

Conserved fermentation-associated genomic architecture and lineage-associated variation in lactic acid bacteria

Bogovid Živković

bogovid1981@gmail.com

Article

Keywords: lactic acid bacteria, comparative genomics, Lactiplantibacillus, fermentation, pyruvate metabolism, metabolic capacity

Posted Date: September 1st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10676934/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Lactic acid bacteria (LAB) display extensive genomic and metabolic diversity, yet the extent to which this diversity affects the genomic architecture of central fermentation functions remains unclear. We analysed 507 curated bacterial genomes using standardized Prokka annotation and a manually defined framework of seven fermentation-associated modules comprising 28 unique biological markers. Within *Lactiplantibacillus* (n = 462), 27 of 28 markers were universally detected, while *pfl* occurred in 459 genomes (99.35%). All 28 markers were detected in every genome of the four multi-genome species *L. plantarum* (n = 411), *L. pentosus* (n = 21), *L. argentoratensis* (n = 16), and *L. paraplantarum* (n = 9). Across multi-genome LAB genera, acetate-associated capacity was completely conserved, whereas lactate formation and alternative pyruvate-fate modules showed the greatest lineage-associated variation. These results reveal a highly conserved fermentation-associated genomic framework accompanied by concentrated variation in selected downstream metabolic functions. The findings support a distinction between conserved genomic capacity and realized fermentation phenotype and identify pyruvate routing as a major axis of potential functional diversification among sampled LAB lineages.

Introduction

Lactic acid bacteria (LAB) comprise a physiologically coherent but phylogenetically and ecologically diverse group of Gram-positive bacteria whose carbohydrate metabolism is commonly associated with substantial lactate formation. Comparative genomic studies have shown that LAB evolution combines conservation of essential cellular functions with extensive lineage-specific gene loss, acquisition and niche adaptation [1, 2]. The historical genus *Lactobacillus* itself was recently reorganized into multiple genera, including *Lactiplantibacillus*, reflecting deep genomic diversification that had been obscured by traditional nomenclature [3]. Consequently, a shared fermentative lifestyle does not imply a uniform genomic architecture across LAB lineages.

At the metabolic level, LAB channel carbon through a limited number of central biochemical frameworks, particularly Embden-Meyerhof-Parnas glycolysis and phosphoketolase/pentose-phosphate-associated routes. Classical homofermentative and heterofermentative descriptions capture important differences in carbon and redox balance, but the actual distribution of end products can change with substrate, oxygen availability, pH, growth phase and energetic demand [4, 5]. Pyruvate is especially important as a branch point from which carbon can be directed toward lactate, acetate, ethanol, formate or acetoin-related products. This organization suggests that conserved central carbon-processing capacity may coexist with greater variability in downstream routing functions.

Genome-scale comparisons have begun to connect this metabolic diversity to genomic variation. Niche-associated gene sets have been identified across LAB, while genotype-phenotype studies have shown that carbohydrate utilization can often be associated with particular genomic features [6, 7]. At the same time, these studies also demonstrate the limitations of interpreting gene presence as a direct phenotype:

paralogues, incomplete annotations, regulatory state, transport capacity and environmental conditions can all modify the relationship between encoded functions and observed metabolism.

Lactiplantibacillus plantarum provides a particularly informative model for this problem because it occupies diverse food, plant and host-associated environments and has a large, open and heterogeneous accessory genome. Pan-genome and comparative-genomic analyses have repeatedly documented substantial strain-to-strain variation in this species and in the wider *Lactiplantibacillus* lineage [8–11]. Such genome-wide diversity raises a specific question: are the genes most directly associated with fermentation equally variable, or is there a more conserved functional core embedded within a flexible genomic background?

Systems-level studies reinforce the need to distinguish genomic repertoire from metabolic behaviour. A genome-scale model of *L. plantarum* WCFS1 showed extensive alternative routing and cofactor-dependent flux solutions, whereas physiological analysis under very slow growth demonstrated major state-dependent changes in metabolism and gene expression [12, 13]. These findings support a view in which the genome defines a set of biochemical possibilities rather than a single fixed metabolic output.

Here, we tested whether LAB share a conserved fermentation-associated genomic framework while lineage-associated diversity is concentrated in selected carbon-routing, pyruvate-fate and redox-balancing functions. We analysed a curated collection of 507 genomes using a standardized Prokka annotation workflow and a manually curated framework of seven fermentation-associated functional modules (FM01-FM07). Thirty-one validated catalogue assignments were collapsed to 28 unique biological markers for marker-level analyses. We quantified marker prevalence and module completeness across taxa, with particular attention to the densely sampled *Lactiplantibacillus* lineage and to descriptive inter-genus variation among LAB genera represented by more than one genome.

