liquid culture and incubated at 37°C for 48 hours. 1 mL of the resulting culture was then centrifuged, washed once, and resuspended in equal volume PBS; all in anaerobic conditions. 220 µL of this inoculum were transferred into culture flasks containing 20 mL of EDM1 (70% carbon source reduced version, see **Table S1**) to obtain a starting OD600 of 0.01. 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 throughout the protocol. Briefly, 200 µL of TE buffer containing lysozyme (15mg/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. 700 µL of Buffer RLT 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 three minutes at 15,000rcf and 850 µL of supernatant was transferred to fresh tubes. 590 µL of 80% ethanol was added to each sample 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 10bp IDT for Illumina indices. Supplementary oligonucleotide probes specific to *E. lenta* rRNA and other highly expressed noncoding RNAs were incorporated during Ribo-Zero depletion (**Table S10**). Sequencing was done on a NextSeq 2000 with 2x50bp 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 (Chen et al., 2018) 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 (Kim et al., 2019) 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 (Liao et al., 2019), 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 (Bisanz et al., 2020) using the DEMETER pipeline (Heinken et al., 2020, 2021b). Briefly, DEMETER performs systematic refinement of a draft genome-scale reconstruction, in this case generated through KBase (Arkin et al., 2018). 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 *Eggerthella lenta* DSM 2243 reconstruction

underwent additional refinement of reactions and gene annotations against manually performed comparative genomics analyses (Heinken et al., 2020).

Reconstructions were analyzed using the Cobra Toolbox version 3.0 (Heirendt et al., 2019) in Matlab version 2018b, with the IBM Cplex solver version 128. Defined media concentrations were mapped from compound names to BiGG metabolite IDs (King et al., 2016) and converted to cell uptake rates over the duration of *E. lenta*'s exponential growth phase in batch culture (**Table S1**) using the *concToCellRate* function in the Cobra Toolbox and an approximate cell dry weight of $3.3x10^{-13}$ g, calculated based on colony forming units and dry biomass quantification from two 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/hr.

The collection of *E. lenta* strain reconstructions included two reconstructions of the type strain: the DSM 2243 reconstruction which had undergone additional comparative genomics curation with PubSeed (Overbeek et al., 2014), and a slightly smaller and less refined reconstruction included in the AGORA2 collection (Heinken et al., 2020) 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 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 (Seemann, 2014), GapMind (Price et al., 2022)) and/or by experimental growth or metabolomics data. We also performed several additional custom curations. Transporters were added for metabolites identified with high confidence (Metabolomics Standards Initiative 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 (**Figure 4**). Several pathways were also modified based on growth assay results and/or pathway annotation software (Price et al., 2022) and (Pascal Andreu et al., 2021)). 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 (Price et al., 2022) 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 **Table S3**.

Flux balance analysis (FBA), parsimonious flux balance analysis (pFBA), and flux variability analysis (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 **Figure 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 (Wohlgemuth et al., 2010) implemented in the R package *webchem* (Szöcs et al., 2020). KEGG IDs were then mapped to BiGG IDs using the BiGG database (King et al., 2016) 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 three 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 knock-out 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 non-clonal *E. lenta* genomes

(Bisanz et al., 2020). In this previous analysis, amino acid sequence families were clustered across genomes using ProteinOrtho (Lechner et al., 2011) 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 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 (Bisanz et al., 2020).

Procrustes analysis was performed using the R package *vegan* v2.6-2, with evaluation of significance using the *protest* function (Oksanen et al., 2022). 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 (Bisanz et al., 2020), 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 two 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 (Wu et al., 2021) 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 two 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 (Wickham, 2016). Statistical tests, sample size and standard error are reported in the figures and figure legends. Benjamini-Hochberg

false discovery rate (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 (Wolfer, 2022). 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. Differential abundance analysis for the gnotobiotic mouse intestinal dataset was performed using linear mixed models with the R package *lmerTest* (Kuznetsova et al., 2017), incorporating fixed effects for intestinal site, colonization group, and the interaction between them; and nested random effects for cage and animal. The *difflsmeans* function with Benjamini-Hochberg multiple hypothesis adjustment was used to evaluate the statistical significance of differences of each colonization group vs. germ-free 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* (Sprouffske and Wagner, 2016). 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 (Lenth, 2022).

RNA-Seq statistical analysis

Differential expression analysis was performed using negative binomial generalized linear models implemented in the *DESeq2* package v1.32.0 (Love et al., 2014).

SUPPLEMENTAL INFORMATION TITLES AND LEGENDS

Supplemental Figures

Figure S1. EDM1 chemically defined media supports robust growth of *Eggerthella lenta* and enables sensitive metabolomics profiling. Related to Figure 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 two commonly used media conditions (Brain Heart Infusion supplemented with L-arginine, and ISP-2 media supplemented with L-arginine), compared with three 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.* **D)** Metabolomics profiling of compounds known to be present in the chemically defined media formulation EDM1. 22 media compounds were detected, most of which were not significantly depleted in *E. lenta* cultures over time. **E)** 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 Metabolomics Standards Initiative 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.

Figure S2. Effects of individual media components on growth of *E. lenta* DSM 2243. Related to Figure 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 +/- standard error. Blue text indicates the growth parameters with significantly different values with and without the compound (Wilcoxon rank-sum test, FDR-adjusted $p < 0.2$; r - growth rate k

- carrying capacity, *tmid* - time to mid-exponential, *auc* - area under the empirical curve). **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*.

Figure S3. Environmental acetate concentrations affect growth and metabolite production of three *E. lenta* strains. Related to Figure 2. **A)** Targeted quantification of acetate depletion in E. lenta EDM1 cultures. Acetate was measured at 2-3 time points in supernatant samples from three *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 +/- 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 three *E. lenta* strain isolates in EDM1 with 0, 1, or 10 mM sodium acetate. Mean +/- standard error across three 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 Metabolomics Standards Initiative confidence level for each metabolite annotation (see *Methods*).

Figure S4. Consistent incorporation of acetate across three *E. lenta* strains based on stable isotope-resolved metabolomics. Related to Figure 2. **A)** Growth of *E. lenta* strains in EDM1 with varying levels of sodium acetate (either stable isotope-labeled $^{13}C_2$ or unlabeled). Optical density measurements were taken and supernatant samples were collected at each indicated time point. Mean +/- standard error across

three replicates is shown. **B)** Average trajectories of labeled extracellular metabolites in three different strains of *E. lenta*. Metabolites shown are those with > 50% and > $5x10^4$ average peak area from labeled isotopologues in at least one 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 Metabolomics Standards Initiative annotation confidence level for each metabolite. **C)** Labeled metabolites of known identity in intracellular extracts across three strains of *E. lenta* (data for DSM 2243 matches **Figure 2E**). Each panel shows the average mass isotopologue distribution across three 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 Metabolomics Standards Initiative 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 **Data S1**. The isotopologue color legend is the same as in panel B. **D)** Labeled metabolites of unknown identity across three strains of *E. lenta*. Each panel shows the average mass isotopologue distribution (across three 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.

Figure S5. Stable isotope profiling of *E. lenta* arginine metabolism confirms that arginine is primarily converted to ornithine as an energy source. Related to Figure 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 +/- standard error across four replicates. **B)** In *E. lenta* DSM 2243 cultures grown with 1% $^{13}C_6$ labeled arginine, correspondingly labeled citrulline and ornithine accumulate in supernatants over the course of growth. Curves

show mean +/- standard error across three 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 one isotopologue with a peak area of greater 10^6 in at least one time point, and have a labeled isotopologue with >3% abundance in at least one time point. **D)** Mass isotopologue distributions of intracellular metabolites. Each barplot shows the mean peak areas of isotopologues for each feature at two 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 one 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 panel), compounds with M+5 labeling patterns are mainly restricted to ornithine, citrulline, and a compound of unknown identity (middle panel), and compounds found with high signal as M+6 isotopologues are mainly arginine and citrulline (right-hand panel). 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.

Figure S6. Single-reaction knockout analysis of iEL2243_2 identifies conserved genes across metabolic subsystems. Related to Figure 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 wildtype 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.

Figure S7. Within-species variation in *E. lenta* metabolic profiles across genomes and metabolomes. Related to Figure 4. **A)** Phylogeny of 30 *Eggerthella* strains analyzed in this study. This phylogeny was previously constructed based on core gene alignments using Phylophlan (Bisanz et al., 2020). **B)** Principal components analysis (PCA) of log-transformed metabolite intensity profiles of stationary phase supernatants from 30 *Eggerthella* isolates in EDM1. 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 one 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 one *Eggerthella* isolate in this experiment (FDR-adjusted *p*-value<0.1 and $\log_2$ fold change>0.5). **D)** 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 **Figure 4E**. Gene regions were defined in each genome based on the location of the genes annotated as *tagG* and *tagH* by Prokka. **E)** Distribution of core and accessory reactions across subsystems, based on comparative analysis of metabolic reconstructions of 24 *E. lenta* strain genomes. **F)** 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.

