

# UC San Diego

## UC San Diego Previously Published Works

### Title

Pangenome reconstruction of Lactobacillaceae metabolism predicts species-specific metabolic traits.

### Permalink

<https://escholarship.org/uc/item/6238b6k4>

### Journal

mSystems, 9(7)

### Authors

Ardalani, O

Phaneuf, P

Mohite, O

et al.

### Publication Date

2024-07-23

### DOI

10.1128/msystems.00156-24

Peer reviewed

# Pangenome reconstruction of *Lactobacillaceae* metabolism predicts species-specific metabolic traits

O. Ardalani,<sup>1</sup> P. V. Phaneuf,<sup>1</sup> O. S. Mohite,<sup>1</sup> L. K. Nielsen,<sup>1,2</sup> B. O. Palsson<sup>1,3,4,5,6</sup>

**AUTHOR AFFILIATIONS** See affiliation list on p. 16.

**ABSTRACT** Strains across the *Lactobacillaceae* family form the basis for a trillion-dollar industry. Our understanding of the genomic basis for their key traits is fragmented, however, including the metabolism that is foundational to their industrial uses. Pangenome analysis of publicly available *Lactobacillaceae* genomes allowed us to generate genome-scale metabolic network reconstructions for 26 species of industrial importance. Their manual curation led to more than 75,000 gene-protein-reaction associations that were deployed to generate 2,446 genome-scale metabolic models. Cross-referencing genomes and known metabolic traits allowed for manual metabolic network curation and validation of the metabolic models. As a result, we provide the first pangenomic basis for metabolism in the *Lactobacillaceae* family and a collection of predictive computational metabolic models that enable a variety of practical uses.

**IMPORTANCE** *Lactobacillaceae*, a bacterial family foundational to a trillion-dollar industry, is increasingly relevant to biosustainability initiatives. Our study, leveraging approximately 2,400 genome sequences, provides a pangenomic analysis of *Lactobacillaceae* metabolism, creating over 2,400 curated and validated genome-scale models (GEMs). These GEMs successfully predict (i) unique, species-specific metabolic reactions; (ii) niche-enriched reactions that increase organism fitness; (iii) essential media components, offering insights into the global amino acid essentiality of *Lactobacillaceae*; and (iv) fermentation capabilities across the family, shedding light on the metabolic basis of *Lactobacillaceae*-based commercial products. This quantitative understanding of *Lactobacillaceae* metabolic properties and their genomic basis will have profound implications for the food industry and biosustainability, offering new insights and tools for strain selection and manipulation.

**KEYWORDS** *Lactobacillaceae*, genome-scale metabolic model, pangenome, genome-scale reconstruction

*Lactobacillaceae* are an essential family of highly diverse lactic acid bacteria. It comprises a large number of species that populate a variety of habitats (1). Due to numerous applications in food and pharmaceutical industries, *Lactobacillaceae*-dependent products have a trillion-dollar market size, including dairy, wine, probiotics, and numerous satellite industries (2–4) (see Table S1 for details), indicating their importance in microbial biotechnology and related industries.

Cost-effective DNA sequencing has led to a steadily increasing number of *Lactobacillaceae* genomes deposited in the NCBI database (5). The availability of these genomes has enabled pangenomic studies (1, 6–8). Metabolism is foundational to the industrial uses of *Lactobacillaceae*, and the generation of predictive genome-scale metabolic models (GEMs) constitutes a significant advancement in bioprocess engineering (9). GEMs are based on annotated sequences and use algorithms to forecast cellular behavior and metabolic fluxes under conditions of interest. GEMs have been shown to predict

**Editor** Karoline Faust, Katholieke Universiteit Leuven, Leuven, Belgium

Address correspondence to B. O. Palsson, palsson@ucsd.edu.

The authors declare no conflict of interest.

See the funding table on p. 16.

**Received** 1 February 2024

**Accepted** 17 May 2024

**Published** 26 June 2024

Copyright © 2024 Ardalani et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

optimal growth conditions and gene manipulation targets to increase product yield by integrating complex metabolic pathways. The availability of a large number of *Lactobacillaceae* genome sequences enables pangenome-based metabolic reconstruction to create a comprehensive and predictive set of GEMs.

Despite the large market size, undeniable contribution to today's humankind lifestyle, and availability of large sets of genomes and other omics data, only a few genome-scale models have been reconstructed for *Lactobacillaceae* members so far. Previous reconstruction efforts were limited to several model strains across the whole family, including *Lactobacillus plantarum* (10), *Lactobacillus reuteri* (11), *Lactobacillus casei* (12), *Lactobacillus mesenteroides* (13), and *Oenococcus oeni* (14), and only one multi-strain reconstruction for *L. reuteri* covering for 36 strains (15). Our work aims to build upon these foundational studies.

In this study, 2,446 GEMs were generated across the *Lactobacillaceae* family. This set of models (called a PanGEM) enables the identification of conserved and variable metabolic traits across different family members. The individual GEMs can be used to develop strategies for strain improvement and bioprocess optimization. Moreover, predictive GEMs have a wide range of applications, from understanding the metabolic basis for probiotic properties to producing value-added compounds for the food and pharmaceutical industries. Overall, constructing the PanGEM is important to fully harness *Lactobacillaceae's* biotechnological potential, and it constitutes a new quantitative representation of the genomic basis for this large industry.

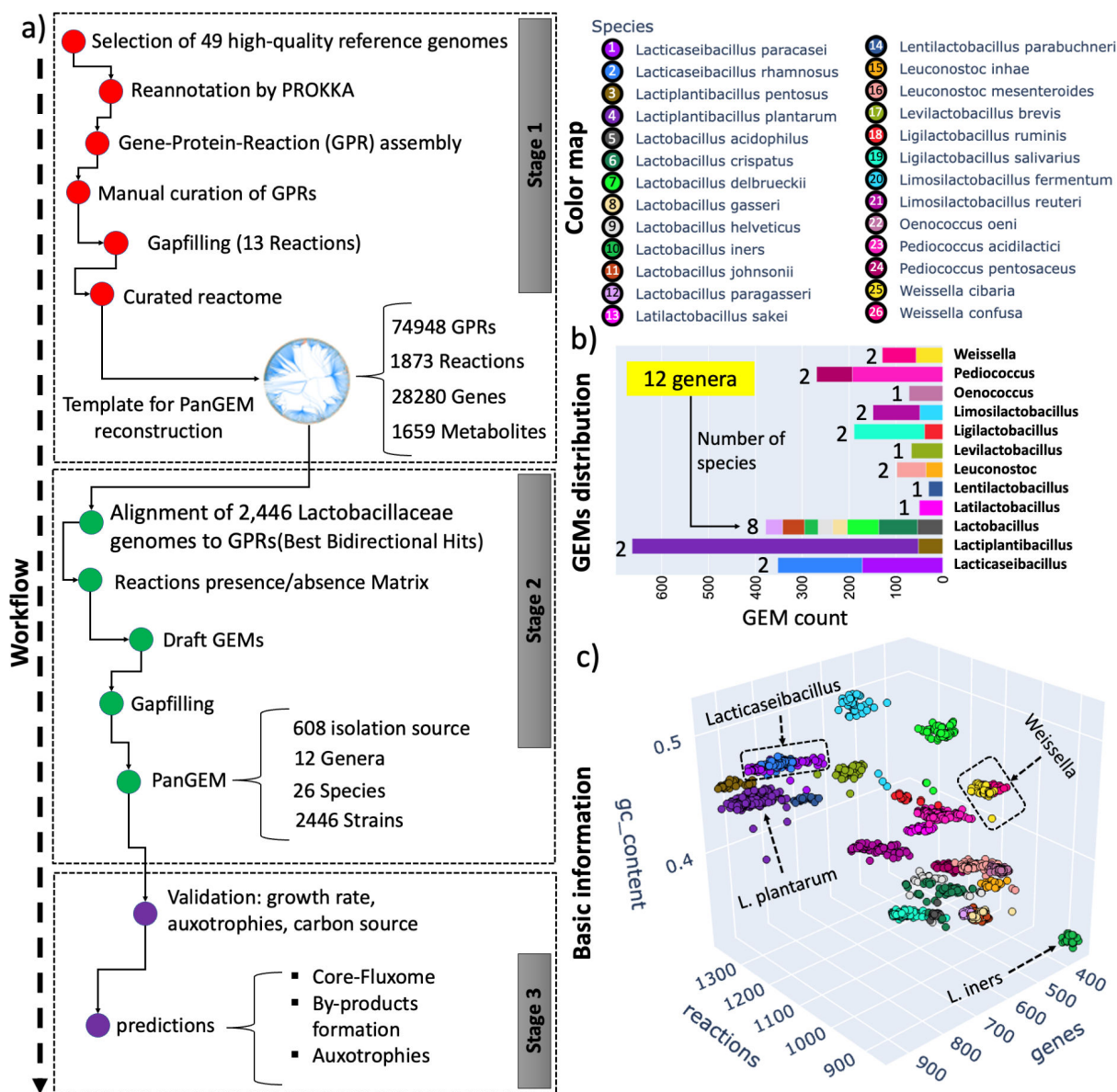
## RESULTS

### Metabolic network reconstruction and genome-scale models for a family of bacteria

First, a metabolic pan-reactome was constructed for use as a template for species-specific metabolic reconstructions. The pan-reactome was reconstructed from 49 high-quality reference genomes, representing 33 species from 9 genera (File S1). The reactome (available at <https://github.com/omidard/LactoPanGEM/blob/main/LBReactome.xml>) contained 75,299 gene-protein-reaction associations (GPRs), 1,873 reactions (File S2), 28,280 genes, and 1,659 metabolites (File S3) (Fig. 1a, stage 1). Reactions were distributed across 231 cellular subsystems. Of 1,873 reactions, 1,516 were gene-associated reactions, whereas the remaining belonged to other types, including exchange, orphan, gap (see Fig. S1 for a detailed gap analysis), demand, and sink reactions. For clarity, it is worth noting that the 75,299 gene-protein-reaction associations result from capturing all possible alleles for a reaction across our reference genomes. Consequently, multiple distinct gene sets, representing different alleles, can be associated with a single metabolic reaction, leading to several GPRs for that specific metabolic reaction.