Results

Construction and validation of the fermentation-associated marker framework

The curated collection contained 507 genomes: 486 core-LAB genomes, 10 fermentative non-LAB genomes, 8 external comparators and 3 *Enterococcus* genomes. All genomes were processed through the same Prokka 1.14.6 annotation workflow, and annotation completion was confirmed for all 507 assemblies. The fermentation catalogue comprised seven modules covering EMP glycolysis (FM01), lactate formation (FM02), phosphoketolase/pentose-phosphate-associated metabolism (FM03), the acetate branch (FM04), the ethanol branch (FM05), alternative pyruvate fates (FM06) and NAD/redox balance (FM07).

After functional curation, 31 catalogue assignments were accepted for analysis. Three biological markers (*ldh*, *ldhD* and *adhE*) participate in more than one module and were therefore represented by

repeated functional assignments but collapsed for marker-level prevalence analysis, yielding 28 unique biological markers. Primary detection was based on exact normalized gene-symbol matches and/or EC-number matches among Prokka CDS features. Product annotations were retained only for post hoc quality control and were not used to rescue primary calls.

Marker-level contrasts against the non-LAB groups were treated as descriptive because the comparator sets were small and taxonomically heterogeneous. The largest positive contrasts within the curated collection involved *pox*, *ldhD*, *alsD*, *npr* and *noxE*, but these patterns were not interpreted as LAB-specific biomarkers. Quality-control review of marker evidence identified 14 high-confidence markers, 6 markers considered high-confidence when product consistency was taken into account, 6 with mixed evidence modes, one gene-only marker requiring additional caution (*adhE*) and one marker with variable detection frequency (*pfl*).

Near-complete conservation of the fermentation-associated repertoire in *Lactiplantibacillus*

Among 462 *Lactiplantibacillus* genomes, 27 of the 28 unique fermentation-associated markers were detected in every genome. The only non-universal marker was *pfl*, which was detected in 459/462 genomes (99.35%) and therefore met the operational NEAR_CORE criterion. The three genomes without a detected *pfl* marker were *L. koreensis* GCF_002970915.1, *L. kimchii* GCF_002970935.1 and *L. brownii* GCF_045862825.1. Each of these species was represented by a single genome, precluding species-level interpretation of the absence.

At the species level, all 28 markers were detected in every genome of the four *Lactiplantibacillus* species represented by multiple genomes: *L. plantarum* (411/411), *L. pentosus* (21/21), *L. argentoratensis* (16/16) and *L. paraplantarum* (9/9). Thus, 457 genomes spanning these four species contained the complete 28-marker repertoire under the standardized detection framework. The result was especially notable for *L. plantarum*, which accounted for 411 genomes yet showed no marker-level presence/absence variation in this targeted set.

Fermentation-module conservation across LAB genera

The 486 core-LAB genomes represented seven genera: *Lactiplantibacillus* (n = 462), *Lactobacillus* (n = 15), *Leuconostoc* (n = 3), *Weissella* (n = 3), *Lactococcus* (n = 1), *Pediococcus* (n = 1) and *Streptococcus* (n = 1). Genus-level module profiles showed complete representation of FM04 in every sampled genus. FM03 was also strongly conserved in the four genera represented by more than one genome, with a mean prevalence of 0.983, followed by FM05 (0.942), FM01 (0.910), FM07 (0.883), FM02 (0.825) and FM06 (0.704).

Lactiplantibacillus showed complete mean marker representation for FM01-FM05 and FM07, while FM06 reached 0.998 because of the three *pfl*-negative genomes. *Lactobacillus* showed partial representation of several modules, with the lowest mean values in FM06 (0.650) and FM05 (0.767). *Leuconostoc* showed its lowest value in FM02 (0.500), whereas *Weissella* showed its lowest value in FM06 (0.500). The

singleton genera were retained in the heatmap for descriptive completeness but were not used in the principal inter-genus variability analysis (Fig. 1).

Inter-genus variability is concentrated in lactate formation and alternative pyruvate fates

To reduce the influence of singleton genera, inter-genus variability was quantified across the four core-LAB genera represented by more than one genome: *Lactiplantibacillus*, *Lactobacillus*, *Leuconostoc* and *Weissella*. Genera were equally weighted in these descriptive calculations. FM02 and FM06 showed the greatest variation, but by different metrics. FM02 had the largest prevalence range (0.500; SD = 0.2283; CV = 0.2768), whereas FM06 had the highest coefficient of variation (CV = 0.2987; range = 0.4984; SD = 0.2102).