Figure S8. Differential abundance analysis of intestinal and serum metabolites of *E. lenta*-monocolonized mice compared to germ-free. Related to Figure 5. **A)** Volcano plots of differential abundance analysis of metabolite features in intestinal contents and serum of gnotobiotic mice monocolonized with one of three *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 two *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 germ-free 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.

Figure S9. Shifts in intestinal amino acid metabolites of *E. lenta*-monocolonized mice compared to germ-free. Related to Figure 6. **A)** Annotated metabolites with the largest shifts in intestinal contents of *E. lenta*-colonized mice compared with germ-free. Metabolites are shown if they were identified based on library comparison and were among the most 600 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.

Supplemental Tables

Table S1. Chemically defined media formulations used in this study. Related to Figure 1 and STAR Methods. Recipes used for preparation of chemically defined media used for experiments in this study. The first two 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.

Table S2. Summarized results of media leave-one-out growth experiments. Related to Figure 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.

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

Table S4. Most highly expressed genes by *E. lenta* DSM 2243 during growth in EDM1. Related to Figure 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.

Table S5. Metabolite features associated with variable *E. lenta* gene families across strains. Related to Figure 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 *Methods*). Gene annotations are listed for association patterns with 20 or fewer candidate gene families.

Table S6. Summary of conserved and strain-variable reactions by subsystem in *E. lenta* strain metabolic reconstructions. Related to Figure 4. Statistics on the distribution of core and accessory reactions across *E. lenta* strain metabolic reconstructions.

Table S7. Genes linked to agmatine utilization by *E. lenta*. Related to Figure 6. A) KEGG annotations of gene families in the agmatine deiminase pathway in *E. lenta* genomes, as previously obtained using GhostKoala. **B)** *E. lenta* DSM 2243 genes with differential expression in response to agmatine sulfate treatment (FDR-adjusted p<0.1 and absolute log2 fold change>1), as estimated by DESeq2.

Table S8. Supplementary oligonucleotide probes used for depletion of highly abundant *E. lenta* noncoding RNAs. Related to STAR 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.

Supplemental Datasets

Data S1. Labeled features detected in stable isotope experiments. Related to Figure 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 **Figures 2**, **S6**, and **S7.**

Data S2. Differentially abundant features across *in vivo* and *in vitro* untargeted metabolomics datasets. Related to Figure 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 at least one 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.

REFERENCES

Alexander, M., Ang, Q.Y., Nayak, R.R., Bustion, A.E., Sandy, M., Zhang, B., Upadhyay, V., Pollard, K.S., Lynch, S.V., and Turnbaugh, P.J. (2021). Human gut bacterial metabolism drives Th17 activation and colitis. Cell Host Microbe *0*. https://doi.org/10.1016/j.chom.2021.11.001.

Almeida, A., Mitchell, A.L., Boland, M., Forster, S.C., Gloor, G.B., Tarkowska, A., Lawley, T.D., and Finn, R.D. (2019). A new genomic blueprint of the human gut microbiota. Nature https://doi.org/10.1038/s41586-019-0965-1.

Amador-Noguez, D., Feng, X.-J., Fan, J., Roquet, N., Rabitz, H., and Rabinowitz, J.D. (2010). Systems-level metabolic flux profiling elucidates a complete, bifurcated tricarboxylic acid cycle in Clostridium acetobutylicum. J. Bacteriol. *192*, 4452–4461.

Arkin, A.P., Cottingham, R.W., Henry, C.S., Harris, N.L., Stevens, R.L., Maslov, S., Dehal, P., Ware, D., Perez, F., Canon, S., et al. (2018). KBase: The United States Department of Energy Systems Biology Knowledgebase. Nat. Biotechnol. *36*, 566–569.

Beloborodov, N.V., Khodakova, A.S., Bairamov, I.T., and Olenin, A.Y. (2009). Microbial origin of phenylcarboxylic acids in the human body. Biochemistry *74*, 1350–1355.

Bess, E.N., Bisanz, J.E., Yarza, F., Bustion, A., Rich, B.E., Li, X., Kitamura, S., Waligurski, E., Ang, Q.Y., Alba, D.L., et al. (2020). Genetic basis for the cooperative bioactivation of plant lignans by Eggerthella lenta and other human gut bacteria. Nat Microbiol *5*, 56–66.

Bisanz, J.E., Soto-Perez, P., Noecker, C., Aksenov, A.A., Lam, K.N., Kenney, G.E., Bess, E.N., Haiser, H.J., Kyaw, T.S., Yu, F.B., et al. (2020). A Genomic Toolkit for the Mechanistic Dissection of Intractable Human Gut Bacteria. Cell Host Microbe 1–14.

Brown, S., Santa Maria, J.P., Jr, and Walker, S. (2013). Wall teichoic acids of gram-positive bacteria. Annu. Rev. Microbiol. *67*, 313–336.

Buttimer, C., Bottacini, F., Shkoporov, A.N., Draper, L.A., Ross, P., and Hill, C. (2022). Selective Isolation of Eggerthella lenta from Human Faeces and Characterisation of the Species Prophage Diversity. Microorganisms *10*, 195.

Cekanaviciute, E., Yoo, B.B., Runia, T.F., Debelius, J.W., Singh, S., Nelson, C.A., Kanner, R., Bencosme, Y., Lee, Y.K., Hauser, S.L., et al. (2017). Gut bacteria from multiple sclerosis patients modulate human T cells and exacerbate symptoms in mouse models. Proc. Natl. Acad. Sci. U. S. A. *114*, 10713–10718.

Chen, J., Wright, K., Davis, J.M., Jeraldo, P., Marietta, E.V., Murray, J., Nelson, H., Matteson, E.L., and Taneja, V. (2016). An expansion of rare lineage intestinal microbes characterizes rheumatoid arthritis. Genome Med. *8*, 43.

1739 Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ
1740 preprocessor. Bioinformatics *34*, i884–i890.

1741 Chiu, H.-C., Levy, R., and Borenstein, E. (2014). Emergent Biosynthetic Capacity in
1742 Simple Microbial Communities. PLoS Comput. Biol. *10*, e1003695.

1743 Chodkowski, J.L., and Shade, A. (2020). Exometabolite Dynamics over Stationary
1744 Phase Reveal Strain-Specific Responses. mSystems *5*, e00493–20.

1745 de Crécy-Lagard, V., El Yacoubi, B., de la Garza, R.D., Noiriel, A., and Hanson, A.D.
1746 (2007). Comparative genomics of bacterial and plant folate synthesis and salvage:
1747 predictions and validations. BMC Genomics *8*, 245.

1748 van Dalen, R., Peschel, A., and van Sorge, N.M. (2020). Wall Teichoic Acid in
1749 Staphylococcus aureus Host Interaction. Trends Microbiol. *28*, 985–998.

1750 DeFelice, B.C., Mehta, S.S., Samra, S., Čajka, T., Wancewicz, B., Fahrmann, J.F., and
1751 Fiehn, O. (2017). Mass Spectral Feature List Optimizer (MS-FLO): A tool to minimize
1752 false positive peak reports in untargeted liquid chromatography-mass spectroscopy (LC-
1753 MS) data processing. Anal. Chem. *89*, 3250–3255.

1754 Devlin, A.S., and Fischbach, M.A. (2015). A biosynthetic pathway for a prominent class
1755 of microbiota-derived bile acids. Nat. Chem. Biol. *11*, 685–690.

1756 Diener, C., Gibbons, S.M., and Resendis-Antonio, O. (2020). MICOM: Metagenome-
1757 Scale Modeling To Infer Metabolic Interactions in the Gut Microbiota. mSystems *5*,
1758 e00606–e00619, /msystems/5/1/msys.00606–00619.atom.

1759 Djoumbou Feunang, Y., Eisner, R., Knox, C., Chepelev, L., Hastings, J., Owen, G.,
1760 Fahy, E., Steinbeck, C., Subramanian, S., Bolton, E., et al. (2016). ClassyFire:
1761 automated chemical classification with a comprehensive, computable taxonomy. J.
1762 Cheminform. *8*, 61.

1763 Duncan, S.H., Barcenilla, A., Stewart, C.S., Pryde, S.E., and Flint, H.J. (2002). Acetate
1764 Utilization and Butyryl Coenzyme A (CoA):Acetate-CoA Transferase in Butyrate-
1765 Producing Bacteria from the Human Large Intestine. Appl. Environ. Microbiol. *68*, 5186–
1766 5190.

1767 Dunphy, L.J., Grimes, K.L., Wase, N., Kolling, G.L., and Papin, J.A. (2021). Untargeted
1768 Metabolomics Reveals Species-Specific Metabolite Production and Shared Nutrient
1769 Consumption by Pseudomonas aeruginosa and Staphylococcus aureus. mSystems *6*.
1770 https://doi.org/10.1128/mSystems.00480-21.

1771 Ernst, M., Kang, K.B., Caraballo-Rodríguez, A.M., Nothias, L.-F., Wandy, J., Chen, C.,
1772 Wang, M., Rogers, S., Medema, M.H., Dorrestein, P.C., et al. (2019). MolNetEnhancer:
1773 Enhanced Molecular Networks by Integrating Metabolome Mining and Annotation Tools.
1774 Metabolites *9*. https://doi.org/10.3390/metabo9070144.