Second, metabolic models for the *Lactobacillaceae* family were constructed. The GPRs were matched against the ORFs in every qualified genome, resulting in 2,446 strain-specific GEMs from 26 species and 12 genera obtained from 608 distinct isolation sources (6) (Fig. S2). *L. plantarum* had the highest number of GEMs (611), while *Lactobacillus iners* and *Lactobacillus parabuchneri* had the lowest (30 each). The genus *Lactiplantibacillus* had the most GEMs (663), while *Lentilactobacillus* had the least (30) (see Fig. 1b). In the *Lactobacillaceae* PanGEM, the genus *Lactobacillus* had the highest diversity with eight species, while *Oenococcus*, *Levilactobacillus*, *Lentilactobacillus*, and *Latilactobacillus* had only one species each, resulting in the lowest diversity within the group.

The PanGEM covers a wide range of metabolic diversity across 26 *Lactobacillaceae* species, with multiple GEMs for each species. The number of reactions per genome ranged from 859 to 1,358, representing 354–944 genes, respectively. On average, *L. iners* had the lowest number of genes and reactions, while *Lactobacillus pentosus* had the highest (see Fig. S3). An in-depth comparison of reaction presence relative to genome length within the *Lactobacillaceae* family is presented in Fig. S4. The analysis reveals a



**FIG 1** Pangenome-scale metabolic reconstructions for *Lactobacillaceae*. (a) The overall three-stage workflow used in this study. Stage 1: reactome reconstruction. To formulate a draft pan-metabolic network reconstruction, 49 *Lactobacillaceae* reference genomes were selected and annotated by Prokka. The draft reactome was manually curated to develop “gold-standard” multi-allelic gene-protein-reaction associations, covering 1,832 reactions and 28,280 metabolic alleles. Stage 2: building strain-specific metabolic reconstructions on a family-wide basis. A total of 2,446 *Lactobacillaceae* genomes from NCBI were aligned with BLAST against the GPRs to identify the corresponding reactions based on sequence similarity. A biomass objective function was added to all draft reconstructions to identify and fill gaps in the metabolic reconstruction. At this step, all GEMs were able to produce biomass and were ready for further analysis. Stage 3: validation and analysis of strain-specific GEMs. Additional validation through iterative refinement was performed to achieve precise and accurate metabolic phenotype predictions. After GEMs passed quality control, analysis was performed to identify common metabolic capabilities and differences across individual *Lactobacillaceae* strains. (b) GEM distribution across the *Lactobacillaceae*. GEMs are classified based on their genus (Y-axis). Different species within each genus are annotated according to the color map; bar plots are annotated with the number of species within each genus. PanGEM contains 12 genera, including *Lactobacillus*, *Ligilactobacillus*, *Limosilactobacillus*, *Lacticaseibacillus*, *Oenococcus*, *Leuconostoc*, *Levilactobacillus*, *Pediococcus*, *Lactiplantibacillus*, *Latilactobacillus*, *Weissella*, and *Lentilactobacillus*. (c) The basic information about the metabolic reconstructions in PanGEM. Color map annotates species, and the three axes depict the number of reactions, the number of genes for each reconstruction, and the GC content of the corresponding genome. These three parameters give distinct clusters for the stains of each species.

strong correlation ( $R^2 = 0.87$ ) between genome size and metabolic reactions. Notably, strains with smaller genomes predominantly exhibit lost reactions in lipid, aminoacid,

and nucleotide metabolism. Each species can be distinguished from others based on three basic properties (number of genes, number of reactions, and GC content of genomes) (Fig. 1c).

## Validation of PanGEM

Once formulated, GEMs were validated against experimentally determined metabolic and physiological traits (16). Growth predictions on 21 carbon sources for three strains of *Limosilactibacillus* (*sakei* DSM 20017, *sakei* LS25, and *sakei* 23k) were compared to experimental results available in the literature (17) (Fig. 2a). Out of 63 simulations, 52 (83%) were accurate, while 2 were false negatives, and 9 were false positives. Both predictions and experimental results were obtained on chemically defined media (CDM). (see CDM formulation and constraints in Table S2).

For further qualitative phenotypic validation of the PanGEM, single omission analysis of 49 compounds in a CDM was simulated by flux-balance analysis (FBA). The computational results were compared to experimental data (13, 18–21) (Fig. 2a). Experimental data were available for six strains within PanGEM, representing four different species and three unique genera. FBA was performed to predict the single omission of 49 compounds from CDM, one by one across six strains. Among 294 simulations, only 8 (2.7%) failed to predict the correct phenotype (false positives) (Fig. 2a).

Growth rate computations showed that the models were capable of quantitative prediction of growth rates that were in agreement with the experimental reports (19, 22–24) (Fig. 2b). Despite the limited growth rate data available in the literature from CDM, validation was performed on six GEMs, and predictions were satisfactory with a mean absolute percentage error of 6.62, root mean squared error of 0.02, and correlation coefficient of 0.93. These metrics illustrate the prediction potential of PanGEM models (Fig. 2b). The sensitivity of predicted growth rates to variations in growth-associated maintenance (GAM), non-growth-associated maintenance (NGAM), and amino acid uptake rates was rigorously examined. While changes in GAM and NGAM parameters did not show an influence on growth rates (Fig. S5), alterations in amino acid uptake rates were shown to be impactful (Fig. S6).

PanGEM could thus be validated against data reported in the literature (Fig. 2c). The validated PanGEM can be used for the analysis of more detailed metabolic traits.

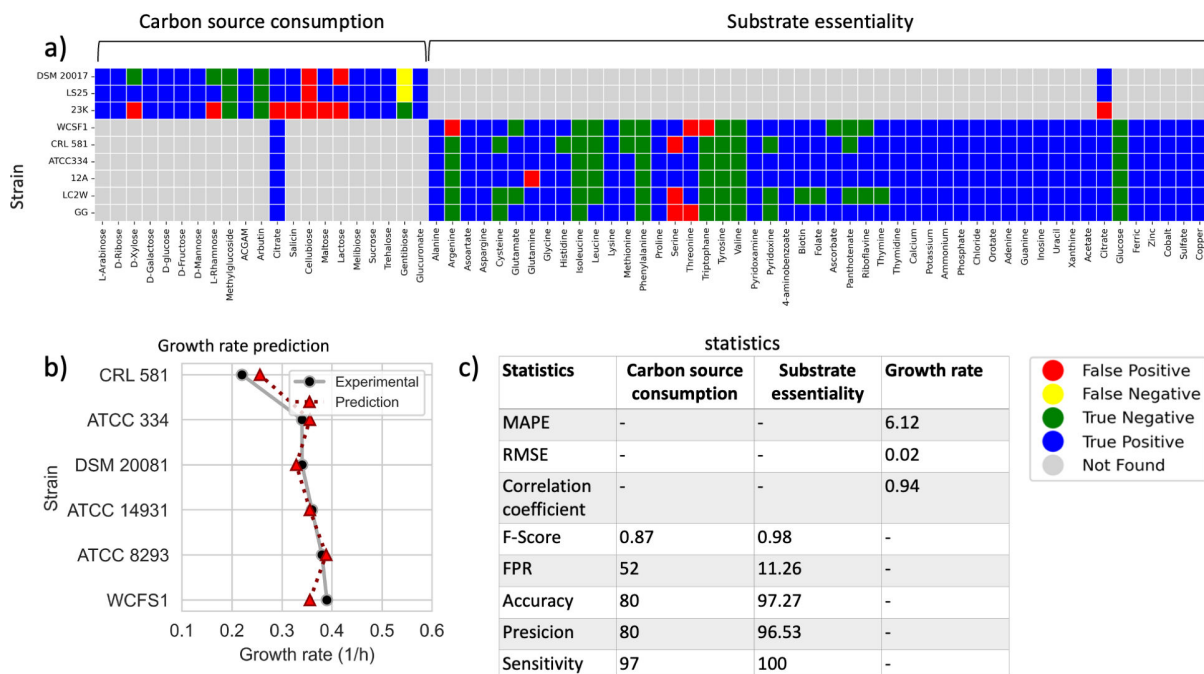
## Classifying strain-specific reactomes in PanGEM

Strain-specific GEMs can be clustered based on the reactions they contain (Fig. 3a). Such clustering highlights metabolic differences between species. The cluster map (Fig. 3a) shows three clusters of reactions: core reactions (common to all strains), accessory reactions (found in many strains), and rare reactions (found in few or even in single strains, then called unique). All strains shared 185 common reactions (Fig. 3b), forming the *Lactobacillaceae* core reactome. There were 1,130 accessory reactions, and they are differentially distributed across the 26 strains (Fig. 3a). Metabolic conservation in each species was assessed by performing a reaction frequency analysis (Fig. 3d). Results indicate that *L. plantarum* had the most metabolically diverse strains, while *Lactobacillus acidophilus* was the most metabolically conserved species (Note S1).

A total of 65 unique reactions were found within 13 species (Fig. 3b, inset bar plot). Among all species, *Lactobacillus ruminis*, a niche-specialized species, had the highest unique reaction count of 19. Next was *L. reuteri* with 11 unique reactions, followed by *Ligilactobacillus salivarius*, with 9 unique reactions (Fig. S7).

## Characterization of allowable flux states using PanGEM

FBA analysis was performed for all 2,446 strain models in PanGEM to better understand *Lactobacillaceae* core metabolism. A subset of 185 reactions spanning 30 distinct metabolic subsystems, common to all GEMs, was identified as the core reactome. Flux variability analysis (FVA) confirmed that all reactions within this subset were active across