FM05 ranked third in inter-genus variability (range = 0.2333; CV = 0.1239), followed by FM07 and FM01. FM03 was highly stable (range = 0.0556; CV = 0.0269), and FM04 showed no detectable inter-genus variability (mean = 1.000; SD = 0; range = 0; CV = 0). Because genus sample sizes were strongly imbalanced and the genus-level summaries were not comparable independent samples, no inferential significance testing was applied. The final module-level synthesis is presented in Table 1.

Table 1
Conservation and inter-genus variability of fermentation-associated functional modules.

Module	Markers	<i>Lactiplantibacillus</i> status	Mean prevalence	SD	Range	CV
FM01	10	10/10 CORE	0.910	0.0747	0.1667	0.0821
FM02	2	2/2 CORE	0.825	0.2283	0.5000	0.2768
FM03	6	6/6 CORE	0.983	0.0265	0.0556	0.0269
FM04	2	2/2 CORE	1.000	0.0000	0.0000	0.0000
FM05	2	2/2 CORE	0.942	0.1166	0.2333	0.1239
FM06	4	3 CORE + 1 NEAR_CORE	0.704	0.2102	0.4984	0.2987
FM07	5	5/5 CORE	0.883	0.0839	0.2000	0.0950

Values for mean prevalence, SD, range and CV were calculated across the four core-LAB genera represented by more than one genome (*Lactiplantibacillus*, *Lactobacillus*, *Leuconostoc* and *Weissella*), with genera equally weighted. Singleton genera were excluded from this variability analysis. CORE and NEAR_CORE are operational dataset-specific categories.

Discussion

The principal result of this study is a hierarchical pattern of conservation rather than a uniform fermentation signature across all LAB. Central fermentation-associated functions were broadly conserved, but variation was concentrated in selected modules involved in lactate formation, alternative pyruvate fates and redox-linked metabolism. The contrast was strongest within *Lactiplantibacillus*, where 27 of 28 markers were universal and *pfl* was present in 99.35% of genomes. This near-invariance is striking because *L. plantarum* and related taxa are known to possess extensive accessory-genome diversity [8–11]. The targeted fermentation repertoire is therefore substantially more conserved than the genome as a whole.

This distinction has an important biological implication. The absence of presence/absence variation among all 28 markers in 411 *L. plantarum* genomes does not imply that these strains have identical fermentation phenotypes. Instead, it suggests that strain-level differences may often arise at other layers of biological organization: allelic variation, enzyme kinetics, copy number, transporter repertoire, transcriptional regulation, cofactor availability and environment-dependent flux partitioning. Genome-scale modelling and slow-growth physiology have already shown that *L. plantarum* can support multiple feasible flux states and can reorganize metabolism extensively without requiring wholesale changes in gene content [12, 13]. Our results extend that perspective by showing that a targeted set of fermentation-associated genes can remain nearly invariant across a genomically diverse lineage.

FM01 and FM03 illustrate the conserved part of this architecture. The EMP framework was complete in *Lactiplantibacillus* and remained highly represented across the multi-genome LAB genera. FM03, which combines phosphoketolase with pentose-phosphate-associated markers, was similarly conserved. Importantly, this should not be interpreted as direct evidence of a heterofermentative phenotype. Classical fermentation categories reflect actual carbon flux and end-product stoichiometry, whereas our analysis detects encoded enzymatic capacity [4, 5]. The widespread presence of FM03 markers is therefore best interpreted as conservation of phosphoketolase/pentose-phosphate-associated metabolic potential rather than as a categorical prediction of fermentation type.

FM02 showed the largest inter-genus prevalence range. This module contains L- and D-lactate dehydrogenase markers, enzymes whose functions extend beyond a simplistic terminal pyruvate-to-lactate step. In *L. plantarum*, NAD-dependent lactate dehydrogenases have been shown experimentally to participate in aerobic lactate utilization during early stationary phase [14]. Consequently, differences in marker representation across genera should not be read as a binary distinction between lactate-producing and non-lactate-producing taxa. They instead indicate lineage-associated differences in the specific LDH repertoire captured by the present annotation framework.

The most biologically diverse module by coefficient of variation was FM06, which includes *pfl*, *alsS*, *alsD* and *pox*. These functions sit around pyruvate, a major metabolic branch point, and provide multiple routes for redistributing carbon, electrons and ATP generation. Pyruvate oxidase in *L. plantarum* has been characterized genetically and physiologically, with its activity strongly influenced by oxygen and carbon catabolite conditions [15, 16]. The acetoin branch provides another example of condition-

dependent routing: pyruvate can be converted efficiently to acetoin under controlled fermentation conditions [17], and the distribution of end products changes with pH and oxygen availability [18]. Acetoin production can also contribute to pH homeostasis by diverting pyruvate away from more acidic products [19]. PFL-related metabolism provides an anaerobic alternative, and experimental studies have linked formate production and pyruvate routing to physiological state and substrate metabolism in *L. plantarum* [20, 21]. Together, these observations explain why presence of FM06 genes should be interpreted as branch-point capacity rather than as a fixed end-product profile.