1775 Fiehn, O., Robertson, D., Griffin, J., and van der Werf, M. (2007). The metabolomics
1776 standards initiative (MSI). Metabolomics.

1777 Franzosa, E.A., Sirota-Madi, A., Avila-Pacheco, J., Fornelos, N., Haiser, H.J., Reinker,
1778 S., Vatanen, T., Hall, A.B., Mallick, H., McIver, L.J., et al. (2018). Gut microbiome
1779 structure and metabolic activity in inflammatory bowel disease. Nature Microbiology
1780 https://doi.org/10.1038/s41564-018-0306-4.

1781 Galgano, F., Caruso, M., Condelli, N., and Favati, F. (2012). Focused review: agmatine
1782 in fermented foods. Front. Microbiol. *3*, 199.

1783 Gardiner, B.J., Tai, A.Y., Kotsanas, D., Francis, M.J., Roberts, S.A., Ballard, S.A.,
1784 Junckerstorff, R.K., and Korman, T.M. (2015). Clinical and Microbiological
1785 Characteristics of Eggerthella lenta Bacteremia. J. Clin. Microbiol. *53*, 626–635.

1786 Girinathan, B.P., DiBenedetto, N., Worley, J.N., Peltier, J., Arrieta-Ortiz, M.L.,
1787 Immanuel, S.R.C., Lavin, R., Delaney, M.L., Cummins, C.K., Hoffman, M., et al. (2021).
1788 In vivo commensal control of Clostridioides difficile virulence. Cell Host Microbe *29*,
1789 1693–1708.e7.

1790 Gu, C., Kim, G.B., Kim, W.J., Kim, H.U., and Lee, S.Y. (2019). Current status and
1791 applications of genome-scale metabolic models. Genome Biol. *20*, 121.

1792 Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and
1793 correlations in multidimensional genomic data. Bioinformatics *32*, 2847–2849.

1794 Haenisch, B., von Kügelgen, I., Bönisch, H., Göthert, M., Sauerbruch, T., Schepke, M.,
1795 Marklein, G., Höfling, K., Schröder, D., and Molderings, G.J. (2008). Regulatory
1796 mechanisms underlying agmatine homeostasis in humans. Am. J. Physiol. Gastrointest.
1797 Liver Physiol. *295*, G1104–G1110.

1798 Haiser, H.J., Gootenberg, D.B., Chatman, K., Sirasani, G., Balskus, E.P., and
1799 Turnbaugh, P.J. (2013). Predicting and Manipulating Cardiac Drug Inactivation by the
1800 Human Gut Bacterium Eggerthella lenta. Science *341*, 295–298.

1801 Han, S., Van Treuren, W., Fischer, C.R., Merrill, B.D., DeFelice, B.C., Sanchez, J.M.,
1802 Higginbottom, S.K., Guthrie, L., Fall, L.A., Dodd, D., et al. (2021). A metabolomics
1803 pipeline for the mechanistic interrogation of the gut microbiome. Nature *595*, 415–420.

1804 Harris, S.C., Devendran, S., Méndez- García, C., Mythen, S.M., Wright, C.L., Fields,
1805 C.J., Hernandez, A.G., Cann, I., Hylemon, P.B., and Ridlon, J.M. (2018). Bile acid
1806 oxidation by *Eggerthella lenta* strains C592 and DSM 2243 ^{T}. Gut Microbes 1–17.

1807 van der Hee, B., and Wells, J.M. (2021). Microbial Regulation of Host Physiology by
1808 Short-chain Fatty Acids. Trends Microbiol. *29*, 700–712.

1809 Heinken, A., Acharya, G., Ravcheev, D.A., Hertel, J., Nyga, M., Okpala, O.E., Hogan,
1810 M., Magnúsdóttir, S., Martinelli, F., Preciat, G., et al. (2020). AGORA2: Large scale

1811 reconstruction of the microbiome highlights wide-spread drug-metabolising capacities
1812 (Systems Biology).

1813 Heinken, A., Basile, A., and Thiele, I. (2021a). Advances in constraint-based modelling
1814 of microbial communities. Current Opinion in Systems Biology *27*, 100346.

1815 Heinken, A., Magnúsdóttir, S., Fleming, R.M.T., and Thiele, I. (2021b). DEMETER:
1816 efficient simultaneous curation of genome-scale reconstructions guided by experimental
1817 data and refined gene annotations. Bioinformatics btab622.

1818 Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S.N., Richelle, A., Heinken, A.,
1819 Haraldsdóttir, H.S., Wachowiak, J., Keating, S.M., Vlasov, V., et al. (2019). Creation and
1820 analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. Nat.
1821 Protoc. *14*, 639–702.

1822 Henry, C.S., DeJongh, M., Best, A.A., Frybarger, P.M., Linsay, B., and Stevens, R.L.
1823 (2010). High-throughput generation, optimization and analysis of genome-scale
1824 metabolic models. Nat. Biotechnol. *28*, 977–982.

1825 Hertel, J., Harms, A.C., Heinken, A., Baldini, F., Thinnes, C.C., Glaab, E., Vasco, D.A.,
1826 Pietzner, M., Stewart, I.D., Wareham, N.J., et al. (2019). Integrated Analyses of
1827 Microbiome and Longitudinal Metabolome Data Reveal Microbial-Host Interactions on
1828 Sulfur Metabolism in Parkinson's Disease. Cell Rep. *29*, 1767–1777.e8.

1829 Huynen, M.A., Dandekar, T., and Bork, P. (1999). Variation and evolution of the citric-
1830 acid cycle: a genomic perspective. Trends Microbiol. *7*, 281–291.

1831 Hylemon, P.B., Harris, S.C., and Ridlon, J.M. (2018). Metabolism of hydrogen gases
1832 and bile acids in the gut microbiome. FEBS Lett. *592*, 2070–2082.

1833 Islam, M.Z., Tran, M., Xu, T., Tierney, B.T., Patel, C., and Kostic, A.D. (2021).
1834 Reproducible and Opposing Microbiome Signatures Distinguish Autoimmune Diseases
1835 and Cancers: A Systematic Review and Meta-Analysis (In Review).

1836 Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based
1837 genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol.
1838 *37*, 907–915.

1839 King, Z.A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J.A., Ebrahim, A.,
1840 Palsson, B.O., and Lewis, N.E. (2016). BiGG Models: A platform for integrating,
1841 standardizing and sharing genome-scale models. Nucleic Acids Res. *44*, D515–D522.

1842 Koppel, N., Bisanz, J.E., Pandelia, M.-E., Turnbaugh, P.J., and Balskus, E.P. (2018).
1843 Discovery and characterization of a prevalent human gut bacterial enzyme sufficient for
1844 the inactivation of a family of plant toxins. Elife *7*, e33953.

1845 Koropatkin, N.M., Cameron, E.A., and Martens, E.C. (2012). How glycan metabolism
1846 shapes the human gut microbiota. Nat. Rev. Microbiol. *10*, 323–335.

1847 Kuznetsova, A., Brockhoff, P.B., and Christensen, R.H.B. (2017). lmerTest Package:
1848 Tests in Linear Mixed Effects Models. J. Stat. Softw. *82*, 1–26.

1849 Lagkouvardos, I., Overmann, J., and Clavel, T. (2017). Cultured microbes represent a
1850 substantial fraction of the human and mouse gut microbiota. Gut Microbes *8*, 493–503.

1851 Lechner, M., Findeiß, S., Steiner, L., Marz, M., Stadler, P.F., and Prohaska, S.J. (2011).
1852 Proteinortho: Detection of (Co-)orthologs in large-scale analysis. BMC Bioinformatics
1853 *12*, 124.

1854 Lenth, R.V. (2022). emmeans: Estimated Marginal Means, aka Least-Squares Means.

1855 Levin, I., Giladi, M., Altman-Price, N., Ortenberg, R., and Mevarech, M. (2004). An
1856 alternative pathway for reduced folate biosynthesis in bacteria and halophilic archaea.
1857 Mol. Microbiol. *54*, 1307–1318.

1858 Li, D., Feng, Y., Tian, M., Ji, J., Hu, X., and Chen, F. (2021). Gut microbiota-derived
1859 inosine from dietary barley leaf supplementation attenuates colitis through PPARγ
1860 signaling activation. Microbiome *9*, 83.

1861 Liao, Y., Smyth, G.K., and Shi, W. (2019). The R package Rsubread is easier, faster,
1862 cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic
1863 Acids Res. *47*, e47.

1864 Liu, Y., Chen, H., Van Treuren, W., Hou, B.-H., Higginbottom, S.K., and Dodd, D.
1865 (2022). Clostridium sporogenes uses reductive Stickland metabolism in the gut to
1866 generate ATP and produce circulating metabolites. Nat Microbiol *7*, 695–706.

1867 Llácer, J.L., Polo, L.M., Tavárez, S., Alarcón, B., Hilario, R., and Rubio, V. (2007). The
1868 Gene Cluster for Agmatine Catabolism of *Enterococcus faecalis* : Study of Recombinant
1869 Putrescine Transcarbamylase and Agmatine Deiminase and a Snapshot of Agmatine
1870 Deiminase Catalyzing Its Reaction. Journal of Bacteriology *189*, 1254–1265.
1871 https://doi.org/10.1128/jb.01216-06.