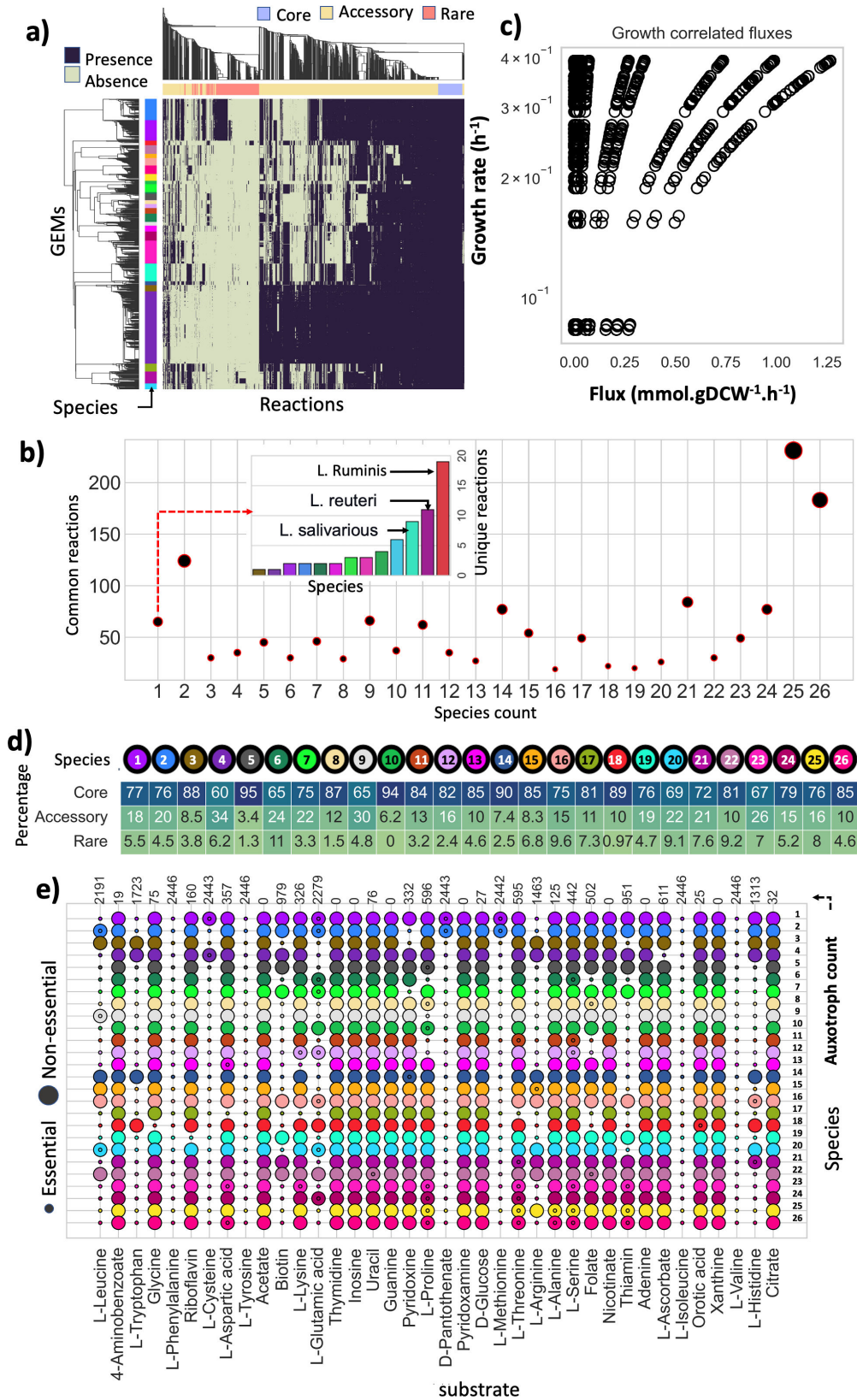


**FIG 2** PanGEM validation. (a) Carbon source utilization: carbon source utilization validation for three strains within PanGEM (*L. sakei* 23K, *L. sakei* DSM 20017, and *L. sakei* LS25) by comparing experimental and predicted results. The capability to grow on 21 carbon sources (X-axis) for three strains (Y-axis) is shown; the color map annotates false/true/positive/negative predictions. Statistical analysis based on a confusion matrix showed an *F*-score of 0.98, a false-positive rate (FPR) of 11.26%, a precision of 96.53%, an accuracy of 97.27%, and a sensitivity of 100%. Substrate essentiality: prediction of auxotrophies. Validation of PanGEM by comparing experimental and predicted results for CDM component essentiality for six strains [*L. plantarum* WCSF1, *Lactobacillus delbrueckii* CRL581, *L. paracasei* ATCC334, *L. paracasei* LC2W, and *Lactocaseibacillus rhamnosus* GG (Y-axis)]. A single-component omission analysis of 49 components of *Lactobacillaceae*-specific CDM was simulated using flux-balance analysis (FBA), and results were compared with experimental data. Orange dots represent false-positive predictions. Blue and green dots represent true negative and true positive, respectively. No false-negative prediction was observed, with an *F*-score of 0.87, a false positive rate of 52%, a precision of 80%, an accuracy of 80%, and a sensitivity of 97%. (b) A comparison of simulated growth rate using FBA on CDM (dashed line-red triangle) with experimental data on the same condition (solid line-black dots) for six strains within PanGEM. CRL 581, *L. delbrueckii* CRL 581; ATCC 8293, *L. mesenteroides* ATCC 8293; DSM 20017, *L. sakei* DSM 20017; DSM 20081, *L. sakei* DSM 20081; LS25, *L. sakei* LS25; 23K, *L. sakei* 23K; WCSF1, *L. plantarum* WCSF1; ATCC334, *L. paracasei* ATCC334; 12A, *L. paracasei* 12A; LC2W, *L. paracasei* LC2W; GG, *Lactocaseibacillus rhamnosus* GG; ATCC 14931, *Limosilactobacillus fermentum* ATCC 14931). (c) Statistics. Statistical analysis of GEM validation results is summarized in the table. For quantitative predictions (growth rates), mean absolute percentage error, root mean squared error, and correlation coefficient were calculated. For qualitative predictions (auxotrophy and C-source utilization), *F*-score, FPR, precision, accuracy, and sensitivity were calculated.

all strains (Fig. S8), leading to their designation as the *Lactobacillaceae* core fluxome (see File S4). Among these reactions, 74 exhibited a correlation coefficient of 1 with respect to the growth rates of the strains, highlighting their fundamental significance as metabolic signatures within the *Lactobacillaceae*.

Among the 30 subsystems of the core fluxome, arginine biosynthesis, cysteine and methionine metabolism, fatty acid biosynthesis, and purine metabolism have the highest fluxes. Other subsystems within the core fluxome were mostly involved in biomass production, such as pyrimidine metabolism, peptidoglycan biosynthesis, aminosugar metabolism, and glycerophospholipid biosynthesis. Nicotinate and nicotine metabolism and pantothenate and CoA biosynthesis were the only subsystems related to cofactor biosynthesis within the core fluxome.

FBA-predicted CDM media component essentiality for each GEM revealed metabolic similarities and differences in *Lactobacillaceae* (Fig. 3e). *L. plantarum* had the lowest auxotrophy count. *Pediococcus acidilactici* and *Weissella cibaria* had the highest number; *L. sakei* and *Lactobacillus paragasseri* had the most consistent auxotrophies. Isoleucine, valine, phenylalanine, and tyrosine were globally essential in all 2,446 models in PanGEM (Note S2).



**FIG 3** Characteristics of the *Lactobacillaceae* reactome. (a) Reaction presence-absence cluster map shows reaction presence and absence calls in each strain (represented by a row) in each species (color-coded as a group of rows). Strain-specific metabolic network reconstructions and reactions are represented by rows and columns, respectively. Row color refers to (Continued on next page)

**FIG 3** (Continued)

species; column color represents core, accessory, and rare reactomes. (b) Shared reactions' distribution across the 26 species. Scatter plot depicts the number of common reactions among different species. The inset bar chart represents the unique reaction count per species. Bars are annotated based on the species color map (see species color map in Fig. 1). (c) The plot visualizes the growth-correlated fluxes of the core reactome, showcasing the maximum flux of 74 reactions with a correlation coefficient of 1 against the growth rate. Each dot on the plot represents a reaction, with the *x*-axis denoting the maximum flux of reactions and the *y*-axis indicating the growth rates. (d) Intra-species reactions frequency analysis. Heatmap depicts species-specific core, accessory, and rare reactomes, color dot represents species (Fig. 1, species color map), and values within the heatmap are the percentage of intra-species core, accessory, and rare reactomes. (e) PanGEM prediction of species auxotrophies in CDM. The figure represents the essentiality of compounds across different strains within a species. Small dots indicate essential compounds, while large dots denote non-essential compounds. When dots are superimposed, it signifies that the compound's essentiality varies among the strains being essential for some but not all. The color of each dot corresponds to a specific species, as detailed in the color map provided in Fig. 1.

The PanGEM enabled the assessment of the composition of the reactome and the range of allowable flux states. The most interesting finding about the reactome composition is the prevalence of species-specific unique reactions. In terms of the core fluxome, reactions exhibited consistent flux across the PanGEM, while certain reactions displayed varying flux values within each individual GEM when compared to the others, such as PPA.

**Exploration of niche adaptation using PanGEM**

PCA analysis indicated distinct metabolic patterns across the *Lactobacillaceae* species, as evidenced by the clustering patterns shown in Fig. 4a. These patterns potentially mirror phylogenetic relationships and adaptations to various ecological niches. For instance, while genus-based clustering suggests phylogenetic similarities, species such as *L. iners* and *L. ruminis* diverge from other members of their genus, hinting at unique metabolic profiles.

*Lactobacillus* species are predominantly found in the gastrointestinal tracts of humans and animals. In contrast, *Lactiplantibacillus* species are associated with plant environments like soils and fermenting vegetables. *Lacticaseibacillus* species, on the other hand, are common in dairy settings. Interestingly, *Lactobacillus ruminis* emerged distinctively in the analysis, possibly indicative of specific adaptations within the ruminant digestive systems. Further investigations are essential to solidify these preliminary observations.

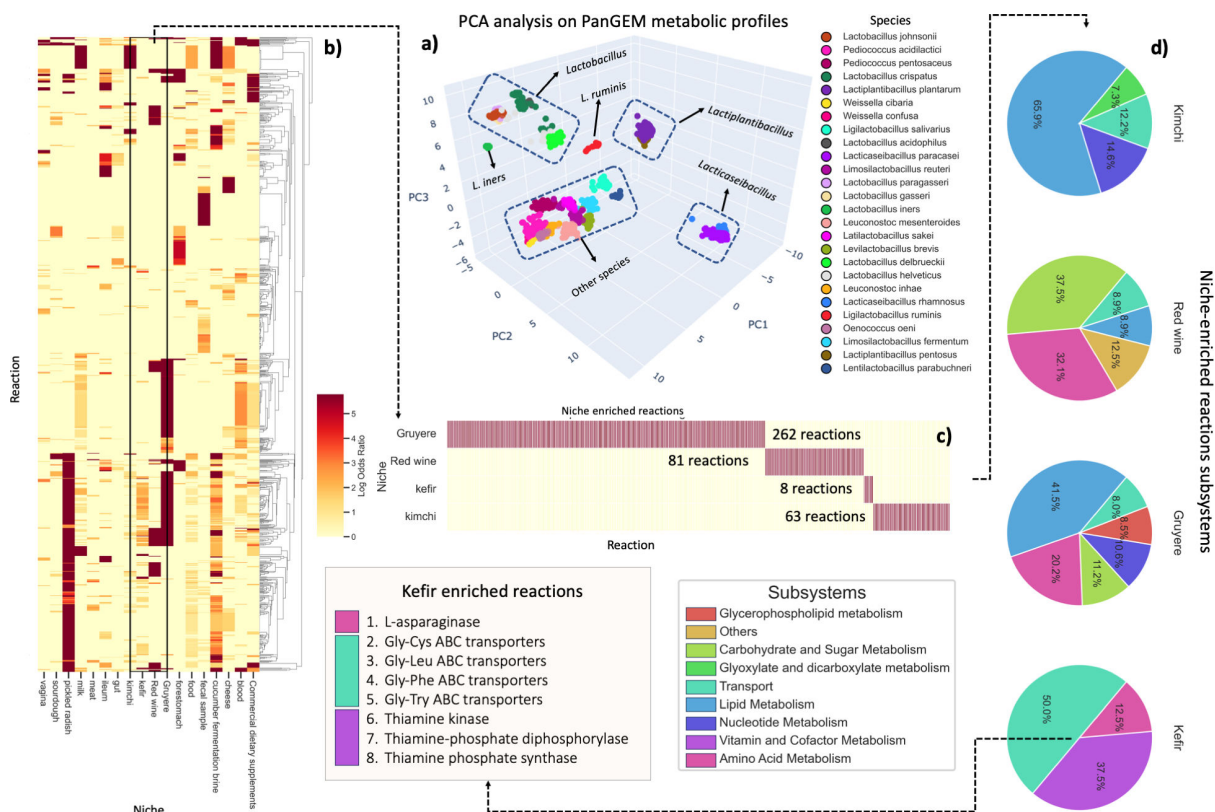
Clustering offers insights into a possible connection between bacterial metabolic profiles and their ecological backdrops. A deeper exploration of these associations could shed light on the unique roles these bacteria undertake in diverse habitats.