FM05 and FM07 further demonstrate the coupling between carbon routing and redox state. The ethanol-associated marker *adhE* is functionally important but context dependent: deletion and regulatory studies in *L. plantarum* have shown that AdhE-mediated ethanol formation can depend strongly on carbon source and on transcriptional control involving redox-responsive regulators [22]. Similarly, NADH oxidase and NADH peroxidase provide routes for oxygen-dependent NADH reoxidation. Classical biochemical work demonstrated coordinated oxygen-consuming reactions involving these enzymes in *L. plantarum* [23], while later aeration experiments showed that oxygen supply modifies lactate-to-acetate conversion, peroxide formation and overall redox metabolism [24]. Such results are consistent with the broader stress physiology of LAB, in which metabolic output emerges from an interaction between encoded functions, regulatory responses and environmental stressors [25].

The complete conservation of FM04 across the sampled genera is also notable. The *pta-ackA* branch links acetyl-CoA/acetyl-phosphate metabolism to acetate formation and ATP generation, and its uniform representation suggests that acetate-associated capacity is part of a broadly conserved metabolic framework in this collection. However, the experimental pyruvate-oxidase literature illustrates why this genomic capacity should not be equated with constitutive acetate production: oxygen availability, carbon repression and growth phase can determine whether the pathway is actually used [15, 16, 24].

Taken together, the results support the concept of a metabolic possibility space. The genomic repertoire constrains which reactions are possible, but it does not uniquely determine which routes carry flux under a given condition. In this sense, conserved marker presence can coexist with substantial phenotypic diversity. The distinction is particularly important for comparative genomics studies that attempt to infer fermentation behaviour from genome annotations alone. Presence/absence analysis is well suited to defining the boundaries of potential capacity, but experimental phenotype, transcript abundance, enzyme activity and metabolic flux remain necessary to identify realized physiology [7, 12, 13].

Several limitations should be considered. First, the collection was intentionally curated rather than population-balanced and was dominated by *Lactiplantibacillus*, particularly *L. plantarum*. Genus-level comparisons are therefore descriptive and should not be interpreted as estimates of global genus frequencies. Second, three LAB genera were represented by single genomes and were excluded from the principal variability ranking. Third, marker calls depended on standardized Prokka annotation and a gene-symbol/EC-number rule rather than sequence-level orthology models for every marker. Post hoc product review supported the major calls, but *adhE* remained a lower-confidence marker because gene-based

annotations encompassed a broader alcohol/aldehyde dehydrogenase product space. Finally, the study contains no transcriptomic, proteomic, metabolomic or experimentally measured fermentation endpoints, so it cannot establish flux or phenotype directly. These limitations are partly offset by uniform annotation, explicit marker definitions, evidence-level hit tables and a fully archived reproducibility workflow.

In conclusion, our results reveal a hierarchical organization of fermentation-associated genomic conservation across the sampled LAB. A highly conserved framework of central carbohydrate-processing, acetate-associated and phosphoketolase/pentose-phosphate-associated functions coexists with greater lineage-associated variation in lactate formation, alternative pyruvate fates and redox-associated functions. The near-invariant 28-marker repertoire observed within *Lactiplantibacillus*, despite the broader genomic diversity known for this lineage, suggests that diversification of fermentation phenotypes may depend substantially on regulation and deployment of a conserved enzymatic repertoire rather than simple gene presence or absence. The genomic repertoire should therefore be viewed as defining a metabolic possibility space rather than a fixed fermentation phenotype. Future integration of transcriptomic, proteomic and experimentally measured fermentation data will be required to determine how this conserved genomic architecture is translated into strain- and condition-specific metabolic behaviour.

Methods

Genome collection and dataset curation

A curated genome collection of 507 bacterial assemblies was used. The frozen dataset snapshot (LAB Genome Landscape v1.0) comprised 486 genomes assigned to the core-LAB group, 10 fermentative non-LAB genomes, 8 external comparators and 3 *Enterococcus* genomes. The core-LAB collection contained 462 *Lactiplantibacillus*, 15 *Lactobacillus*, 3 *Leuconostoc*, 3 *Weissella*, 1 *Lactococcus*, 1 *Pediococcus* and 1 *Streptococcus* genome. Current taxonomy was used in the processed metadata, including the post-2020 *Lactobacillus* reclassification. Genome identity was tracked by a stable internal Genome_ID and the corresponding NCBI RefSeq accession. Assembly paths were validated before analysis, and the dataset was frozen before functional annotation and downstream comparisons.