1872 Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and
1873 dispersion for RNA-seq data with DESeq2. Genome Biol. *15*.
1874 https://doi.org/10.1186/s13059-014-0550-8.

1875 Mager, L.F., Burkhard, R., Pett, N., Cooke, N.C.A., Brown, K., Ramay, H., Paik, S.,
1876 Stagg, J., Groves, R.A., Gallo, M., et al. (2020). Microbiome-derived inosine modulates
1877 response to checkpoint inhibitor immunotherapy. Science eabc3421.

1878 Maini Rekdal, V., Bess, E.N., Bisanz, J.E., Turnbaugh, P.J., and Balskus, E.P. (2019).
1879 Discovery and inhibition of an interspecies gut bacterial pathway for Levodopa
1880 metabolism. Science *364*, eaau6323.

1881 Maini Rekdal, V., Nol Bernadino, P., Luescher, M.U., Kiamehr, S., Le, C., Bisanz, J.E.,
1882 Turnbaugh, P.J., Bess, E.N., and Balskus, E.P. (2020). A widely distributed

metalloenzyme class enables gut microbial metabolism of host- and diet-derived catechols. Elife *9*. https://doi.org/10.7554/eLife.50845.

Massengo-Tiassé, R.P., and Cronan, J.E. (2009). Diversity in enoyl-acyl carrier protein reductases. Cell. Mol. Life Sci. *66*, 1507–1517.

Medlock, G.L., Carey, M.A., McDuffie, D.G., Mundy, M.B., Giallourou, N., Swann, J.R., Kolling, G.L., and Papin, J.A. (2018). Inferring Metabolic Mechanisms of Interaction within a Defined Gut Microbiota. Cell Systems *7*, 245–257.e7.

Molderings, G.J., Kribben, B., Heinen, A., Schröder, D., Brüss, M., and Göthert, M. (2004). Intestinal tumor and agmatine (decarboxylated arginine): low content in colon carcinoma tissue specimens and inhibitory effect on tumor cell proliferation in vitro. Cancer *101*, 858–868.

Monk, J.M., Charusanti, P., Aziz, R.K., Lerman, J.A., Premyodhin, N., Orth, J.D., Feist, A.M., and Palsson, B.O. (2013). Genome-scale metabolic reconstructions of multiple Escherichia coli strains highlight strain-specific adaptations to nutritional environments. Proceedings of the National Academy of Sciences *110*, 20338–20343.

Monk, J.M., Lloyd, C.J., Brunk, E., Mih, N., Sastry, A., King, Z., Takeuchi, R., Nomura, W., Zhang, Z., Mori, H., et al. (2017). iML1515, a knowledgebase that computes Escherichia coli traits. Nat. Biotechnol. *35*, 904–908.

Muñoz-Tamayo, R., Laroche, B., Walter, E., Doré, J., Duncan, S.H., Flint, H.J., and Leclerc, M. (2011). Kinetic modelling of lactate utilization and butyrate production by key human colonic bacterial species. FEMS Microbiol. Ecol. *76*, 615–624.

Noronha, A., Modamio, J., Jarosz, Y., Guerard, E., Sompairac, N., Preciat, G., Daníelsdóttir, A.D., Krecke, M., Merten, D., Haraldsdóttir, H.S., et al. (2018). The Virtual Metabolic Human database: integrating human and gut microbiome metabolism with nutrition and disease. Nucleic Acids Res. gky992–gky992.

Nothias, L.-F., Petras, D., Schmid, R., Dührkop, K., Rainer, J., Sarvepalli, A., Protsyuk, I., Ernst, M., Tsugawa, H., Fleischauer, M., et al. (2020). Feature-based molecular networking in the GNPS analysis environment. Nat. Methods *17*, 905–908.

Oksanen, J., Simpson, G.L., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Solymos, P., Stevens, M.H.H., Szoecs, E., et al. (2022). vegan: Community Ecology Package.

Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., Shukla, M., et al. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. *42*, D206–D214.

Pacheco, A.R., Moel, M., and Segrè, D. (2019). Costless metabolic secretions as drivers of interspecies interactions in microbial ecosystems. Nat. Commun. *10*, 103.

Paik, D., Yao, L., Zhang, Y., Bae, S., D'Agostino, G.D., Zhang, M., Kim, E., Franzosa, E.A., Avila-Pacheco, J., Bisanz, J.E., et al. (2022). Human gut bacteria produce TH17-modulating bile acid metabolites. Nature *603*, 907–912.

Pascal Andreu, V., Roel-Touris, J., Dodd, D., Fischbach, M.A., and Medema, M.H. (2021). The gutSMASH web server: automated identification of primary metabolic gene clusters from the gut microbiota. Nucleic Acids Res. gkab353.

Percy, M.G., and Gründling, A. (2014). Lipoteichoic acid synthesis and function in gram-positive bacteria. Annu. Rev. Microbiol. *68*, 81–100.

Piletz, J.E., Aricioglu, F., Cheng, J.-T., Fairbanks, C.A., Gilad, V.H., Haenisch, B., Halaris, A., Hong, S., Lee, J.E., Li, J., et al. (2013). Agmatine: clinical applications after 100 years in translation. Drug Discov. Today *18*, 880–893.

Price, M.N., Deutschbauer, A.M., and Arkin, A.P. (2022). Filling gaps in bacterial catabolic pathways with computation and high-throughput genetics. PLoS Genet. *18*, e1010156.

Pruss, K.M., Enam, F., Battaglioli, E., DeFeo, M., Diaz, O.R., Higginbottom, S.K., Fischer, C.R., Hryckowian, A.J., Van Treuren, W., Dodd, D., et al. (2022). Oxidative ornithine metabolism supports non-inflammatory C. difficile colonization. Nat Metab *4*, 19–28.

Ravcheev, D.A., and Thiele, I. (2014). Systematic genomic analysis reveals the complementary aerobic and anaerobic respiration capacities of the human gut microbiota. Front. Microbiol. *5*. https://doi.org/10.3389/fmicb.2014.00674.

Roager, H.M., and Licht, T.R. (2018). Microbial tryptophan catabolites in health and disease. Nat. Commun. *9*, 3294.

Rodionov, D.A., Arzamasov, A.A., Khoroshkin, M.S., Iablokov, S.N., Leyn, S.A., Peterson, S.N., Novichkov, P.S., and Osterman, A.L. (2019). Micronutrient Requirements and Sharing Capabilities of the Human Gut Microbiome. Front. Microbiol. *10*, 1316.

Schymanski, E.L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H.P., and Hollender, J. (2014). Identifying small molecules via high resolution mass spectrometry: communicating confidence. Environ. Sci. Technol. *48*, 2097–2098.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics *30*, 2068–2069.

Smith, N.W., Shorten, P.R., Altermann, E., Roy, N.C., and McNabb, W.C. (2020). Mathematical modelling supports the existence of a threshold hydrogen concentration and media-dependent yields in the growth of a reductive acetogen. Bioprocess Biosyst. Eng. *43*, 885–894.

Soto-Perez, P., Bisanz, J.E., Berry, J.D., Lam, K.N., Bondy-Denomy, J., and Turnbaugh, P.J. (2019). CRISPR-Cas System of a Prevalent Human Gut Bacterium Reveals Hyper-targeting against Phages in a Human Virome Catalog. Cell Host Microbe *26*, 325–335.e5.

Sperry, J.F., and Wilkins, T.D. (1976). Arginine, a growth-limiting factor for Eubacterium lentum. J. Bacteriol. *127*, 780–784.

Sprouffske, K., and Wagner, A. (2016). Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves. BMC Bioinformatics *17*, 172.

Strobel, H.J. (2009). Basic laboratory culture methods for anaerobic bacteria. Methods Mol. Biol. *581*, 247–261.

Szöcs, E., Stirling, T., Scott, E.R., Scharmüller, A., and Schäfer, R.B. (2020). webchem : An *R* Package to Retrieve Chemical Information from the Web. J. Stat. Softw. *93*. https://doi.org/10.18637/jss.v093.i13.

Thiele, I., and Palsson, B.Ø. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat. Protoc. *5*, 93–121.

Tramontano, M., Andrejev, S., Pruteanu, M., Klünemann, M., Kuhn, M., Galardini, M., Jouhten, P., Zelezniak, A., Zeller, G., Bork, P., et al. (2018). Nutritional preferences of human gut bacteria reveal their metabolic idiosyncrasies. Nature Microbiology https://doi.org/10.1038/s41564-018-0123-9.

Venturelli, O.S., Carr, A.C., Fisher, G., Hsu, R.H., Lau, R., Bowen, B.P., Hromada, S., Northen, T., and Arkin, A.P. (2018). Deciphering microbial interactions in synthetic human gut microbiome communities. Mol. Syst. Biol. *14*, e8157.

Wang, M., Carver, J.J., Phelan, V.V., Sanchez, L.M., Garg, N., Peng, Y., Nguyen, D.D., Watrous, J., Kapono, C.A., Luzzatto-Knaan, T., et al. (2016). Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat. Biotechnol. *34*, 828–837.