To quantitatively assess the relationship between metabolic profiles and their respective niches, we conducted a PERMANOVA analysis using reaction content Jaccard distances as the response variable. The PERMANOVA revealed significant effects of both isolation source and genus on reaction content variation. Isolation source explained 26.5% ( $R^2$  value of 0.256) of the total variation ( $P < 0.001$ ), while genus ( $R^2$  value of 0.912) explained 91.2% ( $P < 0.001$ ). Despite the stronger effect observed for the genus, the isolation source remained statistically significant. Further analyses were performed to identify niche-enriched reactions from sources with more than 10 strains (detailed in Fig. 4b and File S5).

For illustrative purposes, four niches were examined: kefir, red wine, kimchi, and Gruyere (Fig. 4c). It was observed that most reactions enriched in kimchi and Gruyere are associated with lipid metabolism. In red wine, reactions related to carbohydrate and sugar metabolism were predominantly enriched, while in kefir, an enrichment of transporters was noted (Fig. 4d).

Upon closer examination of kefir-enriched reactions, eight were identified. Of these, four were found to be involved in the transport of dipeptides. These dipeptides have been shown to be produced by yeast and consumed by *Lactobacillaceae* during kefir fermentation (25). Three of the reactions have been associated with thiamine





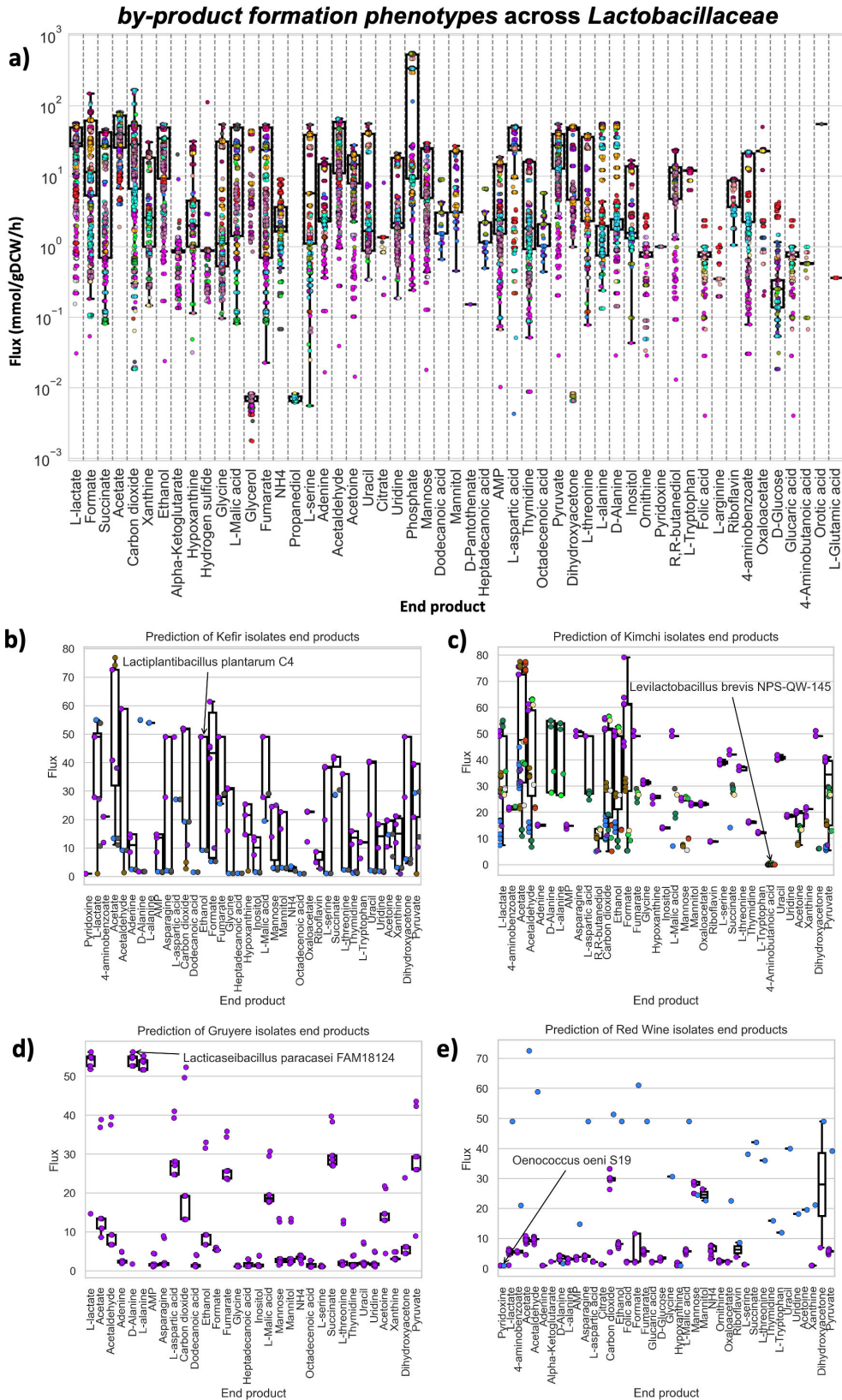
**FIG 4** Niche-enriched reactions. (a) Three-dimensional PCA plot of 2,446 bacterial strains based on their metabolic reactions, color-coded by species. The plot reveals six distinct clusters, including *Lactobacillus*, *Lactoplantibacillus*, and *Lactocaseibacillus* genus, as well as a separate cluster for *Lactobacillus ruminis* and *Lactobacillus iners*, and a final cluster containing all remaining species. (b) Cluster map illustrating niche-enriched reactions specific to isolation sources with more than 10 strains. The color map displays the log odds ratio of reaction prevalence in each niche. (c) Focused cluster map on four niches: kefir, red wine, kimchi, and Gruyere. Only reactions with a log odds ratio above five and exclusive to one niche are shown. (d) The pie charts depict the reaction distribution in kimchi, kefir, red wine, and Gruyere. Each color represents a distinct cellular subsystem, as indicated in the color map. Below, a list details kefir-enriched reactions.

metabolism, in line with the high vitamin B1 content that has been reported in kefir (26). Furthermore, the enrichment of asparaginase in kefir can be correlated with the depletion of asparagine that has been observed during its fermentation (27).

### Biotechnological potential of *Lactobacillaceae* can be discovered using PanGEM

Validated and characterized PanGEM models can be used for a variety of applications (28, 29). Metabolic by-product secretion is a defining metabolic trait of *Lactobacillaceae* and determines their uses in food production. FVA was performed to predict potential metabolite production across *Lactobacillaceae* (Fig. 5). As expected, PanGEM models showed high production rates for lactate and acetate (Fig. 5a). Metabolite connectivity analysis across all metabolic networks of PanGEM showed that the diversity of *Lactobacillaceae* end-products had a strong correlation with the connectivity of pyruvate and glutamate (Fig. S9c).

PanGEM predicted the production of several flavor compounds, such as acetoin, acetaldehyde, pyruvate, succinate, D-alanine, and ethanol, as well as the neurotransmitter gamma-aminobutyric acid (GABA). *L. paracasei* strains were predicted to be major D-Alanine producers. Ethanol producers were mostly plant-based *L. pentosus* and *Lactobacillus brevis*, dairy-based *P. acidilactici*, and commensal *L. paracasei*. Succinate producers include commensal *Lactocaseibacillus rhamnosus*, *P. acidilactici*, and *L. paracasei* isolated from unknown sources. GABA production was predicted in two



**FIG 5** PanGEM prediction of by-product formation phenotypes. (a) Prediction of maximum and minimum by-product secretion rates across PanGEM models. The box plot illustrates the capability of different species for product formation, with species represented by color-coded dots. (b) By-product profiling of kefir isolates. (c) By-product profiling of kimchi isolates. (Continued on next page)

**FIG 5** (Continued)

(d) By-product profiling of Gruyere isolates. (e) By-product profiling of red wine isolates. Across all figures in this panel, dots symbolize species, color-coded according to the species color map shown in Fig. 1. The Y-axis depicts flux ranges, calculated using FVA on CDM. The X-axis displays the secreted metabolites. Key by-products specific to the source of isolation are highlighted, with annotations indicating their predicted optimal producers.

species, commensal *P. acidilactici* and commensal *L. brevis* (Fig. S7a). PanGEM thus predicts the metabolic by-product formation on a species-specific basis.

To understand the biotechnological capabilities of product-specific isolates and how they contribute to desirable organoleptic properties of the final product, FVA was performed on GEMs belonging to isolates from four industrial products, including kimchi, kefir, Gruyere, and red wine. FVA predicted different fermentation profiles for each product. The kefir niche features two types of kefir: Tibetan kefir and a category generally labeled as “Kefir.” Within this niche, a total of 15 isolates have been identified. The isolates include nine strains of *L. plantarum*, two of *L. paracasei*, two of *L. rhamnosus*, one of *L. mesenteroides*, and one of *Pediococcus pentosaceus*. These strains collectively have the capability to produce 40 different metabolites. Among these, ethanol, succinate, acetoin, and lactic acid have been noted as important metabolites due to their roles in kefir’s nutritional profile and fermentation process (Fig. 5b).