Standardized genome annotation

All 507 assemblies were annotated locally using Prokka 1.14.6 with a uniform command structure. Genus and species values were parsed from the curated metadata and supplied to Prokka. Each genome was processed with one CPU per Prokka invocation, with up to two genomes annotated concurrently in the batch runner. Completion was audited using per-genome status files, logs and expected output files; 507/507 genomes completed successfully. Prokka TSV files were used as the primary source for

fermentation-marker screening because they provide locus tag, feature type, gene symbol, EC number and product description in a consistent tabular representation.

Definition and curation of fermentation-associated functional modules

Seven modules were defined to represent functions directly connected to central fermentative carbon routing: FM01, EMP glycolysis; FM02, lactate formation; FM03, phosphoketolase/pentose-phosphate-associated metabolism; FM04, acetate branch; FM05, ethanol branch; FM06, alternative pyruvate fates; and FM07, NAD/redox balance. The catalogue recorded gene symbol, enzyme name, EC number, KEGG orthology identifier, reaction role and interpretation for each entry [27]. Two preliminary generic entries corresponding to non-specific *aco* and *nox* labels were excluded during refinement because they did not unambiguously define a single target function. The validated catalogue contained 31 functional assignments. Because *ldh*, *ldhD* and *adhE* were intentionally assigned to more than one module, these 31 assignments corresponded to 28 unique biological markers for marker-level prevalence analyses.

Marker detection and evidence classification

Only Prokka features annotated as CDS were considered. Gene symbols were normalized by removing a terminal underscore followed by digits (for example, *ldh_1* to *ldh*), and comparison with catalogue symbols was case-insensitive and otherwise exact. EC-number strings were split on slash, semicolon, comma or whitespace separators and compared as sets. A CDS was considered a primary hit when either the normalized gene symbol matched the catalogue gene symbol or the CDS and catalogue shared at least one EC number. Evidence was recorded as gene + EC, gene only or EC only. Product descriptions were retained for post hoc quality control but were not used as a primary matching criterion. All matching loci were retained in the evidence-level hit table, while multiple loci matching the same catalogue entry were collapsed to binary presence for the primary matrix.

Binary marker presence and module coverage

For each genome-marker pair, binary presence was defined as 1 when at least one qualifying CDS was detected and 0 otherwise. Genome-level module coverage was then calculated as the fraction of unique biological markers assigned to a module that were present in that genome:

$$C_{gk} = \frac{\sum_{m \in M} P_{gm}}{n_k}$$

1

where the denominator is the number of unique biological markers assigned to module *k*. Module coverage therefore ranges from 0 to 1.

Taxon-level marker prevalence and operational categories

For each taxon-marker combination, marker prevalence was calculated as the number of marker-positive genomes divided by the total number of genomes in that taxon:

$$F_{tm} = \frac{N_{tm}}{N_t}$$

2

In this expression, the numerator is the number of genomes in the taxon that are positive for the marker and the denominator is the total number of genomes in that taxon. For the *Lactiplantibacillus* landscape, prevalence categories were defined operationally as CORE = 1.00, NEAR_CORE ≥ 0.95 and < 1.00 , VARIABLE ≥ 0.15 and < 0.95 , and RARE < 0.15 . These thresholds are dataset-specific descriptive categories rather than universal bacterial core-genome definitions.

Genus-level module comparisons and descriptive variability

For each genus, module values were calculated as the mean prevalence of the unique markers assigned to that module. The principal inter-genus variability analysis included only core-LAB genera represented by more than one genome: *Lactiplantibacillus* (n = 462), *Lactobacillus* (n = 15), *Leuconostoc* (n = 3) and *Weissella* (n = 3). Genera were equally weighted. For each module, the mean, sample standard deviation, range and coefficient of variation were calculated across the four genus-level module values. The coefficient of variation was defined as:

$$CV_k = \frac{s_k}{\bar{x}_k}$$

3

In this expression, the numerator is the sample standard deviation of the genus-level values for a module and the denominator is their mean. Variability ranking was based primarily on decreasing range, with SD used as a secondary criterion. No inferential significance tests were applied because the collection was strongly imbalanced among genera and the genus-level summaries were not treated as comparable independent samples.

Quality control and interpretation safeguards

Marker evidence was audited using gene/EC evidence modes and product consistency. Representative hits across species were manually reviewed for biologically coherent product annotations and distinct loci. In an additional distinct-locus check focused on five leading markers (*pox*, *ldhD*, *alsD*, *npr* and *noxE*), all 472 genomes containing all five markers contained at least five distinct loci, with no problematic cases. Marker-level confidence categories were used to distinguish straightforward calls from markers requiring product-context review. Because the study measures encoded genomic capacity, module and

marker presence were not interpreted as proof of pathway activity, end-product formation or metabolic flux.