Weidenmaier, C., and Peschel, A. (2008). Teichoic acids and related cell-wall glycopolymers in Gram-positive physiology and host interactions. Nat. Rev. Microbiol. *6*, 276–287.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis.

Wohlgemuth, G., Haldiya, P.K., Willighagen, E., Kind, T., and Fiehn, O. (2010). The Chemical Translation Service--a web-based tool to improve standardization of metabolomic reports. Bioinformatics *26*, 2647–2648.

Wolfer, A. (2022). Short Asynchronous Time-Series Analysis [R package santaR version 1.2.3].

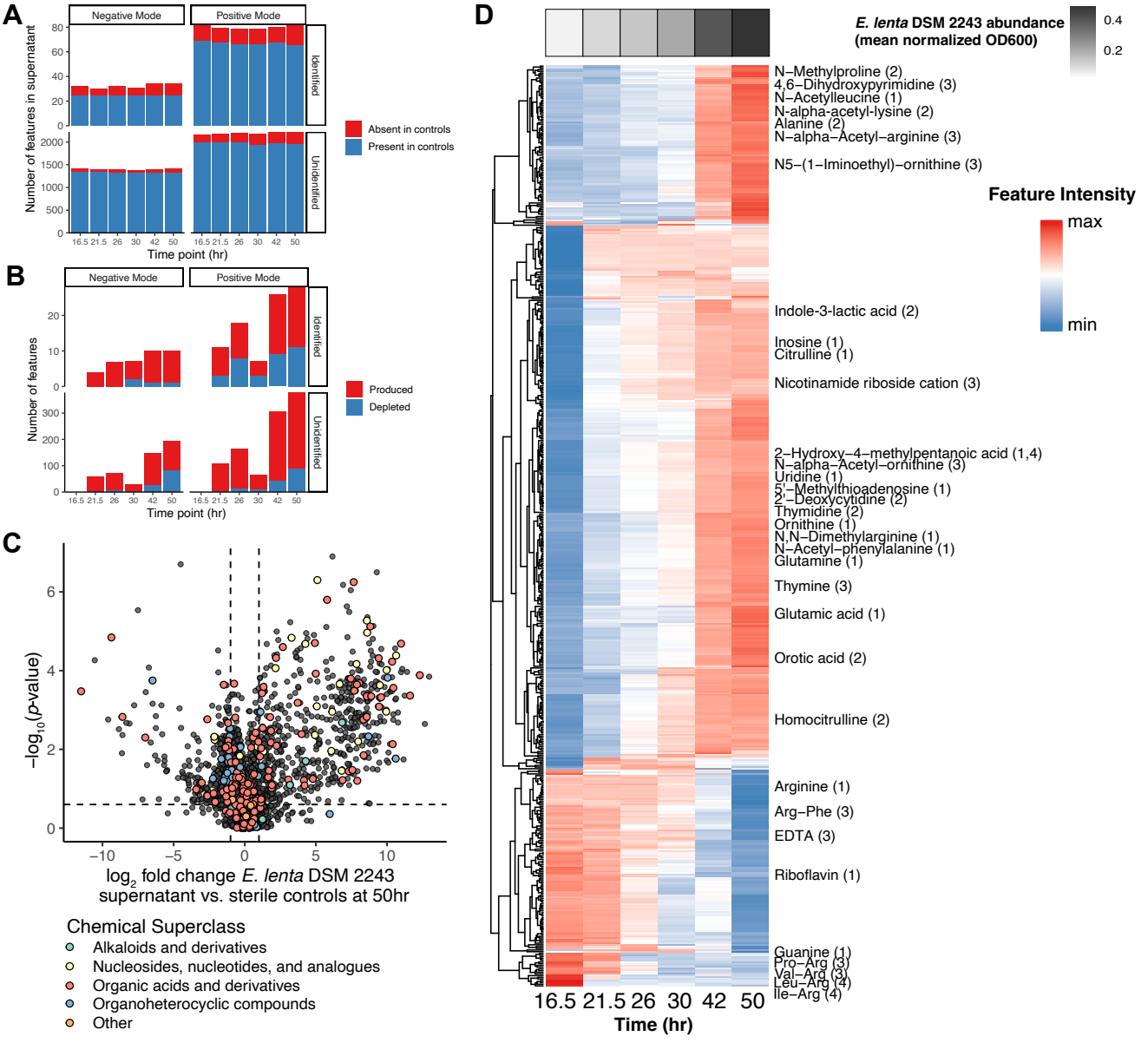
1991 Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan,
1992 L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics
1993 data. Innovation (Camb) *2*, 100141.

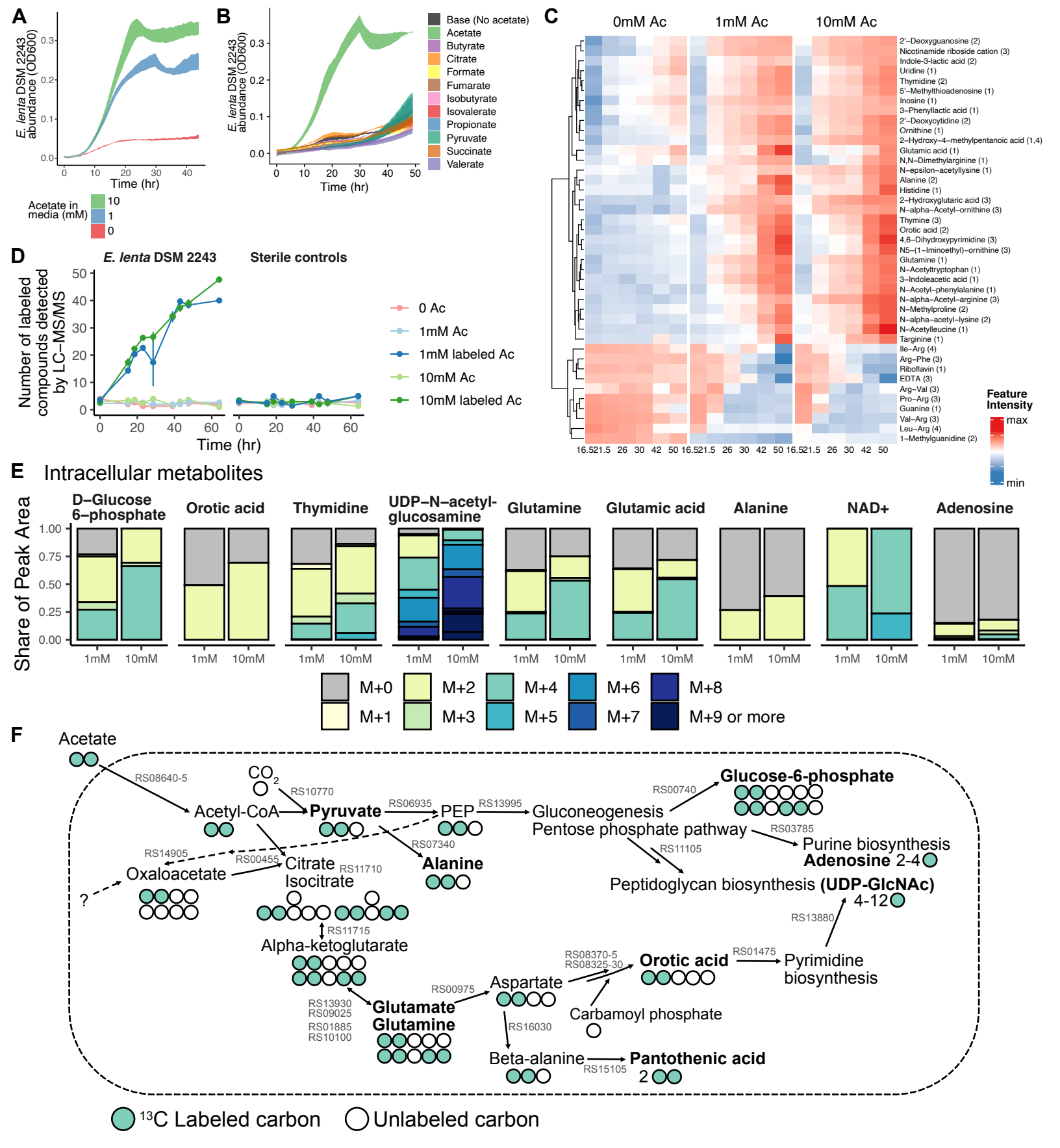
1994 Zhang, G., Mills, D.A., and Block, D.E. (2009). Development of Chemically Defined
1995 Media Supporting High-Cell-Density Growth of Lactococci, Enterococci, and
1996 Streptococci. Appl. Environ. Microbiol. *75*, 1080–1087.