The red wine niche includes a selection of four wine types: Chinese red wine, Nero di Troia red wine, Patagonian red wine, and a category broadly labeled as red wine. Within this selection, there are 22 strains, with 21 identified as *O. oeni* and one as *L. plantarum*. These strains are notable for their broad range of metabolite production. Specifically, among the 40 potentially producible metabolites, pyridoxine, and ornithine stand out for their importance in the wine industry, as illustrated in Fig. 5e.

The kimchi niche includes a collection of 14 unique types, including varieties such as napa cabbage kimchi and white kimchi (baek kimchi). PanGEM includes 74 strains from 10 different species within this profile. The species include *L. plantarum* (34 isolates), *W. cibaria* (12), *L. sakei* (6), *L. paracasei* (6), *L. mesenteroides* (6), *L. brevis* (4), *P. acidilactici* (2), *Weissella confusa* (1), *L. rhamnosus* (1), and *Limosilactobacillus fermentum* (1). These strains are potentially capable of producing 50 metabolites as predicted by FVA, with formate, acetaldehyde, acetoin, succinate, and GABA being particularly noteworthy for their roles in flavor development and potential health benefits in kimchi (Fig. 5c).

In the specialized niche of Gruyere cheese, we have identified a single, crucial bacterial species, *L. paracasei*, represented by 12 isolates. These isolates are responsible for the production of 50 different metabolites. Notably, formate, acetoin, and succinate have been highlighted as important metabolites, contributing significantly to the characteristic flavor and texture that define Gruyère cheese. Also, interestingly, a high production rate for D-alanine was predicted. Since Gruyere is known for its sweet taste (30) and D-alanine is a known amino acid-based sweetener (31), there might be a link between this phenotype and Gruyere’s organoleptic properties (Fig. 5d).

The pangenome metabolic reconstructions and the metabolic traits that they predict for 2,446 strains were thus highly consistent with the characteristics of the genera and species to which they belong. PanGEM thus provides a global atlas of the genetic basis for metabolic traits in the *Lactobacillaceae* family. One can select an isolation source, identify strains specifically from that source, and conduct FVA to construct an end-product profile. This profile can then be a valuable guide for directing subsequent experimental assessments.

## DISCUSSION

*Lactobacillaceae* are widely used in pharmaceutical, fermentation, and beverage industries (Table S1). As producers of various compounds, such as lactic acid, acetic acid, succinic acid, diacetyl, acetoin, GABA, sorbitol, mannitol, butanediol, propanediol, and the vitamin B family, lactic acid bacteria have a strong potential for biotransformation. Despite their enormous industrial potential, prominent market share, and the

availability of large amounts of omics data (32–35), we lack computational models that link genotypes to desirable industrial phenotypes. This study remedies this shortcoming by using pangenome analysis and metabolic reconstruction to characterize the metabolic potential of the *Lactobacillaceae* by generating 2,446 high-quality GEMs across 26 species obtained from 608 isolation sources.

Reactome analysis demonstrated that the core reactome comprises only 8% of the total reactome. In comparison, the accessory and rare reactomes account for 73% and 19%, respectively, indicating significant diversity in metabolic capabilities among *Lactobacillaceae*. Of the 65 unique reactions identified in the reactome, *L. ruminis*, *L. reuteri*, and *L. salivarius* have the highest number. These species commonly reside in the human gut and oral cavity, suggesting that acquiring novel metabolic capabilities may be crucial for bacterial adaptation and survival in complex environments. *Lactobacillus ruminis* demonstrates the highest count of unique reactions, comprising a total of 19. Certain unique reactions, such as TEICEXP, TEICEXP2, and TEICEXP3 related to teichoic acid export, were only found in *Pediococcus acidilactici*, which may be a distinct trait of this species. Previous studies have indicated that the thick cell wall of *P. acidilactici* plays a crucial role in heavy metal accumulation in the gut (36). *P. acidilactici* and *L. brevis* were also responsible for approximately 70% of microbial spoilage in beer due to hop resistance linked to the cell wall and teichoic acid structure (37). Additionally, choline trimethylamine-lyase was found only in *L. plantarum* strains. This enzyme cleaves choline to produce trimethylamine (TMA) and acetaldehyde. TMA causes several disease-associated microbial metabolites (38). Therefore, PanGEM could screen for disease-associated microbial metabolites such as TMA and their producers to avoid or reduce their usage in food industries.

Analysis of the core fluxome shows the possible activity states of core metabolism. Its properties revealed a consistent and narrow flux range (−0.15–1.26 mmol/gDCW/h) for a subset of 74 core reactions, which is predicted to have a maximal correlation with growth rate across *Lactobacillaceae*. However, 111 reactions showed high flux variation (−1,000–1,000) across different strains indicating a flexible part of the core reactome (Fig. S8b).

FBA predicted global auxotrophy, including isoleucine, valine, phenylalanine, and tyrosine, indicating a family-wide lack of complete biosynthetic pathways for these amino acids. This information could be used to generate strain-specific minimal media and screen for wild-type auxotroph strains as a predictive tool to prevent the backslipping of industrially important strains.

FVA was also performed to screen for the biotechnological potential of the species in PanGEM. As expected, PanGEM showed high production rates for lactate and acetate, which are well-known as primary organic acids produced by lactic acid bacteria (39). PanGEM predicted the production of 54 compounds across all strains, some of which are biotechnologically important compounds, consistent with the reports in the literature. These include acetoin (40), acetaldehyde (41), pyruvate (42), succinate (43), D-alanine (44), mannitol (45), formate (46), malate, citrate (46), propanediol (47), butanediol (48), ethanol (49), as well as the neurotransmitter GABA (50), and vitamins such as riboflavin (51) and pyridoxal phosphate (52).

PanGEM's analysis provides insight into niche-specific adaptations among *Lactobacillaceae* species. A targeted examination of enriched reactions in niches like kimchi, Gruyere, red wine, and kefir highlighted distinct metabolic pathways: lipid metabolism in kimchi and Gruyere, carbohydrate processes in red wine, and transport mechanisms in kefir. Notably, within kefir, reactions related to dipeptide transport and thiamine metabolism align with observed fermentation properties and vitamin B1 content. This framework underscores PanGEM's value in identifying reactions that potentially enhance bacterial fitness in specific environments.

Using PanGEM, we investigated the biotechnological capabilities of product-specific isolates and their link to the organoleptic properties of the final product. GEMs for kimchi, kefir, Gruyere, and red wine isolates were analyzed, and PanGEM predicted

the fermentation profile of related strains to identify their role in the quality of the final product. PanGEM contained 74 kimchi isolates belonging to 10 different species predicted to be capable of producing 50 different metabolites, including succinate responsible for umami taste (53), acetic acid as the main organic acid (54), acetaldehyde (55), and acetoin (56). Moreover, PanGEM predicted GABA production, which is one of the targeted metabolites for overproduction during kimchi fermentation (57). Also, red wine isolates, consisting of 21 *Oenococcus oeni* strains and 1 *L. plantarum* strain, were predicted as producers of malate (58), pyridoxine (59), and ornithine (60, 61). The kefir isolates comprised 15 strains from five species, with FVA predicting the production of 40 metabolites, including ethanol (62), succinate (63), acetoin (64), and lactic acid (64). Gruyere isolates all belonged to *L. paracasei*, predicted to be capable of the production of 50 different metabolites, including acetoin (65) and succinate (66). Interestingly, a high production rate for D-alanine was predicted, which could be related to the sweet taste of Gruyere (31).

Although PanGEM shows a high degree of precision in predicting the metabolic traits of *Lactobacillaceae*, transporter annotation remains a challenging aspect in genome-scale metabolic model reconstructions. The accuracy of such annotations directly impacts the prediction of carbon source utilization. While our model demonstrates commendable precision in amino acid essentiality predictions, our findings in carbon source utilization highlight the need for enhanced attention to transporter annotation for some strains. Achieving optimal precision and accuracy will require continued refinement and perhaps even new methodologies to improve transporter annotations. Such improvements are vital for the accurate prediction of metabolic capabilities and will play a key role in future iterations and refinements of GEMs. A fundamental reason for PanGEM's predictive power is the relative simplicity of *Lactobacillaceae* metabolism and that the GPRs reflect its genomic basis well. Another challenge in pangenome-scale reconstructions of metabolism is the absence of comprehensive experimental data regarding biomass composition for every species under study. This limitation necessitates the generalization of biomass objective functions (BOFs), which can introduce varying degrees of predictive inaccuracies in model simulations, directly correlated to the degree of BOF variability among different strains. The development of species-specific BOFs, as opposed to the application of generalized ones, could significantly enhance the precision of GEMs. Despite this potential improvement, the dearth of detailed biomass composition data persists as a fundamental constraint within the field, which could be addressed in future studies. Despite this limitation, PanGEM provides a structured and computable genetic basis for metabolic traits in the *Lactobacillaceae* family.

The economic impact of *Lactobacillaceae* is enormous, enabling industries with an annual turnover exceeding a trillion U.S. dollars. Our study presents a family-wide metabolic reconstruction of *Lactobacillaceae* using pangenome analysis and metabolic reconstruction, resulting in high-quality genome-scale metabolic models. The validated PanGEM enables the discovery and understanding of the biotechnological potential of these bacteria, developing novel applications and screening for disease-associated microbial metabolites in the food industry.

## MATERIALS AND METHODS

### Reactome reconstruction

#### *Data collection*

For the reconstruction of the *Lactobacillaceae*-specific reactome, 49 reference genomes (see Note S4 for reference genome selection rationale) (File S1) were downloaded from NCBI covering 9 genera and 33 species to capture the metabolic diversity of *Lactobacillaceae* as much as possible. We chose the 49 reference genomes based on two main factors: data availability and the presence of existing genome-scale metabolic models.

### **Genome re-annotation**

Initially, all genomes were annotated using the Prokka software (67), and stringent parameters were applied (66). A group of 13 high-quality manually annotated genomes was selected from the NCBI GenBank database as reference genomes for annotation (File S6). The output GenBank files were merged into a single GenBank file and used as the reference genome of the reactome.