Reproducibility and software environment

The complete downstream analysis was regenerated from the 507-genome marker matrix using a dedicated reproducibility script. The regenerated genus-by-module matrix and final module-results table matched the original manuscript-supporting outputs, including all numerical values and final text fields. Figure 1 was regenerated from the reproduced genus-by-module matrix. The public reproducibility package contains the curated public metadata, RefSeq accession manifest, validated catalogues, evidence-level marker hits, presence/absence matrices, result tables, figure files, scripts and software-version record. Key software versions included Python 3.13.13, Prokka 1.14.6, pandas 3.0.5, NumPy 2.5.1 and Matplotlib 3.11.1.

Code availability

The computational workflow used for standardized Prokka annotation, fermentation-marker detection, downstream comparative analyses and generation of Fig. 1 is available in the GitHub repository <https://github.com/bogovid/LAB-Genome-Landscape-Fermentation>, release v1.0.0, and is permanently archived in Zenodo under DOI 10.5281/zenodo.21900316 [26].

Use of a large language model

A large language model (ChatGPT, OpenAI) was used during development of the computational workflow and manuscript preparation for interactive assistance with code development, debugging, organization of analytical steps and language editing. Analytical outputs were generated using the documented local computational workflow, and scientific interpretation, verification of results and responsibility for the final manuscript remained with the author.

Ethics statement

This study used publicly available microbial genome assemblies and did not involve human participants, human-derived samples, live animals or newly collected biological specimens.

Data availability

The genome assemblies analysed in this study are publicly available through NCBI RefSeq. Assembly accession numbers for all 507 genomes, together with the curated public metadata, validated fermentation-marker catalogue, marker presence/absence matrices, supporting result tables and Fig. 1 source data, are archived in Zenodo under DOI 10.5281/zenodo.21900316 [26]. The original RefSeq FASTA assemblies are not redistributed in the archive.

Declarations

Funding

This research received no external funding.

Author contributions

B.Ž. conceived the study, curated the genome collection, designed and executed the computational analyses, interpreted the results, prepared the reproducibility package, and wrote and revised the manuscript.

Additional information

Competing interests

The author declares no competing interests.

Correspondence

Correspondence and requests for materials should be addressed to Bogovid Živković (bogovid1981@gmail.com).

References

1. Makarova, K. et al. Comparative genomics of the lactic acid bacteria. *Proc. Natl. Acad. Sci. USA* **103**, 15611–15616; (2006). 10.1073/pnas.0607117103
2. Makarova, K. S. & Koonin, E. V. Evolutionary genomics of lactic acid bacteria. *J. Bacteriol.* **189**, 1199–1208. 10.1128/JB.01351-06 (2007).
3. Zheng, J. et al. A taxonomic note on the genus *Lactobacillus*: description of 23 novel genera, emended description of the genus *Lactobacillus* Beijerinck 1901, and union of Lactobacillaceae and Leuconostocaceae. *Int. J. Syst. Evol. Microbiol.* **70**, 2782–2858. 10.1099/ijsem.0.004107 (2020).
4. Zaunmüller, T., Eichert, M., Richter, H. & Uden, G. Variations in the energy metabolism of biotechnologically relevant heterofermentative lactic acid bacteria during growth on sugars and organic acids. *Appl. Microbiol. Biotechnol.* **72**, 421–429. 10.1007/s00253-006-0514-3 (2006).
5. Papagianni, M. Metabolic engineering of lactic acid bacteria for the production of industrially important compounds. *Comput. Struct. Biotechnol. J.* **3**, e201210003. 10.5936/csbj.201210003 (2012).
6. O'Sullivan, O. et al. Comparative genomics of lactic acid bacteria reveals a niche-specific gene set. *BMC Microbiol.* **9** 10.1186/1471-2180-9-50 (2009).
7. Buron-Moles, G. et al. Uncovering carbohydrate metabolism through a genotype-phenotype association study of 56 lactic acid bacteria genomes. *Appl. Microbiol. Biotechnol.* **103**, 3135–3152. 10.1007/s00253-019-09701-6 (2019).
8. Choi, S., Jin, G. D., Park, J., You, I. & Kim, E. B. Pan-genomics of *Lactobacillus plantarum* revealed group-specific genomic profiles without habitat association. *J. Microbiol. Biotechnol.* **28**, 1352–