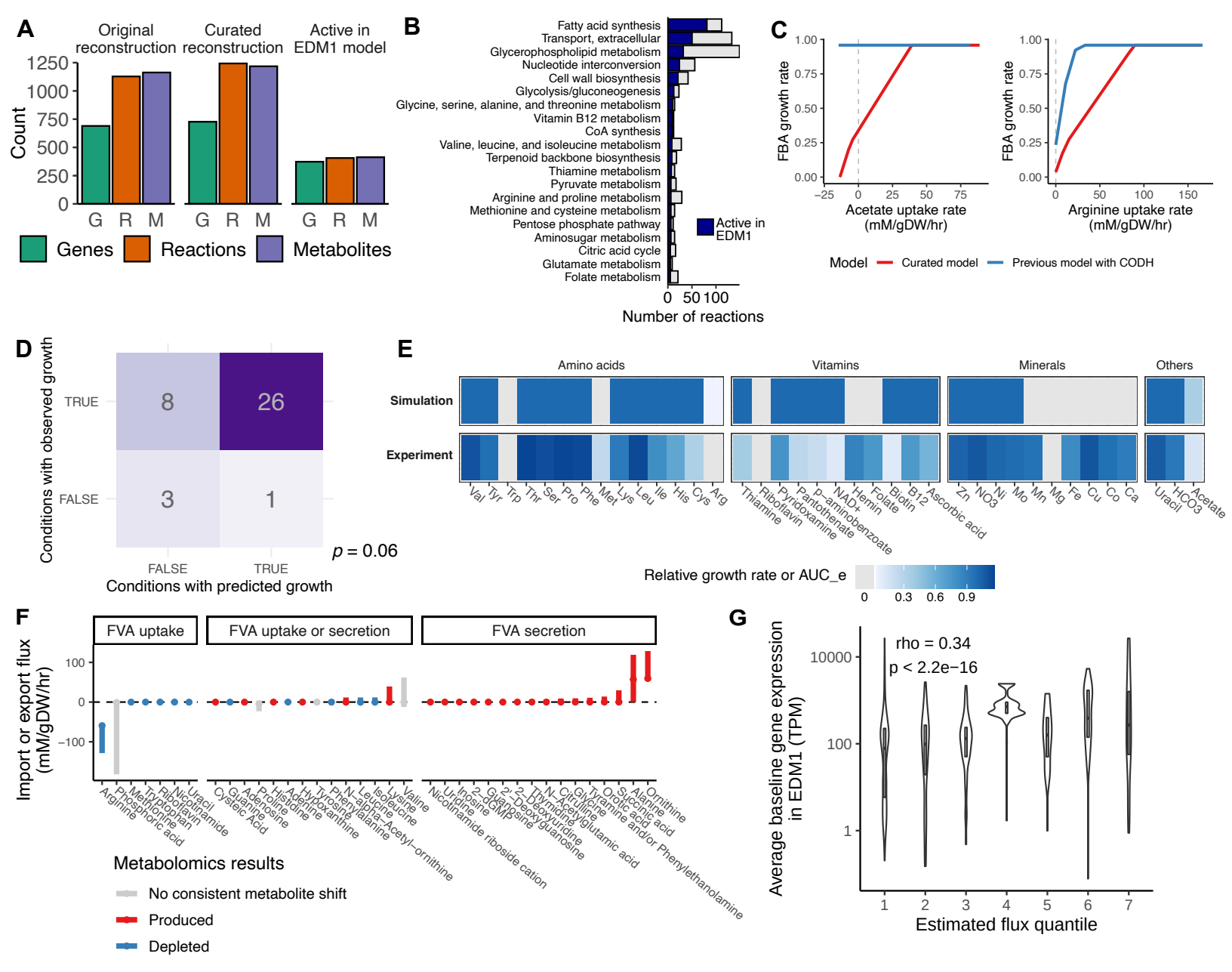
1997 Zhang, Y., Cai, J., Shang, X., Wang, B., Liu, S., Chai, X., Tan, T., Zhang, Y., and Wen,
1998 T. (2017). A new genome-scale metabolic model of Corynebacterium glutamicum and
1999 its application. Biotechnol. Biofuels *10*, 169.

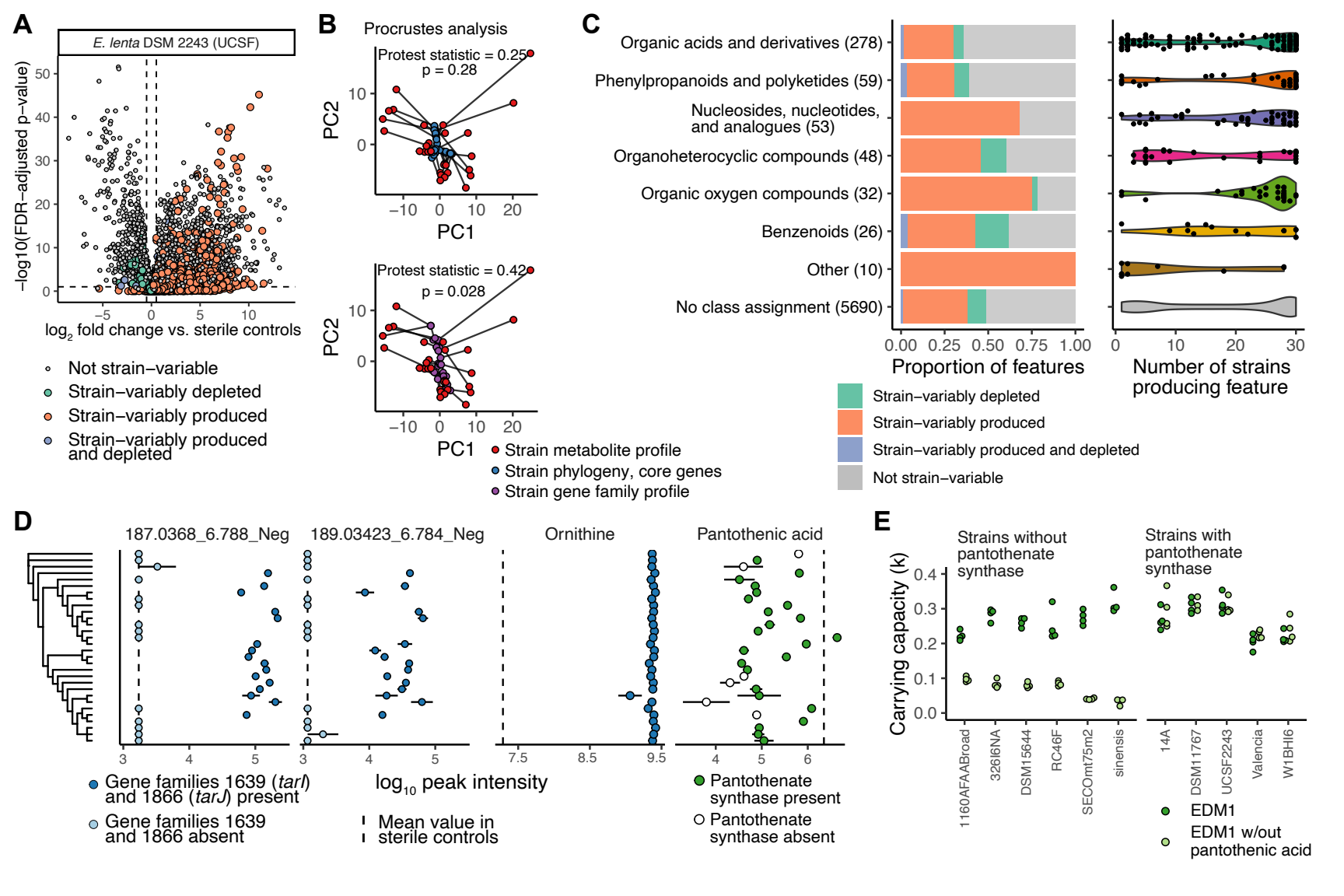
2000 Zhu, Q., Hou, Q., Huang, S., Ou, Q., Huo, D., Vázquez-Baeza, Y., Cen, C., Cantu, V.,
2001 Estaki, M., Chang, H., et al. (2021). Compositional and genetic alterations in Graves'
2002 disease gut microbiome reveal specific diagnostic biomarkers. ISME J.
2003 https://doi.org/10.1038/s41396-021-01016-7.