### **GPR formulation and curation**

Metabolic functions were assigned to the reference genome of the reactome using the Modelseed (68) pipeline to formulate the first draft of reactome. Subsequently, all reactions went through a manual curation process. A cross-validation was performed for each reaction against the KEGG universal reactome. GPR boolean rules were checked against BioCyc (69) and KEGG databases (70). Reactions and metabolite identifiers were mapped against the BIGG database (71). The directionality of each reaction was checked based on (i) Gibbs free energy retrieved from Biocyc, (ii) BIGG reactions due to the availability of information, and (iii) previous *Lactobacillaceae* GEM reconstructions (10, 12, 14, 15, 20, 24). Metabolite charges were curated based on (i) BIGG metabolites, (ii) KEGG metabolites, (iii) ChEBI metabolites, and (iv) previous GEMs. Where information was inconsistent across databases, Marvin's suite was deployed to calculate metabolite charges at physiological pH (7.2). Reactions mass and charge balance were checked using the Cobrapy package (72) and were curated where necessary. Cellular subsystems were assigned to the reactions based on KEGG subsystems.

### **Reactome refinement**

The curated reactome was subjected to an iterative refinement process; for this goal, (i) spontaneous reactions (non-enzymatic reactions) were extracted from KEGG, BIGG, and previous reconstructions and were added to the draft reactome; (ii) exchange reactions were extracted from BIGG and previous reconstructions and were added to the draft reactome; (iii) genes missing from the reactome were identified and assigned to their corresponding reactions manually (Note S5) to form a GPR and included in the draft reactome; (d) a bidirectional blast with a similarity threshold of 60% was performed on the reference genome of reactome against BIGG reconstructions to find and add further missing reactions; (e) a BOF was taken from *L. plantarum* reconstruction (10) as a Universal *Lactobacilli*-BOF; (f) refined reactome was converted to a mathematical model using Cobrapy; (g) a primary gapfilling was performed on the reactome to find and fill metabolic gaps (Fig. S7) within the reactome using fastgapfill algorithm (73). Existence of thermodynamically infeasible cycles (TICs) was checked after curation of reaction directionality in our reactome by conducting a flux variability analysis without any assigned objective function. Having no active fluxes within the network confirmed the absence of TICs in reactome, ensuring the accuracy and reliability of the GEMs derived from this foundation. The output of this step is called *Lactobacillaceae* reactome hereafter and was used as a template for strain-specific GEM reconstruction across *Lactobacillaceae* members. Non-growth-associated maintenance was set to 1 mmol/gDCW/h, this value was calculated based on the mean NGAM of publicly available *Lactobacillaceae* GEMs (Table S3)

### **Lacto PanGEM reconstruction**

#### **Genome collection and re-annotation**

*Lactobacillaceae* genomes were downloaded from NCBI. following quality control steps were performed during genome selection (6):

1. A total of 4,783 genomes of the *Lactobacillaceae* family were retrieved from the NCBI database.
2. Genome Taxonomy Database Toolkit was deployed to re-annotate the taxonomy for all 4,783 genomes (74).
3. Furthermore, quality control and quality assurance (QC/QA) were done to get good-quality genomes. The QC/QA includes the taxonomy, number of contigs (<200), and  $N_{50}$  (>50,000).
4. Species with less than 30 genomes were excluded to maintain sample distribution.
5. Finally, a total number of 2,446 high-quality genomes of the *Lactobacillaceae* family remained to be passed to the multi-strain GEM reconstruction workflow as input (File S7).

Our selection of *Lactobacillaceae* species was based on a specific set of criteria established in our previous study (6). With the continuous growth in the availability of sequenced genomes, some species not included in our current research could be considered for inclusion in future updates of the *Lactobacillaceae* PanGEM.

### **Multi-strain reconstruction**

GPR mapping was conducted using bidirectional best hits with a similarity threshold of 70%. For this goal, quality-controlled genomes from the previous step were used as the target genome, and the reference genome of the reactome was used as a reference. The resultant homology matrix was transformed into a GPR presence-absence matrix (74). Mapped GPRs for each strain were collected to generate draft GEMs. Subsequently, all orphan, exchange, and spontaneous reactions, along with the *Lactobacilli*-specific biomass reaction, were added to all drafts.

### **Gapfilling**

Gapfilling analysis was performed on generated GEMs to ensure their functionalities. Reactome was used as a collection of candidate reactions for gapfilling; all GEMs were gapfilled on CDM. Initially, GEMs were gapfilled using fastgapfill, although it could not return a possible solution for most GEMs. Subsequently, an alternative methodology (Fig. S8), enabling large-scale gapfilling in a feasible timespan, was developed and applied.

## **Lacto PanGEM analysis**

### **Dependencies**

All analyses were performed in Python language deployed on an Azure virtual machine, statistical analysis was done using Python packages, including pandas, numpy, and scipy, and figures were generated using Matplotlib and Plotly. Cobrapy was used for GEM analysis. In all FBA-based simulations, biomass reaction was defined as an objective function.

### **Niche classification**

Isolation sources were collected from NCBI biosample when available, and a total number of 608 distinct isolation sources were identified and classified into nine larger groups including, plant-based, meat-based, dairy-based, commercial, commensal, environmental, beverages, uncategorized (undefined) food, and not reported (Fig. S1).

### **Niche-enriched reactions**

We employed the Scikit-learn package to perform principal component analysis (PCA) on our data set, aiming to reduce its dimensionality for visualization purposes. We set the PCA to produce three principal components. Following this, we used the

KMeans clustering algorithm from Scikit-learn to identify distinct clusters within the data, setting an optimal cluster count of nine based on the nine major niches. We conducted a chi-square test using the Scipy package to investigate the relationship between metabolic profiles and their respective niches. This statistical test allowed us to quantitatively assess the association between the two variables, providing insights into potential niche-specific metabolic adaptations. We calculated the odds ratio for a more granular understanding of the association between specific metabolic reactions and niches. This was achieved by constructing  $2 \times 2$  contingency tables for each reaction-niche combination and applying Fisher's exact test.

### ***Defining family-wide and species-specific core, accessory, and rare reactomes***

A reaction presence-absence matrix was constructed based on the *Lactobacillaceae* reactome and the reaction content of each strain. Family-wide core ( $F > 99\%$ ), accessory ( $15 \leq F \leq 99$ ), and rare ( $0 \leq F \leq 15$ ) reactomes were defined based on reaction frequency ( $F$ ) across PanGEM. To understand the sensitivity of core, accessory, and rare thresholds on the distribution of reactions in reactome categories, we changed the threshold for the core based on the elbow method (98.8%); this did not change the distribution of reactions in reactome categories. Results are included in Fig. S10. Similar parameters were considered for species-specific core, accessory, and rare reactomes. Reactions that were exclusive to a species (whether they could be found in one or multiple strains across a species) were called unique reactions.

### **Validation**

Validation was carried out over three different data types, including carbon source utilization, auxotrophy, and growth rate.

#### ***Carbon source utilization***

The capability of three strains for growing on 21 different C-sources was predicted and compared to the experimental results obtained from the literature. To simulate growing on different C-sources, the exchange reaction of glucose was bound to zero, and the lower bound of the exchange reaction of the target C-source was bound to 1,000 mmol/gDCW/h. Growth was reported if the predicted growth rate was above 0.01/h. FBA was used for growth simulation.

#### ***Auxotrophy***

Auxotrophy prediction was performed over six strains by bounding the lower bound of each of the 49 CDM components to zero. Auxotrophy was reported for a component if the predicted growth rate was below 0.01/h. FBA was used for growth stimulation.

#### ***CDM formulation***

The lower bound of the glucose exchange reaction was taken from the experimentally measured glucose uptake rate of *L. reuteri* (11). Lower bounds of amino acids' exchange reaction were set to 1 mmol/gDCW/h except for glutamate and aspartate, which were set to 2 mmol/gDCW/h (see Fig. S11 and Table S2 for a detailed explanation).

#### ***Growth rate***

The growth rate was predicted for five strains using FBA on CDM. Predicted growth rates were compared to the experimental results obtained from the literature.

### **Prediction of fermentation profile**

Fermentation profile for each strain within PanGEM was predicted by FBA on CDM. Strains were grouped based on isolation sources and species. Strains that were isolated



from four industrially important food products were chosen to predict their fermentation profile and find a link between their metabolic capability and the organoleptic properties of their isolation source. A set of metabolites whose exchange reaction was above zero was reported as a fermentation profile for each strain.

## AUTHOR AFFILIATIONS

<sup>1</sup>Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Lyngby, Denmark

<sup>2</sup>Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, Brisbane, Queensland, Australia

<sup>3</sup>Bioinformatics and Systems Biology Program, University of California, San Diego, La Jolla, California, USA

<sup>4</sup>Department of Pediatrics, University of California, San Diego, La Jolla, California, USA

<sup>5</sup>Center for Microbiome Innovation, University of California, San Diego, La Jolla, California, USA

<sup>6</sup>Department of Bioengineering, University of California, San Diego, La Jolla, California, USA

## AUTHOR ORCIDs

O. Ardalani  <http://orcid.org/0000-0002-0064-3802>

P. V. Phaneuf  <http://orcid.org/0000-0002-4122-6589>

O. S. Mohite  <http://orcid.org/0000-0002-3240-1656>

B. O. Palsson  <http://orcid.org/0000-0003-2357-6785>

## FUNDING

Funder	Grant(s)	Author(s)
Novo Nordisk Foundation Center for Biosustainability	NNF20CC0035580	B. O. Palsson

## DATA AVAILABILITY STATEMENT

All data and scripts are available at <https://github.com/omidard/LactoPanGEM>.

## ADDITIONAL FILES

The following material is available [online](#).

### Supplemental Material

**File S1 (mSystems00156-24-s0001.xlsx)**. Reactome reference genome data.

**File S2 (mSystems00156-24-s0002.xlsx)**. Reactome's reactions.

**File S3 (mSystems00156-24-s0003.xlsx)**. Reactome's metabolite information.

**File S4 (mSystems00156-24-s0004.xlsx)**. Corefluxome.

**File S5 (mSystems00156-24-s0005.xlsx)**. Niche-enriched reactions.

**File S6 (mSystems00156-24-s0006.xlsx)**. Genomes used for re-annotation using PROKKA.

**File S7 (mSystems00156-24-s0007.xlsx)**. *Lactobacillaceae* genome metadata.

**Supplemental information (mSystems00156-24-s0008.docx)**. Supplemental figures, tables, and notes.

## REFERENCES

- Zheng J, Wittouck S, Salvetti E, Franz CMAP, Harris HMB, Mattarelli P, O'Toole PW, Pot B, Vandamme P, Walter J, Watanabe K, Wuyts S, Felis GE, Gänzle MG, Lebeer S. 2020. A taxonomic note on the genus *Lactobacillus*: description of 23 novel genera, emended description of the genus *Lactobacillus beijerinckii* 1901, and union of *Lactobacillaceae* and *Leuconostocaceae*. *Int J Syst Evol Microbiol* 70:2782–2858. <https://doi.org/10.1099/ijsem.0.004107>
- . Global dairy industry - statistics & facts. Available from: <https://www.statista.com/topics/4649/dairy-industry/>. Retrieved 21 Mar 2023.