1359. 10.4014/jmb.1803.03029 (2018).
9. Carpi, F. M. et al. Comprehensive pan-genome analysis of *Lactiplantibacillus plantarum* complete genomes. *J. Appl. Microbiol.* **132**, 592–604. 10.1111/jam.15199 (2022).
 10. Garcia-Gonzalez, N., Bottacini, F., van Sinderen, D., Gahan, C. G. M. & Corsetti, A. Comparative genomics of *Lactiplantibacillus plantarum*: insights into probiotic markers in strains isolated from the human gastrointestinal tract and fermented foods. *Front. Microbiol.* **13**, 854266. 10.3389/fmicb.2022.854266 (2022).
 11. Li, R. & Bi, C. Comparative genomic analysis of *Lactiplantibacillus plantarum*: insights into its genetic diversity, metabolic function, and antibiotic resistance. *Genes* **16**, 869. 10.3390/genes16080869 (2025).
 12. Teusink, B. et al. Analysis of growth of *Lactobacillus plantarum* WCFS1 on a complex medium using a genome-scale metabolic model. *J. Biol. Chem.* **281**, 40041–40048. 10.1074/jbc.M606263200 (2006).
 13. Goffin, P. et al. Understanding the physiology of *Lactobacillus plantarum* at zero growth. *Mol. Syst. Biol.* **6**, 413; (2010). 10.1038/msb.2010.67
 14. Goffin, P., Lorquet, F., Kleerebezem, M. & Hols, P. Major role of NAD-dependent lactate dehydrogenases in aerobic lactate utilization in *Lactobacillus plantarum* during early stationary phase. *J. Bacteriol.* **186**, 6661–6666. 10.1128/JB.186.19.6661-6666.2004 (2004).
 15. Lorquet, F. et al. Characterization and functional analysis of the *poxB* gene, which encodes pyruvate oxidase in *Lactobacillus plantarum*. *J. Bacteriol.* **186**, 3749–3759. 10.1128/JB.186.12.3749-3759.2004 (2004).
 16. Goffin, P. et al. Involvement of pyruvate oxidase activity and acetate production in the survival of *Lactobacillus plantarum* during the stationary phase of aerobic growth. *Appl. Environ. Microbiol.* **72**, 7933–7940. 10.1128/AEM.00659-06 (2006).
 17. Montville, T. J., Hsu, A. H. & Meyer, M. E. High-efficiency conversion of pyruvate to acetoin by *Lactobacillus plantarum* during pH-controlled and fed-batch fermentations. *Appl. Environ. Microbiol.* **53**, 1798–1802. 10.1128/AEM.53.8.1798-1802.1987 (1987).
 18. Tseng, C. P. & Montville, T. J. Enzyme activities affecting end product distribution by *Lactobacillus plantarum* in response to changes in pH and O₂. *Appl. Environ. Microbiol.* **56**, 2761–2763. 10.1128/AEM.56.9.2761-2763.1990 (1990).
 19. Tsau, J. L., Guffanti, A. A. & Montville, T. J. Conversion of pyruvate to acetoin helps to maintain pH homeostasis in *Lactobacillus plantarum*. *Appl. Environ. Microbiol.* **58**, 891–894. 10.1128/AEM.58.3.891-894.1992 (1992).
 20. Lindgren, S. E., Axelsson, L. T. & McFeeters, R. F. Anaerobic L-lactate degradation by *Lactobacillus plantarum*. *FEMS Microbiol. Lett.* **66**, 209–213. 10.1111/j.1574-6968.1990.tb03998.x (1990).
 21. Liu, S. Q., Holland, R., McJarrow, P. & Crow, V. L. Serine metabolism in *Lactobacillus plantarum*. *Int. J. Food Microbiol.* **89**, 265–273. 10.1016/S0168-1605(03)00157-0 (2003).

22. Yang, X. et al. AcrR and Rex control mannitol and sorbitol utilization through their cross-regulation of aldehyde-alcohol dehydrogenase (AdhE) in *Lactobacillus plantarum*. *Appl. Environ. Microbiol.* **85**, e02035–e02018. 10.1128/AEM.02035-18 (2019).

23. Götz, F., Sedewitz, B. & Elstner, E. F. Oxygen utilization by *Lactobacillus plantarum*. I. Oxygen consuming reactions. *Arch. Microbiol.* **125**, 209–214. 10.1007/BF00446878 (1980).

24. Quatravaux, S., Remize, F., Bryckaert, E., Colavizza, D. & Guzzo, J. Examination of *Lactobacillus plantarum* lactate metabolism side effects in relation to the modulation of aeration parameters. *J. Appl. Microbiol.* **101**, 903–912. 10.1111/j.1365-2672.2006.02955.x (2006).

25. Papadimitriou, K. et al. Stress physiology of lactic acid bacteria. *Microbiol. Mol. Biol. Rev.* **80**, 837–890 (2016). 10.1128/MMBR.00076 – 15.