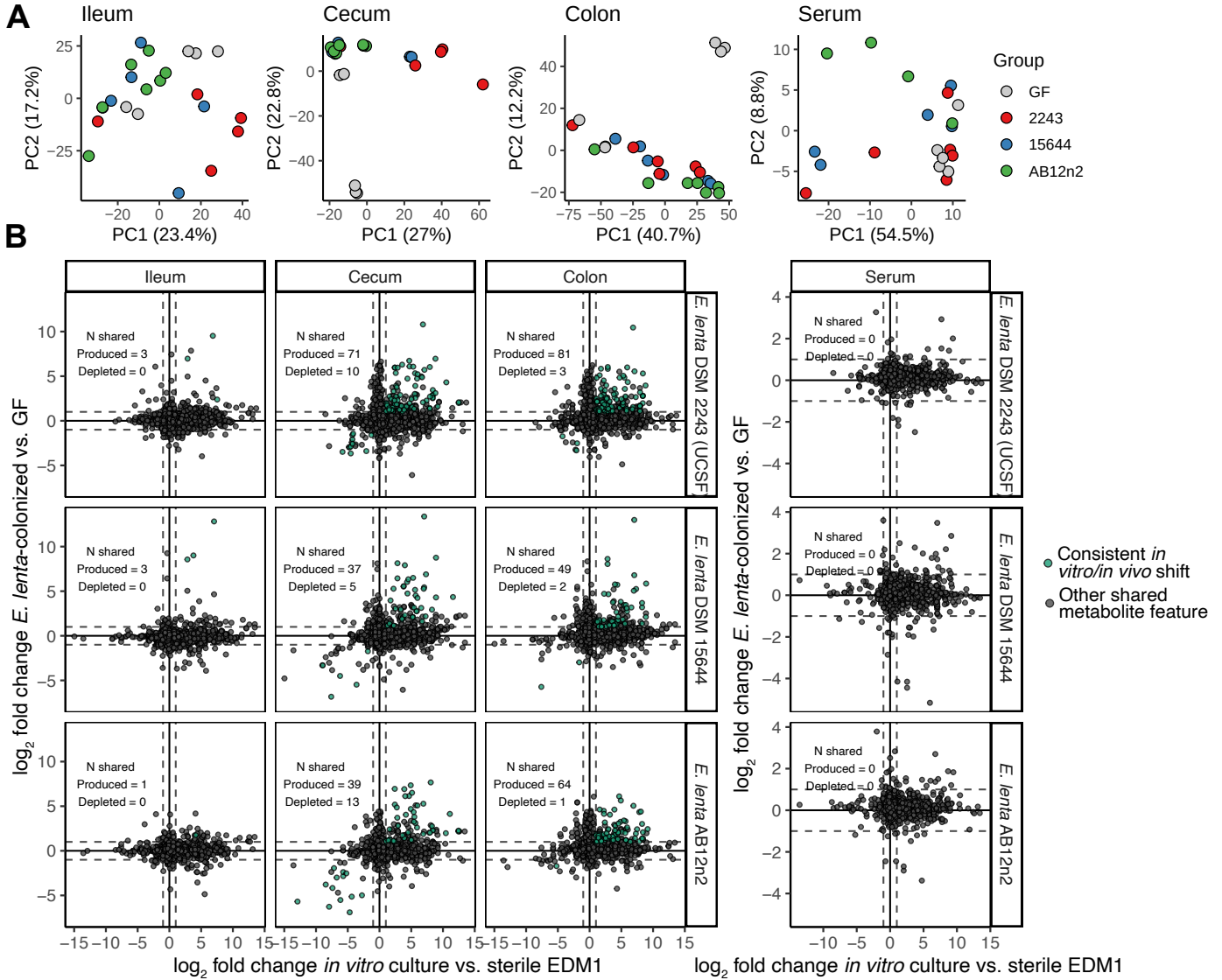
2004

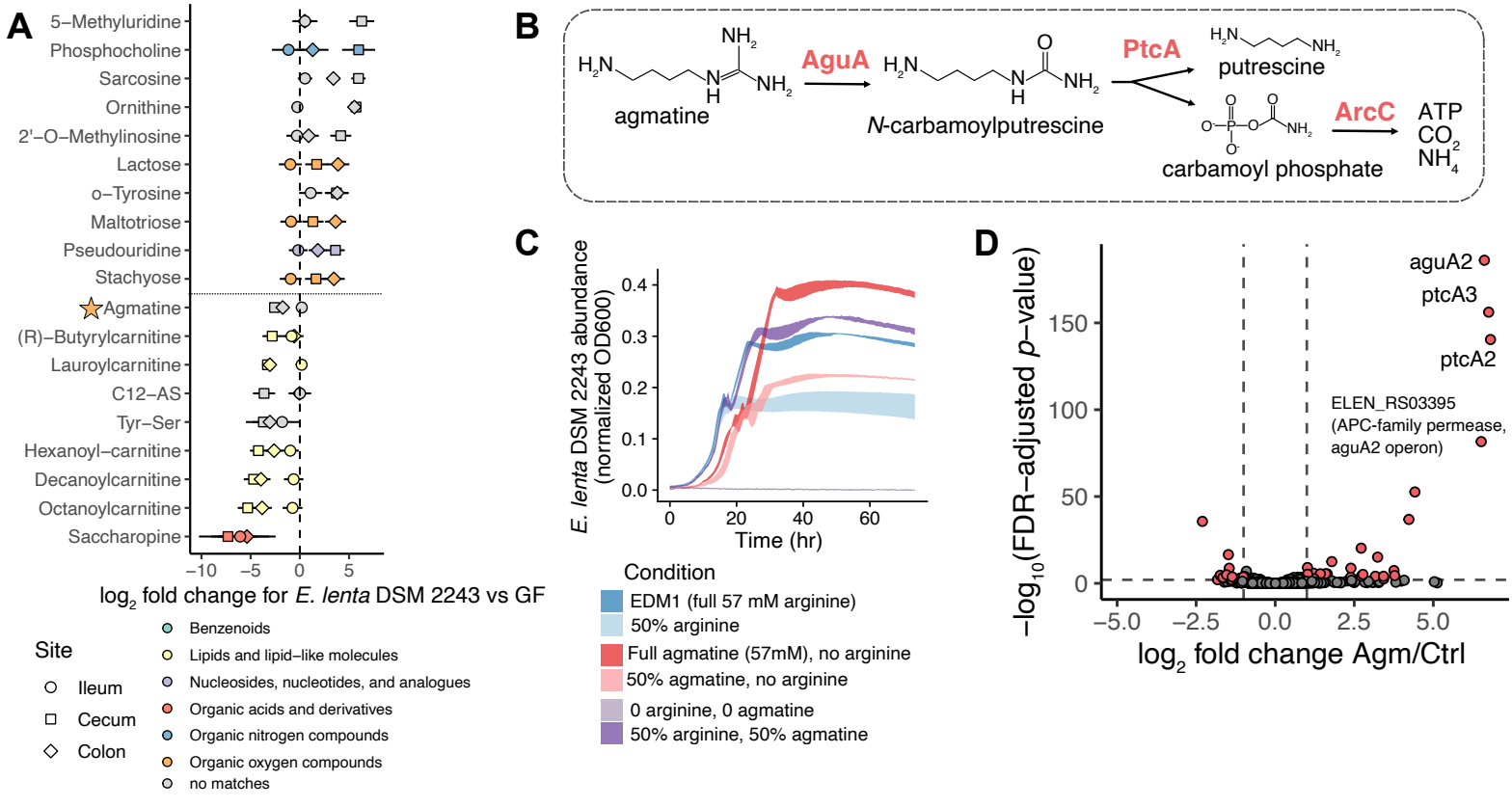

A
Negative Mode
Positive Mode
Identified
Unidentified
Number of features in supernatant
Time point (hr)
Absent in controls
Present in controls
B
Number of features
Produced
Depleted
C
−log10(p-value)
log2 fold change E. lenta DSM 2243 supernatant vs. sterile controls at 50hr
Chemical Superclass
Alkaloids and derivatives
Nucleosides, nucleotides, and analogues
Organic acids and derivatives
Organoheterocyclic compounds
Other
D
E. lenta DSM 2243 abundance (mean normalized OD600)
Feature Intensity
max
min
N−Methylproline (2)
4,6−Dihydroxypyrimidine (3)
N−Acetylleucine (1)
N−alpha−acetyl−lysine (2)
Alanine (2)
N−alpha−Acetyl−arginine (3)
N5−(1−Iminoethyl)−ornithine (3)
Indole−3−lactic acid (2)
Inosine (1)
Citrulline (1)
Nicotinamide riboside cation (3)
2−Hydroxy−4−methylpentanoic acid (1,4)
N−alpha−Acetyl−ornithine (3)
Uridine (1)
5'−Methylthioadenosine (1)
2'−Deoxycytidine (2)
Thymidine (2)
Ornithine (1)
N,N−Dimethylarginine (1)
N−Acetyl−phenylalanine (1)
Glutamine (1)
Thymine (3)
Glutamic acid (1)
Orotic acid (2)
Homocitrulline (2)
Arginine (1)
Arg−Phe (3)
EDTA (3)
Riboflavin (1)
Guanine (1)
Pro−Arg (3)
Val−Arg (3)
Leu−Arg (4)
Ile−Arg (4)
Time (hr)