3. Probiotics market size, trends and global forecast to 2032. Available from: <https://www.thebusinessresearchcompany.com/report/probiotics-global-market-report>. Retrieved 21 Mar 2023.
4. GVR. Wine market size, share. Available from: <https://www.grandviewresearch.com/industry-analysis/wine-market>. Retrieved 21 Mar 2023.
5. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Rapp BA, Wheeler DL. 2002. GenBank. *Nucleic Acids Res* 30:17–20. <https://doi.org/10.1093/nar/30.1.17>
6. Rajput A, Chauhan SM, Mohite OS, Hyun JC, Ardalani O, Jahn LJ, Sommer MO, Palsson B. 2023. Pangenome analysis reveals the genetic basis for taxonomic classification of the lactobacillaceae family. Available from: <https://papers.ssrn.com/abstract=4368218>. Retrieved 12 Apr 2023.
7. Carpi FM, Coman MM, Silvi S, Picciolini M, Verdenelli MC, Napolioni V. 2022. Comprehensive pan-genome analysis of *Lactiplantibacillus plantarum* complete genomes. *J Appl Microbiol* 132:592–604. <https://doi.org/10.1111/jam.15199>
8. Koduru L, Lakshmanan M, Lee YQ, Ho P-L, Lim P-Y, Ler WX, Ng SK, Kim D, Park D-S, Banu M, Ow DSW, Lee D-Y. 2022. Systematic evaluation of genome-wide metabolic landscapes in lactic acid bacteria reveals diet- and strain-specific probiotic idiosyncrasies. *Cell Rep* 41:111735. <https://doi.org/10.1016/j.celrep.2022.111735>
9. Gu C, Kim GB, Kim WJ, Kim HU, Lee SY. 2019. Current status and applications of genome-scale metabolic models. *Genome Biol* 20:121. <https://doi.org/10.1186/s13059-019-1730-3>
10. Teusink B, Wiersma A, Molenaar D, Francke C, de Vos WM, Siezen RJ, Smid EJ. 2006. Analysis of growth of *Lactobacillus Plantarum* WCFS1 on a complex medium using a genome-scale metabolic model. *J Biol Chem* 281:40041–40048. <https://doi.org/10.1074/jbc.M606263200>
11. Kristjansdóttir T, Bosma EF, Branco Dos Santos F, Özdemir E, Herrgård MJ, França L, Ferreira B, Nielsen AT, Gudmundsson S. 2019. A metabolic reconstruction of *Lactobacillus reuteri* JCM 1112 and analysis of its potential as a cell factory. *Microb Cell Fact* 18:186. <https://doi.org/10.1186/s12934-019-1229-3>
12. Xu N, Liu J, Ai L, Liu L. 2015. Reconstruction and analysis of the genome-scale metabolic model of *Lactobacillus Casei* LC2W. *Gene* 554:140–147. <https://doi.org/10.1016/j.gene.2014.10.034>
13. Koduru L, Kim Y, Bang J, Lakshmanan M, Han NS, Lee D-Y. 2017. Genome-scale modeling and transcriptome analysis of *Leuconostoc mesenteroides* unravel the redox governed metabolic states in obligate heterofermentative lactic acid bacteria. *Sci Rep* 7:15721. <https://doi.org/10.1038/s41598-017-16026-9>
14. Mendoza SN, Cañón PM, Contreras Á, Ribbeck M, Agosin E. 2017. Genome-scale reconstruction of the metabolic network in *Oenococcus oeni* to assess wine malolactic fermentation. *Front Microbiol* 8:534. <https://doi.org/10.3389/fmicb.2017.00534>
15. Luo H, Li P, Wang H, Roos S, Ji B, Nielsen J. 2021. Genome-scale insights into the metabolic versatility of *Limosilactobacillus reuteri*. *BMC Biotechnol* 21:46. <https://doi.org/10.1186/s12896-021-00702-w>
16. Thiele I, Palsson BØ. 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc* 5:93–121. <https://doi.org/10.1038/nprot.2009.203>
17. McLeod A, Nyquist OL, Snipen L, Naterstad K, Axelsson L. 2008. Diversity of *Lactobacillus sakei* strains investigated by phenotypic and genotypic methods. *Syst Appl Microbiol* 31:393–403. <https://doi.org/10.1016/j.syapm.2008.06.002>
18. Wegkamp A, Teusink B, de Vos WM, Smid EJ. 2010. Development of a minimal growth medium for *Lactobacillus plantarum*. *Lett Appl Microbiol* 50:57–64. <https://doi.org/10.1111/j.1472-765X.2009.02752.x>
19. Hébert EM, Raya RR, de Giori GS. 2004. Nutritional requirements of *Lactobacillus delbrueckii* subsp. *lactis* in a chemically defined medium. *Curr Microbiol* 49:341–345. <https://doi.org/10.1007/s00284-004-4357-9>
20. Vinay-Lara E, Hamilton JJ, Stahl B, Broadbent JR, Reed JL, Steele JL. 2014. Genome-scale reconstruction of metabolic networks of *Lactobacillus casei* ATCC 334 and 12A. *PLoS One* 9:e110785. <https://doi.org/10.1371/journal.pone.0110785>
21. Sun J, Chen H, Qiao Y, Liu G, Leng C, Zhang Y, Lv X, Feng Z. 2019. The nutrient requirements of *Lactobacillus rhamnosus* GG and their application to fermented milk. *J Dairy Sci* 102:5971–5978. <https://doi.org/10.3168/jds.2018-15834>
22. Diaz-Muñiz I, Steele JL. 2006. Conditions required for citrate utilization during growth of *Lactobacillus casei* ATCC334 in chemically defined medium and cheddar cheese extract. *Antonie Van Leeuwenhoek* 90:233–243. <https://doi.org/10.1007/s10482-006-9078-6>
23. Kim YJ, Eom H-J, Seo E-Y, Lee DY, Kim JH, Han NS. 2012. Development of a chemically defined minimal medium for the exponential growth of *Leuconostoc mesenteroides* ATCC8293. *J Microbiol Biotechnol* 22:1518–1522. <https://doi.org/10.4014/jmb.1205.05053>
24. Özcan E, Selvi SS, Nikerel E, Teusink B, Toksoy Öner E, Çakır T. 2019. A genome-scale metabolic network of the aroma bacterium *Leuconostoc mesenteroides* subsp. *cremoris*. *Appl Microbiol Biotechnol* 103:3153–3165. <https://doi.org/10.1007/s00253-019-09630-4>
25. Pihurov M, Păcularu-Burada B, Cotârleţ M, Grigore-Gurgu L, Borda D, Stănciuc N, Kluz M, Bahrim GE. 2023. Kombucha and water kefir grains microbiomes' symbiotic contribution to postbiotics enhancement. *Foods* 12:2581. <https://doi.org/10.3390/foods12132581>
26. Farag MA, Jomaa SA, Abd El-Wahed A, R. El-Seedi H. 2020. The many faces of kefir fermented dairy products: quality characteristics, flavour chemistry, nutritional value, health benefits, and safety. *Nutrients* 12:346. <https://doi.org/10.3390/nu12020346>
27. Güler Z, Tekin A, Park YW. 2016. Comparison of biochemical changes in kefir produced from organic and conventional milk at different inoculation rates of kefir grains. *J Food Sci Nutr The* 2:008–014. <https://doi.org/10.17352/jfsnt.000003>
28. O'Brien EJ, Monk JM, Palsson BO. 2015. Using genome-scale models to predict biological capabilities. *Cell* 161:971–987. <https://doi.org/10.1016/j.cell.2015.05.019>
29. Bordbar A, Monk JM, King ZA, Palsson BO. 2014. Constraint-based models predict metabolic and associated cellular functions. *Nat Rev Genet* 15:107–120. <https://doi.org/10.1038/nrg3643>
30. Ji T, Alvarez VB, Harper WJ. 2004. Influence of starter culture ratios and warm room treatment on free fatty acid and amino acid in swiss cheese. *J Dairy Sci* 87:1986–1992. [https://doi.org/10.3168/jds.S0022-0302\(04\)70015-6](https://doi.org/10.3168/jds.S0022-0302(04)70015-6)
31. Schiffman SS, Sennewald K, Gagnon J. 1981. Comparison of taste qualities and thresholds of D- and L-amino acids. *Physiol Behav* 27:51–59. [https://doi.org/10.1016/0031-9384\(81\)90298-5](https://doi.org/10.1016/0031-9384(81)90298-5)
32. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A. 2013. NCBI GEO: archive for functional genomics data SETS—update. *Nucleic Acids Res* 41:D991–5. <https://doi.org/10.1093/nar/gks1193>
33. Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2016. GenBank. *Nucleic Acids Res* 44:D67–72. <https://doi.org/10.1093/nar/gkv1276>
34. Deutsch EW, Bandeira N, Perez-Riverol Y, Sharma V, Carver JJ, Mendoza L, Kundu DJ, Wang S, Bandla C, Kamatchinathan S, Hewapathirana S, Pullman BS, Wertz J, Sun Z, Kawano S, Okuda S, Watanabe Y, MacLean B, MacCoss MJ, Zhu Y, Ishihama Y, Vizcaino JA. 2023. The proteomeXchange consortium at 10 years: 2023 update. *Nucleic Acids Res* 51:D1539–D1548. <https://doi.org/10.1093/nar/gkac1040>
35. Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, Gillespie JJ, Gough R, Hix D, Kenyon R, et al. 2014. PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res* 42:D581–91. <https://doi.org/10.1093/nar/gkt1099>
36. Feng P, Yang J, Zhao S, Ling Z, Han R, Wu Y, Salama E-S, Kakade A, Khan A, Jin W, Zhang W, Jeon B-H, Fan J, Liu M, Mamtimin T, Liu P, Li X. 2022. Human supplementation with *Pediococcus acidilactici* GR-1 decreases heavy metals levels through modifying the gut microbiota and metabolome. *NPJ Biofilms Microbiomes* 8:63. <https://doi.org/10.1038/s41522-022-00326-8>
37. Sakamoto K, Konings WN. 2003. Beer spoilage bacteria and hop resistance. *Int J Food Microbiol* 89:105–124. [https://doi.org/10.1016/S0168-1605\(03\)00153-3](https://doi.org/10.1016/S0168-1605(03)00153-3)
38. Roberts AB, Gu X, Buffa JA, Hurd AG, Wang Z, Zhu W, Gupta N, Skye SM, Cody DB, Levison BS, Barrington WT, Russell MW, Reed JM, Duzan A, Lang JM, Fu X, Li L, Myers AJ, Rachakonda S, DiDonato JA, Brown JM, Gogonea V, Lusic AJ, Garcia-Garcia JC, Hazen SL. 2018. Development of a gut microbe-targeted nonlethal thrombotic to inhibit thrombosis potential. *Nat Med* 24:1407–1417. <https://doi.org/10.1038/s41591-018-0128-1>
39. Branen AL, Keenan TW. 1971. Diacetyl and acetoin production by *Lactobacillus casei*. *Appl Microbiol* 22:517–521. <https://doi.org/10.1128/am.22.4.517-521.1971>