26. Živković, B. L. A. B. & Genome Landscape Fermentation-associated genomic architecture in lactic acid bacteria (v1.0.0). *Zenodo* (2026). <https://doi.org/10.5281/zenodo.21900316>

27. Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y. & Ishiguro-Watanabe, M. KEGG: biological systems database as a model of the real world. *Nucleic Acids Res.* **53**. D672–D677. 10.1093/nar/gkae909 (2025).

Figures

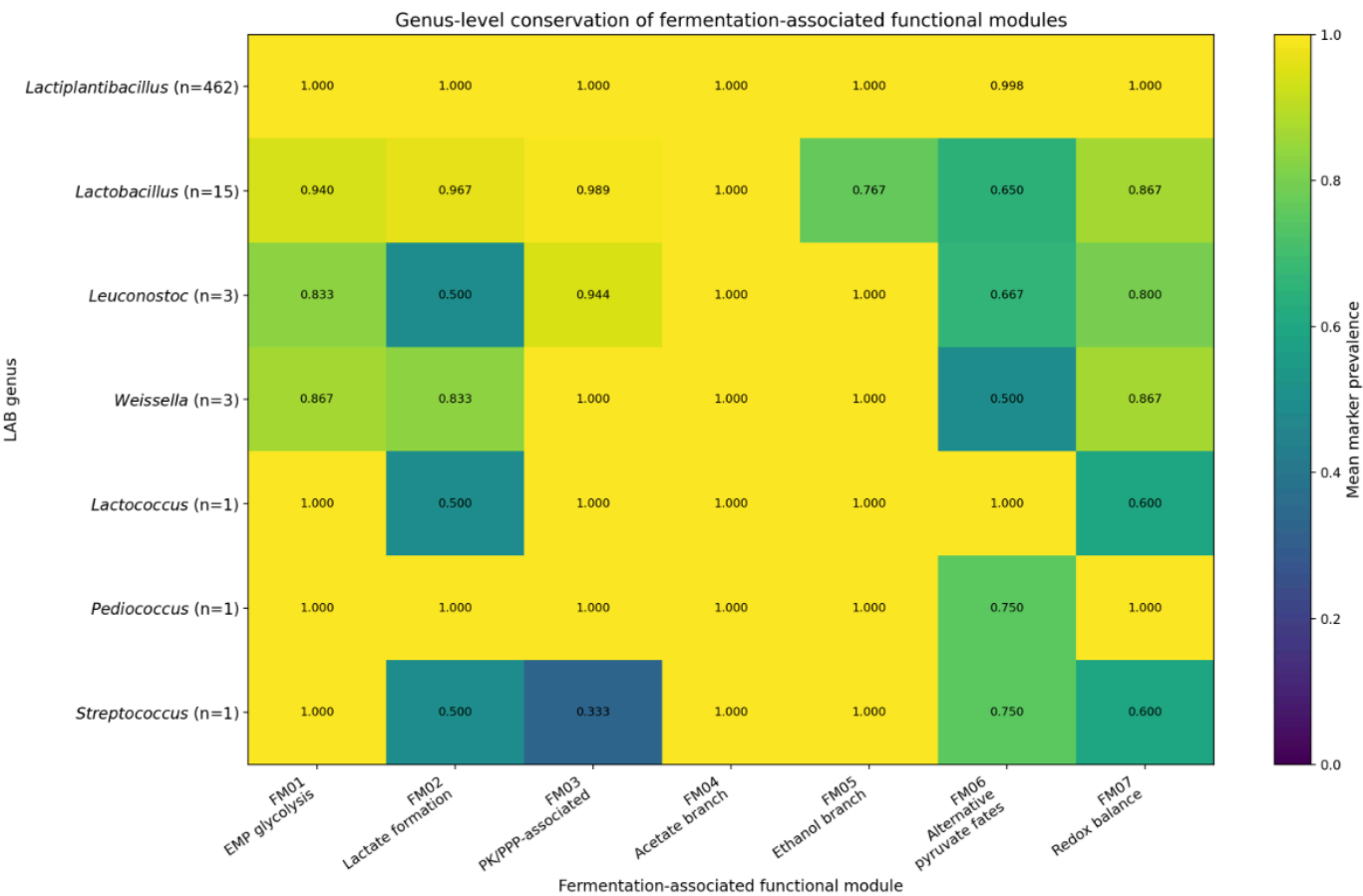


Figure 1

Genus-level conservation of fermentation-associated functional modules across the curated core-LAB collection. The heatmap shows mean marker prevalence for seven fermentation-associated functional modules (FM01-FM07) across sampled LAB genera. Values range from 0 (no marker detection) to 1 (complete marker representation). Genome counts are shown for each genus. Singleton genera are included for descriptive completeness but were excluded from the principal inter-genus variability analysis.