A
E. lenta DSM 2243 abundance (OD600)
Time (hr)
Acetate in media (mM)
10
1
0
B
Base (No acetate)
Acetate
Butyrate
Citrate
Formate
Fumarate
Isobutyrate
Isovalerate
Propionate
Pyruvate
Succinate
Valerate
C
0mM Ac
1mM Ac
10mM Ac
2'–Deoxyguanosine (2)
Nicotinamide riboside cation (3)
Indole-3-lactic acid (2)
Uridine (1)
Thymidine (2)
5'–Methylthioadenosine (1)
Inosine (1)
3–Phenyllactic acid (1)
2'–Deoxycytidine (2)
Ornithine (1)
2–Hydroxy–4–methylpentanoic acid (1,4)
Glutamic acid (1)
N,N–Dimethylarginine (1)
N–epsilon–acetyllysine (1)
Alanine (2)
Histidine (1)
2–Hydroxyglutaric acid (3)
N–alpha–Acetyl–ornithine (3)
Thymine (3)
Orotic acid (2)
4,6–Dihydroxypyrimidine (3)
N5–(1–Iminoethyl)–ornithine (3)
Glutamine (1)
N–Acetyltryptophan (1)
3–Indoleacetic acid (1)
N–Acetyl–phenylalanine (3)
N–alpha–Acetyl–arginine (3)
N–Methylproline (2)
N–alpha–acetyl–lysine (2)
N–Acetylleucine (1)
Targinine (1)
Ile–Arg (4)
Arg–Phe (3)
Riboflavin (1)
EDTA (3)
Arg–Val (3)
Pro–Arg (3)
Guanine (1)
Val–Arg (3)
Leu–Arg (4)
1–Methylguanidine (2)
Feature Intensity
max
min
D
E. lenta DSM 2243
Sterile controls
Number of labeled compounds detected by LC–MS/MS
Time (hr)
0 Ac
1mM Ac
1mM labeled Ac
10mM Ac
10mM labeled Ac
E
Intracellular metabolites
Share of Peak Area
D–Glucose 6–phosphate
Orotic acid
Thymidine
UDP–N–acetyl-glucosamine
Glutamine
Glutamic acid
Alanine
NAD+
Adenosine
1mM
10mM
M+0
M+1
M+2
M+3
M+4
M+5
M+6
M+7
M+8
M+9 or more
F
Acetate
RS08640-5
CO2
RS10770
Acetyl-CoA
Pyruvate
RS06935
PEP
RS13995
Gluconeogenesis
Pentose phosphate pathway
RS00740
Glucose-6-phosphate
RS03785
Purine biosynthesis
Adenosine 2-4
RS11105
Peptidoglycan biosynthesis (UDP-GlcNAc)
4-12
RS13880
RS14905
RS00455
Oxaloacetate
Citrate
Isocitrate
RS11710
RS07340
Alanine
?
RS11715
Alpha-ketoglutarate
RS08370-5
RS08325-30
Orotic acid
RS01475
Pyrimidine biosynthesis
RS00975
Aspartate
Carbamoyl phosphate
RS13930
RS09025
RS01885
RS10100
Glutamate
Glutamine
RS16030
Beta-alanine
RS15105
Pantothenic acid
2
13C Labeled carbon
Unlabeled carbon


A
Original reconstruction
Curated reconstruction
Active in EDM1 model
Count
1250
1000
750
500
250
0
G R M G R M G R M
Genes
Reactions
Metabolites
B
Fatty acid synthesis
Transport, extracellular
Glycerophospholipid metabolism
Nucleotide interconversion
Cell wall biosynthesis
Glycolysis/gluconeogenesis
Glycine, serine, alanine, and threonine metabolism
Vitamin B12 metabolism
CoA synthesis
Valine, leucine, and isoleucine metabolism
Terpenoid backbone biosynthesis
Thiamine metabolism
Pyruvate metabolism
Arginine and proline metabolism
Methionine and cysteine metabolism
Pentose phosphate pathway
Aminosugar metabolism
Citric acid cycle
Glutamate metabolism
Folate metabolism
Active in EDM1
0 50 100
Number of reactions
C
FBA growth rate
1.00
0.75
0.50
0.25
0.00
−25 0 25 50 75
Acetate uptake rate (mM/gDW/hr)
0 50 100 150
Arginine uptake rate (mM/gDW/hr)
Model
Curated model
Previous model with CODH
D
Conditions with observed growth
TRUE
FALSE
8
26
3
1
FALSE TRUE
Conditions with predicted growth
p = 0.06
E
Amino acids
Vitamins
Minerals
Others
Simulation
Experiment
Val Tyr Trp Thr Ser Pro Phe Met Lys Leu Ile His Cys Arg
Thiamine Riboflavin Pyridoxamine Pantothenate p−aminobenzoate NAD+ Hemin Folate Biotin B12 Ascorbic acid
Zn NO3 Ni Mo Mn Mg Fe Cu Co Ca
Uracil HCO3 Acetate
Relative growth rate or AUC_e
0 0.3 0.6 0.9
F
FVA uptake
FVA uptake or secretion
FVA secretion
Import or export flux (mM/gDW/hr)
100
0
−100
Arginine Phosphoric acid Methionine Tryptophan Riboflavin Nicotinamide Uracil
Cysteic Acid Guanine Adenosine Proline Histidine Adenine Hypoxanthine Tyrosine Phenylalanine N−alpha−Acetyl−ornithine Leucine Isoleucine Lysine Valine
Nicotinamide riboside cation Uridine Inosine 2−dGMP 2−Deoxyguanosine N−Acetylglutamic acid Tyramine Glycine Citrulline Orotic acid Succinic acid Alanine Ornithine
Metabolomics results
No consistent metabolite shift
Produced
Depleted
G
Average baseline gene expression in EDM1 (TPM)
rho = 0.34
p < 2.2e−16
10000
100
1
1 2 3 4 5 6 7
Estimated flux quantile


A
E. lenta DSM 2243 (UCSF)
−log10(FDR−adjusted p−value)
log2 fold change vs. sterile controls
Not strain−variable
Strain−variably depleted
Strain−variably produced
Strain−variably produced and depleted
B
Procrustes analysis
Protest statistic = 0.25
p = 0.28
Protest statistic = 0.42
p = 0.028
PC1
PC2
Strain metabolite profile
Strain phylogeny, core genes
Strain gene family profile
C
Organic acids and derivatives (278)
Phenylpropanoids and polyketides (59)
Nucleosides, nucleotides, and analogues (53)
Organoheterocyclic compounds (48)
Organic oxygen compounds (32)
Benzenoids (26)
Other (10)
No class assignment (5690)
Proportion of features
Number of strains producing feature
Strain−variably depleted
Strain−variably produced
Strain−variably produced and depleted
Not strain−variable
D
187.0368_6.788_Neg
189.03423_6.784_Neg
Ornithine
Pantothenic acid
log10 peak intensity
Gene families 1639 (tarI) and 1866 (tarJ) present
Gene families 1639 and 1866 absent
Mean value in sterile controls
Pantothenate synthase present
Pantothenate synthase absent
E
Strains without pantothenate synthase
Strains with pantothenate synthase
Carrying capacity (k)
1160AFAABroad
326I6NA
DSM15644
RC46F
SECOmt75m2
sinensis
14A
DSM11767
UCSF2243
Valencia
W1BHI6
EDM1
EDM1 w/out pantothenic acid


A
Ileum
Cecum
Colon
Serum
PC2 (17.2%)
PC1 (23.4%)
PC2 (22.8%)
PC1 (27%)
PC2 (12.2%)
PC1 (40.7%)
PC2 (8.8%)
PC1 (54.5%)
Group
GF
2243
15644
AB12n2
B
Ileum
Cecum
Colon
Serum
E. lenta DSM 2243 (UCSF)
E. lenta DSM 15644
E. lenta AB12n2
N shared
Produced = 3
Depleted = 0
Produced = 71
Depleted = 10
Produced = 81
Depleted = 3
Produced = 0
Depleted = 0
Produced = 3
Depleted = 0
Produced = 37
Depleted = 5
Produced = 49
Depleted = 2
Produced = 0
Depleted = 0
Produced = 1
Depleted = 0
Produced = 39
Depleted = 13
Produced = 64
Depleted = 1
Produced = 0
Depleted = 0
log2 fold change E. lenta-colonized vs. GF
log2 fold change in vitro culture vs. sterile EDM1
Consistent in vitro/in vivo shift
Other shared metabolite feature


A
5–Methyluridine
Phosphocholine
Sarcosine
Ornithine
2'–O–Methylinosine
Lactose
o–Tyrosine
Maltotriose
Pseudouridine
Stachyose
Agmatine
(R)–Butyrylcarnitine
Lauroylcarnitine
C12–AS
Tyr–Ser
Hexanoyl–carnitine
Decanoylcarnitine
Octanoylcarnitine
Saccharopine
−10
−5
0
5
log2 fold change for E. lenta DSM 2243 vs GF
Site
Ileum
Cecum
Colon
Benzenoids
Lipids and lipid–like molecules
Nucleosides, nucleotides, and analogues
Organic acids and derivatives
Organic nitrogen compounds
Organic oxygen compounds
no matches
B
agmatine
AguA
N-carbamoylputrescine
PtcA
putrescine
carbamoyl phosphate
ArcC
ATP
CO2
NH4
C
E. lenta DSM 2243 abundance (normalized OD600)
0.0
0.1
0.2
0.3
0.4
0
20
40
60
Time (hr)
Condition
EDM1 (full 57 mM arginine)
50% arginine
Full agmatine (57mM), no arginine
50% agmatine, no arginine
0 arginine, 0 agmatine
50% arginine, 50% agmatine
D
−log10(FDR–adjusted p–value)
0
50
100
150
aguA2
ptcA3
ptcA2
ELEN_RS03395
(APC-family permease,
aguA2 operon)
−5.0
−2.5
0.0
2.5
5.0
log2 fold change Agm/Ctrl