40. Wang S, Li S, Zhao H, Gu P, Chen Y, Zhang B, Zhu B. 2018. Acetaldehyde released by *Lactobacillus plantarum* enhances accumulation of pyranoanthocyanins in wine during malolactic fermentation. *Food Res Int* 108:254–263. <https://doi.org/10.1016/j.foodres.2018.03.032>
41. Fucchi V, Coniglio MA, Laghi L, Rescifina A, Caruso M, Stivala A, Furneri PM. 2019. Metabolic characterization of supernatants produced by *Lactobacillus* spp. with *in vitro* anti-legionella activity. *Front Microbiol* 10:1403. <https://doi.org/10.3389/fmicb.2019.01403>
42. Dudley EG, Steele JL. 2005. Succinate production and citrate catabolism by cheddar cheese nonstarter lactobacilli. *J Appl Microbiol* 98:14–23. <https://doi.org/10.1111/j.1365-2672.2004.02440.x>
43. Mutaguchi Y, Ohmori T, Akano H, Doi K, Ohshima T. 2013. Distribution of D-amino acids in vinegars and involvement of lactic acid bacteria in the production of D-amino acids. *Springerplus* 2:691. <https://doi.org/10.1186/2193-1801-2-691>
44. Rodríguez C, Rimaux T, Fornaguera MJ, Vrancken G, de Valdez GF, De Vuyst L, Mozzi F. 2012. Mannitol production by heterofermentative *Lactobacillus reuteri* CRL 1101 and *Lactobacillus fermentum* CRL 573 in free and controlled pH batch fermentations. *Appl Microbiol Biotechnol* 93:2519–2527. <https://doi.org/10.1007/s00253-011-3617-4>
45. Zalán Z, Hudáček J, Štětina J, Chumchalová J, Halász A. 2010. Production of organic acids by *Lactobacillus* strains in three different media. *Eur Food Res Technol* 230:395–404. <https://doi.org/10.1007/s00217-009-1179-9>
46. Oude Elferink SJ, Krooneman J, Gottschal JC, Spoelstra SF, Faber F, Driehuis F. 2001. Anaerobic conversion of lactic acid to acetic acid and 1, 2-propanediol by *Lactobacillus buchneri*. *Appl Environ Microbiol* 67:125–132. <https://doi.org/10.1128/AEM.67.1.125-132.2001>
47. Alan Y. 2019. Culture fermentation of *Lactobacillus* in traditional pickled gherkins: microbial development, chemical, biogenic amine and metabolite analysis. *J Food Sci Technol* 56:3930–3939. <https://doi.org/10.1007/s13197-019-03866-8>
48. Zhang Y, Vadlani PV. 2015. Lactic acid production from biomass-derived sugars via co-fermentation of *Lactobacillus brevis* and *Lactobacillus plantarum*. *J Biosci Bioeng* 119:694–699. <https://doi.org/10.1016/j.jbiosc.2014.10.027>
49. Yunes RA, Poluektova EU, Dyachkova MS, Klimina KM, Kovtun AS, Averina OV, Orlova VS, Danilenko VN. 2016. GABA production and structure of gadB/gadC genes in *Lactobacillus* and bifidobacterium strains from human microbiota. *Anaerobe* 42:197–204. <https://doi.org/10.1016/j.anaerobe.2016.10.011>
50. Mohedano ML, Hernández-Reco S, Yépez A, Requena T, Martínez-Cuesta MC, Peláez C, Russo P, LeBlanc JG, Spano G, Aznar R, López P. 2019. Real-time detection of riboflavin production by *Lactobacillus plantarum* strains and tracking of their gastrointestinal survival and functionality *in vitro* and *in vivo* using mCherry labeling. *Front Microbiol* 10:1748. <https://doi.org/10.3389/fmicb.2019.01748>
51. Rattanaporn S, Saowanit T, Yane T. 2018. Production of pyridoxal 5-phosphate and pyridoxine by *Lactobacillus pentosus* L471-A. *Res J Biotechnol* 13:16–23.
52. Hajeb P, Jinap S. 2015. Umami taste components and their sources in asian foods. *Crit Rev Food Sci Nutr* 55:778–791. <https://doi.org/10.1080/10408398.2012.678422>
53. Shim S-M, Kim JY, Lee SM, Park J-B, Oh S-K, Kim Y-S. 2012. Profiling of fermentative metabolites in kimchi: volatile and non-volatile organic acids. *J Korean Soc Appl Biol Chem* 55:463–469. <https://doi.org/10.1007/s13765-012-2014-8>
54. Hong SP, Lee EJ, Kim YH, Ahn DU. 2016. Effect of fermentation temperature on the volatile composition of kimchi. *J Food Sci* 81:C2623–C2629. <https://doi.org/10.1111/1750-3841.13517>
55. Chun BH, Kim KH, Jeon HH, Lee SH, Jeon CO. 2017. Pan-genomic and transcriptomic analyses of leuconostoc mesenteroides provide insights into its genomic and metabolic features and roles in kimchi fermentation. *Sci Rep* 7:11504. <https://doi.org/10.1038/s41598-017-12016-z>
56. Cho Y-R, Chang J-Y, Chang H-C. 2007. Production of  $\gamma$ -aminobutyric acid (GABA) by *Lactobacillus buchneri* isolated from kimchi and its neuroprotective effect on neuronal cells. *J Microbiol Biotechnol* 17:104–109.
57. Gil-Sánchez I, Bartolomé Suáldea B, Victoria Moreno-Arribas M. 2019. Chapter 6, Malolactic fermentation, p 85–98. In Morata A (ed), *Red wine technology*. Academic Press.
58. Kondrashov A, Ševčík R, Benáková H, Košťálová M, Štípek S. 2009. The key role of grape variety for antioxidant capacity of red wines. *E Spen Eur E J Clin Nutr Metab* 4:e41–e46. <https://doi.org/10.1016/j.ejcnm.2008.10.004>
59. Spano G, Massa S, Arena ME, de Nadra MCM. 2007. Arginine metabolism in wine *Lactobacillus plantarum*: *in vitro* activities of the enzymes arginine deiminase (ADI) and ornithine transcarbamylase (OTCase). *Ann. Microbiol* 57:67–70. <https://doi.org/10.1007/BF03175052>
60. Kuensch U, Temperli A, Mayer K. 1974. Conversion of arginine to ornithine during malo-lactic fermentation in red swiss wine. *Am J Enol Vitic* 25:191–193. <https://doi.org/10.5344/ajev.1974.25.4.191>
61. Gul O, Mortas M, Atalar I, Dervisoglu M, Kahyaoglu T. 2015. Manufacture and characterization of kefir made from cow and buffalo milk, using kefir grain and starter culture. *J Dairy Sci* 98:1517–1525. <https://doi.org/10.3168/jds.2014-8755>
62. Leroi F, Pidoux M. 1993. Characterization of interactions between *Lactobacillus hilgardii* and *Saccharomyces florentinus* isolated from sugary kefir grains. *Journal of Applied Bacteriology* 74:54–60. <https://doi.org/10.1111/j.1365-2672.1993.tb02996.x>
63. Guzel-Seydim Z, Seydim AC, Greene AK. 2000. Organic acids and volatile flavor components evolved during refrigerated storage of kefir. *J Dairy Sci* 83:275–277. [https://doi.org/10.3168/jds.S0022-0302\(00\)74874-0](https://doi.org/10.3168/jds.S0022-0302(00)74874-0)
64. Gilliland SE. 2017. *Bacterial starter cultures for foods*, p 213. CRC Press, Taylor & Francis Group.
65. Crow VL, Turner KW. 1986. Effect of succinate production on other fermentation products in swiss-type cheese. *NZ J Dairy Sci Technol*. Available from: <https://agris.fao.org/agris-search/search.do?recordID=US201301421599>. Retrieved 21 Mar 2023.
66. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>
67. Seaver SMD, Liu F, Zhang Q, Jeffries J, Faria JP, Edirisinghe JN, Mundy M, Chia N, Noor E, Beber ME, Best AA, DeJongh M, Kimbrel JA, D'haeseleer P, McCorkle SR, Bolton JR, Pearson E, Canon S, Wood-Charlson EM, Cottingham RW, Arkin AP, Henry CS. 2021. The modelseed biochemistry database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. *Nucleic Acids Res* 49:D575–D588. <https://doi.org/10.1093/nar/gkaa746>
68. Caspi R, Billington R, Fulcher CA, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Midford PE, Ong WK, Paley S, Subhraveti P, Karp PD. 2019. Biocyc: a genomic and metabolic web portal with multiple omics analytical tools. *Faseb J* 33. [https://doi.org/10.1096/fasebj.2019.33.1\\_supplement.473.2](https://doi.org/10.1096/fasebj.2019.33.1_supplement.473.2)
69. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 45:D353–D361. <https://doi.org/10.1093/nar/gkw1092>
70. Norsigian Charles J, Pusarla N, McConn JL, Yurkovich JT, Dräger A, Palsson BO, King Z. 2020. BIGG models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree. *Nucleic Acids Res* 48:D402–D406. <https://doi.org/10.1093/nar/gkz1054>
71. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. 2013. COBRApy: constraints-based reconstruction and analysis for python. *BMC Syst Biol* 7:74. <https://doi.org/10.1186/1752-0509-7-74>
72. Thiele I, Vlassis N, Fleming RMT. 2014. fastGapFill: efficient gap filling in metabolic networks. *Bioinformatics* 30:2529–2531. <https://doi.org/10.1093/bioinformatics/btu321>
73. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2020. GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 36:1925–1927. <https://doi.org/10.1093/bioinformatics/bt2848>
74. Norsigian C.J, Fang X, Seif Y, Monk JM, Palsson BO. 2020. A workflow for generating multi-strain genome-scale metabolic models of prokaryotes. *Nat Protoc* 15:1–14. <https://doi.org/10.1038/s41596-019-0254-3>